

Data Release Statement
***Gridded Population Estimates Disaggregated from Mozambique's Fourth General
Population and Housing Census (2017 Census), version 1.1***

N.B. Veja abaixo a versão em português

26/11/2020

Abstract

This document outlines the data sets and methods used to provide an estimate of the geographic distribution of the population of Mozambique in 2017. Assumptions, limitations, and use constraints are provided.

Suggested Citation

Bondarenko M, Jones P, Leasure D, Lazar AN, Tatem AJ. 2020. Census disaggregated gridded population estimates for Mozambique (2017), version 1.1. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00672

Data Use Constraints

This data set is jointly owned between University of Southampton in the United Kingdom and the National Institute of Statistics (INE) in Mozambique. These data and accompanying document are licensed under a [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) license, specified in [legal code](#). Contact release@worldpop.org for more information.

Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

Contacts and Data Queries

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

Please contact: WorldPop Email: release@worldpop.org for more information.

1. Introduction

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the Geo-referenced Infrastructure and Demographic Data for Development (GRID3) programme, supported with funding from the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (OPP1182425). Programme partners included the United Nations Population Fund (UNFPA), the Center for International Earth Science Information Network (CIESIN) within the Earth Institute at Columbia University, and the Flowminder Foundation.

This work provides an estimate of the geographic distribution of the population of Mozambique in 2017. The outputs are intended as an interim population product to support ongoing development and operations work until such time as the official 2017 Population and Housing Census results are available in a spatial gridded format. At that time, this interim gridded population layer will be superseded and users will be advised to use the official gridded population release from INE.

2. Input Data

- District level population totals from the 2017 Census were downloaded from INE's website (<http://www.ine.gov.mz/iv-rgph-2017>).
- The district boundaries were those used in the 2017 Census provided by INE as a shapefile.
- Building footprints were provided by the Digitize Africa project of Ecopia.AI and Maxar Technologies (2020). The building footprints represent the year 2018 in most areas of Mozambique (60% coverage from 2018, 15% from 2019, 12% from 2017, 7% from 2016, and 5% from 2015 or earlier; Dooley et al., 2020).
- WorldPop global gridded age-sex proportions (WorldPop et al., 2018). We used this data to produce gridded population estimates for each age-sex group (MOZ_population_v1_1_agesex.zip). The WorldPop gridded age-sex proportions were produced using the methods of Pezzulo et al. (2017) and Carioli et al. (in prep).

Additional geospatial covariates (Lloyd et al., 2019), representing factors related to population distribution, were obtained from the "Global High Resolution Population Denominators Project" (OPP1134076). Building footprint derivatives were prepared by Dooley et al. (2020).

3. Methods Overview

The random forest-based dasymetric mapping approach (Stevens et al., 2015; Bondarenko et al., 2020), using the random forest algorithm (Breiman, 2001), was used to produce a gridded population density data set, with a spatial resolution of 3 arc-seconds (approximately 100 m at the equator). This weighting layer was subsequently used to dasymetrically disaggregate population counts from district level totals into grid cells, based on those grid cells which were deemed “settled”. A settled grid cell was defined based on the presence of one or more buildings in the Maxar Technologies building footprint data (Ecopia.AI and Maxar Technologies, 2020). Zonal sums were calculated for all administrative units in the input data to verify that the disaggregated values add up to the original, observed totals.

We multiplied our gridded population estimates (MOZ_population_v1_1_gridded.tif) by the gridded age-sex proportions in each region to produce the rasters in MOZ_population_v1_1_agesex.zip. This process was completed using the “demographic” function from the peanutButter R package (Leasure et al. 2020b).

