Field testing of pre-Enumeration Areas created using semi-automated delineation approach, Democratic Republic of Congo.

WorldPop, University of Southampton
15/03/2023

Contents

Executive Summary (English)

This report details the main outcomes of the field testing of pre-Enumeration Areas (EAs) created from WorldPop semi-automated EA approach that took place across three test sites in the provinces of Kinshasa and Kongo-Central, Democratic Republic of the Congo in December 2019. The field testing was conducted over four days by the BCR technical staff with participation from UNFPA and WorldPop staff.

Generally, EA boundaries from one census will form the basis for the EAs in the next census, with updates needed to account for new settlements and changes in population density. However, in countries where there hasn't been a census for many years, often due to conflict or insecurity, EA boundaries can be incomplete, outdated, or missing altogether. The delineation of EAs is, therefore, a crucial pre-census activity but can often be particularly challenging and highly resource intensive. Creating EAs requires consideration of population and area size within each unit to ensure that they have approximately equal-sized populations and are a manageable size to be covered by census enumeration staff. To respond to this challenge, WorldPop has developed a semi-automatic approach of delineating pre-EAs to support census cartography. This approach utilises high-resolution gridded population estimates and digitised geographic features, including administrative boundaries, and natural and man-made features, such as rivers and roads, to divide the regions into small areas which are then merged to meet criteria specified for population size and geographic area.

The last census in DRC was conducted in 1984; consequently, a recent, national, digital EA dataset which can be used for cartography planning does not exist. GRID3 is supporting the realisation of a fully digital 2020 round census in the DRC and is working closely with the National Institute of Statistics and the DRC Census Bureau (Bureau Central de Recensement, BCR) to provide technical guidance regarding options for incorporating geospatial methodologies into census planning and census cartography. As the DRC Census Bureau prepares for the 2nd National Population and Housing Census (RGPH2), a new dataset of EA boundaries is needed. As part of GRID3's work with the BCR, a field test was conducted to assess the feasibility of using a semi-automated approach for the delineation of pre-EA boundaries.

A preliminary pre-EA dataset was produced for the three test sites (Site 1: Quartier Kingu, Kinshasa (urban), Site 2: Quartier Dumi, Kinshasa (sub-urban), Site 3: Secteur Kasangulu, Kongo-Central (rural)) that span both rural and urban contexts. The geographic area covered by the three sites totalled 1,190 km^2 and was sub-divided into approximately 312 pre-EAs. The pre-EAs created for the three test sites were classified as classes 1-3 depending on the degree to which the pre-EA boundaries followed visible features (e.g. roads). Class 1 being those pre-EAs with boundaries which fully followed visible features, class 2 boundaries

followed visible features in part, and class 3 which didn't follow visible features at all. A visual assessment was carried out by comparing the pre-EA boundaries with recent high-resolution satellite imagery. A subset of the pre-EAs (15 pre-EAs), covering classes 1, 2 and 3 were selected, and assessed in the field to check how the boundaries related to ground features and their feasibility as units for population enumeration. Class 1 pre-EAs were only found in urban contexts and tended to be bounded fully by roads, which were found to be simple for the field teams to follow. In class 2 and class 3 pre-EAs, the field teams were generally able to follow roads or tracks throughout the pre-EA to reach settlements, and ascertain when they had reached the boundary of the pre-EA using the maps and GPS location indicator on the tablets. The pre-EA boundaries were also created to avoid splitting settlements and therefore even in rural areas, the field teams were able to know where housing units needed to be enumerated.

A range of limitations with this work have been identified, both with the methods and equipment used in the field data collection and the methods and input data used to produce the pre-EA boundaries. Despite the identified limitations and the challenges encountered in the field, the findings from the field test were generally consistent, with the pre-EAs created by the semi-automated approach found to be suitable for population enumeration in the field. Overall the fieldwork was successfully conducted and expectations were met and even exceeded: the BCR found that the pre-EA outputs were found to help facilitate enumeration, as the BCR team could navigate within the pre-EA boundaries and know which housing units to enumerate. The findings of the field test indicate this semi-automated approach to creating pre-EAs has the potential to be used by the BCR to create pre-EAs in preparation for census cartography, and offers large savings in terms of time, labour and cost. Nonetheless, it would be expected that the pre-EA outputs created in the approach are carefully reviewed in the lab, and manually edited as needed prior to census cartography. Then whilst in the field, the pre-EA boundaries should be validated. Limitations associated with input datasets can be addressed through a comprehensive review of existing datasets, incorporating newly available feature extraction datasets as appropriate. Further development of the approach and potential solutions and suggestions to overcome the identified limitations are outlined and discussed in detail in the report.

We expect the findings of the field test in DRC to be transferable to other similar contexts, with the approach having applicability in countries with no recent digital EAs. We also expect the approach could be adapted to update digital EA boundaries in contexts with outdated EA datasets, but this should be explored through further research and testing in such contexts.

Worth noting that in close collaboration with GeoData at the University of Southampton, UNFPA and multiple national statistical offices around the world, WorldPop has now

converted the automatic delineation script to a user-friendly tool which require minimal GIS skill to run.

La note de synthèse (en francais)

Ce rapport détaille les résultats principaux des tests de vérifications sur le terrain des Zones de Dénombrement (ZD) pré-censitaires créés à partir de l'approche de délimitation semi-automatisée développée par WorldPop. Ces tests ont eu lieu sur trois sites pilotes dans les provinces de Kinshasa et Kongo-Central, en République Démocratique du Congo (RDC), en décembre 2019, et ont été effectués par le personnel technique du Bureau Central de Recensement (BCR) avec la participation du personnel du FNUAP et de WorldPop pendant une période de quatre jours.

Généralement, les limites des ZD d'un recensement forment la base des ZD du prochain recensement; des mises à jour sont nécessaires pour tenir compte des nouveaux îlots d'habitation et des changements au niveau de la densité de la population. Toutefois, dans les pays où il n'y a pas eu de recensement pendant plusieurs années, souvent dû aux conflits ou à l'insécurité, les limites des ZD peuvent être incomplètes, obsolètes, ou totalement absentes. La délimitation des ZD est donc une activité pré-censitaire cruciale, mais elle peut souvent être difficile et peut nécessiter beaucoup de ressources. Quand on crée des ZD, on doit tenir compte de la taille de la population et de sa superficie des zones afin de s'assurer qu'elles ont des populations à peu près égales et qu'elles sont d'une taille gérable par le personnel du recensement. Pour répondre à ce défi, WorldPop ont développé une approche semi-automatique de cartographie des ZD pré-censitaires afin de soutenir la cartographie censitaire. Cette approche utilise des estimations de population carroyées à haute résolution et des caractéristiques géographiques numérisées, y compris les limites administratives et les caractéristiques naturelles et artificielles, telles que les rivières et les routes, pour diviser les régions en petites zones qui sont ensuite fusionnées pour répondre aux critères spécifiés en termes de la taille de la population et de superficie de la zone géographique.

Le dernier recensement en RDC a été effectué en 1984; par conséquent, il n'y a pas de jeu de données de ZD numérique national et récent pouvant être utilisé pour la planification de la cartographie. GRID3 soutient la réalisation d'un recensement entièrement numérique en RDC en 2021, et travaille en étroite collaboration avec l'Institut National de la Statistique et le bureau de recensement de la RDC (Bureau Central de Recensement, BCR) afin de fournir des conseils techniques sur les options pour incorporer des méthodologies géospatiales dans la planification du recensement et de sa cartographie. Le bureau de recensement de la RDC se préparant pour le 2ème recensement national de la population et du logement (RGPH2), il est nécessaire d'avoir un nouvel ensemble de données sur les limites des ZD. Dans le cadre du travail de GRID3 avec le BCR, un essai sur le terrain a été effectué pour évaluer la faisabilité d'utiliser une approche semi-automatisée pour la délimitation des limites des ZD pré-censitaires.

Un jeu de données préliminaire des ZD pré-censitaires a été produit pour les trois sites pilotes (Site 1: Quartier Kingu, Kinshasa (urbain), Site 2: Quartier Dumi, Kinshasa (banlieue), Site 3: Secteur Kasangulu, Kongo-Central (rural)) qui couvrent à la fois les contextes ruraux et urbains. La zone géographique couverte par les trois sites correspond à 1,190 km^2 et était divisée en environ 312 ZD pré-censitaires. Les ZD pré-censitaires créées pour les trois sites tests ont été classées en classes 1 à 3 selon le degré auquel leurs limites suivaient les caractéristiques visibles, telles que les routes. La classe 1 était constituée des ZD pre-censitaires avec des limites qui suivaient entièrement les caractéristiques visibles, les limites de la classe 2 suivaient en partie les caractéristiques visibles, et la classe 3 ne suivait pas du tout les caractéristiques visibles. Une évaluation visuelle a été effectuée en comparant les limites des ZD pré-censitaires avec des images satellite haute résolution récentes. Un sous-ensemble des ZD pré-censitaires (15), couvrant les classes 1, 2 et 3, a été sélectionné et évalué sur le terrain pour vérifier comment les limites correspondaient aux caractéristiques du terrain et s'assurer de leur faisabilité comme unités de dénombrement de la population. Les ZD pré-censitaires de la classe 1 ne se trouvaient que dans des contextes urbains et avaient tendance à être entièrement délimitées par des routes qui se sont avérées simples à suivre par les équipes sur le terrain. Dans les ZD pré-censitaires des classes 2 et 3, les équipes de terrain étaient généralement capables de suivre les routes ou les pistes tout au long des ZD pré-censitaires pour atteindre les localités, et de déterminer quand elles avaient atteint les limites des ZD pré-censitaires en utilisant les cartes et l'indicateur de position GPS sur les tablettes. Les limites des ZD pré-censitaires ont également été créées pour éviter de diviser les îlots d'habitations, et par conséquent même dans les zones rurales, les équipes sur le terrain ont pu savoir où les unités d'habitation devaient être dénombrées.

