TecnoLógicas

Populated Anthromes: Construction and Validation of an Analytical Model Applying Metrological Principles in Geocomputation and Socio-Ecological Models

Antromas poblados: construcción y validación de un modelo analítico que aplica principios metrológicos en geocomputación y modelos socioecológicos

Maximiliano S. L. A. Gobbo[1]; Thiago de Oliveira Araujo[1]; Claúdia de Oliveira Faria Salema[1]

[1]National Institute of Metrology, Quality and Technology, Rio de Janeiro - Brazil

Correspondence: maxgobbo91@gmail.com

Received: 08 September 2025 **Accepted:** 26 August 2026 **Available:** 31 August 2026

How to cite / Cómo citar

M. S. L. A. Gobbo, T. de Oliveira Araujo, and C. de Oliveira Faria Salema, "Populated Anthromes: Construction and Validation of an Analytical Model Applying Metrological Principles in Geocomputation and Socio-Ecological Models," *TecnoLógicas*, vol. 29, no. 67, e3672, 2026.
https://doi.org/10.22430/22565337.3672

CC BY NC SA

Abstract

Aligned with the demand for robust environmental analytical models capable of analyzing large volumes of data and providing efficient and comprehensive results for the development of environmental and social public policies, this investigation aimed to structure a scalable analytical model of populated anthromes, focusing on socio-environmental management. Following the structure of the global model, the methodology for constructing and validating the model involved the stages of exploratory analysis, data mining, joining, and plotting of demographic data, aiming to characterize the data applied in regional modeling. Based on this characterization, the spatialization capacity of the analytical model on cartographic maps was evaluated following the methodological guidelines of Geocomputation. Finally, the analytical model was statistically and metrologically analyzed using the national gold standard. The results demonstrated that the analytical model developed in R achieves the same level of precision as the national gold standard. The metrological indices demonstrated the quality and efficiency of the modeling in the distribution and spatialization of demographic data. Furthermore, it was observed that the level of approximation to local reality strongly contributes to the characterization of local spatialities, improving the overall analytical model and significantly characterizing the reality of the territory. Thus, it is concluded that the analytical structure for the construction and validation of the populated anthromes model ensures the quality of the information generated by the modeling and contributes to scalable implementation across the national territory. Additionally, it was verified that the application of metrological principles as a mechanism to guarantee the validity of results and geospatial models is possible and adds value to the technological products.

Keywords

Analytical model, demographic data, populated anthromes, regional model, socio-ecological model.

Resumen

En consonancia con la demanda de modelos analíticos ambientales robustos, capaces de analizar grandes volúmenes de datos y proporcionar resultados eficientes e integrales para el desarrollo de políticas públicas ambientales y sociales, esta investigación tuvo como objetivo estructurar un modelo analítico escalable de poblaciones antropogénicas, centrándose en la gestión socioambiental. Siguiendo la estructura del modelo global, la metodología para su construcción y validación abarcó las etapas de análisis exploratorio, minería de datos, integración y representación gráfica de datos demográficos, con el fin de caracterizar los datos aplicados en el modelado regional. Con base en esta caracterización, se evaluó la capacidad de espacialización del modelo analítico en mapas cartográficos, siguiendo las directrices metodológicas de la geocomputación. Finalmente, el modelo analítico se analizó estadística y metrológicamente utilizando el estándar nacional de referencia. Los resultados demostraron que el modelo analítico desarrollado en R cumple con el mismo nivel de precisión que el estándar nacional de referencia. Los índices metrológicos demostraron la calidad y eficiencia del modelado en la distribución y espacialización de los datos demográficos. Además, se observó que el grado de aproximación a la realidad local contribuye significativamente a la caracterización de las espacialidades locales, mejorando el modelo analítico general y caracterizando de manera sustancial la realidad del territorio. Por lo tanto, se concluye que la estructura analítica para la construcción y validación del modelo antropológico poblacional garantiza la calidad de la información generada por el modelado y contribuye a una implementación escalable en todo el territorio nacional. Adicionalmente, se verificó que la aplicación de principios metrológicos como mecanismo para garantizar la validez de los resultados y los modelos geoespaciales es posible y agrega valor a los productos tecnológicos.

Palabras clave

Modelo analítico, datos demográficos, antromas poblados, modelo regional, modelo socioecológico.

1. INTRODUCTION

The environmental and climate emergencies of the 21st century are driving us towards the development of instruments, tools, policies, and agreements aimed at the survival of human and non-human species and planetary sustainability [1], [2], [3]. Different areas of knowledge have connected to producing cross-cutting methods of analysis, monitoring, and control assisting decision-makers, public and private companies, and the global community in the endeavor of planetary resilience [4], [5]. Within this horizon, we understand that socio-ecological models assume an investigative character regarding the scenarios that are constructed, deconstructed, modified, and re-signified in contemporary times [6], [7], contributing significantly to the development of public policies and international agreements and to the achievement of global sustainable development goals [8], [9], [10], [11].

Considering this, we sought in different socio-ecological models an instrument that would articulate the anthropic factor with the environmental factor, generating an analytical model that could provide pathways for the development of public policies, for territorial and population analyses. Furthermore, it was expected that the model could provide portraits of the socio-environmental state of the Brazilian territory. During the investigation, we came across the anthropogenic biome or anthromes model, whose applicability is directly linked to these premises [12], [13]. The anthromes model (Figure 1) basically uses two geospatial information for its construction, namely population and land use and cover [14], [15].

During the analysis of the global model, we verified that the mapping was produced using data provided by global databases [12], [16], which do not fully describe local spatialities [17], [18], [19] and lack important details for socio-environmental management at the regional scale [20], [21]. From this perspective, the literature suggests that the regionalization of global analytical models can provide more robust ways to characterize spatialities [22], [23]. Complementarily, the application of local data allows the recognition of particularities that provide relevant subsidies for the development of socio-environmental public policies aligned with national premises [24], [25].

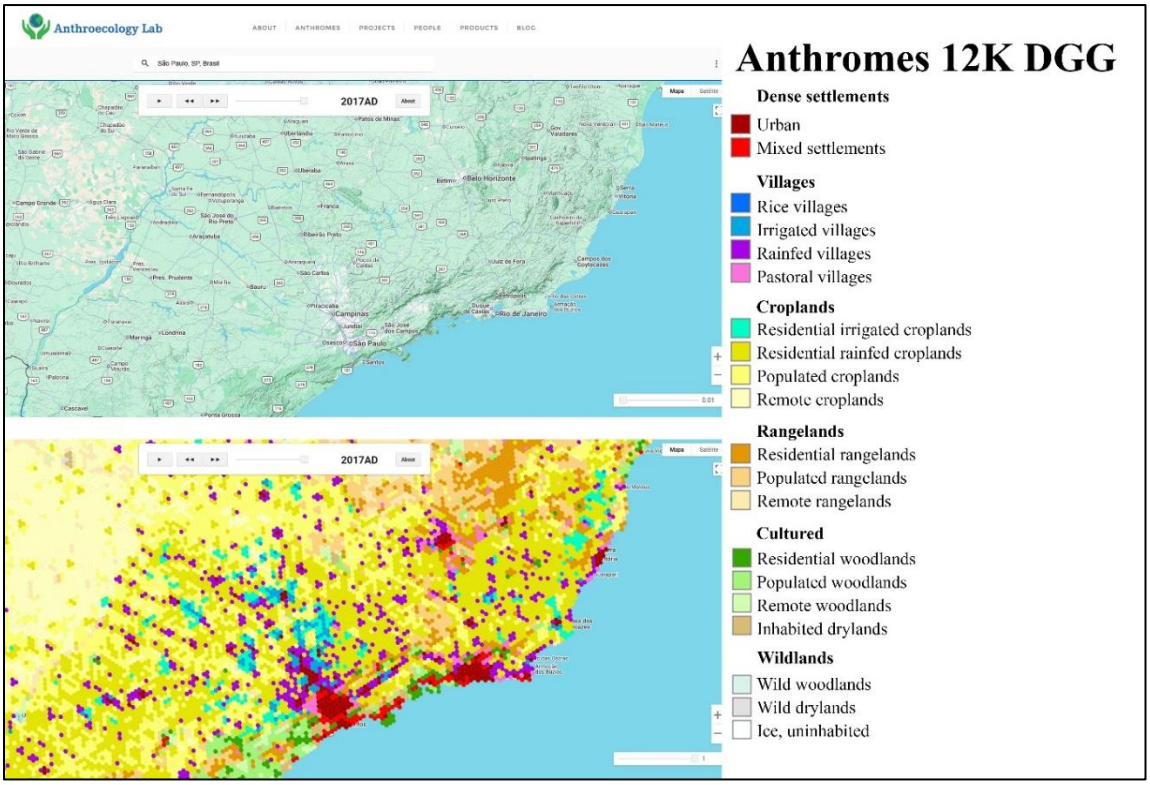


Figure 1. Visualization of the analysis area (pilot study area) through the Anthroecology Lab's global anthromes digital platform. Source: adapted from [26].

Given these guidelines for the regionalization of global analytical models, in previous works we analyzed data produced by the Brazilian Institute of Geography and Statistics (IBGE – Brazil) as sources for the construction of the regional model of anthromes [27], [28]. In these works, we applied the metrological principles of method and tool validation to guarantee the validity of results established in the ISO/IEC 17025 standard [29] as instruments to evaluate the adequacy of the data to the model to be constructed. The results demonstrated that the degree of reliability, precision, and analytical capacity of the methods and results were adequate for the development of the regional model of anthromes.

From this perspective, we proceeded to apply this data to the construction of the regional anthropomorphic model. We fragmented the construction into three analytical stages. The first stage, reported in this manuscript, had as its central objectives the analysis, characterization, mapping, and validation of Brazilian population data in the R software as an input source for analytical modeling at a regional scale, demonstrating our alignment with the latest global mapping of anthropomorphic sites [16]. The second stage involves a similar analysis with land use and land cover data, while the third deals with the production of the regional anthropomorphic model. These two stages are not reported in this research, but the code and products are available on GitHub [30] and RPubs [31], respectively.

Understanding the significant relevance of population data for the construction of the regional anthropomorphic model [16], [32], [33], [34], we structured an analytical model to verify the adequacy of Brazilian data to the guidelines for spatialization and characterization of the distribution of population data and the precision and uncertainty in the cartographic products generated. To this end, following the methodological guidelines of the global model [12], [16] and the technical bases of Geocomputation [35], [36], [37], we developed a specific analytical model for population data, in which, similarly to population distribution models, we attempted to verify whether the data met the assumptions for the construction of the regional anthropomorphic model. Additionally, using metrological principles for validation and guaranteeing the validity of

results [29], [38], we evaluated the analytical model, comparing it to the national gold standard, in order to verify whether the analytical model for populated anthromes met the structural requirements of the national population model [39], [40] and portrayed the population clusters synergistically with the global model [12], [16].

