Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Evaluation of Global Gridded Population Data in Represent Resident Population Distribution in the Brazilian Amazon: The case of the Baixo Tocantins Region

Avaliação das Grades Populacionais Globais para Representar a Distribuição da População Residente na Amazônia Brasileira: o Caso do Baixo Tocantins (PA)

Gustavo Piva Lopes Salgado [1], Ana Paula Dal'Asta [2], Bruno Vargas Adorno [3] e Silvana Amaral [4]

1 National Institute for Space Research, São José dos Campos, Brazil. gustavo.salgado@inpe.br.
ORCID: https://orcid.org/0009-0004-8167-3780
2 National Institute for Space Research, São José dos Campos, Brazil. ana.dalasta@inpe.br.
ORCID: https://orcid.org/0000-0002-1286-9067
3 National Institute for Space Research, São José dos Campos, Brazil. bruno.adorno@inpe.br.
ORCID: https://orcid.org/0000-0003-0302-7834
4 National Institute for Space Research, São José dos Campos, Brazil. silvana.amaral@inpe.br.
ORCID: https://orcid.org/0000-0003-4314-7291

Received: 10.2024 | Accepted: 05.2025

Abstract: This study evaluated four Global Gridded Population datasets (GPWv4, GHS-POP, HRSL, and WorldPop) to represent the distribution of the resident population in the Brazilian Amazon, focusing on their concordance with population data and adherence to a residence address density proxy. The study area was the Baixo Tocantins region, Pará, chosen for its diversity of land uses and coverages, frequency of forest remnants, and dispersed rural population across island and mainland environments. The Global Grids for 2020 were evaluated through: i) concordance analysis with resident population data, by municipality and census tract, derived from municipal estimates for 2020 and the 2022 Demographic Census, and ii) analysis of adherence to residences estimates, using data from the Brazilian National Address Register for Statistical Purposes (CNEFE), based on a residence density proxy. Mean absolute percentage errors evaluated the Grids' estimates and official data. The highest concordance errors in the population data were observed for the GPWv4 and WorldPop grids. WorldPop and GHS-POP underestimate the occurrence of residences in island and riverbank areas. The HRSL grid showed the highest concordance with Census population data and better adherence to residences estimates compared to CNEFE. Thus, in the absence of disaggregated official population data, the HRSL grid is recommended as the preferred option for studies involving population distribution in the Brazilian Amazon.
Keywords: Global Gridded Population. Demographic Modeling. Amazon Region. Adherence Analysis.

Resumo: Este trabalho avaliou quatro Grades Populacionais Globais (GPWv4, GHS-POP, HRSL e WorldPop) para representar a distribuição da população residente na Amazônia brasileira, com foco na concordância com dados populacionais e na aderência a um *proxy* de densidade de endereços de domicílios. A área de estudo é a região do Baixo Tocantins, Pará, por sua diversidade de usos e coberturas da terra, frequência de remanescentes florestais e população rural dispersa por ambientes de ilhas e terra firme. As Grades Globais para 2020 foram avaliadas através de: i) análise de concordância com os dados de população residente, por município e setor censitário, provenientes de estimativa municipal para 2020 e Censo Demográfico de 2022, e ii) análise de aderência às estimativas de domicílios, com referência aos dados do Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE), a partir de uma *proxy* de densidade de domicílios. Erros absolutos percentuais médios avaliaram as estimativas das Grades e os dados oficiais. Os maiores erros de concordância nos dados populacionais foram registrados para as grades GPWv4 e WordPop. WorldPop e GHS-POP subestimam a ocorrência de domicílios nas áreas de ilhas e margens de rios. A grade HRSL apresentou a maior concordância em relação aos dados populacionais e maior aderência às estimativas de domicílios em comparação ao CNEFE. Assim, na ausência de dados populacionais oficiais desagregados recomenda-se a grade HRSL como opção preferencial para estudos que envolvam a distribuição da população na Amazônia brasileira.
Palavras-chave: Grade Populacional Global. Modelagem Demográfica**.** Região Amazônica. Análise de Aderência.

1 INTRODUCTION

With roughly 6.5 million square kilometers of vast tropical rainforests, the Amazon encompasses more than half of the planet's biodiversity and about 20% of the world's freshwater. It also is key in regulating local and regional climate patterns by releasing, daily, around 20 billion tons of water vapor into the atmosphere (Nobre et al., 2009; Fearnside, 2006). At the same time, the biome is home to 26 million inhabitants, about 13.1% of Brazil's population (Instituto Brasileiro de Geografia e Estatística [IBGE], 2022a), distributed across a diversity of socio-spatial units (Cardoso et al., 2023) into 772 municipalities (IBGE, 2022c). With a population density of 5.3 people per km² (IBGE, 2022a), the region has part of its population concentrated mainly in small urban spaces, in addition to a variety of peoples living under the forest canopy, posing challenges for representing population distribution. A further challenge is the vast size of administrative units and territorial divisions used for publishing population data, which assume internal environmental homogeneity and are subject to periodic boundary changes.

In recent decades, various initiatives have developed methods for producing and disseminating estimates of (re)distributed spatial population data at the grid level, known as Population Grids (Dahmm & Rabiee, 2020). A Population Grid consists of standard-sized cells, which can be global, national, regional, or local in scope. Each grid cell can be associated with population-related information, such as: estimates of resident population and ambient population (people present during a given period, such as 24 hours); and population by sex, age, ethnicity, and other demographic characteristics. National demographic censuses constitute the main source of demographic data for the production of grid estimates, from which the quality of these representations also derives. Leyk et al. (2019) suggest that users assess the characteristics and quality of Gridded Population Data in terms of spatial resolution and its suitability for their intended purpose. Ancillary data, particularly derived from satellite imagery, e.g. spatial indicators of population density, can introduce uncertainties by making invisible regional particularities (Leyk et al., 2019; Gonçalves et al., 2021; Silva, 2023).

The growing availability of Gridded Population density estimates has expanded the use of these products for diverse applications, as seen in Renner et al. (2018), Zhou et al. (2020) and Zhao and Wang (2023), increasing the need to assess their accuracy and/or consistency. Studies have sought to evaluate grid quality by examining differences in the methods and the types of ancillary variables used in their construction, as well as the historical and geographic particularities that may complicate modeling (Bai et al. 2018; Archila Bustos et al. 2020; Calka & Bielecka 2020; Chen et al. 2020; Xu et al. 2021; Yin et al. 2021; Kuffer et al. 2022). Silva (2023) developed a Population Grid for the Vale do Paraíba and Litoral Norte Metropolitan Region (São Paulo, Brazil), using auxiliary variables adapted to the regional context, and compared it with global-scale grids, revealing limitations in the latter to represent local conditions. For two Amazonian areas, Amaral et al. (2012) proposed a method to redistribute census tract-level population data into grids, highlighting the challenges of modeling in this region.

Thus, this study evaluates the quality of major Global Grids in representing the distribution of the resident population in the Brazilian Amazon, guided by two research questions: (1) Are the estimates from Global Gridded Population data consistent with official population data for municipalities in the Brazilian Amazon? (2) Are these Global Population Grids viable alternatives to represent population density and distribution at the intra-municipal level in the Brazilian Amazon? To address these questions, focusing on the Baixo Tocantins region (Pará) as the study area, two specific objectives are: (i) Evaluate the agreement between Global Gridded Population estimates and reference data on resident population by municipality; (ii) Analyze the consistency of global grid data with a proxy metric for residence address density at the grid level. By assessing the leading Global Grids – their data sources and construction methodologies – this study advances understanding of their strengths and limitations in representing population distribution in the Amazon.

2 POPULATION GRIDS AND TERRITORIAL POPULATION DISTRIBUTION

The population distribution in regular grids enables the manipulation of population data within a spatially and temporally stable framework, independent of administrative boundaries or census units – which

are subject to change over time. The use of grids also facilitates integration with other multitemporal data sources, as it mitigates statistical issues associated with areal aggregation methods, such as the Modifiable Areal Unit Problem (MAUP)[1]. Additionally, it enhances the availability of intercensal population data for modeling applications that require disaggregated inputs (Archila Bustos et al., 2020).

Lloyd et al. (2017) explain that Population Grids are particularly valuable for public policy planning, allowing the monitoring of population distribution and dynamics, as well as supporting the design of tailored interventions across different territorial scales. The authors further identify specific grid applications, including: disease occurrence estimation and epidemic modeling; resource allocation optimization, natural disaster management, accessibility index modeling, urban planning, poverty mapping, and environmental impact assessments.

Population Grids can be generated through bottom-up (aggregation) or top-down (disaggregation) methods. Bottom-up approaches rely on georeferenced population counts or sample data to model distribution patterns. These are more common for local-scale grids as its smaller geographic extents make such detailed data feasibly available. Top-down methods, conversely, use administrative or census-aggregated data as inputs, disaggregating them into regular grid cells (Dahmm & Rabiee, 2020). These include: Non-auxiliary-data methods (e.g., areal weighting), which redistribute population based solely on the area overlap between grid cells and administrative units (Hallisey et al., 2017); dasymetric techniques (reallocation-based) or estimative/interpolative approaches (e.g., Random Forest modeling). Following Nagle et al. (2013), both require census input data for reallocation/interpolation and auxiliary data to derive weighting parameters. Prominent global top-down Population Grids include Gridded Population of the World (GPW), Global Human Settlement Layer – Population (GHS-Pop), High Resolution Settlement Layer (HRSL) and WorldPop.

The GPW was the first Global Gridded Population data ever produced, with its initial version released in 1995. Since then, it has utilized national administrative boundary maps and water/ice body masks as its primary auxiliary datasets. The latest version (GPWv4) incorporated national census data collected between 2005 and 2014 to generate resident population density estimates for the years 2000, 2005, 2010, 2015 and 2020. These estimates are aligned with the United Nations (UN) World Population prospects, which, in turn, derive from national censuses and population estimates. The population estimates are allocated proportionally to raster cells using a uniform areal weighting method that accounts for intersections with subnational or national boundaries. Specifically, population is disaggregated based on the areal proportion of each administrative unit falling within a given grid cell (Center for International Earth Science Information Network [CIESIN], 2018).

