MAPBIOMAS

Urban Areas mapping

Algorithm Theoretical Basis Document (ATBD)

Collection 11.0 - ATBD Version 1

Team authors

Breno Malheiros de Melo	USP/EESC
Edimilson Rodrigues dos Santos Junior	USP/EESC
Julio Cesar Pedrassoli	USP/FFLCH

São Paulo, 2026

Table of Contents

1. Overview

This document presents the methodology developed to map urban areas across the Brazilian territory from 1985 to 2025, as part of Collection 11 of the MapBiomas Project. Building on the general MapBiomas framework, the urban area mapping process applies supervised classification using the Random Forest algorithm (Breiman, 2001) and annual composite Landsat imagery.

The methodological workflow comprises three main steps: (i) mosaic generation, to obtain annual composites from Landsat imagery; (ii) probability classification, which includes sample preparation, training of classification models, prediction of class probabilities, and application of thresholds for binary urban/non-urban classification; and (iii) post-classification procedures, encompassing spatial and temporal filtering to improve classification consistency and minimize errors. Subsequently, classification results are exported to the MapBiomas workspace for integration with other thematic classifications within the broader land cover and land use mapping framework (Figure 1). Details about each step are provided further and were conducted using Google Earth Engine platform, javascript and python. The codes are openly available in the MapBiomas GitHub repository.

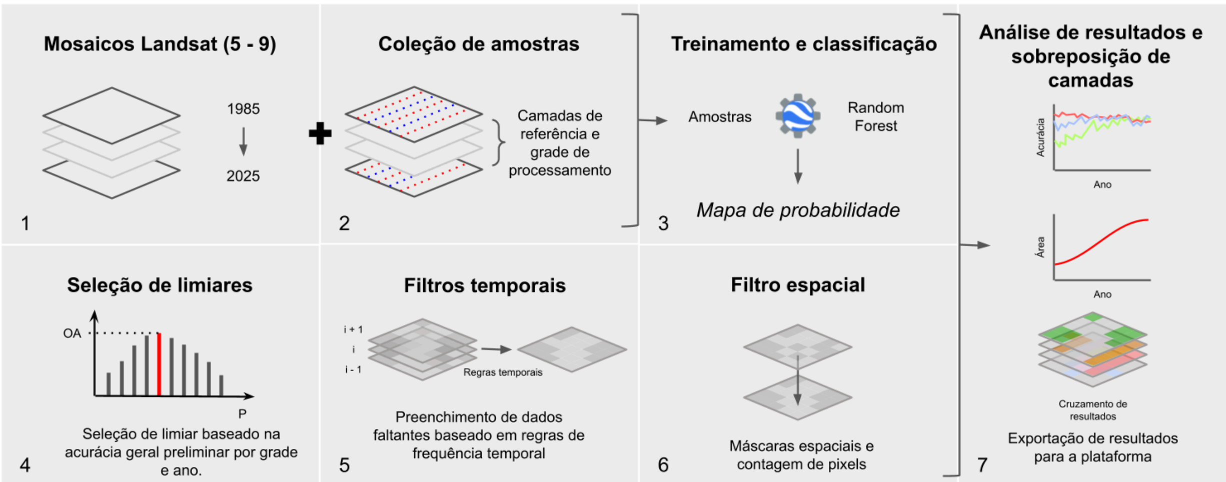


Figure 1. Basic scheme of urban areas classification.

Over successive collections, the urban classification method has been continuously refined through conceptual and methodological improvements. For Collection 11, updates include procedures for obtaining samples based on consolidated results throughout the collections 6 to 10, making it possible to use different classes to identify both urbanized areas and urban vegetation for most of the classified areas. Similar to previous collections, probability thresholding for urban classification was used, as well as for urban vegetation. Additional enhancements were made to improve temporal consistency of the annual maps and reduce classification noise, including the use of a temporal smoothing, probability threshold optimization.

Since Collection 6, the mapped class has been designated as "Urban/Urbanized Area" (UA), replacing the previous label "Urban Infrastructure". This update aligns the nomenclature with terminology commonly used in urban studies, including by (IBGE, 2017). UA are areas with predominance of significant density of buildings, roads and infrastructure. It should be noted that when making external quantitative comparisons, it is crucial to ensure that the chosen concepts are aligned.

2. Landsat image mosaics

Landsat imagery was used throughout the time series, incorporating data from Landsat missions 5, 7, 8, and 9 (see Table S1). Both Surface Reflectance products from each mission were used to compute spectral indices, which were then aggregated annually using median values, percentiles, or composite indices (see Table S2).

Image processing was conducted using annual image collections. For each year, clouds and shadows were masked, scale factors were applied. Spectral indices were computed as additional bands, based on previous Urban Areas mapping products and existing literature. To reduce pixel values within each year into representative values, appropriate statistical reducers based on selected percentiles were applied (see Table S2). Additionally, differences between percentile values were calculated to capture intra-annual variability. The main processing steps were as follows:

1. Filter Landsat Collection scenes by acquisition date on a yearly basis (from 1985 to 2025) and spatially constrained to the Brazilian territory.
2. Mask cloud and cloud shadow pixels in all scenes using quality attributes derived from the CFMASK 2 algorithm[1],accessed through the QA_PIXEL band.
3. Scale surface reflectance values were scaled to the 0–1 range by applying the provided scale factor (−0.2) and offset (0.0000275) as specified in collections' bands description in each reference page.
4. Compute selected spectral indices and spectral mixture fractions for each scene.
5. Apply appropriate reducers to each band and index.
6. Calculate differences between reduced indices to capture intra-annual variability.
7. Composite all processed bands and indices to produce a single annual mosaic.

The selection of bands and indices for urban area classification was analyzed based on the best-performing classification results (see the section Classification Algorithm).

3. Classification based on Random Forest algorithm

Urban areas classification procedures were developed within Google Earth Engine (GEE) based on Landsat imagery and Random Forest algorithm covering the Brazilian territory. The process was divided into several steps, starting with a definition of a spatial scope, satellite imagery, ancillary datasets, and the classification algorithm. This stage was based on selecting the optimal features to provide a temporal consistent urban binary annual classification from 1985 to 2025.

3.1. Spatial scope

This work covers the entire Brazilian territory. However, to avoid unnecessary computation, regions with no signs of urbanization were excluded. Polygons where urban areas are likely to be found were based on existing census tracts (IBGE, 2020). The resulting "search area" was defined using a grid of hexagons that intersect these features, along with topographic sheet codes derived from the World at the Millionth series, scaled to 1:250,000. This regular grid, historically adopted by official agencies for national mapping, were used

[1] CFMask is a multi-pass algorithm that uses decision trees to prospectively label pixels in the scene; it then validates or discards those labels according to scene-wide statistics. It also creates a cloud shadow mask by iteratively estimating cloud heights and projecting them onto the ground. Reference: https://www.usgs.gov/core-science-systems/nli/landsat/cfmask-algorithm .

here as processing units for organizing the computational workflow for urban area mapping (see Figure 2).

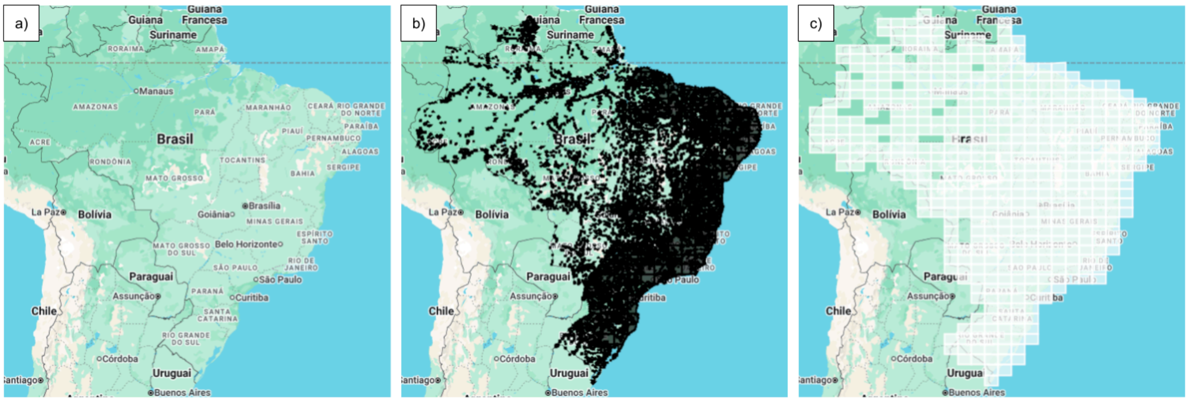


Figure 2. Search area and regular sheets (processing tiles). a) Brazil. b) Search area - where urban areas can be found. c) Processing units - regular grid defining the tiles for processing the classification.

3.2. Samples collection

We used two complementary methods to get samples of urban and not urban areas. An overall approach applied throughout MapBiomas urban areas collections has been the use of external reference datasets and exploratory methods to get samples of urban and not urban areas. To improve our classification and identify urban vegetation, we applied the use of stable areas over time and collections as a simplified approach to get samples of different classes, enabling the use of multiclasses classification. These two approaches are described below.

3.2.1. Overall approach for urban and not-urban samples

Until Collection 10, training samples were automatically generated by first creating a preliminary urban mask from OpenStreetMap road and pathway vectors, which were filtered using nighttime light imagery (NOAA) (Elvidge et al., 2021) and further refined with GHSL (Corbane et al., 2018) built-up maps (for 1985) and Brazil's Third National Inventory urban mappings (for 1994, 2002, and 2010). Candidate urban areas were then defined by buffering the remaining pathways by 100 meters, after which an exploratory classification using NDVI and NDWI masked out vegetated and water bodies. The final urban mask resulted from the intersection of this filtered OSM-based mask with the exploratory classification results, while the non-urban mask was defined as its symmetrical difference; random points were subsequently generated across all 522 tiles and labeled using PostGIS, yielding a large sample set for the years 1985, 1994, 2002, 2010, and 2018 - totaling, for example, roughly 891k non-urban and 533k urban samples in 1985, and 972k non-urban and 614k urban samples in 2018 (3). While this approach has been efficient to obtain urban areas, urban vegetation required other processing steps to proceed the classification. This approach was used to maintain consistency with previous collections regarding urban areas classification.