Based on these analyses, we would be able to verify whether the regional model would substantially describe the local reality, efficiently spatialize the geospatial information of Brazilian spatiality and characterize the distribution of the population in space. In this endeavor, we carried out different analytical stages, from the collection and processing of census data, through plotting and constructing the maps, to the evaluation of the cartographic products, in order to have clarity on the efficiency of the analytical model of populated anthromes as an instrument for spatialization and for validating the input data in the construction of regional anthromes models.

The scripts related to each of the steps are available on GitHub [29] and are published on RPubs [30] for better visualization of the cartographic products. Here we report our analyses in a synthetic and linear way. Therefore, we suggest that the reader read the codes and view the cartographic products for a better understanding of the work. It is also worth noting that, as it is a scalable model, the analytical structure presented in the manuscript can be adjusted to other demographic datasets, serving as an instrument for evaluating and validating this geospatial information on a local scale.

2. METODOLOGY

2.1 Data presentation

The data analyzed in this research are census information from the Brazilian Demographic Census, conducted in 2010 by the Brazilian Institute of Geography and Statistics (IBGE). Until the completion of this research, the data from the last national demographic census, conducted in 2022, had not been fully published on the national platform. Therefore, we justify the use of this 2010 data; however, we indicate that the analytical guidelines can be adjusted to other dates and other global territories.

From the census data, we used vector and tabular data extracted from ".xls" files, both from Microsoft Access 365 and Microsoft Excel 365 files from the IBGE collection. In addition, we used the shapefile of the State of São Paulo (Brazil), model area, for the stages of mapping construction and uncertainty study and validation of cartographic products. We chose the State of São Paulo as a model area considering its political, social, economic, and cultural representativeness throughout the national territory [41], in addition to being the Federative Unit with the highest population concentration [42] and having different social strata distributed throughout the territory [43]. We reiterate that all data are available on GitHub for data validation and replication of analytical routines [30].

2.2 Analytical methodology

We structured the methodology of this research based on the analytical guidelines of Geocomputation and Geoprocessing in R [35], [36], [44], with the aim of structuring a replicable, reproducible, and scalable analytical model for populated anthromes. In this perspective, we employed the following analytical roadmap: 1. Data presentation; 2. Exploratory analysis; 3. Data mining; 4. Data joining; 5. Data plotting; 6. Static mapping; 7. Interactive mapping; 8. Validation and Uncertainty studies of the model (Figure 2). For each of the steps, we employed specific R functions and packages that optimized our analyses and the construction of the cartographic product, consequently. The focal point of this roadmap was to integrate metrological premises, such as guaranteeing the validity of results and validating methods [29], into analytical structures of Geocomputation and socioecological modeling, a unique tool of Human Ecology.

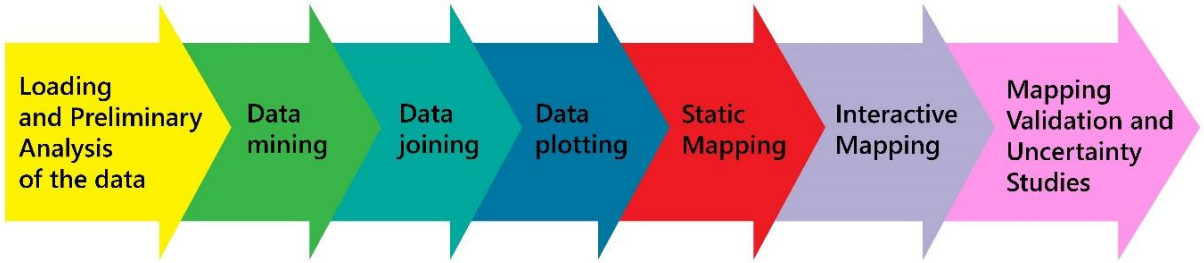


Figure 2. Visual summary of the analytical stages. Source: own elaboration.

2.3 Data loading and exploratory analysis

The data copied from the IBGE platform were stored on the work computer before loading into the R software. We considered only 3 sets of vector data: population01, population02, and population03. These sets contained the following information: the geographic location of the municipalities in the State of São Paulo, the population in each of the subdivisions established in the census, the area and/or population density of each municipality. The vector data were loaded into R using the read_excel() function, and we performed exploratory data analysis [25], using the functions summarized in Appendix 1.

In contrast, we used the shapefile of the localities of all Brazilian census sectors (national document), which was captured from the IBGE database. Therefore, we filtered this data in Microsoft Access 365 (dataset format that structures the shapefile), keeping only the data from the State of São Paulo, which were exported to a Microsoft Excel 365 tabular file (localsp.xls). After filtering, we loaded and performed exploratory analysis of the shapefile using the functions described in Appendix 2, which were extracted from the relevant literature for this scope [35], [36], [44].

2.4 Data mining

This stage of analytical modeling involved, firstly, fragmenting the dataset from the shapefile. The fragmentation of this set into subsets followed the guidelines of the census operation technical manuals [39], [45], that is, we used the different categories of census sectors as guides for the division into subsets. These typologies were present in the NM_CATEGORIA attribute of localsp. The function applied to create the subcategories was the filter() function.

Based on the subcategories created, we evaluated their alignment with the vector datasets population01, population02, and population03. To compare the attributes of the vector data and the subsets of localsp, we applied the names() function, which reports the names of the attributes present in the dataframes. After evaluating the synergy between the vector datasets and the data from localsp, we identified which one was most correlated with the shapefile subsets. With this, we performed the fragmentation of the vector set most synergistically with the shapefile data, following the same orientation of subsets according to the typology of the census sector [39], [45].

2.5 Data merging

In this step, we joined the data from the subsets created from the vector data and from the shapefile data, using the census tract typologies as a reference. For this operation, we followed the analytical guidelines expressed in the literature [35], 36], [37], [46], and Appendix 3 reports the functions used in this phase of the analysis. We should note beforehand that some adjustments to the subsets were necessary, as we will demonstrate in the results and discussion.

2.6 Data plotting

For plotting data on the Cartesian plane, we followed the colors established by Ellis and Ramankutty [15] and which were maintained in the updates of the global anthromes mappings [12], [16]. According to these authors, shades of red would represent populations in urban areas, while earthy tones (orange and brown) would reflect populations in rural areas. Considering the uniqueness of traditional Brazilian populations [21], [47], [48], indigenous and quilombola, we used the shade of blue to represent these populations on the Cartesian plane. Table 1 details the RGB codes for each of the categories of populated anthromes in the territory. The information in Table 1 was the source of information for the creation of the dataframe legend_populatedanthromes, which guided the creation of static and interactive mappings.

Table 1. Territorial typologies defined for anthromes in the State of São Paulo, along with their corresponding RGB color codes. Source: own elaboration.

Brazilian Populated Anthromes	RGB Code
cities	#FF0000
isolatedurbanareas	#FF4747
urbanvillages	#F66969
traditionalcommunities	#9CC2E5
ruralvillage	#ED833B
ruralcore	#DF9B6D
ruralsettlement	#FFD966
settlementproject	#968551

Based on the latitude and longitude information that comprised the 8 typologies of populated anthromes, we plotted the data on the Cartesian plane using the plot() function, as recommended in the literature [12], [16]. We emphasize that the plots were performed individually and collectively, that is, first we plotted each of the 8 populated anthromes and then plotted the 8 sets on a single Cartesian plane, creating a complete image of the distribution of populated anthromes.

2.7 Static mapping

Following the methodological guidelines established in the literature on R [35], [49], we proceeded to create a static map of the populated anthromes of São Paulo. In this analytical stage, we used the shapefile of the polygonal contours of the municipalities of the State of São Paulo (sp_countries.shp), extracted from the IBGE database. We justify the use of this shapefile by the free availability of the product, enabling replicable and reproducible research. Starting from the shapefile, we filtered two attributes of the data to keep only the structures of interest, namely: NM_MUNICIP (name of the municipality) and geometry.

With the filtering, we created the cities_shape set, which contained only the names of the municipalities and their respective geometries. To create the map, we used the ggplot2 package, whose main function was ggplot(), and we combined functions to improve the visual pattern and meet our demands. Examples of complementary functions that integrate the code for creating static mapping include: geom_sf(), geom_point(), labs(), and scale_color_manual().

2.8 Interactive mapping

As a basis for interactive mapping, we used the Open Street Maps (OSM) digital cartographic base, a global collaborative and free mapping project, which can be used by any user and researcher around the world, which is directly aligned with the premises of replication and reproduction of codes and the model for mapping populated anthromes. Regarding the functions applied in this analytical procedure, we used the leaflet() package and the editable resources associated with it, and we will highlight the most relevant ones [36], [46].

Firstly, the addMarkers() function helped in marking the points where data were found in the populated_anthromes set, using the latitude and longitude attributes for plotting. Additionally, the addRectangles() function helped in defining the position and dimension of the squares used to mark the populated anthromes in the State. In the color attribute, we applied the legend_populatedanthromes dataframe to guide the coloring of the squares superimposed on the OSM.

After constructing the map of populated anthromes in the State of São Paulo, we measured the areas covered by the squares representing the anthromes. In this operation, we used the average of the latitude and longitude variables from the `populated_anthromes` dataset as the basis for the calculations. For the width (1) and height (2) calculations of the squares, we applied the first two formulas presented below. Based on the width and height values, we calculated the area of the squares (3) that were represented on the OSM base.

$$\text{Width (longitude)} = 0.03 \times 111.32 \times \cos(\text{mean latitude}) \quad (1)$$

$$\text{Height (latitude)} = 0.03 \times 111.32 \times \text{sen}(\text{mean latitude}) \quad (2)$$

$$\text{Square Area} = \text{Width} \times \text{Height} \quad (3)$$

2.9 Studying validation and uncertainty

After creating static and interactive maps, we carried out certain procedures to evaluate and attest to the quality of the regional model of anthromes [50], [51], [52], [53], [54]. Following the investigative guidelines established in the literature [35], [37], [46], we listed studies to confirm the spatialization of geospatial information (distribution of population data), to evaluate the uncertainty and error associated with the mapping from the perspective of Earth Sciences and Metrology.

For the validation and uncertainty stages of the generated products, it was necessary to establish a gold standard for comparisons [35], [46]. Considering the raw data from IBGE as our national gold standard, we filtered the data from br_locations_2010 (shapefile) using the filter() function, keeping only the data from São Paulo, which we structured with the saopaulo set (gold standard). During the following steps, we used the total number of points (2143) to ensure the highest level of precision and accuracy and reduce mapping errors, aligning the analysis with metrological principles [29], [38]. Furthermore, we defined the geographic coordinate system (CRS) 4326 for the gold standard and for populated_anthromes, our comparator and object of analysis.

