Release Statement

Gridded maps of building patterns throughout sub-Saharan Africa, version 1.1

6 July 2020

Original Release: 6 May 2020

This work was funded by the Bill and Melinda Gates Foundation (BMGF) (INV-002697). Method, coding and data production was led by Claire A. Dooley with additional code provided by Gianluca Boo and Douglas R. Leasure. Oversight was provided by Andrew J. Tatem. Contact release@worldpop.org for more information.

The whole WorldPop group are acknowledged for overall project support. Special thanks go to Rhiannan Price (Maxar) for providing source data details and coordination for this publication, and Maksym Bondarenko (WorldPop) for computer infrastructure and data access support. We acknowledge the efforts of WorldPop's partners (United Nations Population Fund (UNFPA), Center for International Earth Science Information Network (CIESIN) in the Earth Institute at Columbia University, and the Flowminder Foundation) in supporting access to the building footprints and, in particular, we thank Io Blair-Freese (BMGF) and Heather Chamberlain (WorldPop) for providing coordination between WorldPop and BMGF.

CITATION

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. doi:10.5258/SOTON/WP00677

*note this citation will be updated once layers for all 51 countries within sub-Saharan Africa are complete

RELEASE CONTENT

1. * XXX_buildings_v1_1_count.tif
2. * XXX_buildings_v1_1_density.tif
3. * XXX_buildings_v1_1_urban.tif
4. * XXX_buildings_v1_1_total_area.tif
5. * XXX_buildings_v1_1_mean_area.tif
6. * XXX_buildings_v1_1_cv_area.tif
7. * XXX_buildings_v1_1_total_length.tif
8. * XXX_buildings_v1_1_mean_length.tif
9. * XXX_buildings_v1_1_cv_length.tif
10. * XXX_buildings_v1_1_imagery_year.tif
11. bf_metrics_generation_code.R

* Rasters are provided for 51 countries that can be identified using the ISO3 country codes in the filenames in place of XXX above. For example, SOM refers to Somalia. See appendix for full list of countries in this dataset.

FILE DESCRIPTIONS

The geotiff rasters have a spatial resolution of approximately 100m (3 arc seconds). Their coordinate reference system is WGS84. NAs represent grid cells without any building-footprint centroid.

XXX_buildings_v1_1_count.tif

This raster contains counts of buildings that fall within a grid cell. Each buildings was counted in the grid cell that contained the centroid of its building footprint.

XXX_buildings_v1_1_density.tif

This raster contains a measure of the number of buildings per grid cell area in m^2, i.e. building count divided by grid cell area. If needed, the grid cell area can be retrieved by dividing the count raster by the density raster.

XXX_buildings_v1_1_urban.tif

This raster contains an urban / rural classification for each grid cell based on building patterns in their surroundings. A value of 1 indicates an urban grid cell and a value of zero indicates a rural grid cell. NAs indicate grid cells without any building-footprint centroid. Note: Details on the settlement classification are provided in the methods section.

XXX_buildings_v1_1_total_area.tif

This raster contains a grid cell level sum of the building areas for all buildings whose centroid falls inside a grid cell. Total area is given in m^2. Note: Total building area could exceed the area of a grid cell if the centroid of a large building falls within the grid cell.

XXX_buildings_v1_1_mean_area.tif

This raster contains a grid cell level mean of the building areas for all buildings whose centroid falls inside a grid cell. Mean area is given in m^2.

XXX_buildings_v1_1_cv_area.tif

This raster contains a grid cell level coefficient of variation of building areas for all buildings whose centroid falls inside a grid cell. Coefficient of variation is the standard deviation divided by the mean. Note: This measure informs about the heterogeneity in building areas within a grid cell.

XXX_buildings_v1_1_total_length.tif

This raster contains a grid cell level sum of the building lengths for all buildings whose centroid falls inside a grid cell. Here, length refers to a building’s perimeter. Total length per grid cell is therefore a sum of the perimeters of all the buildings whose centroid falls inside a grid cell. Total length is given in metres.

