Release Statement

Modelled gridded population estimates for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo 2018, version 2.0

12 June 2020

Original Release: 12 June 2020

These data were produced by the WorldPop Research Group at the University of Southampton. This work is part of the GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development) project funded by the Bill and Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (OPP1182408). Project partners include WorldPop at the University of Southampton, the United Nations Population Fund (UNFPA), Center for International Earth Science Information Network (CIESIN) in the Earth Institute at Columbia University, and the Flowminder Foundation. The modelling work was led by Gianluca Boo and Edith Darin with the support from Douglas R. Leasure and Claire A. Dooley. Coordination was provided by Heather R. Chamberlain and oversight by Andrew J. Tatem and Attila N. Lazar. The support of the whole WorldPop Research Group is acknowledged. The UCLA-DRC Health Research and Training Program, the Kinshasa School of Public Health (KSPH), and the Bureau Central du Recensement (BCR) coordinated and conducted the two microcensus rounds. The Oak Ridge National Laboratory contributed to the first round of microcensus. We acknowledge the contribution of the many individuals within these institutions.

These data are operational population estimates and are not official government statistics.

CITATION

Boo G, Darin E, Leasure DR, Dooley CA, Chamberlain HR, Lazar AN, Tatem AJ. 2020. Modelled gridded population estimates for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo, version 2.0. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00669

RELEASE CONTENT

1. COD_population_v2_0_gridded.zip
2. COD_population_v2_0_agesex.zip
3. COD_population_v2_0_mastergrid.tif
4. COD_population_v2_0_sql.sql
5. COD_population_v2_0_tiles.zip

FILE DESCRIPTIONS

COD_population_v2_0_gridded.zip

This zip file contains two files:

COD_population_v2_0_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100m grid cell (0.0008333 decimal degrees grid or 3 arc seconds) across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo. The values are the mean of the posterior probability distribution for the predicted population size in each grid cell. NA values represent areas that were mapped as unsettled according to building footprints data. Note: This raster is accompanied by one ancillary file that contains metadata (COD_population_v2_0_gridded.tif.xml).

COD_population_v2_0_uncertainty.tif

This geotiff raster contains estimates of uncertainty in the population estimates for each approximately 100m grid cell across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo. The uncertainty values are the difference between the upper and lower 95% credible intervals of the posterior prediction divided by the mean of the posterior prediction: (upper – lower)/mean. As a consequence, cells with a mean prediction of 0 result in NA uncertainty values. These numbers provide a comparable measure of uncertainty in population estimates across the country. Uncertainty estimates cannot be summed across grid cells to produce an uncertainty measure for a multi-cell area. Uncertainty for multiple cells can be calculated using the cells' posterior predictions. Note: This raster is accompanied by one ancillary file that contains metadata (COD_population_v2_0_uncertainty.tif.xml).

COD_population_v2_0_agesex.zip

This zip file contains 40 geotiff rasters. Each raster provides gridded population estimates for an age-sex group for each approximately 100m grid cell across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the

Congo. We provide 36 rasters for the commonly reported age-sex groupings of sequential age classes for males and females separately. These are labelled with either an "m" (male) or an "f" (female) followed by the number of the first year of the age class represented by the data. "f0" and "m0" are population counts of under 1-year olds for females and males, respectively. "f1" and "m1" are population counts of 1 to 4 year olds for females and males, respectively. Over 4 years old, the age groups are in five year bins labelled with a "5", "10", etc. Eighty year olds and over are represented in the groups "f80" and "m80". We provide four additional rasters that represent demographic groups often targeted by programmes and interventions. These are "under1" (all females and males under the age of 1), "under5" (all females and males under the age of 5), "under15" (all females and males under the age of 15) and "f15_49" (all females between the ages of 15 and 49, inclusive). These data were produced using age-sex proportions estimated at the province level using the microcensus data. The age-sex proportions were applied to the gridded population estimates (COD_population_v2_0_gridded.tif) to allocate the population to the different age-sex classes. While this data represents population counts, values contain decimals, i.e. fractions of people. This is because both the input population data and age-sex proportions contain decimals. For this reason, it is advised to aggregate the rasters at a coarser scale. For example, if four grid cells next to each other have values of 0.25 this indicates that there is 1 person of that age group somewhere in those four grid cells.

