Release Statement

Gridded maps of building patterns throughout sub-Saharan Africa, version 1.0

6 May 2020

Original Release: 6 May 2020

This work was funded by the Bill and Melinda Gates Foundation (BMGF) (INV-002697). Method and data production was led by Claire A. Dooley with oversight provided by Andrew J. Tatem. These data may be distributed using a Creative Commons Attribution Share-Alike 4.0 License. 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, Doug Leasure (WorldPop) for providing feedback on the data, coordinating the data review process and coordinating data access, 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), Attila Lazar (WorldPop) and Heather Chamberlain (WorldPop) for providing coordination between WorldPop and BMGF.

CITATION

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. doi:10.5258/SOTON/WP00666

RELEASE CONTENT

1. * XXX_buildings_v1_0_count.tif
2. * XXX_buildings_v1_0_density.tif
3. * XXX_buildings_v1_0_urban.tif
4. * XXX_buildings_v1_0_total_area.tif
5. * XXX_buildings_v1_0_mean_area.tif
6. * XXX_buildings_v1_0_cv_area.tif
7. * XXX_buildings_v1_0_total_length.tif
8. * XXX_buildings_v1_0_mean_length.tif
9. * XXX_buildings_v1_0_cv_length.tif
10. bf_metrics_generation_code.R

* Rasters are provided for 37 countries that can be identified using the ISO3 country codes in the filenames in place of XXX above. For example, SOM refers to Somalia. Countries included in this dataset are shown in the apprendix.

FILE DESCRIPTIONS

All geotiff rasters have a spatial resolution of approximately 100 m grid cell (3 arc seconds). WGS84 are their reference coordinate system. All rasters contain NAs in grid cells without any building centroids.

XXX_buildings_v1_0_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_0_density.tif
This raster contains a measure of the number of buildings per grid cell area, i.e. building count/pixel area. If needed, the pixel area can be calculated by dividing the count raster by the density raster.

XXX_buildings_v1_0_urban.tif
This raster contains an urban / rural classification for each grid cell based on building patterns in that area. A value of 1 indicates an urban grid cell and a value of zero indicates a rural grid cell. NA values indicate grid cells with zero buildings. See methods section below for details on settlement type classification.

XXX_buildings_v1_0_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. 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_0_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. Given in m^2.

XXX_buildings_v1_0_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. This raster provides a measure in the variation in building area within a grid cell.

XXX_buildings_v1_0_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. Given in metres.

XXX_buildings_v1_0_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. Given in metres.

XXX_buildings_v1_0_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. This raster provides a measure of the variation in building perimeters within a grid cell.

bf_metrics_generation_code.R

This R script provides the code used to generate the all rasters. This should be treated as a guide to accompany the methods overview below. If users would like to run the code, please ensure that the building footprint files are contained within a folder structure that is appropriate for the code.

RELEASE HISTORY

Version 1.0 (6 May 2020)

- This is the original release of the data

ASSUMPTIONS AND LIMITATIONS

- The 'settled' pixels 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. 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. Pixel 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 pixels 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.

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	Version of data release
AGO	1.0
BDI	1.0
BEN	1.0
BFA	1.0
BWA	1.0
CAR	1.0
CIV	1.0
CMR	1.0
COD	1.0
COG	1.0
COM	1.0
CPV	1.0
DJI	1.0
ERI	1.0
ESH	1.0
GHA	1.0
GIN	1.0
GMB	1.0
GNB	1.0
GNQ	1.0
KEN	1.0
LBR	1.0
LSO	1.0
MDG	1.0
MLI	1.0
MOZ	1.0
MWI	1.0

NAM	1.0
NER	1.0
RWA	1.0
SLE	1.0
SOM	1.0
SSD	1.0
SWZ	1.0
TCD	1.0
UGA	1.0
ZMB	1.0