

**Data Release Statement**  
**GRID3 Democratic Republic of Sao Tome and Principe**  
**Settlement Extents,**  
**Version 01.01**

**November 2021**

**Abstract**

This document outlines the methodological approach and data sources used to construct the “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01 dataset.” The dataset consists of settlement extents across Sao Tome and Principe, as well as accompanying population estimates for each settlement extent. Terms of use for these data are provided.

**Dataset citation**

Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2021. GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). 10.7916/d8-bmr9-0s52 .Accessed DAY MONTH YEAR.

**Terms of use**

Users are free to download, store, access, use, copy, adapt, transform, alter, arrange, build upon, distribute and transmit this work and any derivative works based upon the work for humanitarian purposes only. Attribution of the source must be provided, and further distribution of this work or derived work must maintain the same terms of data use and license as set forth in this Terms of Use.

“Humanitarian Purpose” means any non-commercial purpose of supporting operations undertaken to alleviate human suffering caused by events (or series of events) that threaten the health, safety or wellbeing of a large group of people.

Both the data and data release statement are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>) and specified in legal code (<https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode>). These Terms of Use, including the aforementioned Creative Commons license terms, take effect and are legally binding on you, and any on any corporation, government agency or other legal entity for which you are acting as an employee or agent, when you download, access or use the work

and these license terms apply to any further distribution of the work or any derivatives of the work you may make.

The “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01” is a derivative work from Digitize Africa, Ecopia Landbase Africa powered by Maxar Ecopia AI.

Copyright 2021. The Trustees of Columbia University in the City of New York.

### **Contacts and data queries**

GRID3 appreciates feedback regarding this dataset, such as suggestions, discovery of errors, difficulties in using the data, and format preferences.

For dataset-related questions, please send an email to any of the following organisations:

GRID3: [data.queries@grid3.org](mailto:data.queries@grid3.org)



# GRID<sup>3</sup>

GEO-REFERENCED INFRASTRUCTURE AND  
DEMOGRAPHIC DATA FOR DEVELOPMENT

## Table of Contents

Introduction.....	4
Methodological approach.....	4
<i>Input data</i> .....	<b>4</b>
<i>Methods</i> .....	<b>5</b>
<i>Dataset description</i> .....	<b>8</b>
<i>Dataset codebook</i> .....	<b>9</b>
<i>Version History</i> .....	<b>10</b>
Known data limitations and disclaimer.....	10
Acknowledgments.....	11
References.....	11

## Introduction

Settlement extents are polygons representing areas where there is likely a human settlement based on the presence of buildings detected in satellite imagery. Settlement extents are not meant to represent the boundaries of an administrative unit or locality. A single settlement extent may be made up of multiple localities, especially in urban areas. Each settlement extent has an associated population estimate. Provided is information on the common operational boundary that the extent fully resides within along with their associated place codes (PCodes). This document details the methodology applied to produce the dataset “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01.”

This work has been undertaken as part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme. The programme is funded by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office. It is implemented by the Flowminder Foundation, WorldPop at the University of Southampton, the United Nations Population Fund, and the Center for International Earth Science Information Network (CIESIN) at Columbia University.

## Methodological approach

### Input data

Settlement extents and associated population estimates are derived from two primary input datasets: building footprints and population estimates (**Table 1**).

**Table 1.** Input datasets

<i>Data</i>	<i>Description</i>	<i>Source</i>
Building footprints	Building footprints are derived by the Ecopia Tech Corporation using a feature extraction algorithm. The dataset consists of polygon data delineating rooftops present in satellite imagery captured between 2009 and 2019.	Maxar Technologies, Inc. and Ecopia Tech Corporation. 2020. Ecopia Landbase Africa powered by Maxar. Available at <a href="https://digitizeafrica.ai">DigitizeAfrica.ai</a>
WorldPop Population estimates	<u>Population estimates (2020):</u> Top-down modeled population estimates (100-meter resolution) produced by WorldPop at the University of Southampton. Units: Total population per pixel. Extent: Sao Tome and Principe <a href="https://www.worldpop.org/methodology">https://www.worldpop.org/methodology</a>	Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan



