Release Statement

Bottom-up gridded population estimates for Nigeria, version 2.0

17 November 2021

Original Release: 10 July 2019

ABSTRACT

This data release provides gridded population estimates (spatial resolution of 3 arc-seconds, approximately 100 m grid cells) with national coverage for Nigeria, along with estimates of the number of people belonging to various age-sex groups. Version 2.0 is an update to the previous version 1.2 gridded population estimates and is based on more recent and detailed settlement information and a different regional boundary definition. These model-based population estimates most likely represent the time period around 2019, corresponding to the period when the satellite imagery was processed to generate building footprints. Populations are mapped only into areas where residential settlements are predicted.

These data were produced by the WorldPop Research Group at the University of Southampton in collaboration with the National Population Commission of Nigeria. This work was part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme with funding from the Bill and Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (INV 009579, formerly OPP 1182425). Project partners included the United Nations Population Fund, Center for International Earth Science Information Network (CIESIN), a center within the Columbia Climate School at Columbia University, and the Flowminder Foundation. Statistical modelling was led by Chris Jochem and Doug Leasure additional support and oversight from Attila Lazar and Andy Tatem. Chris Lloyd provided the residential building classification. The microcensus data were originally collected by eHealth Africa and Oak Ridge National Laboratory with support from the Bill and Melinda Gates Foundation. The WorldPop group and GRID3 partners are acknowledged for their project support.

The authors followed rigorous procedures designed to ensure that the used data, the applied method and thus the results are appropriate and of reasonable quality. If users encounter apparent errors or misstatements, they should contact WorldPop at release@worldpop.org.

WorldPop, University of Southampton, and their sponsors offer these data on a "where is, as is" basis; do not offer an express or implied warranty of any kind; do not guarantee the quality, applicability, accuracy, reliability or completeness of any data provided; and shall not be liable for incidental, consequential, or special damages arising out of the use of any data that they offer.

RELEASE CONTENT

1. NGA_population_v2_0_gridded.zip
2. NGA_population_v2_0_admin.zip
3. NGA_population_v2_0_agesex.zip
4. NGA_population_v2_0_sql.sql
5. NGA_population_v2_0_mastergrid.tif
6. NGA_population_v2_0_tiles.zip

LICENSE

SUGGESTED CITATIONS

WorldPop and National Population Commission of Nigeria. 2021. Bottom-up gridded population estimates for Nigeria, version 2.0. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00729.

FILE DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984: EPSG 4326).

NGA_population_v2_0_gridded.zip

This zip file contains the following two raster files:

NGA_population_v2_0_gridded.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimates of total population size per grid cell across Nigeria. NA values represent areas that were mapped as unsettled based on a combination of gridded settlement layers derived from building footprints (Dooley et al., 2020) and a predictive model of residential vs non-residential building status (Lloyd et al., 2020). These data are stored as floating point numbers rather than integers to avoid rounding errors in aggregated population totals for larger areas.

NGA_population_v2_0_uncertainty.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100 m), contains estimates of uncertainty in the population estimates for each 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. 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 or through the woprVision application (https://apps.worldpop.org/woprVision/).

NGA_population_v2_0_admin.zip
This zip file contains the following five files with population totals for administrative units in Nigeria. The administrative boundaries used in the model are the same as released with version 1.2 of the population release (WorldPop, 2019). These are not official boundaries. In addition, the gridded population estimates have been summarised to the Ward and Local Government Area (LGA) boundaries available from the GRID3 Nigeria Data Portal (https://grid3.gov.ng/). The attribute tables for the shapefiles and the corresponding comma-separated (*.csv) files contain the estimates of the total population in each polygon and the confidence intervals. The intervals include the mean and standard deviation of the posterior prediction (columns labelled *mean* and *sd*, respectively) as well as the quantiles of the posterior predictions (columns labelled *q025*, *q05*, *q25*, *q50*, *q75*, *q95*, *q975*). The median is *q50* and the 95% credible intervals are described by *q025* and *q975*.

