



26 de julho de 2021

# The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use.

Aluno: Diego Moreira Silva

POPULAÇÃO, ESPAÇO E AMBIENTE

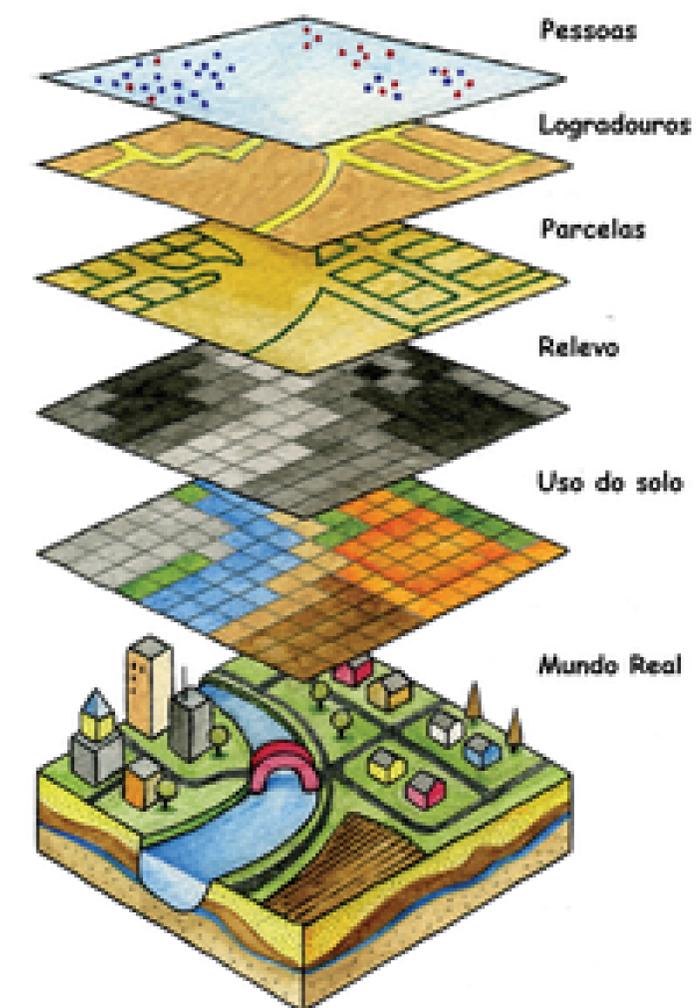
LEYK, Stefan  
et al., 2019

Open Access Earth System  
Science  
Data



Fonte: Home | POPGRID

Uma representação espacial de contagem de população ou densidade populacional por meio de uma grade de células quadrilaterais, cobrindo uma determinada extensão na superfície da Terra (POPGRID, 2020).





A Data Collaborative for  
Settlement, Infrastructure, and Population Data



Improving Accessibility and Documentation  
of Data Sets and Data Services



Encouraging Innovation and Cross-Disciplinary Use

## Disseminar dados gratuitos à todos!

- Melhorar o acesso aos dados;
- Uso de dados para interpretação, sem confusão;
- Identificar e atender às necessidades do usuário final;
- Inovação e uso interdisciplinar;
- Promover a cooperação de produção e harmonização de dados de alta qualidade;
- Conveniência à comunidade que utiliza dados geoespaciais e demográficos.



Fonte: Home | POPGRID

## Nos primórdios ...

- Problemas de agregação

## Atualmente...

- Avanço no Sensoriamento Remoto e revolução dos dados espaciais;
- Complementar (não substituir) os dados do Censo;
- Estimativas mais refinadas para áreas não-administrativas;
- Monitoramento da população (crescimento, mudanças no plano de intervenção).

# Objetivos

- Orientar os usuários finais de forma adequada;
- Descrever as principais fontes de dados de grades populacionais globais e locais;
- Prover um melhor entendimento da natureza desses produtos, a qualidade e os usos apropriados.

Exemplo relacionado a desastres:

A exposição de ma população-alvo requer uma grade populacional que :

- (a) cubra a área de estudo;
- (b) forneça uma unidade analítica significativa (relativo ao tamanho da célula que compõe a grade);
- (c) garante valor necessário em relação ao tempo de interesse;
- (d) estimativa correta da população.

# Distribuição populacional

**(a) Ponderação de área;**

Interpolação simples que não utiliza dados auxiliares.

**(b) Dasimétricos: pesos binários;**

Utiliza variáveis auxiliares binárias, refinando as áreas com pesos.

**(c) Dasimétricos: empiricamente derivado por pesos;**

Uso de vários pesos populacionais que podem ser empiricamente derivado ou baseado em regras definidas

**(d) Dasimétricos: estatisticamente derivado por pesos;**

Modelagem estatística que visa estimar relações que podem ser usadas para modelagem populacional.