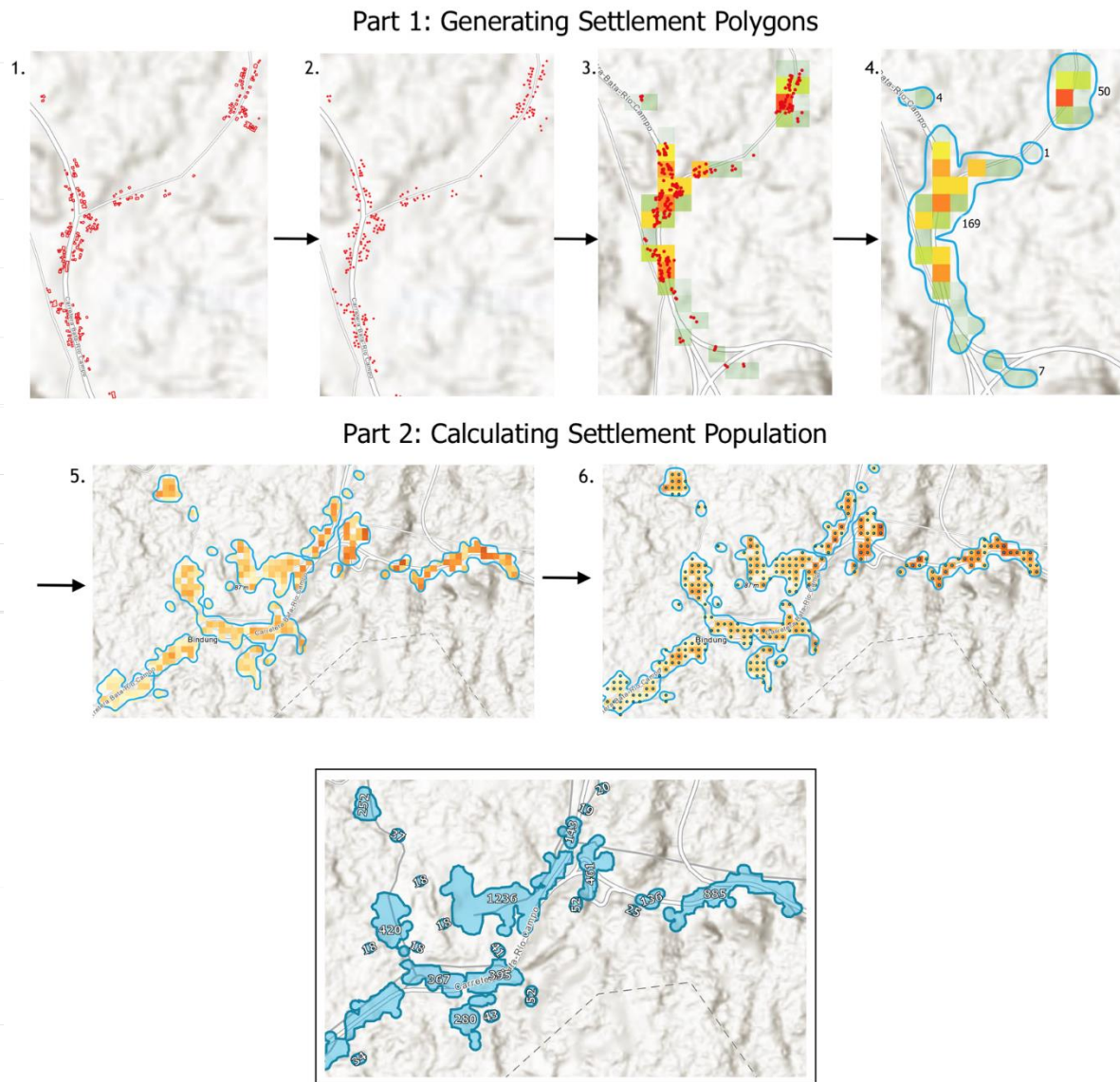
	<p><a href="#">ds/populations</a></p> <p><u>UN-adjusted Population estimates (2020):</u>          Identical to above, but country population totals for Sao Tome and Principe are adjusted to match estimates prepared by: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1.</p>	<p>Africa in 2020 using building footprints. WorldPop, University of Southampton, UK.          doi:10.5258/SOTON/WP00682</p> <p>doi:10.5258/SOTON/WP00683 (UN-adjusted)</p>
Administrative boundaries and Place Codes	<p>The United Nations Office for the Coordination of Humanitarian Affairs Common Operational Datasets for administrative boundaries down to the second subnational level.</p> <p>Common Operational Datasets (CODs) are authoritative reference datasets needed to support operations and decision making for all actors in a humanitarian response. CODs are 'best available' datasets that ensure consistency and simplify the discovery and exchange of key data.</p>	<p>United Nations Office for the Coordination of Humanitarian Affairs (OCHA) Regional Office for West and Central Africa (ROWCA). (Updated 17 August 2020). São Tomé and Príncipe - Subnational Administrative Boundaries. Global Administrative Boundaries Database (GADM - <a href="http://www.gadm.org">www.gadm.org</a>). Accessed 10/30/2021. Available from the Humanitarian Data Exchange:  <a href="https://data.humdata.org/dataset/sao-tome-and-principe-administrative-boundaries">https://data.humdata.org/dataset/sao-tome-and-principe-administrative-boundaries</a></p>

