Per-pixel population estimates in Western Amazon using limited remote sensing and spatial data

Luiz Felipe de Almeida Furtado[1], Luiz Carlos Teixeira Coelho[1,2,3], Maria de Fátima Rodrigues Pereira de Pina[1,4,5], Marília Sá Carvalho[6], Irving da Silva Badolato[1,7]

[1] Universidade do Estado do Rio de Janeiro - Faculdade de Engenharia, Rio de Janeiro, Brazil (luiz.furtado, luiz.coelho, fatima.pina, irving.badolato)@eng.uerj.br
[2] Instituto Municipal de Urbanismo Pereira Passos - Coordenadoria de Informações da Cidade , Rio de Janeiro, Brazil lcteixeiracoelho@prefeitura.rio
[3] Universidade Federal do Rio de Janeiro - Programa de Pós-Graduação em Engenharia Urbana, Rio de Janeiro, Brazil
[4] Fundação Oswaldo Cruz - Centro de Informação Científica e Tecnológica, Rio de Janeiro, Brazil fatima.pina@fiocruz.br
[5] Universidade do Porto - Instituto de Investigação e Inovação em Saúde i3S, Oporto, Portugal - fpina@i3s.up.pt
[6] Fundação Oswaldo Cruz - Programa de Computação Científica, Rio de Janeiro, Brazil marilia.carvalho@fiocruz.br
[7] Universidade do Estado do Rio de Janeiro - Instituto de Matemática e Estatística, Rio de Janeiro, Brazil irving.badolato@pos.ime.uerj.br

Keywords: Random Forest, Machine Learning, Demography, Remote Sensing, Population Estimate.

Abstract

There is a lack of detailed demographic data in the northern Brazil region from the 1980s to the early 2000s. These data are available only at the municipal level, which in northern Brazil corresponds to extensive territorial areas. This data gap may hinder understanding of various human settlement processes in the region, affecting insights into processes such as the expansion of economic activities, deforestation, and even violent conflicts. Machine learning algorithms, such as Random Forest, combined with geospatial data from different sources, can be employed to disaggregate demographic data, transforming the discrete space of municipal polygons into a continuous raster surface. Thus, this study aims to assess the performance of these technologies under limited data availability in scenarios similar to those in the late decades of the twentieth century. To this end, a Random Forest model was implemented and evaluated against both the 2022 Brazilian census data and the WorldPop dataset. The results indicate that the methodology proposed here is a viable solution in data-scarce contexts, yielding estimates comparable to official census figures and to more complex products like WorldPop, while demanding significantly less computational effort. Future research should examine the model's performance across broader and more heterogeneous regions to better assess its generalizability.

1. Introduction

The Northern region of Brazil is characterized by vast municipal territories, low population density, and small urban clusters. This combination poses challenges for academic studies on population dynamics, as well as for the implementation of public policies and urban and socioeconomic planning in the region. Traditional population data often distribute inhabitants homogeneously across municipal areas, obscuring real population distribution and failing to reflect local demographic dynamics.

During the past two decades, significant progress has been made in producing spatially disaggregated population data worldwide. In Brazil, the Brazilian Institute of Geography and Statistics (IBGE), through initiatives such as the National Register of Addresses for Statistical Purposes (CNEFE), enables population disaggregation beyond the municipal and census sector scales. First introduced in the 2000s and refined in the 2010 and 2022 censuses, CNEFE now provides georeferenced data for nearly all urban and rural households in Brazil. This official data set is a milestone for validating household-scale population disaggregation models. However, such data are only available for periods after 2010. For years, the distribution of the population remains homogenized at the municipal level. Dasymetric maps, whilst effective under some circumstances (Eicher and and, 2001), it may not fully apply to the peculiarities of the Brazilian Amazon, where municipalities have territories as large as some countries and the urban clusters are small and/or disperse.

Globally, programs like WorldPop disaggregate population data using spatial datasets (e.g. remote sensing imagery) and machine learning algorithms such as Random Forest (Stevens et al., 2015). This initiative, supported by an extensive research infrastructure, has produced consistent demographic data since the 2000s, particularly for developing countries with limited census resources, and is widely used by professionals in urban development, public health, and environmental sectors. Nevertheless, its estimates are limited to year 2000 and beyond, and employ a grand total of more than 30 covariates, with trees allowed to grow maximally (Stevens et al., 2015), which may severely increase processing time.

