Article

Identifying Urban and Socio-Environmental Patterns of Brazilian Amazonian Cities by Remote Sensing and Machine Learning

Bruno Dias dos Santos [1,2,*], Carolina Moutinho Duque de Pinho [3], Antonio Páez [2] and Silvana Amaral [1]

1 Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil; silvana.amaral@inpe.br
2 School of Earth, Environment & Society, McMaster University, Hamilton, ON L8S 4K1, Canada; paezha@mcmaster.ca
3 Center for Engineering, Modeling and Applied Social Sciences (CECS), Federal University of ABC (UFABC), Santo André 09210-580, Brazil; carolina.pinho@ufabc.edu.br
* Correspondence: bruno.santos@inpe.br

Abstract: Identifying urban patterns in the cities in the Brazilian Amazon can help to understand the impact of human actions on the environment, to protect local cultures, and secure the cultural heritage of the region. The objective of this study is to produce a classification of intra-urban patterns in Amazonian cities. Concretely, we produce a set of Urban and Socio-Environmental Patterns (USEPs) in the cities of Santarém and Cametá in Pará, Brazilian Amazon. The contributions of this study are as follows: (1) we use a reproducible research framework based on remote sensing data and machine learning techniques; (2) we integrate spatial data from various sources into a cellular grid, separating the variables into environmental, urban morphological, and socioeconomic dimensions; (3) we generate variables specific to the Amazonian context; and (4) we validate these variables by means of a field visit to Cametá and comparison with patterns described in other works. Machine learning-based clustering is useful to identify seven urban patterns in Santarém and eight urban patterns in Cametá. The urban patterns are semantically explainable and are consistent with the existing scientific literature. The paper provides reproducible and open research that uses only open software and publicly available data sources, making the data product and code available for modification and further contributions to spatial data science analysis.

Keywords: urban pattern; unsupervised classification; amazon; urban morphology; urban remote sensing

1. Introduction

According to UN-Habitat [1], global population growth has begun to slow and will continue to decelerate in the coming decades. At the same time, population has become increasingly urban: globally, the share of people living in cities doubled between 1950 and 2020, increasing from 25% to 50%. Current projections suggest a slow increase to 58% in the next 50 years, with a decrease in the contribution of cities, semi-dense areas, and rural areas.

In the Brazilian Amazon, one of nature's last frontiers, this trend is bucked to some extent. The most recent demographic census of 2010 revealed that the growth of small- and medium-sized cities continues apace in this region, following the reorganization of the national urban network and a new territorial division of labor [2–4]. The Brazilian Amazon has experienced fast urban population growth, with rates increasing from 37.4% in 1970 to 44.9% in 1980 and 55.2% in 1991. This trend led geographer Bertha Becker [5] to coin the term "urbanized forest". Over 70% of the population in the Amazon lives in urban areas [6], highlighting the relevance of urban issues to regional public policies. Moreover,

the region presents specific characteristics, which include distinctive historical processes of human occupation, climatic and physical features, strong cultural identity, and land and environmental conflicts.

In addition to being labeled "urbanized forests", the Brazilian Amazon also exhibits sharp internal economic divisions at the intra-urban scale. Diverse land use patterns define residential, industrial, and commercial zones. These zones vary in age, location, size, function, and form [7]. This differentiation within cities results from various spatial processes, including state intervention and the impact of private investments. As a result, the objects that compose and define the urban fabric—streets, houses, plots, and blocks— create variegated patterns that interact with the natural environment [8–10].

The need to develop effective policies to manage these regions calls for efficient information systems. In this context, remote sensing has become a valuable tool to complement urban planning [7,11–17]. Recent research by Santos et al. [18] notes that there is a large amount of research that applied remote sensing for identifying urban patterns in Brazilian cities in the southeast, particularly in São Paulo and Rio de Janeiro. Much of this research predominantly used very-high spatial resolution images from private multispectral sensors. Furthermore, automated classification techniques have not yet been widely used. Based on these findings, Santos et al. [18] proposed a framework specifically designed for cities in the Amazon to make use of publicly available data and machine learning.

With this as a background, the present paper aims to produce a classification of Urban and Socio-Environmental Patterns (USEP) in Amazonian cities. We adopt the framework of Santos et al. [18] to identify USEPs in the cities of Santarém, located in western Pará State, and Cametá, located in northeastern Pará State, both situated in the Brazilian Amazon. The study uses remote sensing imagery and machine learning, which are complemented with a field visit to the city of Cametá, where field observations help validate the classification system of USEPs and to make sure that the classes are semantically explainable.

This research presents a comprehensive approach that combines remote sensing data, urban morphological evaluation metrics, and socioeconomic indicators to identify and characterize diverse urban patterns in Amazonian cities. We ensured the transparency and reproducibility of our study by sourcing all data from publicly available repositories. To facilitate collaboration and further analysis, we have made our data product and analysis code accessible through a Zenodo repository linked to our GitHub page [19], aligning with best practices in spatial data science [20,21]. These contributions enhance our understanding of urban development in the Amazon region and provide a foundation for further research and policymaking.

After this introduction, the structure of the article is as follows: Section 2 provides a brief background on the identification of urban patterns in the Brazilian Amazon; Section 3 outlines the study areas; Section 4 introduces our methodology for identifying and characterizing USEPs; Section 5 presents the classification results; Section 6 discusses the main findings; and, finally, Section 7 concludes with final considerations.

2. Background

Remote sensing increasingly plays a significant role in urban mapping, particularly in developing countries [22]. Initiatives such as the Global Urban Footprint (GUF) [23], World Settlement Footprint Evolution (WSF-Evo) [24], and the World Urban Database [25] have emerged to map urban areas globally. However, previous studies have faced challenges in utilizing remote sensing for detailed urban analyses, mainly due to the limited availability of very high-resolution satellite imagery and efficient algorithms.

To address this limitation, Zhu et al. [22] conducted a global-scale study using Sentinel-1 and Sentinel-2 satellite data to identify 17 distinct urban patterns with variations in land use, building density, and verticalization. Their research focused on cities worldwide with populations exceeding 300,000 inhabitants, filling a gap in understanding urban morphology globally.

Additionally, the Global Human Settlement Layer (GHSL) initiative [26] recently provided valuable spatial information about human settlements over time. By utilizing satellite imagery, including Landsat, Sentinel, and the China–Brazil Earth Resources Satellite CBERS-2B, the GHSL offers detailed intra-urban-scale databases. The latest GHSL data package, released in 2022, encompasses features such as multi-temporal built-up area classification, identification of residential and non-residential areas, and average building height.

According to Gonçalves et al. [27], these global databases face some difficulties in identifying less densely built urban areas, for example in the Brazilian Amazon. These authors analyzed the urban extent in seven mapping bases for 2010 using remote sensing data and a regular grid to assess the consistency and agreement between databases. The study focused on six cities in Pará State and found that areas of medium and high building density had over 90% agreement between databases, while the largest discrepancies were observed for lower-density urban patterns.

To the best of our knowledge, only two other studies have examined urban patterns in the Brazilian Amazon using remote sensing: Dal'Asta et al. [28] and Santos et al. [29]. In addition, Cardoso et al. [30] assert the lack of databases and cartographic materials for Amazonian cities. Although there is an abundance of noteworthy local research on Amazonian urban areas, there is not much evidence to suggest that public managers have incorporated these studies into public policy agendas.

Identifying urban patterns in the Amazon is crucial for promoting sustainability in the region [31]. By analyzing different urban patterns, it becomes possible to identify zones with lower and higher environmental impact and that might be more compatible with environmental and cultural living conditions of the region. Additionally, identifying occupation patterns can help to protect local cultures and preserve the cultural heritage of the region.

Dal'Asta et al. [28] visually interpreted CBERS-HRC panchromatic (5 m spatial resolution) and multispectral images (20 m spatial resolution) to identify five typologies of human occupation in western Pará: dense settlement (high-density residential and commercial areas), sparse settlement (low-density residential areas with vegetation between homes), expansion areas (widely sparse low-density residential areas), large nonresidential buildings (e.g., gyms, community centers, factories), and access roads (undeveloped areas around highways and rivers).

Santos et al. [29] identified typologies of precarious settlements in Altamira, Cametá, and Marabá in the state of Pará. To achieve this goal, the authors used the Geographic Object-Based Image Analysis (GEOBIA) and data mining techniques on WPM images from the CBERS-4A satellite (2 m spatial resolution for the panchromatic band and 8 m for the multispectral bands), along with biophysical indices, Gray Level Co-occurrence Matrix metrics, context metrics, and neighborhood metrics. This work was the first study to apply remote sensing to identify precarious settlements in Amazonian cities and classify them into distinct typologies.

The contributions of Santos et al. [29] to identifying urban patterns in Amazonian cities include the following: (i) the development of a methodology for preprocessing and classifying data using open-source software and analyzing free-access data; (ii) the use of imagery provided by the WPM sensor from the CBERS-4A satellite to classify urban land cover using the GEOBIA approach and a machine learning algorithm as a classifier; and (iii) the use of a cellular grid to aggregate data from several types and sources. Moreover, this work concluded that an integrative approach is necessary to identify precarious settlements in Amazonian cities. Unlike many Brazilian metropolitan areas, non-metropolitan Amazonian cities have precarious areas that are not easily distinguishable from non-precarious development.

Based on the recognition that it is necessary to develop urban pattern classification models specifically for Amazonian cities, Santos et al. [18] proposed a framework to classify Urban and Socio-Environmental Patterns (USEPs). A USEP is defined as an ele-

ment in the urban fabric with homogeneous environmental, urban morphological, and socioeconomic characteristics.

The USEPs framework is built on the use of publicly available satellite imagery with high and medium spatial resolution and the adoption of automated classification techniques that do not rely on prior classifications or monitoring. The framework is based on two assumptions: (i) distinctive urban patterns can be observed within a city based on the environment, socioeconomic factors, and urban morphology; and (ii) the combination of remote sensing imagery, demographic census, and Volunteered Geographic Information (VGI). In this research, we validate the framework proposed by Santos et al. [18].

Although not strictly in the remote sensing tradition, it is worth mentioning several urban morphology studies that have made significant methodological contributions to the identification of urban patterns in cities of the Brazilian Amazon. For instance, studies focused on Belém–PA investigate the transformations that occurred in the land tenure system [32], urban expansion [33,34], identification of morphological patterns [35] and evaluation of building cycles in plots [36]. Gomes and Cardoso [37] used the Conzenian concepts of the morphological region and peripheral belts to understand Santarem's formation and expansion process. Such studies describe how real estate market agents act in the Pará cities of Belém, Marabá, and Parauapebas.

Some authors have been calling for new approaches to urban studies, which understand the city as a space dependent on ecosystem resources [38] and allow the coexistence between humans and nature [39]. However, especially for Amazonian cities, new approaches are needed to value traditional and native practices already facing disappearance situations.

3. Study Areas

The study areas combined encompass approximately 187 km^2 and are situated on the municipal centers of Santarém (143 km^2) and Cametá (44 km^2) (Figure 1). Both cities are in the state of Pará, within the Brazilian Amazon. The study site also includes rural areas in the vicinity of both cities. The delimitation of Cametá follows the delineation by Santos et al. [29]. For Santarém, we adopted the delimitation method described by Gonçalves et al. [27], which utilizes nighttime light images to identify potentially populated areas.

Both cities were founded by Portuguese Jesuits in the 16th century with Portuguese territorial expansion as a backdrop. This expansion aimed to extract and export valuable products from the Amazon rainforest to Europe, in a process known in Brazil as "drugs of the hinterlands" (*Drogas do Sertão*, in Portuguese). Condiments, medicines, ornaments, and construction materials were extracted by indigenous people, stored by the Jesuits, and shipped to Portugal through Belém, forming an urban network commanded by the capital of Para [40].

Santarém is located at the confluence of the Tapajós and Amazonas Rivers, in an area of fertile soil, dense forest, and high biodiversity. Evidence of human presence in the region dates back as early as the 10th century [37,41]. Santarém emerged as a result from the combination of settlements established by indigenous, Portuguese, and black populations. According to [37], the myth that it was an uninhabited region with infinite resources contributed to the migration of people to the Amazon region at different times, such as northeasterners escaping drought in 1915, rural settlers in the 1970s, and recent workers involved in infrastructure projects such as highways and reservoirs. As of 2021, Santarém has an estimated population of 308,339 inhabitants, covering an area of 17,898.389 km^2, with a population density of 12.87 inhabitants per km^2 [42].

According to research on Regions of Influence of Cities (REGIC) [43], Santarém is the most relevant municipality in the western region of Pará and is classified as a Regional Capital. Regional Capitals are urban centers with a high concentration of management activities relative to their region. Santarém holds a central position in the state of Pará and plays a crucial role in providing medium- and low-complexity health services to the surrounding municipalities. In addition, it is part of the Metropolitan Region of

Santarém, established in 2012, along with Belterra and Mojuí dos Campos. The area between these three cities is contested by various land uses, including agro-extractivist settlements, soybean plantations, rural communities, private condominiums, and housing developments [37].

Cametá served as the initial point for Portuguese colonization in the Baixo Tocantins region and became a significant facilitator of Catholicism expansion through Jesuit camps aimed at converting indigenous peoples [44,45]. The Portuguese colonization involved multiple attempts to enslave indigenous populations and import enslaved Africans, resulting in the formation of fugitive black communities known as Quilombos and indigenous communities in forested areas in the Cametá region, which were difficult to access. These communities played an important role in the region's territorial occupation. The land occupation in the area also involved the cultivation of cocoa and later sugar cane as strategic endeavors. In both cases, the Portuguese attempted to establish plantation systems based on monoculture, slave labor, and large land properties, with a focus on foreign markets [46].

According to REGIC [43], Cametá is classified as a Local Center. The estimated population of Cametá municipality was 140,814 inhabitants in 2021, distributed in an area of 3081 km^2 and a demographic density of 39.23 inhabitants per km^2 [42].