XXX_buildings_v1_1_mean_length.tif

This raster contains a grid cell level mean of the building lengths for all buildings whose centroid falls inside a grid cell. Here, length refers to a building’s perimeter. Mean length per grid cell is therefore the mean perimeter length of all the buildings whose centroid falls inside a grid cell. Mean length is given in metres.

XXX_buildings_v1_1_cv_length.tif

This raster contains a grid cell level coefficient of variation of building lengths for all buildings whose centroid falls inside a grid cell. Coefficient of variation is the standard deviation divided by the mean. Here, length refers to a building’s perimeter. Note: This measure informs about the heterogeneity in building lengths within a grid cell.

XXX_buildings_v1_1_imagery_year.tif
This raster contains the year of the imagery used to extract the building footprints for each grid cell. Grid cells containing building footprints extracted using imagery from different years were allocated the imagery year used for the majority of buildings in the grid cell. When different years were equally represented, the most recent year was selected. Note: The year of the imagery enables identification of areas where the building footprint metric layers may not provide an up-to-date representation of the current distribution of buildings.

bf_metrics_generation_code_v1.1.R
This R script provides the code used to generate the rasters. The code is intended to be used as a guide to accompany the methods overview below. If users would like to run the code, it will need to be edited to suit the file structure of the building footprints.

RELEASE HISTORY

Version 1.1 (6 July 2020)

- The rasters in the original release (v1.0) except for those for COD remain unchanged but are now labelled v1.1 to reflect the version of the release. In the original release, COD (v1.0) only covered part of the country, in v1.1 COD rasters have national coverage. The rasters for Central African Republic were previously labelled as ‘CAR’ in the file names. These have been edited to ‘CAF’, in line with the official ISO3 country codes. Rasters for an additional 14 countries are included in v1.1. Also, v1.1 includes imagery year rasters for all 51 countries (imagery rasters were not included in the original release).

Version 1.0 (6 May 2020)

- This is the original release of the data

See appendix for country-specific version history.

ASSUMPTIONS AND LIMITATIONS

- The ‘settled’ grid cells included in all rasters is defined by where the building footprints datasets contain polygons. We have not masked out any areas, nor do we perform any quality checks of the building footprints data prior to processing.
- Caution should be taken when using the rasters in combination with national boundaries as the building footprints data may not cover some areas within user boundaries at edge locations.
- Our classification of ‘urban’ is very simplistic and should be used only if the rules applied for classification (see methods section below) meet the user’s needs.
- Our classification of ‘urban’ is partially based on number of contiguous cells in clusters of settled cells. Cell area differs across the globe in the WGS84 projection and so our rule of number of cells will result in a different threshold area for different locations.

SOURCE DATA

These data were created using building footprints provided by the Digitize Africa project of Ecopia.AI and Maxar Technologies (2020). Digitize Africa is a two-year project funded by the Bill and Melinda Gates Foundation to map buildings and roads in 51 countries across sub-Saharan Africa using satellite imagery and artificial intelligence (AI) to support humanitarian assistance and sustainable development. Maxar provided their Vivid satellite imagery mosaics (50 cm resolution) and Ecopia.AI generated the building footprints using their artificial intelligence-assisted feature extraction techniques.