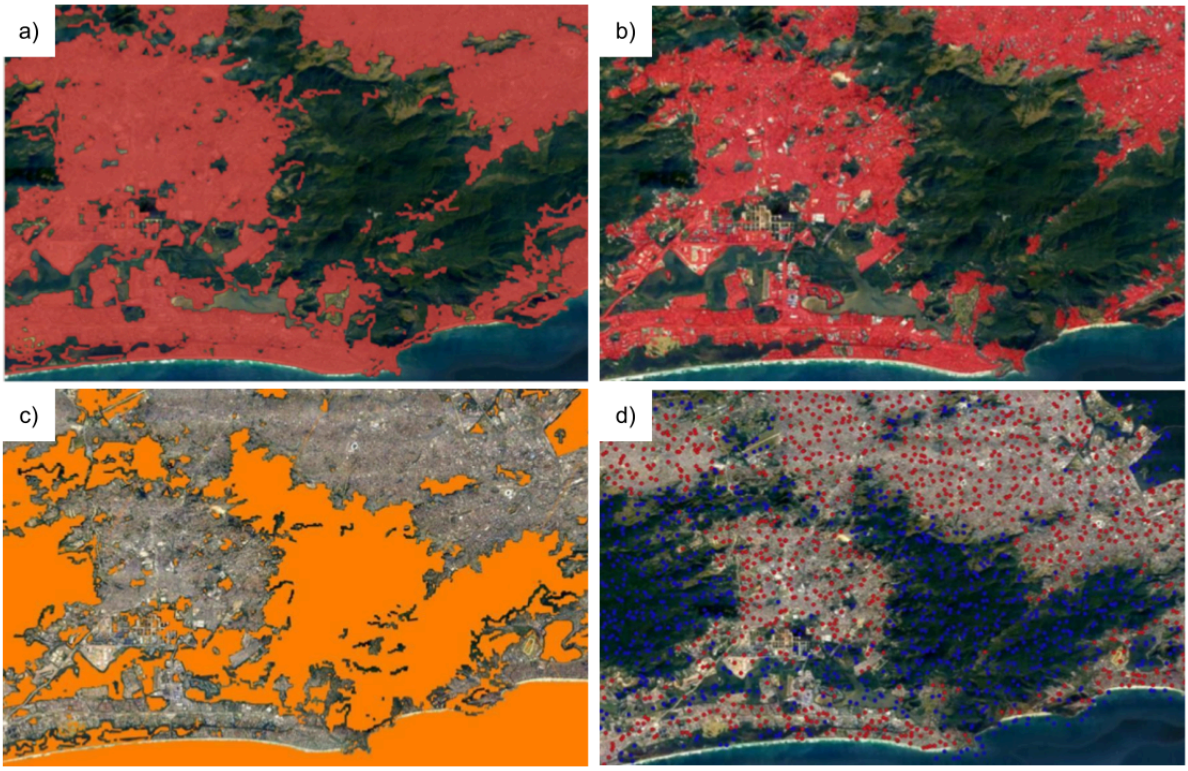


Figure 3. Sequential simplified steps to obtain urban and not urban samples exemplified for Rio de Janeiro - RJ, Brazil. a) Exploratory classification results. b) Final urban mask for (in red). c) Final not-urban mask (in orange). d) Random points divided by urban areas (red) and non-urban areas (blue).

3.2.2. Simplified approach for urban and not-urban samples

Samples of multiclasses were gathered following three basic steps: (i) simplification of classes from MapBiomas collections (from Collection 6 to 10), (ii) gathering stable classes over time and across collections, and (iii) creating random samples for each time range considered. We simplified the classes based on the MapBiomas Legend Level 1, aggregated into natural vegetation, farming, water, and urban areas (Table 1) – we did not consider other non-vegetated areas.

A first set of samples was based on the entire time series considered, from 1985 to 2020. To consider new land use dynamics, we progressively added new samples for the ranges of 1993 to 2020, 2001 to 2020, and 2009 to 2020. Random points were generated and labeled according to the simplified class of the stable pixel at each time range. These samples were then assigned to specific classification years according to their time range. The stable classes until 2020 were applied for the classification of the remaining years up to 2025. This approach was used combined with the overall approach to enable a multiprobability classification for urban vegetation mapping.

Table 1. Simplification of classes used to get samples for multiclasses classification.

Simplified classes	**Aggregated classes**
Natural vegetation	Forests, Herbaceous and Shrubby Vegetation
Farming	All the classes of farming

Water	All the classes of water
Urban areas	Urban areas

3.3. *Classification algorithm*

The Random Forest algorithm implemented in Google Earth Engine (smileRandomForest) was applied to map urban areas in MapBiomas Collection 11.0 using training samples of urban and non-urban areas. Random Forest parameters were set to 120 trees and 5 minimum leaf populations. In the Random Forest algorithm, the output mode was set to muli-probability, resulting in a classification image assigning to each pixel its probability of being one of the classes in Table 1.

Sequential steps were followed to optimize the classification parameters Google Earth Engine. Based on a sub-set of the processing units covering Brazilian capitals (a total of 65 tiles; see the section Spatial scope), sample selection and mosaic refinement were conducted. For that, the years of 1985, 1990, 2000, 2010 e 2020 were used. For each year and unit, different sample quantities (see the section Samples selection) and mosaic compositions (see the section Feature Space) were analyzed through a Random Forest classification (see the section Random Forest training and classification) with a view to define adequate parameters to be used for the entire Brazilian territory and temporal series. These parameters were assumed as the sample's quantities and mosaic composition with best classification performance.

Both sample selection and mosaic refinement were guided by the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) metric (Bradley, 1997; Fawcett, 2006). The ROC curve is a graphical representation used to evaluate the performance of binary classification models. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across various threshold values (equations 1 and 2, respectively), allowing assessment of the model's ability to discriminate between urban and non-urban areas. The AUC summarizes this performance into a single value: the closer it is to 1, the better the model is at distinguishing between the two classes.

$$TPR = TP / (TP + FN) \qquad \text{Eq. 1}$$

$$FPR = FP / (FP + TN) \qquad \text{Eq. 2}$$

Where:

TP = True Positives (urban correctly classified)

FP = False Positives (non-urban wrongly classified as urban)

FN = False Negatives (urban wrongly classified as non-urban)

TN = True Negatives (non-urban correctly classified)

After defining sample quantities and mosaic composition, a Random Forest classification was applied considering the whole time series to generate annual urban classification probabilities for each processing grid. These results were subsequently harmonized temporarily (from 1985 to 2025, annually, see the section Temporal smoothing), and a cutoff threshold was estimated per grid to produce a binary classification of urban areas (see the section Urban areas binary classification).

3.3.1. Samples selection

Sample selection refers to the process of determining the optimal number of urban and non-urban samples required for effective classification. The optimal quantity was defined as the point beyond which increases in sample size did not yield significant improvements in classification accuracy. Based on experimental tests, a range of 30 to 300 randomly selected urban samples was evaluated iteratively using a standard mosaic (see Table S3). A fixed class balance of 1:2 (urban to non-urban samples) was adopted, based on prior results.

To calculate the ROC curve, AUC, and additional accuracy metrics for each sample size, a separate validation set of 400 urban and 800 non-urban samples (also randomly selected) was used.

Considering the results (see Figure S1), the selected urban samples quantity to classify each processing grid were 220 units. Further details of the classification process are provided in the section Random Forest training and classification.

3.3.2. Feature Space

To support urban classification, several mosaic compositions were evaluated using predefined sets of Landsat bands, spectral indices, and spectral mixture components. Each mosaic set was tested to identify the most effective combination for classification purposes (see Table S4 for details of each configuration). The evaluated mosaics included the following list and their combination:

- Bands - Composed of the original Landsat spectral bands from the Surface Reflectance collection.
- Indices 1 - Comprising basic indices related to vegetation, water, soil, and urban areas, calculated from Landsat Surface Reflectance data.
- Indices 2 - Included indices derived from thermal band, using Landsat Raw data.
- Mix 1 - Spectral mixture bands obtained from Spectral Mixture Analysis (SMA) (see Table S2 and Table S4).
- Mix 2 - Additional spectral mixture bands (see Table S2 and Table S4).
- Bands used in MapBiomas Collection 9 - complete mosaic as defined in Table S1 complemented with the Automated Water Extraction Index (AWEIsh) and the Soil Adjusted Vegetation Index (SAVI).

The results show that the model performed well even for simpler mosaics (see Figure S2). A mosaic similar to what was used in Collection 9 was selected for consistency with previous MapBiomas collections. Therefore, the feature space selected to characterize Urban Areas for MapBiomas Collection 11.0 is the dataset of urban and non-urban points trained with the complete mosaic, with no differences from the one used for urban areas of Collection 9 of MapBiomas (see the example of Figure 4).

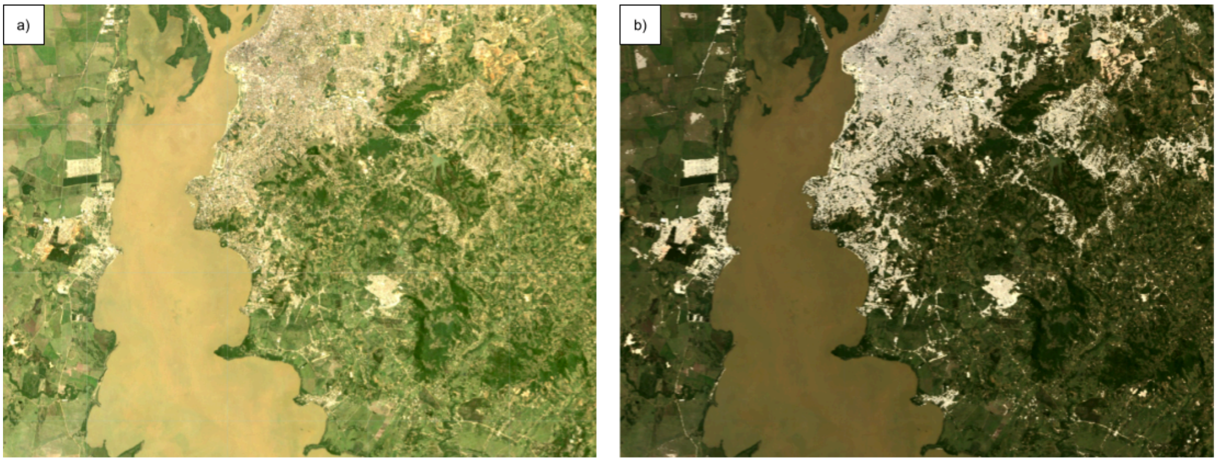


Figure 4. Classification example using the selected mosaic. Porto Alegre - RS, Brazil. a) Landsat mosaic. b) Classification probability overlaid with the mosaic - The white areas are the most probable to be urban.

Samples selection was used with the assumption that once a point was urban, it remained urban for the following years. Therefore, images of 1985 up to 1993 were used to train the dataset of 1985, resulting in one feature space per year per tile. Likewise, images of 1994 up to 2002 were used to train the dataset of 1994, images of 2003 up to 2009, to train the dataset of 2003, images of 2010 up to 2017, to train the dataset of 2010 and images of 2018 up to 2023, to train the dataset of 2018.

3.3.3. Random Forest training and classification

The following Random Forest classification procedures were applied to support sample selection, mosaic refinement, and binary classification of urban areas. First, samples were based on the processing grid (see the section Spatial scope). To ensure a diverse spectral representation and optimize sample availability, the training area for each grid unit was defined as its surrounding neighborhood intersecting the previously defined search area. Within this area, a subset of available samples was randomly selected for training.

The model training was implemented using a moving window approach: a block of nine grid units (3×3) was used, where the central unit served as the classification target, and the nine neighboring units provided the training data (Figure 5). For evaluation purposes, the final validation was conducted using MapBiomas validation samples (see MapBiomas website).

