Release Statement

Gridded datasets of building count and area metrics for the UN Africa Region, version 1.0

15th March 2024

Original Release: 15th March 2024

There has been a rapid growth in the availability of multi-country building footprint datasets over recent years, due to developments in spatiotemporal resolution of satellite imagery, computing power and machine-learning algorithms. In many settings, these datasets enable new insights into the built environment, settlement patterns and associated populations, but can be difficult to work with at national or continental scale, due to the millions of individual polygon features involved. To support work on high-resolution gridded population estimates, and to facilitate comparisons between building footprint data products, we have developed a set of gridded (raster) outputs for building count and area metrics. These datasets cover the UN Africa region, and include metrics for Google, Microsoft and OpenStreetMap building footprints.

These datasets 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 (INV-045694). Project partners included the Center for International Earth Science Information Network in the Earth Institute at Columbia University, and the Flowminder Foundation. Geospatial data processing, and dataset production were led by Heather R. Chamberlain. Support in methods development was provided by Chris Jochem. Oversight of the work was provided by Attila N. Lazar and Andrew 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.

RELEASE CONTENT

1. * Africa_XXX_buildings_v1_0_centroid_count.tif
2. * Africa_XXX_buildings_v1_0_count.tif
3. * Africa_XXX_buildings_v1_0_total_area.tif
4. * Africa_XXX_buildings_v1_0_min_area.tif
5. * Africa_XXX_buildings_v1_0_max_area.tif
6. * Africa_XXX_buildings_v1_0_mean_area.tif

* Datasets are provided for building metrics from three sources of building footprint datasets that can be identified by the source of the data product in the filename (in place of XXX above): Googlev2, Microsoft or OSMJan23.

LICENSE

SUGGESTED CITATION

Chamberlain, H.R., Jochem, W.C., Lazar, A.N. and Tatem, A.J. 2024. Gridded datasets of building count and area metrics for the UN Africa Region, version 1.0. WorldPop, University of Southampton. DOI:10.5258/SOTON/WP00776

FILE DESCRIPTIONS

All spatial data files consist of rasters in geotiff format, in geographic coordinate system WGS84 (World Geodetic System 1984: EPSG 4326). The rasters have a spatial resolution of 3 arc seconds (0.00083333 decimal degrees).

Africa_XXX_buildings_v1_0_centroid_count.tif
Raster grid cell values reflect the count of building footprints in a grid cell, where the building footprint location is determined by its geometric centroid. NoData values represent grid cells without any building footprint centroids.

Africa_XXX_buildings_v1_0_count.tif
Raster grid cell values reflect the count of building footprints in a grid cell, where any part of a building footprint that intersects with any part of the grid cell is included in the count. NoData values represent grid cells without any building footprints.

Africa_XXX_buildings_v1_0_total_area.tif
Raster grid cell values reflect the summed area of all building footprint polygons within each grid cell, considering only the part of each polygon that is within a grid cell. The unit of total area is m^2. NoData values represent grid cells without any building footprints.

Africa_XXX_buildings_v1_0_min_area.tif
Raster grid cell values reflect the area of the smallest building footprint polygon within each grid cell, considering all building footprints polygon that are fully within or overlap with a particular grid cell. The unit of minimum area is m^2. NoData values represent grid cells without any building footprints.

Africa_XXX_buildings_v1_0_max_area.tif
Raster grid cell values reflect the area of the largest building footprint polygon within each grid cell, considering all building footprints polygon that are fully within or overlap with a grid cell. The unit of maximum area is m^2. NoData values represent grid cells without any building footprints.

Africa_XXX_buildings_v1_0_mean_area.tif
Raster grid cell values reflect the mean area of all building footprint polygons in a grid cell, considering all building footprints that intersect with the grid cell. The unit of mean area is m^2. NoData values represent grid cells without any building footprints.

RELEASE HISTORY

Version 1.0 (15 March 2024) [doi:10.5258/SOTON/WP00776]

- Original release of the data set.

SOURCE DATA

Gridded building metrics related to building footprint count and area were calculated for three building footprint data products:

- Google Open Buildings v2
- Microsoft building footprints (as available in January 2023)
- OpenStreetMap buildings (as available in January 2023)

Google Open Buildings v2 (Sirko et al., 2021) were downloaded from https://sites.research.google/open-buildings/#download on 10th January 2023. Data is stored in S2 tiles, with 127 tiles covering the UN Africa region. Data were downloaded in a zipped CSV format (extension .csv.gz), with geometry information included in WKT format in GCS EPSG:4326 (WGS84).

Microsoft building footprints were downloaded from the relevant GitHub repositories (https://github.com/microsoft/Uganda-Tanzania-Building-Footprints, https://github.com/microsoft/KenyaNigeriaBuildingFootprints, https://github.com/microsoft/GlobalMLBuildingFootprints) on 10th January 2023. The building footprints in the country-specific datasets for Tanzania, Uganda, Kenya and Nigeria were in zipped geoJSON files. The building footprints in the Microsoft “global” data product were further subset into multiple tiles per country, using a quadkey identifier. All tiles in the UN Africa region were downloaded and unzipped.

OpenStreetMap buildings (OpenStreetMap Contributors, 2023) were downloaded in country-specific files from the Geofabrik Downloads site (http://download.geofabrik.de/) in Shapefile format on 9th January 2023 for the majority of countries. Data for Mayotte, Madeira and Reunion were only available as raw data files in PBF format, as part of data extracts for Portugal and France. OpenStreetMap buildings in PBF format were converted to Shapefiles before further data processing.