Un certain nombre de limites à ce travail a été identifié, à la fois avec les méthodes et le matériels utilisés dans la collecte de données sur le terrain et les méthodes et les données d'entrée utilisées pour produire les limites des ZD pré-censitaires. Malgré les limites identifiées et les défis rencontrés sur le terrain, les résultats du test sur terrain étaient généralement cohérents, les ZD pré-censitaires créées par l'approche semi-automatisée se sont révélées appropriées pour le dénombrement de la population sur le terrain. Dans l'ensemble, le travail sur le terrain a été mené avec succès, le test a répondu aux attentes et les a même surpassées: le BCR a constaté que les résultats des ZD pré-censitaires ont facilité le dénombrement, car l'équipe du BCR pouvait naviguer dans les limites des ZD pré-censitaires et savoir les unités d'habitation à dénombrer. Les résultats du test sur le terrain indiquent que cette approche semi-automatisée de création des ZD pré-censitaires a le potentiel d'être utilisée par le BCR pour créer des ZD pré-censitaires dans la préparation de la cartographie du recensement, et permettrait des économies considérables en termes de temps, de travail et de coût. Néanmoins, on doit s'attendre à ce que les résultats des ZD pré-censitaires créés dans l'approche soient examinés minutieusement en laboratoire, et modifiés manuellement si nécessaire avant la cartographie du recensement. Ensuite, les limites des ZD pré-censitaires

doivent être validées sur le terrain. Les limitations associées aux jeux de données d'entrée peuvent être résolues par un examen complet des jeux de données existants, en incorporant les jeux de données d'extraction des caractéristiques nouvellement disponibles, le cas échéant. Le développement ultérieure de l'approche et les solutions et suggestions potentielles pour surmonter les limites identifiées sont décrites et discutées en détail dans ce rapport.

Nous nous attendons à ce que les résultats du test sur le terrain en RDC soient transférables à d'autres contextes similaires, l'approche pouvant être appliquée dans des pays sans ZD numériques récentes. Nous nous attendons également à ce que l'approche puisse être adaptée pour mettre à jour les limites numériques des ZD dans des contextes avec des ensembles de données de ZD obsolètes, mais cela devrait être exploré par de nouvelles recherches et des tests dans de tels contextes.

Il convient de noter qu'en étroite collaboration avec GeoData de l'Université de Southampton, l'UNFPA et plusieurs bureaux nationaux de statistique à travers le monde, WorldPop a maintenant converti le script de délimitation automatique en un outil convivial qui nécessite une compétence SIG minimale pour fonctionner.

Introduction and background

GRID3 is supporting the realisation of a fully digital 2020 round census in the DRC. The last census in DRC was conducted in 1984; consequently, a recent, national, digital Enumeration Area (EA) dataset which can be used for cartography planning does not exist. As the DRC Census Bureau prepares for the 2nd National Population and Housing Census, a new dataset of EA boundaries is needed. 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.

Enumeration Areas are the operational geographic unit for the collection of census data and are defined early in the census process by the national statistical agency. The set of all EAs of a country constitutes a partition of that country, with EAs not overlapping with each other and nest to the administrative boundaries. EAs are often used for dissemination and analysis of census data or used as a national sampling frame for various types of surveys. However, in some countries, EAs may not have been created or remain incomplete or are outdated or inaccessible.

Creating and updating EAs is one of the most essential but challenging tasks in the preparation of a national census. Commonly, this is done by digitising small geographic units on high-resolution satellite imagery or by physically walking along the boundaries of EAs. Both techniques are highly time, cost and labour intensive. In addition, the task of creating EAs requires taking into account population and area size within each unit since there are minimum and maximum population and area constraints for each EA. As such, before delineation of EA boundaries, the number of persons living in an area and their geographic distribution needs to be estimated. This is an optimisation problem that a computer could provide big support.

To respond to this challenge, WorldPop at the University of Southampton is developing a semi-automatic designation tool of delineating pre-EAs 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 uses publicly available natural and administrative boundaries. The approach can be used to create pre-EAs 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 validated on the ground during the cartographic exercise and the household listing prior to the actual census unless insecurity restricts access. Hence, 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.

GRID3 engagement in the Democratic Republic of the Congo through the GRID3 partner organisations (UNFPA, WorldPop, Flowminder and CIESIN) includes working closely with the National Institute of Statistics and the DRC Census Bureau to provide technical guidance regarding options for the best ways to incorporate geospatial methodologies into census planning and census cartography implementation. As part of this project, the feasibility of using a semi-automated approach for pre-EA delineation, developed by WorldPop and Flowminder was investigated. During a previous meeting in June 2019, a light field assessment of the approach was conducted in two wards of the Town Kinshasa (Lingwala and Gombe). The results were satisfactory but we needed a more comprehensive test both in terms of the diversity of terrains and the number of features tested. To this end, three sites, reflecting the variety of situations that can be encountered on the field, were selected by BCR to generate pre-EAs and conduct the field assessment (Figure 1).

- Site 1: Quartier KINGU, Commune de Selembao, Kinshasa (Urban)
- Site 2: Quartier DUMI, Commune de Maluku, Kinshasa (Sub-urban)
- Site 3: Secteur KASANGULU, Territoire Kasangulu, Kongo-Central (Rural)

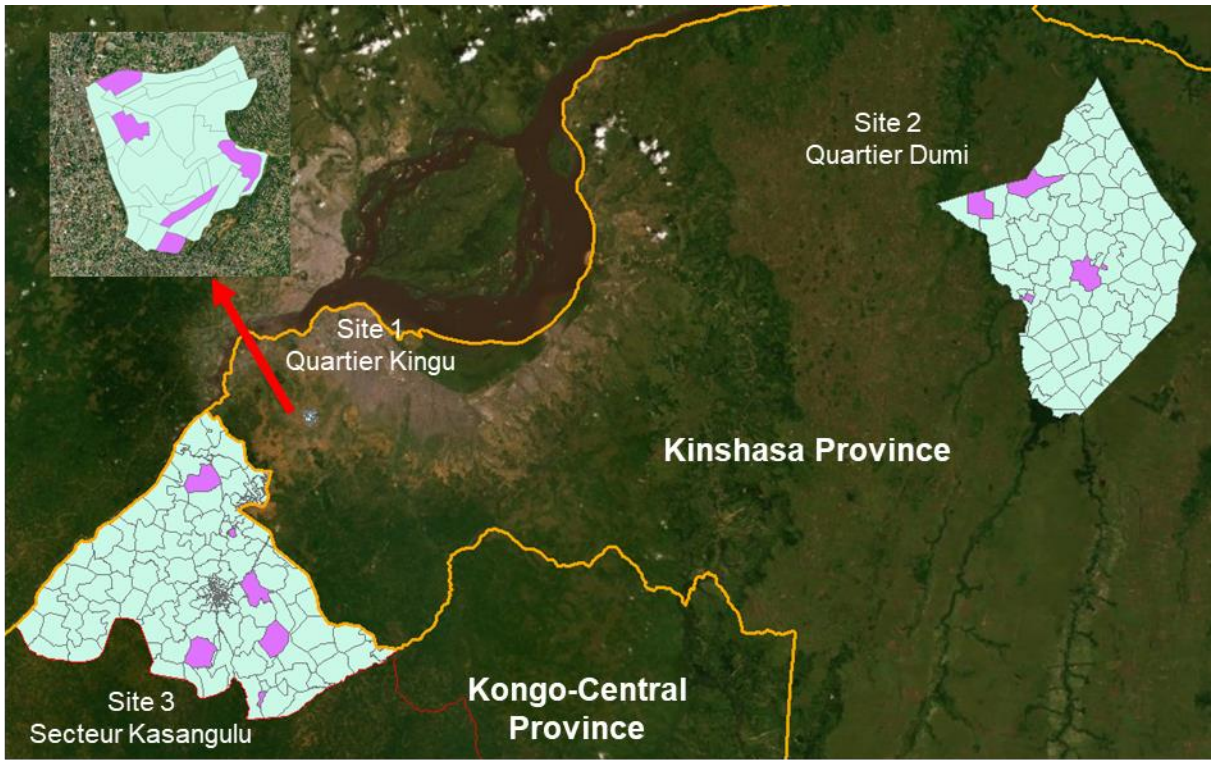

Figure 1. Locations of the sites in DRC where the fieldwork was conducted.

Overall purpose: The field testing of pre-EA outputs took place over four days in Kinshasa and Kongo-Central provinces, DRC in December 2019. The field testing was conducted by the BCR technical staff with participation from UNFPA and WorldPop staff. This field test was intended to provide an opportunity to see if the outputs from the semi-automated EA approach would be useful for the DRC BCR in preparing for a census, as well as provide insights to WorldPop and UNFPA as to the advantages, limitations and areas for further development in the approach.

Specific objectives:

1. For BCR colleagues to develop familiarity with the approach to the creation of pre-EA units, specifically focusing on the outputs and methodology. This also provides an opportunity for UNFPA colleagues to further develop their knowledge and understanding.
2. To undertake a visual assessment through comparison with recent high-resolution satellite imagery for all the pre-EAs across the three selected sites in DRC (in total, the area covered by the three sites was sub-divided into 312 pre-EAs).
3. To undertake, for a subset of pre-EAs, an assessment exercise in the field to gather information on the appropriateness of boundaries in relation to ground features and the size of the pre-EA with regards to the population within a unit.
4. To assess if the generated pre-EAs meet the BCR's criteria, including specified population thresholds, pre-EA boundaries nesting within administrative boundaries and respecting obstacles such as rivers.
5. To understand what skills, steps and considerations are required for the BCR to make use of this methodology more widely, and its applicability at the national scale.
6. To estimate the potential value in applying the approach to other countries' contexts.

Field test methods and data collection

Laboratory Phase

Creation of pre-EA boundaries

Data Inputs

The main data inputs for this approach include GRID3 (version 1.0) high-resolution gridded population estimates (Boo et al. 2021) at approximately 100m spatial resolution, produced for five western provinces (Kinshasa, Kongo Central, Kwilu, Kwango and Mai-Ndombe), digitized geographic features including administrative boundaries (provided by the BCR) and natural and man-made geographic features such as roads, waterways and railways (extracted from OpenStreetMap).

Methodology

Step 1: Splitting

The aim of the splitting process is to partition the Area Of Interests (AOIs) into regions that are as small as possible so that the subsequent merging process has enough flexibility to combine them into optimal pre-EAs. Linear features (both natural and man-made) such as roads, railways and waterways, from existing digitised geospatial sources such as OpenStreetMap, are combined with existing administrative boundaries to subdivide regions into small units. The splitting process has not been incorporated into the current Python script yet, therefore, several ArcGIS tools and techniques were used to conduct the splitting process. The estimated population for each small unit is then computed from resampled gridded population data (approximately 10m spatial resolution) using Zonal Statistics. In the future development of the approach, all the splitting steps and computing population for the split unit will be automated.

In the current field test sites, approximately 2,700 units were created during the splitting process in which 101 units were in Kingu, 650 units in Dumi and 1,944 units in Kasangulu.

Step 2: Merging

When the estimated population size and area have been calculated for all small units created in the splitting process, the units are then merged until they reach a threshold (specified in terms of estimated total population and unit area). This merging step is implemented using a Python script. The maximum population threshold was set to 1,200 people in urban areas, 1,000 people in rural areas and the upper size limit of an area to 9km^2. The script uses the population, area and compactness constraints to merge the small regions. In addition to that, we have produced another version where the population constraints were set the same, but area constraint was enlarged to 20 km^2. Both datasets

were shared with BCR and after an appropriate investigation by the BCR team, the first version was preferred and used in the field test as it was decided that the 9 km^2 area constraint would be easier to manage in the field.