(a) Areal weighting

| Unit A, n=54 |   |   | Unit B, n=36 |   |   |
|--------------|---|---|--------------|---|---|
| 6            | 6 | 6 | 4            | 4 | 4 |
| 6            | 6 | 6 | 4            | 4 | 4 |
| 6            | 6 | 6 | 4            | 4 | 4 |

(c) Dasymetric: empirically derived weights e.g., land use classes (different grey tones)

| Unit A, n=54 |     |     | Unit B, n=36 |   |   |
|--------------|-----|-----|--------------|---|---|
| 4            | 4   | 4   | 3            | 3 | 3 |
| 4            | 8.5 | 8.5 | 7.5          | 3 | 3 |
| 4            | 8.5 | 8.5 | 7.5          | 3 | 3 |

(b) Dasymetric: binary weights Grey: built-up land Hashed: no built-up

| Unit A, n=54 |      |      | Unit B, n=36 |   |   |
|--------------|------|------|--------------|---|---|
| 0            | 0    | 0    | 0            | 0 | 0 |
| 0            | 13.5 | 13.5 | 18           | 0 | 0 |
| 0            | 13.5 | 13.5 | 18           | 0 | 0 |

(d) Dasymetric: statistically derived weights (informed by multiple ancillary variables):  $\hat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$

| Unit A, n=54 |     |     | Unit B, n=36 |   |   |
|--------------|-----|-----|--------------|---|---|
| 5            | 5.5 | 5.5 | 3            | 2 | 2 |
| 5            | 7   | 7   | 9            | 3 | 2 |
| 5            | 7   | 7   | 9            | 3 | 3 |

Fonte: LEYK, Stefan et al. , 2019

# Métodos

- **Interpolação de área (realocação populacional)**

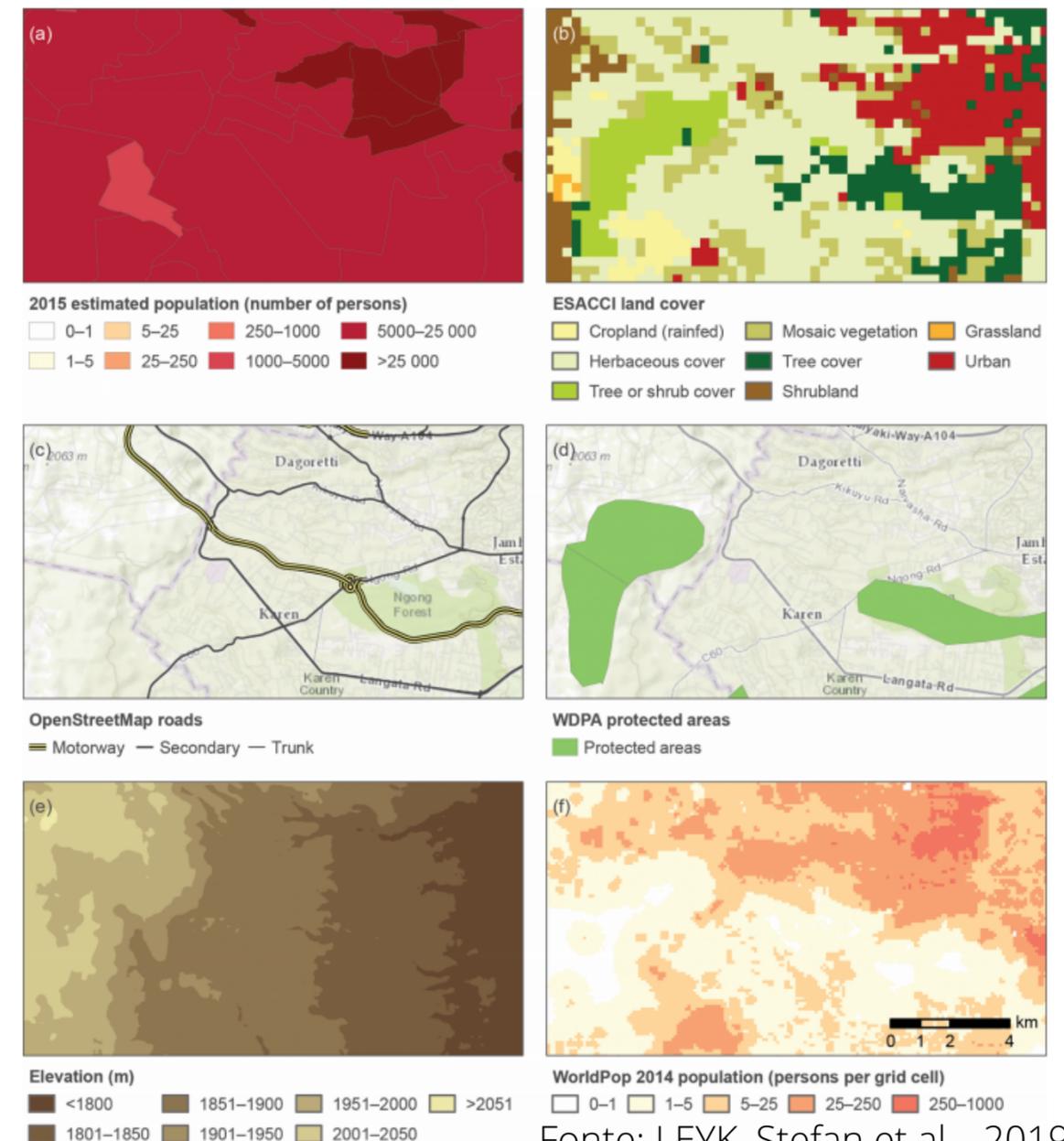
Redistribui determinada fonte de dados em grid de células alvo baseado na proporção da sobreposição com dados não-auxiliares. **Densidade dada por pesos iguais.**