The GHS-Pop provides global estimates of resident population from 1975 to 2020 at five-year intervals, along with projections for 2025 and 2030. The GPWv4 is its input, disaggregated via a binary dasymetric method into refined grid cells, based on built-up area density and classification from the corresponding Global Human Settlement Layer Built (GHSL-Built) dataset for each target year. Cells marked as "NoData" in GHSL-Built are references to zero population (uninhabited). GPWv4 cells classified as "uninhabited" but containing built-up areas in 2020 were reassessed using high spatial resolution imagery (Google Maps® and Bing®), with discrepancies between GPWv4 data and observed residential evidence corrected during validation (Schiavina et al., 2023).

The HRSL provides population estimates for 2015 and 2020 using a volumetric dasymetric disaggregation method that redistributes population based on building volumetric density. Developed through a four-stage process, it first employed GPW data modeled by Center for International Earth Science Information Network (CIESIN) – Columbia University, to project population growth at national and subnational levels. Next, Facebook Connectivity Lab utilized computer vision algorithms (edge detection and two independent CNNs – SegNet and FeedbackNet) to analyze 50 cm-resolution DigitalGlobe® imagery, detecting built-up areas, including those individual buildings in rural areas. Then, building density within 30x30m grid cells was calculated by quantifying structure size and distribution, before finally redistributing population using a building-volume dasymetric approach to generate the final density map (Tiecke et al., 2017).

[1] *Different forms of data aggregation can lead to varying results, influencing a study's conclusions and interpretations. This occurs because the choice of aggregation units can affect data variability and the spatial distribution of observations (OPENSHAW, 1984).*

The WorldPop grid employs a comprehensive set of auxiliary variables related to human settlement distribution in the landscape to establish statistical relationships between these variables and official census or population estimates (WorldPop, n.d.). These variables include settlement locations, land use/cover data, road and building maps, healthcare facility sites, nighttime light data, vegetation mapping, topography, and other spatial indicators, which are integrated into a semi-automated Random Forest regression model. This machine learning approach generates a weighting surface that guides the redistribution of national-level census counts, enabling high-resolution population mapping (Lloyd et al., 2017).

A synthesis of the construction characteristics and origins of the major Global Population Grids, available with resident population estimates for 2020, is presented in Table 1.

Table 1 – Main Global Population Grids for resident population data available for the year 2020.

Grid	Author	Method	Population data	Ancillary data
GPWv4	CIESIN (2018)	Areal weighting	National population censuses adjusted with intercensal estimates	National administrative boundaries Waterbody/permanent ice masks
GHS-POP	European Commission, Joint Research Centre [JRC]: Carioli et al. (2023)	Binary dasymetric	GPWv4 grid	Maps of built-up area density (*GHSL-Build*, adjusted with UN urban areas estimates)
HRSL	Facebook Connectivity Lab & CIESIN (2018)	Volumetric dasymetric	GPWv4 grid	Maps of built-up areas obtained from satellite images (50cm spatial resolution)
WorldPop	University of Southampton: Bondarenko et al. (2020)	Random Forest modelling	GPWv4 grid	Land use, buildings, night time light, vegetation, etc.

Prepared by: The authors (2024).

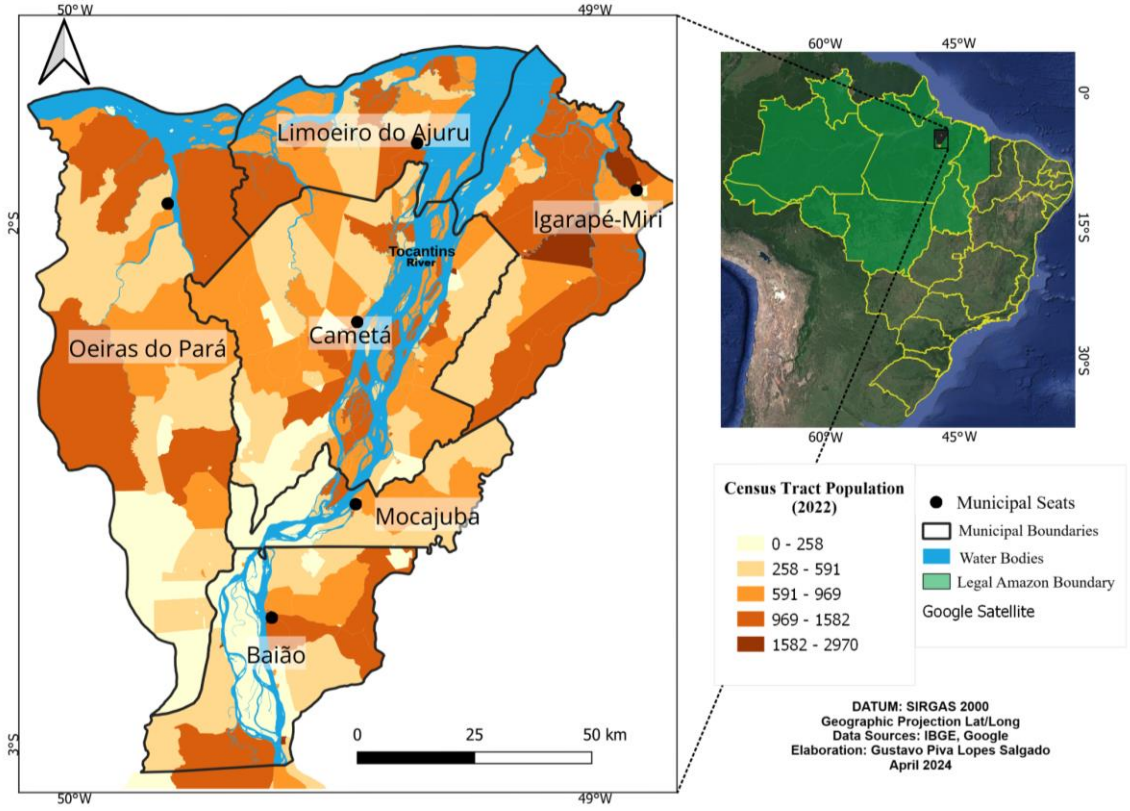
In the brazilian context, the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística [IBGE]) created the country's first national statistical grid using 2010 Census data, including population and residence information. By employing GNSS (Global Navigation Satellite System) devices to geolocate rural residences and integrating urban block face data with census mapping, the 2010 Census achieved unprecedented address-level precision. However, a significant portion of address records lacked geographic coordinates, making pure bottom-up aggregation insufficient for complete statistical grid generation (IBGE, 2016). Hence, IBGE implemented a hybrid approach combining bottom-up aggregation for census tracts with less than 50% missing location data and top-down disaggregation for tracts exceeding this threshold (IBGE, 2016). The disaggregation method followed a preference order: 1) road-proximity dasymetrics where road data existed, 2) land-use/cover dasymetrics when available, and 3) areal weighting as a last resort. The final output featured a 200m-resolution grid for urban areas and 1 km-resolution for rural zones, with each cell coded with its associated method and uncertainty (IBGE, 2016).

3 METHODOLOGY

3.1 Study Area

The study area is located in the Baixo Tocantins region of Pará state, approximately 90 km from the capital Belém, encompassing the municipalities of Baião, Cametá, Igarapé-Miri, Limoeiro do Ajuru, Mocajuba, and Oeiras do Pará (Figure 1). Only the central-northern portion of Baião municipality was included to ensure greater homogeneity in natural land-use characteristics. This transitional zone between fluvial and *terra firme* areas exhibits spatial organization of human activities shaped by environmental and socioeconomic interactions, which directly influence population mobility and distribution patterns (Figure 1).

Figure 1 – Study area location.



Prepared by: The authors (2024).

Among the six municipalities, Cametá stands as the most populous with over 130,000 inhabitants and the highest population density (43.55 inhab/km² as seen in Table 2), yet recorded the region's second-lowest growth rate between 2010-2022. In contrast, Baião showed the study area's highest population growth, likely linked to economic activities that attracted migrants (IBGE, 2022a). Oeiras do Pará, Limoeiro do Ajuru and Mocajuba had the smallest populations (each with 27,000-34,000 inhabitants in 2010-2022) alongside increasing urbanization rates. However these cities had divergent growth patterns – population increases in Limoeiro do Ajuru and Oeiras do Pará versus stagnation in Mocajuba. While Mocajuba has the highest proportion of urban dwellers, Igarapé-Miri, Limoeiro do Ajuru and Oeiras do Pará remain predominantly rural, with Oeiras do Pará notably exhibiting the lowest density (8.79 inhab/km²).

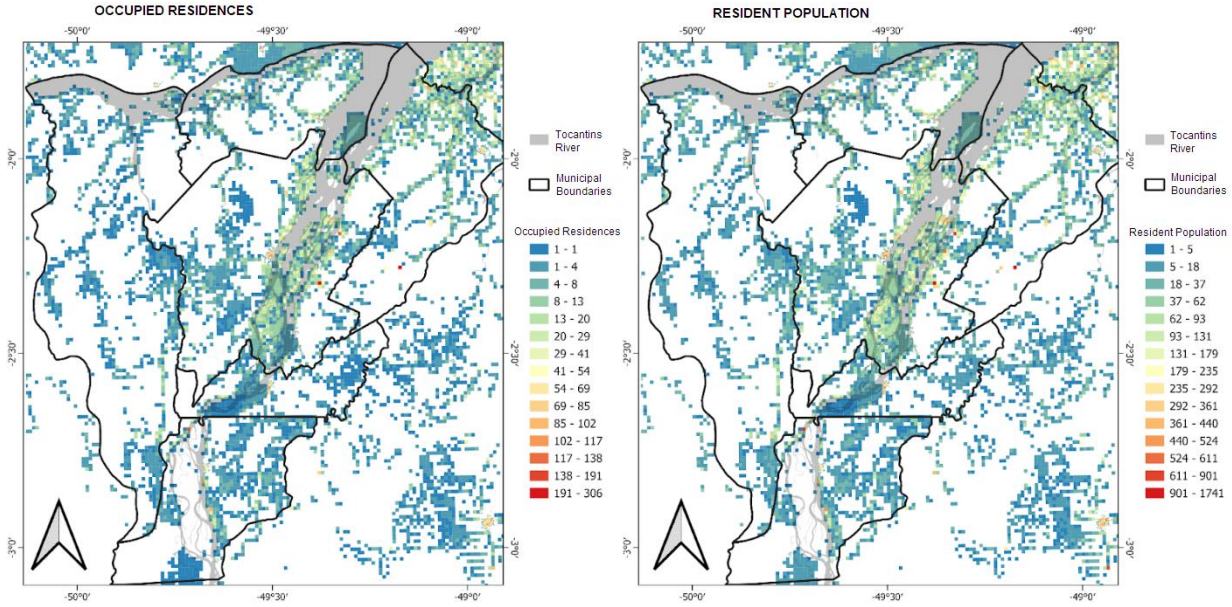
Table 2 – Demographic data of the municipalities in the study area.

