Data Release Statement
GRID3 São Tomé and Príncipe Settlement Extents, Version 02
May 2023

Abstract

This document outlines the methodological approach and data sources used to construct the *GRID3 São Tomé and Príncipe Settlement Extents, Version 02* dataset. The dataset consists of settlement extents across São Tomé and Príncipe, as well as associated settlement types based on the degree of urbanisation and a probability estimate that each settlement extent is a false positive. The terms of use for these data are also included in this document.

Dataset Citation

Center for International Earth Science Information Network (CIESIN), Columbia University and Novel-T. 2023. *GRID3 São Tomé and Príncipe Settlement Extents, Version 02*. Palisades, NY: Geo-Referenced Infrastructure and Demographic Data for Development (GRID3). https://doi.org/10.7916/02fw-7f81. 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 licence as set forth in the terms of use.

"Humanitarian purposes" is defined as a non-commercial purpose related to supporting operations specifically undertaken to alleviate human suffering caused by events (or series of events) that threaten the health, safety, or well-being of a large group of people.[1]

Both the data and data release statement are licenced under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) licence (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 licence terms, take effect and are legally binding on you (and 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. These licence terms apply to any further distribution of the work or any derivatives of the work you may make.

[1] Ecopia DigitizeAfrica Humanitarian License v1.0 https://digitizeafrica-85c854-e8b1cbfcaf49c6907.webflow.io/digitizeafrica-downsampled-derivatives-license-terms)

The *GRID3 São Tomé and Príncipe Settlement Extents, Version 02* is a derivative work from Digitize Africa, Ecopia Landbase Africa powered by Maxar Ecopia AI under the Ecopia DigitizeAfrica Humanitarian License (https://digitizeafrica-85c854-e8b1cbfcaf49c6907.webflow.io/digitizeafrica-downsampled-derivatives-license-terms)

Copyright 2023. The Trustees of Columbia University in the City of New York.

Contacts and Data Queries

GRID3 appreciates feedback regarding this dataset, including suggestions, discovery of errors, difficulties in using the data, and format preferences. Please contact CIESIN at info@ciesin.columbia.edu.

GRID3
GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT

Contents

Introduction	4
Data and Methodology	4
Dataset Description(s)	13
Version History	15
Known Data Limitations	16
Disclaimer	16
Acknowledgments	16
References	16

I. Introduction

This document details the methodology used to produce the dataset *GRID3 São Tomé and Príncipe Settlement Extents, Version 02*.

Settlement extents are polygons representing areas where the presence of buildings detected by satellite imagery suggest that there is likely a human settlement. Settlement extents are geographic features not meant to represent and may not be consistent with administrative unit or locality boundaries. For example, a single settlement extent may be made up of multiple localities, especially in urban areas.

This work has been undertaken as part of the Geo-Referenced Infrastructure and Demographic Data for Development (GRID3) programme (INV-009579). 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 Project at the University of Southampton, the United Nations Population Fund, and the Center for International Earth Science Information Network (CIESIN) at Columbia University.

II. Data and Methodology

Input Data

Multiple input data were used to create the *GRID3 São Tomé and Príncipe Settlement Extents, Version 02*. The table below lays out the input data used for each settlement extent attribute. The settlement attributes are accessible for each settlement extent polygon through the attribute table.

Table 1. *Input datasets*

Settlement Extent Attribute	Input Data	Description	Source
Settlement Extent SHAPE and Building Count Ranges	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 2021).	Maxar Technologies, Inc. and Ecopia Tech Corporation. 2021. Ecopia Landbase Africa powered by Maxar. Available at https://www.ecopiatech.com Version 2.