4. Release Content and File Descriptions

- **MOZ_population_v1_1_README.pdf**
- **MOZ_population_v1_1_gridded.tif**
This GeoTIFF raster (.tif), at a spatial resolution of 3 arc-seconds (approximately 100m at the equator), contains estimated population counts per grid cell across Mozambique. The projection is the geographic coordinate system WGS84 (World Geodetic System 1984). ‘NoData’ values represent areas that were mapped as unsettled based on building footprints provided by the Digitize Africa project of Ecopia.AI and Maxar Technologies (2020). These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated populations for larger areas.
- **MOZ_population_v1_1_agesex.zip**
This zip file contains 40 rasters in GeoTIFF format. Each raster provides gridded population estimates for an age-sex group. We provide 36 rasters for the commonly reported age-sex groupings of sequential age classes for males and females separately. The files are labelled with either an “m” (male) or an “f” (female) followed by the number of the first year of the age class represented by the data. “f0” and “m0” are population counts of under 1 year olds for females and males, respectively. “f1” and “m1” are population counts of 1 to 4 year olds for females and males, respectively. Over 4 years old, the age groups are in five year bins labelled with a “5”, “10”, etc. Eighty year olds and over are represented in the groups “f80” and “m80”.

We provide an additional four rasters that represent aggregated demographic groups of special interest for development programmes and interventions. These are “under1” (all females and males under the age of 1), “under5” (all females and males under the age of 5), “under15” (all females and males under the age of 15) and “f15_49” (all females between the ages of 15 and 49, inclusive). These data were produced using the peanutButter R package (Leasure et al. 2020b) which multiplied the gridded population estimates (MOZ_population_v1_1_gridded.tif) by gridded age-sex proportions that differ by region (WorldPop et al. 2018, Pezzulo et al. 2017).

While each data file represents population counts, individual values contain decimals, i.e. fractions of people. This is because we do not estimate which grid cell each individual in a given age group occupies. For example, if four grid cells next to each other have values of 0.25 this indicates that there is one person of that age group somewhere in those four grid cells.

- **MOZ_population_v1_1_mastergrid.tif**
This geotiff raster contains a one for each grid cell across the study area where buildings occurred and the model was applied to estimate population size (cell size = 0.0008333 decimal degrees; approximately 100m).
- **MOZ_population_v1_1_sql.sql**
This SQLite database contains estimates of population size in each grid cell. This database is source data for the woprVision web application (<https://wopr.worldpop.org/woprVision>) and it can be queried using the wopr R package (Leasure et al. 2020a).

The SQL database contains a single table (Nhat) that includes the population predictions. This table contains the following columns:

- “cell” contains a cell ID to identify the location. Cell IDs correspond to those the cell IDs of MOZ_population_v1_1_mastergrid.tif.
- “x” and “y” columns contain WGS84 coordinates for the centroid of the grid cell.
- “Pop” column contains a population estimate for each grid cell.
- “agesexid” column contains the region ID for the age-sex proportions.
- “area” contains the total settled area in hectares. This corresponds to the total building area from Dooley et al. (2020) in their raster MOZ_buildings_v1_1_total_area.tif.

- **MOZ_population_v1_1_tiles.zip**
This tiled web map allows for rapid display of the approximately 100 m gridded population estimates across the study area (i.e. MOZ_population_v1_1_gridded.tif). These can be used to develop web applications for the model results. The tiles were created in XYZ format (i.e.

compatible with Leaflet) with full coverage of the study area for the zoom levels 1 to 14. These tiles are source data for the woprVision web application (<https://wopr.worldpop.org/woprVision>).

5. Release History

- Version 1.1 – 26 November 2020
The model was refit with updated building footprints (Ecopia.AI & Maxar Technologies 2020).
- Version 1.0 – 16 June 2020 (This is the original limited release.)

6. Assumptions and Limitations

- i. This data set was produced based on the district-level 2017 Population and Housing Census data provided by the National Institute of Statistics (INE) (<http://www.ine.gov.mz/iv-rgph-2017>), which is publicly available for download. It is intended as an interim product for use by GRID3 stakeholders for planning and humanitarian response in Mozambique.
- ii. District level population counts were disaggregated into grid cells, using the random forest model. However, we acknowledge that the random forest model works best with high counts of administration units. Mozambique only has 161 districts, and we recognise that this is at a coarse resolution and may impact the accuracy of the disaggregated population at 100m.
- iii. District level population totals are currently not available for Zambezia province. Instead, an earlier gridded data set obtained from INE was used. For consistency with other districts, the data were re-aggregated to the boundary and used in the random forest model. We acknowledge that aggregating and re-disaggregating the same data will impact the accuracy of the estimates produced.
- iv. Disaggregation of the 2017 Census district population totals to cells containing buildings in the building footprint layer may result in differing output per district in this interim population raster when compared to the totals published on the INE website. This is due to rounding errors, but also because some of the building footprints cross district boundaries. To help mitigate this and prevent the double-counting of data around district boundaries, the centroids of the buildings were used to assign the building to a single district.
- v. The official administrative boundaries were used for this work and they may not match other datasets.
- vi. There is currently no way to distinguish between residential and non-residential buildings in the Maxar building footprint data, therefore, population will be allocated to large industrial buildings and warehouses, as well as to residential buildings.