Municipality	2010 total population	2010 rural share	2022 total population	2022 rural share	Exponential growth rate (2010-2022)	2022 population density (hab / km²)
Baião	36,882	50 %	51,641	47 %	2.80%	13.73
Cametá	120,896	57 %	134,184	49 %	0.87%	43.55
Igarapé-Miri	58,077	55 %	64,831	55 %	0.92%	32.47
Limoeiro do Ajuru	25,021	76 %	29,569	70 %	1.39%	19.84
Mocajuba	26,731	32 %	27,198	29 %	0.14%	31.22
Oeiras do Pará	28,595	61 %	33,844	54 %	1.40%	8.79

Source: IBGE (2010; 2022a).

Figure 2 presents the distribution of occupied residences and resident population in the study area according to the IBGE 2010 Statistical Grid data (IBGE, 2016). The population is concentrated along riverbanks, in the islands of Tocantins river, and urban areas - particularly in municipal seats – while extensive areas remain free from residents or residences.

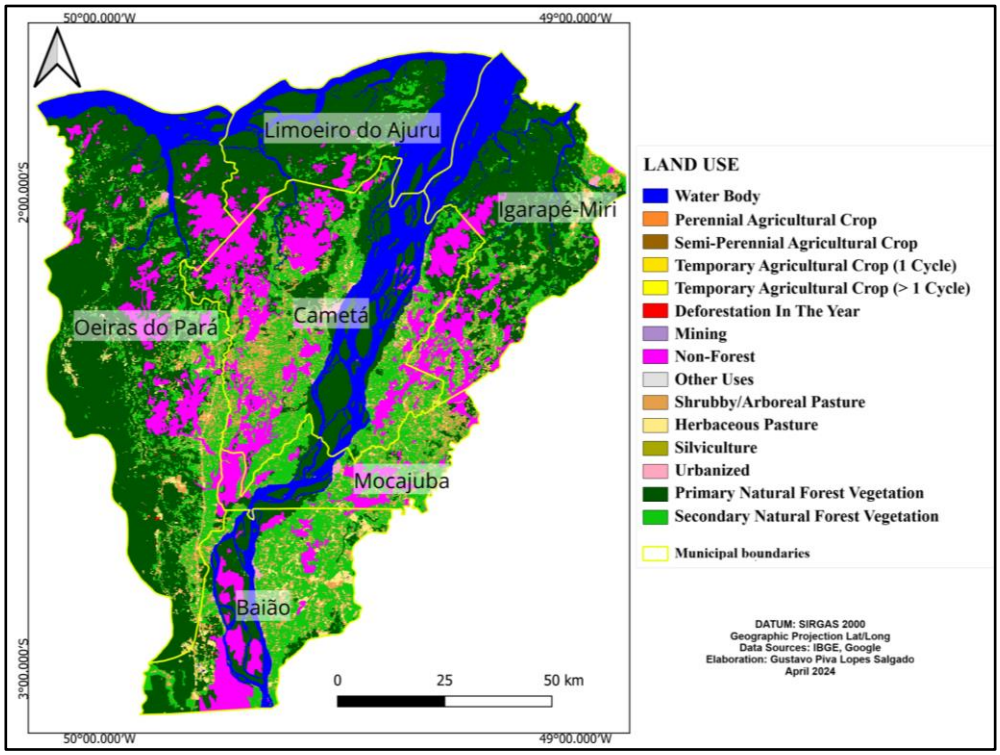
Figure 2 – Occupied residences and resident population in the study area following the IBGE 2010 Statistical Grid.



Prepared by: The authors (2024).

Primary dense ombrophilous forest covers 43.5% of the study area (INPE, 2023), primarily concentrated in the islands and floodplains of the Tocantins river (Figure 3). Notably, approximately 92% of primary forest deforestation in the region occurred before 2007 (INPE, 2023). Natural areas with non-forest phytophysiognomy, locally known as *campos da natureza* (natural grasslands), account for 12.87% of the region (INPE, 2023).

Figure 3 – Land use and cover in the study area, following TerraClass project (INPE & EMBRAPA, 2020).



Prepared by: The authors (2024).

According to Souza et al. (2021), economic activities in Baixo Tocantins vary across *terra firme* (upland), *várzea* (floodplain), and island environments. In *terra firme* and *várzea* areas, beyond cocoa, cassava, and black pepper production, there is cultivation and harvesting of açaí. The past decade was marked by the introduction of planted and irrigated açaí systems in *terra firme* zones (Souza, 2024). Pasturelands occur,

particularly in Baião municipality. Along riverbanks, especially in riverside communities, açaí extraction coexists with other commercially lesser-valued palms like buriti, babaçu, and inajá. The islands are dominated by forest cover with intensively managed açaí stands, which alongside shrimp fishing, are the core economic activities (Souza et al., 2021). These agrarian systems, primarily peasant-driven, are vital to local populations, reflecting historical-geographical processes that shape the landscape (Souza et al., 2019). However, a fraction of these activities are small-scale, often tied to extractive or agroforestry systems under forest canopies, which are frequently underrepresented in remote sensing-based land-use mappings (Souza et al., 2019, 2023).

Consequently, the region poses significant challenges for mapping occupied areas to assist gridded demographic representation, characterized by extensive forest cover, agroforestry production systems, uneven rural population distribution, environmental heterogeneity, and unique riverine dynamics that influence population mobility and settlement patterns. These factors are further complicated by large census tracts (averaging 27.35 km²) and extreme population density variations (0.14 to 15,400 inhab/km² across tracts) (IBGE 2022a, 2022b), creating substantial obstacles for data disaggregation into regular grid cells.

3.2 Materials and Methodological Procedures

3.2.1 MATERIALS

Four Global Population Grid datasets were selected for analysis based on two key criteria: containing resident population estimates and having available data for 2020. The datasets used in our analysis are detailed in Table 3.

Table 3 – Data used in this study.

Data	File	Version	Format	Spatial resolution
Municipal population estimates	POP2020_20220905.xls	2020	Tabular	N/A
Census - Municipal level	tabela_2_1_PA.xlsx	2022		
Census - tract level	Agregados_preliminares_por_setores_censitarios_PA.csv			
IBGE statistical grid	grade_id75.shp	2010	Vector Polygon	1 km (rural areas) e 200 m (urban areas)
Census tract - Pará state	PA Malha _ Preliminar _ 2022.shp	2022		N/A
Municipal mesh - Pará state	PA Municípios _ 2022.shp			
Brazilian National Registry of Addresses for Statistical Purposes (CNEFE)	15 _PA.csv	2022	Vector Point	N/A
Gridded Population of World (GPW)	gpw_v4_population_count _rev11_2020_30_sec.tiff	2020	Raster	30 arcsec (~1 km at equator)
WorldPop	bra_ppp_2020_UNadj_constrained.tiff	2020	Raster	3 arcsec (~100 m at equator)
Global Human Settlement Pop Layer (GHS-POP)	GHS_POP_E2020_GLOBE_R 2023A_4326_3ss_V1_0_R10_C13.tiff	2020	Raster	
	GHS_POP_E2020_GLOBE_R 2023A_4326_3ss_V1_0_R10_C14.tiff	2020		
High-Resolution Settlement Layer (HRSL)	population_bra_northeast_2018-10-01.tiff	2020	Raster	1 arcsec (~ 30 m at equator)

Prepared by: The authors (2024).

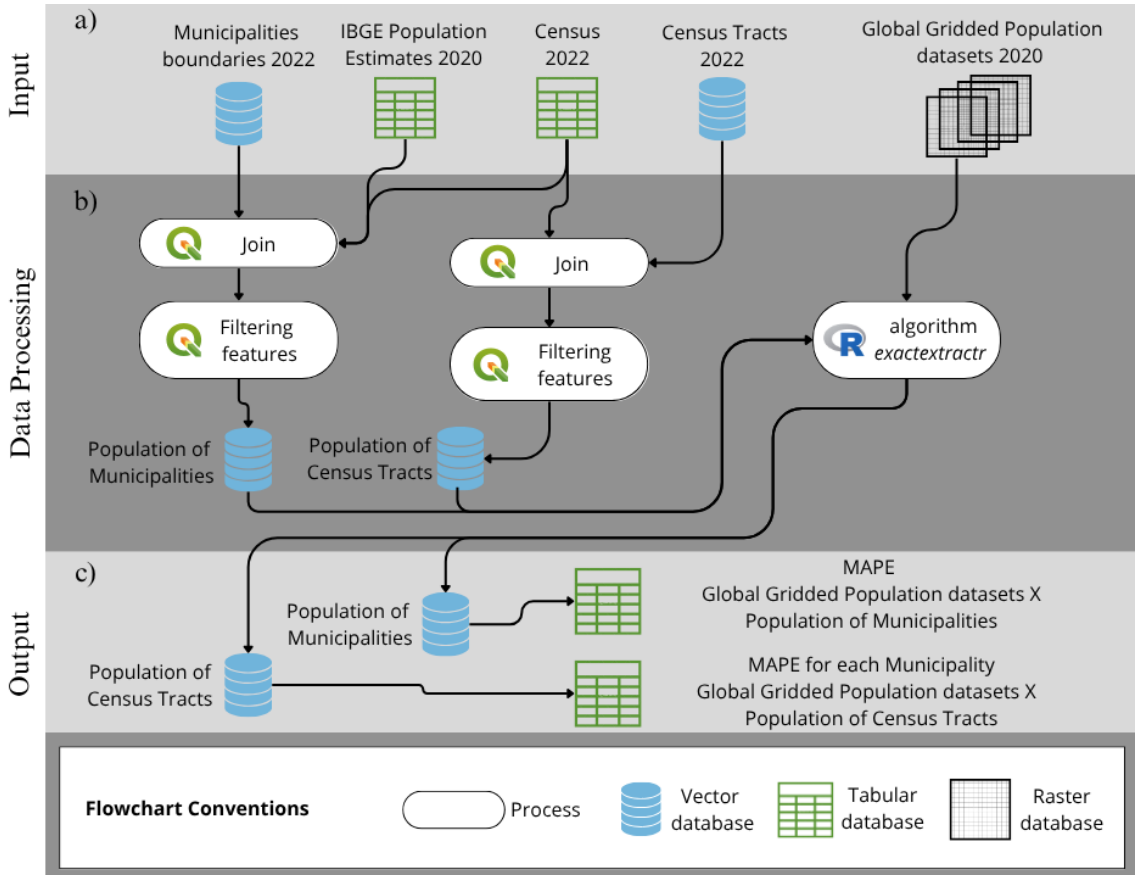
3.2.2 METHODS

The population density estimates from Global Gridded Population datasets were analyzed for both concordance and adherence to the most recent census data (2022 Demographic Census). The concordance analysis compares datasets using consistent measurement units – in this case, population count estimates – across different grids and time periods. The adherence analysis evaluates how well-suited is the alignment between population distribution and a strongly correlated proxy data, represented here by residence counts. All spatial data processing was conducted using RStudio (a programming environment for R) alongside the QGIS geographic information system.