Several studies have employed similar methodologies to generate per-pixel population count and density estimates. For instance, (Qiu et al., 2020) utilized a Random Forest model with a relatively concise set of ancillary data, though their analysis was restricted to Zhengzhou and incorporated social variables such as points-of-interest and building footprints. Meanwhile, (Jiang

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-157-2026

et al., 2023) applied a dual-attention neural network to estimate population from remote sensing imagery, but this deep learning approach substantially increases computational demands. Additionally, their reliance on nighttime lights data, which suffers from temporal gaps and coarse spatial resolution.

Other approaches include the work of (Sanchez-Cespedes et al., 2024), who combined raster and vector data (including social cartography variables) within a Bayesian framework to estimate populations in hard-to-reach regions of Colombia. Similarly, (Campos et al., 2020) employed regression models based on Landsat ETM+ imagery and Brazilian census tracts to derive post-census population distributions in Contagem, Minas Gerais. While effective, these studies focus on post-2000 data and integrate multiple data layers, increasing computational complexity. Moreover, they primarily aim to project future population distributions rather than reconstruct historical estimates. Notably, (Qiu et al., 2020)—like WorldPop—relies on Random Forest but incorporates over thirty variables, including social cartography data, further complicating the model. Finally, regression models calibrated for a single urban area may not generalize well to other regions, as highlighted by (Yagoub et al., 2024).

Therefore, despite such advances, three caveats still persist: there is a data gap for periods prior to the 2000s, algorithms take too much processing resources and time - with several layers and variables, and models are not customized to the idiosyncrasies of settlements in the Amazon and to the lack of georeferenced databases for the region. These end up hindering the reconstruction and analysis of critical population dynamics in northern Brazil. For example, intense migration to the region during the 1980s and 1990s drove agricultural frontier expansion, deforestation, land conflicts, and increased violence. Disaggregated population data for this period are essential to understanding these dynamics, and their integration with recent data could extend the temporal scope of demographic studies, allowing deeper insights on the subjects.

This study proposes a lightweight Random Forest model to disaggregate population data for Rondônia, northern Brazil. Trained on 2022 data and validated against CNEFE records, the model addresses the scarcity of georeferenced historical data (1980s–1990s vs. 2010s–2020s) by relying on minimal, replicable ancillary data layers spanning 40+ years. We hypothesize that high accuracy for 2022 would permit extrapolation to earlier decades. Beyond producing reliable demographic estimates for understudied periods, the study aims to provide a simple, accessible framework adaptable to diverse contexts—even with limited computational resources.

2. Materials and Methods

2.1 Population and census data

We used the IBGE total population for both municipal and census sectors boundaries as reference data. Total municipal population data were employed for training the Random Forest model, while census sector population was used exclusively for validation and assessment of the results.

2.2 Covariates

Covariates used in our Random Forest model were selected based on WorldPop's methodology (Stevens et al., 2015), which iteratively refines their Random Forest models using fewer covariates in successive iterations, retaining only those with demonstrated importance. Additionally, covariate selection prioritized availability of the data for the 1980s–1990s period. Variables such as roads or municipal/district headquarters, for example, were excluded due to data unavailability or the need for manual editing (e.g., vectorization of analog maps).

The covariates we used include: (a) Proximity images derived from MapBiomas land use and cover classes (LUCC), (b) Residential and non-residential built-up areas from the Global Human Settlement Layer (GHSL), (c) Spectral indices derived from Landsat-8/OLI imagery (NDVI: Normalized Difference Vegetation Index; NDBI: Normalized Difference Built-up Index), and (d) elevation and slope data from the Shuttle Radar Topography Mission (SRTM). Except for the proximity images, all covariates were sourced via Google Earth Engine (GEE) and processed using python and Jupyter Notebook in Google Colab.