GRID3_NGA_population_v2_0_admin_LGA.shp
This shapefile contains the Local Government Area (LGA) boundaries from the GRID3 Nigeria portal. Note this file is accompanied by ancillary files (.dbf, .prj, .shx).

GRID3_NGA_population_v2_0_admin_Ward.shp
This shapefile contains the Ward boundaries from the GRID3 Nigeria portal. Note: this file is accompanied by ancillary files (.dbf, .prj, .shx).

NGA_population_v2_0_admin_level1_boundaries.shp
This shapefile contains the state boundaries used in the model. Note this file is accompanied by ancillary files (.cpg, .dbf, .prj, .sbn, .sbx, .shx).

NGA_population_v2_0_admin_level2_boundaries.shp
This shapefile contains the Local Government Area boundaries used in the model. Note: this file is accompanied by ancillary files (.cpg, .dbf, .prj, .sbn, .sbx, .shx).

GRID3_NGA_population_v2_0_admin_LGA.csv
This comma-separated text file contains summary statistics of the posterior predictions for population in Local Government Areas (LGA) in Nigeria (see GRID3 LGA admin unit boundaries, above).

GRID3_NGA_population_v2_0_admin_Ward.csv
This comma-separated text file contains summary statistics of the posterior predictions for population in Wards in Nigeria (see GRID3 Ward admin unit boundaries, above).

NGA_population_v2_0_admin_level0.csv
This comma-separated text file contains summary statistics of the posterior predictions for the total population Nigeria. The national population total is the sum of all lower administrative units.

NGA_population_v2_0_admin_level1.csv
This comma-separated text file contains summary statistics of the posterior predictions for the total population of the 36 states and the Federal Capital Territory in Nigeria (see admin level 1 boundaries, above).

NGA_population_v2_0_admin_level2.csv
This comma-separated text file contains summary statistics of the posterior predictions for the total populations of the 774 Local Government Areas in Nigeria (see admin level 2 boundaries, above).

NGA_population_v2_0_agesex.zip
This zip file contains 40 geotiff rasters at a spatial resolution of 3 arc-seconds (approximately 100 m). Each raster provides gridded population estimates for an age-sex group per grid cell across Nigeria. 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 by 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 from the WorldPop project (WorldPop et al., 2018). The age-sex proportions were applied to the gridded population estimates (NGA_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

NGA_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 source data for the woprVision web application (https://wopr.worldpop.org/woprVision) and it can be queried using the wopr R package (Leasure et al. 2020).

The SQLite database contains a single table (Nhat) that includes the population predictions. This table contains the following columns:

- “cell” contains a cell ID to identify the location. Cell IDs correspond to those the cell IDs of NGA_population_v2_0_mastergrid.tif.

- "x" and "y" columns contain WGS84 coordinates for the centroid of the grid cell.
- "Pop" column contains a comma-separated string of population estimates which are the MCMC samples from the predicted posterior distribution for the population estimate in that grid cell.
- "agesexid" column contains the region ID for the age-sex proportions that are provided in NGA_population_v2_0 _agesex_table.csv and NGA_population_v2_0 _agesex_regions.tif.
- "area" contains the total residential settled area in hectares. This corresponds to the total building area from Dooley et al. (2020) in their raster NGA_buildings_v1_1 _total_area.tif and residential classification model following the methods of Lloyd et al. (2020).

NGA_population_v2_0_mastergrid.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100 m), contains 1s for each grid cell across the study area. NAs represent grid cells considered to be unsettled or outside the study area. Cell IDs from this raster correspond to the "cell" column in the SQLite database (above).

NGA_population_v2_0_tiles.zip

This tiled web map allows for rapid display of the gridded population estimates across the study area (i.e. NGA_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 study area for the zoom levels 1 to 14. These tiles are source data for the woprVision web application (https://wopr.worldpop.org/woprVision).