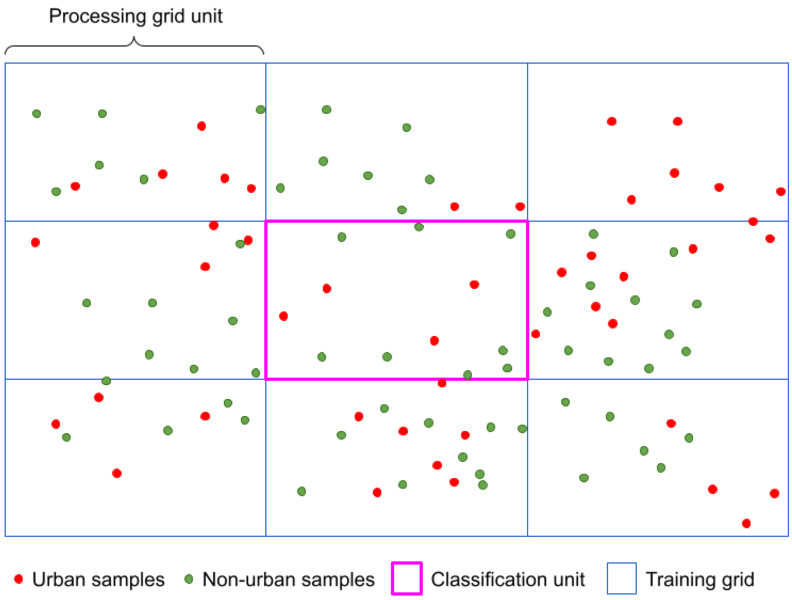


Figure 5. Processing grid for classification scheme.

3.3.4. *Temporal smoothing*

Following classification using the selected samples and mosaic configuration, the probability results were temporally smoothed. This was done by calculating the mean urban classification probability over five-year intervals throughout the entire time series. The procedure aimed to enhance temporal consistency while simplifying post-classification processes. An example illustrating the impact of this approach is presented in FIgure 6.

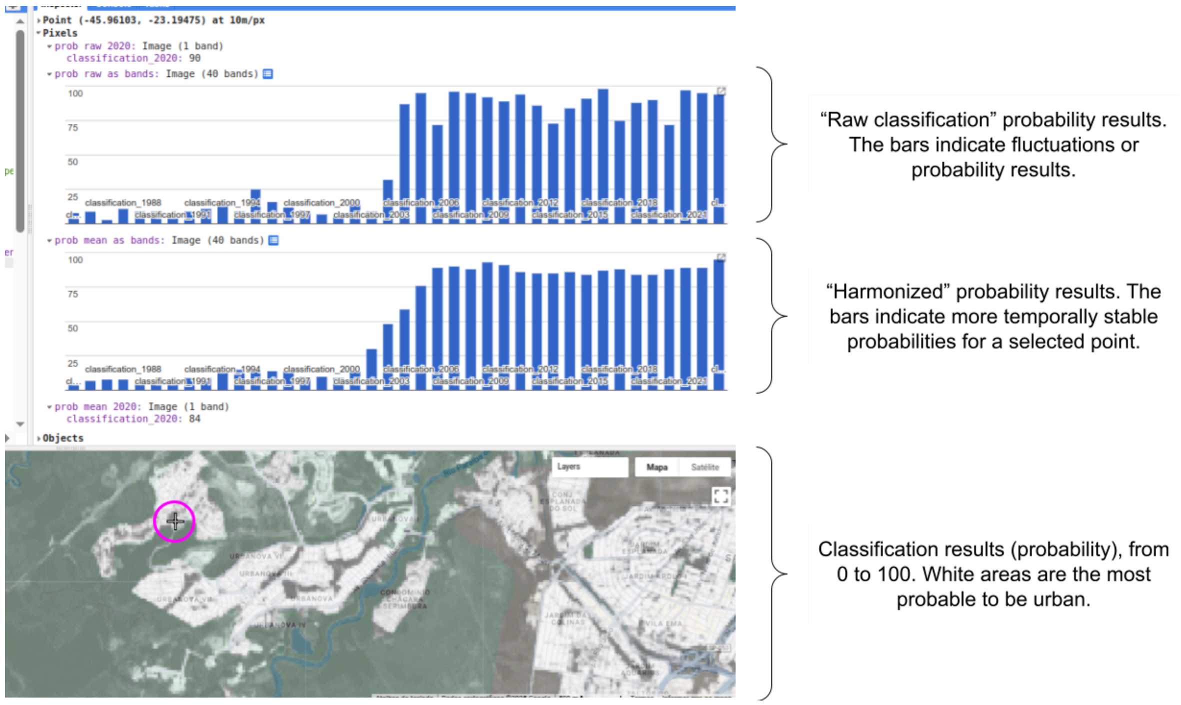


Figure 6. Example of temporal harmonization impact. São José dos Campos - SP, Brazil.

3.3.5. Urban areas binary classification

To produce the final binary classification of urban areas, a thresholding strategy based on ROC curve analysis and probability distribution of urban samples was implemented. For each processing grid unit and year, the optimal classification threshold was estimated by calculating the ROC curve and evaluating the classification probabilities associated with urban reference samples. After comparing the results, the value corresponding to the 15th percentile of urban sample probabilities was selected as the cutoff for binary classification. This iterative procedure was applied across all grid units and time steps. The final threshold value used for each grid was computed as the mean cutoff value derived from the time series, ensuring temporal consistency in the classification. This threshold value was then applied to the probability smoothed results, finally providing a binary classification of urban areas (Figure 7).

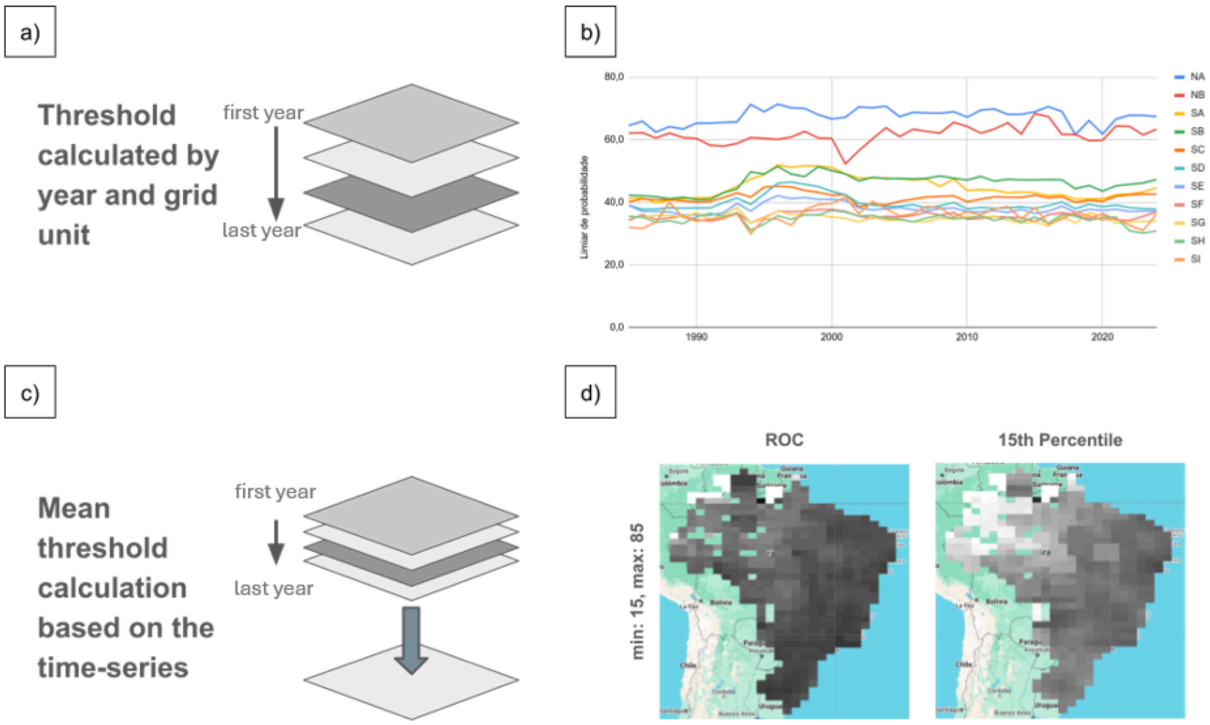


Figure 7. Summary of binary classification procedures. a) Threshold calculation based on ROC curve and percentiles of urban classification gathered from urban samples not used during the classification. b) Example of temporal variation of cut-off values for horizontal sets of grids (covering same latitudes). c) Calculation of mean thresholds (both for ROC and selected percentile) by grid unit based on time-series results. d) Final cut-off values for each approach.

4. Post-classification procedures

Post-classification procedures were developed to improve urban area classification by addressing the inherent noise in the temporal remote sensing data, the limitations imposed by the 30-meter Landsat spatial resolution, and common confusions among land cover types and urban areas (Herold et al., 2003; Lu and Weng, 2007). These procedures also accounted for the typical patterns of urban and settlement configurations in Brazil. The final output is a binary raster distinguishing urban from non-urban areas.

4.1. Gap fill

A gap-filling procedure (Figure 8) was applied to correct no-data pixels by replacing them with the most frequent (mode) class value from neighboring years within a 3-year temporal window (one year previous the year in analysis, and two years forward). Assuming classification quality improves toward recent years, this filtering is applied backward through time. For boundary years (first and last years of the time series), the window was adapted to include all available years, keeping the window size: for the first year, the three subsequent years were used; for 2023, the years 2021, 2022 and 2024, and for 2025, the years 2021, 2022 and 2023 were used to fill no-data values and maintain temporal consistency.