After applying the merging technique to the splitting units, 312 pre-EAs were created for the three sites in which 30 pre-EAs were in Kingu, 79 pre-EAs in Dumi and 203 in Kasangulu.

For more detailed information about methodology and data preparation, please see this report.

Classification of pre-EAs

The pre-EA units created for the field test were classified into three types:

- Class 1: Pre-EAs with boundaries that fully follow features visible on the ground e.g. roads
- Class 2: Pre-EAs with boundaries that that in part follow features such as roads
- Class 3: Pre-EAs with boundaries that do not follow visible features (due to lack of features in existing input datasets)

Field phase

Once the pre-EAs has been created and checked in the lab, the next step is to verify on the field that they actually make sense and respect all the specified constraints. For this exercise, the pre-EAs selected for population enumeration and boundary checking in the field included EAs from each of the three classes. In total, 12 pre-EAs were enumerated by the BCR across the three sites. In addition, boundary checking was undertaken for 15 pre-EAs at the three sites. The number of pre-EAs checked and enumerated at each site is summarised in table 1.

Table 1: The number of pre-EAs where boundaries were checked, and population enumerated at each site (figure 2) with their classes.

	Boundary checks			Population enumeration		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Site 1 (Kingu)	3	2		1	1	

Site 2 (Dumi)		2	2		2	2
Site 3 (Kasangulu)		4	2		4	2
Total	**3**	**8**	**4**	**1**	**7**	**4**

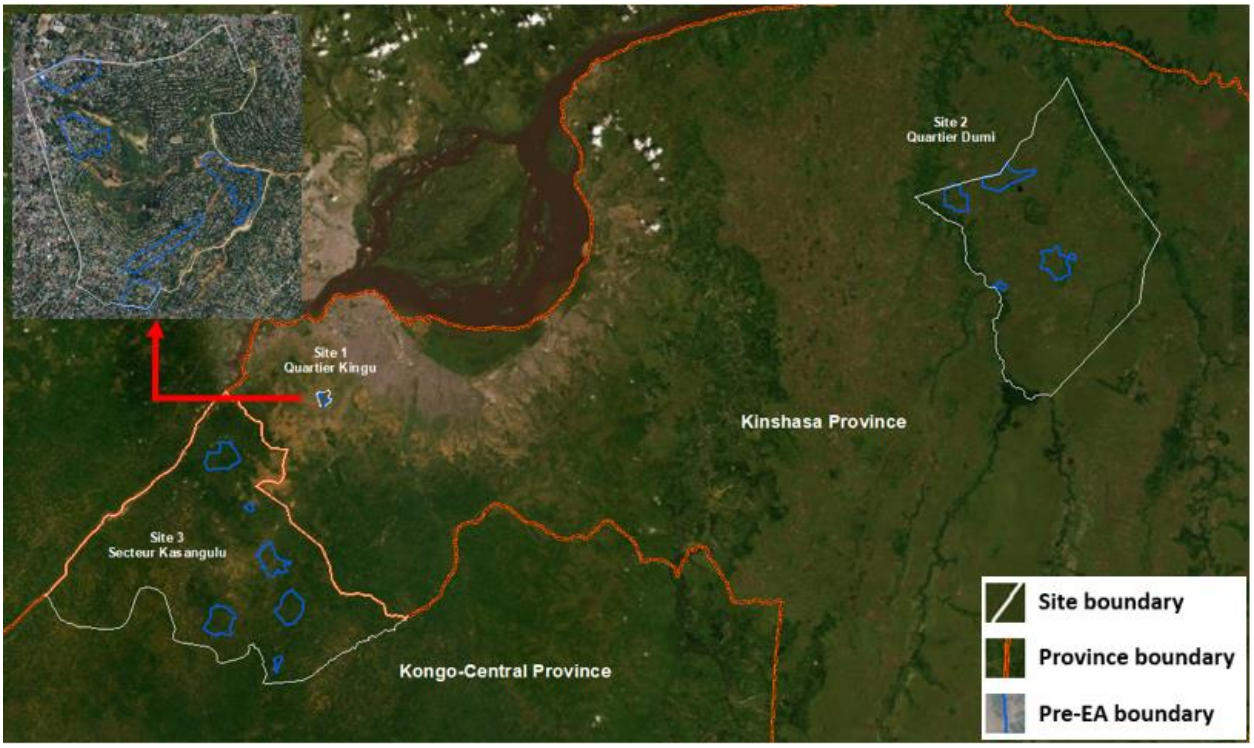

Figure 2. Shows the location of the selected pre-EAs across the three sites in Kinshasa and Kongo-Central, DRC.

Field protocols

The field teams had tablets and handheld GPS devices, supplemented by paper maps and forms as necessary. The tablets were Fujitsu Stylistic M532 running v4.0/1 Android, owned by the BCR, withCSPro (version 7.3) software for collecting household enumeration data and SW Maps application was used for displaying maps, navigating and collecting point data. The handheld GPS units used to record continuous tracks were Garmin eTrex 10, loaned from the University of Southampton.

Protocol for boundary checking

The aim of checking the pre-EAs in the field was to ensure that the boundaries form a unit within which the teams can visit all buildings and ultimately fully enumerate the population. The three classes of pre-EAs were considered in conducting checks on the boundaries in the field as it was necessary to vary the approach to checking the boundaries in each class. For each pre-EA, irrespective of the class, the handheld GPS units were used to collect continuous tracks of the movements taken by field teams to understand their route in relation to the pre-EA boundaries. For each pre-EA, the following data were collected: (1) a waypoint for the start location (point), (2) the route taken by the cartographers in checking the pre-EA boundaries – this was recorded as a GPS track (line) and (3) a waypoint for the end location (point). The SW Maps application was used to record features and comments which needed to be recorded in the process of checking boundaries, for example obstacles, deviations from a boundary or buildings that were crossed by a boundary. The data from the handheld GPS units and SW Maps application were transferred on to the supervisor's laptop at the end of each day.

Class 1 pre-EA boundaries are those that fully follow features visible on the ground such as roads or rivers. An accessible starting point was identified, navigated to and recorded. On the handheld GPS unit, a new track was created. To check these boundaries, the field staff tried to follow the boundary as closely as possible, based on the route that an enumeration team would be likely to take. If field staff encountered an obstacle or situation which meant that they could not continue following the pre-EA boundary, a waypoint was recorded. Once the field staff had followed the boundary to completion/as far as was possible, they returned to the start location, recorded another waypoint and stopped and saved the track created.

Class 3 pre-EAs are those that do not follow visible features, due to a lack of existing data on features in these areas. Class 3 pre-EAs typically were in more rural locations where there may be few roads or where existing data coverage of features may be poor. Due to the lack of digitised visible features in these areas, Voronoi polygons were created around settlements to subdivide areas into small polygons, whilst ensuring that boundaries avoided crossing through small settlements as far as possible. This meant that class 3 pre-EA boundaries were typically composed of a series of straight lines, often passing through rural areas of dense vegetation. To check the boundaries in these areas, the field teams identified a road for access into the pre-EA and started at the point where a road intersected with the pre-EA boundary (as shown by the live GPS location on the tablets displaying the boundary overlaid on satellite imagery. The field staff would then travel along the road, recording any buildings, settlements or other signs of habitation, repeating this process for any adjoining roads, tracks or paths and throughout the remainder of the

pre-EA, continuing up to the edge of the pre-EA where any roads, tracks or paths intersect the boundary. This process was intended to check that the pre-EA boundaries were suitable for enumeration teams to be able to fully enumerate the population within the pre-EA.

Class 2 pre-EAs, which are those that in part follow features such as roads, i.e. were a mix between class 1 and class 3. Therefore, an approach adapted from those outlined above for class 1 and class 3 was used, with boundaries followed as far as possible where they reflected visible features (like for class 1), and roads, tracks and paths within the pre-EA units followed throughout to ensure all settlements or inhabited locations were visited and recorded.

Protocol for household enumeration

Household enumeration was conducted solely by BCR staff. The BCR has prepared different instructions for its field team considering their responsibility on the ground to enumerate the household.

Results

The field test of the pre-EA outputs was successfully conducted with data collected to check boundaries and enumerate households across the three field sites. The data collected included GPS locations for households, GPS tracks for the routes taken by the field team on the ground and points of interest where the field team took note of a specific location. Recorded locations included the start and endpoint of the route taken on the ground, settlements (in rural areas), specific issues requiring deviation from the boundary or where the boundaries cut through settlements or individual buildings.

It would not be expected for the GPS tracks of the field team's movement to exactly match the pre-EA boundaries, due to the boundaries only following roads or tracks in some areas. Rather the GPS tracks were used to provide an indication of which parts of the pre-EA boundaries and settlements were checked. As can be seen in the following figures, the field team might have taken different ways and paths to check all sides of the pre-EAs. In some cases, the field team also had to leave and re-enter the pre-EAs several times to make sure that all the necessary sides of the pre-EAs were checked, report if any houses were cut by the pre-EA boundaries and ensure that all the households within the pre-EA are accessible and can be enumerated with no barriers. The results on validating individual

pre-EAs in the field and the feedback from the field team with necessary suggestions for each site are summarised in the following section.

Site 1: Quartier KINGU, Commune de Selembao, Kinshasa (Urban)

The first site was located in the south-west of Kinshasa urban area with a total area of approximately 3 km^2. The generated pre-EAs were all class 1 and 2 as their boundaries fully or partially followed clear roads and paths on the ground. The site had steep terrain, with unstable, sandy soil. In some places, houses were only accessible by small paths. The field teams encountered some practical issues due to difficult terrain and lack of permission to collect data in some pre-EAs. There were difficulties in checking some sections of the pre-EA boundaries due to the nature of the location with steep slopes, erosion and cliffs present in some areas.

In total, data were collected for five pre-EAs at this site: households in two pre-EAs were fully enumerated and boundaries were validated for all five pre-EAs. The results on validating individual pre-EAs in the field and the feedback from the field team with necessary suggestions are summarised for each pre-EA in the following section.

Figure 3. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 304 in the Kingu site.

1. The pre-EA boundary was very clear and there were no issues on the ground in finding them. In addition, the field teams did not report any issues with buildings being cut by the pre-EA boundary.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them.
3. Some buildings were either empty or non-residential, therefore, their GPS locations were not collected.
4. Few GPS household locations are outside the pre-EA boundary, likely due to data collection issues (e.g. with poor GPS signal and the age of the tablet).
5. On the west side, the pre-EA boundary followed an unpaved road, which was blocked by a gate and the team had to back-track and to change their route to check the other side of the pre-EA.

