Release Statement

Bottom-up gridded population estimates for Nigeria, version 1.2

29 March 2020

Original Release: 10 July 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 (CIESIN) in the Earth Institute at Columbia University, and the Flowminder Foundation. Statistical modelling was led by Dr. Doug Leasure and Dr. Chris Jochem with oversight from Professor Andy Tatem. In-country implementation was led by Dr. Tracy Adole. Oak Ridge National Laboratories (ORNL), eHealth Africa, and the Bill and Melinda Gates Foundation collected microcensus data and produced the settlement map used as inputs for this work. The whole WorldPop group and GRID3 partners are acknowledged for overall project support.

SUGGESTED CITATIONS

WorldPop (School of Geography and Environmental Science, University of Southampton). 2019. Bottom-up gridded population estimates for Nigeria, version 1.2. https://dx.doi.org/10.5258/SOTON/WP00655

* 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

RELEASE CONTENT

1. NGA_population_v1_2_gridded.zip
2. * NGA_population_v1_2_agesex.zip
3. NGA_population_v1_2_admin.zip
4. NGA_population_v1_2_sql.sql
5. NGA_population_v1_2_mastergrid.tif
6. NGA_population_v1_2_tiles.zip

Note: Files available at https://wopr.worldpop.org may be slightly different than the files available from the DOI links in the citations above. See the section "Version History" below for more information.

FILE DESCRIPTIONS

NGA_population_v1_2_gridded.zip

This zip file contains two raster files:

NGA_population_v1_2_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100 m grid cell across Nigeria. 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). These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated population totals for larger areas (i.e. as occurred in version 1.0).

NGA_population_v1_2_uncertainty.tif

This geotiff raster contains estimates of uncertainty in the population estimates within each approximately 100 m grid cell across Nigeria. 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. NA values represent areas that were mapped as unsettled by LandScanHD v1.1 (ORNL 2018).

NGA_population_v1_2_agesex.zip

This zip file contains 36 geotiff rasters (.tif) and their metadata (.xml). Each raster provides gridded population estimates for an age-sex group. The file names refer to the age-sex group represented by the raster. Age-sex group labels beginning with "f" are female populations and labels beginning with "m" are male populations. The age group labels refer to the first year of the age range. For example: "f_0" is females less than one year old. "f1" is females 1 to 4 years old. "f5" is females 5 to 9 years old. "f10" is females 10 to 14 years old. This pattern continues for each 5 year interval up to 80. "f80" is females greater than 80 years old. The labelling is the same for males: "m0", "m1", "m5", "m10", … , "m80". These data were produced using the gridded age-sex proportions from WorldPop et al. (2018b). We multiplied our gridded population estimates (NGA_population_v1_2_gridded.tif) by the gridded age-sex proportions to produce NGA_population_v1_2_agesex.zip.

Note: Summing population estimates for all age-sex groups within a pixel will not exactly match the total population estimates from "NGA_population_v1_2_gridded.zip" due to slight rounding errors, but these discrepancies are very small (i.e. less than 0.00002).

NGA_population_v1_2_admin.zip

This zip file contains population totals for administrative units in Nigeria. States are considered level 2 administrative units and local government areas are considered level 3 administrative units. The administrative boundaries (i.e. the two ESRI polygon shapefiles) were obtained from eHealth Africa in September of 2018. **These are not official government boundaries.** The attribute tables for the shapefiles and the corresponding .csv spreadsheets contain estimates of the total population sizes in each polygon and the confidence intervals. This includes the mean of the posterior prediction (column *mean*) and the quantiles of the posterior prediction (columns *q025*, *q05*, *q25*, *q50*, *q75*, *q95*, *q975*). The median is *q50* and the 95% credible intervals are described by *q025* and *q975*.

This zip file contains the following five files:

NGA_population_v1_2_ admin_level2_boundaries.shp

This shapefile contains state boundaries. Note: this file is accompanied by four ancillary files (.dbf, .prf, .shp.xml, and .shx).

NGA_population_v1_2_ admin_level3_boundaries.shp

This shapefile contains local government area boundaries. Note: this file is accompanied by four ancillary files (.dbf, .prf, .shp.xml, and .shx).

NGA_population_v1_2_admin_level0.csv

This spreadsheet contains summary statistics of the posterior prediction for the total population of Nigeria. The national population total was the sum of all state totals using the boundaries described above.

NGA_population_v1_2_admin_level2.csv

This spreadsheet contains summary statistics of the posterior predictions for the total populations of all 37 states in Nigeria (see description of state boundaries above).

NGA_population_v1_2_admin_level3.csv

This spreadsheet contains summary statistics of the posterior predictions for the total populations of all 774 local government areas in Nigeria (see description of local government area boundaries above).

