Release Statement

Census disaggregated gridded population estimates for Zambia, based on the 2022 Census of Population and Housing preliminary results, version 2.0.

4th June 2024
Original Release: 4th June 2024

These data were produced by the WorldPop Research Group at the University of Southampton and the Zambia Statistics Agency (ZamStats). This work is part of the GRID3 (Geo-Referenced Infrastructure and Demographic Data for Development) project funded by the Bill and Melinda Gates Foundation (INV-045694).

The national gridded population dataset for Zambia (approximately 100m x 100m grid cell resolution) was created by disaggregating ward-level population numbers collected during the Zambia 2022 Census of Population and Housing. This work was undertaken collaboratively by GRID3 partners and ZamStats. The following individuals were closely involved in this work: Sikufele Mubita, Webster Sikalumbi, Salome Naluyele, Welani Simwinga and Hildah Chileshe (ZamStats), Heather Chamberlain and Thomas Abbott (WorldPop/GRID3), Olena Borkovska, Dr Garikai Membele and Chisenga Abel Musuka (GRID3).

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, they should contact WorldPop at release@worldpop.org.

WorldPop, the 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. ZMB_population_2022_v2_0_gridded_total.tif
2. ZMB_population_2022_v2_0_mastergrid.tif
3. ZMB_population_2022_v2_0_gridded_male.tif
4. ZMB_population_2022_v2_0_gridded_female.tif

LICENSE

SUGGESTED CITATION

WorldPop and The Zambia Statistics Agency. (2024) Census disaggregated gridded population estimates for Zambia, based on the 2022 Census of Population and Housing preliminary results, version 2.0. University of Southampton. DOI: 10.5258/SOTON/WP00783

FILE DESCRIPTIONS

All spatial data files are in geographic coordinate system WGS84 (World Geodetic System 1984: EPSG 4326).

ZMB_population_v2_0_gridded_total.tif

This geotiff raster contains estimates of total population for each approximately 100m grid cell (0.0008333 decimal degrees, 3 arc seconds) across Zambia. NA values represent grid cells considered as not being populated, based on an absence of household points in the household location data collected by ZamStats as part of the census.

ZMB_population_v2_0_mastergrid.tif

This geotiff raster provides a binary classification of grid cells considered to be populated. For each approximately 100m grid cell (0.0008333 decimal degrees, 3 arc seconds), values of 1 are assigned based on the household points in the household location data collected by ZamStats as part of the census. Values of 0 indicate grid cells that did not contain households and were therefore assumed to be unpopulated. NA values indicate grid cells considered as outside the study area as defined by the national ward boundary dataset.

ZMB_population_v2_0_gridded_male.tif

This geotiff raster contains estimates of total male population for each approximately 100m grid cell (0.0008333 decimal degrees, 3 arc seconds) across Zambia. NA values represent grid cells considered as not being populated, based on an absence of household points in the household location data collected by ZamStats as part of the census.

ZMB_population_v2_0_gridded_female.tif

This geotiff raster contains estimates of total female population for each approximately 100m grid cell (0.0008333 decimal degrees, 3 arc seconds) across Zambia. NA values represent grid cells considered as not being populated, based on an absence of household points in the household location data collected by ZamStats as part of the census.

RELEASE HISTORY

Version 2.0 (this release – 4th June 2024) [doi: 10.5258/SOTON/WP00783]

- This version of the dataset is based on enumerated population counts from the 2022 National Census preliminary results. A "top-down" modelling approach has been used to spatially disaggregate population counts from ward-level to provide gridded estimates of population (total, male and female only). No age-structured population estimates are included.

Version 1.0 (7th April 2020) [doi: 10.5258/SOTON/WP00662]

- Original release of modelled (census-independent) population estimates for Zambia. These modelled estimates were developed using a "bottom-up" modelling approach, during the intercensal period, and utilised multiple sources of routinely collected geolocated survey data. This release included gridded population estimates for a range of age-sex classes.

SOURCE DATA

The key datasets used to produce the gridded population estimates were:

- **Population counts** from the 2022 National Census (preliminary results) at **ward-level**, including total population and disaggregated in terms of sex (Source: ZamStats).
- **Administrative boundaries** in shapefile format for wards (Source: ZamStats).
- **Household locations**, represented as point locations, as collected during the 2022 National Census (Source: ZamStats). Aside from location, no additional attribute information was included in this dataset.
- **Additional geospatial covariate datasets** representing factors related to population distribution and density. These included gridded building metrics (building count, building total area, building mean area, building density, building total perimeter and building mean perimeter) derived from Ecopia "year 2" building footprints (Dooley et al., 2021).

In addition, a sample of household locations for 10 districts, with associated population counts, was used in validating the gridded population estimates at a range of spatial scales.

METHODS OVERVIEW

The approach to create gridded population estimates included the following stages:

Data pre-processing: Checking and cleaning of the National Census data was undertaken by ZamStats as part of the census process. Subsequently, the mastergrid for this work was created based on the national ward boundaries dataset, with grid cell alignment matching the WorldPop Global 2000-2020 population data products (WorldPop and CIESIN, 2018). In the mastergrid, only grid cells with one or more household points were included in the population disaggregation process (i.e. the gridded population estimates were spatially constrained to grid cells with one or more household points). In the supplied mastergrid file (**ZMB_population_v2_0_mastergrid.tif**), grid cells within the national extent were assigned a value of 1 if the grid cell had one or more household points, and a value of 0 if there were no household points within the grid cell. All geospatial covariates were adjusted to have the same grid cell alignment and spatial extent.

