Release Statement

Census disaggregated gridded population estimates for Senegal (2020), version 1.0

12 January 2022

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 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, Center for International Earth Science Information Network in the Earth Institute at Columbia University, and the Flowminder Foundation. Sarchil Qader (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 and Chris Jochem advised on the modelling. The National Agency for Statistics and Demography (ANSD) of Senegal released the updated yearly census-based total population projection and population projection by sex (female and male) using the results of the 2013 census of population (population projection). In addition, the digital Commune boundaries and infrastructure points were shared by ANSD.

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

SEN_population_v1_0_gridded.tif
SEN_population_v1_0_sex.zip
SEN_population_v1_0_mastergrid.tif

LICENSE

SUGGESTED CITATIONS

Qader S. H., Abbott T., Boytinck, E., Kuepie, M., Lazar A. N., Tatem A. J. 2022. Census disaggregated gridded population estimates for Senegal (2020), version 1.0. University of Southampton. doi:10.5258/SOTON/WP00730

FILE DESCRIPTIONS

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

SEN_population_v1_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 Senegal. 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.

SEN _population_v1_0_sex.zip

This zip file contains the following two raster files:

SEN_population_v1_0_gridded_female.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimates of the total female population size per grid cell across Senegal. 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.

SEN_population_v1_0_gridded_male.tif

This geotiff raster, at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimates of the total male population size per grid cell across Senegal. 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.

SEN_population_v1_0_mastergrid.tif
This geotiff raster contains 1s for each approximately 100m grid cell (0.0008333 decimal degrees) across the study area. Zero values indicate grid cells that did not contain buildings and were therefore assumed to be unpopulated. NAs show grid cells considered as outside the study area.

RELEASE HISTORY

Version 1.0 (12 January 2022) doi:10.5258/SOTON/WP00730

- Original release of this data set.

SOURCE DATA

- Digital Commune boundary and their projected population totals for 2020 based on the 2013 Population and Housing Census were provided by ANSD in a shapefile format.
- Gridded building patterns (building count, building area, building mean area, building density, building length, building and mean length) were derived from building footprints by Dooley and Tatem (2020).
- Infrastructure points (latitude and longitude position of administrative buildings, mosques, schools, commerce, etc.) were provided by ANSD in a shape file format.
- 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, schools, health facility, market place, place of worship, local roads, main roads, railway station, road intersection, and built settlement, and night-time lights), were obtained from the “Global High-Resolution Population Denominators Project” (OPP1134076).

METHODS OVERVIEW

Pre-processing: Overlapping Commune boundaries and existing gaps were fixed to ensure accurate linkage between population count and spatial unit. The building pattern and other covariates were adjusted so that extent precisely covers ANSD’s national boundary. To consider the non-residential buildings, new covariates were produced by combining the rasterised Ecopia building footprint layer with the rasterised ANSD infrastructure layers (i.e. the ratio of residential and non-residential buildings, the number of residential buildings).

Modelling: Following the Random Forest (RF)-based dasymetric mapping approach (Stevens et al., 2015; Bondarenko et al., 2021, the RF algorithm (Breiman, 2001) was used to model Commune 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 0.986% of the total population input variance.

The gridded population estimates were then combined with the ANSD Commune sex proportions for Senegal using ANSD Commune boundaries to produce gridded population estimates for each sex group (female and male).

ASSUMPTIONS AND LIMITATIONS

This dataset was produced based on the projected population totals for 2020 derived from the 2013 Population and Housing Census provided by ANSD. Although the observed population totals were projected to 2020, the total population in each Commune boundary is still an estimate and it might not 100% correspond to the true population on the ground. This might be due to the unquantified uncertainties of the subnational population projection assumptions.

Population estimates are constrained within the settled area derived from building footprints (Dooley and Tatem, 2020). We assumed that the building footprint data is accurate, and that each building polygon corresponds to a building structure. In addition, the distribution of the buildings might not represent the current building landscape in the country because satellite imageries from different years were used to extract building footprints (e.g. due to cloud coverage) (Dooley and Tatem, 2020). The building footprint does not provide any information on the building use, hence it was combined with the infrastructure point locations to adjust the estimated population totals. However, the infrastructure point location marks the location of businesses and not necessarily the non-residential buildings and also the points do not align perfectly with the building footprints. Despite these shortcomings, this makes the results more accurate.

WORKS CITED

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

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

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.r-project.org/package=popRF.

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