3.2.2.1 Concordance Analysis

The 2022 Demographic Census data, available at municipal and census tract levels, enabled analyzing agreement between Global Population Grid estimates within these spatial units (Figure 4). Using the zonal statistics function from the R package exactextractr[2], we aggregated Global Grid estimates to match municipal and census tracts boundaries within the study area. This calculation accounted for area-weighted intersections, where each grid cell's population value was proportionally allocated based on the overlapping area between the cell and administrative boundaries.

Figure 4 – Workflow for concordance analysis.



Prepared by: The authors (2024).

Using tabular data containing both the Global Grid population estimates and reference values at municipal and census tract levels, we calculated the Mean Absolute Percentage Error (MAPE) for each Grid relative to these reference datasets. The MAPE metric (Equation 1) assesses precision by measuring the average absolute deviation between reference and estimated values, standardized through ratio between

[2] *It is an implementation of zonal statistics that can handle grid cells that are partially covered by a polygon, weighting the coverage area and extracting values using that weighting (BASTON, 2023).*

estimated and reference values.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Grid_i - Reference_i}{Reference_i}\right| \quad (1)$$

For the municipal-level analysis, two official reference datasets were used: (1) the 2022 Demographic Census results (IBGE, 2022a), representing Brazil's most recent population count and thus the most reliable comparison parameter, despite its two-year discrepancy from the Global Grids' 2020 reference year; (2) the municipal population estimated by IBGE on July 1, 2020 (IBGE, 2020), which corresponds to official projected values and serves as the demographic basis adopted by the Global Grids – making it essential for evaluating how well the grids align with official population estimates. Importantly, this study does not evaluate the accuracy of IBGE or Global Grid estimates against a "true" population, as municipal projections for 2020 inherently face limitations in capturing abrupt growth rate changes in small areas (Campos, 2017). These estimates depend not only on fertility and mortality but significantly on migration, which can exhibit substantial fluctuations that abruptly alter growth trajectories (Smith & Morrison, 2005). Since global grids derive directly (e.g., GPWv4) or indirectly from UN estimates, their quality is contingent on the underlying official data. Thus, comparing IBGE's 2020 municipal estimates with grid-aggregated values serves solely to assess: (a) the consistency of UN-based global projections (UN, 2015) with Brazil's national statistical office figures, and (b) potential data loss when disaggregating national estimates to municipal boundaries (IBGE, 2020).

3.2.2.2 Adherence Analysis

The primary and most widely used variable for explaining resident population distribution is built infrastructure, which serves as a proxy in Global Gridded Population datasets like WorldPop, GHS-POP, and HRSL. The concept of 'resident population' relates to human habitation patterns, indirectly measured through satellite imagery by detecting spectral signatures of impervious surfaces (Zha et al., 2003; Lu & Weng, 2006), building volumes (Tiecke et al., 2017), and nighttime lights (Elvidge et al., 1997).

Brazil's most detailed population location data comes from Brazilian National Registry of Addresses for Statistical Purposes (CNEFE), a georeferenced database of national addresses cataloged since 2005 and continuously updated upon demand from IBGE surveys, mostly the demographic censuses (IBGE, 2024; 2024a). The CNEFE dataset provides information on both the spatial concentration and distribution of addresses, as well as their functional classification, categorized as follows: 1 (Private residences), 2 (Collective residences), 3 (Agricultural establishments), 4 (Educational institutions), 5 (Health facilities), 6 (Other establishments), 7 (Buildings under construction), and 8 (Religious establishments) (IBGE, 2023). These data informed the IBGE's 2010 Statistical Grid, where rural residences are represented by CNEFE geolocated points and urban residences by block face vectors with associated residence counts (IBGE, 2016).

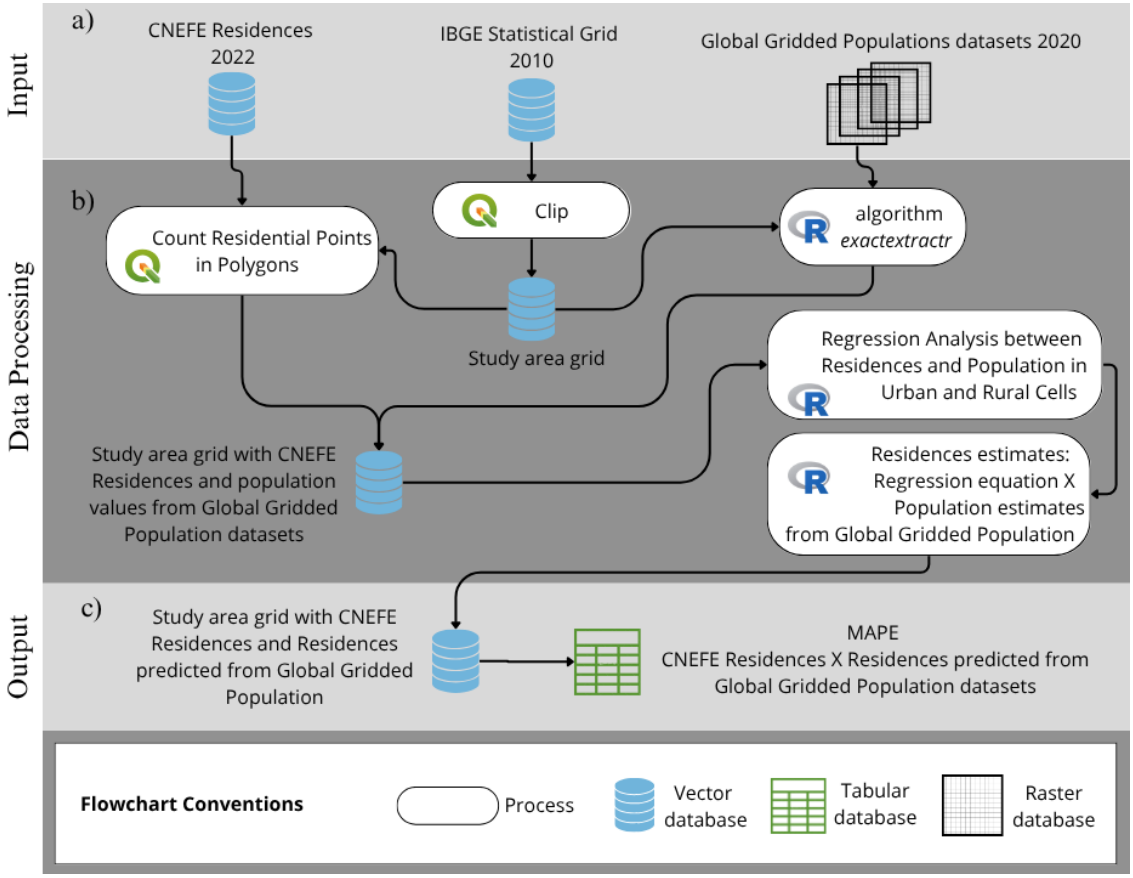
Regarding the quality of geographic coordinates in the CNEFE for the 2022 Demographic Census, accuracy was ensured through a rigorous collection, validation, and correction process. Enumerators collected coordinates using Global Navigation Satellite System (GNSS) receivers embedded in their Mobile Data Collection Devices. While imprecisions exist due to equipment limitations and environmental conditions during collection, the root mean square error (RMSE) of coordinates was 5.84 meters under ideal conditions, extending up to 11.71 meters under normal conditions, and potentially exceeding this threshold in dense urban areas or hard-to-access locations (IBGE, 2024a). Quality assurance involved validation techniques that adjusted and corrected coordinates when necessary, using proximity to reference points like block faces and neighboring addresses (IBGE, 2024b). The CNEFE also implemented a geocoding process that assigned valid coordinates, resorting to fallback methods – such as previous coordinates or block faces midpoints – when primary methods failed (IBGE, 2024b). However, 99.68% of Brazil's addresses were successfully geocoded at the address-level without requiring additional correction, with Pará state achieving a 97.44% success rate (IBGE, 2024b). The positional quality of CNEFE data demonstrates high precision and coverage, making it

reliable for this study's purposes.

Thus, regarding that built-up areas and residences serve as proxies for population distribution, we analyzed the adherence of Global Grid estimates to the 2022 CNEFE residence registry data within the study area, incorporating both rural and urban classifications from IBGE's Statistical Grid. This adherence analysis essentially converted the variables into a residence density indicator.

The workflow for adherence analysis is summarized in Figure 5. First, we clipped IBGE's 2010 Statistical Grid of Pará state to our study area boundaries while preserving its original dimensions (1 x 1 km for rural zones and 200 x 200 m for urban areas). From the 2022 CNEFE address database, we filtered only geocoded points classified as either private residences (Code 1) or collective residences (Code 2). The clipped grid was then populated with residence counts using QGIS's "Count Points in Polygons" spatial operator. Subsequently, we aggregated Global Gridded Population datasets into the study area grid cells using the “exactextractr” algorithm, which performs area-weighted summation of raster pixel values based on their proportional overlap with each target grid cell.