2.10 Examination of data properties

To ensure data quality and mapping accuracy [55], we performed tests on the properties of the populated_anthromes and saopaulo sets [49]. First, we noted that both sets had the same geographic coordinate system (CRS) and the same spatial dimensions of latitude and longitude in their structures. From this perspective, we verified the minimum and maximum latitude and longitude limits (coordinates) of the sets, with the aim of proving that both sets

represented the same territorial area in the mapping [35]. This verification was performed using the range() function, which returns the referenced geographic limits.

2.11 Summary statistics

To perform statistical analyses of cartographic products, we structured a statistical grid of 400 cells (20 by 20) over the territory of São Paulo, following practices already consolidated in this area of investigation [36], [37]. We used the saopaulo dataset (raw data from IBGE) to construct the grid. We converted this dataset into a spatial dataset, which was named saopaulo_spatial. This dataset consisted of the minimum and maximum values of X and Y, representing longitude and latitude, respectively. Subtracting the minimum value from the maximum, for both X and Y, we arrived at the lateral dimensions of the grid. With these values established, we divided them by 20 and arrived at the values of size_cell_x and size_cell_y, representing the width (longitude) and height (latitude) of the square of the statistical grid. We note that these calculations considered the sphericity of the Earth, since the formulas used are derivations of the Haversine Formula [49]. The overall structure of the Haversine formula (4) is represented below.

$$a = sin^2 \times \left(\frac{\Delta lat}{2}\right) + \cos(lat_1) \times \cos(lat_2) \times sin^2\left(\frac{\Delta lon}{2}\right)$$
$$c = 2 \times atan2\left(\sqrt{a}, \sqrt{1-a}\right) \quad (4)$$
$$d = R \times c$$

Where: Δlat is the difference in latitude between the two points, Δlon is the difference in longitude between the two points, lat1 and lat2 are the latitudes of the two points in radians, R is the radius of the sphere (for example, the average radius of the Earth).

Through the calculations, we dimensioned the squares of the statistical grid. The grid values were presented in Table 2.

Table 2. Values obtained during the calculation of height, width, and area for the 20 × 20 (400-cell) statistical grid. Source: own elaboration.

Dimensions	Size
Width (cell_size_x_km)	45.57963 km
Height (cell_size_y_km_	11.34617 km
Area of each quadrant of the grid in km²	517.1542 km²
Average dimension of the slides of the grid square in km	22.74102 km

With the dimensions defined for the quadrants (cells) of the grid, we structured the statistical grid by applying the raster() function, creating the quadrant_grid object for the calculations that would follow. The objective was to evaluate the distribution behavior of the points of populated_anthromes (object of analysis) when compared to that of saopaulo (gold standard) [35], [49].

After establishing the grid, we counted the points in each grid cell that covered both sets using the rasterize() function. After counting the points in each cell, we structured two dataframes with the number of points in each of the cells of the two sets (saopaulo and populated_anthromes). The point count data was also useful for the next analytical steps.

2.12 Confusion matrix structuring for statistical analyses

First, we combined the two datasets associated with point counting (saopaulo_countpoints and anthromes_countpoints). The purpose of combining the two datasets was to structure the confusion matrix for statistical analyses [36], [56]. In it, we aligned the point numbers in each cell of the statistical grid (20x20) of the two datasets, to establish the following relationships:

- True Positives (TPs): quadrants of the statistical grid where the number of points was equal between saopaulo_countpoints and anthromes_countpoints;
- True Negatives (TNs): quadrants of the statistical grid where there were no points represented in either saopaulo_countpoints or anthromes_countpoints;
- False Negatives (FNs): quadrants of the statistical grid where the number of points in saopaulo_countpoints was greater than the number of points in anthromes_countpoints;
- False Positives (FPs): quadrants of the statistical grid where the number of points in saopaulo_countpoints was less than the number of points in anthromes_countpoints.

To structure these relationships, we use the if()/else() function. If the number of points were equal and different from 0 in the two sets, we determine if the values were equal (TP) or different (TN), greater (FN) or less (FP) in saopaulo_countpoints when compared to anthromes_countpoints. If the number of points were the same in both sets and different from 0, the TP column would receive the value of 1 and the TN, FP, and FN columns the value of 0.

If the number of points were equal to 0 in both sets, the TN column would receive the value of 1 and TP, FN, and FP the value of 0. Additionally, if the number of points in the grid quadrant was greater in saopaulo_countpoints than in anthromes_countpoints, the FN column would also receive the value of 1 and the TP, TN, and FP columns the value of 0. On the other hand, if the number of points was less in saopaulo_countpoints than in anthromes_countpoints, the FP column would receive the value of 1 and the TP, TN, and FN columns the value of 0.

We note that in the statistical grid there were areas that did not represent the data from São Paulo, that is, lines where “NA” appeared. These lines were replaced by 0 in the combined set using the function [is.na(combined_set)] <- 0, to allow statistical calculations from the confusion matrix, namely: sensitivity, specificity, accuracy, and overall mapping error.

2.13 Sensitivity, specificity, accuracy, and error of map

Based on the confusion matrix, we proceeded to the sensitivity, specificity, accuracy, and overall error analyses, where we integrated the analytical precepts of Environmental and Earth Sciences [35], [36], [46] with metrological fundamentals to guarantee the validity and quality of the results [28], [29], [38]. Firstly, the calculation of the sensitivity of the populated mantle analytical model aimed to evaluate the model's ability to identify true positives (TPs) [57], [58], that is, to assess whether the populated mantle analytical model could efficiently identify the points present in the pilot area (gold standard). The estimation of this metric was made using (5).

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

Where TP represents true positives and FN represents false positives. Continuing with the statistical analyses, we move on to the analysis of the specificity of the analytical model. This metric refers to the model's ability to identify mapped points that were not part of the comparison standard (gold standard) [46], [49]. From this perspective, the specificity analysis aimed to identify whether the analytical model of populated anthromes (populated_anthromes) was able to point out which points were not included in the same cell of the statistical grid as the comparator (saopaulo).

Through the confusion matrix, we verified the proportion of points identified as True Negatives (TNs) in relation to the number of False Positives (FPs). In other words, we analyzed the proportion between the quadrants in which the number of points of the gold standard was lower when compared to the quadrants generated by the analytical model of populated anthromes (FPs). The formula used in the specificity estimate is represented by (6).

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

Where TN represents the number of quadrants identified as True Negatives and FP the number of False Positives.

Moving on to statistical analyses, we delve into the overall accuracy metric of the populated anthromes analytical model. According to the literature [36], [46], this metric aims to identify the proportion of points that were correctly identified by the populated anthromes analytical model when compared to the gold standard. In other words, this metric estimates the number of True Positives (TPs) and True Negatives (TNs) relative to the total number of cells in the grid. The overall accuracy calculation is done using the formula (7):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

Where TP + TN, which represented the grid cells in which the number of points was equal in the statistical grid generated by the populated anthromes model and in the gold standard, was divided by the total number of cells in the statistical grid, that is, by the sum of the cells with equal (TPs and TNs) and different (FPs and FNs) number of points between the sets.

Sequentially, we move on to calculating the overall error, which, according to the literature [36], [46], indicates the proportion of areas incorrectly classified by the model. In other words, this metric uses the proportion between False Positives (FPs) and False Negatives (FNs) compared to the total number of cells in the statistical grid. The calculation of the overall error is given by the formula (8):

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{8}$$

Where we have the sum of FPs and FNs, areas with a divergent number of points between the quadrants generated through the analytical modeling of populated anthromes and the comparator, divided by the total number of cells with points.

This set of metrics allowed the analysis of the quality of geographic information produced by the populated anthromes model and allowed the correlation between the validation and error assumptions of Environmental and Earth Sciences and the metrological fundamentals for validation and quality assurance of results.

3. RESULT AND DISCUSSION

Following the analytical sequence proposed in the methodology of this research, we present our results and discussion along the same lines of reasoning, allowing the analysis to evolve as the data are correlated. According to the literature, this presentation format facilitates the interpretation and monitoring of the investigative process in geocomputation and geoprocessing in R [35], [36], [44]. Thus, we invite the reader to follow the path to the construction of the analytical model of populated anthromes in R.

3.1 Exploratory analysis

The exploratory analysis of the vector datasets (population01, population02, and population03) revealed important features that were essential for continuing our analyses and, consequently, for the construction and validation of the analytical model of populated anthromes. The exploratory analysis showed that population01 contained attributes referring to the typologies of census tracts. For each attribute, the number of inhabitants in each typology of a particular municipality was presented. On the other hand, the results indicated that population02 contained the same attributes plus two others that complemented the data: "population density" and "total area". In contrast, the analytical results indicated that population03 contained other numerical attributes: "relative total (%)" and "relative in the municipal seat", representing statistical proportions of the population of São Paulo and not population dimensions.

Through analytical results, we found that these three sets (population01, population02, and population03) were structured in tabular format (rows and columns). The columns detailed the quantitative or qualitative values associated with each attribute (first row of the table) for each municipality (rows). From the quantitative evaluation of the sets, we identified that population01 was composed of 645 rows and 10 columns, while population02 was structured in 645 rows and 8 columns, and population03 in 645 rows and 10 columns.

In contrast, the exploratory analysis of the shapefile (localsp.xls), named localsp, revealed that the dataframe consisted of 2.142 rows and 21 variables distributed in columns, which represented spatial attributes such as latitude, longitude, altitude, municipality names, among others. Additionally, the analytical results demonstrated that localsp had integrated identifier attributes, such as ID and CD_GEOCODIGO. According to the technical documentation of the census operation [39], [45], the IDs were associated with the different land use and land cover typologies of the same municipality and, therefore, varied for a single city. On the other hand, CD_GEOCODIGO referred to the geographic area identifier code, which was established by the IBGE to refer to a specific polygon (census tract) in Brazilian territory.

Comparatively, we observed, through the analytical results, that the vector data consisted of 645 lines, while the shapefile consisted of 2.142 lines. Analyzing these values, we verified that this substantial difference was linked to the level of aggregation of the vector data when compared to the shapefile [39], [45]. From this perspective, we understand that localsp detailed the territorial grid of the State of São Paulo in a more robust way, fractioning the territorial portions at the microdata level, while population01, population02, and population03 aggregated the data by municipalities. With the results obtained in the exploratory analysis and the preliminary conclusions about the analyzed sets, we obtained the necessary basis to proceed with the joining of the vector data (population01/02/03) to the shapefile (localsp), following the analytical modeling expressed in the literature [12], [16], [44], [46].

3.2 Data mining

During the data mining of the localsp shapefile, we fragmented the dataset into 8 subsets, represented in Table 3. Each of these sets represented a census tract category, both urban and rural, in the State of São Paulo. To confirm the data fragmentation, we plotted the data from the 8 subsets on the Cartesian plane, using the 'latitude' and 'longitude' attributes associated with each of them. In this analytical step, we employed the plot() function, as suggested in the literature for confirming the spatialization of the data [35],[36], [46].