MapBiomas was selected for its full Google Earth Engine (GEE) integration and automated annual LULC mapping since 1985, offering efficient large-scale analysis despite being more conservative than PRODES (Neves et al., 2020, Maurano and Escada, 2019). The GHSL, developed by the European Commission, was chosen for its similar timespan (using Landsat and Sentinel data from 1975 on) and proposal to identify settlments with more than 50,000 inhabitants (Melchiorri, 2022). NDVI and NDBI indices were also chosen for their level of relevance to previous studies, and direct correlation to natural areas and water bodies - where population is mostly absent. Finally, the SRTM elevation dataset was adopted for its balance between accuracy and GEE compatibility, despite not being the highest-resolution option available (Uuemaa et al., 2020).

All covariates were resampled to a pixel size of 100 x 100m. This ensures that the generated model is compatible with the Worldpop data and at the same time allows for faster and more efficient execution in the Google Colab and Google Earth Engine environments. All covariates used in this study are from the year 2022, except the GHSL layer (year 2020) and SRTM, from the early 2000's.

2.2.1 Distance Images Distance images represent pixel distances (in kilometers) to reference features, with pixel values thus increasing with distance from these features. These were generated using Python's distancerasters library, which employs the Haversine formula in order to calculate distance. Based on WorldPop's covariate importance and Amazonian regional characteristics, seven LUCC classes were selected: (1) Water, (2) Forest, (3) Savanna, (4) Grassland, (5) Urban, (6) Anthropic, and (7) Secondary Vegetation. Classes 1–5 were sourced from MapBiomas' Land Use/Cover Collection (v9), while classes 6–7 came from its Deforestation/Secondary Vegetation Collection. MapBiomas provides annual data from 1985 to 2024, which is critical for temporal analysis.

2.2.2 Global Human Settlement Layer The European Space Agency's (ESA) Global Human Settlement Layer (GHSL) project offers a group of several products that delivers detailed human settlement and population mappings. We used the 2020 Global Built-up Surface product (P2023A) for both Residential and Non-Residential covers. This data is available in 5-year intervals from 1975 onward, with projections to 2030.

2.2.3 Spectral Indices NDVI and NDBI (both well known and easy-to-use indices (Tamiminia et al., 2020)) for 2022 were

calculated using USGS Landsat-8 Level 2, Collection 2, Tier 1 surface reflectance imagery. Processing steps included (a) Selecting all available scenes for 2022, (b) converting pixel digital numbers to surface reflectance, (c) applying cloud/noise filters on each available scene, (d) calculating NDVI and NDBI for each scene, and (e) calculating median pixel values to generate synthetic NDVI/NDBI composite layers for the year 2022.

2.2.4 Shuttle Radar Topography Mission Elevation (meters) and slope (degrees) layers were derived from the Shuttle Radar Topography Mission (SRTM) version 4 digital elevation model.

2.3 Dasymetric Mapping

In order to train the Random Forest model, a dasymetric population mapping was created. Total population values were then allocated to Mapbiomas urban clusters boundaries. This process involved two stages:

Stage 1: For each municipality, the total area of urban clusters within its territory was calculated. Each cluster's proportional area was then squared and normalized to create a weight layer, emphasizing larger urban areas even if their sizes were similar and proportional. The municipality's total population was multiplied by these weights to allocate population proportions to individual urban clusters.

Stage 2: The value of population assigned to each cluster was distributed homogeneously across its pixels. The geometric centroids of the clusters were calculated, and the inverse normalized distance from each centroid was derived (maximizing values for centroid-proximal pixels and minimizing them for distant ones). The product of the homogeneous population layer and the distance layer resulted in a radial population distribution around cluster centroids, where most of the population lays in the central areas and diminishes in periferical areas.

2.4 Random Forest

The Rondônia Random Forest Model (RRFM) was trained with 500 trees (following [2]), default mtry (square root of total covariates), and a maximum tree depth of 10. A stratified random sample of 100,000 points was used: half of the total points for each urban class areas (MapBiomas) and non-urban regions. Of these, 80 percent of the total points were allocated to training.