Figure 5 – Workflow for adherence analysis.



Prepared by: The authors (2024).

To operationalize the adherence analysis, we quantified the statistical relationship between occupied residences ('DOM_OCU') and resident population ('POP') variables from IBGE's 2010 Statistical Grid. Regression analysis was performed with DOM_OCU as the independent variable and POP as the dependent variable, with separate models fitted for rural and urban grid cells. To satisfy regression model assumptions, the equations between residences and population in IBGE's 2010 Statistical Grid required residuals demonstrating both normality and homoscedasticity (constant variance), ensuring valid confidence intervals for statistical hypothesis testing (Breusch & Pagan, 1979). Under the assumption that this statistical relationship remained substantially unchanged from the 2010 Census baseline to 2022, the fitted regression equation was applied to predict residence counts from Global Grid data, which were then compared against actual 2022 CNEFE residence data (IBGE, 2022a). This approach provided an indirect evaluation of Global Population Grid’s estimates alignment with real population distribution patterns at intraregional and intramunicipal scales.

A simple linear regression framework was adopted with transformations applied to both the predictor (residences) and the response (population) variables as required to achieve compliance with statistical assumptions. Separate regression models were evaluated for urban and rural areas. The final urban area model (Equation 2) required only a quadratic transformation of the response variable. For rural areas (Equation 3), in addition to the quadratic transformation of the response variable, it was necessary to weight the model samples based on the inverse of the residual variance from an initial linear regression model. This weighting scheme specifically addressed heteroscedasticity in the initial model's residuals, ensuring that observations with higher residual variance exerted less influence on parameter estimates. The resulting robust model satisfied both normality and homoscedasticity requirements for valid statistical inference.

$$POP^2 = 4.866047 + (0.170471 * DOM\_OCU) \quad (2)$$

$$POP^2 = 2.73807 + (0.33765 * DOM\_OCU) \quad (3)$$

The regression models for urban and rural areas yielded determination coefficients (R^2) of 0.93 and 0.90, respectively. which confirm the models are well-fitted to the data. By inverting the equations to isolate the 'residences' term, we derived predictive equations capable of estimating residence densities from population densities – such as those provided by Global Grids. The residence prediction equations, using population density as input for urban and rural areas respectively, are:

$$Predicted\ residences\ in\ urban\ areas = \frac{Population\ Grid\ Estimates^{\,2} - 4.866047}{0.170471} \quad (4)$$

$$Predicted\ residences\ in\ rural\ areas = \frac{Population\ Grid\ Estimates^{\,2} - 2.73807}{0.33765} \quad (5)$$

The resulting regression equations were applied to estimate 2020 residence densities for both urban and rural areas using population estimates from Global Grids. This generated a novel dataset serving as a directly comparable proxy to the CNEFE 2022-derived residence densities per grid cell. Using these outputs, we calculated Mean Absolute Percentage Errors (MAPEs) to evaluate error distributions at the cellular grid scale.

4 RESULTS AND DISCUSSION

4.1 Concordance Analysis

Table 4 presents the differences between population estimates derived from Global Grids and official data obtained from both IBGE's 2020 Municipal Estimates and the 2022 Demographic Census. The Global Population Grids show smaller discrepancies relative to IBGE's 2020 Municipal Estimates, as all analyzed grids directly or indirectly rely on UN population estimates for 2020.

Table 4 – Mean Absolute Percentage Errors (MAPE) between population estimates from global grids and official reference data at municipal level.

2020 Population Grid	2020 IBGE estimate	2022 Census
GHS-Pop	4.02%	6.64%
GPWv4	11.42%	12.88%
HRSL	1.75%	4.77%
WorldPop	22.96%	24.34%

Prepared by: The authors (2024).

GPWv4's modeling (last updated in 2018) is based on census counts and historical/future population

projections from the "UN World Population Prospects: Revision 11" (CIESIN, 2018), which itself incorporates national statistical office data, including IBGE's (UN, 2015). While UN estimates are adjusted nationally, they may include finer administrative levels depending on data availability (CIESIN, 2018). For Brazil, GPWv4's raster datasets were built from 2010 census tract boundaries and projected to 2020, then disaggregated to grid level (CIESIN, 2018). Since GHS-Pop, HRSL, and WorldPop use GPWv4 as input, they inherit the same 2020 reference projections. Consequently, projections originating from the same statistical offices – although generated at different times (2015 for those adopted by the Global Grids and 2020 for IBGE estimates used as reference here) – tend to converge toward similar values, given their shared data sources, methods, and the common 2020 projection year.

Despite serving as input for other grids, GPWv4 showed the second-highest errors: 11.42% against IBGE 2020 Estimates; 12.88% against 2022 Census, highlighting that 1 km spatial resolution challenges municipal-level representation. WorldPop exhibited the largest deviations: 22.96% (IBGE, 2020) and 24.34% (2022 Census) likely due to its heavy reliance on auxiliary variables (some non-indicative of regional population) and its Random Forest model, which does not preserve input population volumes (Lei et al., 2023). HRSL outperformed others with minimal 1.75% discrepancy from IBGE 2020, and the lowest deviation from 2022 Census counts. Its success likely comes from 30m spatial resolution (vs. GHS-Pop's 100m), centimeter-scale satellite imagery, convolutional neural networks for built-area detection, enabling a more precise dasymetric modeling.

Table 5 summarizes the MAPE values between Global Grids and 2022 reference data at the census tract level. This finer territorial comparison revealed even higher error magnitudes, underscoring the challenges these datasets face in representing population distributions at intramunicipal scales. Despite this, HRSL demonstrated the lowest discrepancies across all municipalities, followed by GHS-Pop. This advantage was particularly evident in Cametá, Igarapé-Miri, and Limoeiro do Ajuru. HRSL and GHS-Pop had nearly identical error margins in Oeiras do Pará. GPWv4 showed the highest errors in four of six municipalities, outperformed only by WorldPop in: Limoeiro do Ajuru (highest rural population percentage) and Mocajuba (lowest rural population percentage).

Table 5 – Mean Absolute Percentage Errors (MAPEs) between 2020 global population grids and 2022 Brazilian census tract data at each municipality.

Municipalities	GHS-Pop	GPWv4	HRSL	WorldPop
Baião	51.81 %	256.74 %	46.38 %	94.57 %
Cametá	48.14 %	135.19 %	47.28 %	103.20 %
Igarapé-Miri	46.28 %	65.03 %	44.10 %	62.47 %
Limoeiro do Ajuru	25.81 %	53.87 %	23.71 %	79.85 %
Mocajuba	62.37 %	156.47 %	49.73 %	191.77 %
Oeiras do Pará	42.67 %	99.52 %	42.51 %	84.90 %

Prepared by: The authors (2024).

As noted by Doxsey-Whitfield et al. (2015), areal weighting's primary advantage is preserving input data volumes, while its key limitation is pixel-level accuracy variability – a direct function of input administrative/census unit sizes. This is particularly problematic where input units are large, as evidenced by GPWv4's performance: its 1 km-resolution pixels (disaggregated via areal weighting) had the second-highest MAPE against municipal estimates (Table 4) and the largest errors in 4 out of 6 municipalities at census-tract level (Table 5).

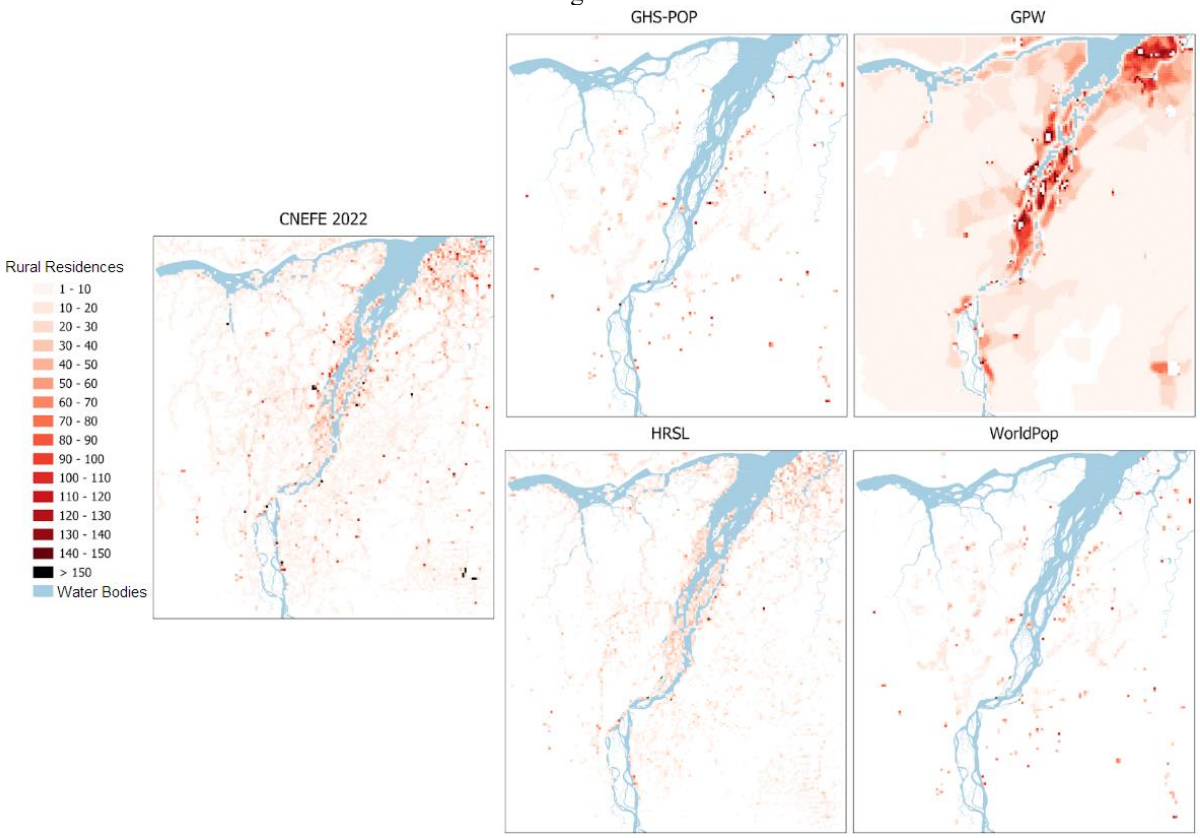
Except for WorldPop, all Global Grids achieved lower errors in Limoeiro do Ajuru. This municipality had the highest share of rural population, stable temporal demographics, and homogeneous spatial distribution. These factors likely facilitate predictability of population for the grid estimates. Conversely, Baião, Cametá, and Mocajuba exhibited the highest discrepancies across all datasets, attributable to rapid urban growth (2010-2022) in Baião; a more consolidated urbanization, despite slow growth, in Mocajuba. The level of urbanization

and its dynamics may contribute to a more heterogeneous distribution of the population, complicating prediction. As a common challenge, these three municipalities are in part covered by non-forest natural physiognomies. These natural grassland areas' spectral signatures – with strong bare soil contributions during dry seasons – often resemble urbanized surfaces in satellite imagery. Such land-cover misclassification propagates errors into population redistribution within Global Grids that rely on built-up area data.