Imputation of building footprint metrics was necessary for grid cells with household points but no building footprints.

Imputation of building footprint metrics: Building footprint metric values were imputed for grid cells with household points but no building footprints. For all grid cells with household points and building footprints, the count of household (HH) points per building footprint (BF) centroid (A = n_{HH} /n_{BF}) was calculated. Then a 3x3 focal window was applied to calculate the mean value of A, considering the 8 surrounding grid cells. For all grid cells with households but no building footprints, the value of A was taken from the nearest grid cell with building footprints that is within the same GHS-SMOD class (rural/urban classification). The count of building footprints was then imputed for these grid cells by dividing the count of household points by A (n_{BF} = n_{HH} / A). The same logic was applied to impute total area of building footprints and total perimeter of building footprints, by first calculating (i) n_{HH} / total area of building footprints and (ii) n_{HH} / total length of building footprints per grid cell. From the imputed values of counts, total area and total perimeter of building footprints, the mean building footprint area, mean building footprint perimeter and building footprint density (count / grid cell area) were then derived.

Modelling: The gridded estimates of total population (**ZMB_population_v2_0_gridded_total.tif**) were created using a variation of WorldPop's "top-down" modelling approach (Wardrop et al., 2018). Specifically, we utilised a Random Forest (RF)-based dasymetric mapping approach (Stevens et al., 2015) with the popRF R package (Bondarenko et al., 2021), based on the Breiman (2001) random forest algorithm. This was used to spatially disaggregate the ward total population, by modelling population density at the ward level with a combination of geospatial covariates and then using the model to estimate population density in each grid cell (0.0008333 decimal degrees, 3 arc seconds). The model could explain 95.3% of the input ward population variance. Outputs were constrained to grid cells that were considered to be populated, based on the presence of household points in the household location data collected by ZamStats as part of the 2022 census.

Gridded estimates for male (**ZMB_population_v2_0_gridded_male.tif**) and female population (**ZMB_population_v2_0_gridded_female.tif**) were calculated by multiplying the estimated total population per grid cell by the proportion male/female recorded in the census for each ward.

Validation: A final stage of this work compared the gridded population estimates (**ZMB_population_v2_0_gridded_total.tif**), with household-level population data in a sample of 10 districts. Within these districts, aggregated population totals were calculated from both the gridded population estimates and the household points, for grid cells at spatial resolutions ranging from 100m to 20km.

Data processing and analysis was carried out using ArcGIS Pro v3.0, R 4.3, and RStudio 3.0.

ASSUMPTIONS AND LIMITATIONS

The approach used to provide gridded population estimates in line with the 2022 National Census of Population and Housing, utilised ward-level population counts from the preliminary census results (Zambia Statistics Agency, 2022) and associated spatial boundaries. The "top-down" spatial disaggregation approach means that the grid cell values within each ward, in the resulting gridded population estimates, sum up to the total input ward population totals. The estimates are "constrained", such that population has only been allocated to grid cells with one or more households, as recorded during the 2022 census. In this work it is assumed that the household dataset is accurately georeferenced and represents all residential households at the time of the 2022 census. In locations which have experienced recent, 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.

To model population density, we considered a range of geospatial covariates, including building metrics (Dooley et al., 2021) derived from Ecopia building footprints (Ecopia.ai and Maxar Technologies, 2020). The Ecopia building footprint product for Zambia was extracted from satellite imagery predominantly dating from 2019 and thus there is a degree of temporal mismatch with the National Census in 2022. A degree of spatial mismatch was identified between the building footprint and household point datasets, with some grid cells having only building footprints but no household points (potentially due to buildings being non-residential in their use) and conversely, some grid cells having household points but no building footprints. To enable the building footprint metrics to be included as geospatial covariates for modelling population density, imputation of building footprint metrics was necessary for grid cells that had household points but no building footprints.

ACKNOWLEDGEMENTS

We thank the management and staff at the Zambia Statistics Agency who were involved in undertaking the 2022 Census of Population and Housing, in particular the Census Office team led by Mr Frank Kakungu and his deputies: Ms Chola Nakazwe Daka and Mr Palver Sikanyiti. Access to census data was provided to WorldPop at the University of Southampton in accordance with the data sharing agreement established as part of the GRID3 project. We acknowledge and thank the Zambia Statistics Agency staff who participated in the GRID3 gridded population data workshop in May 2023, which initiated this collaborative work to create a national gridded population dataset for Zambia, in line with the 2022 Census.

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, GitHub, url: https://github.com/wpgp/popRF.

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

Dooley, C.A., Leasure, D.R., Boo, G. and Tatem, A.J. (2021). Gridded maps of building patterns throughout sub-Saharan Africa, version 2.0. University of Southampton: Southampton, UK. Source of building footprints “Ecopia Vector Maps Powered by Maxar Satellite Imagery”© 2020/2021. https://dx.doi.org/10.5258/SOTON/WP00712

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

Stevens F.R., Gaughan A.E., Linard C., Tatem A.J. (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.

Wardrop, N.A., Jochem, W.C., Bird, T.J., Chamberlain, H.R., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V. and 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(14), pp.3529-3537, https://doi.org/10.1073/pnas.1715305115.

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).

Zambia Statistics Agency (2022). Preliminary Report - 2022 Census of Population and Housing, *Zambia Statistics Agency*, https://www.zamstats.gov.zm/wp-content/uploads/2023/05/2022-Census-of-Population-and-Housing-Preliminary.pdf.