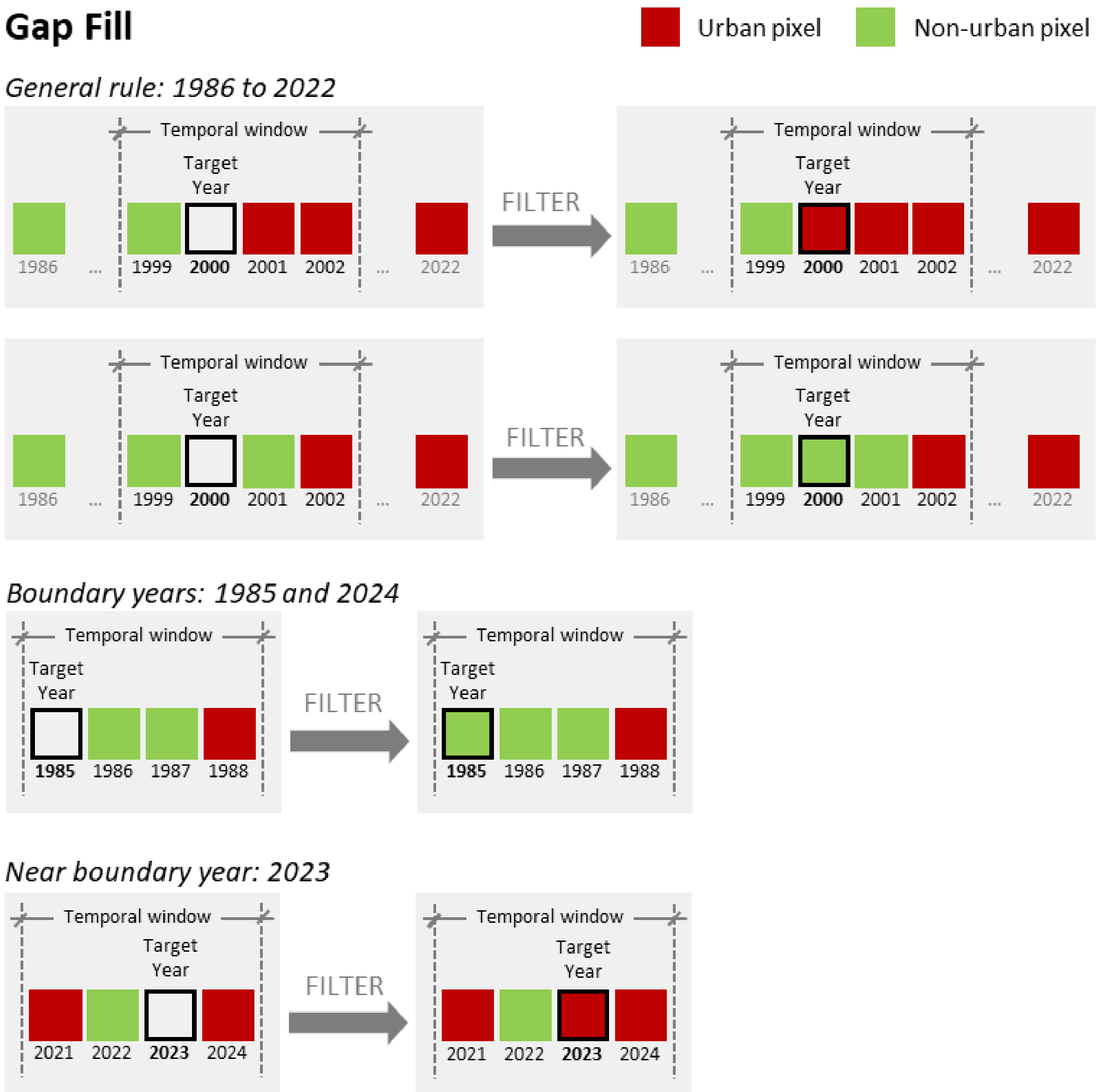


Figure 8. Gap-filling procedure.

4.2. Temporal filter

Pixels in the time series can exhibit one of three behaviors: a unique transition from non-urban to urban, stable urban status throughout, or stable non-urban status throughout.

The temporal filter is designed to preserve these patterns with minimal interference, ensuring that transitions are singular and stable states are consistently maintained over time.

To operationalize this objective, the following steps were taken:

(i) Eliminate isolated urban pixels (Figure 9), i.e. pixels with values differing from their neighbors within a 5-year window (function TempFilter_wMask_5yearsunique): unique urban pixel values are identified within a 5-year window centered on the target year. If the target year's value differs from the majority of the surrounding years, it is corrected to match the majority. Following the same assumption and logic applied in the gap-filling procedure, this filter is applied backward through time. For boundary year (1985) and near boundary years (1986, 2025), the 5-year window is adjusted to include all available neighboring years. The year of 2025 will remain unchanged in all cases.

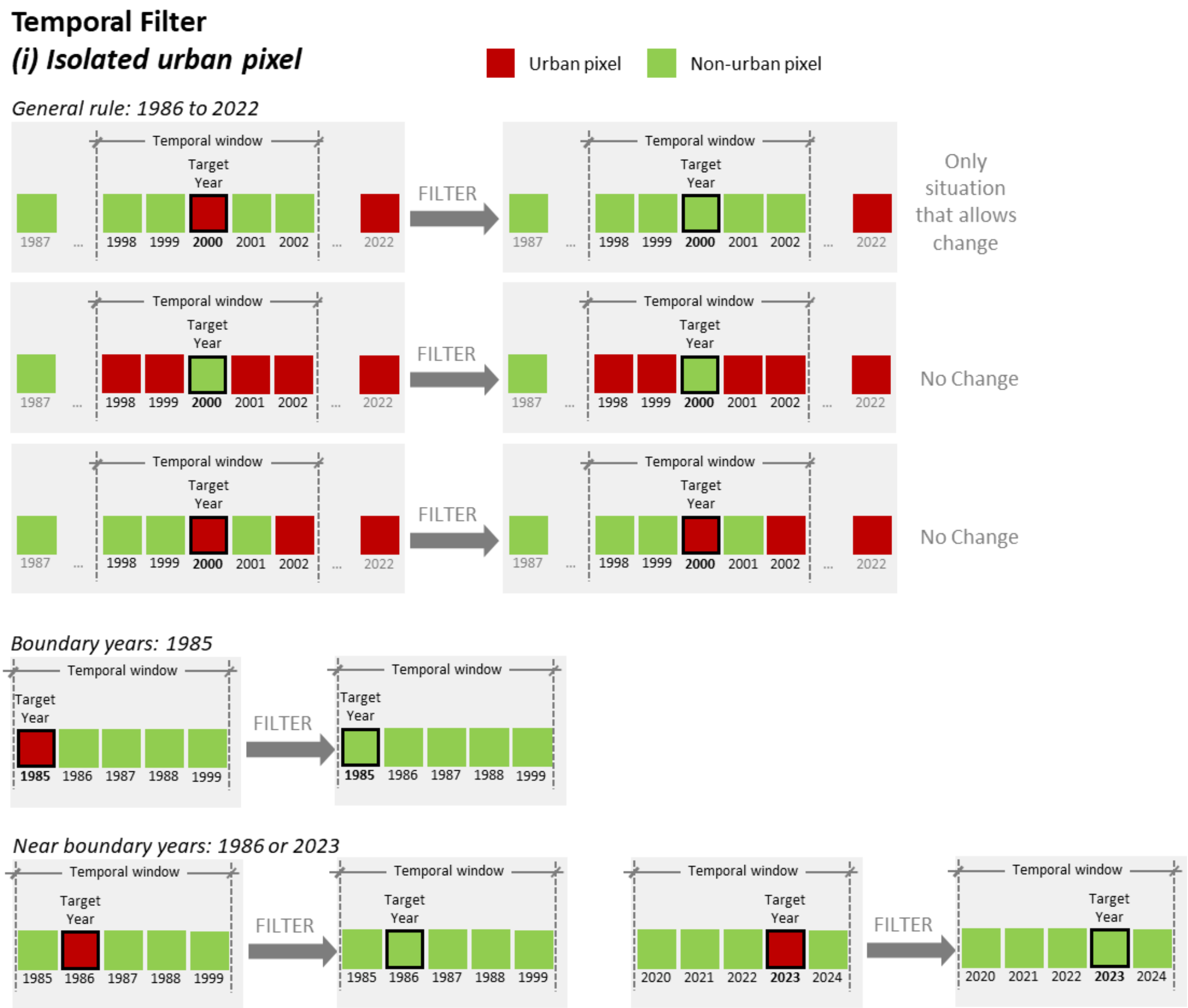


Figure 9. Isolated urban pixels.

(ii) Identify breakpoints (function getBreakpoints) (Figure 10): For each year, a pixel is classified as breakpoints if it: (1) exhibits a valid transition from non-urban to urban in that year; (2) remains urban for at least half of the remaining years in the time series; and (3) its urban persistence is higher than one, i.e. it is urban for at least one subsequent year.

- Valid transitions (function getTransitions_valid) are detected by comparing classifications between consecutive years (the target year and the previous year). Pixels that change from non-urban to urban are marked as valid transitions, assessed annually from 1986 to 2025. Pixels classified as urban in 1985 are assumed to originate from a valid transition.
- Urban persistence is calculated by pixel, for each year, consisting in the number of subsequent years - including the target year - during which the pixel remains classified as urban (function getUrbToEnd).

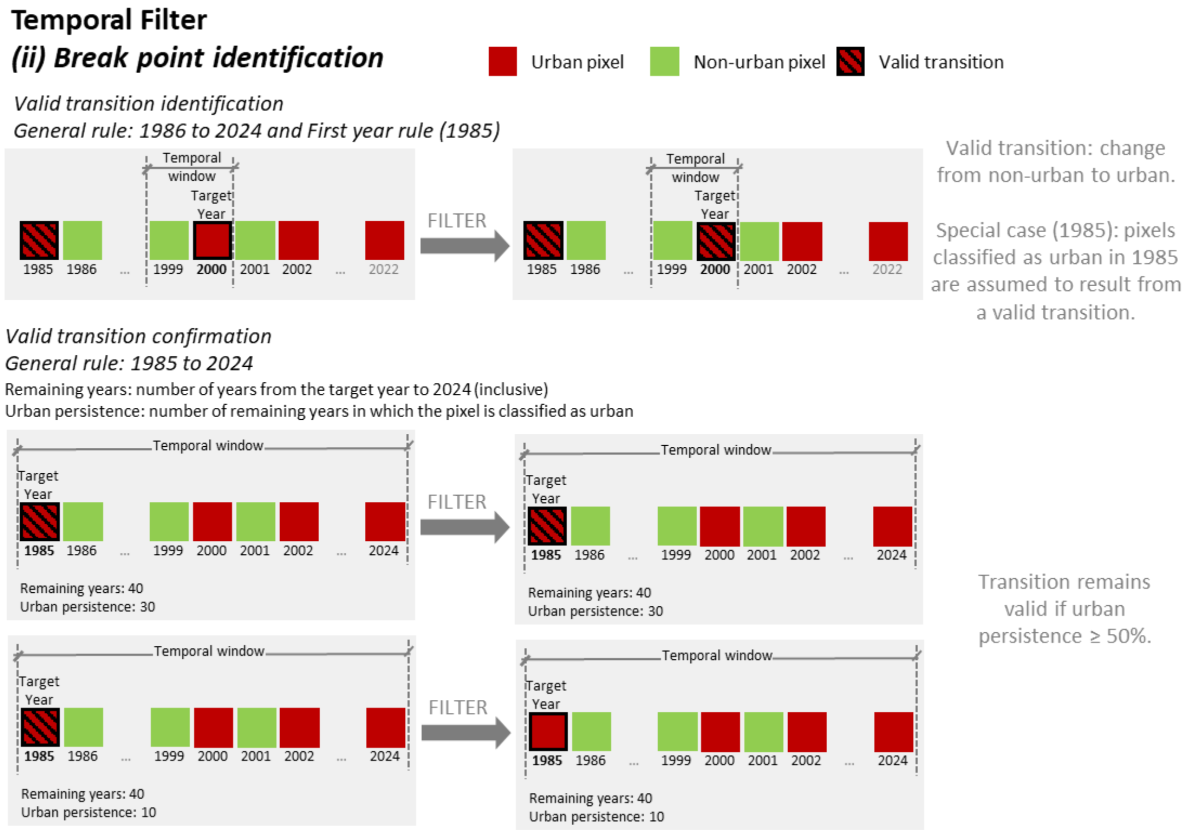


Figure 10. Break point identification.

(iii) Select the first breakpoint and accumulate forward (function accumulateForward) (Figure 11): This final step iterates, for each pixel, through the time-ordered list of breakpoints to identify the first occurrence of a breakpoint. From this point onward, the pixel is consistently classified as urban for all subsequent years, ensuring temporal consistency after the initial transition.