Both study areas exhibit distinct spatial dynamics of occupation, characterized by varying sizes, population growth rates, roles within the urban network, and proportions of urban and rural population. These differences make these cities ideal study areas for applying the classification model as they demonstrate how the model performs in cities with diverse characteristics. Furthermore, both cities are located in regions that offer potential opportunities for future field visits by researchers from the Laboratory of Socioenvironmental Investigations (LiSS) [47].
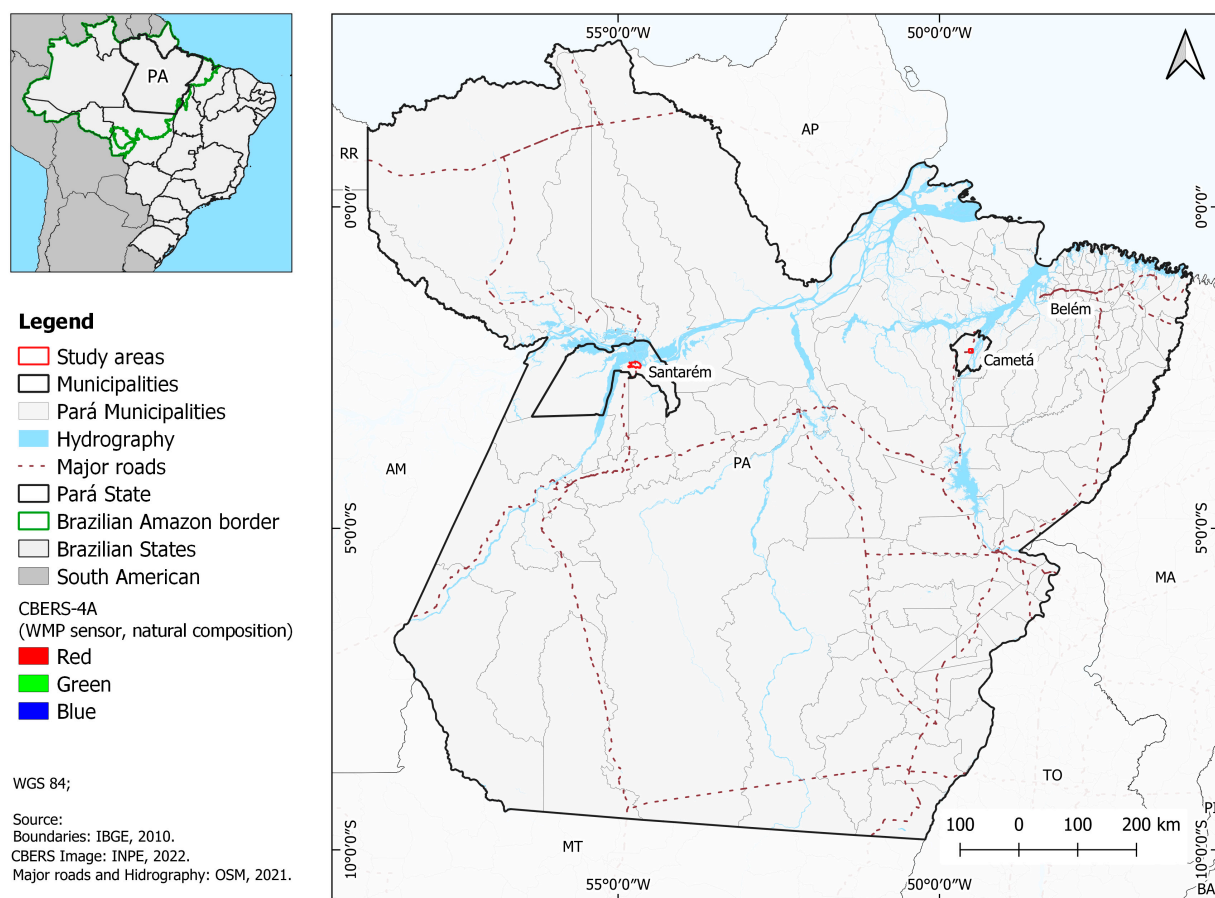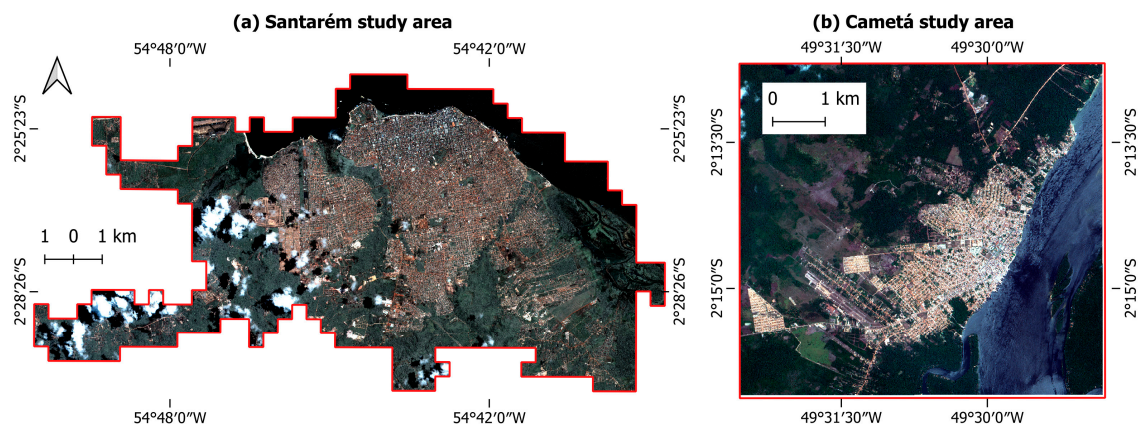

Figure 1. *Cont.*

(a) Santarém study area   **(b) Cametá study area**

Figure 1. Study areas: (**a**) Santarém and (**b**) Cametá [6,48,49].

4. Materials and Methods

The USEPs model classification is based on remote sensing and machine learning techniques, used to analyze environmental, urban morphological, and socioeconomic dimensions. The variables are formed based on assessment criteria and grouped under the three dimensions of analysis. Additionally, the approach differs from prior methods that solely concentrate on either urban morphology or socioeconomic factors as they strive to integrate both dimensions and incorporate an environmental focus.

The framework organizes a database and creates assessment criteria based on three dimensions of analysis. After defining the assessment criteria, variables are created based on these criteria and integrated into a cellular grid (Figure 2). Unsupervised classification techniques are applied to identify clusters in the data, initially using only environmental variables and later using urban morphology variables. Although these two dimensions may be related, the decision to split them into separate clusters is intended to highlight the unique environmental characteristics of Amazonian cities and to assess how these characteristics vary across different urban patterns. In this way, the two analysis dimensions are used as levels of information. The resulting clusters represent these information's levels and provide insights into the environmental and urban nature of the study area.
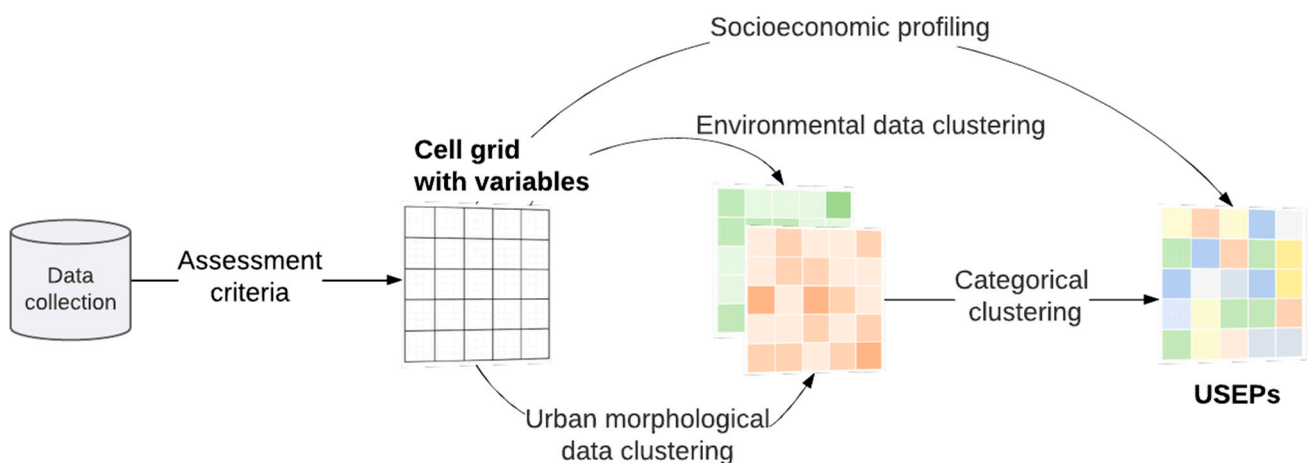
Figure 2. Methodology schema based on Santos et al. [18] to identify the Urban and Socio-Environmental Patterns.

Next, a new clustering is performed using only the results of the previous two clusterings but this time utilizing a categorical clustering algorithm. Categorical clustering is preferred due to the potentially large number of combinations between the two layers of information, making manual clustering impractical. At the end of this process, the

resulting clusters synthesize information from the environmental and urban morphological dimensions. Unsupervised classification is appropriate in this research as it allows for the classification of urban patterns directly from the data, without requiring prior knowledge of the study area or previous labelling of the analytical units. Ending the process, we profile the clusters with socioeconomic indicators.

We can summarize the USEPs model classification methodology in six steps. The first step defines data collection to analyze the cities. The second step consists of defining assessment criteria for the environmental, urban morphological, and socioeconomic dimensions. The third step involves creating the variables for the assessment criteria and integrating each variable into a cellular grid. The fourth step consists of obtaining clusters generated through unsupervised classifications and identifying the USEPs. In the fifth step, we use socioeconomic indicators and a decision tree algorithm to characterize the USEPs. Finally, the sixth step involves evaluating the classifications. We tested the methodology in Santarém and evaluated it in Cametá. This research included a field visit to Cametá during the evaluation process.

4.1. Data Sources

For this work, we used the following:

- WPM images from the CBERS-4A satellite: orthorectified images, one panchromatic and one multispectral, dated in 2020. The WPM sensor provides panchromatic and multispectral images simultaneously. The panchromatic images have 2 m of spatial resolution, with a spectral range between 0.45 and 0.90 μm. Multispectral images have a spatial resolution of 8 m, with spectral bands: blue (blue, 0.45–0.52 μm), green (green, 0.52–0.59 μm), red (red, 0.63–0.69 μm), NIR (near infrared, 0.77–0.89 μm). The radiometric resolution of the images is 10 bits. The imaged swath width is 92 km, and the revisit period is 31 days [48];
- Urban land cover classification from the amazonULC package [50]. The amazonULC package is a project that makes land cover classification maps available for some Brazilian Amazonian cities. Imagery from the CBERS-4A satellite's WPM sensor was used for a classification model that includes the GEOBIA method, data mining strategies, and the Random Forest machine learning algorithm. The maps present the following land cover classes: "Shrub Vegetation", "Herbaceous Vegetation", "Water", "Exposed Ground", "High Gloss Cover", "Ceramic Cover", "Fiber Cement Cover", "Asphalt Road", "Terrain Road", "Cloud" and "Shadow";
- Digital Elevation Models (DEM) and their derivations: a DEM (elevation), a slope grid (in percentage), and a vertical curvature grid of the relief, obtained from the TOPODATA portal [51]. The images have a spatial resolution of 30 m and a radiometric resolution of 8 bits;
- Road network: road data generated by Volunteered Geographic Information (VGI), extracted from OpenStreetMap [49], available in vector format (line) with different types of roads, bikeways, and pedestrian paths. As it is a VGI source data, the road network has no date information and no elaboration scale;
- Multitemporal GHSL-BUILT image: images from the Global Human Settlement Layer program, with multitemporal information about the built-up area [52]. The Global Land Survey (GLS) Landsat image collection (GLS1975, GLS1990, GLS2000, and Landsat 8) was the basis for this data construction. We used images with 30m spatial resolution, 8 bits radiometric resolution, classified into the following classes: built-up area before 1975, the built-up area between 1975 and 1990, the built-up area between 1990 and 2000, the built-up area between 2000 and 2014, no built-up area, water, and no data;
- Census data: data from the 2010 Demographic Census, provided by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) [6], available in comma-separated-values (.csv) and aggregated by census sector, in addition to the 2010 and 2020 census tracts, available in shapefile format (.shp).

The following software was used:

- QGIS 3.18 [53]: for database preparation and construction of the thematic maps;
- Python Programming Language [54]: for data preparation and mining, classifying clusters, and identifying the USEPs;
- TerraView 5.6.3, with GeoDMA 2.0.1 [55] and TerraHidro [56] plugins: for preprocessing the satellite images and generating the cell surface, extracting attributes, filling the cell surface, and building Height Above the Nearest Drainage (HAND).
- DepthMapX [57]: for the construction of axial maps.
- Figure 3 shows how the different data and software were used to identify the USEPs in the study areas.
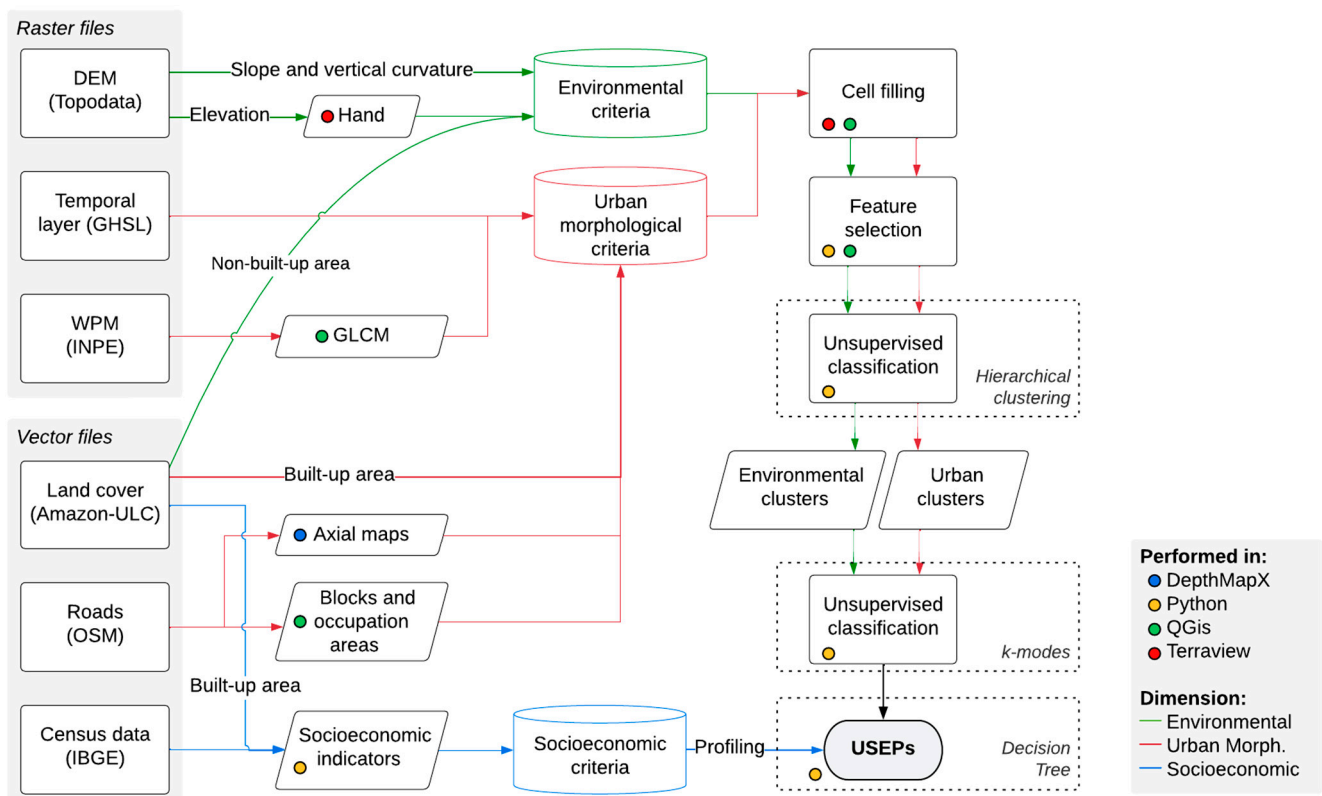

Figure 3. Expanded methodology schema to identify the Urban and Socio-Environmental Patterns.