7. Acknowledgments

Funding for the development and dissemination of this data set was provided by the Bill & Melinda Gates Foundation and the United Kingdom's Foreign, Commonwealth & Development Office (OPP1182425).

Patricia Jones and Donna Clarke (WorldPop) supported the generation of inputs for the application of the random forest-based dasymetric mapping approach developed by Stevens et al. (2015). The population disaggregation was done by Maksym Bondarenko (WorldPop), using the random forest population modelling R scripts (Bondarenko et al., 2020). The age-sex rasters, SQL database, and map tiles were created by Doug Leasure (WorldPop). The administrative boundaries and population totals were provided by Arlindo Charles, at INE (National Institute of Statistics) in Mozambique. Claire Dooley and Chris Jochem reviewed the population raster and its documentation, and provided comments and suggestions. Attila N. Lazar (WorldPop) coordinated the work with substantial engagement and translation support from Sandra Baptista (CIESIN). The whole WorldPop group and GRID3 partners are acknowledged for overall project support.

8. References

- Bondarenko M, Nieves JJ, Stevens FR, Gaughan AE, Tatem A, Sorichetta A. 2020. wpgpRFPMS: Random Forests population modelling R scripts, version 0.1.0. University of Southampton: Southampton, UK. <https://dx.doi.org/10.5258/SOTON/WP00665>
- Breiman, L. 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Carioli A, Pezzulo C, Hanspal S, Hilber T, Hornby G, Kerr D, Tejedor-Garavito N, Nielsen K, Pistolesi L, Adamo S, Mills J, Nieves JJ, Chamberlain H, Bondarenko M, Lloyd C, Yetman G, Gaughan A, Stevens F, Linard C, James W, Sorichetta A, Tatem AJ. In prep. Population structure by age and sex: a multi-temporal subnational perspective.
- Dooley CA, Boo G, Leasure DR, Tatem, AJ. 2020. Gridded maps of building patterns throughout sub-Saharan Africa, version 1.1. WorldPop, University of Southampton. Source of building footprints: Ecopia Vector Maps Powered by Maxar Satellite Imagery (C) 2020. doi:10.5258/SOTON/WP00677
- Instituto Nacional de Estatística (Moçambique). 2019. Dados Definitivos do IV Recenseamento Geral da População e Habitação, 2017. <http://www.ine.gov.mz/iv-rgph-2017>
- Leasure DR, Bondarenko M, Darin E, Tatem AJ. 2020a. wopr: An R package to query the WorldPop Open Population Repository, version 0.4.0. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00688
- Leasure DR, Dooley CA, Bondarenko M, Tatem AJ. 2020b. peanutButter: An R package to produce rapid-response gridded population estimates from building footprints, version 0.3.0. WorldPop Research Group, University of Southampton. <https://dx.doi.org/10.5258/SOTON/WP00681>

- Lloyd CT, Chamberlain H, Kerr D, Yetman G, Pistolesi L, Stevens FR, Gaughan AE, Nieves JJ, Hornby G, MacManus K, Sinha P, Bondarenko M, Sorichetta A, Tatem AJ. 2019. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 3(2), 108-139. <https://dx.doi.org/10.1080/20964471.2019.1625151>
- Ecopia.AI. and Maxar Technologies 2020. Digitize Africa data.
<http://digitizeafrica.ai>
- Pezzulo C, Hornby GM, Sorichetta A, Gaughan AE, Linard C, Bird TJ, Kerr D, Lloyd CT, Tatem AJ. 2017. Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Sci. Data* 4:170089.
<https://dx.doi.org/10.1038/sdata.2017.89>
- Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* 10(2), e0107042.
<https://doi.org/10.1371/journal.pone.0107042>
- 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 & Melinda Gates Foundation (OPP1134076).