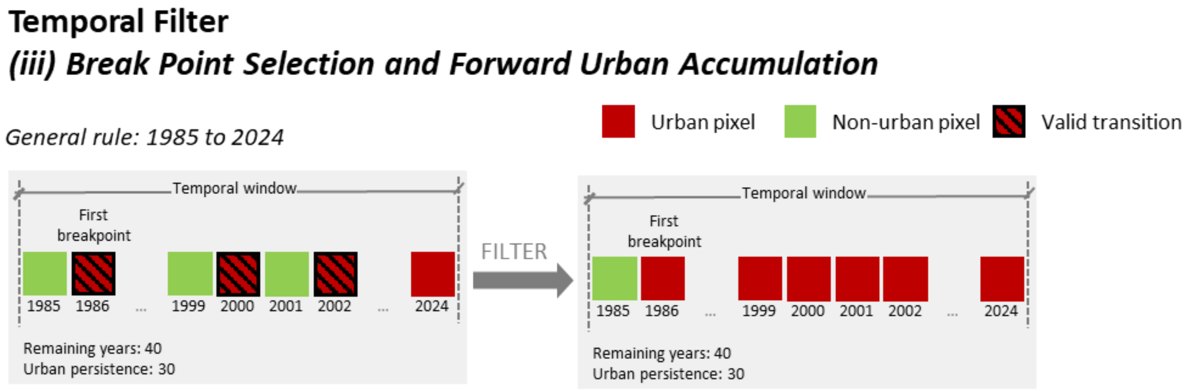


Figure 11. Temporal filter: break point selection and forward urban accumulation.

This combination of steps ensures that breakpoints represent sustained and meaningful urbanization events rather than noise or transient changes.

4.3. Spatial filter

The spatial filter was developed to reduce commission errors by applying a spatial mask to exclude areas unlikely to be urban, and to improve spatial coherence by eliminating small holes within urban areas and scattered, sparse settlements.

To build the spatial mask, high-resolution ancillary datasets were used, including:

- Index of Roads and Infrastructure v2 (IRS) (Justiniano et al., 2022): The IRS defines urban limits according to roads and infrastructure density, derived from Open Street Maps datasets, used with a threshold of greater than or equal to 500.
- Urban areas (IBGE, 2022): This is the reference data for urban areas in Brazil, and was obtained through visual interpretation of Sentinel-2/MSI imagery (10m of spatial resolution for year 2019), supplemented by higher-resolution data where necessary. For the spatial mask, all urban classes — including high-density and low-density urban areas, and vacant urbanized areas — were included. Polygons classified as “other urban equipment,” which typically correspond to areas characterized exclusively by non-residential establishments, were excluded.
- Google Open Buildings v3 (Sirko et al., 2021): Building footprints inferred in May 2023 from high-resolution (50 cm) satellite imagery. Building polygons with confidence ≥65% were buffered by 25 meters to better capture the surrounding built-up environment and rasterized to match the 30-meter resolution of MapBiomas data. Small internal holes (≤5 connected pixels) were filled, and small isolated clusters (≤22 connected pixels) were removed to reduce noise and enhance spatial coherence.
- Favelas and urban communities (IBGE, 2020): This dataset, which serves as a basis for the most recent Demographic Census of Brazil, was used to complement the other datasets, particularly in regions that were underrepresented due to spatial resolution or other limitations of the previous datasets.

All ancillary datasets were converted into binary raster and combined using an inclusive rule: a pixel is included in the spatial mask if it is identified as urban in at least one of these datasets. This spatial mask was then applied sequentially to the temporal filter for each year, excluding areas classified as urban that are likely commission errors—such as bare soil within agricultural fields and rocky outcrops.

The final post-classification step applies filters based on connected pixel counts. For urban holes, a threshold of 280 pixels (approximately 25 hectares) was used, according to IBGE's definition of vacant urban spaces (IBGE, 2022). Areas smaller than this threshold are, therefore, incorporated into the urban final classification.

Conversely, urban noise, defined as isolated small clusters of urban pixels, was reclassified as non-urban. An empirical threshold of 44 pixels (approximately 4 hectares) was applied. These clusters typically correspond to non-urban structures, such as agricultural buildings, infrastructure, or transportation facilities and were removed.

The final product of these procedures is a set of annual raster datasets, spanning from 1985 to 2025, mapping urban areas across the Brazilian territory. This dataset was integrated with other thematic maps to compose MapBiomas Collection 11 Land Cover and Land Use.

5. Comparison with previous collections

For each MapBiomas collection, the classification methodology is updated and the entire time series is reprocessed, resulting in a recalculation of the mapped class areas. As a result, variations in the total mapped urbanized area are expected between collections, as illustrated in Figure 1**2.** For Collection 11, the total mapped urbanized area starts at approximately 1.90 million hectares in 1985 and increases over time, reaching around 4.44 million hectares by 2025.

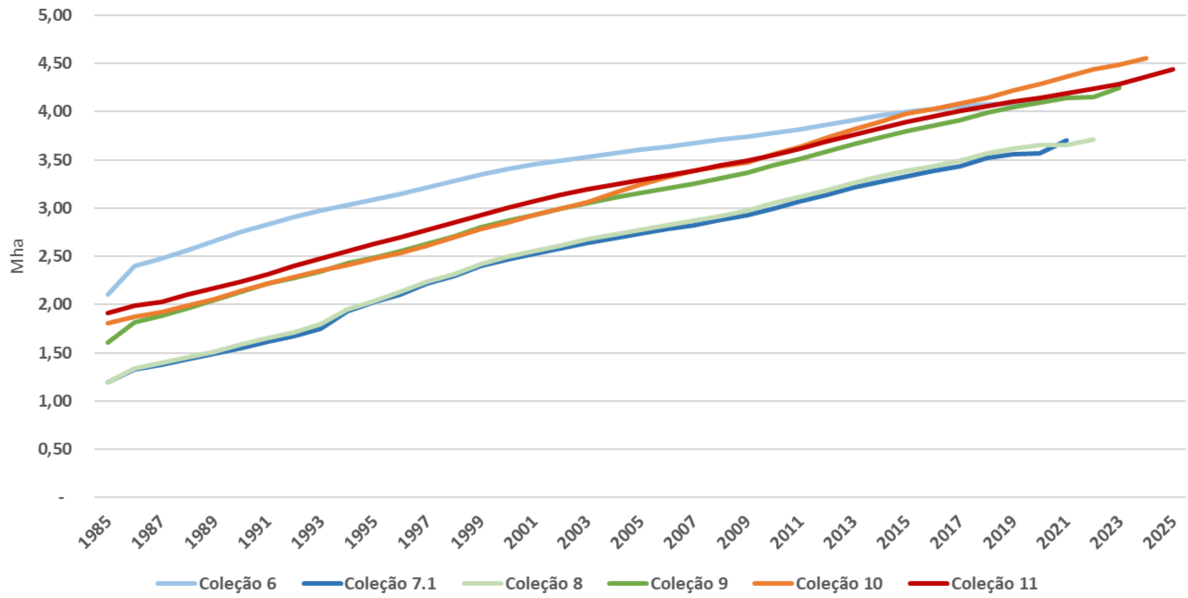
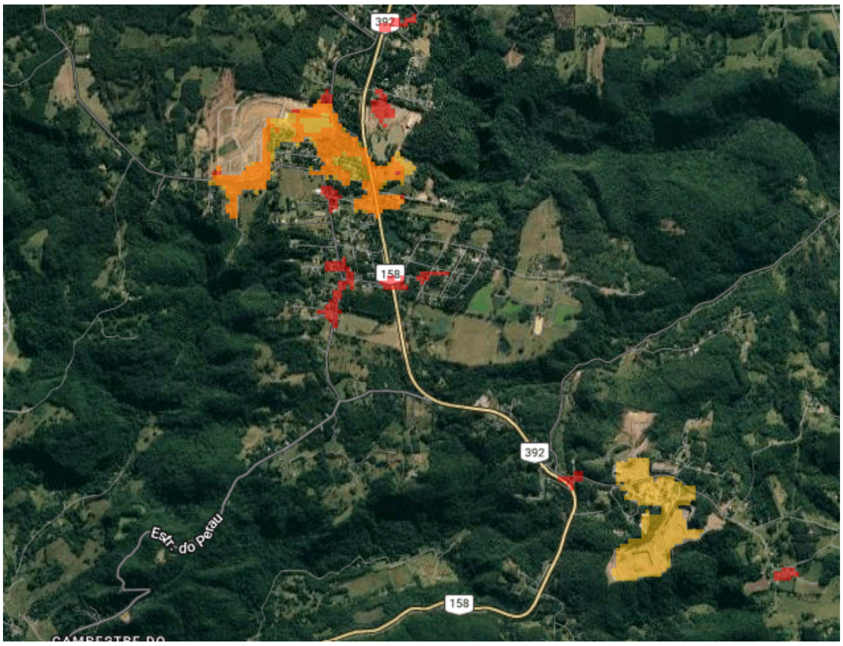


Figure 12. Total urban area (in millions of hectares) mapped by year in MapBiomas Collections 11 and earlier collections.

Figure 12 shows the differences among MapBiomas collections, highlighting how estimates of urban areas have changed with successive methodological updates. Collection 6 is the most distinct in the earlier part of the time series, presenting substantially higher urban-area estimates than Collections 7.1 and 8 between 1985 and the late 1990s. For example, in 1985, Collection 6 mapped 2.10 million hectares of urban area, while Collections 7.1 and 8 mapped only 1.19 and 1.20 million hectares, respectively. Collections 7.1 and 8 follow very similar trajectories throughout their overlapping period, with nearly identical values in most years, indicating strong consistency between these two versions. From Collection 9 onward, the estimated urban area increases relative to Collections 7.1 and 8, especially in the initial years of the series. Collections 10 and 11 show closely aligned trajectories until around the 2000, but Collection 11 reports slightly higher values in most years up to 2009. After 2010, however, Collection 10 generally estimates larger urban areas than Collection 11, particularly between 2014 and 2024. In the most recent period, Collection 11 provides the longest and most updated series, reaching 4.44 million hectares in 2025, while Collection 10 reaches 4.55 million hectares in 2024. Overall, the comparison indicates that differences among collections are more pronounced in the earlier years and tend to narrow over time, although relevant divergences remain between the most recent collections.

A qualitative comparison between Collections 10 and 11 highlights clear improvements in spatial detail. As shown in 13, Collection 11 presents a slightly more refined delineation of urban area, particularly along the urban-rural interface. Compared to Collection 10, Collection 11 more accurately captures the built-up footprint, reducing commission errors in adjacent vegetated areas (Figure 13 [A]) and in areas of bare soil (Figure 13 [B]).

[A]



[B]

Figure 13A and 13B. Comparison of urban area mapping between Collections 10 and 11 in two locations: [A] Rio de Janeiro (RJ) and [B] Santa Maria (RS). In each location, yellow represents the urban area mapped in Collection 10, red represents Collection 11 and orange represents both collections.

6. Validation Strategies

6.1. Accuracy Analysis

Following MapBiomas LULC validation strategy, the error assessment analysis was conducted using ~75,000 samples per year, labeled according to MapBiomas LULC classes by experts after the visual interpretation of Landsat data, MODIS-NDVI times series, and high-resolution imagery from Google Earth (when available). The accuracy analysis was based on Stehman (Stehman, 2014; Stehman and Foody, 2019) using the population error matrix and the global, user, and producer accuracies.