4.2. Assessment Criteria

For each dimension, environmental, urban morphological, and socioeconomic, we defined assessment criteria to build the variables. The criteria seek to emphasize the main aspects of each dimension that are relevant to Amazonian cities.

4.2.1. Environmental Dimension

The environmental dimension helps identify urban patterns in Amazonian cities, with a focus on periurban characteristics and addressing limitations of morphological analysis methodologies. Two competing urban patterns are recognized in the region [37]: one rooted in traditional indigenous knowledge and the other reflecting urban-industrial practices causing environmental degradation. Traditional settlements tend to have closer proximity to rivers and more vegetation, while urban-industrial settlements have more significant changes in topography and higher building densification.

As criteria for the environmental dimension, we use the following: (a) area of the land cover classes "Shrub Vegetation", "Herbaceous Vegetation", "Water", and "Exposed Ground"; (b) slope (in percent) and the vertical curvature base; and (c) the HAND built with the DEM image and according to Rennó et al. [58]. The land cover classes, such

as vegetation and water, have a direct impact on human health and provide important ecosystem services, such as regulating pollution and microclimates [59]. The slope indicates areas prone to flooding [60], while curvature helps understand erosive and hydrological processes that affect hillside orientations [61]. HAND indicates the elevation of settlements and their potential flood risk [58].

4.2.2. Urban Morphological Dimension

The urban morphological dimension provides an understanding of the physical form of the cities, including their building patterns, streets, lots, and blocks. It also enables the identification of key players and processes involved in city formation. The morphological analysis methods proposed are adapted from Conzen [62,63] and Morpho [64], taking into account the unique characteristics of Amazonian cities, which differ greatly from traditional industrialized societies.

Therefore, we build variables based on the following elements of analysis: Accesses, which consider road infrastructure and waterways as means of transportation; Blocks and Occupation Areas, which are defined based on the road network and river boundaries and include occupation areas where block boundaries are not present; Roofs, which replace the traditional "building" element in the morphological analysis due to the difficulty of obtaining high-resolution building imagery.

As in the case of Morpho [64], we provide bases for integrated morphological analysis, not focusing on a specific element of the urban form. Figure 4 presents the evaluation criteria and how they are related to the urban elements.
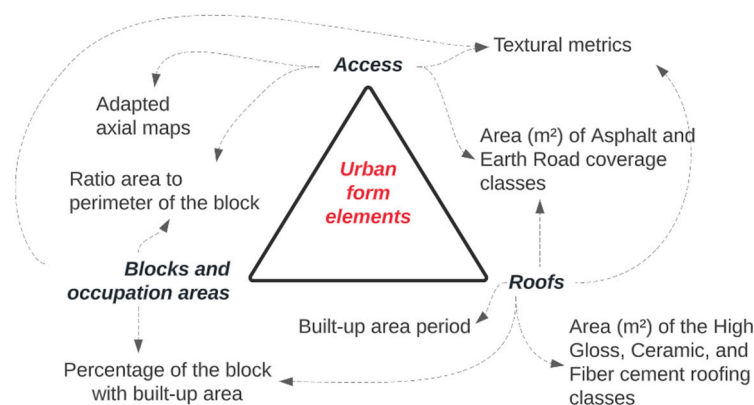

Figure 4. Evaluation criteria of the urban morphological dimension, based on the following elements: Accesses, Blocks and Occupation Areas, and Roofs.

We used Spatial Syntax to analyze spaces of movement, as well as social and economic activities. This approach utilizes techniques and computational models to associate quantitative values with mathematical expressions for spatial analysis. The technique quantifies the relationships within a road network, revealing areas of natural movement where the configuration of the road network itself has the potential to concentrate the movement of people, benefiting developments and other activities [65]. It is also possible to assess the accessibility potential of roads using axial lines, which represent the intersection of line segments in the road system of the studied area [66,67]. Table 1 presents the assessment criteria used in this dimension, with their descriptions.

Table 1. Assessment criteria of the urban morphological dimensions and their descriptions.

Assessment Criteria	Description
Adapted Axial Maps	We applied adapted axial maps of integration and choice on a regional scale using a 50 km buffer around the study area. The metrics were developed using the Access base, consisting of the road network and perennial rivers. Rivers were represented by a hexagonal grid with 500 m and 2 km hexagons. The following parameters were used: Number of bins—16, Metrics—choice, integration, node count, and total depth, Radius—500 m, 1 km, 5 km, and all accesses.
Ratio area to perimeter of the block	Jacobs [68] states that short block lengths lead to more diversity of use along the streets and make it easier for the population to move around, fostering diversity of use of buildings.
Roofing Class Area	The area of the High Gloss, Ceramic, and Fiber cement roofing classes identifies areas with a higher or lower proportion of roof types among settlements and areas of higher or lower building density.
Built-up Area Period	We use the built-up area period to reflect the urban palimpsest, which refers to overlapping periods of construction reflecting the ideologies that guided land use over time. The urban form shows the record of civil and public actions. The period of the area was obtained from the GHS-BUILT Multi-temporal database [52].
Percentage of Block with Built-up Area	Identifies areas of higher and lower building density.
Road Coverage Class Area	The area of Asphalt and Terrain Road coverage classes identifies the proportion and type of roads in settlements.
Textural Metrics	Textural metrics obtained from Gray Level Co-occurrence Matrices (GLCMs) [69] are used by several authors to identify urban differences at the local scale.

4.2.3. Socioeconomic Dimension

The socioeconomic dimension characterizes the conditions of the households, their surroundings, and the resident population. This dimension uses the socioeconomic indicators proposed by Santos et al. [70] after adaptations from indicators provided by the IBGE [8] (Table 2). These adapted indicators take into account cultural practices and experiences with alternative water and sewage treatment techniques to establish what constitutes adequate housing conditions in the region, adopting only the base of the demographic census universe aggregated at the census tract level.

Table 2. Indicators of the socioeconomic dimension and their descriptions.

Assessment Criteria	Description
Percentage of households with inadequate sewage disposal	We adopted the following as inadequate: the landfilling of garbage on the property; thrown garbage in empty lots, rivers, lakes, seas, or other destinations without collection.
Percentage of households with inadequate water supply	We adopted the following as inadequate: the existence of households without a toilet; the disposal to a rudimentary septic tank; the disposal to a ditch; the disposal to a river, lake, sea, or other destinations without collection.
Percentage of households with adequate garbage disposal	We adopted the following as inadequate: water supply by well or spring outside the property (or village), rainwater stored in a way other than in cisterns, and supply only by tank trucks.
Total people per household	Number of people divided by the number of households
Average monthly income of private households	Total monthly household income divided by the total number of residences.
Young Dependency Ratio	Total young population (under 15 years old) divided by the economically active population (between 15 and 65 years old).

Table 2. *Cont.*

Assessment Criteria	Description
Old Dependency Ratio	Total elderly population (over 65 years old) divided by economically active population (between 15 and 65 years old).
Sex Ratio	Total male population divided by the total female population
Household heads who are over 10 years old and literate	Population responsible for the household over 10 and literate.

4.3. Creating and Integrating Variables into Cell Grids

After defining the assessment criteria, we integrated all data into a regular grid of 100×100 m square cells (10,000 m^2). These cells were used as reference units to aggregate data from various sources. The cell space was created from the IBGE statistical grid [71] using a cell size commonly used to identify urban patterns [72–75]. We decided to use a cellular grid because this spatial structure can be easily generated for any study area and is the most suitable spatial unit to integrate data from different formats and sources. Furthermore, cells have the advantage of being spatial temporally stable because they are not subject to changes in their physical boundaries [76].

To fill the cells, we extracted the landscape metrics "Class area" (measured in ha) and "Number of patches" from the land cover base using the GeoDMA plugin [55]. To avoid misclassification, we excluded areas that were blocked by clouds and/or shadow in the remotely sensed images, when the blocked area was greater 0.4 ha. After this, with the cell grid prepared, we separated the cells that had the presence of built-up areas (occupied area) and the cells without the presence of built-up areas (non-occupied area) (Figure 5).
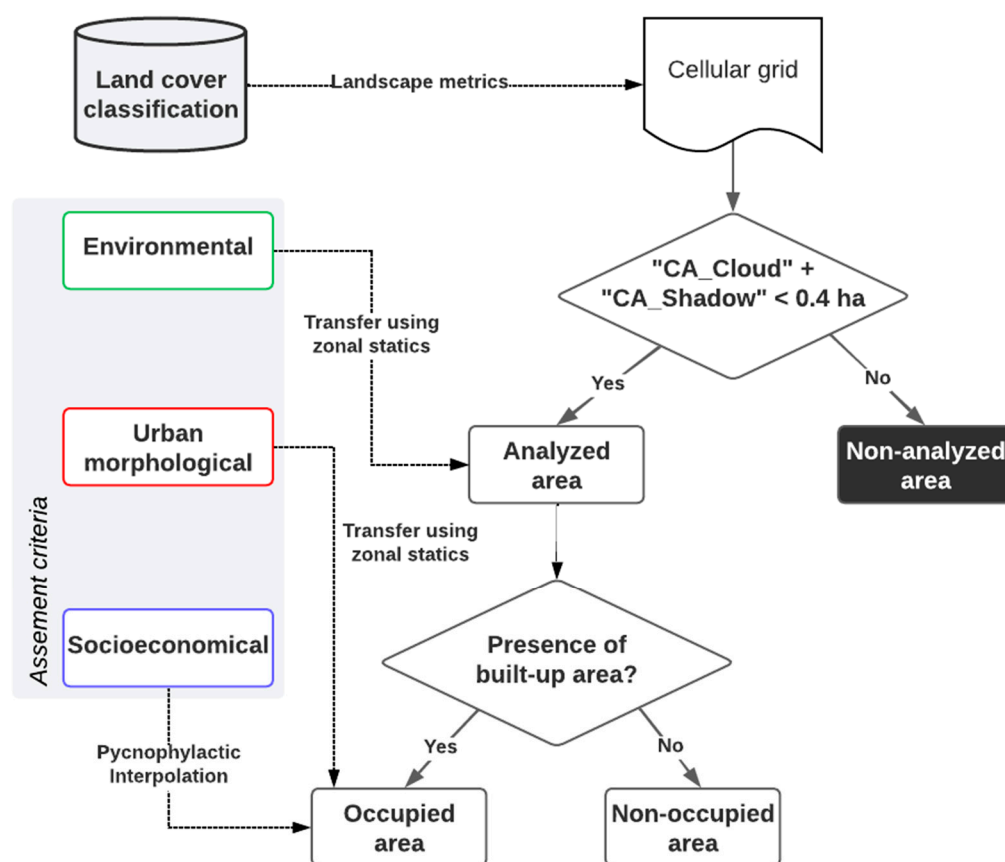

Figure 5. Division of the cell grid into occupied and unoccupied area ("CA_Cloud" and "CA_Shadow" are the "Class area" of the "Cloud" and "Shadow" classes within the cell).

Next, we transferred all data to the grid cell. Data related to urban morphological or socioeconomic dimensions were transferred only to the occupied area, while data from the environmental dimension were transferred to all areas analyzed. Socioeconomic data were transferred grid cells using Pycnophylactic Interpolation [77], following the methodology of Santos et al. [70]. This technique allows us to change the unit of support while preserving the total mass of the variable. For vector-type data, we adopted the area-weighted average operator for polygons and average for lines, both methods available in the GeoDMA plugin [55]; for variables in matrix format, we extracted the values of mean, maximum, minimum, sum, standard deviation, range, and variance [53].

After assembly of all data in grid cells, we filled in possible missing values and normalized the variables, removing highly correlated variables. At this point, we generated a map for each feature and eliminated the variables that showed limited ability to differentiate between clusters based on a visual analysis. The outcome of this process was eight environmental variables, eleven urban morphological variables, and nine socioeconomic variables.

4.4. Clustering Process

To identify clusters in the environmental and urban morphological dimensions, we compared several unsupervised classification algorithms, namely k-means, Self-Organizing Maps (SOMs), and hierarchical clustering. After comparing the results, we found that hierarchical clustering [78] was the most effective approach. While k-means and Self-Organizing Maps classified homogeneous settlements with significant cluster variability, hierarchical clustering had higher generalization capabilities, resulting in almost the entire settlement falling within the same urban pattern type.

Furthermore, hierarchical clustering is easier to parameterize because it does not require a previously defined number of clusters, unlike k-means. Additionally, hierarchical clustering was less sensitive to parameter changes compared to Self-Organizing Maps. The hierarchical clustering algorithm works by creating a dendrogram, a tree-like representation of instances where the leaves represent the instances, and their clustering is based on similarity. The clustering progresses until all instances are connected to form a single trunk. The analyst then evaluates the dendrogram and decides how many clusters to create by cutting the tree at a desired point. We defined the number of clusters after analyzing the dendrograms and plotting the cluster maps. We used the Scikit-learn python library [79] with parameters set to agglomerative clustering, ward linkage, and Euclidean affinity.

After clustering of environmental and urban morphological data, we have two categorical variables that need combining. Following the approach of Cunha et al. [80], we decided to use the k-modes algorithm to cluster our categorical variables. The k-modes is a clustering method developed by Huang [81] based on the evaluation of cluster modes instead of cluster means. The k-modes algorithm determines clusters from categorical data by using the simple matching distance as a measure of dissimilarity and minimizing the cost function by adjusting the modes of clusters. The simple matching distance between two vectors can be calculated as the total number of mismatches between the corresponding attribute categories of the two objects. The "iteration" parameter in the k-modes algorithm determines the number of rounds the algorithm undergoes to converge on a solution. During each iteration, the algorithm updates the cluster assignments and centroids. This process continues until a convergence criterion is satisfied, indicating that the algorithm has reached a stable solution. For the estimation of the clustering algorithm, we used the k-modes python library [82], parameters set to 300 iterations, and "Hang" as the method for initialization. The clusters generated by the k-modes algorithm integrate the information from the environmental and urban morphological dimensions.

4.5. Socioeconomic Profiling

The result of the procedure described above is a set of environmental–urban clusters. Following Victoriano et al. [83], we used a decision tree to obtain socioeconomic profiles of the clusters. Decision trees are advantageous due to their simple, explicit, binary partitions,

which make them a "white-box" model. We treated the environmental–urban clusters as a categorical dependent variable and the socioeconomic indicators as the independent variables that determined the partitions. The outcome of the decision tree was a partition of the sample into terminal nodes that gave the socioeconomic profile of each cluster.

To define the parameters of the decision trees, we used a grid search method with cross-validation to determine the parameters that return the best accuracy value. In this technique, several hyperparameters and many values are selected to be tried. The algorithm evaluates all possible combinations between the different hyperparameter values, using cross-validation and a performance measure. Grid Search uses cross-validation, dividing the training base into k parts (folds), and the model is trained and evaluated k times. For each iteration, the algorithm selects a part (fold) that will serve as an evaluation and trains the model on the other k-1 parts.