Declaração de Lançamento de Dados
Estimativas de População em Grelha Desagregadas do IV Recenseamento Geral da População e Habitação de Moçambique (Censo 2017), versão 1.1

26/11/2020

Resumo

Este documento descreve os conjuntos de dados e métodos usados para fornecer uma estimativa da distribuição geográfica da população de Moçambique em 2017. São fornecidas premissas, limitações, e restrições de uso.

Citação Sugerida

Bondarenko M, Jones P, Leasure D, Lazar AN, Tatem AJ. 2020. Census disaggregated gridded population estimates for Mozambique (2017), version 1.1. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00672

Restrições de Uso de Dados

Este conjunto de dados é de propriedade conjunta entre a Universidade de Southampton no Reino Unido e o Instituto Nacional de Estatística (INE) em Moçambique. Esses dados podem ser distribuídos usando uma [licença Creative Commons Atribuição 4.0 Internacional \(CC BY 4.0\)](#), especificado no [código legal](#). Entre em contacto com release@worldpop.org para obter mais informações.

Os usuários podem usar, copiar, distribuir, transmitir, e adaptar o trabalho para fins comerciais e não comerciais, sem restrições, desde que seja fornecida uma atribuição clara da fonte.

Contactos e Consultas de Dados

O WorldPop aprecia feedback sobre esse conjunto de dados, como sugestões, descoberta de erros, dificuldades no uso dos dados, e preferências de formato.

Por favor entre em contacto: WorldPop Email: release@worldpop.org para obter mais informações.

1. Introdução

Estes dados foram produzidos pelo Grupo de Pesquisa WorldPop da Universidade de Southampton. Este trabalho fez parte do programa Dados Geo-referenciados das Infra-estruturas e Demográficos para o Desenvolvimento (GRID3), apoiado com financiamento da Fundação Bill & Melinda Gates e do Foreign, Commonwealth & Development Office do Reino Unido. Os parceiros do programa incluíram o Fundo das Nações Unidas para a População (UNFPA), o Centro da Rede Internacional de Informações sobre Ciências da Terra (CIESIN) no Earth Institute da Universidade Columbia, e a Fundação Flowminder.

Este trabalho fornece uma estimativa da distribuição geográfica da população de Moçambique em 2017. Os resultados destinam-se a um produto provisório da população para apoiar o trabalho contínuo de desenvolvimento e operações até que os resultados oficiais do Censo da População e Habitação de 2017 estejam disponíveis em formato espacial em grelha (raster). Futuramente, essa camada provisória da população em grelha será substituída e os usuários serão aconselhados a usar a versão oficial da população em grelha divulgada pelo INE.

2. Dados de Entrada

- Os totais da população no nível distrital do Censo 2017 foram baixados do site do INE (<http://www.ine.gov.mz/iv-rgph-2017>).
- As divisões distritais foram os utilizados no Censo 2017 fornecido pelo INE como um arquivo em formato shapefile.
- As pegadas de edifícios foram fornecidas pelo projecto Digitize Africa da Ecopia.AI e Maxar Technologies (2020). As pegadas dos edifícios representam o ano 2018 na maioria das áreas de Moçambique (60% de cobertura a partir de 2018, 15% a partir de 2019, 12% a partir de 2017, 7% a partir de 2016, e 5% a partir de 2015 ou antes; Dooley et al., 2020).
- Proporções globais de idade-sexo em grelha do WorldPop (WorldPop et al., 2018). Usamos esses dados para produzir estimativas populacionais em grelha para cada grupo de idade e sexo (MOZ_population_v1_0_agesex.zip). As proporções de idade e sexo em grelha do WorldPop foram produzidas usando os métodos de Pezzulo et al. (2017) e Carioli et al. (em preparação).

Covariáveis geoespaciais adicionais (Lloyd et al., 2019), representando fatores relacionados à distribuição da população, foram obtidas no "Projecto Global de Denominação Populacional de Alta Resolução" (OPP1134076). Os derivados das pegadas de edifícios foram preparados por Dooley e Tatem (2020).