The accuracy results are published on the MapBiomas website[2] and are reproduced here for the Urban Area class in Figure 14 (Producer's Accuracy) and Figure 15 (User's Accuracy).

[2] https://brasil.mapbiomas.org/en/estatistica-de-acuracia/

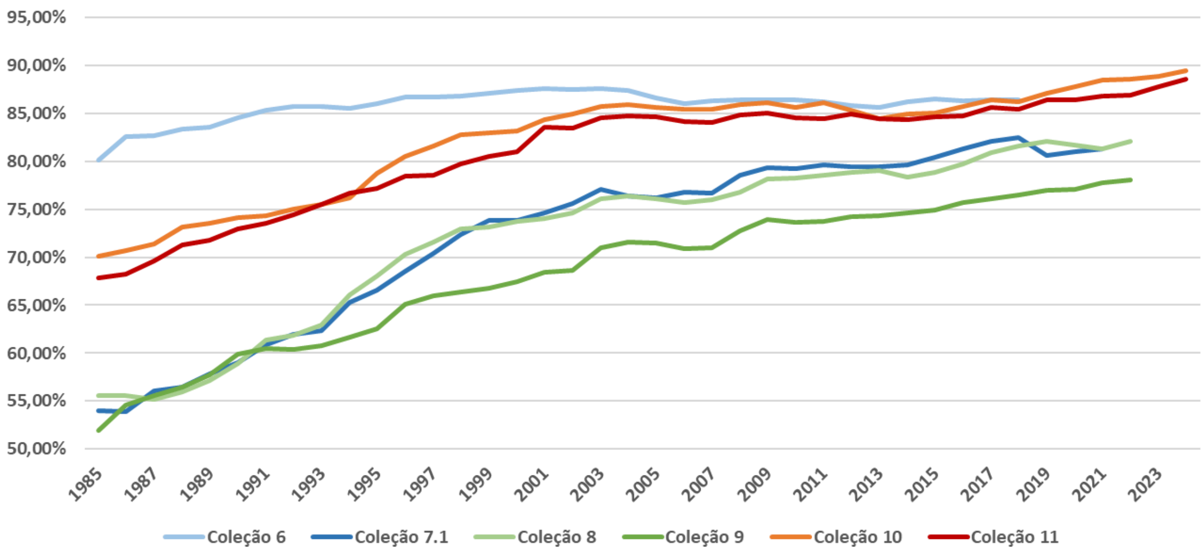


Figure 14**.** Producer's accuracy.

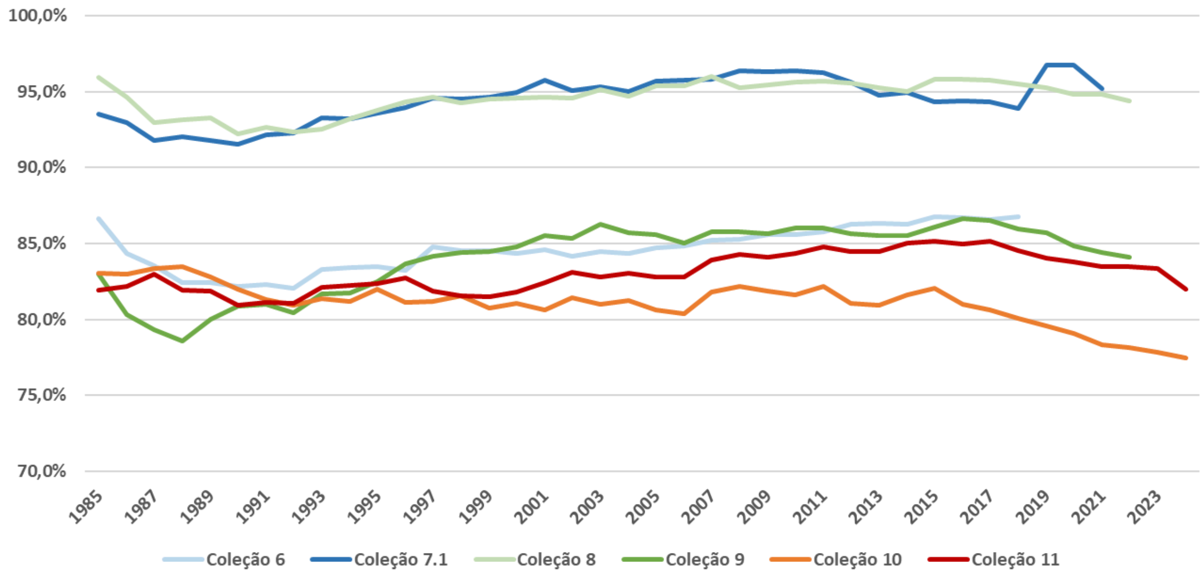


Figure 15. User's accuracy.

In terms of Producer's Accuracy, Collections 10 and 11 show similar performance throughout the time series, with both presenting a consistent increasing trend. Collection 11 remains slightly below Collection 10 in most years, although the difference gradually decreases toward the end of the series. By 2024, Producer's Accuracy reaches approximately 88.5% in Collection 11, compared with 89.5% in Collection 10, indicating that the new collection maintains a comparable capacity to capture urban areas while keeping omission errors at a similarly low level.

For User's Accuracy, Collection 11 shows a clear improvement over Collection 10, particularly from the mid-1990s onwards. While both collections present similar values in the early years, Collection 11 increasingly outperforms Collection 10 over time. In the most recent years, User's Accuracy remains around 82–84% in Collection 11, whereas it decreases to approximately 77–79% in Collection 10. Overall, Collection 11 improves the reliability of urban-area identification by reducing commission errors, while largely preserving the high Producer's Accuracy achieved by Collection 10.

6.2. Comparison with reference maps

MapBiomas Collection 11.0 were compared to two urban area maps: (1) the World Settlement Footprint (WSF) produced by Deutsches Zentrum für Luftund Raumfahrt (DLR) (Marconcini et al., 2020) and (2) Brazil Urbanized Areas produced by IBGE, Instituto Brasileiro de Geografia e Estatística (IBGE, 2022).

WSF is a 10m resolution binary mask outlining the extent of human settlements globally derived by means of 2014-2015 multitemporal Landsat-8 and Sentinel-1 imagery, using different classification schemes based on Support Vector Machines. It is available at Earth Engine Data Catalog[3].

Quantitative analysis (Table 2) indicates that the urbanized area mapped by MapBiomas Collection 11.0 in 2015 totals 3,842,147 hectares, exceeding the 3,421,914 hectares mapped by the World Settlement Footprint (WSF). MapBiomas exclusively maps 1,014,071 hectares, whereas WSF exclusively maps 593,839 hectares. The two datasets share an overlapping area of 2,828,076 hectares, corresponding to 82.6% of the total area mapped by WSF across Brazil. At the biome level, overlap relative to WSF ranges from 80.5% in the Caatinga to 85.0% in the Amazon. The Cerrado (84.8%) and Pampa (84.2%) also show comparatively high agreement, while the Atlantic Forest (81.8%) and Pantanal (81.7%) present intermediate values. Thus, overlap exceeds 80% in all six Brazilian biomes.

Table 2. Comparison of Urban Area Mapping between MapBiomas Collection 11 and the World Settlement Footprint-WSF for the Year 2015.

Biome	Total Area Mapped by Col. 11 (in ha)	Total Area Mapped by WSF (in ha)	Area Only Mapped by Col. 11 (in ha)	Area Only Mapped by WSF (in ha)	Overlapping Area (in ha)	Overlap Relative to WSF (in %)
Amazon	421175	307698	159497	46019	261678	85,04
Caatinga	445975	335600	175861	65487	270113	80,49
Cerrado	857472	694352	268448	105327	589025	84,83
Atlantic Forest	1980189	1952023	384103	355937	1596086	81,77
Pampa	131276	126136	25089	19948	106188	84,19
Pantanal	6059	6106	1074	1120	4986	81,66

Brazil Urbanized Areas is a visual interpretation of urban features, identified according to the elements of specific shape (geometry of objects) and pattern (spatial arrangement), using Sentinel 2 imagery, with spatial resolution of 10m, supplemented by

[3] https://developers.google.com/earth-engine/datasets/catalog/DLR_WSF_WSF2015_v1

higher-resolution data where necessary. It is available in shapefile format at IBGE's website[4]. The mapped urban land use types include: "Urbanized Area," categorized into two classes — high density and low density —, "Other Urban Facilities," and "Vacant Urbanized Areas."

The comparison with IBGE's 2019 Urbanized Areas dataset indicates that MapBiomas Collection 11.0 maps a smaller total urbanized extent (3). MapBiomas maps 4,042,893 hectares, compared with 4,592,407 hectares identified by IBGE. The area mapped exclusively by MapBiomas amounts to 627,172 hectares, whereas IBGE exclusively maps 1,176,686 hectares. The overlapping area totals 3,415,721 hectares, corresponding to 74.4% of the total area mapped by IBGE across Brazil. At the biome level, the highest overlap occurs in the Pantanal (86.0%), followed by the Cerrado (81.3%). The Amazon (74.5%), Atlantic Forest (73.7%), and Pampa (73.0%) show intermediate levels of agreement, while the Caatinga has the lowest overlap, at 66.7%.

Table 3. Comparison of Urban Area Mapping between MapBiomas Collection 11 and the Urbanized Areas (IBGE, 2022) for the Year 2019.

Biome	**Total Area Mapped by Col. 11 (in ha)**	**Total Area Mapped by IBGE AU (in ha)**	**Area Only Mapped by Col. 11 (in ha)**	**Area Only Mapped by IBGE AU (in ha)**	**Overlapping Area (in ha)**	**Overlap Relative to IBGE AU (in %)**
Amazon	437715	476083	83052	121420	354662	74,5
Caatinga	476903	555043	106602	184742	370301	66,72
Cerrado	906457	897509	177234	168286	729223	81,25
Atlantic Forest	2079937	2486448	248659	655170	1831279	73,65
Pampa	135689	170720	11110	46141	124579	72,97
Pantanal	6192	6605	515	927	5677	85,96

7. Urban Module

The Urban Module of the MapBiomas platform provides a comprehensive mapping of urbanized areas in Brazil from 1985 to 2025, based on Collection 11 data. It offers access to urbanization data by year and in 5-year intervals, as well as information on urban vegetation and urbanization patterns relative to slope and elevation from the nearest drainage. The methodology used to generate each of the module's layers is detailed in the following sections.

[4] https://www.ibge.gov.br/geociencias/cartas-e-mapas/redes-geograficas/15789-areas-urbanizadas.html?=&t=acesso-ao-produto

7.1. Urban vegetation

The urban vegetation layer provides a detailed breakdown of the class Urban Area (ID=24) from the MapBiomas Collection 11 land use and land cover data, along with natural vegetation remnants that lie inside the urban domain but outside the urban class. The final product is a multi-year dataset (1985–2025) with three thematic components: intra-urban vegetation (green areas inside the urban footprint), peri-urban vegetation (natural vegetation classes within the urban domain but outside the urban class), and aggregated vegetation (the union of both). Non-vegetated surfaces, water bodies, and farming areas are excluded from the peri-urban and aggregated layers, ensuring that the product represents only natural vegetation cover (forest and herbaceous/shrubby formations). Additionally, a vegetation availability index is calculated for each year, representing the percentage of the total urban area covered by intra-urban vegetation.