- **Dasimétrico**

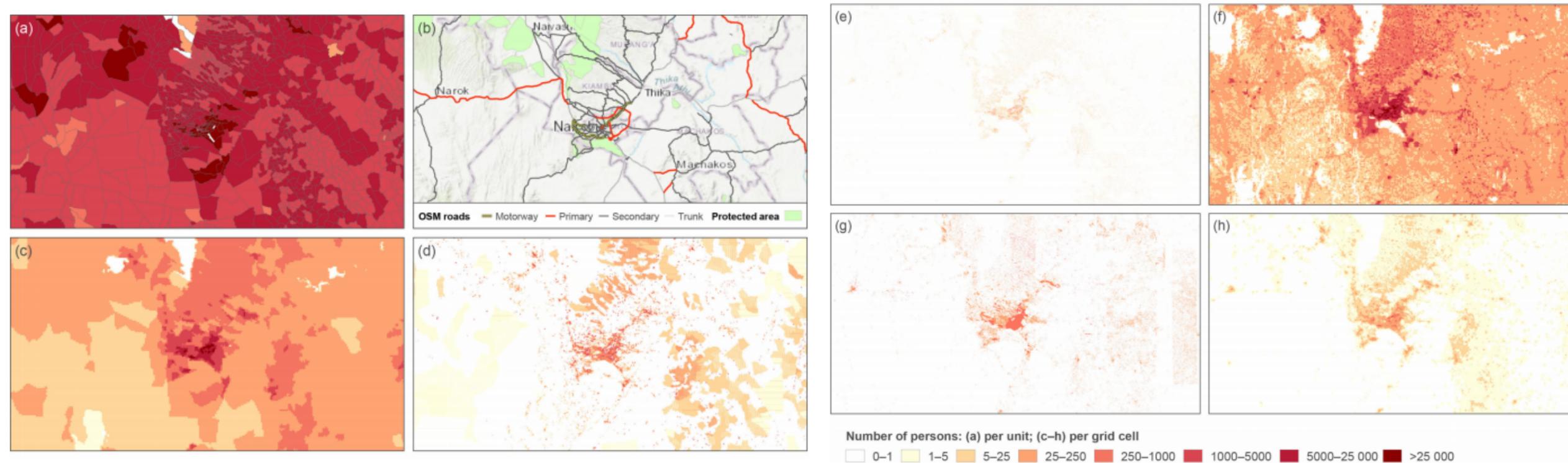
Interpolação da área fonte que utiliza variáveis auxiliares para célula alvo. O volume é preservado e maior representatividade da distribuição da população. **Restringe possíveis ocorrências na unidade original (limite da proporção na mesma unidade).**

- **Método híbrido**

Possibilidade de utilizar muitas variáveis de mesma forma; é possível com inferência estatística; e melhor acurácia. Segundo Leyk et al apresenta resultados mais promissores.