Figure 4. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 298 in the Kingu site.

1. The team had confirmed that the pre-EA boundary was easily identifiable and recognizable by the enumerators in the field and they could easily find out when they had reached the edge of the pre-EA boundary. However, due to the nature of the location (very steep, no roads, few paths and issues with erosions), the team had to take different ways to check all sides of the pre-EA.
2. There were no issues reported by the field teams of buildings being cut by the pre-EA boundary.
3. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. Only one GPS household location seems to be outside the pre-EA, likely due to data collection issues (e.g. with poor GPS signal and the age of the tablet).
4. There appears to have been a few buildings at the far east of the pre-EA that GPS locations were not recorded by household enumerators. Similarly in the central west part of the pre-EA, however these buildings may be non-residential.

Figure 5. Shows the points of interest, GPS track and the boundary for pre-EA 283 in the Kingu site.

1. The pre-EA boundary was clear and there were no issues on the field in finding them. In addition, none of the buildings were cut by the pre-EA boundary.

2. The North-West corner was a compound with a lot of trees, so the team has shifted the starting point to the corner of the road.
3. The team was following a clear road to check the boundary until the point marked in the south where there was bad erosion and the team had to divert through a residential area. However, the team confirmed that all households can be accessible within the pre-EA.
4. The households were not enumerated for this pre-EA as they were told by the local authority, this pre-EA does not belong to the Quartier Kingu and the necessary prior authorization required for household enumeration had not been obtained.

Figure 6. Shows the points of interest, GPS track and the boundary for pre-EA 293 in the Kingu site.

1. The pre-EA boundary was clear and there were no issues on the field in finding them. However, there was a small part of the boundary where the team could not walk because of erosion.
2. None of the buildings were cut through the pre-EA boundary.
3. Household enumeration was not conducted, however the team had confirmed that they have not anticipated issues with enumerating within the pre-EA boundary.
4. This pre-EA was not part of the selected pre-EAs priority list that was supposed to be checked, but the team had added it since they had more time in the field.

Figure 7. Shows the points of interest, GPS track and the boundary for pre-EA 306 in the Kingu site.

1. The team had confirmed that the pre-EA boundary could be recognised by the enumerators in the field and they could find out when they had reached the edge of the pre-EA boundary. In addition, none of the buildings were cut through the pre-EA boundary. However, due to the nature of the place (e.g. steep slope and erosion) and lack of roads and path, the team had to take different paths to check the pre-EA boundary - in some places leaving the pre-EA and then returning in order to continue.
2. The households were not enumerated but the team confirmed that there were no anticipated issues with enumerating within the boundary.

Site 2: Quartier DUMI, Commune de Maluku, Kinshasa (Suburban)

The second site was located in the north-east of Kinshasa province with a total area of approximately 530 km^2. The generated pre-EAs were all class 2 or 3 as no pre-EA boundary followed a clear road or path on the ground in its entirety, instead the pre-EA boundaries either did not follow visible features on the ground or partially followed visible features such as roads and paths on the ground. Site 2 was rural, centred around the town of Dumi. Outside Dumi, roads were unpaved with poor accessibility and areas of dense vegetation between small settlements. Practical issues were encountered in the field surrounding difficult terrain, lack of internet and poor coverage of the mobile network. There were difficulties in checking some of the pre-EA boundaries due to the nature of the location such as bush and cliff but this did not cause issues for data collection as they were not settled.

In total, four pre-EAs were validated in the field for this site in which households were fully enumerated and boundaries were validated. The results on validating individual pre-EAs in the field and the feedback from the field team with necessary suggestions for each pre-EA are summarised below.

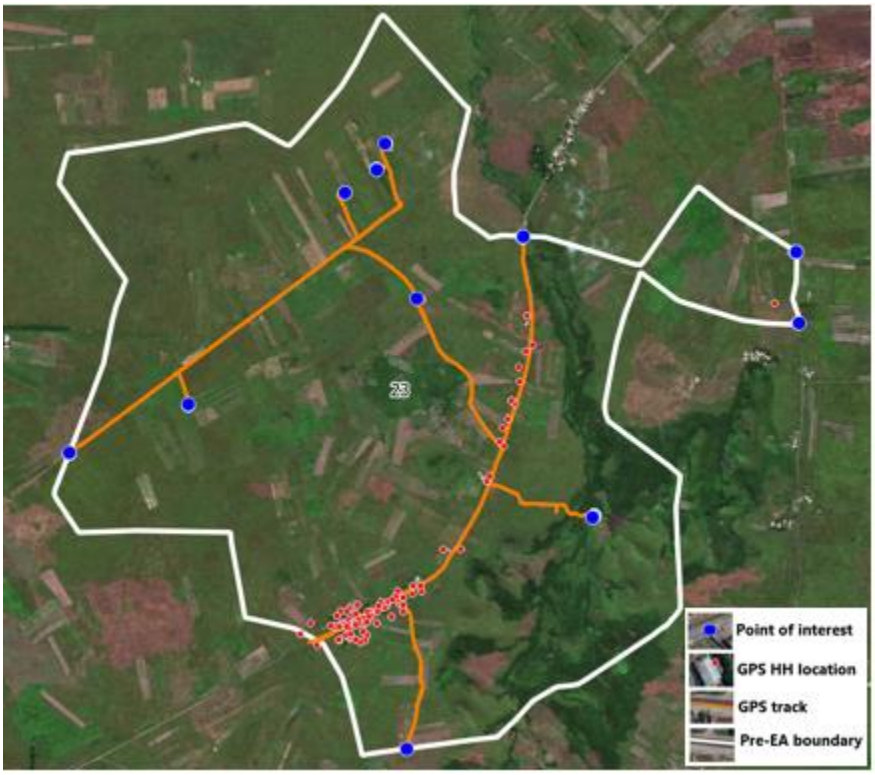

Figure 8. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 023 in the Dumi site.

1. As a class 3 pre-EA, none of the boundaries followed features such as roads. However, the extent of the pre-EA could be recognised using the GPS location, imagery and boundaries on the tablets. The enumerators in the field could easily find out when they had reached the pre-EA boundary, for example, if travelling along a road or path which intersected the boundary. In addition, none of the buildings were cut through by the pre-EA boundary.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. All the settlements and buildings were checked and a few of them were abandoned.
3. As the pre-EA was very rural, most of the boundaries were located in the bush, but this did not cause issues as teams travelled along paths and roads as described in point 1. The odd extra section on the east was successfully surveyed (although no GPS tracks of the route taken were recorded), but it would be simpler if such odd shapes will be avoided in future pre-EA creation.
4. There were some technical issues with the tablets.

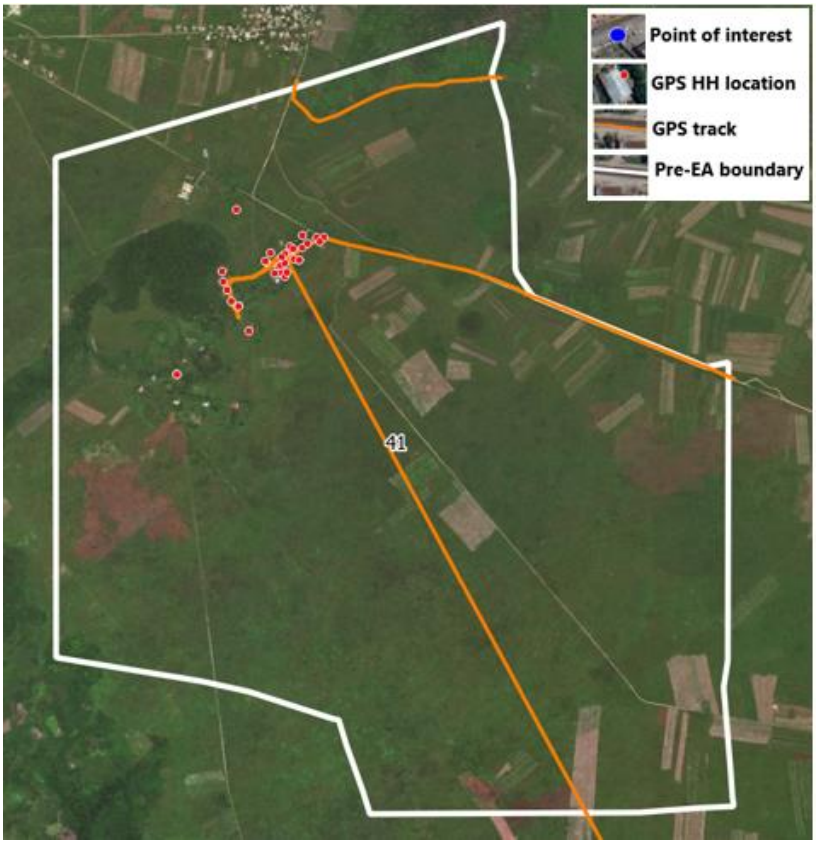

Figure 9. Shows the distribution of GPS household locations, GPS track and the boundary for pre-EA 041 in the Dumi site.

1. As a class 3 pre-EA, none of the boundaries followed features such as roads. However, the extent of the pre-EA could be recognised using the GPS location, imagery and boundaries on the tablets. The enumerators in the field could easily find out when they had reached the pre-EA boundary, for example, if travelling along a road or path which intersected the boundary. In addition, none of the buildings were cut through by the pre-EA boundary. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. However, the majority of the area is uninhabited.
2. There were technical issues with the tablets which complicated the data collection.
3. The team has recommended adding further informative layers such as settlement names, roads, etc to the MBTiles to provide context and better guidance in the field.

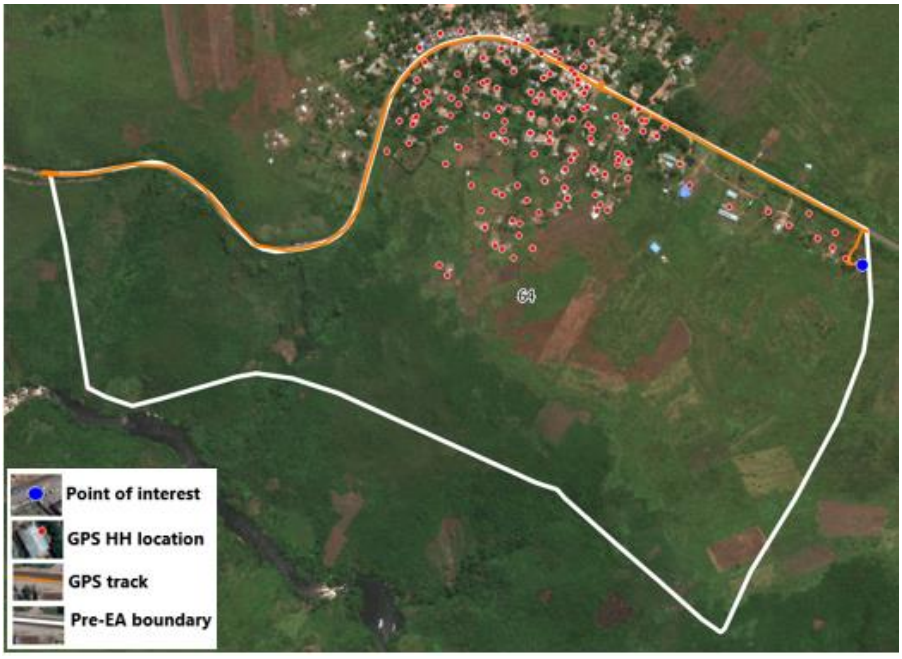

Figure 10. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 064 in the Dumi site.