Through the grid search, we selected the best decision tree to profile the clusters. We adopted a maximum value of 100 parameter combinations (iterations), F1-Score as a performance metric, and 5-fold cross-validation. For the decision tree estimation, we used the Scikit-learn python library [79].

4.6. Evaluation

Given that the patterns were derived from an unsupervised classification technique, our study sought to ascertain the semantic significance of each USEP. Specifically, we investigated the potential for the analytical explanation of the USEPs and their capacity to represent the various typologies in Santarém and Cametá, based on the scientific literature [37,84,85].

According to Cardoso and Lima [86], the land register aspect has determined the urban expansion in Amazonian cities by the conversion of areas of rural use into urban areas. Thus, we used the 2010 and 2020 census track bases to evaluate the existence of one (or more) USEPs in areas classified as rural in 2010 but urban in 2020. Therefore, from the evaluation with the overlap of the 2010 and 2020 census tracks, we adopted the manual decision tree in Figure 6, classifying the cells between "Old urban" (areas considered urban already in 2010), "New urban" (areas considered urban only in 2020), and "Rural" (rural areas in 2020).
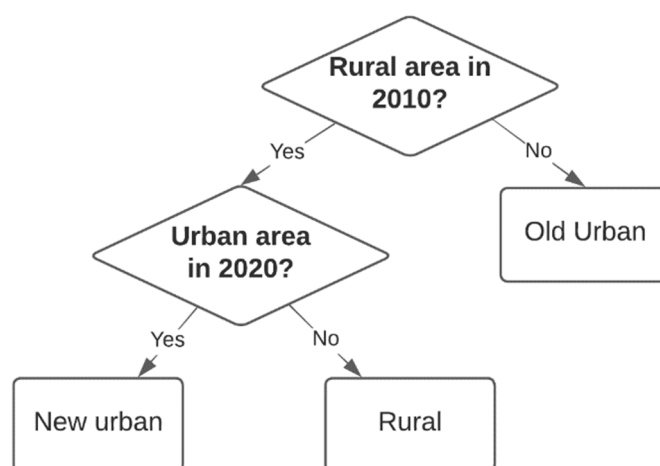

Figure 6. Decision tree for classifying the cellular grid into "Old urban", "New urban", and "Rural".

In addition, we also used 2020′s AGSN (from the Portuguese *Aglomerados Subnormais*), a dataset on irregular settlements, to assess the existence of patterns that can be considered precarious. According to the IBGE [87], an AGSN is a form of irregular occupation of land owned by others—public or private—for residential purposes, characterized by an irregular urban pattern, lack of essential public services, and location in areas with occupation restrictions. The AGSNs are distributed both in rural and urban areas.

To further validate our findings, we conducted a field visit to Cametá. The field visit took place between 17 and 18 November 2022. The field team used a field sheet to collect information about the density of the settlements, the width and condition of the streets, street arborization, street lighting conditions, and materials used in the building of the houses. GPS equipment and cameras registered the geographical coordinates and general aspects of the physical environment, respectively.

5. Results

5.1. USEPs Identification in Santarém

Concerning the Santarém study region, 93% of the area was deemed suitable for analysis, while the remaining 7% could not be analyzed due to the prevalence of clouds and/or shadows in the Amazon-ULC land cover database. Of the area analyzed, 46% was classified as "occupied" (63 km^2), indicating that the cells contained a built-up area according to the Amazon-ULC land cover database. The remaining 54% of the analyzed area was classified as "non-occupied", comprising areas of dense vegetation or rivers.

A series of experiments and visual analyses were carried out to ascertain the optimal number of clusters. Through the examination of the dendrograms (Figure 7), we identified six clusters for the environmental dimension and seven clusters for the urban morphological dimension. We determined this based on the significant reduction in the Euclidean distance, which served as the dissimilarity index, with a reasonable number of observations in each cluster, and a visually coherent spatial pattern in the study area.
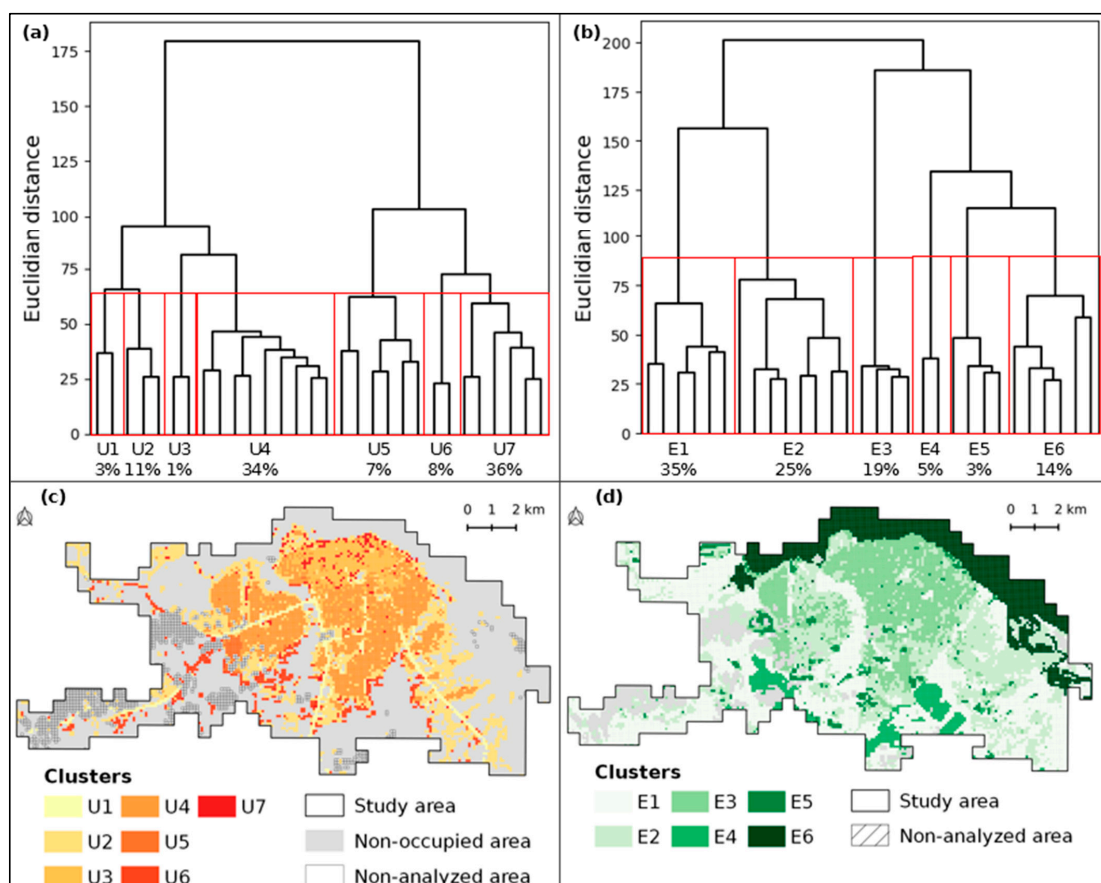

Figure 7. Dendrograms and maps used to define the ideal cluster quantity: (**a**) dendrogram of the urban morphological dimension variables; (**b**) dendrogram of the environmental dimension variables; (**c**) map with the distribution of urban morphological dimension clusters; (**d**) map with the distribution of environmental dimension clusters. The red boxes show the clusters obtained in each unsupervised classification.

Upon deciding the number of environmental and urban morphological clusters, we applied the k-modes algorithm to conduct an unsupervised classification over the two categorical variables based on the prior clustering. After several experiments, we settled for seven final clusters for Santarém, considering the best trade-off between identifying meaningful clusters and their number. Each final cluster can be interpreted as a distinct USEP, with the socio-economic profiling conducted in the subsequent subsection. Figure 8 presents the USEP map for Santarém.
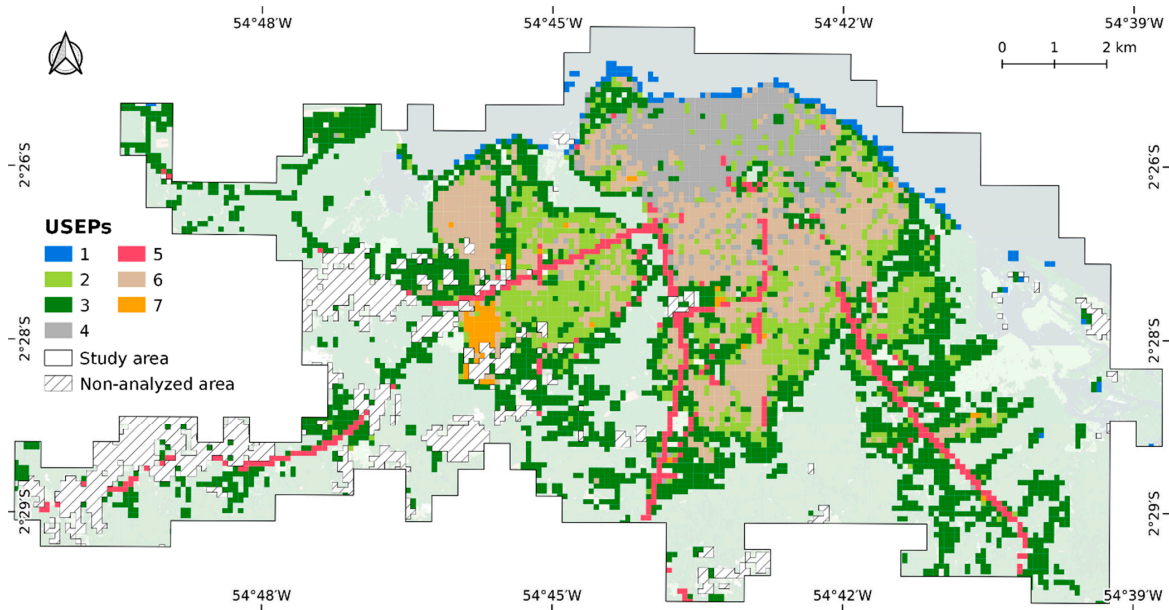
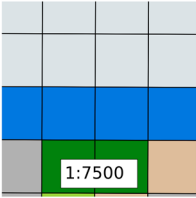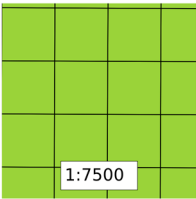**Figure 8.** Map of the USEPs identified for the city of Santarém.
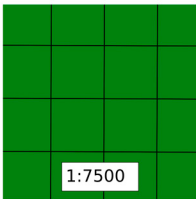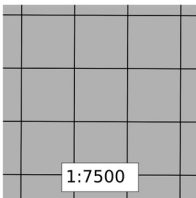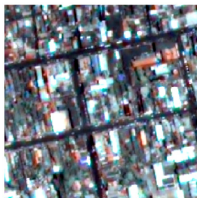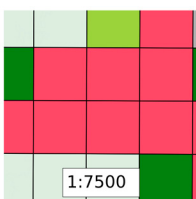
5.2. Profiling the USEPs in Cametá

In order to designate the various Urban Socio-Environmental Patterns (USEPs), Figure 9 illustrates the mean values of the selected features employed during the clustering process. The distinct shades indicate the relative variability of the features across the USEPs. The seven clusters were labeled based on their predominant environmental and urban morphological characteristics, namely, "USEP 1–Riverside", "USEP 2–Medium-density", "USEP 3–Periurban", "USEP 4–High integration", "USEP 5–Main roads", "USEP 6–High-density informal", and "USEP 7–Housing complex". Table 3 presents a detailed description of the environmental and urban morphological characteristics of each USEP in Santarém.

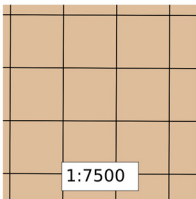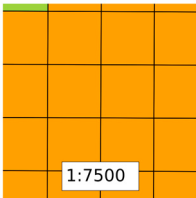| USEP	HAND (m)	Water (ha)	Shrub vegetation (ha)	Herbaceo vegetation (ha)	Access choice (10e6)	Ceramic (ha)	Fiber cement (ha)	High gloss (ha)	Area per perimeter of the block	Building density	Terrain roads (ha)
Riverside	5.96	0.78	0.05	0.03	116	0.00	0.04	0.04	8.42	0.09	0.00
Medium-density	14.59	0.00	0.07	0.36	35	0.06	0.13	0.02	36.33	0.52	0.26
Periurban	13.45	0.01	0.43	0.27	51	0.02	0.03	0.00	95.40	0.16	0.08
High integration	13.18	0.01	0.01	0.08	150	0.04	0.36	0.12	33.78	0.83	0.07
Main roads	12.64	0.00	0.30	0.23	794	0.03	0.07	0.01	88.14	0.27	0.12
High-density informal	16.13	0.00	0.01	0.11	67	0.07	0.27	0.01	27.39	0.80	0.38
Housing complex	9.47	0.00	0.01	0.03	7	0.70	0.02	0.01	30.72	0.84	0.14

Figure 9. Average values of a selected group of features for each USEP. The variation of hue in each column represents the variation of attribute values among the clusters.

Table 3. USEPs of Santarém, comparison with the natural composition WPM image (2 m spatial resolution), and their description.

USEP	WPM RGB (3, 2, 1)	Pattern Description
		USEP-1: Riverside. This pattern is characterized by a built-up area along the waterfront with high connectivity and infrastructure for navigation. Most of the area is occupied by the land class "water."
		USEP-2: Medium-density. This pattern is located on the periphery of Santarém's urban area with low integration and accessibility. It has medium building density and rectangular-shaped blocks with buildings that have fiber cement roofs. The area has a significant herbaceous vegetation cover, which presents about 35% of the cell, and is situated at an average slope of 5.3%.
		USEP-3: Periurban. This pattern is situated far from the center of Santarém, usually bordering highways or close to the river, presenting moderate accessibility. It has low building density and no specific observed roofing type. The area has the highest vegetation cover among the identified patterns, with shrubs and herbaceous vegetation accounting for over 40% of the area. It is located at an elevation of 13.5 m and features the steepest slope among the identified patterns, measuring 6.4%.
		USEP-4: High integration. This pattern is located in central areas close to the Santarém waterfront, boasting highly integrated access routes and well-maintained asphalt roads. It has small, densely built blocks with a regular shape and a high proportion of buildings with fiber cement roofs. The area is situated on flat terrain, with a slight slope of 4.3%, and features large warehouses, suggesting commercial and logistic activities in the region.
		USEP-5: Main roads. This pattern pertains to the primary interconnecting highways in the wider region outside of Santarém's city center. This pattern features both asphalt and dirt road access, with the highest level of connectivity and integration. In general, the area closer to the urbanized zone exhibits regular, densely built blocks, whereas building density and conformity decrease closer to rural areas.
		USEP-6: High-density informal. This pattern is located on the periphery of downtown Santarém, with expansion largely guided by highways. Access is moderately integrated, with most of the roads being dirt roads. This pattern exhibits medium regularity and high-density blocks, with block sizes varying between 100 and 220 m in length. It features a high proportion of buildings with fiber cement roofs, but buildings with ceramic tops are also present. It has no vegetation within the blocks, situated on flat terrain, with the highest average elevation (16 m) among the identified patterns. The high building density, irregular block shapes, and unpaved streets suggest an informal settlement type.
		USEP-7: Housing complex. This pattern describes a recent housing complex located far from the city center, built after 2010, in areas that were recently converted from rural to urban use. Access to these areas is poorly integrated, and the roads are paved. The developments in this class consist of regular, large blocks with high construction density. The buildings are generally small and lack backyards, featuring ceramic roofs, with some exceptions that have high-gloss roofs. This pattern is situated in areas of low elevation, on flat terrain with no slope. Vegetation within the blocks is absent.