METHODS OVERVIEW

All building footprint datasets were downloaded, either in country-specific zipped files (Microsoft and OpenStreetMap) or tiles (Google v2). As the input datasets did not utilise a common set of

national boundaries, initial data processing was done on a per-country or tile basis, based on the spatial extents of the source building footprint datasets.

The initial data processing consisted of calculating the building area and count metrics at a spatial resolution of 3 arc seconds. The spatial extents of the initial outputs was determined by the source building footprint files, with a common spatial resolution and grid cell alignment (based on the WorldPop Global 100m mastergrid (WorldPop et al., 2018)). The gridded building metrics were calculated using open-source Python (v3.11.0) code, adapted from Foks et al. (2020) and Heris et al. (2020), utilising the Rasterio v1.3.4 (Gillies et al., 2022a), Shapely v.2.0.0 (Gillies et al., 2022b), Geopandas v0.12.2 (Jordahl et al., 2022) and Fiona v1.8.22 (Gillies et al., 2022c) libraries.

Subsequent data processing then addressed the different spatial extents of the gridded building metric files, and mosaiced these to create a single Africa-wide raster for each source building footprint data product and metric. Similar gridded building metrics have previously been produced based on Ecopia building footprints (Dooley et al., 2020).

ASSUMPTIONS AND LIMITATIONS

- The spatial extent of the Africa-wide gridded building metric rasters are determined by the spatial extent of the source building footprint datasets. Therefore the spatial extent of rasters will vary for the same metric between different sources of building footprints.
- It is possible to have overlaps between building footprint polygons in each of the source datasets. We have not accounted for these in calculating gridded metrics since the overall impact is considered to be minimal.
- The Google v2 building footprints include a confidence score for each building footprint. We have not applied a threshold to this confidence score, and have therefore included all Google footprints in the calculation of building metrics.
- Google has not released v2 building footprints for some countries (and some sub-national administrative units), primarily those affected by conflict. Countries known to have no Google v2 building footprints include: Libya, Morocco, Mali, Chad and South Sudan. Sub-national areas without Google v2 building footprints include: Cabo Delgado (Mozambique), North Kivu (Democratic Republic of the Congo), Central Darfur (Sudan) and Borno state (Nigeria).
- Microsoft building footprints do not include a clear versioning system. The Microsoft data used for the calculation of building metrics are those available in January 2023. This consisted of country-specific datasets for Nigeria, Kenya, Tanzania and Uganda, and data for all other countries from the “global” dataset. There is some overlap in coverage between country-specific datasets and the “global” dataset in some locations. No building footprints were available for Cabo Verde, and there are considerable gaps in coverage in many countries, including no building footprints for the entirety of Kinshasa (DR Congo).
- Buildings from OpenStreetMap consist of features that have been tagged in OpenStreetMap as ‘buildings = *’. Given that OpenStreetMap features are created by a community of thousands of mappers, there is likely to be some variation in how buildings

are represented as features in this data. We have observed that OpenStreetMap buildings in Western Sahara, consistently have multiple buildings represented as a single polygon, often covering the full extent of a city block in urban areas. Homogeneity in spatial coverage of OSM buildings is expected given its nature as a volunteered geographic information (VGI) data product.

- The date of satellite imagery used for feature extraction or digitisation of building footprints is not known at a granular level, with the only information available from data producers being an indication of the range of years from which satellite imagery was sourced.

WORKS CITED

Dooley, C. A., Boo, G., Leasure, D.R. and Tatem, A.J. (2020). Gridded maps of building patterns throughout sub-Saharan Africa, version 1.1. 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/WP00677

Foks, N., Heris, M. P., Bagstad, K. J. & Troy, A. A. (2020). A Code for Rasterizing Microsoft's Building Footprint Dataset. U.S. Geological Survey https://doi.org/10.5066/P9XZCPMT.

Gillies, S. & others (2022a). Rasterio: geospatial raster I/O for Python programmers (v1.3.4). Mapbox. https://github.com/rasterio/rasterio

Gillies, S., van der Wel, C., Van den Bossche, J., Taves, M. W., Arnott, J., Ward, B. C., & others. (2022b). Shapely (v2.0.0). Zenodo. https://doi.org/10.5281/zenodo.7428463

Gillies, S., Buffat, R., Arnott, J., Taves, M.W., Wurster, K., Snow, A.D., Cochran, M., Sales de Andrade, E. and Perry, M. (2022c). Fiona (v1.8.22). https://github.com/Toblerity/Fiona

Heris, M. P., Foks, N. L., Bagstad, K. J., Troy, A., & Ancona, Z. H. (2020). A rasterized building footprint dataset for the United States. Scientific data, 7(1), 207, https://doi.org/10.1038/s41597-020-0542-3.

Jordahl, K., Van Den Bossche, J., Fleischmann, M., McBride, J., Wasserman, J., Richards, M., & Ren, C. (2022). geopandas/geopandas: v0.12.2 (v0.12.2). Zenodo. https://doi.org/10.5281/zenodo.7422493

OpenStreetMap contributors (2023). Planet dump [Data file from 09 JAN 2023]. Retrieved from https://planet.openstreetmap.org, accessed from https://download.geofabrik.de/africa

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/WP00647