The methodology is implemented entirely in Google Earth Engine and comprises three main steps: (1) definition of the urban domain mask; (2) generation of intra-urban and peri-urban vegetation layers; and (3) final assembly and calculation of availability indices. Each step is detailed below.

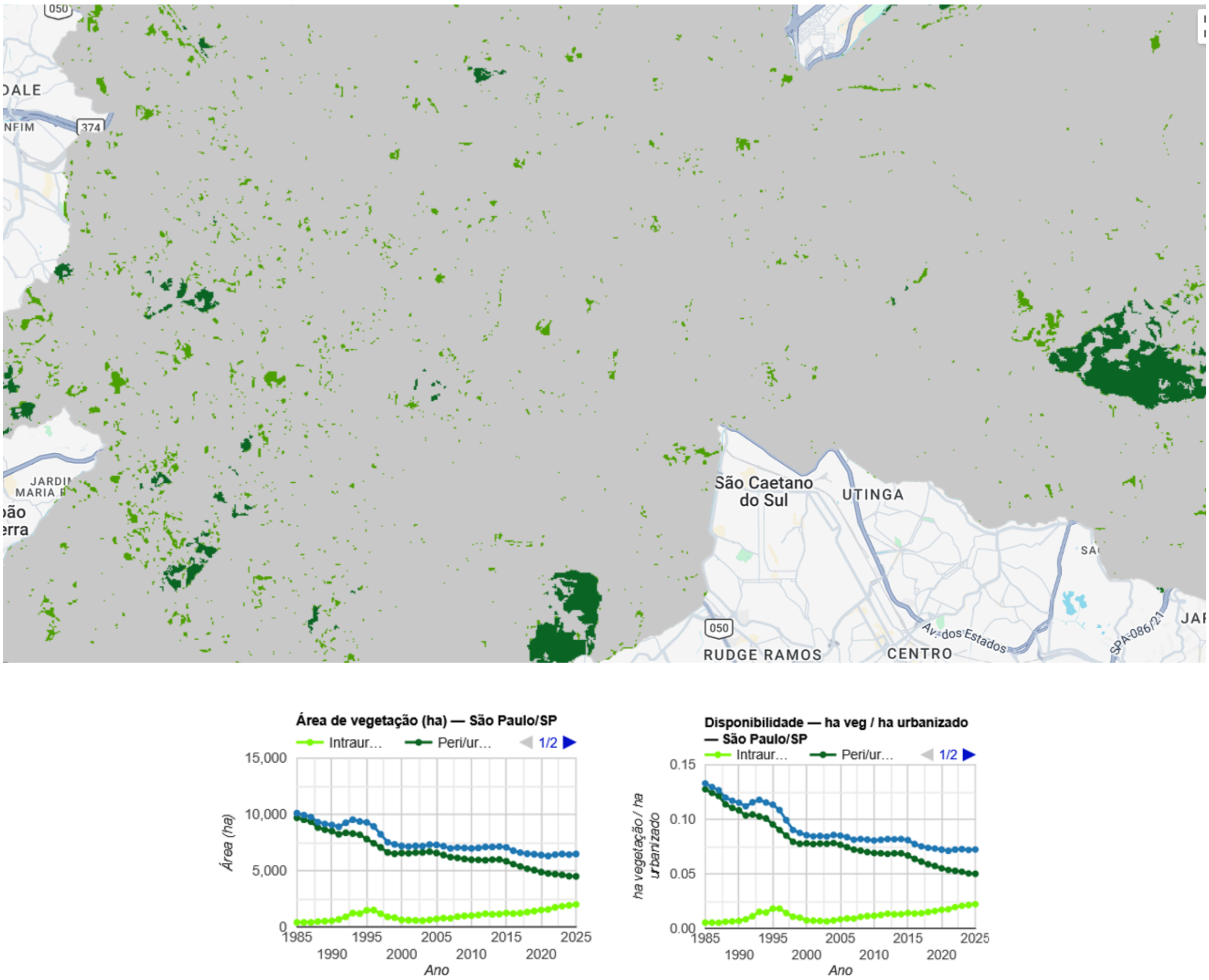


Figure 16. A portion of the city of São Paulo in 2025, showing: Urban Areas (gray) intra-urban vegetation layer data (light green) and peri-urban vegetation data (dark green), followed by the time series charts of São Paulo showing the urban vegetation area over time and it’s availability (vegetation/urbanized area for each year).

7.1.1. *Procedures*

7.1.1.1. Definition of the Urban Domain Mask

The analysis is restricted to a specific spatial domain defined by the intersection of two datasets:

- Urban Census Tracts (uct): A feature collection of official urban census tract boundaries, provided by the Brazilian Institute of Geography and Statistics (IBGE). These polygons delineate the legally defined urban perimeter of each municipality.
- Most Recent Urban Area: The urban class (ID=24) from the MapBiomas Collection 11 classification for the year 2025, which represents the most up-to-date mapping of built-up areas.

The urban domain mask is created as the union (logical OR) of these two sources. This mask ensures that all pixels that are either within the official urban perimeter or classified as urban in the latest MapBiomas product are included in the analysis. This approach avoids excluding newly urbanised areas that may fall outside the census tract boundaries, while also capturing non-urban voids that are entirely surrounded by the city. All subsequent operations are clipped to this mask.

7.1.1.2. Generation of Intra-Urban and Peri-Urban Vegetation

Intra-urban vegetation was derived from a pre-existing MapBiomas asset: projects/mapbiomas-brazil/assets/LAND-COVER/COLLECTION-11/URBAN/vegintraurb. This asset contains annual binary maps (1985–2025) indicating the presence of vegetation within urban areas. The classification was produced using a Random Forest classifier applied to the full Landsat time series, with spectral indices and textural metrics as predictors. The classifier was trained on a comprehensive set of reference samples, specifically designed to distinguish vegetated from non-vegetated surfaces applied inside the urban footprint. Following the classification, a temporal consistency filter was applied to the intra-urban vegetation results to ensure that only pixels exhibiting persistent vegetation cover over multiple years were retained, reducing commission errors caused by seasonal variations or classification noise. The output is a binary raster where value 1 represents intra-urban vegetation and 0 represents non-vegetated urban surfaces (paved or built-up).

Peri-urban vegetation is generated on-the-fly from the MapBiomas Collection 11 annual land use and land cover maps. Within the urban domain mask, pixels that are not classified as Urban Area (ID=24) are evaluated. To compose the peri-urban vegetation component, only classes representing natural vegetated cover are retained. These include:

- Forest: classes 1 (Forest Formation), 2 (Savanna Formation), 3 (Mangrove), 4 (Floodable Forest), 5 (Wooded Sandbank Vegetation), and 6 (Wetland).
- Herbaceous and Shrubby Vegetation: classes 10 (Grassland), 11 (Herbaceous Sandbank Vegetation), 12 (Hypersaline Tidal Flat), 29 (Rocky Outcrop), and 50 (Herbaceous Shrubby Vegetation).

Classes representing farming (14 – Pasture; 15 – Agriculture; 18 – Tree Plantation; 19 – Mosaic of Uses), water (25 – River, Lake and Ocean; 26 – Aquaculture; 33 – Wetland Water), and non-vegetated surfaces (21 – Beach, Dune and Sand Spot; 22 – Mining; 23 – Photovoltaic Power Plant; 30 – Other non-Vegetated Areas) are explicitly excluded from the peri-urban layer. No additional temporal or spatial filters are applied to the peri-urban vegetation component, as the MapBiomas Collection 11 land use and land cover product already undergoes rigorous temporal consistency checks and validation by the MapBiomas team. The annual LULC maps are produced using a systematic approach that includes temporal smoothing and cross-year consistency analysis, ensuring that the classified vegetation classes are reliable and temporally coherent without further filtering at this stage.

This strict selection ensures that the peri-urban component, and consequently the aggregated vegetation layer, contains only natural vegetation cover (forest and herbaceous/shrubby formations), omitting managed agricultural areas, barren surfaces, rocky outcrops, and aquatic features. The intra-urban component, having undergone its own temporal consistency filter, complements the peri-urban component to produce a coherent and conservative representation of all natural vegetation within the urban domain.

7.1.1.3. Final Assembly and Availability Indices

In the final step, the vegetation dataset is separated into its original components using the urban area mask:

- Intra-urban vegetation: the pixels from the vegintraurb asset (with temporal consistency filter already applied) that fall within the Urban Area (ID=24) class.
- Peri-urban vegetation: the reclassified natural vegetation pixels (Forest and Herbaceous and Shrubby Vegetation) that fall outside the Urban Area class but within the urban domain mask (i.e., the voids).

The aggregated vegetation layer is the union of these two components, representing all natural vegetated cover within the urban domain.

For each year, the code calculates the vegetation availability index as: Availability (%) = (area of intra-urban vegetation / total urban area) × 100. Where the total urban area corresponds to the area covered by the Urban Area class (ID=24) within the urban domain mask. This metric provides a quantitative summary of greenness within the city, complementing the spatial distribution shown in the map.

The final legend is hierarchical:

- Level 1 – Urban Area
    - Non-vegetated Urban Area (ID 240) – paved or built-up surfaces.
    - Vegetated Urban Area (ID 241) – all intra-urban green spaces (parks, squares, gardens, etc.).
- Level 1 – Non-urban Area (voids) – this level exclusively comprises natural vegetation classes, as farming, water, and non-vegetated surfaces are excluded from the vegetation product.
    - Forest – aggregated from Forest Formation, Savanna Formation, Mangrove, Floodable Forest, and Wooded Sandbank Vegetation.
    - Herbaceous and Shrubby Vegetation – aggregated from Grassland, Herbaceous Sandbank Vegetation, Hypersaline Tidal Flat, Rocky Outcrop, and Herbaceous Shrubby Vegetation.
      All processing is performed within the user-selected municipality, using the municipality boundary for clipping. The intra-urban vegetation asset is used

directly with its pre-applied temporal consistency filter, while the peri-urban component is derived from the temporally consistent MapBiomas Collection 11 LULC classifications.

7.2. Urbanization periods

The Urbanization Periods layer depicts the spatial and temporal dynamics of urban expansion in Brazil from 1985 to 2025 derived from the MapBiomas Collection 11 land-use and land-cover time series and aggregated into 5-year intervals.

For each year, pixels classified as “Urban Area” (ID = 24) were extracted and assigned a temporary value corresponding to the year of observation. These annual layers were combined into a multi-band image, and a pixel-wise minimum was computed to identify the earliest year each pixel was urbanized. The resulting raster was then reclassified into 5-year periods, producing a single-band image where each value represents the urbanization period of that pixel, according to Table 4.

Table 4. Complete Legend of the Urbanization Periods Layer.