To obtain socioeconomic profiles of the USEPs, a decision tree was trained. Due to the COVID-19 pandemic, the most recent and available demographic census data in Brazil is from 2010. Because of this, it was necessary to exclude "USEP 7–Housing complex" from the socioeconomic analysis, as this urban pattern was created after 2010. We also excluded "USEP 1–Riverside" and "USEP 5–Main roads" from the analysis, as both especially compromise the access and uninhabited areas, referring to the river and the roads, respectively.

Figure 10 presents the results of the decision tree. The tree was partitioned a total of 13 times, resulting in 14 terminal nodes with different compositions of urban patterns. The first partition of the tree separates the patterns in terms of household income, separating to the left the samples with average income below BRL 1207.41 and to the right the samples with average income above this value. On the left side of the tree, the "USEP 2–Medium-density" and "USEP 3–Periurban" patterns are predominant, while on the right side, the "USEP 4–High integration" and "USEP 6–High-density informal" patterns are predominant.

Decision tree nodes:

- Household income R$ 1,207.41 — Less than / Greater than
- Left: Dependency (old) 4.5% | Right: Dependency (young) 22.5%
- Left branch: Household income R$ 757.93 ; Dependency (old) 5.3%
- Right branch: Dependency (old) 5.8% ; Sex ratio 105.8%
- Terminal-level nodes: Inadequate sewage 98.78% | Inadequate sewage 28.70% | Dependency (young) 27.7% | Inadequate sewage 5.17% | Household density 4.18 p. | Inadequate garbage 2.35% | Inadequate sewage 66.20%

Node composition (columns of terminal nodes, left to right):

USEP															
USEP-2	7%	91%	53%	40%	6%	13%	2%	6%	19%	8%	21%	27%	17%	13%	33%
USEP-3	89%	9%	29%	40%	75%	19%	98%	15%	15%	4%	38%	2%	58%	79%	8%
USEP-4	1%		2%	1%	1%			79%	30%	80%	15%	17%	6%	1%	
USEP-6	4%		16%	19%	19%	69%			36%	8%	26%	55%	19%	6%	58%
	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Row proportions:

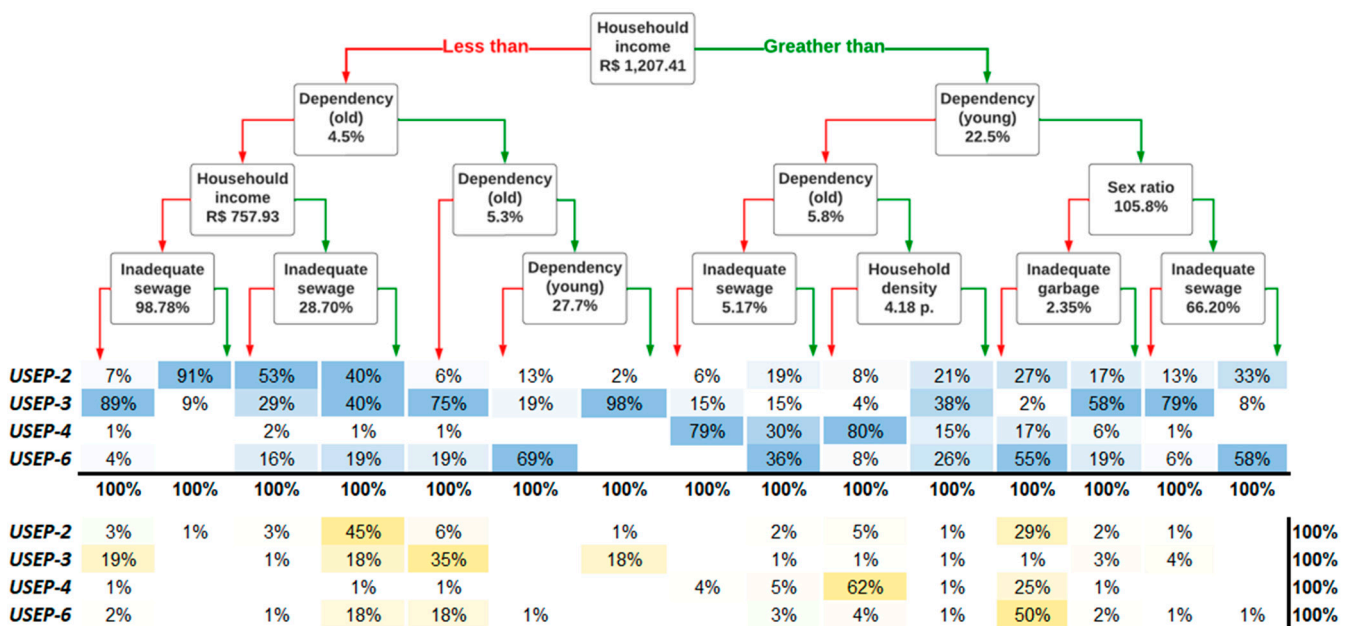| USEP																Total
USEP-2	3%	1%	3%	45%	6%		1%		2%	5%	1%	29%	2%	1%		100%
USEP-3	19%	1%	18%	35%	18%				1%	1%	1%	1%	3%	4%		100%
USEP-4	1%		1%	1%			4%		5%	62%	1%	25%	1%			100%
USEP-6	2%	1%	18%	18%	1%				3%	4%	1%	50%	2%	1%	1%	100%

Figure 10. Decision tree used to characterize the USEPs of Santarém based on socioeconomic indicators.

On the left side of the tree, where the average household income is below BRL 1207.41, new partitions help to differentiate the patterns. In the first two branches from left to right are the samples that present an elderly dependency ratio below 4.5% and an average household income below BRL 757.93—these two values are the lowest registered in the entire tree. After that, the samples were split based on the inadequacy rate of the sewage treatment service between the "USEP 2–Medium-density" and "USEP 3–Periurban" patterns. However, the last division related to the inadequacy of sewage treatment starts from a very high value (98.78%), so it is possible to see that both patterns have high inadequacy rates of this public service.

The "USEP 3–Periurban" category tends to have an elderly dependency ratio of between 4.5% and 5.3%, indicating a low number of elderly (people above 65 years old) relative to the total population. In addition, this pattern is associated with a youth dependency ratio above 27.7%, indicating a high number of young people (under 15) relative to the total population. On the left side of the tree, the "USEP 4–High integration" category is practically non-existent in the branches. In turn, "USEP 6–High-density informal" appears with a high proportion only in the branch referring to an elderly dependency ratio above 5.3% and a youth dependency ratio below 27.7%. These two patterns appear more frequently on the other side of the tree, where the average household income exceeds BRL 1207.41.

"USEP 4–High integration" has a strong presence in the branch of higher income, with the lowest youth ratio (below 22.5%) and lowest sewage treatment service inadequacy ratio (below 5.17%), representing almost 80% of this branch. It also has an 80% representation in the branch with a lower proportion of young people (below 22.5%), a higher proportion of elderly (above 5.8%), and low household density (below 4.18 persons per household).

"USEP 6–High-density informal" tends to have a higher proportion of young people than "USEP 4–High integration", but still lower than that recorded among "USEP 2–Medium-density" and "USEP 3–Periurban". It also has a strong presence in the rightmost branch (almost 60%), demarcated by a youth dependency ratio above 22.5%, a sex ratio above 105% (meaning a higher proportion of males compared to females), and a sewage treatment inadequacy rate above 66.20%.

By analyzing Figures 10 and 11, "USEP 4–High integration" presents the greatest urbanity of the study area. As is typical in the central regions of urban areas, "USEP 4–High integration" has better income conditions and sanitation services, as well as lower domiciliary density. Concerning the resident population, this pattern tends to have a higher proportion of the elderly population and a lower proportion of the young population when compared to the other patterns.
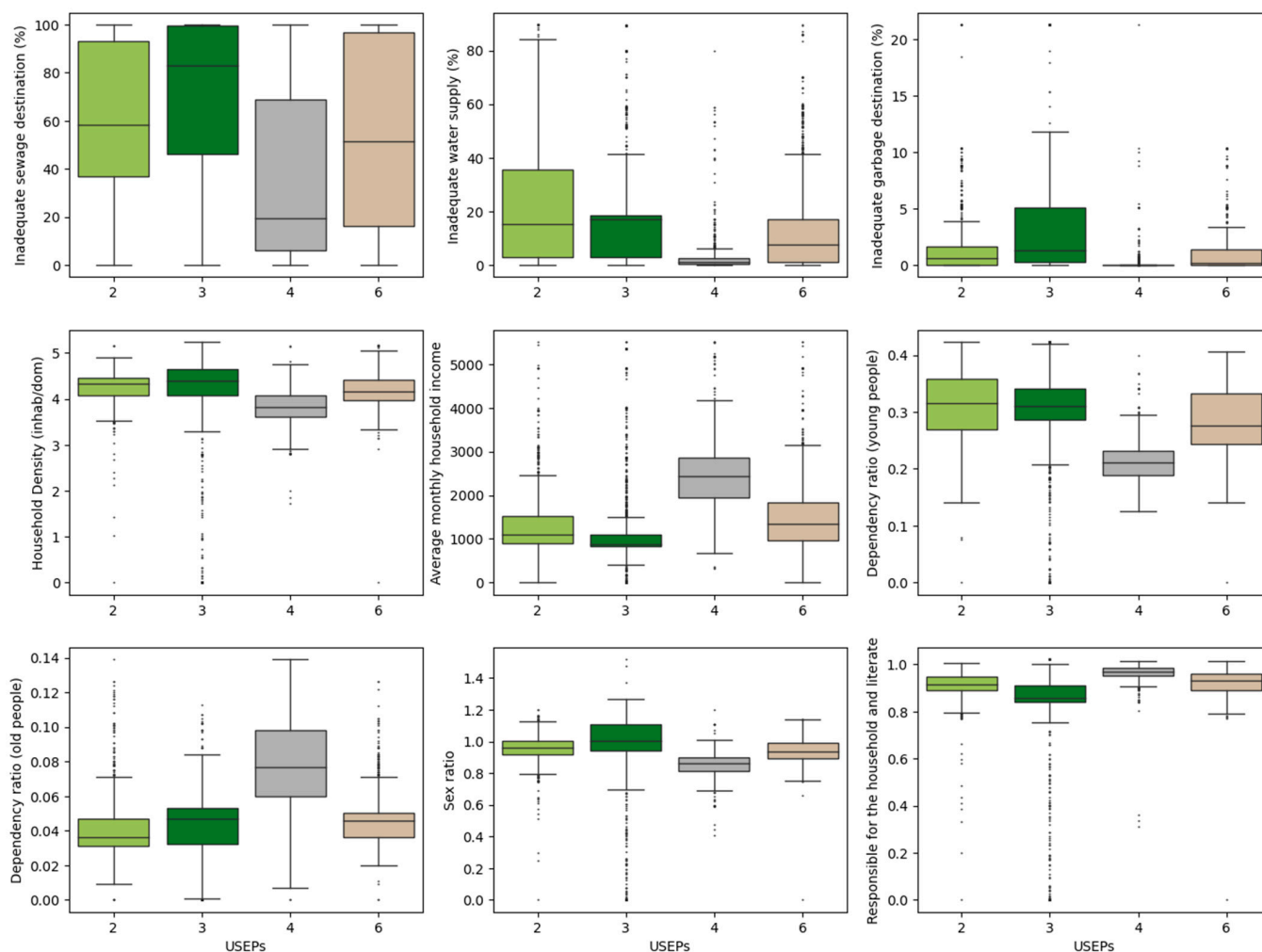

Figure 11. Bloxpot of the socioeconomic indicators of "USEP 2–Medium-density" (light green), "USEP 3–Periurban" (green), "USEP 4–High integration" (grey), and "USEP 6–High-density informal" (light brown).

Household income and sanitation services both decrease as the patterns become more rural, as seen in "USEP 3–Periurban". In this case, there is a lower fraction of the elderly

population and a higher fraction of youth in relation to the total population. In addition, the higher sex ratio makes explicit a higher proportion of the male population when compared to the female population. Regarding household density, this pattern presents the highest amount of people per household.

Thus, "USEP 2–Medium-density" and "USEP 6–High-density informal" are patterns of intermediate characteristics, with "USEP 2–Medium-density" approaching the profile of "USEP 3–Periurban" and "USEP 6–High-density informal" approaching the profile of "USEP 4–High integration". "USEP 6–High-density informal", although being in a more urbanized area, presents low levels of sanitation services. Approximately 50% of this pattern does not have access to adequate sewage treatment or disposal systems. In addition, 75% of the households have an inadequate water supply which is comparable to the rural area of Santarém.