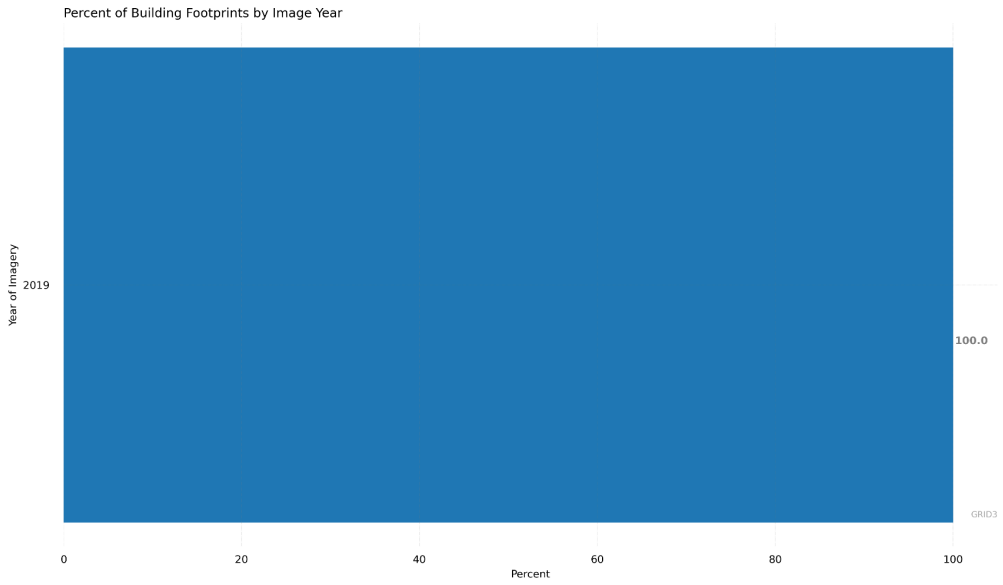
Settlement Extent Attribute	Input Data	Description	Source
Degrees of Urbanisation (DOU_level1; DOU_level2)	WorldPop population estimates	Population estimates (2020): Bottom-up modeled population estimates (100-meter resolution) produced by WorldPop at the University of Southampton. Units: Total population per pixel. Extent: São Tomé and Príncipe https://www.worldpop.org/methods/populations	Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. "Census/projectiondisaggregated 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
		Unconstrained UN-adjusted population estimates (2020): Country population totals for São Tomé and Príncipe are adjusted to match estimates prepared by the United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, Online Edition. Rev. 1.	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. https://hub.worldpop.org/geodata/listing?id=69
	European Commission's Joint Research Centre	Degree of urbanisation population and settlement analysis toolkit. The toolkit contains scripts and processes for creating a degree of urbanisation layer.	GHSL - Global Human Settlement Layer https://ghsl.jrc.ec.europa.eu/tools.php

Settlement Extent Attribute	Input Data	Description	Source
Place Codes (P-codes)	OCHA common operational boundaries	The United Nations Office for the Coordination of Humanitarian Affairs Common Operational Datasets 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.	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 - www.gadm.org). Accessed 10/30/2021. Available from the Humanitarian Data Exchange: https://data.humdata.org/dataset/sao-tome-and-principe-administrative-boundaries
Predicted False Positives (is_fp; prob_fp)	Google	Open Buildings	W. Sirko, S. Kashubin, M. Ritter, A. Annkah, Y.S.E. Bouchareb, Y. Dauphin, D. Keysers, M. Neumann, M. Cisse, J.A. Quinn. Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283, 2021. https://sites.research.google/open-buildings/
	Facebook	High Resolution Settlement Extent	Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL Copyright 2016 DigitalGlobe. Accessed 11/10/2022. Data shared under: Creative Commons Attribution International. https://registry.opendata.aws/dataforgood-fb-hrsl/
	Microsoft	Bing Global ML Building Footprints	Accessed 10/10/2022 https://github.com/microsoft/GlobalMLBuildingFootprints

Imagery Year

Imagery from multiple years is used in the feature extraction of the building footprints.

Figure 1. *Percent of building footprints by imagery year .*



The date of satellite imagery used in the feature extraction varies across a country. This table shows the percentage of building footprints by imagery year for São Tomé and Príncipe.

