Release statement

Gridded Disaggregated Population Estimates for Uganda, version 2.0.

20 November 2023

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 United Nations Children's Fund (UNICEF) - Population Modelling for use in Routine Health Planning and Monitoring project (contract no. 43335861). Projects partners included the Uganda Unicef Regional and Country Offices, WorldPop research group at the University of Southampton and the Center for International Earth Science Information Network in the Columbia Climate School at Columbia University. Assane Gadiaga (WorldPop) led the input processing and the modelling work following the Random Forest (RF)-based dasymetric mapping approach developed by Stevens et al. (2015). Thomas Abbott supported the covariates processing work. In-country engagements were done by Catherine Kabahuma, Silvia Reen, and Maria Muniz (Unicef). Using the 2014 census data from the Uganda Bureau of Statistics (UBOS), the US Census Bureau released the census-based total population projections, population by age and sex and digital District boundaries. Duygu Cihan helped in the preparation of these input population data. Attila N Lazar, Edith Darin, Maksym Bondarenko and Heather Chamberlain advised on the modelling procedure. The work was overseen by Attila N Lazar and Andy J Tatem.

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

UGA_population_v2_0_gridded.tif

UGA_population_ v2_0_agesex.zip

UGA_population_ v2_0_mastergrid.tif

LICENSE

SUGGESTED CITATIONS

Gadiaga A. N., Abbott T. J., Darin E., Lazar A. N., Chamberlain H., Tatem A. J. 2023. Census disaggregated gridded population estimates for Uganda (2022), version 2.0. University of Southampton. 10.5258/SOTON/WP00769

FILES DESCRIPTIONS

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

UGA_population_v2_0_gridded.tif

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

UGA_population_v2_0_agesex.zip

This zip file contains 40 GeoTIFF raster files representing estimated population counts for specific age and sex groups within grid cells of approximately 100m. We provide 36 raster files 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 raster files 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 “f1549” (all females between the ages of 15 and 49, inclusive).

These data were produced *post-hoc* by multiplying the total population counts provided in the *UGA_population_v2_0_gridded.tif* raster with the age and sex proportions for each sub-county. 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.

UGA_population_v2_0_mastergrid.tif

This geotiff raster contains the rasterised administrative units used to perform the population disaggregation, with a spatial resolution of approximately 100m grid cell (0.0008333 decimal degrees). The pixel values are IDs referring to the administrative boundary polygons (sub-counties) that match the corresponding units in the input population data.

RELEASE HISTORY

Version 1.0 (12 September 2020) doi:10.5258/SOTON/WP00682-
Original release of this data set.

Version 2.0 (20 November 2023) doi:10.5258/SOTON/WP00769 - The model was updated by using Ecopia building footprints version 2, combined with 2022 projected population totals at the district level.

SOURCE DATA

- Digital District-level boundary and their projected population totals and age/sex group totals for 2022 based on the 2014 Population and Housing census were provided by the US Census Bureau in a shapefile format.
- Gridded building patterns (building count, building total area, building mean area, building area variance, building density, building length, building mean length and building length variance) were derived from the latest Ecopia building footprints (Ecopia.AI and Maxar Technologies, 2020).
- Additional geospatial covariates, representing factors related to population distribution (distance to land cover maps, slope and elevation, motorized friction surface, walking friction surface, travel time to city, distance to coastline, protected areas, schools, health facility, place of worship, local roads, main roads, road intersection, and built settlement, and night-time lights), were created using the data sources listed in the appendix. Because building footprints contain different structure types, information on the residential status of the buildings were derived from the Global Human Settlement Layer (GHSL) and used to improve the prediction of gridded population counts.

METHODS OVERVIEW

Modelling: Following the Random Forest (RF)-based dasymetric mapping approach (Stevens et al., 2015), the popRF 'R' package (Bondarenko et al., 2021) based on Breiman (2001) algorithm was used to model District total population density as a combination of the geospatial covariates and then to estimate the total population density in each approximately 100 m grid cell (0.0008333 decimal degrees grid or 3 arc seconds). The set of the geospatial covariates included in the model could explain 82.6% of the total input population data variance. The list of used covariates is in the Appendix.