4.2 Adherence Analysis

The distribution of predicted residences for Global Gridded Population datasets, derived through rural-cell regression modeling, is presented in Figure 6.

Figure 6 – 2022 CNEFE Residence reference and 2020 Global Grid Residence Predictions using linear regression in rural grid cells.



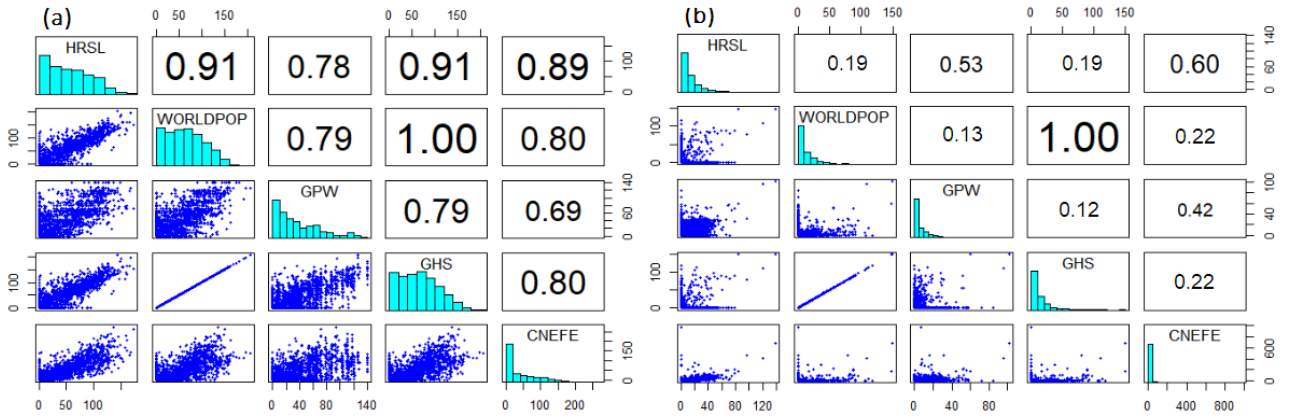
Prepared by: The authors (2024).

The analysis reveals that the HRSL-derived product most accurately represented the 2022 CNEFE residence density distribution, particularly in capturing settlement patterns along riverside and island areas – the region's primary population concentration zones. The GPWv4 grid also successfully identified residence clusters in these riparian and insular environments, though with less spatial precision. In contrast, both GHS-Pop and WorldPop systematically underestimated residence densities in these critical zones, indicating fundamental limitations in their modeling approaches for this specific regional context.

While GHS-Pop demonstrated strong performance at municipal scales (making it suitable for regional analyses) and achieved results comparable to HRSL at census tract level, its consistent underestimation in fluvial and island areas makes it inappropriate for fine-scale studies. WorldPop showed the poorest performance across all evaluation metrics, making it particularly unreliable for population distribution studies in the Amazon, especially in regions with similar rural-urban configurations.

Statistical validation (Figure 7) shows HRSL achieved the highest correlation coefficients with 2022 CNEFE reference data: 0.89 for urban cells (200m resolution) and 0.60 for rural cells (1km resolution). GPWv4 showed the weakest urban correlation (0.69), while GHS-Pop and WorldPop both produced the lowest rural correlations (0.22 each). These results quantitatively confirm HRSL's superior ability to replicate observed settlement patterns across both urban and rural landscapes in the study area.

Figure 7 – Histograms (diagonal), scatter plots (blue), and Correlation Coefficients (R) between predicted residences from Global Grids (2020) and official CNEFE data (2022) for urban (a) and rural (b) grid cells in the study area.



Prepared by: The authors (2024).

According to Bondarenko et al. (2020), the WorldPop disaggregation method is advantageous for providing highly accurate spatial distribution when settlements and buildings are precisely mapped, avoiding population predictions in uninhabited areas. However, while this approach preserves total population volumes, its strength becomes a limitation when buildings are misclassified (inclusion errors) or omitted (omission errors). These errors can lead to: (1) underestimation of population, by allocating people from correctly identified cells to those with inclusion errors; or (2) overestimation of population, by distributing population from cells with omission errors to correctly identified ones. This explains why WorldPop and GHS-Pop grids underestimated residence predictions in island and riverside zones while overestimating in *terra firme* environments. As noted by Gonçalves et al. (2021), auxiliary urban infrastructure data may not adequately capture built environments in forested riparian and island areas, as these lack the characteristic housing patterns of dense human settlements. False positives may also occur when natural non-forest vegetation areas are misidentified as urban surfaces by image classification algorithms.

Table 6 presents the MAPE values for residence predictions across urban and rural strata. For urban cells, errors varied substantially: HRSL showed the lowest error (11%), followed by GHS-Pop and WorldPop (both 32%), with GPWv4 performing significantly worse (93%). Rural cells exhibited higher errors across all grids – GPWv4 (106%), WorldPop (78%), GHS-Pop (79%), and HRSL (34%) – confirming that rural areas pose greater challenges for residence prediction accuracy. HRSL maintained superior performance in both contexts.

Table 6 – Mean Absolute Percentage Errors (MAPE) in residence density predictions derived from Global Population Grids versus CNEFE 2022 reference data across urban and rural strata.

	GHS-Pop	**GPWv4**	**HRSL**	**WorldPop**
Urban grid cells	32%	93%	11%	32%
Rural grid cells	79%	106%	34%	78%

Prepared by: The authors (2024).

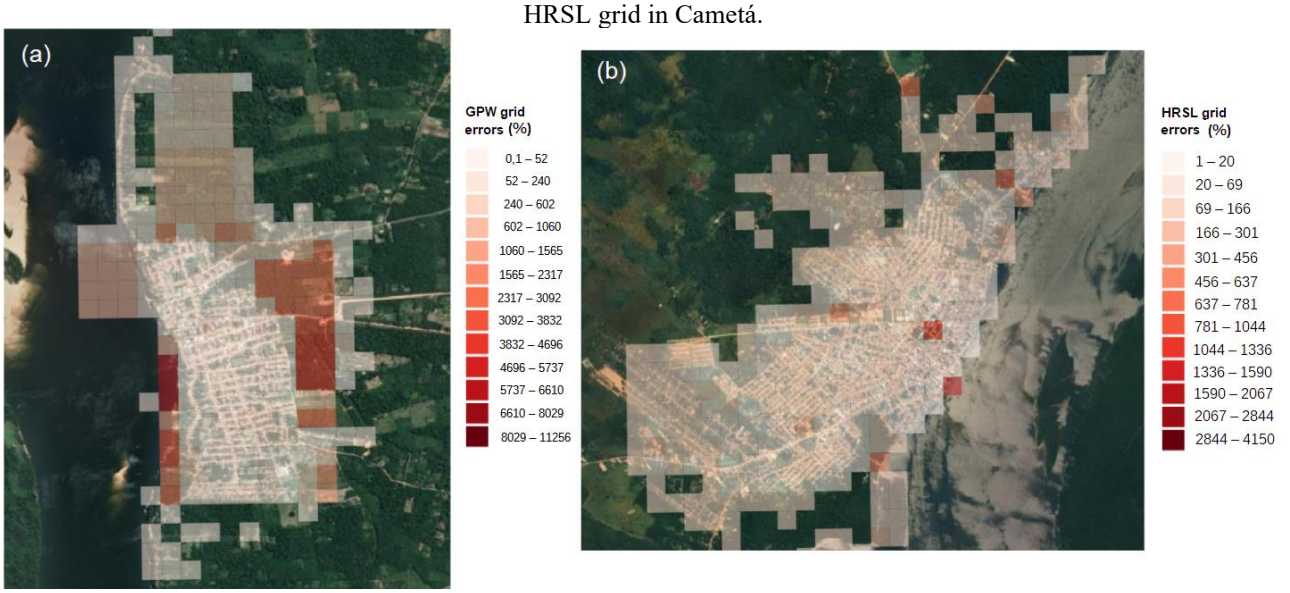
These findings align with international literature demonstrating how the accuracy of gridded population models varies according to distinct patterns of built-up or impervious surface concentrations mapped by their auxiliary datasets. Population cluster detection performs optimally in high-density urban areas, with accuracy progressively declining along urban-rural gradients where built environments become more dispersed (Harvey, 2002; Leyk et al., 2018; Silva, 2023). The reduced accuracy of grids in rural areas comes from challenges in differentiating impervious surfaces within auxiliary data due to spectral reflectance similarities with other landscape elements – particularly grass-dominated land cover (Wickham et al., 2013) and bare soil (Weng & Hu, 2008). These limitations underscore the critical need for: (1) models capable of handling spectral mixing during built-area estimation, and (2) sensors with enhanced spatial resolution to better delineate these environments in auxiliary datasets, thereby improving Global Gridded Population

representativeness in rural contexts.

For rural areas, the GPWv4 grid showed the second-highest correlation coefficient (0.42), yet it presented the highest mean percentage error for rural cells (106%). The correlation coefficient indicates that the strength and direction of GPWv4 estimates follow the same trend as CNEFE data, and with even greater consistency than GHS-Pop and WorldPop grids. However, the high MAPE reveals that, on average, these estimates are numerically farther from real values, likely due to systematic bias from the grid's low resolution and the areal weighting population redistribution method that only used water masks as auxiliary data (without infrastructure data to guide density modeling).