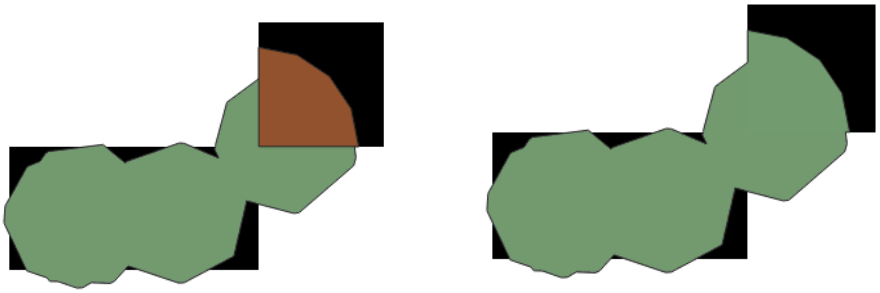
Methods

The settlement extents are derived from Ecopia's building footprints. Rust, Python, GDAL, and LibGEOS were used for data analysis and processing. The code is open source GNU General Public License v3.0 and is available from Novel-T's github site (https://github.com/novelt/building-aggregation-tool).

The data product derived from the Ecopia building footprints must be generalised to approximately 100 meters, according to Ecopia's terms of use. To accomplish this, building footprint features within .001 decimal degrees of each other (approximately 130 metres) are converted to multipart polygons. The multipart polygons are then buffered by 50 metres. All polygons are then unioned and dissolved. Finally, any inner rings caused by open areas, such as parks or sports fields, are filled.

Since the centre point of raster cells are used for calculating zonal statistics, settlement extents are modified to cross the centre point of any worldpop 100m raster cell that a building footprint falls within. In the event that a settlement does not cross the centre point, a corner buffer is merged with the settlement extent to extend the settlement extent past the centre point.

Figure 2. *Example of a corner buffer added to a settlement extent*



Example of geographic edits to facilitate comprehensive coverage of building footprints that fall within settlement extents. The brown corner buffer has been added so that the settlement extent crosses the centre of the raster grid cell (in black).

Settlement Extent Classification

A classification based on the European Commission's Joint Research Centre's degree of urbanisation model[2] was used to classify the settlement extents, and these classifications are included in the dataset. The methodology has been endorsed by the United Nations Statistical Commission and combines total population and population density to classify degrees of urbanisation . WorldPop's population layers were the population source used to classify settlement extents based on the "degree of urbanisation model". In most cases, the input population layers are constrained to settled areas at a 100m resolution. For some countries (i.e. Nigeria, Zambia, Sierra Leone, South Sudan, Democratic Republic of Congo (partial), Mozambique, and Burkina Faso), settled area populations are estimated based on bottom-up methodologies. For other counties, settled area populations are estimated based on a top-down methodology.[3] When population values were not available from either bottom-up or top-down constrained layers, a population value was imputed based on WorldPop's unconstrained datasets.[4]

---

[2] European Commission, FAO, UN-Habitat, OECD, The World Bank. 2020. Applying the Degree of Urbanisation. A methodological manual to define cities, towns and rural areas for international comparisons. https://ec.europa.eu/eurostat/documents/10186/11395216/DEGURBA-manual.pdf/3a6bab6a-3fb1-4261-ad5b-e604cb67dc0d#:~:text=Applying%20the%20Degree%20of%20Urbanisation%20%E2%80%94%20A%20methodological%20manual%20to%20define,FAO)%2C%20the%20United%20Nations%20Human

[3] WorldPop methodologies, https://www.worldpop.org/methods/populations/

[4] Constrained vs unconstrained WorldPop methodologies, https://www.worldpop.org/methods/top_down_constrained_vs_unconstrained/

Table 2. *Degrees of urbanisation*

Level 1	Level 2
Urban Center	Urban Center
Urban Cluster	
	Dense Urban Cluster
	Semit-Dense Urban Cluster
	Suburban
Rural Cluster	
	Rural
	Low Desnsity Rural
	Very Low Density Rural

Note: Information from European Commission, FAO, UN-Habitat, OECD, The World Bank. 2020. Applying the Degree of Urbanisation. A methodological manual to define cities, towns and rural areas for international comparisons, p. 26. © European Union/FAO/UN-Habitat/OECD/World Bank, 2020

Two levels exist within the degrees of urbanisation and are discussed below.