5.3. USEPs Identification for Cametá and Field Visit

The methodology was replicated for the city of Cametá, and we found eight USEPs (as shown in Figure 12), one more than in Santarém. Compared to Santarém, the city of Cametá has a lower density, less urban sprawl, and smaller population size. The highway in Cametá was built parallel to the river, causing urban growth to be concentrated along the sides of the city and parallel to the river, with little penetration into the interior of the territory. Despite the differences between the two cities, we identified some similarities in the urban patterns formed by "USEP 3–Periurban", "USEP 4–High integration", and "USEP 5–Main roads". These patterns have similar characteristics in terms of integration and construction density in both cities.
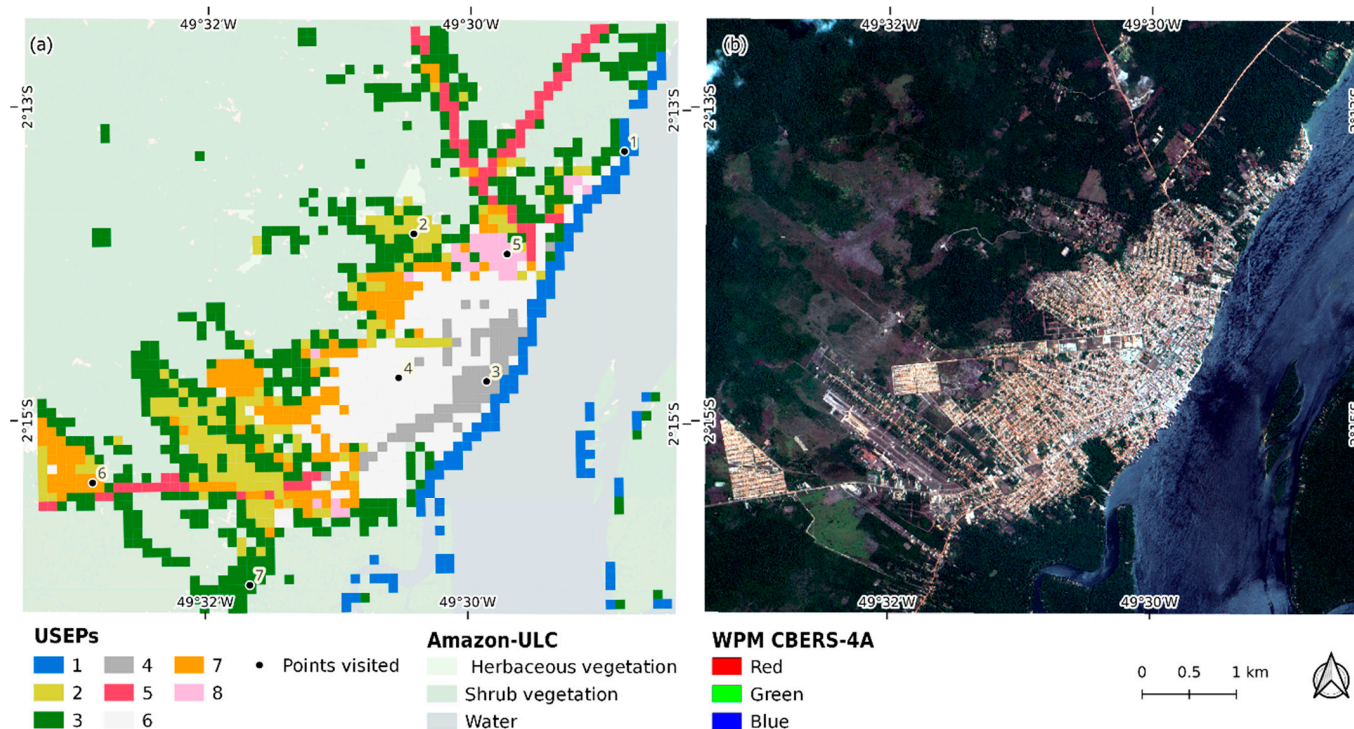

Figure 12. (**a**) Map of the USEPs identified for the city of Cametá (**b**) and WPM image.

We also identified specific urban patterns for Cametá, such as "USEP 1–Riverside and riparian communities", "USEP 2–Low-density", "USEP 6–Medium-density", "USEP 7–Housing complex", and "USEP 8–Medium-density informal". However, all urban patterns in Cametá, except for "USEP 4–High integration", have low levels of attendance to sanitation services. This lack of coverage in public services for water supply, sewage

treatment, and garbage collection is common throughout the city, regardless of whether it is a consolidated urban area or a more rural area.

However, the "USEP 1–Riverside and riparian communities" include areas with both high and low coverage of sanitation services. This is because this urban pattern's main characteristic is the presence of water in the inner cell, encompassing the more touristic edge of the river, as well as riverside houses with exclusive access by boats. Unlike the "USEP 1–Riverside" pattern of Santarém, the "USEP 1–Riverside and riparian communities" not only includes areas with infrastructure for navigation but also inhabited areas. It is a USEP of high integration and connectivity of its accesses, motivated mainly by the accessibility provided by the river. In the riverfront area, there are high-end housing buildings separated into lots within blocks, and there are shops. There is vegetation present on the riverfront and within the lots.

On the islands near the town of Cametá, "USEP 1–Riverside and riparian communities" includes wooden houses built on the banks of the Tocantins River in a traditional riverine pattern (Figure 13). The river is the main way to access the city and other settlements on the mainland. In general, these houses are inserted within a community formed by other houses in a floodable area, around a community center. The house is positioned facing the river, and behind the house, there is a yard. After the yard, there are working areas dedicated to extractive activities where a single crop is generally planted, usually açaí. The figure shows residences in the riparian settlements and visualization on satellite images.
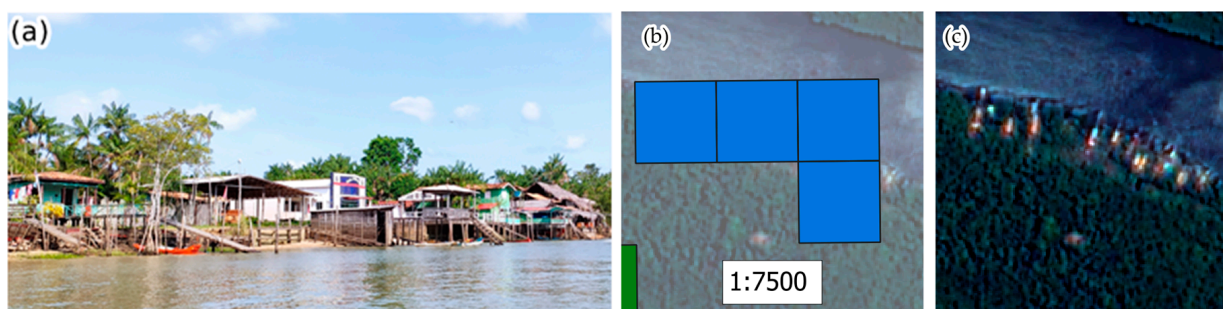

Figure 13. (**a**) Photograph of houses in a riparian community; (**b**) Result of our classification and identification of "USEP 1–Riverside and riparian communities"; (**c**) Visualization of the same cells of 11 (**b**) on the CBERS-4A satellite WPM image (2 m spatial resolution, natural composition).

Regarding housing conditions, "USEP 1–Riverside and riparian communities" has the highest housing density (about 4.80 persons per house) along with "USEP 3–Periurban". It also has a high proportion of the young population (28.5%) and the second-highest proportion of the elderly (5.9%) among the standards.

"USEP 2–Low-density" is characterized by areas of low building density, with built-up area classes representing only about 13% of the block or the occupation block on average. These areas are recent, spontaneous, and of low financial standards, resulting from the conversion of rural to urban use. Houses are made of unfinished masonry or good-quality wood, and public lighting is provided clandestinely by residents using spotlights on rustic electric poles (Figure 14b). These areas are not integrated with the rest of the city, with dirt roads as the only means of transportation. This pattern has low-income levels (an average of BRL 878.17 per household), a high proportion of young people (28.5%), and a low proportion of elderly people (5.9%). Moreover, this pattern has the highest sex ratio, with 117 men for every 100 women.

Figure 14. Photographs from the field visit. (**a**) Point 1: "USEP 1–Riverside and riparian communities"; (**b**) Point 2: "USEP 2–Low-density"; (**c**) Point 3: "USEP 4–High integration"; (**d**) Point 4: "USEP 6–Medium-density"; (**e**) Point 5: "USEP 7–Housing complex"; and (**f**) Point 6: "USEP 8–Medium-density informal".

"USEP 3–Periurban" in Cametá includes areas of traditional extractive settlements (Figure 15a,b). These settlements are of the *terra firme* type (dry land), with a clear demarcation between the plantation area, the housing area, and the monoculture area. In this case, roads determine transportation, and they are generally connected to regional highways. This pattern has the lowest household income (an average of BRL 813 per household).
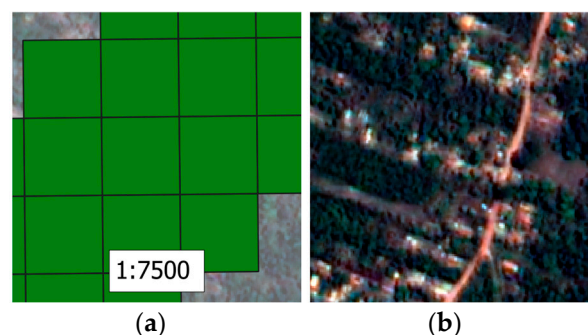

Figure 15. (**a**) Result of our classification and identification of "USEP 3–Periurban"; (**b**) Visualization of the same cells of 13 (**a**) on the CBERS-4A satellite WPM image (2 m spatial resolution, natural composition).

The "USEP 4–High integration" pattern in Cametá represents the best socio-economic conditions with an average household income above BRL 2180 and a literacy rate of 95%. The housing density is the lowest at 4.4 people per household, and there is a higher proportion of elderly people (8%) and a predominance of the female population. The central area of the city is where commercial activities, public health, and education buildings are concentrated (Figure 14c). The streets are paved and have regular layouts, and public lighting is available. Houses are generally made of masonry with good finishing, and there are three- to four-story buildings.

"USEP 6–Medium-density" is an area of medium-high density located in the outskirts of the central area of Cametá. The streets are paved near the center and unpaved as they

move away, with trees lining the streets. The houses are made of masonry with a good exterior finish, and the area is well-integrated with the rest of the city (Figure 14d). Similar to USEP 4, the female population predominates.

"USEP 7–Housing complex" consists mainly of medium-density housing complexes, including those from the "*Minha Casa, Minha Vida*" program. The streets have poor paving and lighting, and the electric grid has clandestine connections (Figure 14e). There is minimal greenery in the area, and the houses are built with unfinished masonry on plots of approximately 10×20 m. The area is almost exclusively residential, with a population density of 4.73 residents per household, a predominance of the male population, and the highest proportion of young people (32.24%). The income level is one of the lowest (BRL 916 per household) among the different patterns.

Finally, the "USEP 8–Medium-density informal" pattern consists of a spontaneous, medium-density occupation with vegetation inside the block. The area has grown informally towards the BR-422 highway, with unpaved and narrow streets and low public lighting, and it is poorly integrated with the rest of the city. The occupation is relatively recent, with a little more than ten years of occupancy, and is continuously expanding and densifying construction. There is no basic sanitation in the area, and the houses are mostly made of masonry with a population density of 4.43 people per house. The income level is below the municipality's average, around BRL 1025 per household, and there is a high proportion of young people (32.18%) and the smallest proportion of elderly people (below 4%).

6. Discussion

6.1. Santarém's Results

According to Figure 15a, the urban patterns in the occupied area of Santarém are dominated by the "USEP 3–Periurban" class (42%), followed by "USEP 6–High-density informal" (20%) and "USEP 2–Medium-density" (16%). Higher building density patterns, such as "USEP 4–High integration", only make up one-third of the area. The distribution of these patterns differs within the city, with the "High integration" pattern concentrated in the older urban areas, while the "USEP 7–Housing Complex" pattern is mostly found in recently converted urban areas. Rural areas exhibit patterns of low building density.

In Santarém's occupied area, most cells are classified as "Old urban" (78%), with smaller proportions classified as "New urban" (18%) and "Rural" (4%). The "Old urban" areas are characterized by a mix of urban patterns, with the "USEP 4–High integration", "USEP 2–Medium-density", and "High-density informal" patterns predominating. On the other hand, the "New urban" areas show a higher concentration of the "USEP 7–Housing Complex" pattern, indicating a pattern of recent conversion from rural to urban use. Rural areas display patterns associated with low building density, such as "USEP 1–Riverside", "USEP 3–Periurban", and "USEP 5–Main Roads".

A significant portion (36%) of the occupied area in Santarém overlaps with AGSN. The classes "USEP 3–Periurban", "USEP 6–High-density informal", and "USEP 2–Medium-density" are most found within the AGSN. However, the "USEP 2–Medium-density" class is 34% more likely to be in a precarious area compared to other cells, while the "High integration" class is 62% less likely to be in a precarious area. This aligns with the socioeconomic characteristics of these patterns, with "USEP 4–High integration" having higher income and better sanitation indicators.

Two classes, "USEP 6–High-density informal" and "USEP 5–Main roads", have a higher likelihood of overlapping with the AGSN. Over 45% of the cells in the "High-density informal" class overlap with the AGSN. This pattern exhibits characteristics such as high building and population density, lack of tree planting and paving, and lower levels of sanitation coverage and income, meeting the criteria of an AGSN. On the other hand, although "USEP 3–Periurban" has many areas overlapping with the AGSN (33% of its cells), it has a lower probability compared to the overall occupied area, suggesting that its overlap is due to its distribution rather than inherent precariousness.

The alignment plan implemented by Portuguese colonizers in the Santarém region aimed to structure urban expansion by adhering to the Land Law of 1850 [37]. Although not fully executed, this plan established a pattern of urban occupation with regularly shaped, high-density blocks. Our classification of urban patterns, particularly the "USEP 4–High integration" class, overlaps with this plan, indicating some agreement between the observed features and the historical alignment plan.

Our results align with the findings of previous studies. The "USEP 2–Medium-density" and "USEP 3–Peri-urban" classes represent more traditional occupation styles [37] characterized by unplanned layouts, adapted roads, and multi-purpose backyard areas. "USEP 6–High-density informal" resembles organized informal settlements resulting from the conversion of rural land, leading to poverty and limited opportunities. "USEP 7–Housing Complex" resembles formal settlements with standardized typologies and alterations to the natural landscape [84].

According to Tourinho [85], Santarém is a mid-sized Amazonian city that combines traditional and riverside influences with the impacts of highways. Our classification is consistent with this model, which places high-income populations, services, and commerce near the riverfront ("USEP 4–High integration") and then successively expands into areas of medium-low income and high density ("USEP 6–High-density informal"), low income and high or medium density ("USEP 2–Medium-density"), and low income and low density ("USEP 3–Periurban"). The only difference between this model and our classification is the inclusion of a high-density, low-income urban pattern ("USEP 7–Housing Complex") on the outskirts of the city.

6.2. Cametá's Results

According to the Figure 16d, in the occupied area of Cametá, the "Periurban" class has the highest prevalence (35%), followed by "Medium-density" (19%), "Low-density" (12%), "Housing Complex" (12%), "Riverside and riparian" (9%), "Main roads" (6%), "High integration" (5%), and "High-density informal" (2%). Low-density classes account for over 56% of the area, indicating a lower level of density compared to Santarém.

In Cametá, the distribution of urban patterns differs from Santarém. The "Old urban" class represents 56% of the area, while "New urban" and "Rural" make up 18% and 27%, respectively. Similar to Santarém, "USEP 4–High integration" and "USEP 2–Medium-density" are concentrated in the "Old urban" part, while "USEP 7–Housing Complex" is mostly in "Old urban" with a smaller portion in "New urban".

In Cametá, only 17% of the occupied area overlaps with an AGSN and the distribution of AGSNs among patterns is more balanced compared to Santarém. However, assessing how the patterns are distributed along the AGSNs provides interesting information. Over 85% of "USEP 8–High-density informal" and 31% of "USEP 7–Housing complex" are in AGSN areas, which are high rates compared to the rest of the cells. In contrast, only 4% of "USEP 4–High integration" overlaps with an AGSN. Similar to Santarém, although 25% of the cells that overlap an AGSN are from "USEP 3–Periurban", the class has a lower likelihood of overlapping when compared to the rest of the occupied area (30% less). This suggests that most of the AGSNs overlap the periurban areas not because this class is usually precarious but because of its high proportion and distribution in the study area.