The Random Forest algorithm was selected due to its seamless integration with Google Earth Engine, flexibility, and proven reliability as a machine learning (ML) regressor for remote sensing applications (Belgiu and Drăguţ, 2016, Phan et al., 2020, Magidi et al., 2021). Its widespread use and reproducibility across diverse scenarios make it a practical choice. Compared to other ML algorithms, Random Forest often delivers superior performance (Ouchra et al., 2023), balancing high accuracy with effective modeling of non-linear relationships while mitigating overfitting risks (Srivastava and Kumar, 2013). Additionally, tree depth optimization was implemented to enhance computational efficiency and accommodate the constraints of free platforms like Google Colab (Yang et al., 2022), providing faster training and prediction and even less overfitting. Also, it was expected to yield reasonably good results (Duroux and Scornet, 2018).

2.5 Post-Processing

A correction factor was applied to mitigate population overestimation that occur in such models [2]. For each municipality, the factor was calculated as the ratio of reference population to the estimated population. This ratio was then uniformly applied to all pixels within the municipality.

2.6 Validation

Validation occurred in two stages: first, the total population estimates from RRFM and both WorldPop Brazil Constrained and Unconstrained (WPBrC and WPBrUn, respectively) were compared against IBGE population census data, at census sectors level. The Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 and the mean Error Bias (the mean between the difference between predicted and reference values) were computed. In addition, spatial error patterns were assessed.

3. Results

3.1 Covariates Importance

The mean decrease in node impurity (MDI) highlights (Figure 1) the GHSL Residential areas (ghsl_res) as the dominant predictor (MDI = 0.687), followed by Distance to Savanas (dst_savanas) with MDI = 0.080, Elevation (0.076), Distance to Grasslands (dst_grasslands) (0.055), and Distance to Water Bodies (dst_water) (0.044). Remaining variables exhibited negligible influence (MDI ¡ 0.02): Distance to Secondary Forests (dst_secondary) GHSL Non-Residential areas (ghsl_nres), Slope, NDVI and NDBI showing minimal to near-zero importance (MDI equal or less to 0.004). The observed MDI values reflect important characteristics of the urban occupation in the Amazon Basin: the occupied areas are mainly regions with low elevation, close to rivers and floodplain areas. These wetlands, such as várzeas and igapós, have as their predominant vegetation forests with smaller trees and areas occupied by grasses.

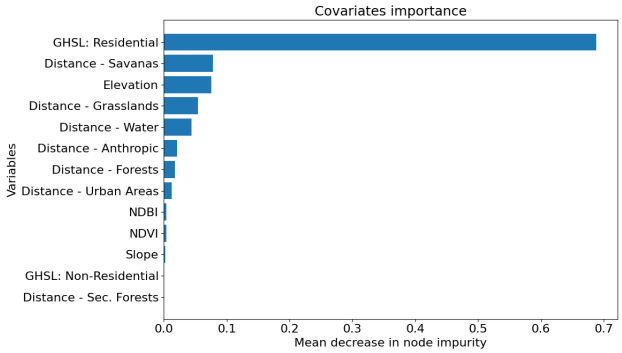

Figure 1. Mean decrease in node impurity (MDI) observed for the Rondônia Random Forest Model (RRFM).

3.2 Model Comparisons

The Rondônia Random Forest Model (RRFM), along with the WorldPop Brazil Constrained (WPBrC) and Unconstrained (WPBrUn) models, were compared against 2022 IBGE census data. This comparison is partially limited due to methodological discrepancies between the models: (a) WPBrC and WPBrUn use projected population data for 2020 based on the 2010

