Release Statement: 21/03/2019 Nigeria Population Estimates v1.1

Release Date: 21/03/2019

Release Content

GRID3_NGA_PopEst_v1_1_mean_float.tiff

- Raster of mean population estimates (~100 m grid)

GRID3_NGA_PopEst_v1_1_uncert.tiff

- Raster of uncertainty estimates (~100 m grid). Uncertainty is defined as the difference between the upper 95% confidence interval and the lower 95% confidence interval divided by the mean prediction.

GRID3_NGA_PopEst_v1_1_table_national.csv

- Table with population totals and confidence intervals for Nigeria nationally.

GRID3_NGA_PopEst_v1_1_table_state.csv

- Table with population totals and confidence intervals for Nigerian states.

GRID3_NGA_PopEst_v1_1_table_LGA.csv

- Table with population totals and confidence intervals for Nigeria local government areas.

This document explains the GRID3 Nigeria population model. It describes the datasets that have been used to build the model, highlights the model results and explains how the data should be used.

This work was undertaken by the Flowminder Foundation and WorldPop at the University of Southampton as part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) initiative in Nigeria. GRID3 is a multi-country and multi-donor partnership that aims to support collection, storage and application of geospatial data for informed decision-making across target countries.

The initiative is funded by the Bill & Melinda Gates Foundation (BMGF) and the United Kingdom's Department for International Development (DFID). It is implemented by WorldPop/Flowminder (WPFM) and the United Nations Population Fund (UNFPA), and coordinated by the Center for International Earth Science Information Network (CIESIN).

Version history

Nigeria Population Estimates v1.1 contains floating point values (with decimal places), instead of integers (rounded to the nearest whole number) released with v1.0.

Five LGAs were missing from the previously released tables of population totals (version 1.0, 20th November 2018). The missing LGAs were Ifelodun (Osun state), Irepodun (Osun state), Benue (Obi state), Surulere (Lagos state), and Bassa (Kogi state). These population totals were mistakenly combined with LGAs having the same name in other states. This caused those LGAs to have inflated totals: Ifelodun (Kwara state), Irepodun (Kwara state), Obi (Nasarawa state), Surulere (Oyo state), Bassa (Plateau state). The current data release corrects this mistake. This error also affected the state totals which have now been corrected. The error did not affect the underlying gridded population estimates.

We have identified two LGAs in Nasarawa state with significantly overlapping boundaries: Keana and Awe LGAs. We are investigating how this may have affected the gridded population estimates and the population totals for these LGAs. In the meantime, results from these two LGAs should be used with caution.

Note: The administrative boundaries used here do not reflect official government boundaries. They were obtained from eHealth Africa in October 2018 and were the most recent operational administrative boundaries available at the time.

Methodology

This section provides a brief summary of the methodology on how the population estimates are derived.

Approach

Building on previous population estimation work in Nigeria (Weber et al. 2018), these results take the model-based approach to estimating populations described by Wardrop et al. (2018). The model combines information on population density from small microcensus surveys (n=1,142) with high-resolution geospatial datasets, and analyses the relationships between population characteristics and geospatial data in the microcensus enumeration zones. This provides a basis for extrapolating population estimates to areas where no census has been done but covariate data are observed.

The model used is 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, regions, states, and local government 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 100 metre grid cells throughout the country. 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 Covariates

Important datasets for the current GRID3 model include settlement types from Weber et al. (2018), schools from eHealth Africa, global gridded population estimates from WorldPop 2014, and household sizes derived from Demographic and Health Survey (DHS) data. The covariates that are currently included in the model are:

- **The WorldPop Global gridded population estimates**: The 2014 estimates are being used currently, but more recent estimates will be used as they become available (Stevens et al., 2015).
- **Settlement Type**: The settlement type raster for Nigeria derived from data within the LandScan HD: Nigeria version 1.1 geodatabase (Oak Ridge National Laboratory, 2018) and described in Weber et al. (2018).
- **Settlement context in surrounding areas**: Derived from data in the LandScan HD: Nigeria version 1.1 geodatabase (Oak Ridge National Laboratory, 2018) and the Weber et al. (2018) map of settlement types.
- **Gridded household size**: Interpolated from Demographic and Health Survey data (National Population Commission and ICF International, 2014)
- **School density**: Derived from eHealth in Nigeria via http://geopode.world/.