**Methods**

Both the settlements’ extents and classification are derived solely from Ecopia’s building footprints. Data analysis and processing were achieved entirely using Esri’s ArcGIS software (Pro version 2.7.3), its native module arcpy, and open source python library pandas 1.3.0.

The center points of building footprint features are converted to a raster grid (3 arc-second resolution) of building densities (i.e. total number of building points within each cell), referred to throughout this document as “building densities.” The ArcGIS Pro “Contour” tool was used to generate shell-up contours from the building density grid to delineate settled versus non-settled areas. The shell-up method includes contours that start at the lower

bounds, but includes all grid cells with building densities to the upper bounds of the grid. For example, a shell-up contour of 10 would include all grid cells with a building density of 10 or more. Contours with a building density of one or more are used to create the settlement extent polygons (see **Figure 1**).



**Figure 1:**

**Part 1** - Illustration of methods used to generate settlement extents: 1) Building footprints (Ecopia/Maxar) are projected to map and 2) Building footprints are converted to points. 3) Building points are converted to a building density raster layer at 100m resolution. Values represent the total number of building points within each cell. 4) Shell-up contours are generated around building density cells, and building counts within each settlement extent are summed. Contour geometry is smoothed to produce final settlement extent polygons.

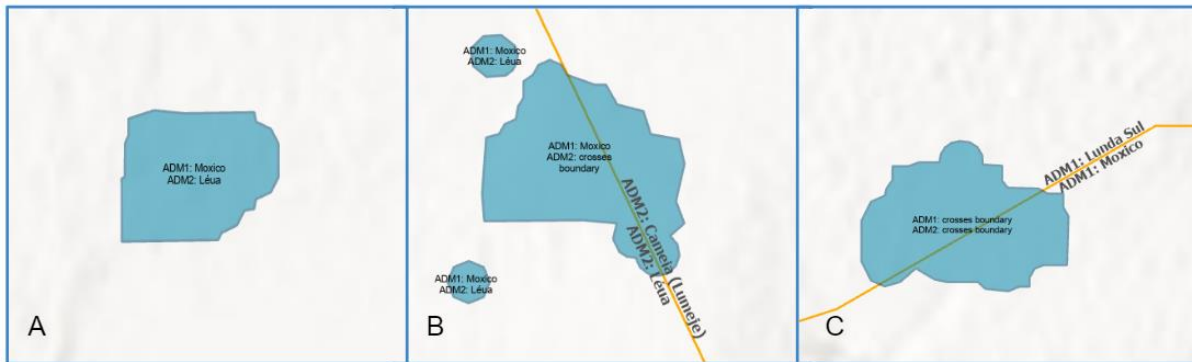
**Part 2** - Illustration of methods to generate population estimates for each settlement extent: 5) A population raster (100m, WorldPop) is added to the map and 6) cells are converted to points. Each population point is assigned to the nearest settlement extent and population values are summed to produce population estimates. Steps 5 and 6 are performed once for a WorldPop raster and once for a UN-adjusted WorldPop raster. **Bottom panel** represents final settlement extents with associated population count (rounded to nearest integer for clarity).

The GRID3 Democratic Republic of Sao Tome and Principe settlement extents are then classified (based on building density) into three classes: built-up areas (BUA), small settlement areas (SSA), and Hamlets (Barau et al., 2014).

**Table 2.** Settlement extent classification

Settlement class	Description
Built-up areas (BUAs)	A built-up area (BUA) is generally an area of urbanisation with moderately-to-densely-spaced buildings and a visible grid of streets and blocks. BUAs are characterised as polygons that maintain a 100 m <sup>2</sup> building density of 13 or more across an area greater than or equal to 0.4 km <sup>2</sup> .
Small Settlements (SSAs)	A small settlement (SSA) is a settled area of permanently inhabited structures and compounds of roughly a few hundred to a few thousand inhabitants. The housing pattern in SSAs is an assemblage of family compounds adjoining other similar habitations. Small settlement areas are characterised as polygons containing 50 or more buildings and not a BUA.
Hamlet	A hamlet is a collection of several compounds or sleeping houses in isolation from small settlements or urban areas. Hamlets are characterised as polygons containing between 1 and 49 buildings.

