Release Statement
Census disaggregated gridded population estimates for Sierra Leone (2015), version 2.0

02 June 2021

Notice: This document, data, and license are for review purposes only and may not be distributed further until this notice has been removed.

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the Geo-referenced Infrastructure and Demographic Data for Development project (GRID3) 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 (UNFPA), Center for International Earth Science Information Network (CIESIN) in the Earth Institute at Columbia University, and the Flowminder Foundation. We acknowledge the whole WorldPop Research Group and GRID3 partners for overall project support. Edith Darin (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 and Heather Chamberlain (WorldPop) supported the covariates processing work. Statistics Sierra Leone enumeration area-level population count were digitally mapped and shared by DSTI (the Directorate of Science, Technology & Innovation, Sierra Leone). Jolynn Schmidt (CIESIN) processed the enumeration area boundaries.

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. SLE_population_v2_0_gridded.tif
2. SLE_population_v2_0_agesex.zip
3. SLE_population_v2_0_mastergrid.tif
4. SLE_population_v2_0_tiles.zip
5. SLE_population_v2_0_sql.sql

SUGGESTED CITATION

WorldPop and Statistics Sierra Leone. 2021. Census disaggregated gridded population estimates for Sierra Leone (2015), version 2.0. University of Southampton. [doi: 10.5258/SOTON/WP00714]

RELEASE HISTORY

Version 2.0 (02 June 2021)

- The original input population data of version 1.0 contained errors in assigning population records to sections (Sierra Leone administrative division level 4). **The original data release should no longer be used**. In version 2.0, the enumeration area population as provided by DSTI was used to model the high-resolution population density and the DSTI chiefdom totals were used as a reference input for the disaggregation.
- The mastergrid was updated to reflect the exact boundaries as provided by DSTI. This changes the extent of the mastergrid and subsequent gridded populations from version 1.0.
- The boundaries used to link district-level age-sex proportions and gridded population were updated to reflect official district boundaries provided by DSTI.

Version 1.0 (16 November 2020)

- Added mastergrid, tiles, and SQL database to the release.

Version 1.0 (29 May 2020) [doi:10.5258/SOTON/WP00668]

- Original release of this data set.

FILE DESCRIPTIONS

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

SLE_population_v2_0_gridded.tif

This geotiff raster contains estimates of total population size for each approximately 100 m grid cell (0.0008333 decimal degrees grid or 3 arc seconds) across Sierra Leone. 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.

SLE_population_v2_0_agesex.zip

This zip file contains 40 geotiff rasters that contain counts of people for each age-sex group in each approximately 100 m grid cell across the study area.

We provide 36 rasters for the commonly reported age classes for males and females separately. These files are labelled with either an “m” (male) or an “f” (female) followed by 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 cohorts labelled with a “5”, “10”, etc. Eighty year olds and over are represented in the groups “f80” and “m80”. We provide four additional

rasters that represent demographic groups often needed 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).

In addition to the gridded population estimates for specific age-sex groups, we provide a spreadsheet of age-sex proportions (SLE_population_v2_0_agesex_table.csv). The regions are defined in the accompanying raster (SLE_population_v2_0_agesex_regions.tif).

SLE_population_v2_0_mastergrid.tif
This geotiff raster defines the study area and the resolution/alignment of the gridded outputs (~100 m grid cells; 0.0008333 decimal degrees). Grid cells outside the study area contain values of NA. Grid cells inside the study area with at least one building contain values of 1 (building counts from Dooley et al 2020). Grid cells inside the study area with no buildings contain values of 0.

SLE_population_v2_0_tiles.zip
This tiled web map allows for rapid display of the 100 m gridded population estimates across the study area. These can be used to develop web applications based on the model results. The tiles were created in XYZ format (compatible with Leaflet) with full coverage for zoom levels 1 to 14. These tiles are source data for the woprVision web application (https://apps.worldpop.org/woprVision/).