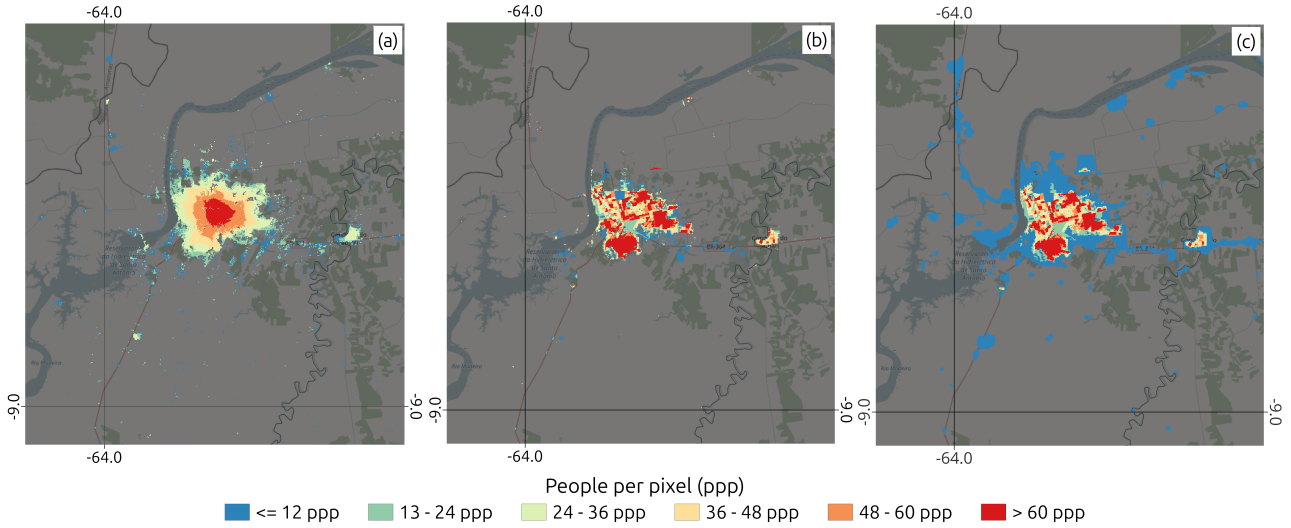

Figure 2. Visual assessment of Porto Velho estimated people per pixel for each one of the models: (a) Rondônia Random Forest Model - RRFM, (b) Worldpop Brazil Constrained - WPBrC and (c) Worldpop Brazil Unconstrained - WPBrUn.

Census, rather than the 2022 Census itself, and (b) these models also employ several covariates from distinct years (e.g., 2015 land use/cover classes, multi-year road networks). The models were compared using the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), in addition to the coefficient of determination (R^2) and error bias.

3.2.1 MAE and RMSE The RRFM showed a slightly lower Mean Absolute Error (MAE) than WPBrC but higher than WPBrUn: 259.2, 275.8, and 252.2 persons per pixel (ppp), respectively. For Root Mean Squared Error (RMSE), RRFM yielded higher values than both WorldPop models: 477.8 ppp versus 407.4 (WPBrC) and 387.8 (WPBrUn). This suggests that while RRFM achieves comparable average accuracy to WorldPop models, it exhibits greater error dispersion, likely due to localized extreme variations.

The tendency for RRFM showing greater error dispersion can be related to the fact that the model employs, as the Random Forest training layer, an population dasymetric layer where the values are . Thus, RRFM disagregates population radially 2, and when the data is compared with the IBGE sectors, the model's error increases.

This dispersion, concentrated in high-density urban areas, may stem from two factors. The first one is linked to dasymetric mapping limitations: while RRFM's independent variable (urban cluster-based population allocation) distributes population radially based on the centroid of each urban cluster 2, WorldPop models integrate IBGE census sectors as inputs, providing higher demographic granularity and heterogeneity. The latter may be connected to the reduced number of covariates: The RRFM's limited number of covariates (compared to WorldPop) may struggle to explain population variability in complex Amazonian landscapes.

3.2.2 Coefficient of Determination All models exhibited low R^2 values: -0.48 (RRFM), -0.08 (WPBrC), and 0.02 (WPBrUn). These results indicate that none of the models sufficiently explain population variability, particularly the RRFM. Even WorldPop models, which utilize significantly more covariates, achieved R^2 values near zero, underscoring the inherent challenges of modeling heterogeneous populated regions like the Amazon region.

3.2.3 Error Bias All models displayed an overall tendency to overestimate population per pixel. The WPBrUn showed the highest bias (+79.9), followed by RRFM (+43.8) and WPBrC (+39.2). Alongside spatial error distribution analysis, it is possivel to assess that both RRFM and WPBrC have high error variability, with both over- and underestimations across census sectors, and WPBrUn shows predominantly positive error distributions, reflecting systemic overestimation.