The population estimates for each settlement extent were calculated using population estimates from WorldPop. Individual, country-constrained population raster cells and their associated population estimate (100m resolution) were converted to individual points and joined to the nearest settlement extent. Finally, the values within a settlement extent were summed to obtain the total population for each settlement extent. Settlement extents from population estimates can be found in the “Population” and “Pop\_UN\_adj” fields.



**Figure 2:** Panel A illustrates the names for both the ADM1 and ADM2 boundaries are provided when a settlement extent resides within a single boundary. Panel B illustrates that only the ADM1 boundary name is provided when a settlement crosses over an ADM2 boundary. Panel C illustrates that neither the ADM1 or ADM2 boundary name is provided when a settlement crosses over an ADM1 boundary.

The administrative boundary names represent the administrative level that the settlement extent resides fully within, down to the second subnational level. If the settlement extent crosses over an administrative boundary then the extent is assigned to the level above and the value “crosses boundary” is used for the administrative boundary name. Settlement PCodes are generated based on the PCode for the administrative level that the boundary resides fully within and a six digit number unique to that administrative level. If an extent falls completely outside of the farthest administrative boundary then, the extent admin level data is set to NA.

### *Dataset description*

The data are in geodatabase format and consist of a single-feature class. An example of data is shown on Figure 3.

Extent: Sao Tome and Principe: Admin Level 0 Boundaries. The overall extent of the layer is limited to the overall extent of the building footprint dataset and may not reflect the extent of official administrative boundaries.