First level of degree of urbanisation classes and general criteria of delineation:

- Urban centres: Clusters of cells with combined populations of ≥ 50,000 inhabitants, where each cell has a density of ≥ 1,500 inhabitants km^2.
- Urban clusters: Clusters of cells with combined populations of ≥ 5,000 inhabitants, where each cell has a density ≥ 300 inhabitants km^2.
- Rural clusters: All cells not belonging to urban clusters or urban centres.

Second level of degree of urbanisation classes and general criteria of delineation:

- Urban Centres: Urban Centres do not change from level 1 to level 2 and are still defined by clusters of cells with combined populations of ≥ 50,000 inhabitants, where each cell has a density of ≥ 1500 inhabitants km^2.
- Dense Urban Cluster: A subcategory of Urban Clusters. It consists of contiguous cells with a density of at least 1,500 inhabitants per km^2, and has between 5,000 and 50,000 inhabitants in the cluster.
- Semi-Dense Urban Cluster: A subcategory of Urban Clusters. It consists of contiguous grid cells with a density of at least 300 inhabitants per km^2, has at least 5,000 inhabitants, and is at least 2km from borders of Urban Centres or Dense Urban Clusters.
- Suburban: A subcategory of Urban Clusters. It includes Urban Cluster cells that are not already classified as Dense or Semi-Dense Urban Clusters. Suburban grid cells are within 2km of a Dense Urban Cluster or an Urban Center.

- Rural: A subcategory of Rural Clusters. It consists of contiguous cells (8-connectivity cluster) with a density of at least 300 inhabitants per km^2 and has between 500 and 5,000 inhabitants.
- Low Density Rural: A subcategory of Rural Clusters. It consists of grid cells with a density of at least 50 inhabitants per km^2 of permanent land.
- Very Low Density Rural: Grid cell that is in Rural Clusters but not part of Rural and Low Density Rural clusters.

Table 3. *Settlement classification summary*

		Minimum population size of the cluster of cells (settlement size)			No minimum population size criterion (not an entity)
		≥50,000	50,000 - 5,000	5,000 - 500	
Population density of cells, inhabitants per km^2	≥1500	Urban Centre	Dense Urban Cluster	Rural Cluster	
	300 - 1500		Semi-Dense Urban Cluster		Suburban or peri urban grid cells
	300 - 50				Low Density Rural grid cells
	<50				Very Low Density Rural grid cells

Note: From "Figure 7.1: Schema for the grid cell classification for level 2 of the degree of urbanisation classification, p. 44" European Commission, FAO, UN-Habitat, OECD, The World Bank. 2020. Applying the Degree of Urbanisation. A methodological manual to define cities, towns and rural areas for international comparisons. © European Union/FAO/UN-Habitat/OECD/World Bank, 2020