Table 3. Localsp subsets based on census tract category. Source: own elaboration.

Census tract category	Subset formed
Cities	sp_cities
Isolated Urban Areas	sp_isolatedurbanareas
Urban Villages	sp_urbanvillages
Rural Village	sp_ruralvillage
Rural Core	sp_ruralcore
Settlement	sp_settlement
Settlement Project	sp_settlementproject
Tradiconal Communities	sp_traditionalcommunities

Based on the 8 subsets created from localsp, we evaluated the alignment of these categories (Table 3) with the attributes present in the vector datasets population01, population02, and population03. We used the names() function to compare the names of the attributes that made up the datasets. Through the results, we found, based on the attributes that made up these vector dataframes, that the attributes present in population01 were the closest to the subsets of localsp. Assuming these results, we fragmented the population01 data into 8 subsets, following the same logic applied to localsp.

After fragmenting the population01 data, we compared the subsets from localsp and population01. The objective was to verify the synergy of the number of rows present in each of the subsets (Table 4).

Table 4. Comparing the Number of Lines in Population01 Sets and Corresponding LocalSP Sets. Source: own elaboration.

Population Cluster Categories	Data from population01	Data from localsp
Cities	645	645
Isolated Urban Areas	645	985
Urban Villages	645	295
Traditional Communities	645	12
Rural Villages	645	104
Rural Core	645	30
Settlement	645	61
Settlement Project	645	9

Based on the results presented in Table 4, we found that a significant portion of the municipalities in São Paulo did not have certain census typologies, as we verified in the lines of traditional communities, rural core, settlement, and settlement project, and which was revealed through the large numerical divergence in Table 4 [35], [59]. On the other hand, in the case of isolated urban areas, we observed that the number of microdata (localsp) is significantly greater than the vector data (population01). Considering this, we verified that this occurs due to the repetitive presence of isolated urban areas in the same municipality, as suggested by the relevant literature [36], [46]. Therefore, the numerical divergence between the two sets did not invalidate the subsequent stages of our analysis.

3.3 Data merger

Firstly, as we highlighted in the previous item, there was a discrepancy between the number of rows in the localsp and population01 subsets (Table 4). Therefore, it was necessary to make some adjustments to the data before correlating them. The first adjustment concerns

the localsp subsets with a number of rows equal to or less than 645. Regarding the lower numbers, we understand that there was no population described for a given census sector typology or that it had been aggregated into another demographic typology, as suggested in the literature [23], [60]. Concerning the equal number of rows, we understand that all sampling points had their populations integrated.

Another aspect to be noted refers to census typologies where the number of lines in localsp was greater than in population01. The only case present in the sets was observed in isolated urban areas (AUI) (Table 4), a sector that, according to the IBGE technical documentation [40], represents "an area defined by law and separated from the district seat [municipality] by a rural area or by another legal boundary". Based on the extract, we understand that a municipality may have more than one AUI integrated into its urban perimeter. However, the data on the AUIs of the municipalities in São Paulo were aggregated, preventing fragmentation by municipal AUIs.

From this perspective, to enable the construction of the analytical model of populated anthromes, we applied the numerical population value of the municipal AUI present in the population01 subset to all AUIs of the same municipality in the localsp subset. Thus, we indicate here a limitation in this research and in the sample used, overestimating the population numerically in these sectors. However, as a means of enabling the analytical modeling of populated anthromes, this was the chosen path to proceed with data production, highlighting here a research limitation.

After considering the factors, we joined the data based on the functions reported in Appendix 3 (data joining) and following the methodological guidelines expressed in the literature [35]-[37], [46]. We were able to connect the data from the population01 subsets to those from localsp, adding the respective populations (numerical values) to the different census typologies, a fact of paramount importance for consolidating the different typologies of populated anthromes [12], [16]. In this way, we structured 8 new sets, where the census information and the geospatial information of the census sectors that integrated them were located. The sets resulting from the correlation of the subsets were: cities, urbanvillages, isolatedurbanareas, traditionalcommunities, ruralvillage, ruralcore, ruralsettlement and settlementproject.

3.4 Data plotting

After establishing the 8 sets representing the populated anthromes in the State of São Paulo, we proceeded to plotting the data. First, we merged the 8 sets into a single set, which was named populated_anthromes. In this set, we inserted a column for entering the census tract typologies. For plotting the sets on the Cartesian plane, we chose the representation format using the descriptor "pch", defining the value "15", which represents the shape of filled squares.

The colors for filling the squares were entered into the 'colors' dataset, based on the census sector typology and its associated RGB code (Table 1), and defined using the "col" descriptor during the coding of the plot() function in R. Additionally, the "cex" descriptor defined that the square size would be based on the population dimension, using the 'population' attribute of each subset of populated anthromes, meeting the geocomputation principles in this area [35], [36], [46]. The resulting plot is shown in Figure 3.

The analytical result represented in Figure 3 revealed that the plotting of the data from populated_anthromes occurred correctly, allowing the integration of different typologies into a single image, meeting the guidelines established in the literature [12], [14], [16]. Furthermore, we noted that the legend followed the established color scheme for the different territorial categories, aligning with the guidelines for creating plots with geospatial data [36], [46], [49]. Therefore, we confirmed the adequacy of the analytical model of populated anthromes in the distribution of data in territorial spatiality through Figure 3, contributing to the validation of the modeling as an analytical tool.

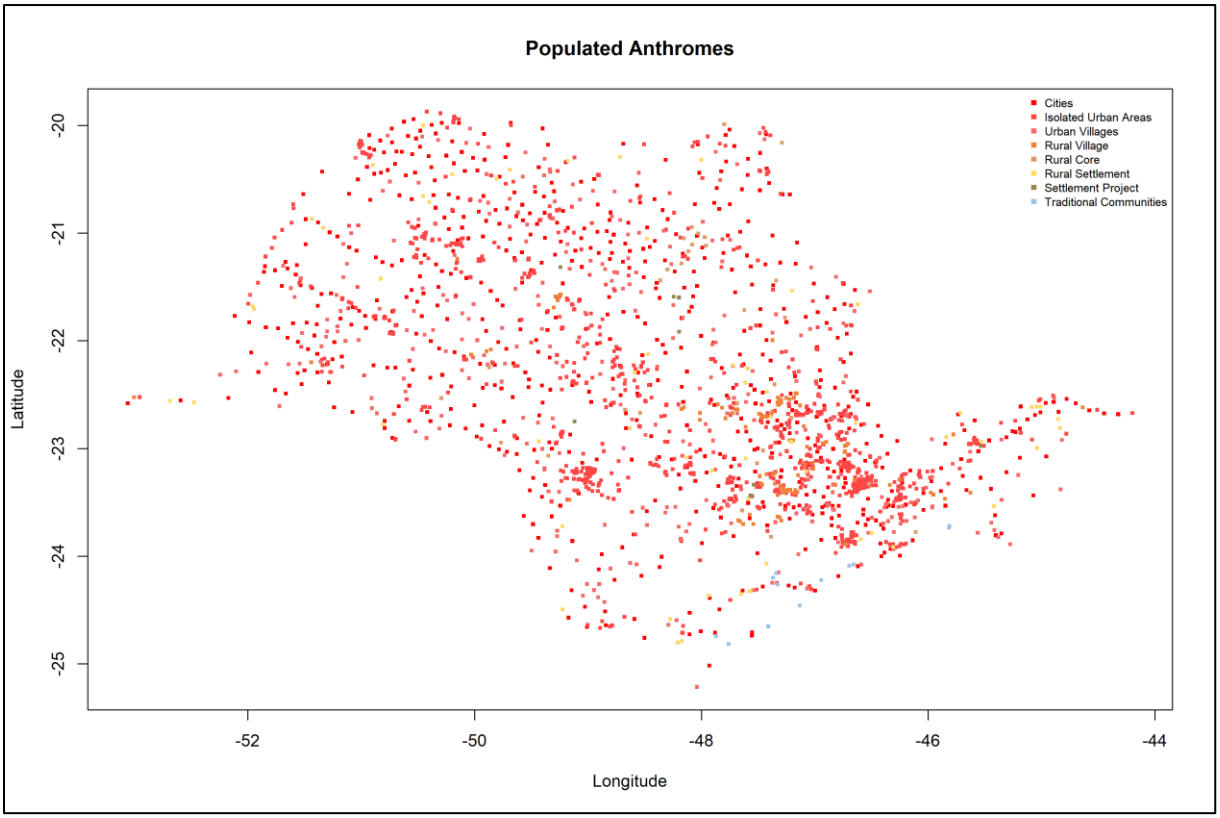


Figure 3. Distribution of populated anthromes in São Paulo, categorized into 8 sets according to IBGE data. Source: own elaboration.

3.5 Static mapping of populated anthromes

To create the static mapping of populated anthromes in São Paulo, a fundamental tool for evaluating the spatial distribution of population data and the accuracy of data distribution across the territory [36], [46], [61], we first used the territorial base where the data would be overlaid. Therefore, the first file loaded was the cities_shape shapefile. Subsequently, we overlaid the populated_anthromes data and defined different arguments for the mapping through complementary functions, such as defining colors, mapping vertices (X and Y), titles, and subtitles. The code for creating the mapping can be accessed on GitHub [30] and RPubs [31], where the digital static mapping is also available. Figure 4 illustrates the static mapping of populated anthromes in the State of São Paulo.

The analytical results reported through the static mapping of populated anthromes in São Paulo (Figure 4) revealed that the populated_anthromes data were overlaid on the cities_shape shapefile as expected, meeting the requirements established by the literature for the construction of static maps [12], [16]. The results also demonstrated that the analytical modeling of populated anthromes followed the guidelines for the distribution of sample points, according to the longitude and latitude information described in the populated_anthromes dataset, as recommended by the geocomputation literature [35], [36], [37]. Furthermore, as indicated by research in this area [46], [49], maintaining the colorimetric standard in the data mapping demonstrated compliance with the premises of visual standardization.