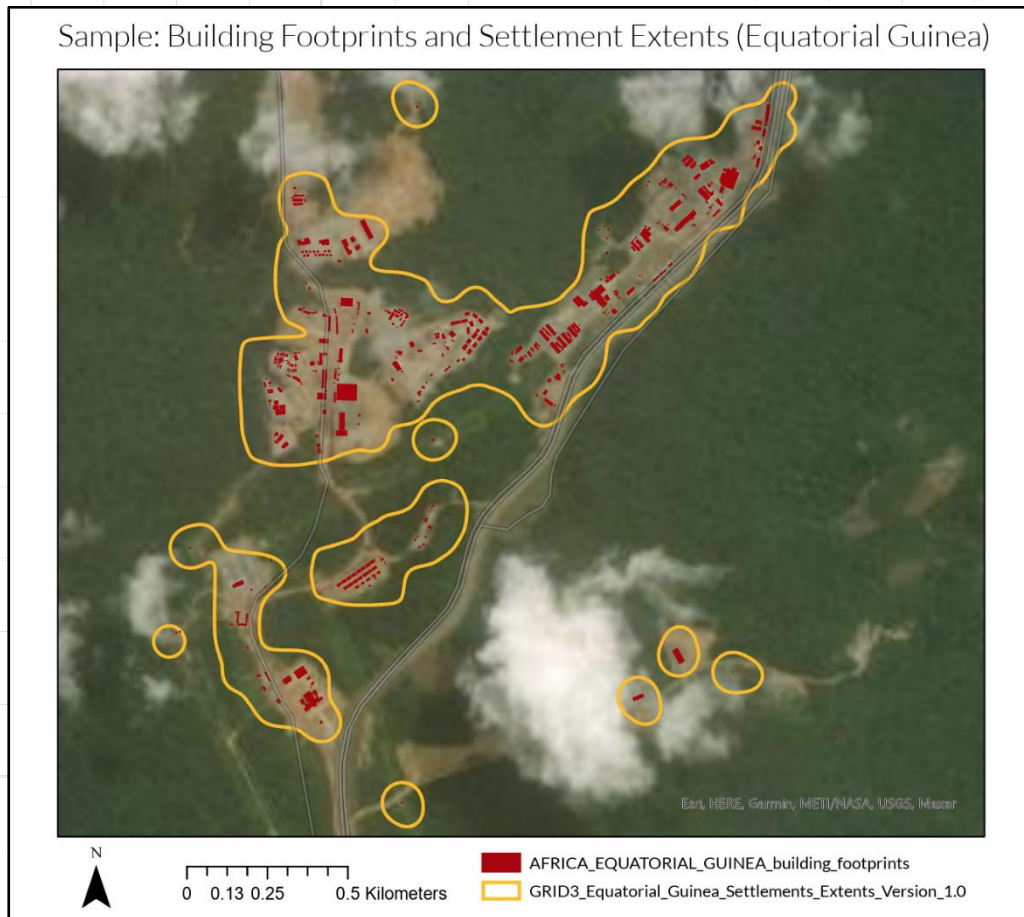
Coordinate system: GCS WGS 1984

The WorldPop top-down constrained population estimates 2020 (Population) uses, for each country, the highest admin level official population totals of the 2000 and 2010 census rounds. These are publicly available and can be mapped to associated boundaries, and project them to 2020. These projected values then are disaggregated statistically to 100x100m resolution using a set of detailed geospatial datasets to disaggregate them to grid cell-based counts. The estimates are constrained to settlements based on the satellite-derived building footprint data from Maxar/Esocopia for the 51 African countries, and based on a built settlement growth model of WorldPop for the remaining countries.

The Population Counts / Constrained Individual countries 2020 UN adjusted (100m resolution) population estimates (Pop\_UN\_adj) recognises that the United Nations produce their own estimates of national population totals. WorldPop, in order to provide flexibility to users, adjusted the number of people per pixel of its top-down constrained population



estimates nationally to match the corresponding official United Nations population estimates (i.e. 2019 Revision of World Population Prospects).



**Figure 3:** Sample map depicting a settlement extent data layer, with building footprint layer for reference. Note: Building footprint layers are not included in this data product.

### *Dataset codebook*

Field	Description
OBJECTID	Default ESRI field: Unique sequential numeric identifier maintained by the database.
Shape	Default ESRI field: Geometry type
MGRS_Code	Unique name generated using the Military Grid Reference System
Country	Country name
ISO	Three-letter country ISO code
Type	The settlement type as defined by a built-up area, small-settlement area, or hamlet

Population	Total population estimate for settlement extent derived from constrained WorldPop raster
Pop_UN_adj	Total population estimate for settlement extent derived from constrained and UN-adjusted WorldPop raster
adm0_pcode	The place code for administrative level 0
adm1_name	The name of administrative level 1
adm1_pcode	The place code for administrative level 1
adm2_name	The name of administrative level 2
adm2_pcode	The place code for administrative level 2
settl_pcode	The place code for the settlement extent
Shape_Length	Default ESRI field: The shape length in geographic coordinates (decimal degrees)
Shape_Area	Default ESRI field: The shape area in geographic coordinates (decimal degrees)

### **Version History**

This data product contains all information contained in the previous “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01” product, with updates. Updates in this version include: revised terms of use and license, Boundary and Place Codes for each settlement extent.

### **Known data limitations and disclaimer**

The “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01” has not been visually inspected for false negatives. A settlement may exist for a location where there is no building footprint data and consequently no settlement extent. This may lead to settlements not being identified. Likewise, the “GRID3 Democratic Republic of Sao Tome and Principe Settlement Extents, Version 01.01” has not been visually inspected for false positives. A settlement polygon may have been falsely identified as an actual settlement. The type or use of buildings (e.g. residential, non-residential, mixed use, shed, etc) are not known. Boundary information is based on OCHA’s common operational datasets, and is not authoritative.

CIESIN, Columbia University, and GRID3 follow procedures designed to ensure that data disseminated by the project are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact GRID3 at [data.queries@grid3.org](mailto:data.queries@grid3.org).

CIESIN, Columbia University, and their sponsors do not guarantee the accuracy, reliability, or completeness of any data provided. We provide these data without warranty of any kind whatsoever, either expressed or implied, and shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided.

## Acknowledgments

Funding for the development and dissemination of this dataset was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office.

## References

Barau, I., et al. 2014. "Improving Polio Vaccination Coverage in Nigeria Through the Use of Geographic Information System Technology, *The Journal of Infectious Diseases*, Volume 210, Issue suppl\_1, Pages S102–S110, <https://doi.org/10.1093/infdis/jiu010>

Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. "Census/projection-disaggregated gridded population datasets for 51 countries across sub-Saharan Africa in 2020 using building footprints." WorldPop, University of Southampton, UK. doi:10.5258/SOTON/WP00682; doi:10.5258/SOTON/WP00683

Esri Inc. 2020. ArcGIS Pro (Version 2.7.3). Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>

Maxar Technologies, Inc. and Ecopia Tech Corporation. 2020. Ecopia Landbase Africa powered by Maxar, website - "DigitizeAfrica.ai".

McKinney, Wes. 2010. Data Structures for Statistical Computing in Python (Pandas), Proceedings of the 9th Python in Science Conference, 51-56