Settlement Status: Year 1 - Year 2 Comparison

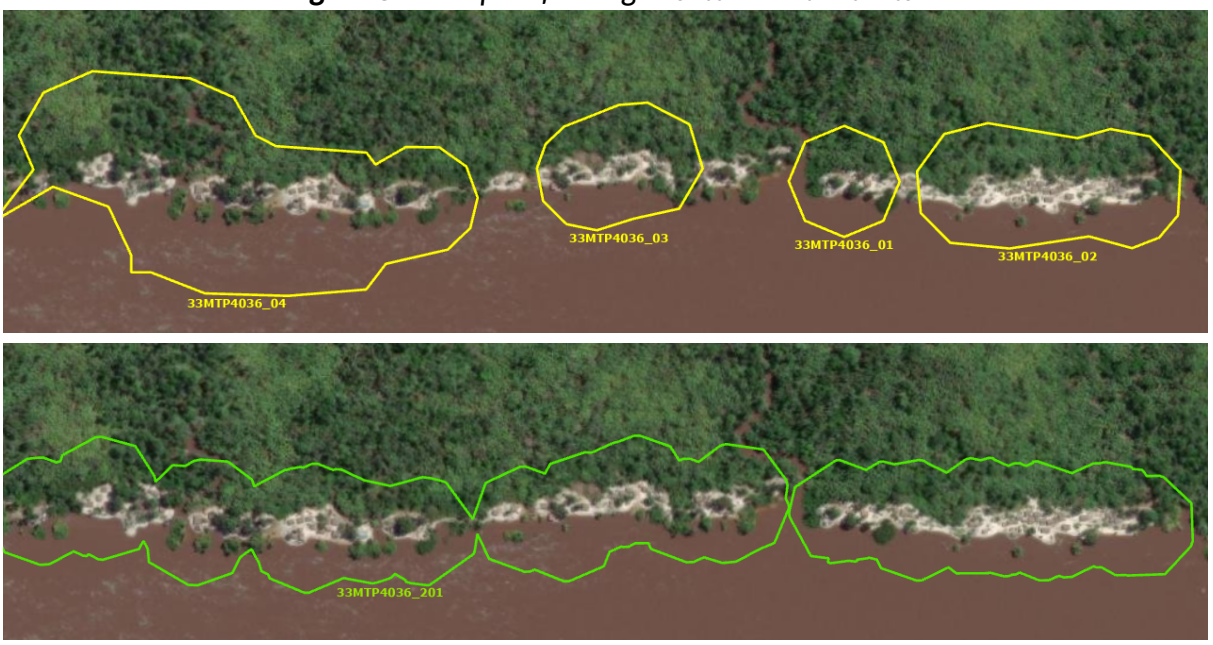
Settlements evolve, they grow, merge, shrink, and split. The settlement status value identifies whether a settlement is existing, has diverged from the previous version or is new. To obtain the settlement status, the percentage of overlap between the Version 01 and Version 02 of the settlement extents was calculated. In order to be considered an "existing" settlement", the percentage of overlap must be more than 50 in the following cases: settlements intersect between the two versions as one-to-many, many-to-one, or many-to-many.

Some settlements have a one-to-one intersection. This happens when a settlement from Version 01 shrinks or is identical to a settlement from Version 02, or if the Version 01 settlement grew without merging with any other settlements around it. In this case, settlements must intersect each other by more than 10 percent to be considered an "existing" settlement.

If a Version 02 settlement intersects with a Version 01 settlement but does not meet either criteria described above, it is marked as "diverged." A diverged extent has changed enough that it is no longer considered the same settlement as in Version 01. Reasons for this include settlements have merged, settlements have split, or the Version 02 data has more precise information due to improved imagery or improved feature extraction.

A settlement is considered "new" if it does not intersect with any extents from the Version 01 settlements at all.

Figure 3. *Example of diverged settlement extents*



Example where the settlement extents from Version 02 (lower green) have diverged from the extents from Version 01 (upper yellow). The settlements have grown causing them to merge together into a single larger settlement. Since the percentage of overlap between any single Version 01 settlement and the newer merged Version 02 settlement is less than 50% the settlement is marked as "diverged" and given a new unique mgrs code.

Place Codes (P-codes)

Place codes (P-codes) are unique identifiers used in common operational datasets for emergency response. They are created based on a set of guidelines defined by OCHA.[5] Settlement P-codes for existing settlements are maintained from the Version 01 settlements. P-codes for new or derivative settlements are generated based on the P-code for the administrative level that the settlement resides fully within and a six digit number unique to that administrative level.

[5]OCHA Information Mangment Wiki - Implementing P-Codes
https://sites.google.com/site/ochaimwiki/geodata-preparation-manual/p-code-guidelines

Identifying False Positives

In some cases, the underlying data used for creating the settlement contains errors. For example, if some crops or rock formations could be falsely identified as a settled area. This type of error is called a "false positive" or "false settlement." A model was created to help identify false positives.

Training Data

A random sample of at least 300 records was drawn from the settlements and visually compared to two sources of imagery: the ESRI imagery basemap and Google Earth Pro imagery. For the latter, multiple dates of imagery were often examined. If no settlements were detected in both sources of imagery, the settlement was marked as a false positive. In addition to the random sample, all of the settlements typed as Urban under the previous scheme (year 1 method) were reviewed to locate any large-area false positives. The results were an unbalanced sample of false positives and true positives, with a strong bias toward true positives.