The resolution issue also affects urban cells (Figure 8a), where high-density values available at coarse resolution generate larger errors when disaggregated to finer grids – particularly around built-up areas that were contained within single 1 km GPWv4 cells. Figure 8b illustrates error variability in the HRSL grid, with most values concentrated between 1% and 69%, which was lower than that observed in the other analyzed Global Population Grids. The analysis revealed that the highest errors are heterogeneously distributed across Cametá's urban area with a slight tendency to cluster in peripheral zones, suggesting HRSL's – albeit reduced – limitation in capturing accurate densities in low-population density areas and urban-rural transition zones. Some central/port-area cells also showed elevated errors, indicating satellite imagery's difficulty in capturing complex, dense building patterns in these zones.

Figure 8 – Absolute Percentage Errors (APEs) of residence predictions in urban areas: (a) GPWv4 grid in Baião, and (b) HRSL grid in Cametá.



Prepared by: The authors (2024).

5 FINAL REMARKS

The quality of major Global Population Grids for representing population distribution in the Baixo Tocantins region was evaluated through concordance with official population data and adherence to residence distribution patterns. While 2022 census data is available at municipal and tract levels, a robust quality assessment of Global Grids requires direct comparison with official population data (dis)aggregated into standardized units – such as the official Statistical Grid produced from the 2010 Census.

The adherence analysis employed an indirect approach using released 2022 CNEFE data, introducing a novel evaluation method that treats residence locations as population distribution proxies. This detailed analysis revealed specific under/overestimation patterns linked to local population distribution characteristics and new understandings into Global Grid limitations when representing both fine-scale population distributions and context-specific settlement patterns.

This study revealed significant variations in Global Gridded Population datasets accuracy across municipalities, highlighting the critical importance of grid selection for research outcomes. However, these findings have limited generalization, as they are intrinsically related to the specific characteristics and unique features of the Baixo Tocantins region. Specifically, the remote sensing data and methods used for population

modeling in regions like Baixo Tocantins face inherent limitations in representing residence infrastructure pattern, territorial occupation specificities that shape auxiliary variable mapping. These global models are primarily calibrated for industrial-capitalist urbanization patterns, which may not reflect local realities. This way, extrapolations to other regions – even within the Brazilian Amazon – require caution, particularly for areas with substantially different: rural-urban composition, population density profiles, infrastructure development, land use/cover dynamics. Nevertheless, this study establishes a valuable framework for assessing the strengths and limitations of major Global Grids in contexts that diverge from conventional urban concentration patterns, such as traditional Amazonian settlement schemes.

The HRSL grid demonstrated consistent superiority across all evaluated areas and experimental tests. Its methodology and input data proved better suited to capture the unique characteristics of population distribution in the study region. When the 2022 Census Statistical Grid data becomes available, replicating this study will be essential to validate findings against refined official reference data.

This work advances critical discussions about the representativeness of Global Gridded Population data in underrepresented contexts like the Brazilian Amazon. Also it provides methodological frameworks for assessing grid limitations and potentials, as well as empirical results and methodological benchmarks to support informed selection of the most reliable population distribution products for regionally specific contexts.

Acknowledgments

This study was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES – Finance Code 001), the Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq, process number: 130659/2023-1) and the Harmonize project, funded by the Wellcome Trust Foundation. (Grant No. 224694/Z/21/Z).

Authors Contributions

Conceptualization: G. P. L. S., A. P. D. e S. A.; Data curation: G. P. L. S.; Formal analysis: G. P. L. S.; Acquisition of funding: none; Research: G. P. L. S.; Methodology: G. P. L. S.; Project management: G. P. L. S., A. P. D. e S. A.; Resources: G. P. L. S.; Software: G. P. L. S.; Supervision: A. P. D. e S. A.; Validation: G. P. L. S.; Visualization: G. P. L. S. Writing: G. P. L. S., A. P. D; Writing -revision and editing: G. P. L. S., A. P. D., B. V. A. e S. A.

Conflicts of Interest

The authors state that there is no conflict of interest.

References

Amaral, S., Gavlak, A. A., Escada, M. I. S. & Monteiro, A. M. V. (2012). Using remote sensing and census tract data to improve representation of population spatial distribution: case studies in the Brazilian Amazon. *Population and Environment,* 34, 142–170. https://doi.org/10.1007/s11111-012-0168-2.

Archila Bustos, M. F., Hall, O., Niedomysl, T. & Ernstson, U. (2020). A pixel level evaluation of five multitemporal global gridded population datasets: a case study in Sweden, 1990–2015. *Population and Environment, 42*, 255-277. https://doi.org/10.1007/s11111-020-00360-8.

Bai, Z., Wang, J., Wang, M., Gao, M., & Sun, J. (2018). Accuracy assessment of multi-source gridded population distribution datasets in China. *Sustainability*, 10(5), 1363. https://doi.org/10.3390/su10051363

Baston, D. (2023). *exactextractr version* (v0.10.0) [R package]. GitHub. https://github.com/isciences/exactextract.

Bondarenko, M., Kerr, D., Sorichetta, A. & Tatem, A. J. (2020). *Census/projection-disaggregated gridded*

population datasets, adjusted to match the corresponding UNPD 2020 estimates, for 183 countries in 2020 using Built-Settlement Growth Model (BSGM) outputs [Data set]. WorldPop: University of Southampton, UK. https://hub.worldpop.org/geodata/summary?id=49921

Breusch, T .S. & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica,* 47(5), 1287–1294. https://doi.org/10.2307/1911963

Calka, B., & Bielecka, E. (2020). GHS-POP accuracy assessment: Poland and Portugal case study. *Remote Sensing*, 12(7), 1105. https://doi.org/10.3390/rs12071105

Campos, J. (2017). *Estimativas populacionais a partir de dados orbitais de média resolução espacial: aplicações em municípios da Região Metropolitana de Belo Horizonte. [*Tese de doutorado, Universidade Federal de Minas Gerais] Repositório digital de teses e dissertações da UFMG. https://repositorio.ufmg.br/bitstream/1843/FACE-B27P7Y/1/tese___j_rvis_campos.pdf.

Cardoso, A. C. D., Dal'Asta, A. P. & Monteiro, A. M. V. (2023). O que é o urbano na Amazônia contemporânea? Implicações para a vigilância em saúde no bioma. *Cadernos de Saúde Pública*, 39(9), e00129723. https://doi.org/10.1590/0102-311XPT129723

Carioli, A., Schiavina, M., Freire, S. & MacManus, K. (2023). *GHS-POP R2023A - GHS population grid multitemporal (1975-2030)*. European Commission, Joint Research Centre (JRC) [Dataset]. https://doi.org/10.2905/2FF68A52-5B5B-4A22-8F40-C41DA8332CFE

Center for International Earth Science Information Network [CIESIN], Columbia University. (2018). *Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45Q4T5F.

Chen, R., Yan, H., Liu, F., Du, W., & Yang, Y. (2020). Multiple Global Population Datasets: Differences and Spatial Distribution Characteristics. *ISPRS International Journal of Geo-Information*, 9(11), 637. https://doi.org/10.3390/ijgi9110637.

Dahmm, H. & Rabiee, M. (2020). *Leaving no one off the map: A guide for gridded population data for sustainable development*. Thematic Research Network on Data and Statistics (TReNDS) of the UN Sustainable Development Solutions Network (SDSN), in support of the POPGRID Data Collaborative. https://www.popgrid.org/sites/default/files/documents/Leaving_no_one_off_the_map.pdf

Doxsey-Whitfield, E., MacManus, K., Adamo, S. B., Pistolesi, L., Squires, J., Borkovska, O. & Baptista, S. R. (2015). Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography, 1*(3), 226-234. https://doi.org/10.1080/23754931.2015.1014272

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R. & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing, 18*(6), 1373-1379. https://doi.org/10.1080/014311697218485

Facebook Connectivity Lab & Center for International Earth Science Information Network [CIESIN] - Columbia University. (2018). *High Resolution Settlement Layer (HRSL) - Source imagery for HRSL © 2016 DigitalGlobe - Brazil Population Density Maps and Demographic Estimates (2015-2020) [Data set]*. https://data.humdata.org/dataset/brazil-high-resolution-population-density-maps-demographic-estimates.

Fearnside, P. M (2006). Desmatamento na Amazônia: dinâmica, impactos e controle. *Acta Amazonica*, Manaus, 36(3), 395-400. https://doi.org/10.1590/S0044-59672006000300018

Gonçalves, G. C., Oliveira, L. M. de, Dal'Asta, A. P. & Amaral, S. (2021). Geoinformação Para a Visibilidade das Áreas Urbanas de Cidades Amazônicas. *Revista Geoaraguaia, 11*(Especial), 149-165. https://periodicoscientificos.ufmt.br/ojs/index.php/geo/article/view/12742.

Hallisey, E., Tai, E., Berens, A., Wilt, G., Peipins, L., Lewis, B., Graham, S., Flanagan, B. & Lunsford, N. (2017). Transforming Geographic Scale: A Comparison of Combined Population and Areal Weighting to Other Interpolation Methods. *International Journal of Health Geographics*, 16(29).

https://doi.org/10.1186/s12942-017-0102-z

Harvey, J. T. (2002). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, 23(10), 2071-2095. https://doi.org/10.1080/01431160110075901

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2010). *Censo Demográfico 2010 [Dataset].* Rio de Janeiro: IBGE. https://censo2010.ibge.gov.br.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2016). *Grade Estatística IBGE 2010*. Rio de Janeiro: IBGE. https://biblioteca.ibge.gov.br/visualizacao/livros/liv102043.pdf

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2020). Estimativas de população residente por município, data de referência em 1º de Julho de 2020. *Diário Oficial da União em 27 de agosto de 2020*. https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=28674.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2022a). *Censo Demográfico 2022 [Dataset]*. Rio de Janeiro: IBGE. https://censo2022.ibge.gov.br.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2022b). *Malha de setores censitários do Pará - Edição 2022 [Dataset]*. Rio de Janeiro: IBGE. https://www.ibge.gov.br/geociencias/downloads-geociencias.html?caminho=organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2022_preliminar/setores/shp/UF.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2022c). *Malha territorial municipal brasileira - Edição 2022 [Dataset].* Rio de Janeiro: IBGE. https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2023). *Cadastro Nacional de Endereços para Fins Estatísticos (CNEFE) do Censo 2022*. Rio de Janeiro: IBGE. https://www.ibge.gov.br/estatisticas/sociais/populacao/38734-cadastro-nacional-de-enderecos-para-fins-estatisticos.html.