Cametá follows a similar urban pattern to other riverine towns where road access has become more important than waterways. Like the model of Tourinho [85], in our work, the highest-income population resides in the area between the river and the parallel highway, which also accommodates commercial activities along the waterfront. Moving away from the highway, there is a decrease in income and building density, transitioning through areas of medium income and density, informal medium-density areas, and eventually lower income and density areas. However, like Santarém, Cametá does not have a high-density, low-income urban pattern on the outskirts of the city.
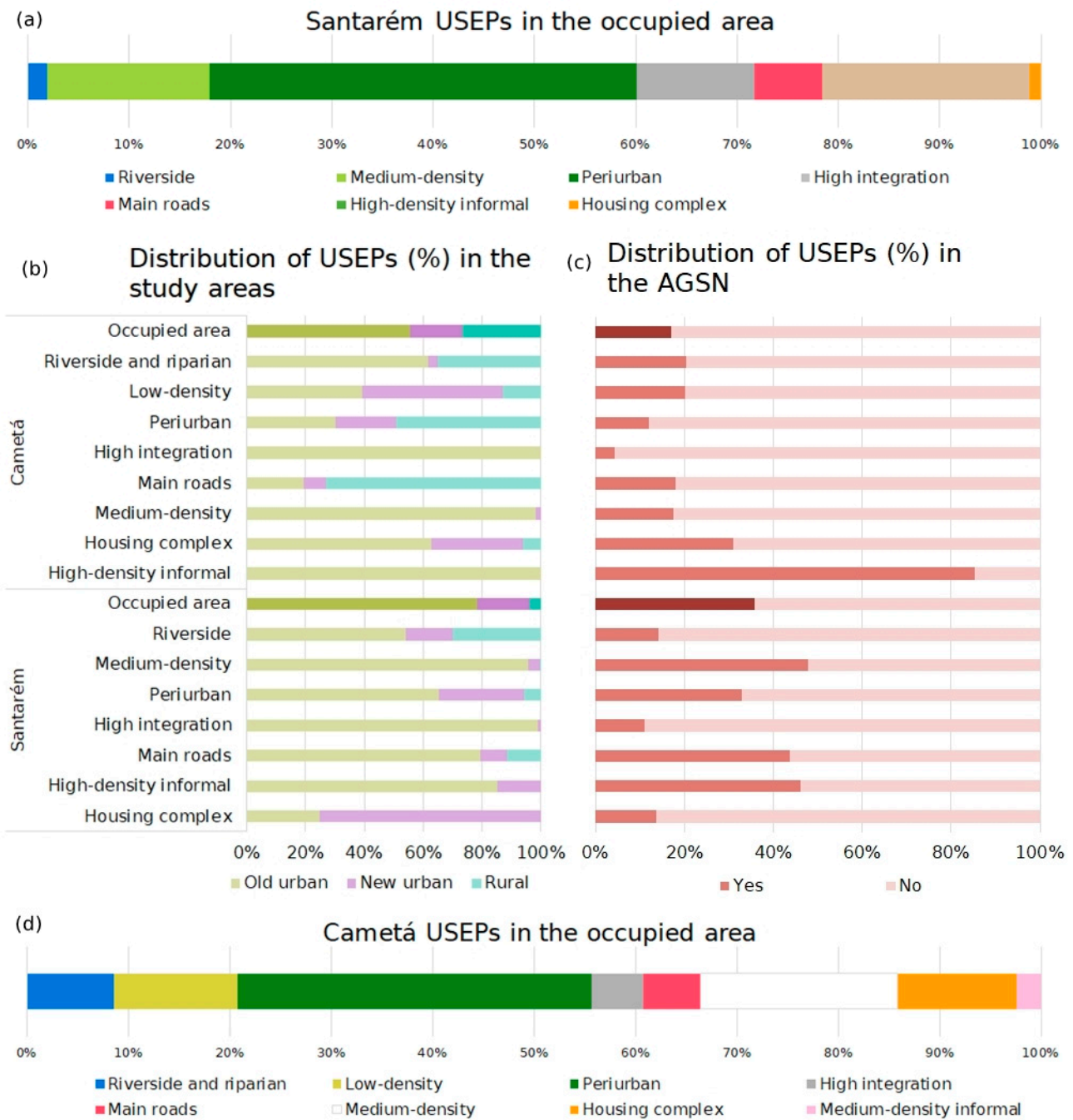
Figure 16. Graphs of the distribution of the USEPs along the occupied areas in Santarém (**a**) and Cametá (**d**). Distribution of USEPs in the study areas (**b**) and intersected with Subnormal Agglomerations (AGSN) (**c**).

6.3. Study Limitations and Considerations about the Classification Model

Our study has some limitations. In Cametá, the classification did not differentiate between the Riverside and Riparian classes, which refer to occupations along the river and the islands, respectively. Both classes were combined under "USEP 1–Riverside and riparians" due to their shared characteristic of being near water. To address this, we recommend identifying new clusters for mixed classes in future studies to allow better differentiation.

Another limitation is the use of the 2010 census data for constructing the socioeconomic indicators due to the cancellation of the 2020 census amid the COVID-19 pandemic and budget constraints. However, the socioeconomic indicators can be updated once new

census data becomes available. Despite these limitations, our study contributes to the understanding of sanitation conditions in Amazonian cities. Some urban patterns still exhibit low sanitation coverage, highlighting the need for improved services beyond central areas. Moreover, we adapted methodologies for analyzing urban morphology in Amazonian cities, incorporating rivers as transportation routes, occupation areas, and environmental variables.

The separate clustering of environmental and urban morphological dimensions allowed for a meaningful explanation of the USEPs. These two analysis dimensions are used as levels of information compartmentalization, where the clusters of each dimension represent these levels and provide information about the environment and urban nature of the study area. We decided to separate these two dimensions to explicitly highlight the environmental characteristics of these Amazonian cities. Utilizing unsupervised classification methods was crucial in the absence of comprehensive databases and prior mappings, revealing distinct patterns in the data. Although classifications were conducted separately, certain patterns showed high similarity, particularly in central areas, periurban regions, roads, and housing complexes. This suggests the existence of regional urban patterns that are replicated across different cities, varying in size and population.

Cametá, in comparison to Santarém, exhibits more rural characteristics and lower socioeconomic indicators. Its patterns are less dense, and while Santarém has higher construction density despite higher income levels, Santarém is more similar to the Belém Metropolitan Area, presenting areas of more regular blocks and streets, as well as more areas of informal settlements and intersections with the AGSN. The promotion of housing complexes, which deviates from the traditional Amazonian way of living, raises concerns in both cities.

The use of machine learning algorithms played a central role in our methodology, surpassing manual classification models and enabling the identification of the USEPs. We combined machine learning algorithms with expert knowledge obtained through field research. This combination allows a better understanding of the results produced by machine learning models, particularly when the inner workings of the algorithm are not well defined. Experts contribute by guiding the selection of important variables and identifying data sources, while machine learning algorithms facilitate updates and application in different areas [88].

However, some machine learning algorithms may require significant computational power and an ample amount of training data, which can be insufficient for different urban patterns in Brazilian municipalities, including Amazonian cities [29].

Regarding the data sources, two datasets were essential: amazon-ULC, which provided the land cover classifications with intra-urban classes; and the Topodata MDE bases, which enabled the construction of evaluation criteria for the environmental dimension. Finally, our work relied on publicly accessible data and free software, facilitating its replication in other cities by researchers.

7. Conclusions

The main objective of this study was to identify USEPs in Amazonian cities, specifically in Santarém and Cametá in Brazil, through the framework of Dos Santos et al. [18]. The USEPs were characterized by their environmental, urban morphological, and socioeconomic characteristics, including building density, integration, access road conditions, and terrain conditions. The study adapted morphological evaluation metrics and socioeconomic indicators to the unique reality of Amazonian cities. The findings confirmed the existence of diverse urban morphological patterns in Amazonian cities and highlighted the importance of specific classification models for accurate identification.

Future research directions are also suggested as follows: (i) conduct a new socioeconomic characterization with data from the demographic census scheduled to be published in 2023, (ii) study the identification of urban patterns in different study areas in a combined manner, identifying clusters that are common in all cities, (iii) evaluate the possibility of

plotting temporal trajectories of the identified urban pattern, and (iv) expand the application of the USEPs classification framework to other Amazonian cities.

To conclude, the Amazon region is of global relevance due to its unique environmental characteristics and biodiversity. Identifying urban patterns in the Amazon provides a foundation for sustainable urban planning and protects the region's cultural heritage. This work can aid in the proposition of more effective public policies to improve the urban population's quality of life and promote sustainable urban development in the region.

References

1. UN-HABITAT. *World Cities Report 2022: Envisaging the Future of Cities*; UN-HABITAT: Nairobi, Kenya, 2022.
2. da Trindade, S.-C.C., Jr. Cidades Médias Na Amazônia Oriental: Das Novas Centralidades à Fragmentação Do Território. *Rev. Bras. Estud. Urbanos Reg.* **2011**, *13*, 135. [CrossRef]
3. da Trindade, S.-C.C., Jr. A Cidade Dispersa: Os Novos Espaços de Assentamentos Em Belém e a Reestruturação Metropolitana. Ph.D. Thesis, Universidade de São Paulo, São Paulo, Brazil, 1998.
4. Leitão, K.O. A Dimensão Territorial Do Programa de Aceleração Do Crescimento: Um Estudo Sobre o PAC No Estado Do Pará e o Lugar Que Ele Reserva à Amazônia No Desenvolvimento Do País. Ph.D. Thesis, Universidade de São Paulo, São Paulo, Brazil, 2009.
5. Becker, B.K. Undoing Myths: The Amazon-an Urbanized Forest. *Man Biosph. Ser.* **1995**, *15*, 53.
6. IBGE. *Censo Demográfico 2010*; IBGE: Rio de Janeiro, Brazil, 2011.
7. Corrêa, R.L. Diferenciação Sócio-Espacial, Escala e Práticas Espaciais. *Revista Cidades* **2007**, *4*, 62–72. [CrossRef]
8. IBGE. *Tipologia Intraurbana: Espaços de Diferenciação Socioeconômica Nas Concentrações Urbanas Do Brasil*; IBGE: Rio de Janeiro, Brazil, 2017.
9. Soja, E. Uma Concepção Materialista Da Espacialidade. In *Abordagens Políticas da Espacialidade*; Becker, B., da Costa, R.H., Silveira, C.B., Eds.; UFRJ: Rio de Janeiro, Brazil, 1983; Volume 1, pp. 22–74.
10. Santos, M. *O Espaço Do Cidadão*; Edusp: São Paulo, Brazil, 2007; Volume 8, ISBN 8531409713.
11. Kurkdjian, M. de L.N. de O. Aplicações de Sensoriamento Remoto Ao Planejamento Urbano. In *Simpósio Latino Americano de Sensoriamento Remoto*; INPE: Cartajena, Colombia, 1993; pp. 1–15.
12. Netzband, M.; Stefanov, W.L.; Redman, C. *Applied Remote Sensing for Urban Planning, Governance and Sustainability*; Springer Science & Business Media: New York, NY, USA, 2007; ISBN 3540680098.
13. de Almeida, C.M. Aplicação Dos Sistemas de Sensoriamento Remoto Por Imagens e o Planejamento Urbano Regional. *Arq. Urb* **2010**, *1*, 98–123.
14. Weng, Q.; Quattrochi, D.A. *Urban Remote Sensing*; CRC Press: Boca Raton, FL, USA, 2018; ISBN 1315166615.
15. Andries, A.; Morse, S.; Murphy, R.; Lynch, J.; Woolliams, E.; Fonweban, J. Translation of Earth Observation Data into Sustainable Development Indicators: An Analytical Framework. *Sustain. Dev.* **2019**, *27*, 366–376. [CrossRef]
16. Kuffer, M.; Pfeffer, K.; Persello, C. Special Issue "Remote-Sensing-Based Urban Planning Indicators". *Remote Sens.* **2021**, *13*, 1264. [CrossRef]
17. Weil, K.E. Planejamento Empresarial. *RAE* **1975**, *1*, 66–67. [CrossRef]
18. dos Santos, B.D.; Maciel, R.R.; Kampel, M.; de Pinho, C.M.D.; Amaral, S. State-of-the-Art and Framework for Identifying Urban Patterns by Remote Sensing Data. *Rev. Bras. Cartogr.* **2023**, *1*, 1–19.

19. dos Santos, B.D.; de Pinho, C.M.D.; Amaral, S.; Paez, A. USEPs Data. *Zenodo* **2023**, 1.

20. Brunsdon, C.; Comber, A. Opening Practice: Supporting Reproducibility and Critical Spatial Data Science. *J. Geogr. Syst.* **2021**, *23*, 477–496. [CrossRef]

21. Desjardins, E.; Higgins, C.D.; Páez, A. Examining Equity in Accessibility to Bike Share: A Balanced Floating Catchment Area Approach. *Transp. Res. Transp. Environ.* **2022**, *102*, 103091. [CrossRef]

22. Zhu, X.X.; Qiu, C.; Hu, J.; Shi, Y.; Wang, Y.; Schmitt, M.; Taubenböck, H. The Urban Morphology on Our Planet–Global Perspectives from Space. *Remote Sens. Environ.* **2022**, *269*, 112794. [CrossRef]

23. Esch, T.; Marconcini, M.; Felbier, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M.; Taubenböck, H.; Müller, A.; Dech, S. Urban Footprint Processor—Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1617–1621. [CrossRef]

24. Marconcini, M.; Metz-Marconcini, A.; Üreyen, S.; Palacios-Lopez, D.; Hanke, W.; Bachofer, F.; Zeidler, J.; Esch, T.; Gorelick, N.; Kakarla, A.; et al. Outlining Where Humans Live, the World Settlement Footprint 2015. *Sci. Data* **2020**, *7*, 242. [CrossRef]

25. Stewart, I.D.; Oke, T.R.; Krayenhoff, E.S. Evaluation of the 'Local Climate Zone' Scheme Using Temperature Observations and Model Simulations. *Int. J. Climatol.* **2014**, *34*, 1062–1080. [CrossRef]

26. Schiavina, M.; Melchiorri, M.; Pesaresi, M.; Politis, P.; Freire, S.; Maffenini, L.; Florio, P.; Ehrlich, D.; Goch, K.; Tommasi, P. *GHSL Data Package 2022*; European Commission: Luxembourg, 2022.

27. Gonçalves, G.C.; de Oliveira, L.M.; D'Asta, A.P.; Amaral, S. Geoinformação Para a Visibilidade Das Áreas Urbanas de Cidades Amazônicas. *Rev. Geoaraguaia* **2021**, *11*, 149–165.