Models

The samples of false and true positives were used to evaluate three different models to predict false positives: decision tree classifier, random forest classifier, and XGBoost classifier (binary logistic). The data were split into training and test data (80 percent/20 percent sample sizes, respectively) and the models fitted with parameter tuning. Details for each country are shown in Jupyter Notebooks that include the source code and data inputs.

Results varied by country, but overall the XGBoost classifier performance was the best, with F1 statistics over 0.9 for major, minor, and weighted F1 measures. The major F1 measure is important for unbalanced data, as it will reflect the performance of predicting the underrepresented category. After model tuning, the XGBoost model was re-trained on the full set of reviewed data (all false and true positives). The resulting model was used to predict false positives for the complete dataset.

Table 4: *Training model statistics*

São Tomé and Príncipe	Model Selected: XGBoost
False positives identified in review	27 of 355 settlements reviewed
Training Accuracy	1
Macro F1-Score	1
Micro F1-Score	1
Weighted F1-Score	0.999
Total predicted false positives	42 of 743 settlements

Training Model Statistics for São Tomé and Príncipe. Decision tree and random forest models were also evaluated, but the XGBoost model performed best across all measures. The scores are based on model predictions made with test data: 20% of records that were not used in training the models.

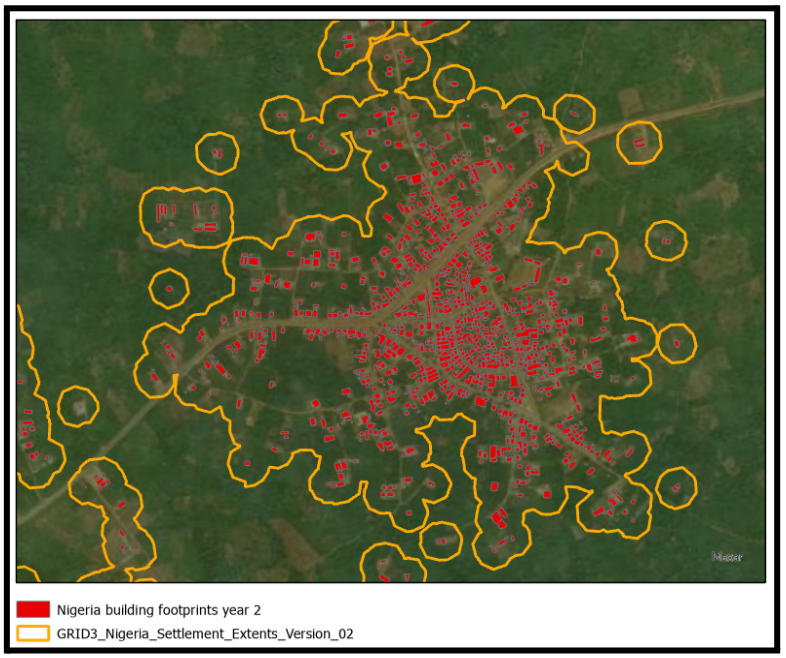
III. Dataset Description(s)

The settlement extent data are in geodatabase format and consist of a single-feature class. An example of data is shown on Figure 4.

Extent: São Tomé and Príncipe. The overall extent of the layer is limited to the overall extent of the building footprint dataset and does not reflect the extent of official administrative boundaries.

Coordinate system: GCS WGS 1984

Figure 4. *Sample building footprints and settlement extents*



Sample map depicting a settlement extent data layer, with building footprint layer for reference. Note that 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
country	Country name
iso	Three-letter country ISO code
bld_count	Building count present within a settlement polygon, represented as a range. The ranges are "1-50," "51-100," "101-250," "251-1000," and "1001 and up." The unit is the number of buildings.
dou_level1	Categorical classification of settlement polygons based on the first level of Degrees of Urbanisation (DoU). The categories are: “urban centers,” “urban clusters,” and “rural clusters.”
dou_level2	Categorical classification of settlement polygons based on the second level of Degrees of Urbanisation (DoU). The categories are: “dense urban cluster,” “semi-dense urban cluster,” “suburban,” “rural,” “low density rural,” and “very low density rural.”