The gridded population estimates were then combined with the age/sex proportions calculated from the projections for Uganda (population projection) to produce gridded population estimates for each sex group (female and male) at regular age intervals.

ASSUMPTIONS AND LIMITATIONS

These datasets were produced based on the 2022 projected district-levelpopulation totals for Uganda using the 2014 Population and Housing Census. Although the enumerated population totals have been projected to 2022, the estimate of population in each District may not accurately reflect the current population, given the time elapsed since the last census and the necessary assumptions made in projecting the population estimates.

The gridded population estimates are constrained within the settled areas derived from gridded building metrics. We assumed that the building footprint data (Ecopia.AI and Maxar Technologies, 2020), from which the gridded building metrics were derived, is accurate and that each building polygon corresponds to a building structure. In addition, the distribution of buildings might not represent the current building landscape because of the necessity to use satellite imagery from different years in extraction of the building footprints (e.g. due to cloud coverage). The images used for building extraction were predominantly from 2018-2020. In locations which have recently experienced rapid settlement changes, for example, establishment of new settlements, rapid urban

growth or abandonment of settlements, the population estimates are likely to be less accurate. Efforts were made to use recent and up-to-date covariates that match the temporal domain of the input population data. However, mismatches still remain for some covariates between the year of the input population data (e.g. building footprints, some distance-based covariates) , which may lead to less accurate population estimates.

Lastly, we advise that although these raster datasets were produced through very rigorous statistical processes already outlined above, they should be used with caution especially when used in combination with national boundaries because the gridded population data may not cover some areas within user boundaries at edge location.

WORKS CITED

Breiman, L. Random forests. Mach. Learn. 45, 5–32 (2001).

Bondarenko M., Nieves J.J., Forrest R.S., Andrea E.G., Jochem C., Kerr D., and Sorichetta A. (2021): popRF: Random Forest-informed Population Disaggregation R package DOI:10.5258/SOTON/WP00715

Ecopia.AI and Maxar Technologies. 2020. Digitize Africa data. *http://digitizeafrica.ai*

Stevens, F. R., Gaughan, A. E., Linard, C. & Tatem, A. J. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10, e0107042 (2015). *https://doi.org/10.1371/journal.pone.0107042*

US Census Bureau. Subnational Population Data by Geographic Area. *https://www.census.gov/geographies/mapping-files/time-series/demo/international-programs/subnationalpopulation.html.* Accessed November 2, 2023.

APPENDIX

List of covariates

Covariate name	Source	Link
Buildings area (coefficient of variation)	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings length (coefficient of variation)	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings mean area	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings mean length	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings total area	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings total length	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings count	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Buildings density	WorldPop/Ecopia	Ecopia Map Platform (ecopiatech.com)
Distance to coastline	WorldPop	https://www.worldpop.org/project/categories?id=16
Distance to Water bodies	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to place of education	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to cultivated areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to woody areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to shrub area edges	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to herbaceous areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to sparse vegetation areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to aquatic vegetation areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to Urban area	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to bare areas	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to local roads	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to major roads	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to health points	Open Street Map	https://www.geofabrik.de/data/download.html
Slope	Worldpop	https://www.worldpop.org/project/categories?id=14
Elevation	WorldPop	https://www.worldpop.org/project/categories?id=15
Nighttime lights VIIRS	WorldPop	https://www.worldpop.org/project/categories?id=17
Distance to protected areas	Worldpop	https://www.worldpop.org/project/categories?id=18
Distance to primary road intersection	OSM/WorldPop	https://www.worldpop.org/geodata/listing?id=33
Distance to waterways	OSM/WorldPop	https://www.worldpop.org/geodata/listing?id=33

Motorized friction surface	Malaria Atlas Project	https://malariaatlas.org/research-project/accessibility-to-healthcare/
Total built-up surface	Global Human Settlement Layer	Global Human Settlement - Download - European Commission (europa.eu)
Non-residential built-up surface	Global Human Settlement Layer	Global Human Settlement - Download - European Commission (europa.eu)
Walking friction surface	Malaria Atlas Project	https://malariaatlas.org/research-project/accessibility-to-healthcare/