RELEASE HISTORY

- **Version 2.0** (17 November 2021)[https://dx.doi.org/10.5258/SOTON/WP00729]
    - Refinement of gridded population estimates using more recent settlement data based on building footprints.
    - Predictions of residential and non-residential buildings incorporated in the settlement map.
    - A different regional boundary definition was used in the model, corresponding with Nigerian statistical regions.
    - Representative of the year 2019.

- **Version 1.2** (15 September 2020)
    - A peer-reviewed article (Leasure et al., 2020b) was added to describe the statistical methods that were developed to produce the population estimates (https://doi.org/10.1073/pnas.1913050117).

- **Version 1.2** (20 May 2020)
    - Gridded population estimates were added to NGA_population_v1_2_agesex.zip for the following demographic groups: children under 1, children under 5, children under 15, and women 15 to 49 years of age.

- **Version 1.2** (26 March 2020) [https://dx.doi.org/10.5258/SOTON/WP00661]
    - 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.
    - Representative of the time period from 2016 to 2017.

- **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 mostly likely represent the time period of 2018 to 2019, reflecting when the majority of the satellite imagery scenes were processed to produce building footprints. This settlement data primarily determines the spatial distribution of the gridded population estimates. The updated settlement layer used in Version 2.0 means that the predicted population distribution has changed compared with previous data releases. The variation in population density is modelled based on microcensus data, which were collected between 2016 and 2017. This model assumes that population densities observed during the earlier time period are still representative of the more recent period. Similarly, the age and sex structures data come from an earlier time point and were assumed to be stable.

Furthermore, the population estimates assume that no people live in areas that are predicted to be either unsettled or non-residential. Compared with version 1.2, the residential/non-residential classification used in the version 2.0 estimates is, in general, less restrictive. It considers more areas to be potential residential in order to avoid excluding and underestimating the population in urban centres. This could lead to over- or under-estimates in misidentified areas.

Due to differences in the spatial data of administrative boundaries, population estimates may be missing from some areas near the Nigerian border.

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 (Weber et al., 2018). 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.

WorldPop Global Gridded Age-Sex Proportions (WorldPop et al., 2018)

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_v2_0_gridded.tif) by the gridded age-sex proportions to produce NGA_population_v2_0_agesex.zip.

Settlement Data

Settled grid cells and the number of buildings per grid cell were defined as a raster layer (Dooley et al., 2020). These gridded layers were derived from building footprint polygons extracted from high-resolution satellite imagery (Ecopia.AI and Maxar Technologies, 2020). Because the building footprint-derived settlement layers do not identify different structure types, we predicted the proportion of residential and non-residential buildings (e.g. warehouses, industrial areas). Note that mixed-use structures are classified as residential. The classification method used is described by Lloyd et al. (2020). Additionally, we used a gridded neighbourhood classification map (Jochem et al., 2020) as a predictor of population density. The six different settlement types in this map reflect differences in local morphological patterns of the building footprints.

Demographic and Health Survey (National Population Commission Nigeria and ICF International, 2019)

We used the household sizes from the 2018 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.

Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar (SAR) (Mullissa et al., 2021)

We extracted single co-polarization bands (VV and VH) from the analysis-ready datasets in Google Earth Engine (Gorelick et al., 2017). SAR data can provide information on surface type and roughness which can be related to building density or building volume, and we use this as a predictor in the model of population density.

METHODS OVERVIEW

Building on the previous gridded population model for Nigeria (Leasure et al., 2020b; WorldPop, 2019), we implemented a model-based approach to generate a "bottom-up" estimate (Wardrop et al., 2018) of the population size in a gridded format (approximately 100 m x 100 m) for all of Nigeria. The statistical model is based on the relationship between population, observed at 1142 microcensus survey locations collected in 2016 – 2017, and high-resolution geospatial datasets with national coverage. These relationships provide a basis to predict the population in areas where no populations were enumerated.