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2024a). *Censo Demográfico 2022 - Coordenadas geográficas dos endereços no Censo Demográfico 2022 : nota metodológica n. 01.* Rio de Janeiro: IBGE. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102063

Instituto Brasileiro de Geografia e Estatística [IBGE]. (2024b). *Censo Demográfico 2022 - Cadastro Nacional de Endereços para Fins Estatísticos: Nota metodológica n. 04.* Rio de Janeiro: IBGE. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102091

Instituto Nacional de Pesquisas Espaciais [INPE]. (2023). *Programa de Monitoramento da Amazônia e Demais Biomas [PRODES]: Desmatamento anual [Dataset]*. https://terrabrasilis.dpi.inpe.br/downloads/

Instituto Nacional de Pesquisas Espaciais [INPE] & Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. (2020). *Uso e cobertura da terra do Bioma Amazônia - TerraClass [Data set].* http://terrabrasilis.dpi.inpe.br/downloads/

Kuffer, M., Owusu, M., Oliveira, L., Sliuzas, R. & van Rijn, F. (2022). The missing millions in maps: Exploring causes of uncertainties in global gridded population datasets. *ISPRS International Journal of Geo-Information*, 11(7), 403. https://doi.org/10.3390/ijgi11070403

Lei, Z., Xie, Y., Cheng, P. & Yang, H. (2023). From auxiliary data to research prospects, a review of gridded population mapping. *Transactions in GIS*, 27(1), 3–39. https://doi.org/10.1111/tgis.13020

Leyk, S., Uhl, J. H., Balk, D. & Jones, B. (2018). Assessing the accuracy of multi-temporal built-up land layers across rural-urban trajectories in the United States. Remote Sensing of Environment, 204, 898-917. https://doi.org/10.1016/j.rse.2017.08.035

Leyk, S., Gaughan, A. E., Adamo, S. B., de Sherbinin, A., Balk, D., Freire, S., Rose, A., Stevens, F. R., Blankespoor, B., Frye, C., Comenetz, J., Sorichetta, A., MacManus, K., Pistolesi, L., Levy, M., Tatem, A. J. & Pesaresi, M. (2019). The spatial allocation of population: a review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11(3), 1385–1409.

https://doi.org/10.5194/essd-11-1385-2019

Lloyd, C. T., Sorichetta, A. & Tatem, A. J. (2017). High resolution global gridded data for use in population studies. *Scientific Data, 4*(1), 170001. https://doi.org/10.1038/sdata.2017.1

Lu, D. & Weng, Q. (2006). Use of impervious surfaces in urban land-use classification. *Remote Sensing of Environment, 102*(1-2), 146-156. https://doi.org/10.1016/j.rse.2006.02.010

Nagle, N. N., Buttenfield, B. P., Leyk, S. & Spielman, S. (2013). Dasymetric Modeling and Uncertainty. *Annals of the American Association of Geographers*, 104(1), 80-95. https://doi.org/10.1080/00045608.2013.843439

Nobre, C. A., Obregón, G. O., Marengo, J. A., Fu, R. & Poveda, G. (2009). Characteristics of Amazonian Climate: Main Features. In: Keller, M., Bustamante, M., Gash, J. & Silva Dias, P. (Ed.) *Amazonia and Global Change*. American Geophysical Union. https://doi.org/10.1029/2009GM000903

Openshaw, S. (1984). *The modifiable areal unit problem.* Norwich: Geo Books.

Renner, K., Schneiderbauer, S., Pruß, F., Kofler, C., Martin, D. & Cockings, S. (2018). Spatio-temporal population modelling as improved exposure information for risk assessments tested in the Autonomous Province of Bolzano. *International Journal of Disaster Risk Reduction, 27*, 470–479. https://doi.org/10.1016/j.ijdrr.2017.11.011

Schiavina, M., Melchiorri, M., Pesaresi, M., Politis, P., Carneiro Freire, S. M., Maffenini, L., Florio, P., Ehrlich, D., Goch, K., Carioli, A., Uhl, J., Tommasi, P. & Kemper, T. (2023). *GHSL data package 2023* (JRC133256). Publications Office of the European Union. https://doi.org/10.2760/098587

Silva, D. (2023). *VaLin-Pop: uma nova grade populacional para a região metropolitana do Vale do Paraíba e Litoral Norte, SP.* [Dissertação de Mestrado, Instituto Nacional de Pesquisas Espaciais] Repositório digital de teses e dissertações do INPE. http://mtc-m21d.sid.inpe.br/col/sid.inpe.br/mtc-m21d/2023/10.19.20.37/doc/thisInformationItemHomePage.html

Smith, S. K. & Morrison, P. A. (2005). Small-Area and Business Demography. In: Poston, D. L. & Micklin, M. (Orgs.) *Handbook of Population*. Handbooks of Sociology and Social Research. Springer, Boston, MA. https://doi.org/10.1007/0-387-23106-4_26

Souza, A. R., Escada, M. I. S., Marujo, R. de F. B. & Monteiro, A. M. V. (2019). Cartografia do Invisível: Revelando a Agricultura de Pequena Escala com Imagens Rapideye na Região do Baixo Tocantins, PA. *Revista Do Departamento De Geografia*, *38*, 137-153. https://doi.org/10.11606/rdg.v38i1.151603

Souza, A. R., Adorno, B. V., Gonçalves, G. C., Bragion, G. da R., Oliveira, K. D., Escada, M. I. S., Reis, M. S., Sant'Anna, S. J. S. & Amaral, S. (2021). *Paisagens e uso da terra em núcleos populacionais e estabelecimentos rurais da região do Baixo Tocantins - Pará*. São José dos Campos-SP: INPE. 81 p. Relatório Técnico de Atividade de Campo de 2018 e 2019 dos Cursos de Pós-Graduação em Sensoriamento Remoto e em Ciências do Sistema Terrestre do INPE, em parceria com a Universidade Federal do Pará. http://mtc-m21d.sid.inpe.br/col/sid.inpe.br/mtc-m21d/2021/06.18.16.50/doc/thisInformationItemHomePage.html.

Souza, A. R., Amaral, S., Dal'Asta, A. P., Escada, M. I. S. & Monteiro, A. M. V. (2023) Modelos dasimétricos para desagregar dados do Censo Agropecuário. In: *Anais do XX Simpósio Brasileiro de Sensoriamento Remoto*. São José dos Campos : INPE. http://marte2.sid.inpe.br/col/sid.inpe.br/marte2/2023/05.16.12.12/doc/156100.pdf

Souza, A. R. (2024). *Os sistemas do açaí na paisagem amazônica: elementos para análise da economia do açaí da região do Baixo Tocantins-PA.* [Tese de Doutorado, Instituto Nacional de Pesquisas Espaciais]. Repositório digital de teses e dissertações do INPE. https://mtc-m21d.sid.inpe.br/col/sid.inpe.br/mtc-m21d/2024/09.03.22.32/doc/publicacao.pdf.

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., Kilic, T., Murray, S., Blankespoor, B., Prydz, E. B. & Dang, H. H. (2017). *Mapping the world population one building at a time*. arXiv. https://doi.org/10.48550/arXiv.1712.05839.

United Nations [UN] (2015). *UN World Population Prospects: Revision 11 [Data Booklet]*.

https://www.un.org/en/development/desa/publications/world-population-prospects-2015-revision.html.

Weng, Q. & Hu, X. (2008). Medium Spatial Resolution Satellite Imagery for Estimating and Mapping Urban Impervious Surfaces Using LSMA and ANN. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2397-2406. https://doi.org/10.1109/TGRS.2008.917601

Wickham, J. D., Stehman, S. V., Gass, L., Dewitz, J., Fry, J. A., & Wade, T. G. (2013). Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment, 130*, 294–304. https://doi.org/10.1016/j.rse.2012.12.001

WorldPop. (n.d.). *WorldPop methods.* University of Southampton. Retrieved [June 2024], from https://www.worldpop.org/methods/

Xu, Y., Ho, H. C., Knudby, A. & He, M. (2021). Comparative assessment of gridded population data sets for complex topography: A study of Southwest China. *Population and Environment*, 42, 360–378. https://doi.org/10.1007/s11111-020-00366-2

Yin, X., Li, P., Feng, Z., Yang, Y., You, Z. & Xiao, C. (2021). Which Gridded Population Data Product Is Better? Evidences from Mainland Southeast Asia (MSEA). *ISPRS International Journal of Geo-Information*, 10(10), 681. https://doi.org/10.3390/ijgi10100681

Zha, Y., Gao, J. & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing, 24*(3), 583-594. https://doi.org/10.1080/01431160304987

Zhao, T. & Wang, W. (2023). Coordination Dynamics between Population Change and Built-Up Land Expansion in Mainland China during 2000–2020. *Sustainability*, *15*(22), 16059. https://doi.org/10.3390/su152216059

Zhou, Y., Ma, M., Shi, K. & Peng, Z. (2020). Estimating and Interpreting Fine-Scale Gridded Population Using Random Forest Regression and Multisource Data. *ISPRS International Journal of Geo-Information, 9*(6), 369. https://doi.org/10.3390/ijgi9060369

Main Author Biography

Gustavo Piva Lopes Salgado was born in 1990 in the city of Campinas, São Paulo. He holds a bachelor's degree in Geography from the University of Campinas (UNICAMP -2020), a specialist in Geoprocessing from the SENAC University Center of São Paulo (2023) and is a master's candidate in Remote Sensing from the National Institute for Space Research (INPE) in São José dos Campos/SP. He currently works as an interpreter of satellite images in the Annual Monitoring Project for the Suppression of Brazilian Native Vegetation (PRODES) in INPE. He develops process automation, selection and visual analysis of images, data validation, as well as performing spatial data analysis.