Quality checks completed (details available on request):

1. Convergence statistics
2. Sensitivity analysis for priors
3. Initial model comparisons
4. Posterior predictive check (i.e. Bayesian p-value)
5. In-sample residuals analysis: bias, imprecision, and inaccuracy
6. Plots:
    a. Estimated covariate effects
    b. In-sample observed vs. predicted (with CI)
    c. In-sample residuals vs. predicted
    d. Spatial correlogram and Moran's I for model residuals
7. Maps:
    a. In-sample prediction error
    b. Residual variance
8. Cross-validation residuals: bias, imprecision, and inaccuracy
9. Plots:
    a. Cross-validation observed vs. predicted (with CI)
    b. Cross-validation residuals vs. predicted
10. Maps
    a. Cross-validation prediction error

11. Comparisons with existing population estimates (e.g. UN Pop Division; US Census; national government numbers) and explanation of differences.

Known data limitations:

It is recommended to use the pre-calculated table totals where aggregation is required (i.e. summing pixel values to calculate a population total for an area). The aggregation used the eHealth administrative boundaries (as of October 2018). Using other administrative boundaries will result in different population totals. **Please refer to the tables for pre-calculated confidence intervals.**

Aggregation of the mean population raster using a zonal statistics tool will not provide confidence intervals. If zonals statistics are calculated, the mean population raster should be used as the snap raster and the zonal boundaries should be rasterized using the same cell size as the mean population raster to avoid introducing error in population totals from the zonal statistics calculation.

Population estimates are missing from some areas near the Nigerian border as no data exist for some settlement data and geospatial covariates in these areas. Future updates will expand population estimates to include these areas wherever possible.

The current model assumes that no people live in areas classified as "non-residential settlements". This assumption is necessary because no microcensus data have been collected from these areas historically. Microcensus surveys have recently been conducted in Lagos and Kaduna states that included non-residential settlements and similar microcensus surveys will soon be conducted in Enugu state. This is an initial step towards estimating population sizes in areas mapped as non-residential rather than assuming they are zero.

Future versions of the model will work to improve population estimates by addressing these known issues. Work will also be ongoing to perform model assessments, include new microcensus data as they become available, and identify new geospatial data that are good predictors of population density. Feedback from programme partners and data users will be important to help find areas needing improvement and to work towards solutions.

Acknowledgements:

We thank Oak Ridge National Laboratory for providing the microcensus data and settlement maps used in the model.

This work was supported with funding from the Bill & Melinda Gates Foundation (OPP1106427, OPP1182408).

We thank the WorldPop "Global High-Resolution Population Denominators" Project, funded by the Bill & Melinda Gates Foundation (OPP1134076), for providing access to the 2014 population estimates.

We also thank eHealth Africa for their work on conducting the microcensus.

References

Denwood, Matthew 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.

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.

Oak Ridge National Laboratory. (2018). *LandScan HD: Nigeria version 1.1* . [Dataset].

Plummer, Martyn. 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 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). Public Library of Science: 1–22. doi: 10.1371/journal.pone.0107042 .

Weber, Eric M., Vincent Y. Seaman, Robert N. Stewart, Tomas J. Bird, Andrew J. Tatem, Jacob J. McKee, Budhendra L. Bhaduri, Jessica J. Moehl, and Andrew E. Reith. 2018. "Census-Independent Population Mapping in Northern Nigeria." *Remote Sensing of Environment* 204: 786–98. doi: https://doi.org/10.1016/j.rse.2017.09.024 .

Release Check Record:

Modelling and development: Doug Leasure 21.03.2019
Product: Michael Harper, 21.03.2019
Documentation: Michael Harper 21.03.2019
Science: Andy Tatem, 21.03.2019
Communications: Sophie Delaporte, 21.03.2019
Director: Cathy Riley, 21.03.2019
CIM: Tracy Adole, 21.03.2019