COD_population_v2_0_mastergrid.tif

This geotiff raster contains 1s for each settled approximately 100m grid cell (0.0008333 decimal degrees) across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo. 0 values indicate grid cells that were considered unsettled and thus not containing people. NAs show grid cells considered as outside the above-mentioned provinces.

COD_population_v2_0_sql.sql

This SQLite database contains samples (n=10,000) from the Bayesian posterior predictions of population size in each grid cell. These can be used to derive the posterior distribution for population totals for larger areas that contain more than one grid cell. This database is the source data for WorldPop tools (e.g. https://apps/wprldpop.org/woprVision) used to display and analyze the model results. Note that these 10,000 samples do not necessarily produce a fully converged posterior distribution. The fully converged Bayesian model contained three MCMC chains. We limited the SQL database to 10,000 samples due to file size considerations (the SQL database is approximately 50 GB).

COD_population_v2_0_tiles_population.zip

This tiled web map allows for rapid display of the approximatively 100m gridded population estimates across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo (COD_population_v2_0_gridded.tif). These can be used to develop web applications for the model results. The tiles were created in XYZ format (i.e. compatible with Leaflet) with full coverage of the above-mentioned provinces for the zoom levels 1 to 14.

RELEASE HISTORY

Version 2.0 (27 May 2020)

- This is a major model update based on finer resolution input data.
    - The settled extent is no longer derived from settlement data [7] but from building footprints data [8].
    - Population estimates for the different age-sex groups are no longer derived from existing age-sex proportions [9] but from the original microcensus survey data [10].
- Gridded population estimates were added for individual age-sex groups (COD_population_v2_0_agesex.zip).
- Uncertainty tiles "COD_population_v1_0_tiles_uncertainty.zip" were removed because they were discontinued for use in WorldPop web applications (e.g. https://apps/wprldpop.org/woprVision).

Version 1.0 (20 May 2019) [https://dx.doi.org/10.5258/SOTON/WP00658]

- Original release of the population dataset.

ASSUMPTIONS AND LIMITATIONS

The assumptions and limitations are as follows:

- These population estimates represent the time period between 2017 and 2018 corresponding to when the two rounds of microcensus surveys were conducted. We assume the data represents early 2018, however, we cannot pinpoint an exact time because the input data was collected at different time points. We also cannot assign a specific month to the dataset for the same reason.
- Population estimates are constrained within the area derived from building footprint data [8] and the province boundaries of the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces [11]. We assume that the building footprint data is accurate and that each building polygon corresponds to a building structure. We also assume that the province boundaries are accurate.
- The model assumes that no people live outside of the settled area defined above. The extent of the settled area could be significantly underestimated in regions

where the urbanization patterns are more dynamic and the imagery used to derive the building footprint data is outdated.

- A comparison of COD_population_v1_0_mastergrid.tif (defined based on settlement data [7]) and COD_population_v2_0_mastergrid.tif (defined based on building footprints data [8]) can enable the quick identification of potential omissions in the building footprint data. In the south of the Kongo Central province, for instance, a number of settlements are present in the settlement data but not in the building footprints data probably due to the date of the imagery used for the feature extraction.
- The model assumes that each building polygon is potentially residential. This can introduce uncertainty in regions of greater urban complexity, where residential, non-residential, and mixed-use buildings are juxtaposed.
- The age-sex proportions are modelled at the province level based on the microcensus survey data [10]. This aggregation is meant to overcome known issues associated with low sample size. The mean proportions are applied to the gridded population estimates to allocate the population to the different age-sex classes. We could not compute ad store the posteriors for the predicted population counts within each age-sex class because of file size considerations (the SQL database is approximately 1.2TB).

SOURCE DATA

The key datasets used to produce the modelled population estimates are:

- Two rounds of microcensus surveys (n=926) conducted across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo in 2017 and 2018 [10]. Survey weights, household GPS locations, and population and age-sex totals were used.
- Building footprints for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces in the Democratic Republic of the Congo [8]. Vector polygons were used to derive metrics based on building count, area, perimeter, and distance to the nearest building.
- Settlement extent classification for the Democratic Republic of the Congo consisting of built-up areas (BUA), small settlement areas (SSA), and hamlets (hamlets) [12]. Two settlement classes were derived from this dataset — urban (comprising BUA) and rural (comprising SSA and hamlets).
- Administrative boundaries at the province, territory (territoire), sector or chiefdom (secteur or chefferie), and commune level for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces [11].