3. Visão Geral dos Métodos

A abordagem de mapeamento dasimétrico baseado em floresta aleatória (Stevens et al., 2015; Bondarenko et al., 2020), usando o algoritmo floresta aleatória (Breiman, 2001), foi utilizado para produzir um conjunto de dados de densidade populacional em grelha (raster), com uma resolução espacial de 3 arco-segundos (aproximadamente 100 m no equador). Essa camada de ponderação foi subsequenteiramente usada para desagregar dasimetricamente as contagens da população dos totais do nível distrital em células da grelha, com base nas células da grelha que foram consideradas “assentadas”. Uma célula de grelha assentada foi definida com base na presença de um ou mais edifícios nos dados de pegadas de edifícios da Maxar Technologies (Ecopia.AI e Maxar Technologies, 2020). As somas zonais foram calculadas para todas as unidades administrativas nos dados de entrada para verificar se os valores desagregados somam os totais observados originais.

Multiplicamos nossas estimativas de população em grelha (MOZ_population_v1_1_gridded.tif) pelas proporções de idade e sexo em cada região para produzir os rasters em MOZ_population_v1_1_agesex.zip. Esse processo foi concluído usando a função “demográfica” do pacote R peanutButter (Leasure et al. 2020b).

4. Conteúdo do Lançamento de Dados e Descrições dos Arquivos

- **MOZ_population_v1_1_README.pdf**
- **MOZ_population_v1_1_gridded.tif**
Este raster no formato GeoTIFF (.tif), com resolução espacial de 3 arco-segundos (aproximadamente 100m no equador), contém contagens de população estimadas por célula da grelha, abrangendo Moçambique. A projeção é o sistema de coordenadas geográficas WGS84 (World Geodetic System 1984). Os valores de 'NoData' (SemDados) representam áreas que foram mapeadas como sem assentamento com base nas pegadas de edifícios fornecidas pelo projecto Digitize Africa da Ecopia.AI e Maxar Technologies (2020). Esses dados são armazenados como números de ponto flutuante em vez de números inteiros para evitar erros de arredondamento em populações agregadas para áreas maiores.
- **MOZ_population_v1_1_agesex.zip**
Este arquivo zip contém 40 rasters no formato GeoTIFF. Cada raster fornece estimativas populacionais em grelha para um grupo de idade e sexo. Fornecemos 36 rasters para os grupos de idade e sexo comumente relatados de classes de idade sequenciais para pessoas masculinas e femininas separadamente. Os arquivos são rotulados com um “m” (masculino) ou um “f” (feminino) seguido pelo número do primeiro ano da

classe etária representada pelos dados. "f0" e "m0" são contagens populacionais de menores de um ano para pessoas femininas e masculinas, respectivamente. "f1" e "m1" são contagens populacionais de 1 a 4 anos para pessoas femininas e masculinas, respectivamente. Acima de 4 anos, as faixas etárias estão em grupos de cinco anos rotuladas com "5", "10", etc. As pessoas de oitenta anos ou mais estão representadas nos grupos "f80" e "m80".

Nós fornecemos mais quatro rasters que representam grupos demográficos agregados de interesse especial para programas e intervenções de desenvolvimento. São "under1" (abaixo de 1; todas as pessoas femininas e masculinas com menos de 1 ano), "under5" (abaixo de 5; todas as pessoas femininas e masculinas com menos de 5 anos), "under15" (abaixo de 15; todas as pessoas femininas e masculinas com menos de 15 anos) e "f15_49" (todas as pessoas femininas entre 15 e 49 anos, inclusive). Esses dados foram produzidos usando o pacote peanutButter R (Leasure et al. 2020b) que multiplicou as estimativas populacionais em grelhas (MOZ_population_v1_1_gridded.tif) por proporções de idade e sexo em grelhas que diferem por região (WorldPop et al. 2018, Pezzulo et al. 2017).

Embora cada arquivo de dados represente as contagens populacionais, os valores individuais contêm números decimais, ou seja, frações de pessoas. Isso ocorre porque não estimamos qual célula da grelha cada indivíduo em uma determinada faixa etária ocupa. Por exemplo, se quatro células da grelha próximas umas das outras tiverem valores de 0,25, isso indica que há uma pessoa dessa faixa etária em algum lugar nessas quatro células da grelha.