The RRFM's lower bias magnitude compared to WPBrUn suggests that temporally synchronized covariates (aligned with 2022 census data) improve mean estimates despite persistent challenges in outlier regions.

3.3 Spatial Error Analysis Highlights

High-magnitude errors occur mainly in clustered/densely populated areas (i.e., areas tightly stacked with census sectors within urban zones). Visual analysis confirmed that (1) RRFM have greater error dispersion, evidenced by the increased number of census sectors with saturated colours, associated to higher error values; (2) WPBrUn showed widespread overestimation, with the highest frequency of positive errors across census sectors, and (3) both RRFM and WPBrC share a tendency of mixed over- and underestimations, reflecting the challenges of capturing the Amazon's demographic complexity. Table 1 and Figure 3 show the aforementioned results.

Model	MAE	RMSE	R^2	Bias
RRFM	259.2	477.8	-0.48	43.8
WPBrC	275.8	407.4	-0.08	39.2
WPBrUn	252.2	387.8	0.02	79.9

Table 1. Error metrics for Rondônia Random Forest Model (RRFM), Worldpop Brazil Constrained (WPBrC) and Worldpop Brazil Unconstrained (WPBrUn), when compared to IBGE Census Data for 2022.

4. Discussion

The analysed models showed minor differences in RMSE, MAE, and error trends overall. Thus, the mean population values disaggregated by each model are similar. Therefore,