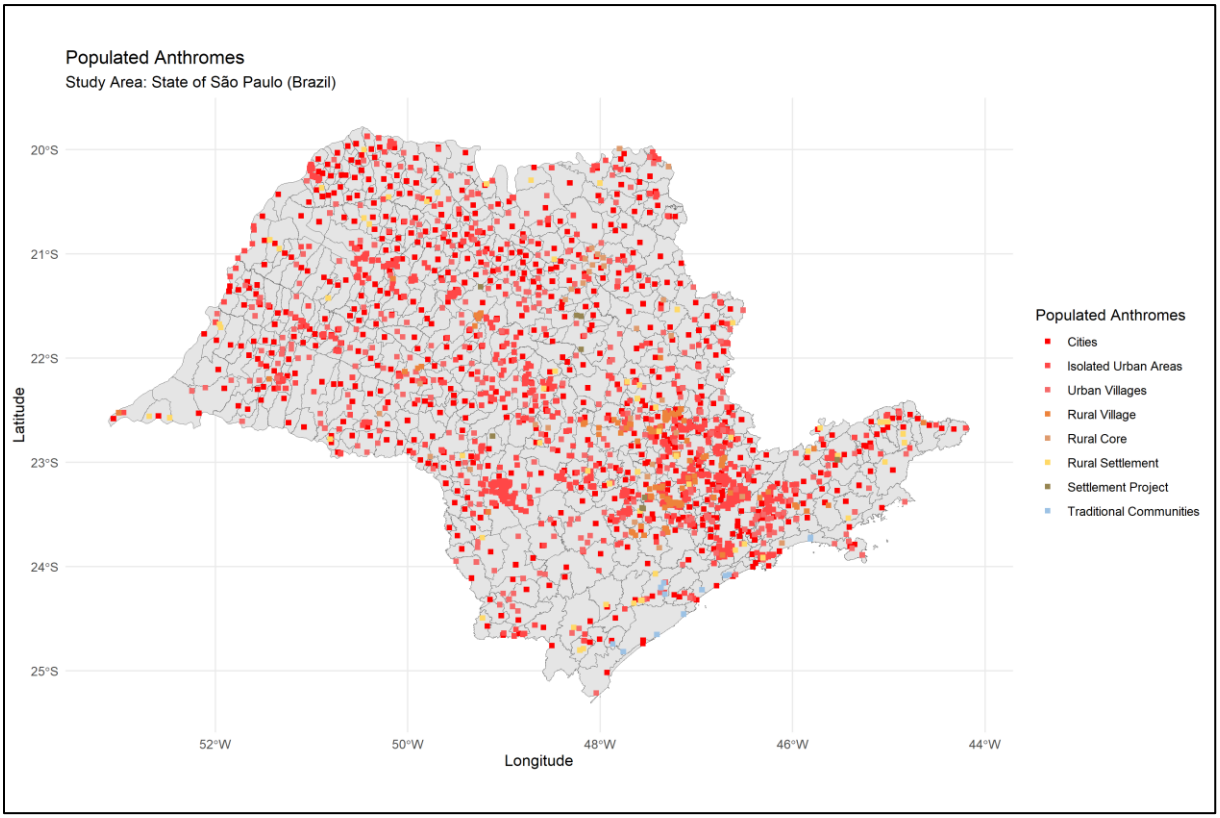


Figure 4. Map showing populated anthromes in São Paulo. Dot shades represent different anthropogenic typologies. Source: own elaboration.

However, we observed that some points relating to populated anthromes (squares in the mapping) extended beyond the areas of the cities_shape shapefile. This aspect was considered in the study of the uncertainty in mapping populated anthromes, a study that we present later. In the next section, we transpose the static mapping into interactive mapping, to structure a map that could be integrated with technological services, such as websites and the GitHub and RPubs repositories.

3.6 Interactive mapping of populated anthromes

As a data source for structuring the leaflet() function, we used the populated_anthromes dataset. Additionally, the colors were defined based on the legend_populatedanthromes dramaframe (Table1) (color table). During the coding of the model in R, we faced the challenge of establishing a measure, in degrees, to guide the addRectangles() function of the leaflet() package. Consulting the literature [35], [37], [49], we observed that the value of 0.03 (positive and negative) had been used in similar research and, during testing, we verified that for smaller degrees, the squares did not cover the surface of some cities. Therefore, we maintained the value of 0.03 to dimension the squares that would represent the populations in the different populated anthropogenic biomes of the State of São Paulo.

As predicted, the first interactive mapping of Brazilian populated anthromes, generated through the analytical model of populated anthromes, is represented in Figure 5, depicting the model area of the State of São Paulo and the populated anthromes present in it. It should be noted that Figure 4 is a screenshot of the virtual model. Therefore, we suggest that it be consulted on RPubs [31] or GitHub [30] for better visualization and detail.

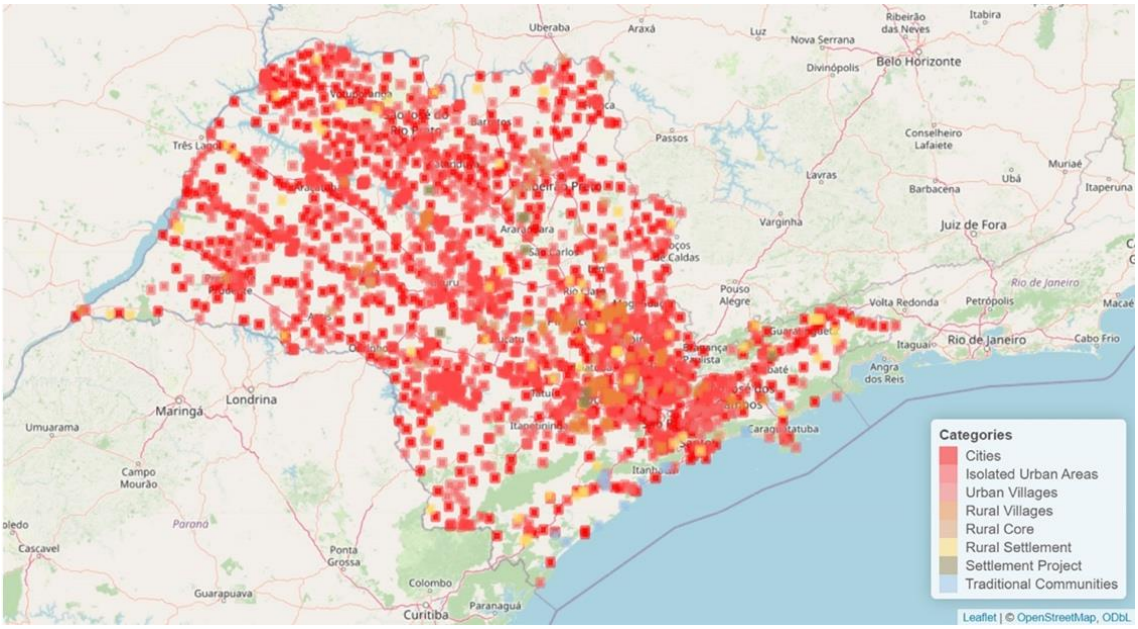


Figure 5. Map illustrating populated anthromes in São Paulo. Each square represents an anthropogenic sector sized ±0.03°; colors correspond to distinct anthropogenic typologies. Source: own elaboration.

Notably, the result illustrated by the interactive mapping of populated anthromes in São Paulo (Figure 5) revealed that the analytical model of populated anthromes adequately represented the anthropogenic sectors distributed throughout the territory of the State of São Paulo. To verify the size of the area covered by the squares in the interactive mapping, we applied (1), (2) y (3). The results of the application were:

- *Square width (longitude) in kilometers: 3.020237 km*
- *Square height (latitude) in kilometers: 1.425164 km*
- *Average square area in square kilometers: 4.304335 km²*

The results demonstrated that the analytical model of populated anthropic structures significantly approximated the information to the local reality and was able to reveal important territorial features for socio-environmental management [12], [16]. This approximation points to a better characterization of the local reality [32], which favors the application of this analytical modeling in socio-ecological studies [44], [62] and in its future application in the construction of the regional model of anthropic structures [63], [64]. We note that up to this point, validation and uncertainty studies of the mapping have not been carried out; this is the subsequent stage of the research, aiming to estimate the quality of the results generated throughout the investigation.

3.7 Mapping validation and uncertainty studies

3.7.1 Examination of data properties

Applying the range() function to the analysis of saopaulo (gold standard) and populated_anthromes (comparator), we obtained the following minimum and maximum latitude and longitude values (Table 5).

Table 5. Minimum and maximum geographic coordinates for the Populated_Anthromes and SãoPaulo sets, determined via the range() function. Source: own elaboration.

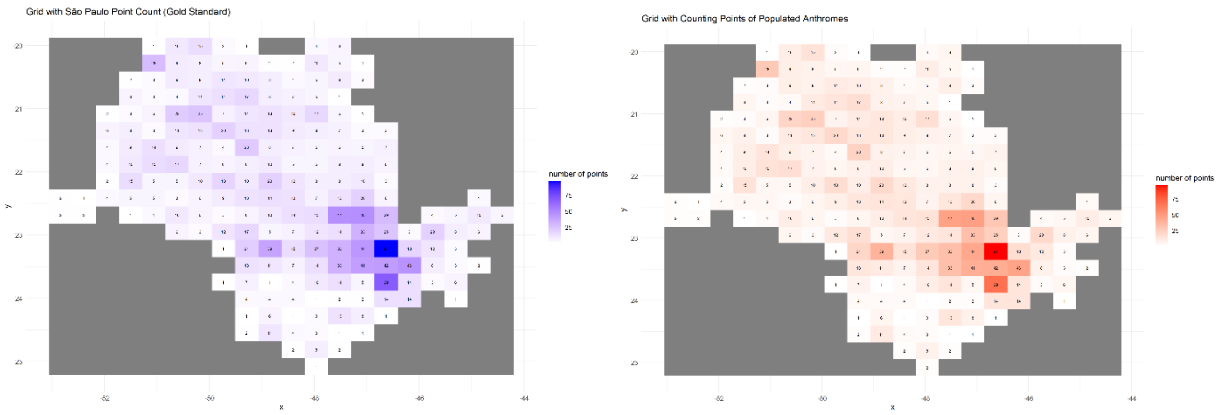
Minimum and Maximum Values	Maximum	Minimum
Maximum and minimum latitude of populated_anthromes	-25.21507	-19.87297
Maximum and minimum latitude of saopaulo	-25.21507	-19.87297
Maximum and minimum longitude of populated_anthromes	-53.05865	-44.149936
Maximum and minimum longitude of saopaulo	-53.05865	-44.19936

The information from the range() function confirmed that the minimum and maximum limits, both for longitude and latitude, were the same for both sets, meeting the assumptions for spatialization of geographic information [62], [65]. Continuing with the verification [35], we used the summary() function to analyze extreme statistical values associated with the two sets, also considering the latitude and longitude information of both sets. The results reported by the summary() function showed small differences in the third quartile of latitude (0.01) and in the median of longitude (0.01), a fact that, according to the literature [56], [57], reflects a small distortion in the mapping. Evaluating these results, we found that the difference was associated with the number of points described by the two sets. In populated_anthromes there were 2141 points, while saopaulo was composed of 2143.

Investigating the two sets, we identified that the missing lines in populated_anthromes referred to two isolated urban areas (AUIs) in the municipality of Estiva Gerbi (Moji Minirim district area). As previously discussed, the aggregation of data in AUIs directly reflects the absence of these lines referring to the municipalities, since small urban areas are considered as a single area in the census data (population01) [39], [45]. Therefore, the error (or spatial limitation) is due to data agglomeration.