SLE_population_v2_0_sql.sql
This SQLite database contains estimates of population size in each grid cell. This database is source data for the woprVision web application (https://apps.worldpop.org/woprVision/) and it can be queried using the wopr R package.

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 the cell IDs (numbered left to right from top-left) of SLE_population_v2_0_mastergrid.tif.
- "x" and "y" columns contain WGS84 coordinates for the centroid of the grid cell.
- "Pop" column contains a population estimate for each grid cell.
- "agesexid" column contains the region ID for the age-sex proportions.
- "area" contains the total settled area in square meters (Dooley and Tatem, 2020).

SOURCE DATA

- Enumeration Area (EA) and chiefdom population totals based on the 2015 Population and Housing Census (Statistics Sierra Leone, 2015) were provided by DSTI in a shapefile format.
- Gridded building patterns (building count and density; mean, coefficient of variation and total building area per pixel; mean, coefficient of variation and total building length per pixel; satellite imagery year) were derived from building footprints by Dooley and Tatem (2020).
- WorldPop Global Gridded Age-Sex Proportions (WorldPop et al 2018).

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, distance to coastline and protected areas, night-time lights), were obtained from the "Global High Resolution Population Denominators Project" (OPP1134076).

METHODS OVERVIEW

Pre-processing: Overlapping EA units were merged together to ensure accurate linkage between population count and spatial unit.

Modelling: Following the Random Forest (RF)-based dasymetric mapping approach (Stevens et al., 2015; Bondarenko et al., 2020), the RF algorithm (Breiman, 2001) was used to model EA population density as a combination of the geospatial covariates and then to estimate the population density in each approximately 100 m grid cell (0.0008333 decimal degrees grid or 3 arc seconds). The model could explained 91.3% of the population input variance, compared with 83.37% when the building footprint-derived covariates were not included (and 51.9% in version 1.0).

The population density layer estimated from the RF modelling was subsequently used as a weight to disaggregate population counts from chiefdom into grid cells. Chiefdom totals rather than corrected EA totals were used for three reasons:

(1) The chiefdom boundaries were less prone to topological issues (overlapping and slivers) than the EA boundaries dataset.

(2) The merging of overlapping EAs led to some units crossing over administrative boundaries. Using these population totals would have resulted in mismatches between official census totals and aggregated gridded estimates.

(3) Some EAs do not contain any settled areas according to the building footprints layer despite containing population according to the census. Using EA level data would mean losing that EA population from the map whereas considering chiefdom level data reassigns those populations to other settled areas in the corresponding chiefdom.

The gridded population estimates were then combined with the WorldPop district age-sex proportions for Sierra Leone (WorldPop et al. 2018) using DSTI district boundaries 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).

ASSUMPTIONS AND LIMITATIONS

This dataset was produced based on the 2015 Population and Housing Census data provided by DSTI, Sierra Leone to the GRID3 programme. The observed population totals were not projected to 2020, and thus reflect the 2015 population distribution. The aggregated gridded estimates at chiefdom level might not exactly match the official chiefdom population totals due to small spatial overlap and slivers in the chiefdom boundaries data.

Population estimates are constrained within the settled area derived from building footprints (Dooley and Tatem, 2020). We assume that the building footprint data is accurate and that each building polygon corresponds to a building structure.

Some areas in the country have higher than average people per grid cells, despite being in rural areas. This may be due to some uncertainty around the input data, and more precisely the ratio between official chiefdom population totals and the amount of settled area from the building footprints (cf. Fig 1).

