Release statement

Gridded disaggregated population estimates for Mozambique, version 2.0.

09 August 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 Mozambique 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). Amy Bonnie supported the covariates processing work. In-country engagement were done by Katia Quinhas, Sandra Baptista, Maria Muniz. The National Bureau of Statistics of Mozambique (INE) released the updated yearly census-based total population projection and projection by age-groups, and sex (female and male) using the results of the 2017 national census. In addition, the District-level administrative boundaries were shared by INE. Attila N Lazar and Edith Darin 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

MOZ_population_v2_0_gridded.tif

MOZ_population_v2_0_agesex.zip

MOZ_population_v2_0_mastergrid.tif

LICENSE

SUGGESTED CITATIONS

Gadiaga A. N., Bonnie A. L., Lazar A. N., Darin E., Tatem A. J. 2023. Census disaggregated gridded population estimates for Mozambique (2022), version 2.0. University of Southampton. doi:10.5258/SOTON/WP00764

FILES DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984) with a spatial resolution of approximately 100m grid cell (0.0008333 decimal degrees).

MOZ_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 Mozambique. NA values represent areas that were mapped as unsettled based on gridded building patterns derived from building footprints (after 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.

MOZ_population_v2_0_agesex.zip

This zip file contains 40 GeoTIFF rasters representing estimated population counts for specific age and sex groups within grid cells of approximately 100m. 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 old 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 years bins labelled with a “5”, “10”, etc. Eighty-year-old and over are represented in 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 “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 *MOZ_population_v2_0_gridded.tif* raster and age and sex proportions derived from the INE age-sex projections for each District. 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.

MOZ_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 (districts) that match the corresponding units in the input population data.

RELEASE HISTORY

Version 2.0 (09 August 2023) doi:10.5258/SOTON/WP00764- The model was updated by using Ecopia building footprints version 2, combined with 2022 subnational projections.

Version 1.1 (26 November 2020) doi:10.5258/SOTON/WP00672- The model was refit with updated building footprints (Ecopia.AI & Maxar Technologies 2020).

Version 1.0 (16 June 2020) This is the original limited release.

SOURCE DATA

- Digital District boundary and their projected population totals and age/sex group totals for 2022 based on the 2017 Population and Housing census were provided by the INE statistical office in a shapefile format (population projection) .
- 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 (Lloyd et al., 2019), representing factors related to population distribution (distance to land cover maps, mean precipitation and temperature, slope and elevation, motorized friction surface, walking friction surface, travel time to city, distance to coastline, protected areas, health facility, local roads, main roads, railway station, road intersection, and built settlement, and night-time lights), were created using the data sources listed in the appendix, and used to improve the prediction of the 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 model could explain 97% of the total population input variance. The list of used covariates is listed in the Appendix.

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

ASSUMPTIONS AND LIMITATIONS

This dataset was produced based on the projected 2022 population totals for Districts derived from the 2017 Population and Housing Census. Although the enumerated population totals have been projected to 2022, the estimate of population in each District may not 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

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, _Comprehensive R Archive Network (CRAN)_, url:https://cran.rproject.org/package=popRF.

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

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, C. A. and Tatem, A.J. 2020. Gridded maps of building patterns throughout sub-Saharan Africa, version 1.0. University of Southampton: Southampton, UK. Source of building Footprints "Ecopia Vector Maps Powered by Maxar Satellite Imagery"© 2020. https://dx.doi.org/10.5258/SOTON/WP00666.

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

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), 108-139.

https://dx.doi.org/10.1080/20964471.2019.1625151

Muñoz Sabater, J., (2021): ERA5-Land monthly averaged data from 1950 to 1980. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.68d2bb30

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.

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

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). https://dx.doi.org/10.5258/SOTON/WP00646

APPENDIX

List of covariates

Covariate name	Source	Link
Motorized friction surface	Malaria Atlas Project	https://malariaatlas.org/research-project/accessibility-to-healthcare/
Walking friction surface	Malaria Atlas Project	https://malariaatlas.org/research-project/accessibility-to-healthcare/
Distance to artificial surface edges	WorldPop	https://hub.worldpop.org/geodata/listing?id=56
Distance to main roads	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to roads	WorldPop	https://hub.worldpop.org/geodata/listing?id=31
Distance to OSM major road intersections	WorldPop	https://hub.worldpop.org/geodata/listing?id=33
Distance to Water bodies	Open Street Map	https://www.geofabrik.de/data/download.html
Distance to IUCN strict nature reserve and wilderness area edges	WorldPop	https://hub.worldpop.org/geodata/listing?id=35
Distance to ESA-CCI-LC inland water	WorldPop	https://hub.worldpop.org/geodata/listing?id=61
Distance to OSM major waterways	WorldPop	https://hub.worldpop.org/geodata/listing?id=34
Distance to herbaceous areas	WorldPop	https://hub.worldpop.org/geodata/listing?id=56
Distance to cultivated areas 2015	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to woody areas 2015	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to shrub area edges 2015 (130)	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to sparse vegetation areas 2015	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to aquatic vegetation areas 2015	WorldPop	https://www.worldpop.org/project/categories?id=14
Distance to bare areas 2015	WorldPop	https://www.worldpop.org/project/categories?id=14
Current average total annual precipitation	Copernicus	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=form
Current average annual temperature	Copernicus	https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=form
Slope	Worldpop	https://www.worldpop.org/project/categories?id=14
Elevation	WorldPop	https://hub.worldpop.org/project/categories?id=14
Distance to open-water coastline	WorldPop	https://hub.worldpop.org/project/categories?id=14

Nighttime lights 2020 VIIRS	National Centers for Environmental Information(NOAA)	https://ngdc.noaa.gov/eog/viirs/index.html
Buildings area (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)