Release Statement

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

13 February 2020

Original Release: 20 May 2019

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the GRID3 project with funding from the Bill and Melinda Gates Foundation and the United Kingdom's Department for International Development (OPP1182408). Project partners included the United Nations Population Fund, Center for International Earth Science Information Network in the Earth Institute at Columbia University, and the Flowminder Foundation.

CITATION

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

RELEASE CONTENT

1. COD_population_v1_0_gridded.zip
2. COD_population_v1_0_sql.sql
3. COD_population_v1_0_mastergrid.tif
4. COD_population_v1_0_tiles_population.zip
5. COD_population_v1_0_tiles_uncertainty.zip

FILE DESCRIPTIONS

COD_population_v1_0_gridded.zip

This zip file contains two raster files:

COD_population_v1_0_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100 m grid cell (0.0008333 decimal degrees grid) 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 by LandScanHD v1.1 (ORNL 2018).

COD_population_v1_0_uncertainty.tif
This geotiff raster contains estimates of uncertainty in the population estimates within each approximately 100 m 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. These numbers provide a comparable measure of uncertainty in population estimates across the country.

COD_population_v1_0_sql.sql
This SQLite database contains samples (n=9,000) 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 used to display and analyze these model results. Note that these 9,000 samples do not necessarily produce a fully converged posterior distribution.

COD_population_v1_0_mastergrid.zip
This geotiff raster contains the mastergrid used to define where population estimates were made. It contains a value of 1 in areas that are mapped as residential and values of 0 in unsettled and non-residential areas (based on ORNL 2018). This raster can be used to identify cell IDs of a location in the study area which can be used to lookup population estimates in the SQL database.

COD_population_v1_0_tiles_population.zip
This tiled web map allows for rapid display of the approximately 100 m gridded population estimates across the study area. These can be used to develop web applications for these model results.

COD_population_v1_0_tiles_uncertainty.zip
This tiles web map allows for rapid display of the approximately 100 m gridded estimates of uncertainty across the study area. These can be used to develop web applications for these model results.

RELEASE HISTORY

Version 1.0 (20 May 2019)

- Original release of the Democratic Republic of the Congo population dataset

ASSUMPTIONS AND LIMITATIONS

To retrieve totals and uncertainty measures at the province level, it is recommended to use the pre-calculated figures reported in the *GRID3_DRC_PopEst_v1_0_province.shp* data. Aggregation based on the raster *GRID3_DRC_PopEst_v1_0_mean.tiff* will not provide correct confidence intervals.

Population estimates are constrained within the settled area derived from LandScan HD DRC data and the province boundaries of the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces provided by the Bureau Central du Recensement (BCR) of DRC. The LandScan HD DRC data was produced based on satellite imagery collected during the period 2008-2016. The current model also assumes that no people live outside of the settled area defined above. The extent of settled area could be significantly underestimated where the urbanization patterns are more dynamic and the imagery is outdated. These patterns are also expected to involve the transition of settlement types (e.g., from rural to urban) in peri-urban areas. For the previous reasons, the population model is potentially underestimating population both at the pixel and province level.

Whether the current version of the model addresses most of our concerns, future versions will focus on 1) providing a model-based update of the settled area through a dedicated model component; and 2) developing further model assessments, including the selection of additional model covariates that could improve the current population estimates. Future versions of the model will also aim at 3) including gridded age and gender breakdowns, first using pre-calculated demographic pyramids produced by WorldPop and then the original microcensus data. In these directions of future improvements, feedback from programme partners and data users will be important to help identify areas in need of improvement.

SOURCE DATA

Important datasets for the current model include raster data with a spatial resolution of 0.0008333 decimal degrees. In detail:

- The *population estimates* derived from WorldPop Global data for 2018 (WorldPop 2018). Focal mean with a window of approximately two kilometres.
- The *density of residential roads* derived from OpenStreetMap data for 2018 (OSM 2018). Focal mean with a window of approximately two kilometres.
- The *travel-time to cities* derived from Malaria Atlas Project data for 2015 (Malaria Atlas Project 2015).