NGA_population_v1_2_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 used to display and analyze these 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 with 50,000 samples each. We limited the SQL database to 10,000 samples due to file size considerations (the SQL database is over 140 GB).

NGA_population_v1_2_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.

NGA_population_v1_2_tiles.zip

This tiled web map allows for rapid display of the 100 m gridded population estimates across Nigeria. These can be used to develop web applications for these model results. The tiles were created in XYZ format (i.e. compatible with Leaflet) with full coverage of Nigeria for zoom levels 1 to 14.

RELEASE HISTORY

Version 1.2 (26 March 2020)

- Gridded population estimates were added for individual age-sex groups (NGA_population_v1_2_agesex.zip).
- The SQL database "NGA_population_v1_2_sql.sql" that is used in WOPR applications was updated to remove unnecessary data (e.g. covariate values, names of administrative units).
- Population tiles were updated with a revised color palette. This file was renamed from "NGA_population_v1_2_tiles_population.zip" to "NGA_population_v1_2_tiles.zip".
- Uncertainty tiles "NGA_population_v1_2_tiles_uncertainty.zip" were removed because they were discontinued for use in WorldPop web applications (e.g. https://apps.worldpop.org/woprVision).

Version 1.2 (10 July 2019) [https://dx.doi.org/10.5258/SOTON/WP00655]

- The previous release contained a few grid cells with erroneously high population estimates that resulted from the way the statistical model was summarised (based on 1,000 samples from posterior predictions as opposed to 10,000 samples used here).
- This update changes the population estimates slightly in every grid cell. State and LGA totals have changed marginally but remain within 1% of previous estimates.

Version 1.1 (22 February 2019) [https://dx.doi.org/10.5258/SOTON/WP00657]

- Updated to include floating-point rasters rather than integer rasters to resolve rounding errors when calculating population totals for larger areas (e.g. zonal sums)

Version 1.0 (11 November 2018) [https://dx.doi.org/10.5258/SOTON/WP00656]

- Original release of Nigeria population dataset

ASSUMPTIONS AND LIMITATIONS

These population estimates represent the time period of 2016 to 2017 corresponding to when the microcensus surveys were conducted. This model assumed that zero people lived in areas that were mapped as unsettled by LandScanHD v1.1 (ORNL 2018). The settlement map was based on satellite imagery mostly from 2014 with some images as old as 2010. Our population estimates assumed that zero people lived in areas that were mapped as unsettled based on these images.

Our population estimates assumed that no people lived in areas classified as non-residential settlements (e.g. industrial and commercial areas). This assumption was necessary because no microcensus data (defined below) were available from these areas to estimate the expected distribution of population densities.

Population estimates are missing from some areas near the Nigerian border because no data existed for some geospatial covariates in these areas.

SOURCE DATA

Nigeria Microcensus Survey (ORNL 2018)

This microcensus data set comes from household surveys at 1,142 locations in 15 states of Nigeria in 2016 and 2017. These locations represented a random sample stratified by settlement type. Each survey cluster contained about 3 hectares of settled area and the total number of people living in each household was recorded. We used the total number of people in each survey area (i.e. cluster) as the response variable in our statistical model.

LandScanHD v1.1 (ORNL 2018)

We used the Nigeria settlement classification from this data set (see Weber et al. 2018) to classify each grid cell into five different settlement types: urban (A), urban (B), urban (D), urban (F), rural (M), and non-residential (Z). We used this as a predictor of population density.

Nigeria Administrative Boundaries (eHealth Africa 2018, personal communication)
We obtained these boundaries from eHealth Africa in September of 2018. They represent the boundaries for 37 states and 774 local government areas. These are not official government boundaries.

WorldPop Global Gridded Population Estimates (WorldPop & CIESIN 2018)
We used WorldPop's 2014 gridded population estimates for Nigeria (the most recent available at the time) to develop covariates. WorldPop derived these gridded estimates from projections of the 1991 and 2006 Nigeria population and housing census. Columbia University's Center for International Earth Science Information Network (CIESIN) extrapolated the population totals for local government areas (level 2 administrative units) into future years (Lloyd et al., 2019) using subnational growth rates calculated at the state level (level 1 administrative units). WorldPop then disaggregated these state-level population projections to a 100 m grid using a machine learning approach (Stevens et al. 2015). We used these gridded estimates as a predictor of population density.

WorldPop Global Gridded Age-Sex Proportions (WorldPop et al 2018b)
We used WorldPop gridded age-sex proportions for Nigeria to produce gridded population estimates for each age-sex group. The WorldPop gridded age-sex proportions were produced using the methods of Pezzulo et al. (2017) and Carioli et al. (in prep). We multiplied our gridded population estimates (NGA_population_v1_2_gridded.tif) by the gridded age-sex proportions to produce NGA_population_v1_2_agesex.zip.