Field	Description
status	The settlement feature as compared with Version 01 of the settlement extents. The status values are “existing”, “new”, or “diverged”.
is _fp	Predictive value for whether the settlement polygon is a false settlement. Value “1” represents a predicted false settlement; value “0” represents a predicted true settlement.
prob_fp	The probability that the settlement is a false positive. Range is between 0 and 1 with values near zero being low probability and values near one being high.
pcode	A place code for the settlement extent based on OCHA’s Common Operational Datasets (COD) boundary's classification.
mgrs	Unique name generated using the Military Grid Reference System[6]
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)
date	The date of the processing
area_m2	The area of the settlement extent in square metres

IV. Version History

This data product is based on an updated building footprints feature extraction produced by “Digitize Africa, Ecopia Landbase Africa powered by Maxar Ecopia AI.” Some settlement extents have remained the same while others have been removed or added based on the new feature extraction. Updates in this version include:

1. The degree of urbanisation has replaced the previous classifications of built-up areas, small settlement areas, and hamlets
2. Boundary names have been removed, since the OCHA dataset is not an official set of boundaries and may not be accurate
3. Building count ranges have been included
4. Predicted false positives have been included
5. Population data have been removed until new constrained population numbers are available
6. Settlement status has been included, as it pertains to Version 01 of the settlement extents

[6] Military Grid Reference System (MGRS) data homepage http://mgrs-data.org

V. Known Data Limitations

The *GRID3 São Tomé and Príncipe Settlement Extents, Version 02* has the following data limitations:

1. It 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.
2. P-codes are generated using OCHA common operational boundary information and are not authoritative
3. The age of satellite imagery varies across a country, and so the settlement extents may not represent current settlement extents
4. Settlement extents represent both residential and non-residential areas
5. Settlement extents are not official settlement boundaries and may include multiple administrative units in a single boundary

VI. Disclaimer

CIESIN, Columbia University follows 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 CIESIN at info@ciesin.columbia.edu.

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.

VII. 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.

VIII. 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

Bing Global ML Building Footprints. Accessed 10/10/2022 https://github.com/microsoft/GlobalMLBuildingFootprints

Bondarenko M., Kerr D., Sorichetta A., and Tatem, A.J. 2020. "Census/projectiondisaggregated 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/enus/arcgis/products/arcgis-pro/overview

European Commission, FAO, UN-Habitat, OECD, The World Bank. 2020. Applying the Degree of Urbanisation. A methodological manual to define cities, towns and rural areas for international comparisons. https://ec.europa.eu/eurostat/documents/10186/11395216/DEGURBA-manual.pdf/3a6bab6a-3fb1-4261-ad5b-e604cb67dc0d#:~:text=Applying%20the%20Degree%20of%20Urbanisation%20%E2%80%94%20A%20methodological%20manual%20to%20define,FAO)%2C%20the%20United%20Nations%20Human

Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL Copyright 2016 DigitalGlobe. Accessed 11/10/2022. Data shared under: Creative Commons Attribution International. https://registry.opendata.aws/dataforgood-fb-hrsl/

GHSL - Global Human Settlement Layer https://ghsl.jrc.ec.europa.eu/tools.php

Maxar Technologies, Inc. and Ecopia Tech Corporation. 2021. Ecopia Landbase Africa powered by Maxar. Available at https://www.ecopiatech.com, Version 2

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

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 - www.gadm.org). Accessed 10/30/2021. Available from the Humanitarian Data Exchange: https://data.humdata.org/dataset/sao-tome-and-principe-administrative-boundaries

W. Sirko, S. Kashubin, M. Ritter, A. Annkah, Y.S.E. Bouchareb, Y. Dauphin, D. Keysers, M. Neumann, M. Cisse, J.A. Quinn. Continental-scale building detection from high resolution satellite imagery. arXiv:2107.12283, 2021. https://sites.research.google/open-buildings/

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/WP00660