METHODS OVERVIEW

The same method was applied to each country separately. All data processing was carried out in R (R Core Team, 2013). For each country building footprint polygons were converted into centroid points in UTM projection using the *st_centroid* function in the *sf* R package (Pebesma, 2018). The centroid points were then re-projected to WGS84 using *sf's st_transform* function so that their corresponding cell IDs in the WGS84 projected raster master grid could be identified. We used the WorldPop country master grids (WorldPop et al. 2018) which defined the extent and resolution of the output building footprint metric layers, in addition to the coordinate reference system. Note that the master grid for South Africa (ZAF) was cropped to exclude Prince Edward Islands as the building footprints did not include these islands. Building centroid cell IDs were obtained using the *cellFromXY* function in the *raster* R package (Hijmans & van Etten, 2012). Building counts per grid cell were calculated by simply summing the number of centroids for each cell ID. Building density per grid cell was calculated by dividing the building count in a cell by the area of the cell. Grid cell area was calculated using the *area* function in the *raster* R package. The three building area related rasters and the three building length related rasters were all derived using the attributes data ("Shape_Area" and "Shape_Length") contained in the .gdb building footprints dataset, by applying sum (total area or length), mean (mean area or length) and standard deviation divided by mean (cv area or length). For the urban rasters we classified grid cells as urban (as opposed to rural) by applying two simple rules that were consistent across countries. If cells were part of a grouping of contiguous cells with: 1) 1,500 or more cells, and; 2) 5,000 or more buildings. The *clump* function in the *raster* R package was used to identify clusters of contiguous cells, and cells were classified as being in the same cluster if they were directly adjacent horizontally, vertically or diagonally. The year of the imagery used to extract the building footprints was obtained from original data for each building and then the most common year was recorded for each grid cell. In the few cases where there were the same number of buildings with different imagery years, the most recent is given.

REFERENCES

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

Hijmans, R.J. & van Etten, J. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. http://CRAN.R-project.org/package=raster

Pebesma, E. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1), 439–446. doi: 10.32614/RJ-2018-009, https://doi.org/10.32614/RJ-2018-009.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

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/WP00651

Appendix

ISO3 country code	Country name	Version 1.0 release (all metric layers except imagery year)	Version 1.1 release (all metric layers)
AGO	Angola	1.0	1.1
BDI	Burundi	1.0	1.1
BEN	Benin	1.0	1.1
BFA	Burkina Faso	1.0	1.1
BWA	Botswana	1.0	1.1
CAF	Central African Republic	1.0 (labelled CAR)	1.1
CIV	Côte d'Ivoire	1.0	1.1
CMR	Cameroon	1.0	1.1
COD	Democratic Republic of the Congo	1.0 (partial spatial coverage)	1.1
COG	Congo	1.0	1.1
COM	Comoros	1.0	1.1
CPV	Cabo Verde	1.0	1.1
DJI	Djibouti	1.0	1.1
ERI	Eritrea	1.0	1.1
ESH	Western Sahara	1.0	1.1
ETH	Ethiopia		1.1
GAB	Gabon		1.1
GHA	Ghana	1.0	1.1
GIN	Guinea	1.0	1.1
GMB	Gambia	1.0	1.1
GNB	Guinea-Bissau	1.0	1.1
GNQ	Equatorial Guinea	1.0	1.1
KEN	Kenya	1.0	1.1
LBR	Liberia	1.0	1.1
LSO	Lesotho	1.0	1.1
MDG	Madagascar	1.0	1.1
MLI	Mali	1.0	1.1
MOZ	Mozambique	1.0	1.1
MRT	Mauritania		1.1
MUS	Mauritius		1.1
MWI	Malawi	1.0	1.1
NAM	Namibia	1.0	1.1
NER	Niger	1.0	1.1
NGA	Nigeria		1.1
REU	Réunion		1.1
RWA	Rwanda	1.0	1.1
SDN	Sudan		1.1
SEN	Senegal		1.1
SLE	Sierra Leone	1.0	1.1

SOM	Somalia	1.0	1.1
SSD	South Sudan	1.0	1.1
STP	Sao Tome and Principe		1.1
SWZ	Eswatini	1.0	1.1
SYC	Seychelles		1.1
TCD	Chad	1.0	1.1
TGO	Togo		1.1
TZA	Tanzania		1.1
UGA	Uganda	1.0	1.1
ZAF	South Africa		1.1
ZMB	Zambia	1.0	1.1
ZWE	Zimbabwe		1.1