- **MOZ_population_v1_1_mastergrid.tif**
Este raster geotiff contém um para cada célula da grelha na área de estudo onde ocorreram edifícios e o modelo foi aplicado para estimar o tamanho da população (tamanho da célula = 0,0008333 graus decimais; aproximadamente 100m).
- **MOZ_population_v1_1_sql.sql**
Esta base de dados SQLite contém estimativas do tamanho da população em cada célula da grelha. Esta base de dados é fonte de dados para a aplicação web woprVision (<https://wopr.worldpop.org/woprVision>) e pode ser consultada utilizando o pacote wopr R (Leasure et al. 2020a).

A base de dados SQL contém uma única tabela (Nhat) que inclui as previsões da população. Esta tabela contém as seguintes colunas:

- "cell" contém um ID de célula para identificar a localização. Os IDs de célula correspondem aos IDs de célula de MOZ_population_v1_1_mastergrid.tif.

- As colunas "x" e "y" contêm as coordenadas WGS84 para o centróide da célula da grelha.
 - A coluna "Pop" contém uma estimativa da população para cada célula da grelha.
 - A coluna "agesexid" contém o ID da região para as proporções de idade e sexo.
 - A coluna "area" contém a área total ocupada em hectares. Isto corresponde à área total de edifícios de Dooley et al. (2020) em seu raster MOZ_buildings_v1_1_total_area.tif.
- **MOZ_population_v1_1_tiles.zip**
Este mapa web em mosaico permite a visualização rápida das estimativas de população em grelha de aproximadamente 100 m em toda a área de estudo (ou seja, MOZ_population_v1_1_gridded.tif). Estes podem ser utilizados para desenvolver aplicativos web para os resultados do modelo. Os ladrilhos foram criados em formato XYZ (ou seja, compatível com Leaflet) com cobertura total da área de estudo para os níveis de zoom 1 a 14. Estes azulejos são dados de fonte para o aplicativo web woprVision (<https://wopr.worldpop.org/woprVision>).

5. Histórico do Lançamento

- Versão 1.1 - 26 de novembro de 2020
O modelo foi executado novamente com pegadas de edifícios actualizadas (Ecopia.AI & Maxar Technologies 2020).
- Versão 1.0 - 16 de Junho de 2020 (Este é o lançamento limitado original.)

6. Suposições e Limitações

- i. Este conjunto de dados foi produzido com base nos dados do Censo 2017 de População e Habitação, fornecidos pelo Instituto Nacional de Estatística (INE) (<http://www.ine.gov.mz/iv-rgph-2017>), publicamente disponível para download. Ele se destina como um produto intermediário para uso das partes interessadas do GRID3 no planeamento e resposta humanitária em Moçambique.
- ii. As contagens populacionais no nível distrital foram desagregadas em células da grelha, usando o modelo de floresta aleatória (*random forest model*). No entanto, reconhecemos que o modelo de floresta aleatória funciona melhor com altas contagens de unidades de administração. Moçambique possui apenas 161 distritos, e reconhecemos que esta baixa resolução pode ter um impacto na precisão da população desagregada a 100m.
- iii. Atualmente, os totais da população no nível distrital não estão disponíveis para a província da Zambézia. No lugar, um conjunto de dados em grelha

- anterior obtido do INE foi usado. Para consistência com outros distritos, os dados foram agregados novamente aos limites e utilizados no modelo de floresta aleatória. Reconhecemos que agregar e desagregar os mesmos dados afetará a precisão das estimativas produzidas.
- iv. A desagregação dos totais da população do Censo 2017 no nível distrital em células que contêm edifícios na camada de pegadas de edifícios pode resultar em resultados diferentes por distrito nesse raster interina da população em comparação com os totais publicados no site do INE. Isso ocorre devido a erros de arredondamento, mas também porque algumas das pegadas de edifícios ultrapassam os limites do distrito. Para ajudar a mitigar isso e impedir a contagem dupla de dados em torno dos limites dos distritos, os centróides dos edifícios foram usados para atribuir o edifício a um único distrito.
 - v. As divisões administrativas oficiais foram usadas para este trabalho e podem não corresponder a outros conjuntos de dados.
 - vi. Atualmente, não há como distinguir entre edifícios residenciais e não residenciais nos dados de pegadas de edifícios do Maxar; portanto, a população será alocada para grandes edifícios industriais e armazéns, bem como para edifícios residenciais.