28. Dal'Asta, A.P.; Brigatti, N.; Amaral, S.; Escada, M.I.S.; Monteiro, A.M.V. Identifying Spatial Units of Human Occupation in the Brazilian Amazon Using Landsat and CBERS Multi-Resolution Imagery. *Remote Sens.* **2012**, *4*, 68–87. [CrossRef]

29. dos Santos, B.D.; de Pinho, C.M.D.; Oliveira, G.E.T.; Korting, T.S.; Escada, M.I.S.; Amaral, S. Identifying Precarious Settlements and Urban Fabric Typologies Based on GEOBIA and Data Mining in Brazilian Amazon Cities. *Remote Sens.* **2022**, *14*, 704. [CrossRef]

30. Cardoso, A.C.D.; Lima, J.J.F.; Ponte, J.P.X.; Ventura, R.D.S.; Rodrigues, R.M. Morfologia Urbana Das Cidades Amazônicas: A Experiência Do Grupo de Pesquisa Cidades Na Amazônia Da Universidade Federal Do Pará. urbe. *Rev. Bras. Gestão Urbana* **2020**, *12*, e20190275. [CrossRef]

31. Indvik, K.B.; Gallo, I.; Quintana, M.; Blanc, F.; do Canto, Ó. *Amazon Cities and Sustainable Urban Development*; Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ): Brasília, Brazil, 2018.

32. de Abreu, P.V.L.; Lima, J.J.F.; Fischer, L.R.D.C. Aforar, Arrumar e Alinhar: A Atuação Da Câmara Municipal de Belém Na Configuração Urbano-Fundiária Da Cidade Durante o Século XIX. *An. Mus. Paul. História Cult. Mater.* **2018**, *26*, e29. [CrossRef]

33. Lima, A.; Lima, J.J.F. O Plano Urbano Da Villa Do Pinheiro e a Exploração Gomífera Na Amazônia Oriental (1869–1906). In Proceedings of the Anais da 8a Conferência da Rede Lusófona de Morfologia Urbana PNUM 2019, Maringá, Brasil, 21–23 August 2019; Volume 1.

34. Lima, A.P.C.; Rodrigues, R.M. O Superdimensionamento Da Grelha: Análise de Padrão de Subdivisão Fundiária e Suas Adaptações Aos Usos Do Solo No Núcleo Histórico de Icoaraci, Belém-PA. In *Morfologia Urbana: Território, Paisagem e Planejamento, Proceedings of the Anais da 6a Conferência da Rede Lusófona de Morfologia Urbana PNUM 2017, Sintra, Portugal, 24–25 Augest 2017*; UFES Vitória: Vitória, Brazil, 2017; pp. 98–108.

35. Souza, R.D.P.; Galvão, L. Formas Da Produção Habitacional Na "Nova Belém": Estudo Comparativo Dos Diferentes Tipos de Produção Habitacional Ao Longo Da Av. Augusto Montenegro, Belém (PA) e Suas Tendências de Consolidação. *An. XV Enanpur* **2013**, *15*, 1–18.

36. Ventura Neto, R.D.S. A (Trans) Formação Socioespacial Da Amazônia: Floresta, Rentismo e Periferia. Ph.D. Thesis, Universidade Estadual de Campinas, Campinas, Brazil, 2017.

37. Gomes, T.D.V.; Cardoso, A.C.D. Santarém: O Ponto de Partida Para o (Ou de Retorno) Urbano Utopia. *Urbe. Rev. Bras. Gestão Urbana* **2019**, *11*, e20170219. [CrossRef]

38. Swyngedouw, E.; Acselrad, H. Acselrad, H. A Cidade Como Un Híbrido: Natureza, Sociedade, e 'Urbanizaçao-Cyborg. In *A Duração das Cidades–Sustentabilidade e Risco nas Políticas Urbanas*; DP&A Editora: Rio de Janeiro, Brazil, 2001.

39. Cameron, R.W.F.; Blanuša, T.; Taylor, J.E.; Salisbury, A.; Halstead, A.J.; Henricot, B.; Thompson, K. The Domestic Garden—Its Contribution to Urban Green Infrastructure. *Urban Urban Green* **2012**, *11*, 129–137. [CrossRef]

40. Corrêa, R.L. A Periodização Da Rede Urbana Da Amazônia. *Rev. Bras. Geogr.* **1987**, *4*, 39–68.

41. Roosevelt, A. Uma Memória Histórica Da Pesquisa Arqueológica No Brasil (1981–2007). *Bol. Mus. Para. Emílio Goeldi. Ciências Hum.* **2009**, *4*, 155–170. [CrossRef]

42. IBGE, Cidades e Estados. Available online: https://www.ibge.gov.br/cidades-e-estados/pa/santarem.html (accessed on 20 February 2022).

43. IBGE. *Regiões de Influência Das Cidades 2018*; IBGE: Rio de Janeiro, Brazil, 2020.

44. Malheiro, B.C.P.; da Trindade Júnior, S.-C.C. Entre Rios, Rodovias e Grandes Projetos: Mudanças e Permanências Em Realidades Urbanas Do Baixo Tocantins (Pará). *História Rev.* **2009**, *14*, 6.

45. Oliveira, K.D. Entre a Várzea e a Terra Firme: Estudos de Espaços de Assentamentos Tradicionais Urbanos Rurais Na Região Do Baixo Tocantis. Master's Thesis, UFPA, Belém, Brazil, 2020.

46. BRASIL. *Plano Territorial de Desenvolvimento Rural Sustentável Território Águas Emendadas–DF*; Ministério do Desenvolvimento Agrário: Brasilia, Brazil, 2006; Volume 1.

47. LISS Grupo De Pesquisa. Liss Inpe. Available online: https://www.lissinpe.com.br/ (accessed on 2 June 2023).

48. INPE. Câmeras Imageadoras CBERS-4A. Available online: http://www.cbers.inpe.br/sobre/cameras/cbers04a.php (accessed on 8 January 2022).

49. OSM. OpenStreetMap. Available online: https://www.openstreetmap.org/#map=4/-15.13/-53.19 (accessed on 1 November 2021).

50. dos Santos, B.D.; de Pinho, C.M.D.; Amaral, S.; Paez, A. AmazonULC Data Package 2023. Available online: http://10.5281/zenodo.7749057 (accessed on 25 April 2023).

51. Valeriano, M.d.M.; Rossetti, D.d.F. Topodata: Brazilian Full Coverage Refinement of SRTM Data. *Appl. Geogr.* **2012**, *32*, 300–309. [CrossRef]

52. Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Freire, S.; Halkia, M.; Julea, A.; Kemper, T.; Soille, P.; Syrris, V. Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014. *Publ. Off. Eur. Union* **2016**, 1–62.

53. Team, Q.D. QGIS Geographic Information System 2021. Version Number: 3.18. Available online: https://issues.qgis.org/projects/qgis/wiki/QGIS_Citation_Repository (accessed on 5 June 2023).

54. van Rossum, G. *Python Reference Manual*; Department of Computer Science, Centrum voor Wiskunde en Informatica: Amsterdam, The Netherlands, 1995.

55. Körting, T.S.; Garcia Fonseca, L.M.; Câmara, G. GeoDMA—Geographic Data Mining Analyst. *Comput Geosci* **2013**, *57*, 133–145. [CrossRef]

56. Abreu, E.S.; Rosim, S.; Rennó, C.D.; Oliveira, R.D.F.; Jardim, A.C.; Ortiz, J.D.O.; Dutra, L.V. Terrahidro—A Distributed Hydrological System to Delimit Large Basins. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 546–549.

57. Turner, A. Depthmap: A Program to Perform Visibility Graph Analysis. In Proceedings of the 3rd International Symposium on Space Syntax, Citeseer, Atlanta, 7–11 May 2001; Volume 31, pp. 12–31.

58. Rennó, C.D.; Nobre, A.D.; Cuartas, L.A.; Soares, J.V.; Hodnett, M.G.; Tomasella, J.; Waterloo, M.J. HAND, a New Terrain Descriptor Using SRTM-DEM: Mapping Terra-Firme Rainforest Environments in Amazonia. *Remote Sens. Environ.* **2008**, *112*, 3469–3481. [CrossRef]

59. Braga, B.; Hespanhol, I.; Conejo, J.G.L.; Mierzwa, J.C.; de Barros, M.T.L.; Spencer, M.; Porto, M.; Nucci, N.; Juliano, N.; Eiger, S. *Introdução à Engenharia Ambiental: O Desafio do Desenvolvimento Sustentável*; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2005.

60. Souza, C.M.M.; Montero, L.S.; Liesenberg, V. Análise de Urbanização Em Áreas Declivosas, Como Uma Das Etapas Da Avaliação Ambiental Estratégica (AAE), Visando o Desenvolvimento Local. In Anais. In Proceedings of the XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 21–26 April 2007; pp. 21–26.

61. Galera, R.A.; Brito, C.O.; Campos, F.S.; Antunes, J.S.; Canil, K. Modelo Digital de Curvatura Côncava Para Determinação de Unidades Geotécnicas de Aptidão à Urbanização. In Proceedings of the XVIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, 28–31 May 2017.

62. Conzen, M.R.G. Alnwick, Northumberland: A Study in Town-Plan Analysis. *Trans. Pap.* **1960**, *3*, 122. [CrossRef]

63. Conzen, M.R.G. *Thinking about Urban Form: Papers on Urban Morphology, 1932–1998*; Peter Lang: New York, NY, USA, 2004; ISBN 3039102761.

64. Oliveira, V. Morpho: A Methodology for Assessing Urban Form. *Urban Morphol.* **2013**, *17*, 21–33. [CrossRef]

65. Hillier, B.; Penn, A.; Hanson, J.; Grajewski, T.; Xu, J. Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environ. Plann B Plann* **1993**, *20*, 29–66. [CrossRef]

66. de Medeiros, V.A.S. Urbis Brasiliae Ou Sobre Cidades Do Brasil: Inserindo Assentamentos Urbanos Do País Em Investigações Configuracionais Comparativas. Ph.D. Thesis, Universidade de Brasília, Brasília, Brazil, 2006.

67. do Carmo, C.L.; Junior, A.A.R.; Nogueira, A.D. Aplicações Da Sintaxe Espacial No Planejamento Da Mobilidade Urbana. *Ciência Eng.* **2013**, *22*, 29–38. [CrossRef]

68. Jacobs, J. *Morte e Vida de Grandes Cidades*, 3rd ed.; WMF Martins Fontes: São Paulo, Brazil, 2011; ISBN 978-85-7827-421-4.

69. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **1973**, *6*, 610–621. [CrossRef]

70. dos Santos, B.D.; Amaral, S.; de Pinho, C.M.D.; Paéz, A. Adapting Socioeconomic Indicators And Identifying Intra-Urban Typologies in Santarém—PA. In Proceedings of the Anais do XX Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brazil, 2–5 April 2023; pp. 1941–1944.

71. D'Antona, Á.; Bueno, M.D.C.; Dagnino, R. Using Regular Grids for Spatial Distribution of Census Data for Population and Environment Studies in Brazil. In Proceedings of the Population Association of America Annual Meeting Program 2011, Washington, DC, USA, 31 March–2 April 2011.

72. dos Santos, B.D.; de Pinho, C.M.D.; de Jesus, T.B. Níveis de Consolidação de Assentamentos Precários A Partir de Dados de Sensoriamento Remoto. In *Proceedings of the XIX Simpósio Brasileiro de Sensoriamento Remoto*; Santos, Brazil, 14–17 April 2019, INPE: São José dos Campos, Brazil, 2019; pp. 3224–3227.

73. Gonçalves, G. *Identificação de Assentamentos Precários Na Região Do Grande ABC: Uma Abordagem Estatística*; Trabalho de Conclusão de Curso, UFABC: Santo André, Brazil, 2018.
74. CDHU. *Ufabc Desenvolvimento e Aplicação de Metodologia Para Identificação, Caracterização e Dimensionamento de Assentamentos Precários*; CDHU: São Bernardo do Campo, Brazil, 2018.
75. Feitosa, F.D.F.; Vasconcelos, V.V.; de Pinho, C.M.D.; da Silva, G.F.G.; da Silva Gonçalves, G.; Danna, L.C.C.; Lisboa, F.S. IMMerSe: An Integrated Methodology for Mapping and Classifying Precarious Settlements. *Appl. Geogr.* **2021**, *133*, 102494. [CrossRef]
76. Feitosa, F.D.F.; da Cunha, L.F.B.; Gonçalves, G.D.S.; da Silva, G.F.G.; Simões, P.R. Modelagem para a identificação de núcleos urbanos informais: Uma proposta metodológica. In *Núcleos Urbanos Informais: Abordagens Territoriais da Irregularidade Fundiária E da Precariedade Habitacional*; Krause, C., Denaldi, R., Eds.; Editora Instituto de Pesquisa Econômica Aplicada (Ipea): Brasília, Brazil, 2022; Volume 4, pp. 114–144. ISBN 978-65-5635-044-8.
77. Tobler, W.R. Geographical Filters and Their Inverses. *Geogr. Anal.* **1969**, *1*, 234–253. [CrossRef]
78. Murtagh, F.; Contreras, P. Algorithms for Hierarchical Clustering: An Overview. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 86–97. [CrossRef]
79. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
80. da Cunha, L.F.B.; Feitosa, F.D.F.; Paez, A.; de Pinho, C.M.D. Exploring the Diversity of Informal Settlements in Brazil: A Clustering Approach to Qualitative Field Survey Data. In *Proceedings of the Simpósio Brasileiro de Geoinformática (GEOINFO)*; São José dos Campos, Brazil, 30 November 2022, Rosim, S., Santos, L.B.L., Pereira, M.D.A., Eds.; Instituto Nacional de Pesquisas Espaciais (INPE): São José dos Campos, Brazil, 2022; pp. 63–73.
81. Huang, Z. Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Min Knowl Discov* **1998**, *2*, 283–304. [CrossRef]
82. de Vos, N. Kmodes Categorical Clustering Library 2021. Available online: http://pypi.org/project/kmodes/ (accessed on 24 February 2023).
83. Victoriano, R.; Paez, A.; Carrasco, J.-A. Time, Space, Money, and Social Interaction: Using Machine Learning to Classify People's Mobility Strategies through Four Key Dimensions. *Travel Behav. Soc.* **2020**, *20*, 1–11. [CrossRef]
84. Cardoso, A.C.D.; de Melo, A.C.; Gomes, T.D.V. O Urbano Contemporâneo Na Fronteira de Expansão Do Capital: Padrões de Transformações Espaciais Em Seis Cidades Do Pará, Brasil. *Rev. Morfol. Urbana* **2016**, *4*, 5–28. [CrossRef]
85. Tourinho, H.L.Z. Estrutura Urbana de Cidades Médias Amazônicas: Análise Considerando a Articulação Das Escalas Interurbana e Intraurbana. Ph.D. Thesis, UFPE, Recife, Brazil, 2011.
86. Cardoso, A.C.D.; Lima, J.J.F. *Tipologias e Padrões de Ocupação Urbana Na Amazônia Oriental: Para Que e Para Quem*, 1st ed.; UFPA: Belém, Brazil, 2006.
87. IBGE. *Aglomerados Subnormais 2019: Classificação Preliminar e Informações de Saúde Para o Enfrentamento à COVID-19*; IBGE: Rio de Janeiro, Brazil, 2020.
88. Kuffer, M.; Pfeffer, K.; Sliuzas, R. Slums from Space—15 Years of Slum Mapping Using Remote Sensing. *Remote Sens.* **2016**, *8*, 455. [CrossRef]