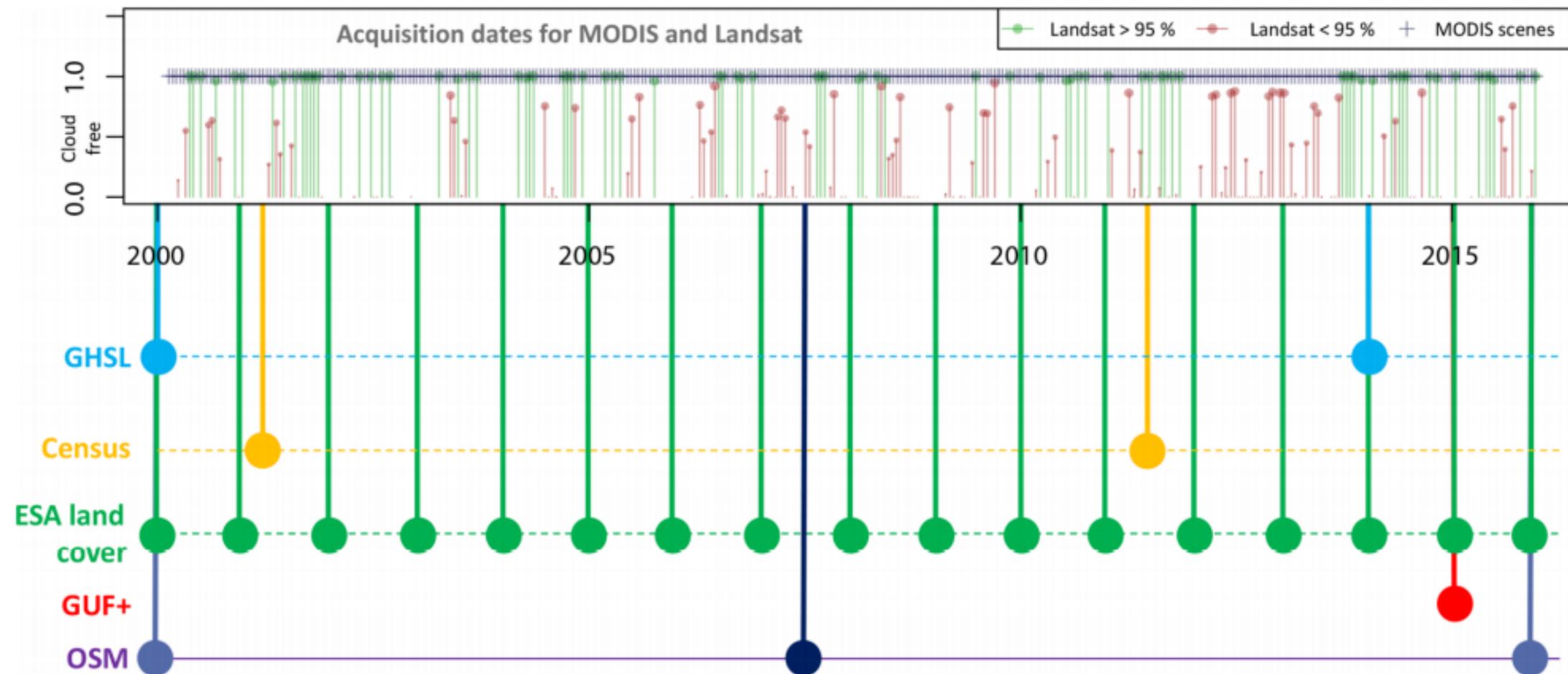
# Métodos



**Figure 4.** Illustration of population input, exemplary ancillary data and different outcome data for a larger region around Nairobi, Kenya: (a) Kenya National Bureau of Statistics, Population and Housing Census 2009, level 5 population units (Center for Development and Environment, Kenyan Atlas Project); (b) World Topographic Map (Esri, HERE, FAO, GeoBase) with roads (OpenStreetMap; © OpenStreetMap contributors 2019; distributed under a Creative Commons BY-SA License) and the World Database of Protected Areas (March 2019 Release); (c) Gridded Population of the World version 4, revision 10, UN adjusted 2015 population count (1 km); (d) Global Human Settlement Layer 2015 population count (250 m); (e) High-Resolution Settlement Layer 2015 population count (30 m); (f) LandScan 2015 population count (1 km); (g) Esri World Population Estimate 2016 population count (150 m); (h) WorldPop 2014 population count (100 m).

Fonte: LEYK, Stefan et al., 2019

# Métodos



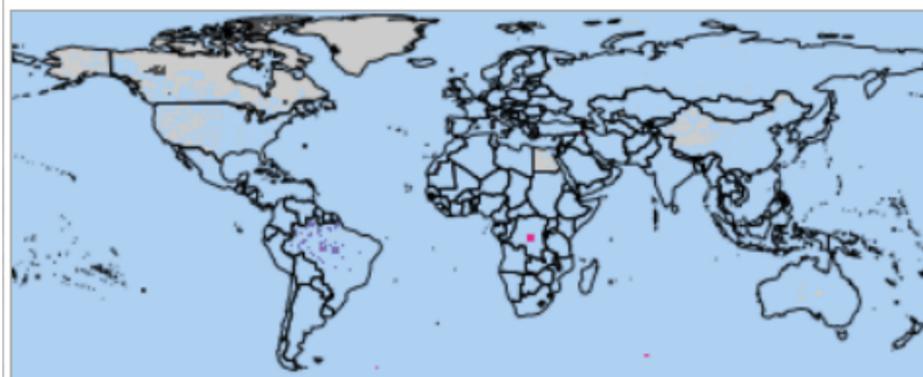
Fonte: LEYK, Stefan et al. , 2019

# Incertezas

- Diferentes métodos empregam diferentes variáveis e operam em suposições variadas, que podem resultar em diferentes estimativas;
- Entender as relações e tendências (o que demonstrar e conhecer o dado);
- É importante reconhecer o erro que acompanha a estimativa resultante de tais abordagens de redistribuição;
- Os erros ou incertezas tendem a ocorrer em estudos mais complexos que envolvem muitos dados (censo e auxiliares).

# Esforços de produção de dados populacionais globais e continentais

GPWv4 Data Quality Indicators, v4.11 Data Context



Fonte: Gridded Population of the World (GPW), v4 | SEDAC (columbia.edu)



Fonte: [landscan.ornl.gov](http://landscan.ornl.gov)



**GHS-POP**

Fonte: [ghsl.jrc.ec.europa.eu/ghs\\_pop.php](http://ghsl.jrc.ec.europa.eu/ghs_pop.php)



Fonte: [www.popgrid.org/data-providers](http://www.popgrid.org/data-providers)

# Perguntas importantes!

- 1 Qual a importância do grid populacional com refinamento no seu trabalho?
- 2 Qual é a população alvo do seu estudo?
- 3 O grid populacional está sendo usado para modelar outros dados?
- 4 Possui abordagem de mudança no tempo?
- 5 Como esses conjuntos de dados foram usados anteriormente?



Fonte: Google imagens

# High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020

SORICHETT  
A, A. et al.  
2015

[www.nature.com/scientificdata](http://www.nature.com/scientificdata)

Fonte: Sorichetta et al, 2015

- A América latina é uma das regiões mais urbanizada do mundo. Espera-se que sua população atual, de 630 milhões, suba 25% até o ano de 2050;
- Monitoramento das mudanças, apoio ambiental e aplicações de saúde e planos de intervenção;
- Desagregação de dados oriundos de agências oficiais para 28 países da América Latina e Caribe;
- Para cada país foi realizado um conjunto de distribuição da população para os anos de 2010, 2015 e 2020 (previsão);

SCIENTIFIC DATA 

OPEN

SUBJECT CATEGORIES

» Geography

» Malaria

» Sustainability

» Environmental sciences

## High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020

Alessandro Sorichetta<sup>1,2</sup>, Graeme M. Hornby<sup>3</sup>, Forrest R. Stevens<sup>4</sup>, Andrea E. Gaughan<sup>4</sup>, Catherine Linard<sup>5,6</sup> & Andrew J. Tatem<sup>1,7,8</sup>

Received: 16 June 2015  
Accepted: 07 August 2015  
Published: 01 September 2015

The Latin America and the Caribbean region is one of the most urbanized regions in the world, with a total population of around 630 million that is expected to increase by 25% by 2050. In this context, detailed and contemporary datasets accurately describing the distribution of residential population in the region are required for measuring the impacts of population growth, monitoring changes, supporting environmental and health applications, and planning interventions. To support these needs, an open access archive of high-resolution gridded population datasets was created through disaggregation of the most recent official population count data available for 28 countries located in the region. These datasets are described here along with the approach and methods used to create and validate them. For each country, population distribution datasets, having a resolution of 3 arc seconds (approximately 100 m at the equator), were produced for the population count year, as well as for 2010, 2015, and 2020. All these products are available both through the WorldPop Project website and the WorldPop Dataverse Repository.