1. The team had confirmed that the pre-EA boundary on the north side was easily identifiable and recognisable by the enumerators in the field. In the north, the pre-EA boundary was in-line with the main road. The remainder of the boundary did not follow visible features and was mostly in vegetated areas with forest and cliffs. Despite this, the field teams could easily find out where they had reached the edge of the pre-EA boundary, as only the northern part adjacent to the road was settled. In addition, none of the buildings were cut through the pre-EA boundary.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them.
3. There are a few GPS household locations outside the pre-EA boundary (to the north), likely due to data collection issues (e.g. with the poor GPS signal, the age of the tablet and the external GPS was not available).

Figure 11. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 076 in the Dumi site.

1. As a class 2 pre-EA, some of the boundaries followed features such as roads. However, the extent of the pre-EA could be recognised using the GPS location, imagery and boundaries on the tablets. The enumerators in the field could easily find out when they had reached the pre-EA boundary, for example, if travelling along a road or path which intersected the boundary. In addition, none of the buildings were cut through the pre-EA boundary.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. However, the majority of the area is uninhabited.
3. On the south-east side (where the most easterly POI point is located), the pre-EA boundary separated a farm compound into two. Preferably, the entire settlement should be within the same pre-EA if the maximum population threshold allows. In the future development of the approach, new ideas will be incorporated to avoid this (see discussion section).
4. The team had difficulty in navigating to the pre-EA but once they were within the pre-EA boundary, there was no issue. It has been recommended to create MBTiles for a larger area to provide the context of the surrounding area and add more informative layers to the MBTiles for better guidance.

Site 3: Secteur KASANGULU, Territoire Kasangulu, Kongo-Central (Rural)

The third site was located in the north of Kongo-Central province with a total area of approximately 660 km^2. The generated pre-EAs were class 2 or 3 as no pre-EA boundary followed a clear road or path on the ground in its entirety, instead the pre-EA boundaries either did not follow visible features on the ground or partially followed visible features such as roads and paths. Site 3 covered the full Secteur of Kasangulu, which covers the town of Kasangulu and the surrounding area.

Practical issues were also encountered in the field surrounding difficult terrain, lack of internet and poor coverage of the mobile network. There were difficulties in checking some of the pre-EAs boundaries due to the nature of the location such as bush and cliff but they did not cause issues as they were not settled. It was discovered whilst in the field that some of the selected pre-EAs are in areas affected by insecurity, with abandoned villages. In these areas, inhabitants have moved to spend the night in nearby towns, but return to the land during the day to work, for example, to tend crops. In these areas, many villages and houses were abandoned and so enumerated population counts were very low.

In total, six pre-EAs were validated in the field for this site in which households were fully enumerated in all of them and boundaries were validated for all the pre-EAs. The results

of validating individual pre-EAs in the field and the feedback from the field team with necessary suggestions for each pre-EA are summarised below.

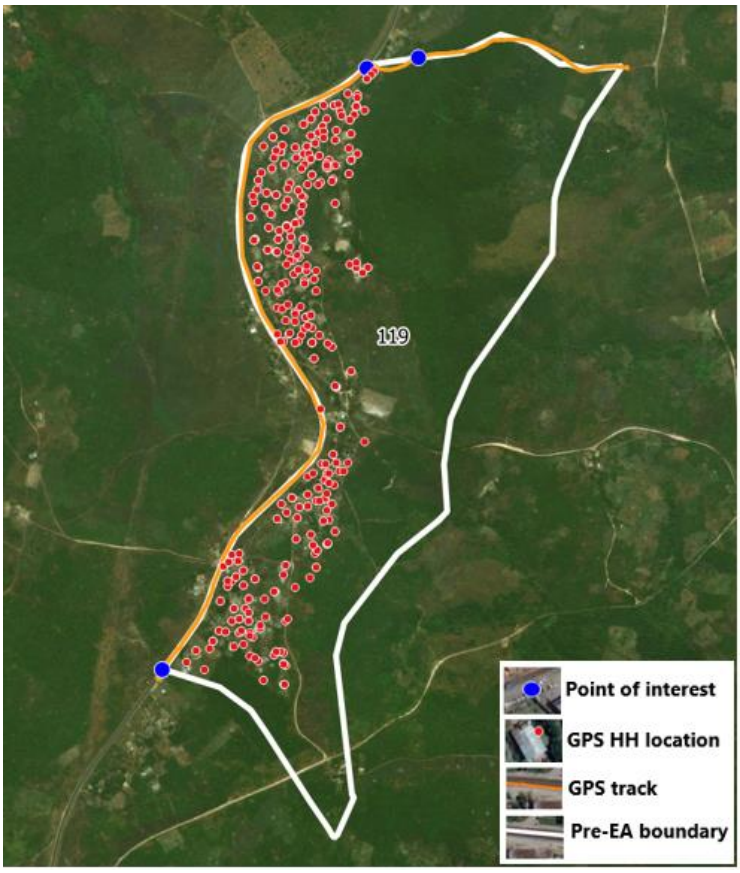

Figure 12. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 119 in the Kasangulu site.

1. The pre-EA boundary was very clear for the north and west side and there were no issues on the field in finding them since the boundary was following the main road. The area of settlement within the pre-EA was concentrated in the north and west also. The majority of eastern and southern parts of the pre-EA boundary were in bush and they could not be followed but this did not cause issues especially as the areas close to the boundaries were not inhabited.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them.
3. At the most westerly of the two POI locations in the north of figure 10, the boundary cut through two buildings. This may be due to the date of satellite imagery used by others in digitising input datasets (e.g. OSM).
4. The team had reported some technical issues on the tablets.

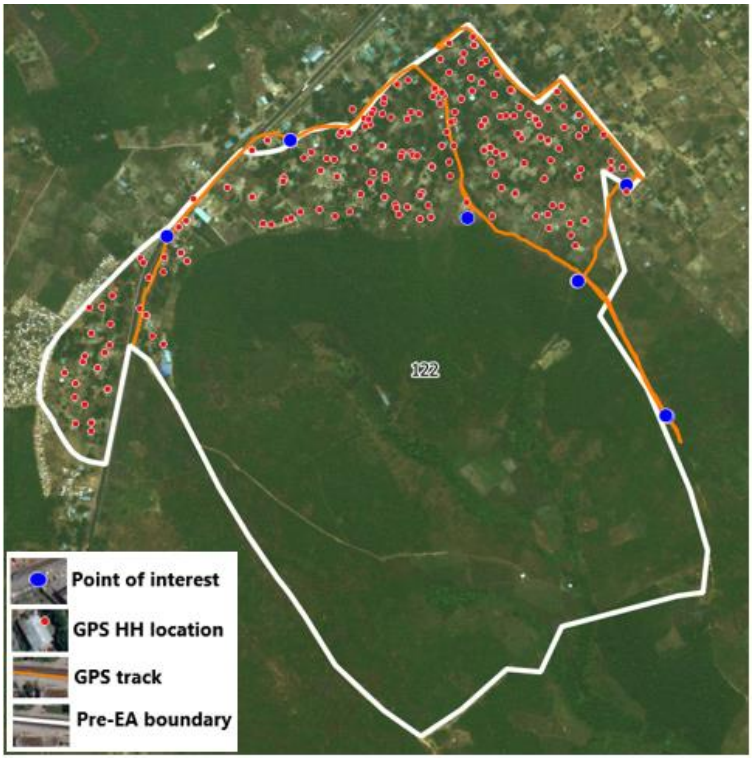

Figure 13. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 122 in the Kasangulu site.

1. As a class 2 pre-EA, some of the boundaries followed features such as roads, particularly in the northern parts of the pre-EA where the settled area was concentrated. The team had confirmed that the pre-EA boundary was easily identifiable and recognizable by the enumerators in the field and they could easily find out where they had reached the edge of the pre-EA boundary.
2. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. Although part of the pre-EA was a cemetery (in the north-west of the pre-EA) and moving around it was not easy, the team successfully enumerated all the households.
3. The pre-EA boundary cut through a small structure in the east of the pre-EA and it was included in the household enumeration.

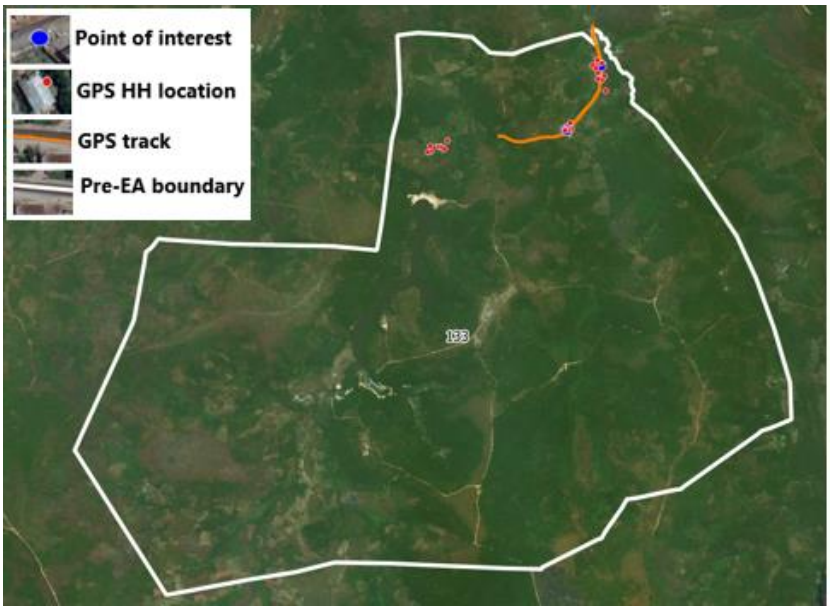

Figure 14. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 133 in the Kasangulu site.

1. Through conversation with local people, it was ascertained by the field team that the majority of the area is uninhabited due to insecurity, with the population having moved to nearby towns where they stay overnight but may return to this area during the day (e.g. for agricultural purposes).
2. The households for this pre-EA were enumerated and the enumerators had no issues in accessing them, however, given that the majority of the area is uninhabited, few data points were collected.