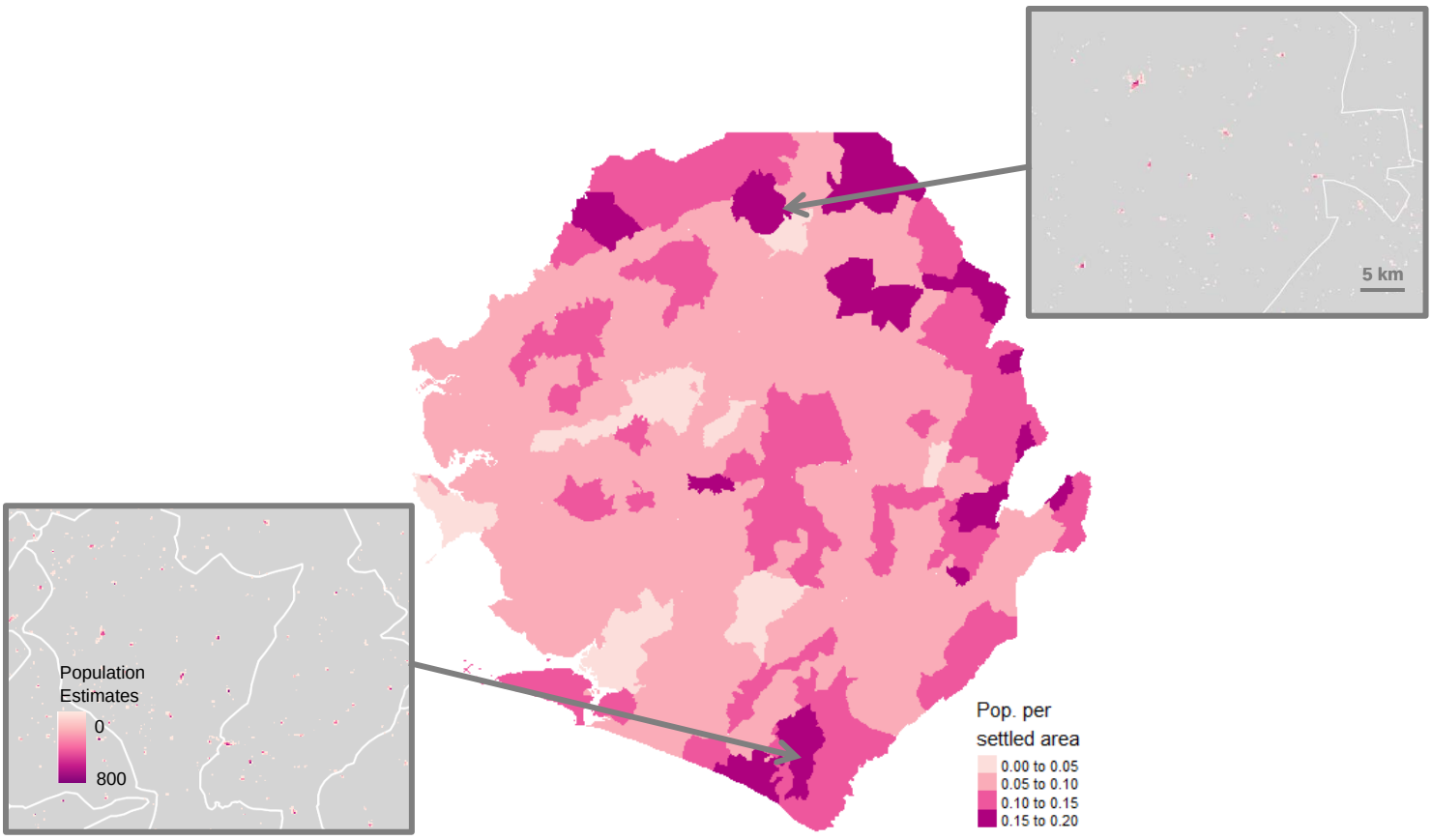

Figure 1: Map of the chiefdom population density per settled area (m^2) with two close-ups on the estimated 100 m x 100 m gridded population.

Furthermore, since the building footprint does not provide any information on the building use, we assume that every building structure is potentially residential, such that non-residential areas will nevertheless contain population estimates. Additionally, no adjustment was made to adapt the building footprints to reflect the settlement spatial distribution in 2015.

REFERENCES

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, Ves N, Wigley A, Koper P, 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

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

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

Statistics Sierra Leone (2015) Sierra Leone 2015 Population and Housing Census. https://www.statistics.sl/index.php/census/census-2015.html

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