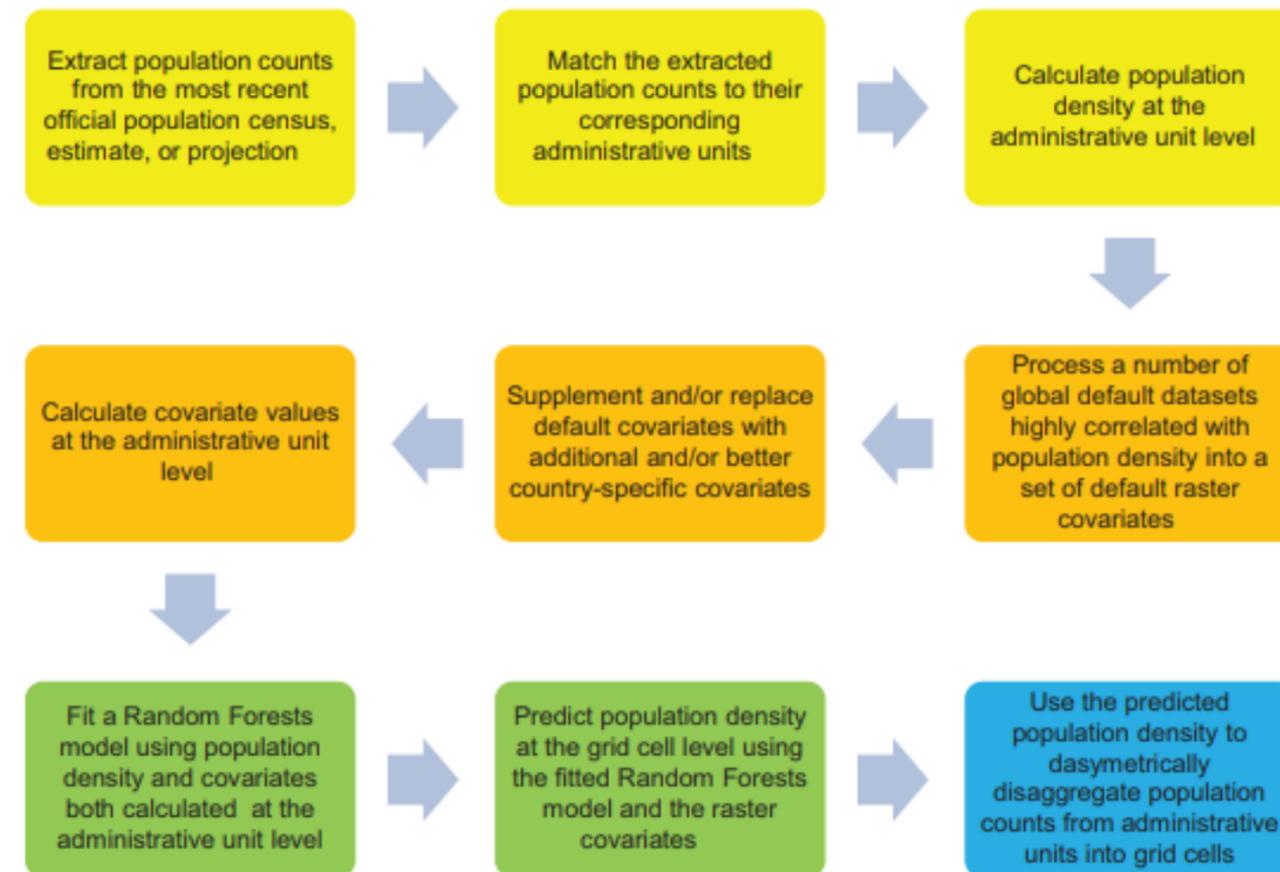
# Metodologia

- Abordagem de Stevens et al.

Random Forest:

- o número de covariáveis a serem randomicamente selecionado em cada nó;
- o número de observações nós nós de cada árvore;
- o número de árvore na floresta.

A variável de densidade populacional e adequação das covariáveis foram calculadas para unidade administrativa e então readequadas para o modelo RF para previsão da densidade populacional para o nível de grid celular com as covariáveis de rasters de resolução espacial de 100 m.



**Figure 1.** Schematic overview of the Random Forest (RF)-based dasymetric mapping approach used to produce the WorldPop Americas datasets (modified from Stevens *et al.*<sup>24</sup>). The preparation of the response variable and covariates is described in the yellow and orange panels, respectively, the RF modelling steps are outlined in the green panels, and the disaggregation of the input population counts from administrative units into grid cells is described in the blue panel.

# Metodologia

- O conjunto de dados foi projetado para 2010, 2015 e 2020 usando taxas da Programa das Nações Unidas para o Desenvolvimento (UNPD) rural e urbana;
- Todo processo sendo feito por meio do código WorldPOP-RF disponível gratuitamente;
- O código é escrito em ambiente R e pacote RF para o modelo de limite da unidade administrativa e predizer o nível de grid celular, Python e o pacote ArcGIS (ArcPy).