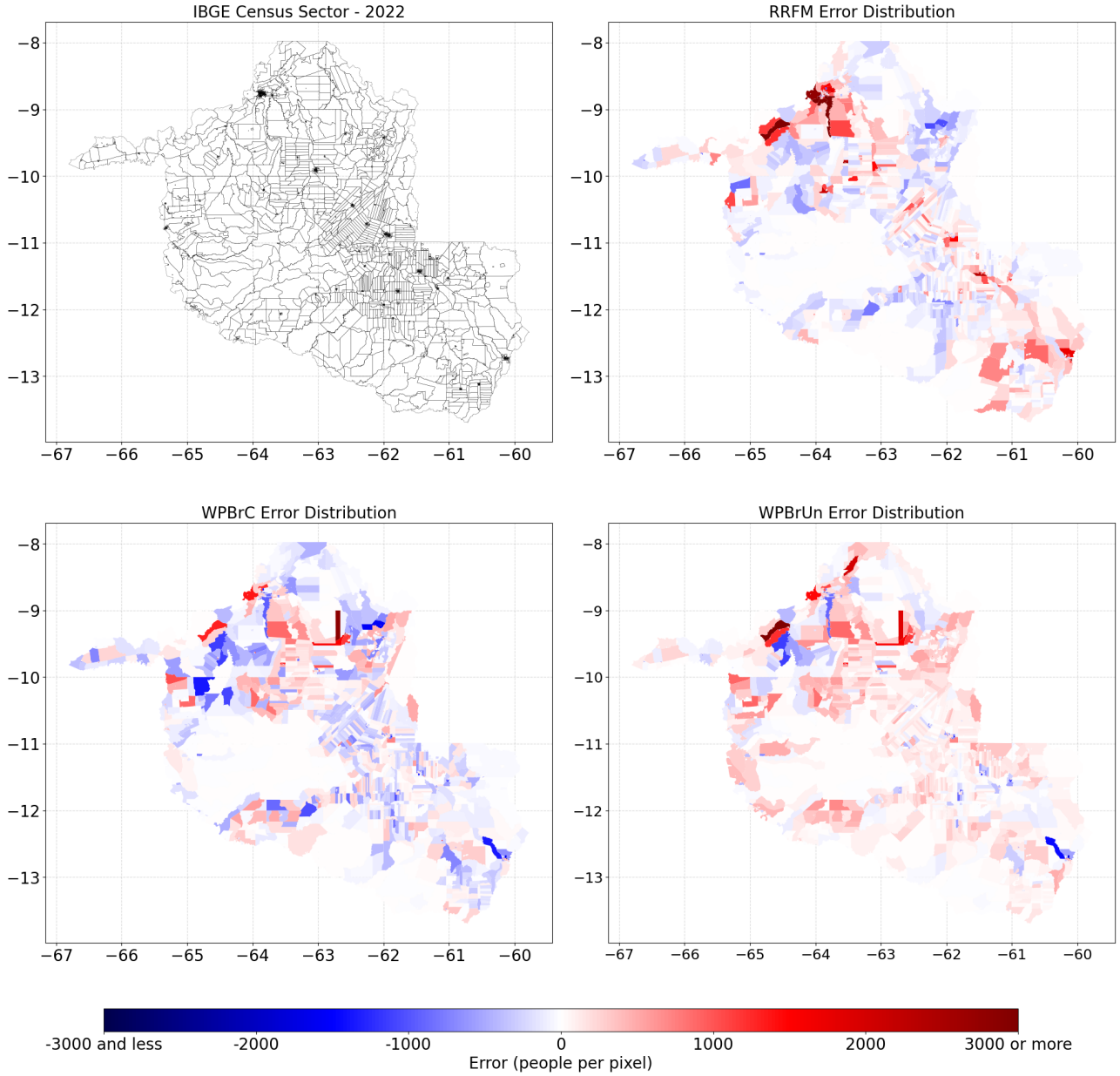

Figure 3. Spatial distribution of observed errors for the models Rondônia Random Forest Model - RRFM, Worldpop Brazil Constrained - WPBrC and Worldpop Brazil Unconstrained - WPBrUn.

RRFM, WPBrC and WPBrUN are equally capable of disaggregating the population with similar values in the expected urban places, with no major differences in the amount of population estimated by them. The observed differences between the models stem primarily from RRFM's current hindered ability to disaggregate the population at finer scales. This is mainly due to its Random Forest training based on a radially distributed population layer rather than IBGE census sectors as input data, as WPBrC and WPBrUn.

Despite this weakness, one of the key strengths of the RRFM is its broad applicability. Both WPBrC and WPBrUn rely on a lot of variables, including local variables such as roads, important landmarks etc. Thus, the models are restricted to specific cities or regions in such those data are available. On the other hand, RRFM doesn't rely on specific social variables, nor is it tailored to a particular city or spatial context. In addition, it relies in only a few open-access spatial datasets, such as Mapbiomas, GHSL, Landsat imagery and SRTM. Thus, as it is based on a few generalized and widely available datasets, it can potentially be applied across an entire Amazonian state or similarly large regions.

From a technical standpoint, the model is based on the Random Forest algorithm, using a max-depth of 10. Despite its relatively low complexity — only eight dependent variables — it shows solid performance when compared to other established models such as Worldpop database, despite the current hindered ability to disaggregate the population at finer scales. This makes it both efficient and accessible, as the processing can be done in a short span of time, using basic cloud-based tools like Google Colab and Google Earth Engine python API integration. Thus this methodology is available for both large and small areas, densely populated or more secluded, requiring less code-intensive user experience as both Python and Google Earth Engine python API are full documented and benefit from a vast online com-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-157-2026 | © Author(s) 2026. CC BY 4.0 License.

munity of users and supporters.

Another important feature is its potential for retrospective analysis. The model uses covariates that have been commonly available since at least the early 1980s. While it's not possible to directly test its accuracy for past periods due to the way official data is aggregated (typically by municipality), it's reasonable to assume that the model would still perform adequately, as for 2022, for historical estimates when compatible covariates are used.

5. Conclusions

This study presents a robust, versatile and adaptable model that estimates per-pixel total population and demonstrates strong potential for broad geographic applications, by not relying on specific local variables and by using generic and widely available datasets.

The model achieves good performance while maintaining low complexity, employing less dependent variables than similar studies. Processing can be completed in a less intensive manner, by using free and accessible cloud-based platforms and covariates that are readily available nation or worldwide.

Additionally, the model shows promise for retrospective applications. Although direct validation for past decades is constrained by the availability of disaggregated data, the choice of covariates—commonly used since at least the 1980s—supports the assumption that the model can reasonably estimate historical patterns as well.

The weaknesses of the RRFM model, mainly the greater saturation of errors in more densely populated areas, can be addressed in future studies with the use of a more detailed dasymetric population mappings as independent variables, or even using census sectors directly, as other models such as Worldpop does, depending on the year for which the model will be generated.