3.7.2 Mapping summary statistics

Based on the point count proposed in the analytical modeling, we were able to structure images that represented the number of points in each cell of the statistical grid in the saopaulo (a) and populated_anthromes (b) sets, as illustrated in Figure 6.



a) b)

Figure 6. Statistical grids of point distributions in São Paulo: a) gold standard dataset and b) populated_anthromes. Color gradients reflect point density; gray cells denote areas outside the study region. Source: own elaboration.

Comparatively, we observe in Figure 6 that some of the points referring to the comparator (a) were counted in other cells in the populated anthromes (b). According to analyses reported in other analytical contexts [66], [67], this difference is associated with the EPSG structure in which the different datasets were found. As we demonstrated in previous topics, the CRS were

different between the two sets (comparator and gold standard), forcing us to adjust for the analyses. This generated small distortions in the distribution of points, as we demonstrated through other mapping metrics below.

However, the analytical results in statistical grid were promising and met the demands of spatialization and point distribution in analytical models at local scale [68], [69], [70]. Thus, these results demonstrated that the analytical model met the intended use of spatializing data in accordance with the gold standard, reinforcing the quality of the structured modeling in this research.

3.7.3 Estimation of sensitivity, specificity, accuracy, and error through the confusion matrix

Using the confusion matrix to statistically and metrologically analyze the results obtained throughout this research, we were able to estimate the sensitivity, specificity, accuracy, and error associated with the analytical model of populated anthromes. Through the sensitivity calculation, we obtained a value of 0.9951691, that is, approximately 99.52% of the gold standard points were recovered by the analytical model of populated anthromes. By this metric, we confirm that the analytical model of populated anthromes was able to identify almost all points represented by the gold standard, following the metrological premises of sensitivity [55] and quality [71] of the analytical model.

In terms of specificity, the result obtained for this metric was 0.9896373, that is, in the quadrants of the statistical grid where there were no points identified by the gold standard, the analytical model of populated anthromes operated with a 98.96% accuracy rate, correctly classifying the absence of points in these areas, as suggested by the literature to ensure the efficiency of the model in terms of specificity [69], [70].

It is also worth pointing out that this value indicated that there were quadrants where the points recovered by the analytical model exceeded the number of points reported in the gold standard. This indication led to an examination of the data properties, where we demonstrated that some points were distorted during the spatial distribution (latitude and longitude of the points on the map) and that in the statistical grid analysis they were framed in different cells than those of the comparator.

Thus, despite these limitations, we observed that this metric also contributed to demonstrating the quality of the analytical model developed, considering the alignment of the metrics obtained with those reported by Pang et al. [63] and Candelária et al. [71]. Furthermore, considering the work and premises of Brewer et al. [72], we found that the results obtained through this metric help demonstrate the suitability of the model for its intended use in metrological terms.

Based on the overall accuracy estimate, we obtained a result of 0.9925, that is, approximately 99.25% of the points recovered by the analytical model of populated mantles are correctly represented in the gold standard. This proportion demonstrates a high rate of correspondence between the data represented by the analytical model and the gold standard, pointing to the quality of the analytical model of populated mantles in the pilot area, considering the conformity of this metric with the results reported for it in other analytical contexts [20], [72], [73].

In contrast, through the calculations of the overall error we obtained an estimate of 0.0075, that is, only 0.75% of the points recovered by the analytical model of populated mantles were performed erroneously when compared to the gold standard. This value indicated the low error rate [74], [75] and pointed to the accuracy of the analytical model of populated mantles, reinforcing the observations on the quality of the modeling in line with the estimates reported in other works that evaluated this metrological metric [66], [68].

The statistical and metrological metrics of the analytical model of populated mantles revealed that the results obtained are promising, as we summarize in Table 6. According to other work involving such metrics in analytical models, whether for the analysis or for the

validation of analytical models, we verified that the estimates for the analytical model of populated mantles aligned with the requirements for suitability for the intended use, reflecting the quality of the distribution of points and the efficiency of their representation in spatialities and territory.

Table 6. Results of four statistical metrics used to evaluate the model for mapping populated anthromes, presented in numbers and percentages (rounded to two decimals). Source: own elaboration.

Metrics	Estimation	Percentage
Sensitivity	0.9951691	99.52%
Specificity	0.9896373	98.96%
Global Accuracy	0.9925000	99.25%
Global Error	0.0075000	0.75%

As we observed, the overall error approached zero significantly, which according to metrological assumptions [38] is expressively positive, considering the low distortion index between the recovery capacity of the analytical model of populated anthromes and the gold standard. On the other hand, we note that sensitivity, specificity, and overall accuracy are close to 1, which according to metrological bases [76] indicates a high capacity of the model to represent the data in accordance with the comparator. Thus, we found that the analytical model of populated anthromes met these metrological assumptions and was able to respond to the specifications of these metrics with significant efficiency.

Moving towards the conclusion of our investigations, we summarized the statistical results produced through the Radar Chart (Figure 7). In it, the four metrics are presented simultaneously on a target, where the center represents the value of 0 and the last (outermost) circle the value of 1. This pattern is commonly used by Metrology to analyze the measurement capacity of methods and/or instruments [38], [55], [76]. Since this is one of the sciences that contextualize our investigations and that we seek to integrate into socio-ecological and geocomputation models to add value and robustness to the results, we have incorporated this graphical modeling into the representation of our results, reinforcing our commitment to aligning these sciences.

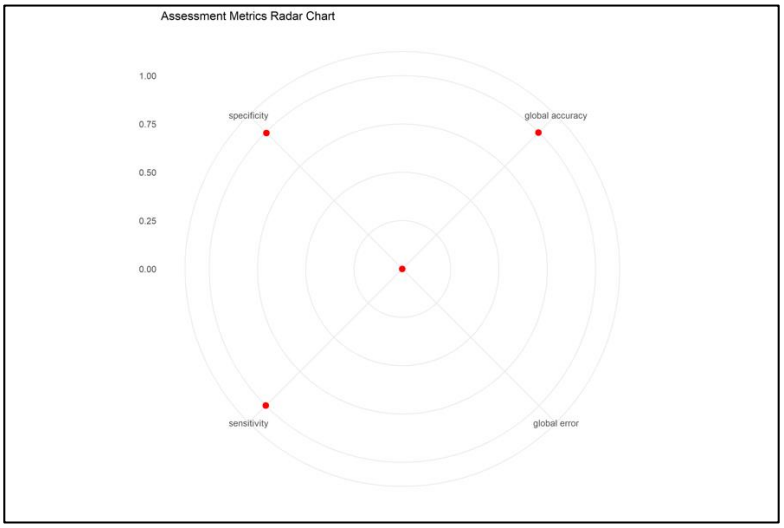


Figure 7. Radar chart displaying the statistical metrics used to validate the mapping model, including sensitivity, specificity, overall accuracy, and global error. Source: own elaboration.

4. CONCLUSIONS

Throughout this study, we developed an analytical model to analyze, characterize, classify, and spatialize population data using the concept of anthromes and applying metrological principles in geocomputation to ensure the quality and validity of the model. During the investigative process to structure the model, we processed and mined data from the Brazilian Demographic Census, classifying the geospatial data according to the anthromes classes established in the global literature on the subject. Subsequently, we plotted the classified data and, sequentially, constructed static and interactive mapping of the populated anthromes in the pilot area, with the aim of visualizing the efficiency of the analytical model in the spatialization and cartographic representation of demographic information.

Once the mapping construction stages were completed, we moved on to statistical and metrological analyses to validate the analytical model of populated anthromes. At this point, we examined the properties of the data, comparing the results from the analytical model of populated anthromes to the information from the gold standard, which was established based on IBGE data. Throughout the investigation, we identified some distortions between the datasets and in the comparative results, revealing limitations in the analytical model, but which did not invalidate the results obtained, mainly considering the statistical metrics analyzed.

In contrast, we showed that the urban-rural-natural structure and gradient of the IBGE data enriches the socio-ecological concept of anthromes, broadening and regionalizing the perspective on the model developed here. From this perspective, the estimates obtained for these metrics showed that the model had significant suitability for the intended use of classification and distribution of points in the cartographic representation of demography, efficiently performing these operations when compared metrologically and statistically to the gold standard. The minor distortions identified at this stage also proved to be limitations of the modeling, but did not invalidate the model, as we discussed throughout the research.

Thus, we found that the R software contributed significantly to the entire analytical process developed here. The functions and packages used throughout the analysis demonstrated the alignment of the products generated in this investigation with global products on anthropomorphic sites. In addition, we saw that the use of the software also contributed significantly to the quality assessment and validation of the analytical modeling of populated anthropomorphic sites. Complementarily, through the analytical scheme developed in this research, we were able to articulate metrological tools to guarantee the validity and quality of results to socio-ecological and geocomputation models, thus adding value to the modeling developed throughout this research.

Finally, based on all the analyses and all the results reported in this research, and considering the globally consolidated products on anthropomorphic sites, we have verified that the analytical model of anthropomorphic sites met the requirements for its application as an investigative instrument on a local scale. Considering the structure of the analytical model developed in R and the high adaptive capacity of structured codes, we created a reproducible and replicable analytical model applicable to other national areas (scalable) and adaptable to other nationalities, given that Brazilian data follows the international standard for vector data structures. Therefore, this research developed an analytical model that can be applied by other researchers for the analysis of demographic data, which articulates metrological bases with socio-ecological and geocomputation models.

5. ACKNOWLEDGMENTS

We would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for the funding allocated to the EcoMetrology Working Group of the

National Institute of Metrology, Quality and Technology (INMETRO-Brazil), with which it was possible to carry out this research. We also thank the Brazilian Institute of Geography and Statistics (IBGE-Brazil) for making the census data available, through which it was possible to carry out all the analytical procedures for the construction and evaluation of the Brazilian anthromes model. Furthermore, thanks are due to the Anthroecology Lab of the University of Maryland, Baltimore Country, USA, in the persons of Dr. Erle C. Ellis and Dr. Nicolas Gauthier, for the bibliographic contribution regarding the modeling of anthropogenic biomes, which allowed the structuring of this research.