Fonte: GitHub - ForrestStevens/WorldPop-RF: This is the home for the RandomForest-based production code for WorldPop's population maps.

## WorldPop-RF

This is the home for the RandomForest-based production code for WorldPop's population maps.

### Prediction of Gridded Population Density

The source stored here represent the Python and R scripts backing the [WorldPop](#) project's current mapping efforts to produce gridded population density estimates of people per ~100 m pixel. These are produced using a Random Forest (RF) model as described in Stevens, et al. (2015). The published article contains a description of the RF model and its covariates, their sources and any metadata collected for each covariate. An estimated prediction weighting layer is used to dasymmetrically redistribute the census counts and project counts to match estimated populations based on UN estimates for the final population maps provided by [WorldPop](#).

Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *PLOS ONE*, 10(2), e0107042. doi:10.1371/journal.pone.0107042

# Metodologia

## Coleta dos dados

- Cada país teve sua contagem para o mais recente e detalhada contagem de população atual e coincidentes para a unidade administrativa correspondente em um ambiente GIS;
- Dados correlacionados com a distribuição populacional (luzes noturnas, usinas de produção de energia, topografia e inclinação, fatores climáticos, uso e cobertura, presença e ausência de vias, leitos de Rio e corpos hídricos, assentamentos urbanos e áreas protegidas);
- Doze rasters globais e dados vetoriais foram identificadas, coletadas, ajustados e processados em um conjunto de covariáveis usados para o modelo e predição.

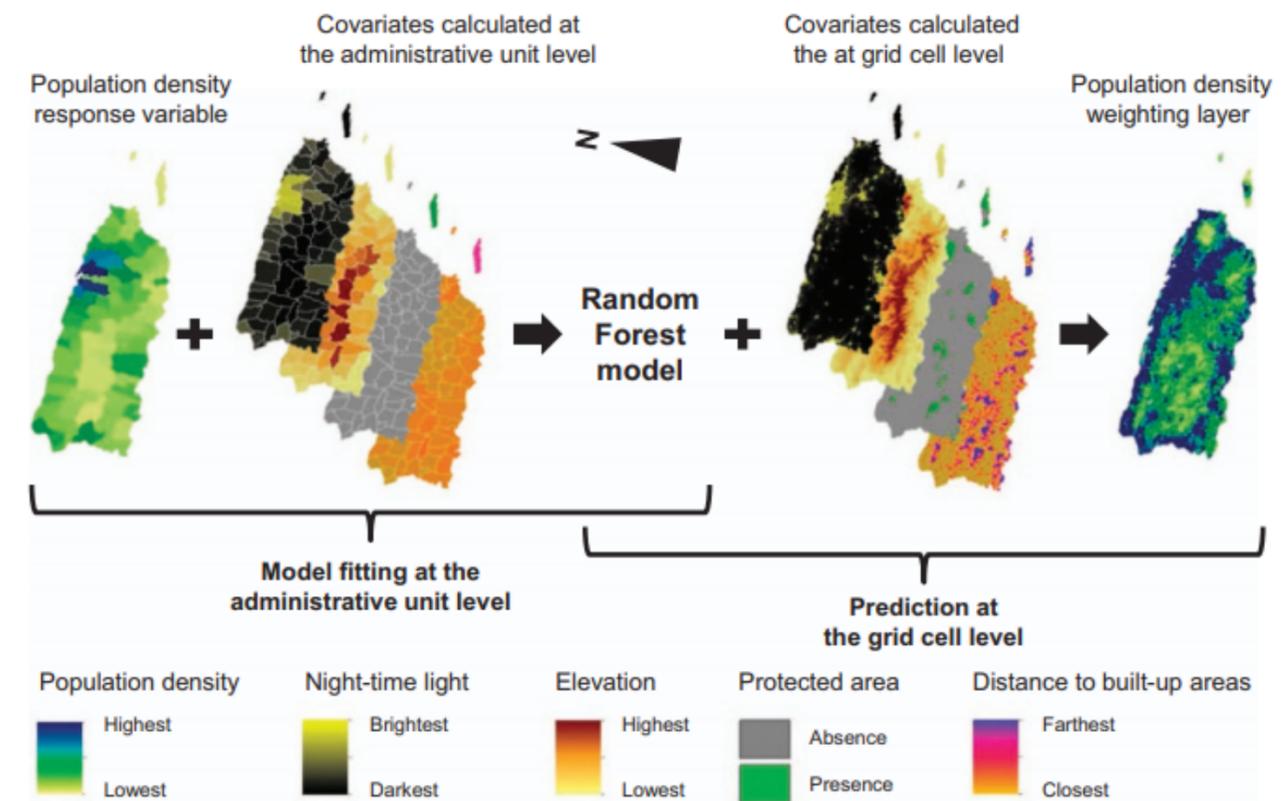
## Preparação dos dados

- Gerar um conjunto de dados representando a variável densidade populacional;
- Obter um conjunto de raster representando a área de estudo para registrar todas as covariáveis raster;
- Produzir um número de covariáveis raster de distância que não foram afetadas pelo efeito de borda devido ao fato que a área de estudo é artificialmente limitada enquanto os processos espaciais não;
- Foram selecionados 6 países para avaliar maior precisão da abordagem de mapeamento dasimétrico;