- The *density of tertiary-sector activities* (e.g. offices, shops, and places of worship) derived from OpenStreetMap data for 2018 (OSM 2018). Focal mean with a window of approximately two kilometres.
- The *settlement extent and type* (i.e., urban, rural and hamlet) derived from LandScan HD DRC data version 1.1 (Oak Ridge National Laboratory 2016).

Additional vector data is:

- The *country, province and local administrative areas boundaries* for the Kinshasa, Kongo-Central, Kwango, Kwilu, and Mai-Ndombe provinces issued by the Report des Limites Administratives (RLA) lead by the Bureau Central du Recensement (BCR 2018).

METHODS OVERVIEW

Building on previous work carried out in Nigeria (Leasure et al. 2019; Weber et al. 2018) and Afghanistan (Clarke et al. 2018; Chamberlain et al. 2019), these population estimates are produced using a model-based approach described by Wardrop et al. (2018).

The population model combines information on population characteristics from small microcensus surveys (n=926) (UCLA and KSPH 2018) with high-resolution geospatial datasets, and estimates the relationships between population densities and geospatial data within the microcensus enumeration zones. An additional model component accounts for the different sample designs (i.e., spatial random sampling and population-weighted sampling) implemented in the two rounds of microcensus surveys. This model set-up provides a basis for extrapolating population estimates to areas where no microcensus survey has been carried out but settled area has been observed.

The population model consists of a hierarchical Bayesian regression within the family of Poisson generalised linear mixed models. The hierarchical form allows for submodels to estimate average population densities for specific settlement types, provinces and local administrative areas. The model also estimates responses of population densities to high-resolution geospatial covariates. These relationships are used to further refine estimates of population densities within settled grid cells across the five provinces.

The model was implemented using JAGS v4.3.0, R v3.5.0, and the R package runjags (Plummer 2003; R Core Team 2013; Denwood 2016).

ACKNOWLEDGEMENTS

We thank the UCLA-DRC Health Research and Training Program, the Kinshasa School of Public Health (KSPH), and the Bureau Central du Recensement (BCR) for coordinating and

conducting the two microcensus rounds. We also thank the Oak Ridge National Laboratory for contributing to the first round of microcensus and providing the settlement data used in this model. We thank OpenStreetMap and its contributors for the work carried out in DRC. This work was supported with funding from the Bill & Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (DFID).

WORK CITED

Denwood, M.J. 2016. runjags: An R package providing interface utilities, model templates, parallel computing methods and additional distributions for MCMC models in JAGS. Journal of Statistical Software 71.9: 1-25.

Clarke, D.J., Chamberlain, H.R., Jochem, W.C., Kerr, D., Wardrop, N., Bird, T., and Tatem, A.J. 2018. High-resolution population mapping for Afghanistan. Report for the Afghanistan Central Statistics Organisation [unpublished].

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., 2019 National population mapping from spatially incomplete enumeration data [draft].

Leasure, D.R., Jochem, W.C., Harper, M., Weber, E.M., Seaman, V., Tatem, A.J. 2019. High resolution population mapping with limited survey data: a hierarchical Bayesian modeling framework to account for uncertainty [draft].

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 . Vol. 124. No. 125.10.

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

Stevens F.R., Gaughan A.E., Linard C., Tatem A.J. 2015. "Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data." PLOS ONE 10 (2). Public Library of Science: 1–22.

Wardrop N.A., Jochem W.C., Bird T.J., Chamberlain H.R., Clarke D., Kerr D., Bengtsson L., Juran S., Seaman V., 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.

Weber E.M., Seaman V., 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.

DATA CITED

Malaria Atlas Project. 2015. A global map of travel time to cities to assess inequalities in accessibility in 2015. Predicted travel time (minutes) to nearest city [Dataset].

Oak Ridge National Laboratory. 2016. LandScan HD DRC version 1.1 [Dataset].

OpenStreetMap contributors. 2018. Residential road data [Dataset].

WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton). 2018. Democratic Republic of the Congo 100 m Population. 2018 estimates of numbers of people per grid square [Dataset].

University of California, Los Angeles and Kinshasa School of Public Health. 2018. Kinshasa, Kongo Central and former Bandundu microcensus survey [Dataset].

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