6. REFERENCES

[1] C. H. Dabard, C. Mann, and B. Martín-López, "An archetype analysis of sustainability innovations in Biosphere Reserves: Insights for assessing transformative potential," *Environ. Sci. Policy,* vol. 153, p. 103674, Mar. 2024. https://doi.org/10.1016/j.envsci.2024.103674

[2] U. Khan, and W. Liu, "The role of internal auditing on corporate governance: its effects of economic and environmental performance," *Environ. Sci. Pollut. Res.*, vol. 30, no. 4, pp. 15112877-112891, Nov. 2023. https://doi.org/10.1007/s11356-023-30363-5

[3] M. V. Prazian, and V. M. Prykhodko, "High-quality education for better sustainability and resiliency," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1254, no. 1, p. 012031, 2023. https://doi.org/10.1088/1755-1315/1254/1/012031

[4] M. Li, P. H. Verburg, and J. van Vliet, "Global trends and local variations in land take per person," *Landsc. Urban Plan.*, vol. 218, p. 104308, Feb. 2022. https://doi.org/10.1016/j.landurbplan.2021.104308

[5] A. Schaffartzik, M. Pichler, E. Pineault, D. Wiedenhofer, R. Gross, and H. Haberl, "The transformation of provisioning systems from an integrated perspective of social metabolism and political economy: a conceptual framework," *Sustain. Sci.*, vol. 16, pp. 1405-1421, Sep. 2021. https://doi.org/10.1007/s11625-021-00952-9

[6] S. Ryder, K. Powlen, M. Laituri, S. A. Malin, J. Sbicca, and D. Stevis, *Environmental Justice in the Anthropocene*. Routledge: London, 2021, p. 358. https://doi.org/10.4324/9781003023609

[7] H. Meng, X. Gao, Y. Song, G. Cao, and J. Li, "Biodiversity arks in the Anthropocene," *Reg. Sustain.*, vol. 2, no. 2, p. 109-115, Apr. 2021. https://doi.org/10.1016/j.regsus.2021.03.001

[8] S. Basu, H. Nagendra, P. Verburg, and T. Plieninger, "Perceptions of ecosystem services and knowledge of sustainable development goals around community and private wetlands users in a rapidly growing city," *Landsc. Urban Plan.*, vol. 244, p. 104989, Apr. 2024. https://doi.org/10.1016/j.landurbplan.2023.104989

[9] I. Merylova, V. Smilka, and G. Kovalska, "Public Spaces in Historic Environment as Urban Fundamentals of Sustainable Development," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1275, no. 1, p. 12026, 2023. https://doi.org/10.1088/1755-1315/1275/1/012026

[10] I. Wayan Wiryawan, "Land Consolidation Process Based on Environmental Conservation and Urban Spatial Law," *IOP Conf. Ser.: Earth Environ. Sci*. vol. 1275, no. 1, p. 012024, 2023. https://doi.org/10.1088/1755-1315/1275/1/012024

[11] J. S. Appannan, R. Mohd Said, T. San Ong, and R. Senik, "Promoting sustainable development through strategies, environmental management accounting and environmental performance," *Bus. Strateg. Environ.*, vol. 32, no. 4, p. 1914-1930, May. 2023. https://doi.org/10.1002/bse.3227

[12] E. C. Ellis, A. H. W. Beusen, and K. K. Goldewijk, "Anthropogenic Biomes: 10,000 BCE to 2015 CE," *Land*, vol. 9, no. 5, p. 129, Apr. 2020. https://doi.org/10.3390/land9050129

[13] E. C. Ellis, "Sustaining biodiversity and people in the world's anthropogenic biomes," *Curr. Opin. Environ. Sustain.*, vol. 5, no. 3-4, p. 368-372, Sep. 2013. https://doi.org/10.1016/j.cosust.2013.07.002

[14] E. C. Ellis, "Anthromes," in *Encyclopedia of the World's Biomes*, M. I. Goldstein, and D. A. Dellasala, Ed., Oxford, U. K.: Elsevier, 2020. p. 5-11. https://doi.org/10.1016/B978-0-12-409548-9.12494-7

[15] E. C. Ellis, and N. Ramankutty, "Putting people in the map: anthropogenic biomes of the world," *Front. Ecol. Environ.*, vol. 6, no. 8, p. 439-447, Oct. 2008. https://doi.org/10.1890/070062

[16] N. Gauthier, *Anthromes 12K DGG (V1) analysis code and R research compendium*. Harvard Dataverse, 2021. https://doi.org/10.7910/DVN/6FWPZ9

[17] I. Ramli, A. Achmad, N. Nizamuddin, A. Izzaty, and I. Irzaidi, "Transformation of land use and land cover and its implication to the local climate in Aceh Besar District, Aceh, Indonesia," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1263, no. 1, p. 012023, 2023. https://doi.org/10.1088/1755-1315/1263/1/012023

[18] A. Sofianto, L. Febrian, O. C. Ambarwati, and T. Risandewi, "Mapping of Regional Strategic Issues in The Development Planning Documents (Case Study in Central Java)," *IOP Conf. Ser: Earth Environ. Sci*., vol. 1263, no. 1, p. 012036, 2023. https://doi.org/10.1088/1755-1315/1264/1/012036

[19] R. V. Leite *et al*. "Large scale multi-layer fuel load characterization in tropical savanna using GEDI spaceborne lidar data," *Remote Sens. Environ.*, vol. 268, p. 112764, Jan. 2022. https://doi.org/10.1016/j.rse.2021.112764

[20] R. Dalagnol *et al*. "Mapping tropical forest degradation with deep learning and Planet NICFI data," *Remote Sens. Environ*., vol. 298, p. 113798, Dec. 2023. https://doi.org/10.1016/j.rse.2023.113798

[21] A. Savitri, A. E. Pravitasari, and V. B. Rosandi, "Dynamics of land cover change, regional development, and its local dependence driving factors in Bojonegoro Regency," *IOP Conf. Ser.: Earth Environ. Sci*., vol. 1263, no. 1, p. 012014-12028, 2023. https://doi.org/10.1088/1755-1315/1263/1/012014

[22] C. E. Vickery, and J. E. Quinn, "Climate-altered Precipitation is more Important than Land Use when Modeling Ecosystem Services Associated with Surficial Processes," *Environ. Manag.*, vol. 72, p. 1216-1227, Dec. 2023. https://doi.org/10.1007/s00267-023-01861-6

[23] F. H. Mardiansjah, S. Ma'rif, and A. Sugiri, "Analyzing Urban Population Growth in the Towns of Non-urban Regions in Java, Indonesia, Using Spatial Analysis," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1264, p. 012012, 2023. https://doi.org/10.1088/1755-1315/1264/1/012012

[24] C. Venier-Cambron, Ž. Malek, and P. H. Verburg, "Avoiding an unjust transition to sustainability: An equity metric for spatial conservation planning," *Proc. Natl. Acad. Sci.*, vol. 120, no. 43, p. e2216693120, Oct. 2023. https://doi.org/10.1073/pnas.2216693120

[25] R. Aditantri, L. A. Rahmi, A. N. Kusumodewi, and A. D. Wibisono, "Growth Centre Interaction through Gravity Model Approach using Geographic Information System," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1264, no. 1, p. 012019, 2023. https://doi.org/10.1088/1755-1315/1264/1/012019

[26] E. Ellis, and Anthroecology Lab, "Anthromes 12K DGG (v1)," anthroecology.org. [Online]. Available: https://anthroecology.org/anthromes/12kdggv1/maps/ge/

[27] M. S. L. A. Gobbo, T. de O. Araujo, and C. O. F. Salema, "Metrology Applied in Decision-Making for Land Planning: A Guide for Critical Analysis to Guarantee the Quality of Land Use and Cover Results," *Mapan*, vol. 38, no. 2, p. 779-794, Dec. 2023. https://doi.org/10.1007/s12647-023-00679-1

[28] M. S. L. Araujo Gobbo, T. de Oliveira Araujo, and C. O. Faria Salema, "Critical Analysis of Demographic Data Based on ISO/IEC 17,025 Standard for the Regionalization of Brazilian Anthromes," *Mapan*, vol. 37, no. 2, p. 83-109, Mar. 2022. https://link.springer.com/article/10.1007/s12647-022-00573-2

[29] ISO/CASCO Committee on conformity assessment, "General requirements for the competence of testing and calibration laboratories," ISO/IEC 17025:2017, International Organization for Standardization, Geneva, Switzerland, Nov. 2017. [Online]. Available: https://www.iso.org/standard/66912.html

[30] M. S. L. Araujo Gobbo, "Brazilian Anthromes," GitHub, 2026. Available: https://github.com/maximilianogobbo/brazilian-anthromes

[31] M. S. L. Araujo Gobbo, "Repository R Pubs - Maximiliano Gobbo." Rpubs, 2026. Available: https://rpubs.com/maximilianogobbo

[32] J. E. Quinn, and E. C. Ellis, "Anthromes," in *Handbook of the Anthropocene: Humans between Heritage and Future,* N. Wallenhorst, and C. Wulf, Ed., Cham, SWI: Springer International Publishing, 2023. p. 203-211. https://doi.org/10.1007/978-3-031-25910-4_32

[33] M. J. C. Walker *et al*., "The Anthropocene is best understood as an ongoing, intensifying, diachronous event," *Boreas*, vol. 52, no. 3, p 1-3, Jan. 2023. https://doi.org/10.1111/bor.12636

[34] E. C. Ellis, "Land Use and Ecological Change: A 12,000-Year History," *Annu. Rev. Environ. Resour.*, vol. 46, no. 1, p. 1-33, Oct. 2021. https://doi.org/10.1146/annurev-environ-012220-010822

[35] R. Lovelace, J. Nowosad, and J. Muenchow, *Geocomputation with R*. London, U. K.: Chapman and Hall/CRC, 2019.

[36] E. Pebesma, and R. Bivand, *Spatial Data Science: With Applications in R*. New York, USA: Chapman and Hall/CRC, 2023. https://doi.org/10.1201/9780429459016

[37] H. Wickham, *Advanced R*. London, U. K.: CRC Press, 2022. https://adv-r.hadley.nz/

[38] G. M. Mahmoud, S. M. Osman, and R. S. Hegazy, "Proposed approach for force transducers classification," *Int. J. Metrol. Qual. Eng.*, vol. 12, p. 3, Jan. 2021. https://doi.org/10.1051/ijmqe/2021001

[39] Instituto Brasileiro de Geografia e Estatística (IBGE), *Metodologia do censo demográfico 2010*. Rio de Janeiro, Brasil: IBGE, 2013. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=281634

[40] Instituto Brasileiro de Geografia e Estatística (IBGE), *Proposta Metodológica para Classificação dos Espaços do Rural, do Urbano e da Natureza do Brasil.* Ministério do Planejamento e Orçamento.