# Metodologia

## • Método

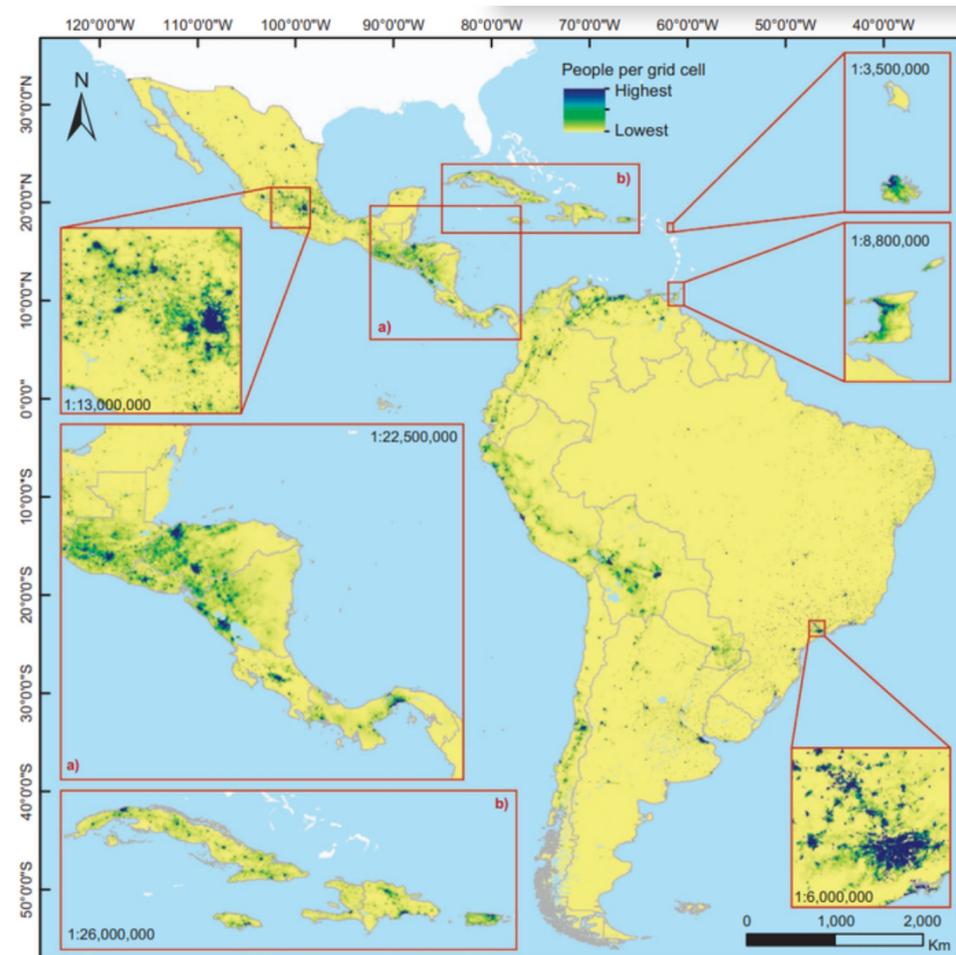
- 1- Estabelecimento de variáveis e covariáveis;
- 2- Ajustar o modelo RF para prever a densidade populacional a nível da célula da grade;
- 3 - desagregação da densidade da população da unidade administrativa e produzir 2 dados de grid populacional (nº estimado de pessoas e por hec para a população do ano);
- 4 - Preparo dos dados (projeção e buffer);
- 5- Validação dos erros para 6 países.



**Figure 2.** Schematic overview of the procedure used to generate population density weighting layers. For illustrative purpose, only 4 out of the 74 covariates considered for Puerto Rico are shown here (the uninhabited Puerto Rican islands of Mona, Monito, and Desecheo are not shown).

Fonte: Sorichetta et al, 2015

# Resultados



Fonte: Sorichetta et al, 2015

| Default dataset                        | Default derived covariate   | Temporal coverage | Type       | Format | Resolution     | Source                              |
|--|-----------------------------|-------------------|------------|--------|----------------|-------------------------------------|
| Suomi NPP-VIIRS                        |                             | 2012              | Continuous | Raster | 15 arc seconds | NOAA <sup>46</sup>                  |
|  | Night-lights' intensity     | 2012              | Continuous | Raster | 100 m          |                                     |
| MODIS Net Primary Production           |                             | 2014/2015         | Continuous | Raster | 30 arc seconds | NASA <sup>48</sup>                  |
|  | Plants' energy productivity | 2014/2015         | Continuous | Raster | 100 m          |                                     |
| WorldClim (BIO <sub>1</sub> )          |                             | 1950–2000         | Continuous | Raster | 30 arc seconds | Hijmans <i>et al.</i> <sup>50</sup> |
|  | Annual Mean Temperature     | 1950–2000         | Continuous | Raster | 100 m          |                                     |
| WorldClim (BIO <sub>12</sub> )         |                             | 1950–2000         | Continuous | Raster | 30 arc seconds | Hijmans <i>et al.</i> <sup>50</sup> |
|  | Annual Precipitation        | 1950–2000         | Continuous | Raster | 100 m          |                                     |
| HydroSheds (3 s GRID: Void-filled DEM) |                             | 2000              | Continuous | Raster | 3 arc seconds  | WWF <sup>52</sup>                   |