METHODS OVERVIEW

The key steps of our approach were as follows:

- Cleaning of household survey data and imputing population in households where a non-response was recorded. Imputation was based on the average household size within each microcensus cluster.
- Extracting the building footprint vector polygons with centroid located within each microcensus cluster. The vector polygons were further subset to retain building footprints located within a radius of approximatively 50m (0.00041665 decimal degrees) distance from the surveyed households.
- Developing a Bayesian statistical model including:
    - A generalized linear mixed model describing the relationship between population counts and building area within the microcensus clusters using three geospatial covariates — mean building area, mean distance to the nearest building, and mean building count within a focal window of approximatively 1Km.
    - A weighted likelihood accounting for the different sampling designs adopted across the two rounds of microcensus surveys, i.e. spatial random sampling in 2017 and population-weighted sampling in 2018.
    - A multinomial model estimating individual age-sex group proportions from the aggregated age-sex population counts for the microcensus clusters within each province.
- Fitting the model to the data, checking model convergence.
- Carrying out 10-fold cross-validation to check model fit.
- Predicting population size and age-sex proportions for each settled pixel across the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces.

The data processing and analysis were carried out using R (v.3.6.0) [1] and JAGS (v4.3.0) [2].
The concept of bottom-up population modelling for estimating population in the absence of recent census data was described by [3]. Similar approaches have been carried out for Afghanistan [4], Nigeria [5,6, 13] and Zambia [14].

WORK CITED

[1] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[2] Plummer M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd international workshop on distributed statistical computing 124(125):10.

[3] Wardrop N.A., Jochem W.C., Bird T.J., Chamberlain H.R., Clarke D., Kerr D., Bengtsson L., Juran S., Seaman V., and Tatem A.J. 2018. Spatially disaggregated

population estimates in the absence of national population and housing census data. Proceedings of the National Academy of Sciences 115, 3529–3537.
[4] Chamberlain H.R.*, Clarke D.J.*, Jochem W.C., Wardrop N., Kerr D., Hellali B., Rasuli J., Nabi M., Rahimi S., Ndyanabangi B., Halimi A., Bashir N., Bird T.J., Bengtsson L., Juran S., and Tatem A.J. 2020. National population mapping from spatially incomplete enumeration data [draft].
[5] Weber E.M., Seaman V.Y., Stewart R.N., Bird T.J., Tatem A.J., McKee J.J., Bhaduri B.L., Moehl J.J., and Reith A.E., 2018. Census-Independent Population Mapping in Northern Nigeria. Remote Sensing of Environment 204, 786–98.
[6] Leasure D.R., Jochem W.C., Weber E.M., Seaman V., and Tatem A.J. 2020. High-resolution population mapping with limited survey data: a hierarchical Bayesian modeling framework to account for uncertainty [in review at Proceedings of the National Academy of Sciences of the United States of America].

DATA CITED

[7] Oak Ridge National Laboratory. 2016. LandScan HD DRC, version 1.1.
[8] Maxar Technologies, Ecopia.AI. 2020. Building Footprints DRC, Digitize Africa data.
[9] WorldPop (School of Geography and Environmental Science, University of Southampton; Department of Geography and Geosciences, University of Louisville; Departement de Geographie, Universite de Namur) and Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00646
[10] UCLA-DRC Health Research and Training Program (University of California, Los Angeles) and Kinshasa School of Public Health. 2017 and 2018. Kinshasa, Kongo Central and former Bandundu microcensus survey data.
[11] Bureau Central du Recensement (BCR). 2018. Report des Limites Administratives (RLA).
[12] Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2020. GRID3 Democratic Republic of the Congo Settlement Extents Version 01, Alpha. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). Source of Building Footprints “Ecopia Vector Maps Powered by Maxar Satellite Imagery” © 2020. https://doi.org/10.7916/d8-cpry-wv37.
[13] WorldPop (School of Geography and Environmental Science, University of Southampton). 2020. Bottom-up gridded population estimates for individual age-sex groups in Nigeria, version 1.2.1. https://dx.doi.org/10.5258/SOTON/WP00661.

[14] WorldPop (School of Geography and Environmental Science, University of Southampton). 2020. Bottom-up gridded population estimates for Zambia, version 1.0. https://dx.doi.org/10.5258/SOTON/WP00662.