Following Leasure et al. (2020b), we developed a Poisson-lognormal regression model. This formulation included a random intercept that estimated the population density in a hierarchy of spatial units, specifically, within settlement types (Jochem et al., 2020), statistical regions (defined as groups of states), states, and local government areas. The model included a linear regression that estimated the effects of three geospatial covariates on population density:

- Household size (NPC and ICF, 2019)
- Sentinel-1 SAR: C-band Synthetic Aperture Radar VV and VH bands (Mullissa et al., 2021)

We used Bayesian inference methods and implemented the model using R (R Core Team, 2021) Stan (Stan Development Team, 2021) and the R interface package RStan (Stan Development Team, 2020). The model-based estimates of the population in the microcensus locations (i.e. the in-sample fit) had a squared correlation (R^2) of 0.47, indicating a reasonable model fit at this spatial scale. Additionally, the credible intervals contained the observed population sizes for most observations, suggesting that the model adequately quantified the uncertainty. The posterior predictions were used to estimate administrative totals (and uncertainties) by aggregating the individual grid cell predictions within the administrative boundaries. As a result of the changing settlement data and residential/non-residential classification, more areas of Nigeria have a predicted population in this data release.

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.

Dooley CA, Boo G, Leasure DR, Tatem AJ. 2020. Gridded maps of building patterns throughout sub-Saharan Africa, version 1.1. WorldPop, University of Southampton. Source of building footprints: Ecopia Vector Maps Powered by Maxar Satellite Imagery (C) 2020. doi:10.5258/SOTON/WP00677. Data available from https://wopr.worldpop.org/?NGA/Buildings/v1.1.

Ecopia.AI and Maxar Technologies. 2020. Digitize Africa data. Ecopia.AI and Maxar Technologies.

Gorelick, N, Hancher, M, Dixon, M, Ilyushchenko, S, Thau, D, & Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. doi:10.1016/j.rse.2017.06.031.

Jochem, WC, Leasure, DR, Pannell, O, Chamberlain, HR, Jones, P, Tatem, AJ. 2020. Classifying settlement types from multi-scale spatial patterns of building footprints. *Environment and Planning B: Urban Analytics and City Science* 48(5), 1161-1179. https://doi.org/10.1177%2F2399808320921208.

Leasure DR, Bondarenko M, Darin E, Tatem AJ. 2021. wopr: An R package to query the WorldPop Open Population Repository, version 1.0.4. WorldPop, University of Southampton. doi: 10.5258/SOTON/WP00716. Source code available from https://github.com/wpgp/wopr.

Leasure DR, Jochem WC, Weber EM, Seaman V, Tatem AJ. 2020b. National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1913050117.

Lloyd, CT, Sturrock, JJ, Leasure DR, Jochem, WC, Lazar, AN, Tatem, AJ. 2020. Using GIS and machine learning to classify residential status of urban buildings in low and middle income settings. *Remote Sensing* 12923), 3847. https://doi.org/10.3390/rs12233847.

Mullissa, A, Vollrath, A, Odongo-Braun, C, Slagter, B, Balling, J, Gou, Y, Gorelick, N, Reiche, J. 2021. Sentinel-1 SAR Backscatter analysis ready data preparation in Google Earth Engine. *Remote Sensing* 13, 1954. doi:10.3390/rs13101954.

National Population Commission (NPC) [Nigeria] and ICF. 2019. *Nigeria Demographic and Health Survey 2018.* Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF.

[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.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Stan Development Team. 2020. RStan: The R interface to Stan. R package version 2.21.2. https://mc-stan.org.

Stan Development Team. 2021. *Stan Modeling Language Users Guide and Reference Manual*, 2.28. https://mc-stan.org.

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.Y., Stewart, R.N., Bird, T.J., Tatem, A.J., McKee, J.J., Bhaduri, B.L., Moehl, J.J., Reith, A.E. 2018. Census-independent population mapping in northern Nigeria. *Remote Sensing of Environment* 204, 786-798. 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 (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076).

WorldPop. 2019. Bottom-up gridded population estimates for Nigeria, version 1.2. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00655.