References

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Campos, J., Rigotti, J. I. R., Baptista, E. A., Monteiro, A. M. V., Reis, I. A., 2020. Population Estimates from Orbital Data of Medium Spatial Resolution: Applications for a Brazilian Municipality. *Sustainability*, 12(9). https://doi.org/10.3390/su12093565.

Duroux, R., Scornet, E., 2018. Impact of subsampling and tree depth on random forests. *ESAIM: Probability and Statistics*, 22, 96–128.

Eicher, C. L., and, C. A. B., 2001. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, 28(2), 125–138. https://doi.org/10.1559/152304001782173727.

Jiang, Y., Huang, Z., Li, L., Dong, Q., 2023. Local–global dual attention network (LGANet) for population estimation using remote sensing imagery. *Resources, Environment and Sustainability*, 14, 100136. https://doi.org/10.1016/j.resenv.2023.100136.

Magidi, J., Nhamo, L., Mpandeli, S., Mabhaudhi, T., 2021. Application of the Random Forest Classifier to Map Irrigated Areas Using Google Earth Engine. *Remote Sensing*, 13(5). https://doi.org/10.3390/rs13050876.

Maurano, L., Escada, M., 2019. Comparação dos dados produzidos pelo prodes versus dados do mapbiomas para o bioma amazônia. *Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto*, INPE, Brazil, 1–4.

Melchiorri, M., 2022. The global human settlement layer sets a new standard for global urban data reporting with the urban centre database. *Frontiers in Environmental Science*, Volume 10 - 2022. https://doi.org/10.3389/fenvs.2022.1003862.

Neves, A. K., Körting, T. S., Fonseca, L. M. G., Escada, M. I. S., 2020. Assessment of TerraClass and MapBiomas data on legend and map agreement for the Brazilian Amazon biome. *Acta Amazonica*, 50(2), 170–182.

Ouchra, H., Belangour, A., Erraissi, A., 2023. Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study. *IEEE Access*, 11, 71127-71142.

Phan, T. N., Kuch, V., Lehnert, L. W., 2020. Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sensing*, 12(15). https://doi.org/10.3390/rs12152411.

Qiu, G., Bao, Y., Yang, X., Wang, C., Ye, T., Stein, A., Jia, P., 2020. Local Population Mapping Using a Random Forest Model Based on Remote and Social Sensing Data: A Case Study in Zhengzhou, China. *Remote Sensing*, 12(10). https://doi.org/10.3390/rs12101618.

Sanchez-Cespedes, L. M., Leasure, D. R., Tejedor-Garavito, N., Cruz, G. H. A., Velez, G. A. G., Mendoza, A. E., Salazar, Y. A. M., Esch, T., Tatem, A. J., and, M. O. B., 2024. Social cartography and satellite-derived building coverage for post-census population estimates in difficult-to-access regions of Colombia. *Population Studies*, 78(1), 3–20. https://doi.org/10.1080/00324728.2023.2190151. PMID: 36977422.

Srivastava, G., Kumar, P., 2013. Water quality index with missing parameters. *International Journal of research in Engineering and Technology*, 2(4), 609–614.

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), 1-22. https://doi.org/10.1371/journal.pone.0107042.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164, 152-170. https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Uuemaa, E., Ahi, S., Montibeller, B., Muru, M., Kmoch, A., 2020. Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sensing*, 12(21). https://doi.org/10.3390/rs12213482.

Yagoub, M., Tesfaldet, Y. T., AlSumaiti, T., Al Hosani, N., Elmubarak, M. G., 2024. Estimating population density using open-access satellite images and geographic information system: Case of Al Ain city, UAE. *Remote Sensing Applications: Society and Environment*, 33, 101122. https://doi.org/10.1016/j.rsase.2023.101122.

Yang, L., Driscol, J., Sarigai, S., Wu, Q., Chen, H., Lippitt, C. D., 2022. Google Earth Engine and Artificial Intelligence (AI): A Comprehensive Review. *Remote Sensing*, 14(14). https://doi.org/10.3390/rs14143253.