Demographic and Health Survey (National Population Commission and ICF International 2014)
We used the household sizes from this survey to create an interpolated map covering Nigeria with estimates of average household sizes for each grid cell. We used this as a predictor of population density.

Map of Schools in Nigeria (Geopode 2018)
We used a map of schools in Nigeria compiled by eHealth Africa to derive gridded estimates of school densities within a 1 km radius. Because the source data were produced with uneven mapping effort among regions, we rescaled our gridded estimates relative to the average school density within a 50 km radius.

METHODS OVERVIEW

Building on previous population estimation work in Nigeria (Weber et al. 2018), we adopt the model-based approach of Wardrop et al. (2018). We developed a statistical model to estimate population sizes for every 100 m grid cell across Nigeria. These estimates were based on relationships observed between microcensus surveys that enumerated people at a sample of 1,142 locations nationally and high-resolution geospatial datasets that have complete national coverage. These relationships provided a basis for extrapolating population estimates to areas where no population data were available and providing reliable estimates of uncertainty.

We developed a hierarchical Bayesian regression model within the family of Poisson generalised linear mixed models. This included a random intercept that estimated population densities for specific settlement types (ORNL 2018) in each region, state, and local government area (boundaries from eHealth Africa 2018, personal communication). The model also included a linear regression that estimated effects of the following geospatial covariates on population densities:

1. Projected, gridded population density (WorldPop 2018a)
2. Household size (National Population Commission and ICF International, 2014)
3. Residential settlement area within 1 km (ORNL 2018)
4. Non-residential settlement area within 1 km (ORNL 2018)
5. School density within 1 km (Geopode 2018)

* 1-2 were rescaled (mean = 0, standard deviation = 1) using: (x – mean(x)) / sd(x)
** 3-5 were rescaled as above but with mean and standard deviation calculated within a 50km radius.

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). Model-based estimates of population densities for microcensus survey areas that were withheld from the model had an r-squared value of 0.46 indicating moderate model fit at the spatial scale of microcensus clusters (approximately 3 hectares of settlement each). The confidence intervals adequately quantified prediction uncertainty (i.e. they included the observed population sizes most of the time). We expected better prediction accuracy for larger areas (e.g. local government areas and states) but we did not have validation data to assess model fit at these scales. Complete model diagnostics are available upon request (release@worldpop.org).

WORKS CITED

Carioli A, Pezzulo C, Hanspal S, Hilber T, Hornby G, Kerr D, Tejedor-Garavito N, Nielsen K, Pistolesi L, Adamo S, Mills J, Nieves JJ, Chamberlain H, Bondarenko M, Lloyd C, Yetman G, Gaughan A, Stevens F, Linard C, James W, Sorichetta A, Tatem AJ. *In prep*. Population structure by age and sex: a multi-temporal subnational perspective.

Denwood MJ. 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

Geopode. 2018. Map of schools in Nigeria. http://geopode.world Accessed September 2018.

Lloyd, C.T., Chamberlain, H., Kerr, D., Yetman, G., Pistolesi, L., Stevens, F.R., Gaughan, A.E., Nieves, J.J., Hornby, G., MacManus, K., Sinha, P., Bondarenko, M., Sorichetta, A., and Tatem A.J., 2019.Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. Big Earth Data, 3(2), 108-139doi:10.1080/20964471.2019.1625151

National Population Commission (NPC) [Nigeria] and ICF International. 2014. *Nigeria Demographic and Health Survey 2013*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International.

[ORNL] Oak Ridge National Laboratory. (2018). *LandScan HD: Nigeria version 1.1*. Oak Ridge National Laboratory: Oak Ridge, TN, USA.

Pezzulo C, Hornby GM, Sorichetta A, Gaughan AE, Linard C, Bird TJ, Kerr D, Lloyd CT, Tatem AJ. 2017. Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. Sci. Data 4:170089 https://dx.doi.org/10.1038/sdata.2017.89

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.

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

Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE* 10 (2):e0107042. https://doi.org/10.1371/journal.pone.0107042

Weber EM, Seaman VY, Stewart RN, Bird TJ, Tatem AJ, McKee JJ, Bhaduri BL, Moehl JJ, and Reith AE. 2018. Census-independent population mapping in northern Nigeria. *Remote Sensing of Environment* 204: 786–98. https://doi.org/https://doi.org/10.1016/j.rse.2017.09.024

WorldPop (www.worldpop.org - 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 (2018a). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). https://dx.doi.org/10.5258/SOTON/WP00645

WorldPop (www.worldpop.org - 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 (2018b). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).