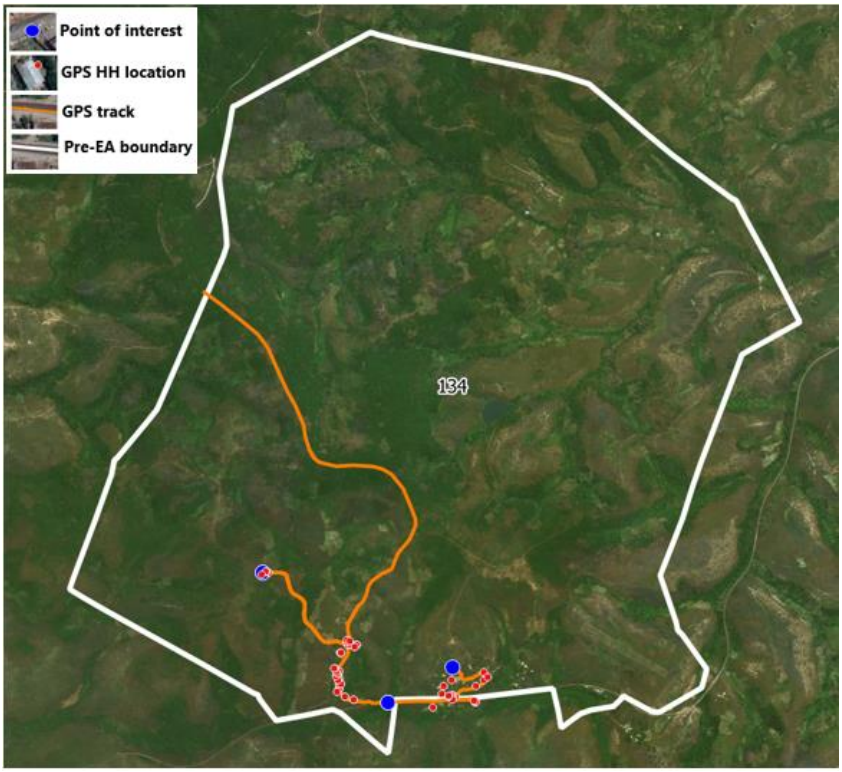

Figure 15. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 134 in the Kasangulu site.

1. Through conversation with local people, it was ascertained by the field team that parts of the area covered by this pre-EA are uninhabited due to insecurity, with the population having moved to nearby towns. The area of settlement with a population was concentrated in the south of the pre-EA.
2. As a class 2 pre-EA, some of the boundaries followed features such as roads, particularly in the southern parts of the pre-EA where the settled area was concentrated. The team was able to follow the boundary in this area and identify when the boundary had been reached, for example where the road intersected with the boundary in the west of the pre-EA.
3. The households for this pre-EA were fully enumerated and the enumerators had no issues in accessing them. The team was aware that there were a few GPS household locations that were collected outside the pre-EA boundary in the south. However, this was a request from the local people to include them when they saw their neighbors were interviewed. The field teams were aware they were enumerating households from outside pre-EA 134.
4. The pre-EA boundary separated the settlement (in the south) into two parts but preferably, the pre-EA should include the entire settlement if the maximum

population threshold allows. In the future development of the approach, new ideas will be incorporated to avoid this (see discussion section).

5. The GRID3 and local team had no prior information about the insecurity issues in this area and better communication should be made in future field visits.

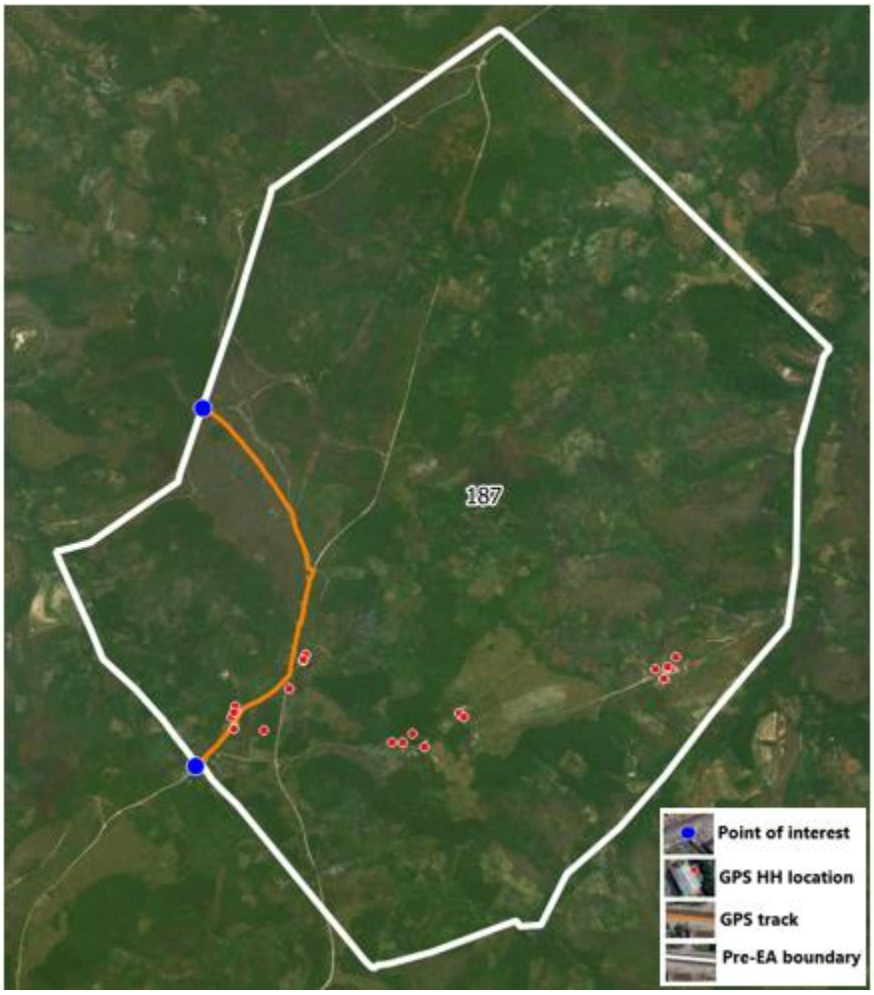

Figure 16. Shows the distribution of GPS household locations, points of interest, GPS track and the boundary for pre-EA 187 in the Kasangulu site.

1. As a class 3 pre-EA, none of the boundaries followed features such as roads. The majority of the area is uninhabited and some villages and buildings are abandoned because of insecurity issues. GPS points were still collected for buildings that were abandoned/empty.

2. The enumerator team did not have a GPS tracker when they were checking some of the settlements. So only the route travelled by the supervisor/cartographer in the west of the pre-EA was recorded.
3. The team confirmed that the pre-EA boundary could be recognised by the enumerators in the field and they could find out when they had reached the edge of the pre-EA boundary.
4. None of the buildings were cut through by the pre-EA boundary.

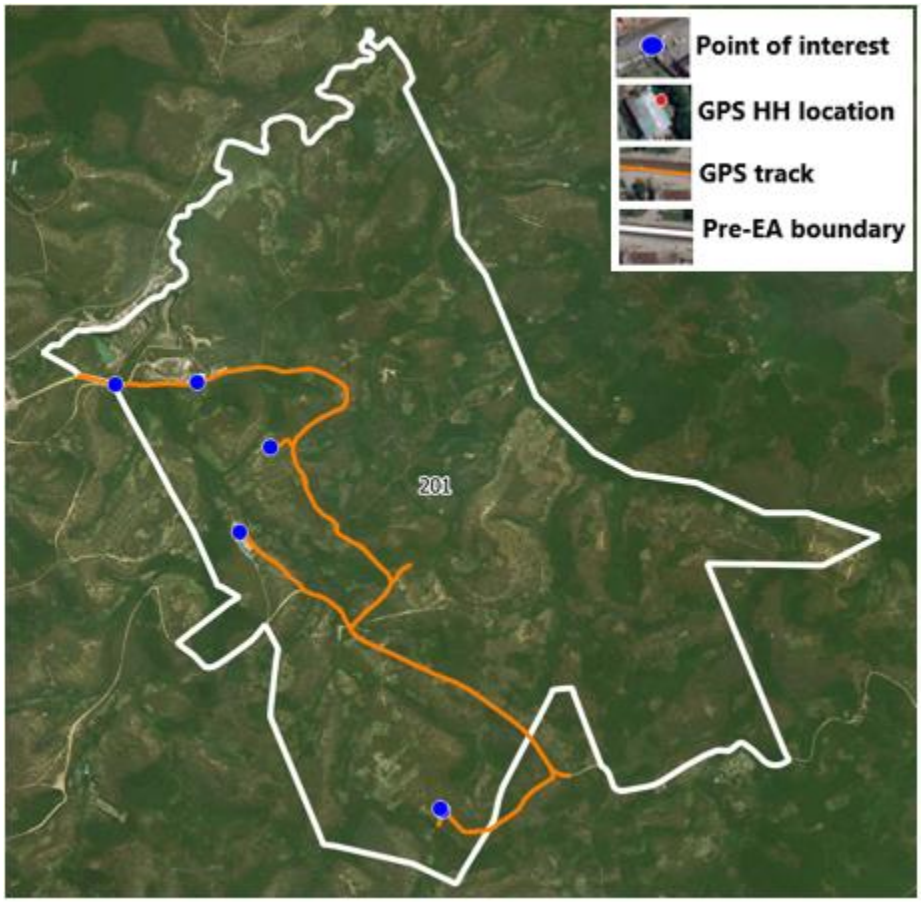

Figure 17. Shows the points of interest, GPS track and the boundary for pre-EA 201 in the Kasangulu site.

1. The team had confirmed that the pre-EA boundary could be recognised by the enumerators in the field and they could find out when they had reached the edge of the pre-EA boundary.
2. The team has checked all the settlements but it found out no one lives in this pre-EA as the buildings in the north-west of the site were associated with a mine and

in the west, with a brick-making facility. All the villages were abandoned because of insecurity issues.

The main objective of the field test was to check the pre-EA boundaries to ensure that the field staff knew where they were in relation to the boundary, and identify which households to enumerate. Alongside boundary checking, BCR staff conducted a household enumeration in selected pre-EAs with population counts recorded. The results are presented in the following section.

Household enumeration

Across the three sites, the BCR conducted household enumeration in 12 pre-EAs, the results of which are provided by the BCR and summarised in table 2 and presented alongside the estimated population, calculated from the GRID3 v1.0 modelled high-resolution gridded population estimates. Considering the small sample size of the pre-EAs that were enumerated in the field, it is not possible to conduct a comparison of these population figures in a statistically robust way, the results are nonetheless included for reference. The enumerated population totals that were collected on the ground are presented with the calculated mean, median, lower and upper (two-tailed 90% confidence interval) population estimates, based on the GRID3 v1.0 high-resolution gridded population data.