7. Agradecimentos

O financiamento para o desenvolvimento e a disseminação desse conjunto de dados foi fornecido pela Fundação Bill & Melinda Gates e pelo Foreign, Commonwealth & Development Office do Reino Unido (OPP1182425).

Patricia Jones e Donna Clarke (WorldPop) apoiaram a geração de entradas para a aplicação da abordagem de mapeamento dasimétrico baseado em floresta aleatória, desenvolvido por Stevens et al. (2015). A desagregação populacional foi realizada por Maksym Bondarenko (WorldPop), usando scripts em R de modelagem populacional floresta aleatória (Bondarenko et al., 2020). Os rasters idade-sexo, o banco de dados SQL, e os blocos de mapas foram criados por Doug Leasure (WorldPop). As divisões administrativas e os totais das populações foram fornecidos por Arlindo Charles no INE (Instituto Nacional de Estatística) em Moçambique. Claire Dooley e Chris Jochem revisaram o raster da população e sua documentação e forneceram comentários e sugestões. Attila N. Lazar (WorldPop) coordenou o trabalho com apoio de engajamento substancial e tradução de Sandra Baptista (CIESIN). Todo o grupo WorldPop e os parceiros do GRID3 são reconhecidos pelo suporte geral ao projecto.

8. Referências

Bondarenko M, Nieves JJ, Stevens FR, Gaughan AE, Tatem A, Sorichetta A. 2020. wpgpRFPMS: Random Forests population modelling R scripts, version 0.1.0.

- University of Southampton: Southampton, UK.
<https://dx.doi.org/10.5258/SOTON/WP00665>
- Breiman, L. 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Carioli A, Pezzulo C, Hanspal S, Hilber T, Hornby G, Kerr D, Tejedor-Garavito N, Nielsen K, Pistolesi L, Adamo S, Mills J, Nieves JJ, Chamberlain H, Bondarenko M, Lloyd C, Yetman G, Gaughan A, Stevens F, Linard C, James W, Sorichetta A, Tatem AJ. In prep. Population structure by age and sex: a multi-temporal subnational perspective.
- Dooley CA, Boo G, Leasure DR, Tatem, AJ. 2020. Gridded maps of building patterns throughout sub-Saharan Africa, version 1.1. WorldPop, University of Southampton. Source of building footprints: Ecopia Vector Maps Powered by Maxar Satellite Imagery (C) 2020. doi:10.5258/SOTON/WP00677
- Instituto Nacional de Estatística (Moçambique). 2019. Dados Definitivos do IV Recenseamento Geral da População e Habitação, 2017.
<http://www.ine.gov.mz/iv-rgph-2017>
- Leasure DR, Bondarenko M, Darin E, Tatem AJ. 2020a. wopr: An R package to query the WorldPop Open Population Repository, version 0.4.0. WorldPop, University of Southampton. doi:10.5258/SOTON/WP00688
- Leasure DR, Dooley CA, Bondarenko M, Tatem AJ. 2020b. peanutButter: An R package to produce rapid-response gridded population estimates from building footprints, version 0.3.0. WorldPop Research Group, University of Southampton. <https://dx.doi.org/10.5258/SOTON/WP00681>
- Lloyd CT, Chamberlain H, Kerr D, Yetman G, Pistolesi L, Stevens FR, Gaughan AE, Nieves JJ, Hornby G, MacManus K, Sinha P, Bondarenko M, Sorichetta A, Tatem AJ. 2019. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*, 3(2), 108–139. <https://dx.doi.org/10.1080/20964471.2019.1625151>
- Ecopia.AI. and Maxar Technologies 2020. Digitize Africa data.
<http://digitizeafrica.ai>
- Pezzulo C, Hornby GM, Sorichetta A, Gaughan AE, Linard C, Bird TJ, Kerr D, Lloyd CT, Tatem AJ. 2017. Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Sci. Data* 4:170089.
<https://dx.doi.org/10.1038/sdata.2017.89>
- Stevens FR, Gaughan AE, Linard C, Tatem AJ. 2015. Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLoS ONE* 10(2), e0107042.
<https://doi.org/10.1371/journal.pone.0107042>
- 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 & Melinda Gates Foundation (OPP1134076).