Rio de Janeiro, Brasil: Instituto Brasileiro de Geografia e Estatística - Coordenação de Geografia, 2023. https://journals.openedition.org/confins/53955?lang=pt

[41] Instituto Brasileiro de Geografia e Estatística (IBGE), *Pesquisa Nacional por Amostra de Domicílios Contínua: notas técnicas.* Rio de Janeiro, Brasil: Ministério da Economia/IBGE, ed. 1.7, 2020. https://biblioteca.ibge.gov.br/visualizacao/livros/liv101708_notas_tecnicas.pdf

[42] Instituto Brasileiro de Geografia e Estatística (IBGE), *Monitoramento da Cobertura e Uso da Terra do Brasil 2016-2018*. Rio de Janeiro, Brasil: IBGE, 2020. https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15831-cobertura-e-uso-da-terra-do-brasil.html

[43] Instituto Brasileiro de Geografia e Estatística (IBGE), *Projeções da população: Brasil e Unidades da Federação – revisão 2018*. Rio de Janeiro, Brasil: IBGE, 2018. https://www.ibge.gov.br/estatisticas/sociais/populacao/9109-projecao-da-populacao.html

[44] R. Biggs, A. de Vos, R. Preiser, H. Clements, K. Maciejewski, and M. Schlüter, *The Routledge Handbook of Research Methods for Social-Ecological Systems*. London and New York: Routledge, 2021. https://www.routledge.com/The-Routledge-Handbook-of-Research-Methods-for-Social-Ecological-Systems/Biggs-deVos-Preiser-Clements-Maciejewski-Schluter/p/book/9781032020761

[45] Instituto Brasileiro de Geografia e Estatística (IBGE), *Censo Demográfico 2010*. Rio de Janeiro, Brasil: IBGE, 2010. https://sidra.ibge.gov.br/pesquisa/censo-demografico/demografico-2010/inicial

[46] H. Wickham, M. Çetinkaya-Rundel, and G. Grolemund, *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd ed. Sebastopol, CA: O'Reilly Media, 2023. https://r4ds.hadley.nz/

[47] L. K. Katherina, "Indonesia: a case study of West Java Province," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1263, no. 1, p. 012015-12026, 2023. https://doi.org/10.1088/1755-1315/1263/1/012015

[48] S. Chowdhury *et al.*, "Protected areas in South Asia: Status and prospects," *Sci. Total Environ.,* vol. 811, p. 152316-152316, Mar. 2022. https://doi.org/10.1016/j.scitotenv.2021.152316

[49] E. C. Anderson, *Making Maps with R: Reproducible Research Course*. U.S.: GitHub, 2021. https://eriqande.github.io/rep-res-web/lectures/making-maps-with-R.html

[50] B. González-Mon *et al.*, "The importance of cross-scale social relationships for dealing with social-ecological change in agricultural supply chains," *J. Rural Stud*. vol. 105, p. 103191-103204, Jan. 2024. https://doi.org/10.1016/j.jrurstud.2023.103191

[51] M. H. R. Sales, S. Bruin, C. de Souza, and M. Herold, "Land Use and Land Cover Area Estimates from Class Membership Probability of a Random Forest Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 1-11, 2021. https://doi.org/10.1109/TGRS.2021.3080083

[52] X. Li *et al.*, "A cellular automata downscaling based 1 km global land use datasets (2010–2100)," *Sci. Bull.*, vol. 61, no. 21, p. 1651-1661, Nov. 2016. https://doi.org/10.1007/s11434-016-1148-1

[53] R. Rose-Redwood, "Introduction: The Limits to Deconstructing the Map," *Cartogr.: Int. J. Geogr. Inf. Geovisualization*, vol. 50, no. 1, p. 1-8, Jan. 2015. https://doi.org/10.3138/carto.50.1.01

[54] L. M. Harris, "Deconstructing the Map after 25 Years: Furthering Engagements with Social Theory," *Cartogr: Int. J. Geogr. Inf. Geovisualization*, vol. 50, no. 1, p. 50-53, Jan. 2015. https://doi.org/10.3138/carto.50.1.10

[55] J. Mittaz, C. J. Merchant, and E. R. Woolliams, "Applying principles of metrology to historical Earth observations from satellites," *Metrologia,* vol. 56, no. 3, May. 2019. http://doi.org/10.1088/1681-7575/ab1705

[56] R. Lovelace, J. Cheshire, and R. Oldroyd, *Introduction to visualizing spatial data in R*. leeds, U.K.: University of Leeds, 2017. https://cran.r-project.org/doc/contrib/intro-spatial-rl.pdf

[57] K. Hartung *et al.*, "Bookkeeping estimates of the net land-use change flux – a sensitivity study with the CMIP6 land-use dataset," *Earth Syst. Dynam.*, vol. 12, no. 2, p. 763-782, Jun. 2021. https://doi.org/10.5194/esd-12-763-2021

[58] G. Kalt *et al.*, "Exploring the option space for land system futures at regional to global scales: The diagnostic agro-food, land use and greenhouse gas emission model BioBaM-GHG 2.0," *Ecol. Model.*, vol. 459, p. 109729-109745, Nov. 2021. https://doi.org/10.1016/j.ecolmodel.2021.109729

[59] M. Staniak, and P. Biecek, "The Landscape of R Packages for Automated Exploratory Data Analysis," *The R Journal*, vol. 11, no. 2, p. 23, Sep. 2019. https://arxiv.org/pdf/1904.02101

[60] A. R. Rakhmatulloh, D. I. K. Dewi, and E. K. Syahri, "Spatial Entropy Model for the Impact Assessment of Built Environment on Urban Street Design," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1264, no. 1, p. 012047, 2023. https://doi.org/10.1088/1755-1315/1264/1/012047

[61] P. Santillán-Carvantes *et al.*, "Spatial characterization of social-ecological systems units for management in Tropical Dry Forests," *Lands. Ecol.*, vol. 38, p. 4303-4323, Dec. 2023. https://doi.org/10.1007/s10980-023-01714-x

[62] S. H. Shah *et al.*, "Connecting physical and social science datasets: challenges and pathways forward," *Environ. Res. Commun.*, vol. 5, no. 9, p. 095007-95020, 2023.

https://doi.org/10.1088/2515-7620/acf6b4
[63] M. Grace Albright, C. Vickery, R. Bower, and J. E. Quinn, "Patterns of land use change, land governance, and the supply of ecosystem services in a multifunctional landscape: A case study from Upstate SC, USA," *J. Land Use Sci.*, vol. 18, no. 1, p. 284-295, Jul. 2023. https://doi.org/10.1080/1747423X.2023.2234903
[64] V. Amelia, S. Sinaga, and A. Bhermana, "Land resource management on environment and sustained basis for agricultural land use planning using landform and land evaluation approach (a case study in North Barito District, Central Kalimantan Province)," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1282, no. 1, p. 012001-12009, 2023. https://doi.org/10.1088/1755-1315/1282/1/012001
[65] Y. Tang *et al.*, "Estimating wheat grain yield by assimilating phenology and LAI with the WheatGrow model based on theoretical uncertainty of remotely sensed observation," *Agr. For. Meteorol.*, vol. 339, p. 109574-109593, Aug. 2023. https://doi.org/10.1016/j.agrformet.2023.109574
[66] K. Barnett, and R. T. Belote, "Modeling an aspirational connected network of protected areas across North America," *Ecol. Appl.*, vol. 31, no. 6, p. e2387, Sep. 2021. https://doi.org/10.1002/eap.2387
[67] S. A. Abdullah, K. H. Abbas, A. H. Harif, R. F. Alaa, and S. Hiba, "Sustainable Urban Distribution of Educational Institutions and Population Density in Baghdad City Using Remote Sensing Techniques," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1202, no. 1, p. 012015, 2023. https://doi.org/10.1088/1755-1315/1202/1/012015
[68] M. Ward *et al.*, "A national-scale dataset for threats impacting Australia's imperiled flora and fauna," *Ecol. Evol.*, vol. 11, no. 17, p. 11749-11761, Sep. 2021. https://doi.org/10.1002/ece3.7920
[69] A. A. Plowright, N. C. Coops, C. M. Chance, S. R. J. Sheppard, and N. W. Aven, "Multi-scale analysis of relationship between imperviousness and urban tree height using airborne remote sensing," *Remote Sens. Environ.*, vol. 194, p. 391-400, Jun. 2017. https://doi.org/10.1016/j.rse.2017.03.045
[70] S. E. H. Pang, J. W. Ferry Slik, D. Zurell, and E. L. Webb, "The clustering of spatially associated species unravels patterns in tropical tree species distributions," *Ecosphere*, vol. 14, no. 6, p. e4589, Jun. 2023. https://doi.org/10.1002/ecs2.4589
[71] S. Schmitt, G. Salzet, F. J. Fischer, I. Maréchaux, and J. Chave, "rcontroll: An R interface for the individual-based forest dynamics simulator TROLL," *Methods Ecol. Evol.*, vol. 14, no. 11, p. 2749-2757, Nov. 2023. https://doi.org/10.1111/2041-210X.14215
[72] L. Du *et al.*, "Forest terrain and canopy height estimation using stereo images and spaceborne LiDAR data from GF-7 satellite," *Geo-spat. Inf. Sci.*, vol. 27, no. 3, p. 811-821, Sep. 2023. https://doi.org/10.1080/10095020.2023.2249037
[73] L. A. Al-Ani, and A. H. Mohammed, "Principal Component Analysis of Selective Bands for Digital Satellite Imagery Classification," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1202, no. 1, p. 012002, 2023. https://doi.org/10.1088/1755-1315/1202/1/012002
[74] D. Pouliot, R. Latifovic, N. Zabcic, L. Guindon, and I. Olthof, "Development and assessment of a 250 m spatial resolution MODIS annual land cover time series (2000–2011) for the forest region of Canada derived from change-based updating," *Remote Sens. Environ.*, vol. 140, p. 731-743, Jan. 2014. https://doi.org/10.1016/j.rse.2013.10.004
[75] K. G. Laraby, and J. R. Schott, "Uncertainty estimation method and Landsat 7 global validation for the Landsat surface temperature product," *Remote Sens. Environ.*, vol. 216, p. 472-481, Oct. 2018. https://doi.org/10.1016/j.rse.2018.06.026
[76] P. J. Brewer *et al.*, "Advances in reference materials and measurement techniques for greenhouse gas atmospheric observations," *Metrologia*, vol. 56, no. 3, p. 034006, 2019. http://doi.org/10.1088/1681-7575/ab1506

DECLARATION IA

The spelling and grammar of this manuscript were reviewed using artificial intelligence (Microsoft Copilot). This tool made it possible to identify spelling and grammar problems and inconsistencies after a complete review of the manuscript in a language other than the authors' native language.

CONFLICT OF INTEREST

This manuscript presents no conflict of interest. All data used in this research are open national data, available to the scientific and social community. The development of the scripts and analytical structures, inspired by global models, are also available for access and were not intended to generate conflicts with the pioneering authors on the subject; on the contrary, the objective was to develop a global and integrative science among the various research centers on the subject.

AUTHORSHIP CONTRIBUTION

Maximiliano S. L. A. Gobbo: Writing this manuscript, Conducted the research, Developed the methodology, Analyzed the results, and arrived at conclusions based on the discussion of the results generated.

Thiago de Oliveira Araujo: Supervised and guided the development of the work. Manuscript review before publication and contributed theoretically to the analysis and discussion of the results.

Cláudia de Oliveira Faria Salema: Supervised and guided the development of the work. Manuscript review before publication and contributed theoretically to the analysis and discussion of the results.