The enumerated population totals in four pre-EAs (pre-EAs 064, 076, 298 and 122) are within the range of the lower and upper population estimates (two-tailed 90% confidence interval). For the remaining pre-EAs enumerated at site 2 (Q. Dumi), the enumerated population was less than the lower population estimate (pre-EAs 023 and 041). Similarly at site 1 (Q. Kingu) the enumerated population for pre-EA 304 was less than the lower population estimate. For pre-EAs at both sites 1 and 2, it should be considered whether the population was fully enumerated during the fieldwork, with potentially some households missed or individuals not counted given that households were only visited on one occasion with no revisit. The information provided in household enumeration may also be unreliable, particularly if the head of household was not present and instead information was provided by a neighbour or others nearby. Considering the distribution of GPS household locations, it seems there could have been some households missed or populations resident in buildings considered to be non-residential, for example within the central part of pre-EA 304 (figure 3).

Greater differences were observed between enumerated and modelled population estimates for the pre-EAs at site 3 (Kasangulu), with considerably less population

enumerated within the pre-EAs of 133, 134, 187 and 201 than predicted by the modelled population estimates. This was due to housing units or whole villages being recently abandoned due to insecurity issues. Our understanding is that the population from these areas is now resident in nearby towns, suggesting considerable population movements. It should also be considered that the modelled population estimates were for 2018 whereas the pre-EAs were enumerated in December 2019. The areas in which population is estimated is determined by a geospatial settlement layer, which is a known limitation with the current version (v1.0) of the GRID3 modelled population estimates, primarily due to the age of the satellite imagery from which the settlement dataset was extracted (in some locations this can be more than five years old). Therefore, considering these factors, differences would be expected between the enumerated and modelled population estimates. We would expect that an updated version of GRID3 modeled population estimates which is currently under development, to be an improvement on existing estimates.

Table 2. The enumerated population totals collected by BCR and mean, median, lower, and upper (two-tailed 90% confidence intervals) population estimates calculated from GRID3 2018 modelled population estimates for the selected pre-EAs across the three sites.

Province	Site	Pre-EA IDs	Area (m^2)	Class	Estimated Population (GRID3 estimates)				Population enumerated by BCR
					mean	median	lower	upper	
Kinshasa	Q. Dumi	023	8,924,043	3	637	615	453	891	396
Kinshasa	Q. Dumi	041	6,830,845	2	464	448	319	669	115
Kinshasa	Q. Dumi	064	889,895	2	929	910	680	1,238	721
Kinshasa	Q. Dumi	076	8,976,656	3	155	147	94	244	103
Kinshasa	Q. Kingu	298	63,135	2	1,054	1,038	779	1,383	845
Kinshasa	Q. Kingu	304	64,526	1	841	825	608	1,126	538
Kongo-Central	S. Kasangulu	119	895,755	2	648	634	457	885	1,178
Kongo-Central	S. Kasangulu	122	645,185	2	700	635	316	1,309	726

Kongo-Central	S. Kasangulu	133	8,787,914	3	213	193	94	404	43
Kongo-Central	S. Kasangulu	134	8,977,448	2	516	468	232	959	106
Kongo-Central	S. Kasangulu	187	8,665,709	3	161	146	67	313	1
Kongo-Central	S. Kasangulu	201	7,590,453	2	190	171	82	363	0

Discussion

This work represents the first attempt to generate semi-automatic pre-EAs in DRC where a digital EA dataset from the previous census is not in existence. A sample dataset was successfully created for three test sites, and a sample of pre-EAs selected for evaluation in the field. This work was conducted with a number of objectives. These included for the BCR to develop familiarity with the semi-automated approach, understanding the relationship between boundaries and visible ground features through conducting an assessment exercise in the field for a subset of pre-EAs and assessing the suitability of criteria for determining pre-EA extent (estimated population etc). Through the field data collection across three sites, data were successfully collected by the BCR field teams to check the pre-EA boundaries for 15 pre-EAs with population enumeration conducted in 12 pre-EAs.

For class 1 pre-EAs, the field teams were generally able to follow the pre-EA boundary with few issues, as these pre-EAs were typically in urban settings, with the pre-EA boundary following roads or tracks. This meant that it was clear to enumerators where they needed to enumerate households. Where boundaries follow visible features (class 1/class 2), there were rarely any issues with knowing where the boundary was and therefore the designation of an area as belonging to a particular pre-EA. This was typically the case in urban areas where the density of linear features such as roads is higher, and geospatial data for these features are more readily available. In contrast, in rural areas, there are far fewer linear features which could be used as boundaries and any geospatial data on these features tends to be less also. To address the lack of linear features in rural areas (predominantly class 2/class 3), Voronoi polygons were created around areas of settlement, such that the space between settlements was the indication of the pre-EA boundary. The combination of available linear features with Voronoi polygons to create boundaries and the use of GPS-enabled tablets with satellite imagery and the boundaries

overlaid, enabled field teams to determine where they needed to visit and enumerate with relative ease.

The 3 class classification was used to ensure that the field teams collected data for a range of pre-EAs. For this field test, the pre-EA boundaries created from the semi-automatic approach were deliberately used in the field by the team, without any prior manual editing in the lab. This was agreed with BCR mainly to see if the outputs are satisfactory to BCR without any prior manual editing and to evaluate the pre-EA boundaries in the field without prior checkings in the lab. Therefore, this might have complicated the field data collection in some areas. To maximise the efficiency, we would suggest having prior-checkings in the lab and manual edits as necessary for future work.

A key consideration in creating boundaries is that they should not cut-through buildings. This was something that was checked by the teams in the field, and among all the pre-EA boundaries that were checked on the ground, only three small structures were cut through by the pre-EA boundaries, likely reflecting changes since the input datasets (e.g. OSM) were digitised, or the satellite imagery used to digitise the input datasets being out of date or of low resolution. It is encouraging that the field test showed that the semi-automated approach resulted in only a few instances where boundaries cut through buildings. It is inevitable that there will always be some instances where this occurs, and appropriate instructions should be included in the data collection protocol for the field teams, as to how to treat such a condition on the ground.

In other cases, in rural areas, it happened in two pre-EAs where a small settlement was split across multiple pre-EAs; the pre-EA boundaries cut the settlement into two parts. Preferably, a pre-EA should include an entire settlement (e.g a village) if the maximum population constraint allows. However, this could happen if the settlement is spread out, with sufficient areas without buildings for it to be separated into multiple Voronoi polygons, or the total population of a settlement is larger than the maximum population threshold for a single pre-EA. In order to avoid splitting settlements that have a population smaller than the maximum population threshold and make sure to include the entire settlement in a pre-EA, further ideas will be considered in the future pre-EA creation. For instance, before using the settlement outline for creating the Voronoi polygons, an aggregation method with a short distance should be applied to aggregate the settlements that are close to each other. By applying this, we will make sure that the small settlements that are close to each other, and possibly were part of the same settlement originally, will be aggregated to one settlement and will be represented in one pre-EA if the maximum population constraint allows.

In addition, as we see in most rural villages (e.g. pre-EA 064 (figure 8), pre-EA 023 (figure 6)), usually there are one or more roads which go through the middle of the village, such that the village spans both sides of the road. In rural areas, if the road datasets are used in the splitting process, it is likely that the village will be split between two or more pre-EAs. To address this, we suggest using Voronoi polygons, which is based on the settlement delineation, and other extracted features except the road to split the rural area in the first initial splitting process. After this step, we can identify the splitting units that have a population larger than the maximum population threshold. Then, the road lines should be only used to split the regions that have a population totals larger than the maximum population threshold after the initial splitting process. Following this, we will make sure that the entire settlements are more likely to be included in one pre-EA and only settlements with a total population larger than the maximum population threshold will be split into two or more pre-EAs.

Some of the pre-EA boundaries were an awkward shape, for example on the east side of pre-EA 023 (figure 6), the boundary narrowed and then spread out again, forming an additional area on the east side. Pre-EA 023 was successfully enumerated without issue, however, such shapes may complicate the field data collection and slow down the process. To address this, further constraints and parameters will be introduced in the future development of the approach. For instance, to avoid creating such a shape, a minimum shared boundary length threshold constraint could be introduced in the future development of the approach. This would mean that only regions with a specified percentage of the shared borders will be considered for merging with their neighbour in the merging process.

An upper threshold of 9km^2 was used when merging units to create the pre-EAs. In urban settings, pre-EAs were generally much smaller than the area threshold as the specified population size threshold was reached first, e.g. the pre-EAs where household enumeration was conducted in Quartier Dumi were less than 0.5km^2 in area. However, in sparsely populated rural areas, pre-EAs which were approximately 9km^2 were commonly created with a low estimated population size. In the most sparsely populated areas, the population. Going forward, decisions are needed as to the most appropriate thresholds for the EA unit area and population size. This is particularly important for rural, sparsely populated areas where a balance is needed between EAs that are manageable in terms of the area they cover, and ensuring EAs don't have very low population counts. This is something that BCR needs to agree on regardless of whether a manual or automated approach is used in creating enumeration areas.

Further discussion is also needed with regards to the existing administrative boundaries within which pre-EAs should nest. No issues were reported with the nesting of pre-EAs in the field sites, however, the area covered by the field sites was relatively small. Issues were noted with the current quartier boundaries at the Kingu site where differences existed between the extent of the area designated as the quartier of Kingu in the BCR quartier boundary dataset and on the ground. Wider exploration of administrative boundary datasets indicated that there seem to be some instances where quartier boundaries overlap in places with Secteur/Province boundaries, e.g. “Kinshasa” Quartiers overlap with “Kongo-Central” Secteurs and Kinshasa Secteur boundaries. Future use of the quartier boundaries in census preparation or for other applications, will require spatial inconsistencies in the boundary datasets to be addressed. In terms of creating pre-EAs, suggestions were made by BCR staff that within urban areas, there could be no sub-division administrative units (e.g. quartiers) specified in terms of input boundaries within which pre-EAs must nest. And instead, following the cartography, the EA boundaries could be updated to reflect information on administrative subdivisions gathered in the field. Alternatively, currently available administrative quartier boundaries could be included as units within which pre-EAs must nest and then the following cartography, the generated pre-EA boundaries could be updated to reflect information on administrative subdivisions gathered in the field.