| ISO code | Area (sqkm) | Total population         | Year | No. of units | Unit name/level | Average spatial resolution | Population data source   | Units data source     |
|----------|-------------|--------------------------|------|--------------|-----------------|----------------------------|--|-----------------------|
| ATG      | 436         | 81,799 <sup>C</sup>      | 2011 | 8            | Parish/1        | 2.6                        | Census Office <sup>G</sup>                                       | GADM <sup>34</sup>    |
| ARG      | 2,804,771   | 40,117,096 <sup>C</sup>  | 2010 | 526          | Department/2    | 3.2                        | INDEC <sup>G</sup>   | IGN <sup>67</sup>     |
| BLZ      | 21,918      | 312,971 <sup>C</sup>     | 2010 | 16           | Subdivision/1   | 9.3                        | Statistical Institute of Belize <sup>G</sup>                     | Meerman <sup>68</sup> |
| BOL      | 1,069,327   | 10,027,262 <sup>C</sup>  | 2012 | 112          | Province/2      | 9.2                        | INEC <sup>G</sup>  | GADM <sup>34</sup>    |
| BRA      | 8,233,131   | 190,732,694 <sup>C</sup> | 2010 | 5565         | Municipality/3  | 0.6                        | IBGE   | IBGE                  |
| CHL      | 756,096     | 16,341,929 <sup>C</sup>  | 2012 | 297          | Municipality/3  | 2.9                        | INE-CELADE <sup>G</sup>  | GADM <sup>34</sup>    |
| COL      | 1,141,261   | 47,661,787 <sup>E</sup>  | 2013 | 1115         | Municipality/2  | 1.0                        | Departamento Administrativo Nacional de Estadística <sup>G</sup> | GADM <sup>34</sup>    |
| CRI      | 51,100      | 4,301,712 <sup>C</sup>   | 2011 | 469          | District/3      | 0.5                        | INEC <sup>G</sup>  | GADM <sup>34</sup>    |
| CUB      | 109,884     | 11,167,325 <sup>C</sup>  | 2012 | 168          | Municipality/2  | 2.0                        | ONE <sup>G</sup>   | GADM <sup>34</sup>    |
| DOM      | 48,070      | 9,445,281 <sup>C</sup>   | 2010 | 155          | Municipality/3  | 1.4                        | ONE <sup>G</sup>   | GADM <sup>34</sup>    |
| ECU      | 257,320     | 14,483,499 <sup>C</sup>  | 2010 | 978          | Parish/4        | 0.5                        | INEC <sup>G</sup>  | GADM <sup>34</sup>    |

# Conclusão

Fonte: Sorichetta et al, 2015

|     |     |   |      |       |    |           |        |          |
|-----|-----|---|------|-------|----|-----------|--------|----------|
| COL | RF  | 2 | 1115 | 0.35  | 84 | —         | —      | —        |
| COL | RF  | 1 | 33   | 1.20  | 75 | 109798.10 | 259.81 | 29361.29 |
| COL | SAW | 1 | 33   | —     | —  | 128372.29 | 303.76 | 36463.22 |
| CRI | RF  | 3 | 469  | 0.40  | 92 | —         | —      | —        |
| CRI | RF  | 2 | 81   | 0.20  | 93 | 4837.37   | 52.96  | 3012.04  |
| CRI | SAW | 2 | 81   | —     | —  | 14463.43  | 158.34 | 7976.94  |
| CUB | RF  | 2 | 168  | 0.33  | 82 | —         | —      | —        |
| DOM | RF  | 2 | 155  | 0.22  | 86 | —         | —      | —        |
| DOM | RF  | 1 | 32   | 0.53  | 62 | 46349.33  | 76.06  | 19461.99 |
| DOM | SAW | 1 | 32   | —     | —  | 101563.30 | 166.67 | 39729.54 |
| ECU | RF  | 4 | 978  | 0.47  | 82 | —         | —      | —        |
| ECU | RF  | 3 | 198  | 0.43  | 77 | 36713.59  | 248.75 | 7243.05  |
| ECU | SAW | 3 | 198  | —     | —  | 60295.60  | 408.52 | 12322.64 |
| SLV | RF  | 2 | 267  | 0.20  | 81 | —         | —      | —        |
| GUF | RF  | 3 | 21   | 2.60  | 59 | —         | —      | —        |
| GTM | RF  | 2 | 335  | 0.24  | 80 | —         | —      | —        |
| GTM | RF  | 1 | 22   | 0.33  | 58 | 51704.90  | 114.13 | 20590.36 |
| GTM | SAW | 1 | 22   | —     | —  | 62299.18  | 137.52 | 26125.56 |
| GUY | RF  | 2 | 116  | 1.10  | 87 | —         | —      | —        |
| HTI | RF  | 4 | 570  | 0.14  | 84 | —         | —      | —        |
| HTI | RF  | 3 | 140  | 0.071 | 90 | 10794.96  | 62.01  | 5493.25  |
| HTI | SAW | 3 | 140  | —     | —  | 18677.50  | 107.29 | 8501.97  |

- Mean Absolute Error (MAE) possui resultados mais baixos do que os valores correspondentes calculados pelo Sample Areal-Weighting (SAW).

# Colocações

## POSITIVOS

- Apresenta boa introdução ao tema;
- Demonstra bons pontos a serem pensados;
- Propõe boas perguntas.

## NEGATIVOS

- Melhor explicitação do funcionamento do método híbrido.

# Referências

LEYK, S. et al. The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, v. 11, n. 3, p. 1385-1409, 2019.

SORICHETTA, A. et al. High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015, and 2020. *Scientific data*, v. 2, n. 1, p. 1-12, 2015.

NIEVES, Jeremiah J. et al. Examining the correlates and drivers of human population distributions across low-and middle-income countries. *Journal of the Royal Society interface*, v. 14, n. 137, p. 20170401, 2017.

# Obrigado !