Overall, it can be considered that the results from the field test were encouraging with the approach performing well at the sites in which it was tested. As noted by senior BCR staff during our concluding meeting after the fieldwork, the sites selected by the BCR include challenging locations which were not included in the first pilot cartographic test due to the challenge presented by inaccessibility, steep terrain, erosion, sparse population and lack of network coverage (everything had to be offline). Despite these challenges, the presented semi-automatic approach and the methods and applications that were used to check boundaries and collect GPS household locations in this field test have delivered encouraging results at all sites.

Limitations

A range of limitations with the field test have been identified. These include limitations with the methods and equipment used in the field data collection and the methods and input data used to produce the pre-EA boundaries.

In terms of the input data used in producing the pre-EAs, OSM was used as the primary data source for roads, waterways and railways. The coverage of these datasets is patchy with good coverage in some areas but poor coverage in others, particularly for roads in

rural areas. This limits the flexibility of the creation of pre-EAs. There is an increasing amount of data on roads and other features becoming available, and comprehensive exploration of potential additional datasets is needed.

The tablets used in the fieldwork were in excess of 7 years old and are not regularly used. This meant there were some technical issues with the tablets functioning as expected, with long delays in getting accurate GPS locations and the batteries in some of the tablets draining quickly. Portable chargers and solar bags were provided by SINFIC to the BCR, however there were issues with getting these to function as expected, and it would be prudent to test these further and potentially find alternatives before further use in the field. In some instances, the location provided through the in-built GPS on tablets took a long time to be found and the location was not accurate, which resulted in some household locations being recorded outside the pre-EA boundaries coverage. For example, in the pre-EA 064 at the Dumi site, household locations recorded by the GPS are on the north side of the road, whereas in reality the fieldworker was located on the south side of the road (the road forms the northern boundary of the pre-EA 064). There were also issues with loading and working with the MBTiles, with issues reported around the MBTiles freezing. The cause of the issue is assumed to be associated with the age of the tablets.

On the tablets, MBTiles of satellite imagery with the pre-EA boundaries overlaid, were used as basemaps. A number of limitations were identified related to this. The MBTiles consisted of the pre-EA boundary polygon, overlaid on satellite imagery for the area covered by the pre-EA. This was found to not provide sufficient detail for navigation purposes. In rural areas, the coverage of the generated MBTiles was too small for some of the pre-EAs, such that there were issues navigating to the pre-EAs from the surrounding area, which made it difficult for the team to find them in the field. MBTiles covering a larger area and with additional features indicated, should be used for future fieldwork.

Practical issues were also encountered in the field surrounding difficult terrain, weather conditions, lack of permission to collect data in some areas, lack of internet and poor coverage of the mobile network. There were difficulties in accessing some of the pre-EAs due to the nature of the location with steep slopes, erosion and cliffs present in some areas. These made the data collection more challenging as the field team had to take different paths to check all parts of the pre-EA boundaries which slowed down data collection. It rained whilst the teams were in the field and the BCR tablets did not have waterproof covers. GRID3 provided resealable plastic bags (Ziploc type) to protect the tablets and this enabled the tablets to be used whilst it was raining. The poor coverage of telecommunication networks and the low speed of internet service had limited the communication within the teams and the daily sending of data. In addition, it was not

possible to work on or update some of the materials or datasets that we had prepared for the fieldwork while we were in DRC because of poor internet connection. The BCR cartography lab does not have internet connectivity. All preparations reliant on internet connection (e.g. creation of MBTiles) should be completed in a location with a strong internet connection, which even within Kinshasa are few and far between.

With regard to authorisation to collect data in the field, a number of issues were encountered. The site at Kingu encompassed the quartier of Kingu, as designated by the BCR quartier boundary dataset. However, the designation of the quartier of Kingu in the BCR dataset was not up to date, as indicated by the local guide, which halted data collection within selected pre-EAs, as authorisation had only been given for data collection within Kingu, and not for Mbala (the neighbouring quartier). Given the very limited time for the field test to be conducted, awareness about conducting the household data collection in the selected sites on the ground was not as comprehensive as it should have been, and further awareness raising should be done for future work.

It was also discovered by the field teams in Kasangulu that there were areas within the selected pre-EAs affected by insecurity. This was not known by the BCR or UNFPA. Although this didn't specifically affect data collection during the field test, given the short duration in the field, further questions around security assessment and greater information gathering/awareness raising from local sources are needed ahead of future field data collection

Considering all the logistical issues and that the approach is still under development, the field test shows that even an initial set of pre-EA boundaries created using the semi-automated EA delineation approach, was found to be suitable to facilitate enumeration in the field. The presented methodology and applications for checking the pre-EA boundaries and data collection are thought to have maximised the overall work efficiency. Initially, it had been anticipated by the BCR, that only two pre-EAs per site can be enumerated within the time period of the field test. However, all the pre-EAs that were selected for Dumi and Kasangulu were fully enumerated (10 pre-EAs) by the BCR team. Only at the Kingu site, were all selected pre-EAs not enumerated. This was not due to time availability but instead because of the misclassification of Quartiers and local authorization issues. Two pre-EAs were fully enumerated but the team indicated that more pre-EAs could have been enumerated if these issues were not encountered.

Conclusion

In conclusion, the field test which aimed to assess the feasibility of using a semi-automated approach for the delineation of pre-EA boundaries in DRC, was successfully conducted. The field test was conducted across three sites in Kongo-Central and Kinshasa provinces and included a range of urban and rural contexts. A sample of 15 pre-EAs was selected and assessed in the field to check how the boundaries related to ground features and their feasibility as units for population enumeration.

Visual assessment of the pre-EA dataset confirmed that the pre-EA boundaries were nested within the existing administrative boundary hierarchy without gaps or overlaps and were in line with BCR's criteria. In the field, the BCR staff found that the pre-EA outputs were suitable for use in enumeration; the field teams could navigate within pre-EAs, know which housing units required enumeration and recognise when they had reached the boundary of a pre-EA. No major issues were reported by the field teams in knowing which housing units to enumerate. The pre-EAs used in the field test were created with area constraints of 9km^2 and population constraints of 1,000 (rural) and 1,200 (urban). These were found to be satisfactory, but further consideration around this may be needed, particularly in regard to sparsely populated areas, where a balance is needed between EAs that are manageable in terms of the area they cover, and ensuring EAs don't have very low population counts.

In order to further enhance the quality of the pre-EA boundaries created using the semi-automated approach we intend to conduct a comprehensive review of currently available input feature datasets (such as roads), as there is an increasing volume of such data being available. We would also suggest that a standardised process is developed in the lab to review pre-EA boundaries created using the semi-automated approach and implement manual edits to the pre-EA boundaries as necessary, before using them in the field.

The findings of the field test indicate this semi-automated approach to creating pre-EAs has the potential to be used by the BCR to create pre-EAs in preparation for census cartography, with the potential to expand for the entire country, and offers large savings in terms of time, labour and cost. We expect the findings of the field test in DRC to be transferable to other similar contexts, with the approach having applicability in countries with no recent digital EAs. We also expect the approach could be adapted to update digital EA boundaries in contexts with outdated EA datasets, but this should be explored through further research and testing in such contexts.

Next steps

Based on the conclusions of the field test, the following next steps have been identified:

- BCR leadership to discuss how the semi-automated approach can be used for census cartography and the associated timelines. Decisions needed around whether to first create a pre-EA dataset specifically for areas included in the second pilot cartographic exercise or create a national pre-EA dataset straight off.
- WorldPop to conduct a comprehensive review of available input datasets, and share the results with the BCR.
- Considering the recent feedback from the BCR and knowledge gained through the field test, plans should be made to introduce several new constraints, parameters and techniques in the future development of the approach.
- The BCR leadership needs to discuss the most appropriate thresholds for the EA unit area and population size. This is particularly important for rural, sparsely populated areas where a balance is needed between EAs that are manageable in terms of the area they cover and ensuring EAs don't have very low population counts. This is something that needs to be agreed on regardless of whether a manual or automated approach is used in creating enumeration areas.
- UNFPA and WorldPop to discuss the future development of the approach and the feasibility of transforming it into a user-friendly tool.
- We expect that the findings of the field test in DRC and the semi-automated approach to create pre-EAs, to be transferable to other countries without recent EAs. We also expect the approach could be employed to update outdated digital EA boundaries, but this should be explored through further testing in this context. To progress this work, opportunities to apply the approach in a range of contexts should be identified (e.g. in other countries without national EA datasets, and in a country where the current EA boundaries are outdated).

Release content and file descriptions

- PreEA_Outputs_TestLocations.shp
  This is shape file for the pre-Enumeration Area outputs generated from the semi-automated preEA approach.

Attribute Table

Column name	Description
Admn1_Name	DRC administrative one name

Region	DRC region name
EA_Number	Unique pre-Enumeration Area number
Area_m2	Pre-Enumeration area in m^2
CLASS	Class of preEA depending on the degree to which the pre-EA boundaries followed visible features
GroundPop	Ground population estimates
PopEst_Med	Model population estimates_Median
PopEst_Mea	Model population estimates_Mean
PopEst_Low	Model population estimates_Lower limit
PopEst_Upp	Model population estimates_Upper limit

Acknowledgement

Funding for the implementation of this ground truthing fieldwork was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (INV-009579, formerly 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 preEAs in the test locations. Freja Hunt was leading the code development in consultation with Sarchil Qader. Sarchil Qader and Heather Chamberlain designed the field data collection material and protocols and trained the BCR team. Sarchil Qader, Heather Chamberlain and Mathias Kuepie supervised the BCR data collector team on the ground. Oversight of the work was provided by Attila N. Lazar and Andrew J. Tatem

We would like to sincerely thank BCR team for conducting the fieldwork. Special thanks to the UNFPA HQ and DRC UNFPA country office for facilitating and coordinating the fieldwork. Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org".

The whole WorldPop group and GRID3 partners are acknowledged for overall project support.

License

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

References

G Boo, R Hosner, PZ Akilimali, E Darin, HR Chamberlain, WC Jochem, P Jones, R Shulungu Runika, HM Kazadi Mutombo, AN Lazar and AJ Tatem. 2021. Modelled gridded population estimates for the Haut-Katanga, Haut-Lomami, Ituri, Kasaï, Kasaï-Oriental, Lomami and Sud-Kivu provinces in the Democratic Republic of the Congo (2021), version 3.0. WorldPop, University of Southampton, Flowminder Foundation, École de Santé Publique de Kinshasa, Bureau Central du Recensement and Institut National de la Statistique. doi:10.5258/SOTON/WP00720

Bureau Central du Recensement (BCR). 2018. Report des Limites Administratives (RLA) [Dataset].

OpenStreetMap contributors. 2019. Roads, Waterways, Railways [Dataset].

WorldPop (School of Geography and Environmental Science, University of Southampton). 2020. Bottom-up gridded population estimates for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo, version 1.0. https://dx.doi.org/10.5258/SOTON/WP00658