Class name	*Class ID*	*Color*	*Color Code*
Urbanized up to 1985	1		#bfbfbf
Urbanized between 1986 and 1990	2		#c0afc8
Urbanized between 1991 and 1995	3		#a8a9db
Urbanized between 1996 and 2000	4		#9f8ed2
Urbanized between 2001 and 2005	5		#7d7ed3
Urbanized between 2005 and 2010	6		#7a6fc9
Urbanized between 2011 and 2015	7		#b05ea6
Urbanized between 2016 and 2020	8		#b73875
Urbanized between 2021 and 2025	9		#930000

7.3. Urban slope

The Urban Slope layer depicts the topographic characteristics of urban areas in Brazil derived from digital elevation model (DEM), classified into four slope categories according to Table 6. The DEM used was the NASA 30m Digital Elevation Model (NASA JPL, 2020), available directly on GEE.

For each pixel classified as "Urban Area" (ID = 24) in the MapBiomas Collection 11 land-use and land-cover product, slope values were extracted from a digital elevation model and classified into four categories based on percentage inclination. The resulting raster represents the predominant slope class of each urban pixel, according to Table 5.

Table 5. Complete Legend of the Urban Slope Layer.

Level	*Class name*	*Class ID*	*Color*	*Color Code*
1	Slope ≤ 10%	132		#ffc6a8

1	Slope between 10% and 20%	133		#ff7a34
1	Slope between 20% and 30%	134		#e63b22
1	Slope > 30%	135		#8c1a14

7.4. *Height above nearest drainage*

The HAND (Height Above Nearest Drainage) layer depicts the relative elevation of urban areas in Brazil in relation to the nearest drainage network, classified into three categories according to Table 6. A globally available HAND model (Donchyts et al., 2016), originally generated with an accumulation threshold of 100 cells and available directly on GEE, was employed. For each pixel classified as "Urban Area" (ID = 24) in the MapBiomas Collection 11 land-use and land-cover product, HAND values were extracted and classified into three categories.

Table 6**.** Complete Legend of the Urban HAND Layer.

Level	*Class name*	*Class ID*	*Color*	*Color Code*
1	HAND ≤ 3 m	142		#9cc8f2
1	3 < HAND ≤ 6 m	143		#41ab5d
1	HAND > 6 m	144		#d9b26f

8. References

As-syakur, A.R., Adnyana, I.W.S., Arthana, I.W., Nuarsa, I.W., 2012. Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. Remote Sens. 4, 2957–2970. https://doi.org/10.3390/rs4102957

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30, 1145–1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Corbane, C., Florczyk, A., Pesaresi, M., Politis, P., Syrris, V., 2018. GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014) - OBSOLETE RELEASE. https://doi.org/10.2905/jrc-ghsl-10007

Donchyts, G., Winsemius, H., Schellekens, J., Erickson, T., Gao, H., Savenije, H., van de Giesen, N., 2016. Global 30m Height Above the Nearest Drainage. https://doi.org/10.13140/RG.2.1.3956.8880

Elvidge, C.D., Zhizhin, M., Ghosh, T., Hsu, F.-C., Taneja, J., 2021. Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. Remote Sens. 13, 922. https://doi.org/10.3390/rs13050922

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett., ROC Analysis in Pattern Recognition 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sens. Environ. 140, 23–35. https://doi.org/10.1016/j.rse.2013.08.029

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Herold, M., Liu, X., Clarke, K.C., 2003. Spatial Metrics and Image Texture for Mapping Urban Land Use. Photogramm. Eng. Remote Sens. 69, 991–1001. https://doi.org/10.14358/pers.69.9.991

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ., The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

IBGE, 2020. Instituto Brasileiro de geografia e Estatística (IBGE). Malha de Setores Censitários [WWW Document]. URL https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html

IBGE, 2017. Áreas urbanizadas do Brasil: 2015, Coordenação de Geografia. IBGE, Rio de Janeiro.

IBGE, C. de M.A. (Ed.), 2022. Áreas urbanizadas do Brasil: 2019. Ibge, Rio de Janeiro, RJ.

Justiniano, E.F., Santos Junior, E.R. dos, Melo, B.M. de, Siqueira, J.V.N., Morato, R.G., Fantin, M., Pedrassoli, J.C., Martines, M.R., Kawakubo, F.S., 2022. Proposal for an index of roads and structures for the mapping of non-vegetated urban surfaces using OSM and Sentinel-2 data. Int. J. Appl. Earth Obs. Geoinformation 109, 102791. https://doi.org/10.1016/j.jag.2022.102791

Kawamura, M., Jayamanna, S., Tsujiko, Y., 1997. Quantitative Evaluation of Urbanization in Developing Countries Using Sa℡lite Data. Doboku Gakkai Ronbunshu 1997, 45–54. https://doi.org/10.2208/jscej.1997.580_45

Key, C.H., Benson, N.C., 2006. Landscape Assessment: Ground Measure of Severity, the Composite Burn Index; and Remote Sensing of Severity, the Normalized Burn Ratio, in: Lutes, D.C., Keane, R.E., Caratti, J.F., Key, C.H., Benson, N.C., Sutherland, S., Gangi, L.J. (Eds.), FIREMON: Fire Effects Monitoring and Inventory System, Gen. Tech. Rep. USDA Forest Service, Rocky Mountain Research Station, Ogden, UT.

Lu, D., Weng, Q., 2007. A survey of image classification methods and techniques for improving classification performance. Int. J. Remote Sens. 28, 823–870. https://doi.org/10.1080/01431160600746456

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., Paganini, M., Strano, E., 2020. Outlining where humans live, the World Settlement Footprint 2015. Sci. Data 7, 242. https://doi.org/10.1038/s41597-020-00580-5

McFEETERS, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. https://doi.org/10.1080/01431169608948714

NASA JPL, 2020. NASADEM Merged DEM Global 1 arc second V001. NASA EOSDIS Land Process. Distrib. Act. Arch. Cent. DAAC Data Set NASADEM_HGT.001. https://doi.org/10.5067/MEASURES/NASADEM/NASADEM_HGT.001

Rikimaru, A., Roy, P.S., Miyatake, S., 2002. Tropical forest cover density mapping. Int. Soc. Trop. Ecol. 43, 39–47.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA NTRS, Texas.

Santos, B.D. dos, Pinho, C.M.D. de, Oliveira, G.E.T., Korting, T.S., Escada, M.I.S., Amaral, S., 2022. Identifying Precarious Settlements and Urban Fabric Typologies Based on GEOBIA and Data Mining in Brazilian Amazon Cities. Remote Sens. 14, 704. https://doi.org/10.3390/rs14030704

Sirko, W., Kashubin, S., Ritter, M., Annkah, A., Bouchareb, Y.S.E., Dauphin, Y., Keysers, D., Neumann, M., Cisse, M., Quinn, J., 2021. Continental-Scale Building Detection from High Resolution Satellite Imagery. https://doi.org/10.48550/arXiv.2107.12283

Small, C., Milesi, C., 2013. Multi-scale standardized spectral mixture models. Remote Sens. Environ. 136, 442–454. https://doi.org/10.1016/j.rse.2013.05.024

Souza, C.M., Roberts, D.A., Cochrane, M.A., 2005. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. Remote Sens. Environ. 98, 329–343. https://doi.org/10.1016/j.rse.2005.07.013

Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. Int. J. Remote Sens. 35, 4923–4939. https://doi.org/10.1080/01431161.2014.930207

Stehman, S.V., Foody, G.M., 2019. Key issues in rigorous accuracy assessment of land cover products. Remote Sens. Environ. 231, 111199. https://doi.org/10.1016/j.rse.2019.05.018

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033. https://doi.org/10.1080/01431160600589179

Zha, Y., Gao, J., Ni, S., 2003. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. Int. J. Remote Sens. 24, 583–594. https://doi.org/10.1080/01431160304987

Supplemental material

Table S1**.** Landsat imagery used in Urban Area mosaics within Google Earth Engine (GEE).

Landsat Mission, Sensor	Collection, Tier	Period	Level	GEE Collection ID	Bands [wavelength]
Landsat 5, TM	Collection 2, Tier 1	1984 to 2012	Surface Reflectance	LANDSAT/LT05/ C02/T1_L2	SR_B1: Blue [0.45-0.52 µm] SR_B2: Green [0.52-0.60 µm] SR_B3: Red [0.63-0.69 µm] SR_B4: Near Infrared [0.77-0.90 µm] SR_B5: Shortwave Infrared 1 [1.55-1.75 µm] SR_B7: Shortwave Infrared 2 [2.08-2.35 µm]
			Raw Images	LANDSAT/LT05/ C02/T1	B4: Near infrared [0.76 - 0.90 µm] B5: Shortwave infrared 1 [1.55 - 1.75 µm] B6: Thermal Infrared 1 (resampled from 60m to 30m) [10.40 - 12.50 µm]
Landsat 7, ETM+	Collection 2, Tier 1	2010 to 2016	Surface Reflectance	LANDSAT/LE07/ C02/T1_L2	SR_B1: Blue [0.45-0.52 µm] SR_B2: Green [0.52-0.60 µm] SR_B3: Red [0.63-0.69 µm] SR_B4: Near Infrared [0.77-0.90 µm] SR_B5: Shortwave Infrared 1 [1.55-1.75 µm] SR_B7: Shortwave Infrared 2 [2.08-2.35 µm]
			Raw Images	LANDSAT/LE07/ C02/T1	B4: Near Infrared [0.77 - 0.90 µm] B5: Shortwave Infrared 1 [1.55 - 1.75 µm] B6_VCID_1: Low-gain Thermal Infrared 1 (resampled from 60m to 30m) [10.40 - 12.50 µm]
Landsat 8, OLI / TIRS	Collection 2, Tier 1	2013 to 2024	Surface Reflectance	LANDSAT/LC08/ C02/T1_L2	SR_B2: Blue [0.45 - 0.51 µm] SR_B3: Green [0.53 - 0.59 µm] SR_B4: Red [0.64 - 0.67 µm] SR_B5: Near Infrared [0.85 - 0.88 µm] SR_B6: Shortwave Infrared 1 [1.57 - 1.65 µm] SR_B7: Shortwave Infrared 2 [2.11 - 2.29 µm]
			Raw Images	LANDSAT/LC08/ C02/T1	B5: Near infrared [0.85 - 0.88 µm] B6: Shortwave infrared 1 [1.57 - 1.65 µm] B10: Thermal infrared 1 (resampled from 100m to 30m) [10.60 - 11.19 µm]
Landsat 9, OLI / TIRS	Collection 2, Tier 1	2021 to 2024	Surface Reflectance	LANDSAT/LC09/ C02/T1_L2	SR_B2: Blue [0.45 - 0.51 µm] SR_B3: Green [0.53 - 0.59 µm] SR_B4: Red [0.64 - 0.67 µm] SR_B5: Near Infrared [0.85 - 0.88 µm] SR_B6: Shortwave Infrared 1 [1.57 - 1.65 µm] SR_B7: Shortwave Infrared 2 [2.11 - 2.29 µm]
			Raw Images	LANDSAT/LC09/ C02/T1	B5: Near infrared [0.85 - 0.88 µm] B6: Shortwave infrared 1 [1.57 - 1.65 µm] B10: Thermal infrared 1 (resampled from 100m to 30m) [10.60 - 11.19 µm]

Table S2. Bands and indices applied for urban areas classification.

Type	Name	Description	Equations (if applicable)	Statistics
Bands	BLUE	Blue	SR_B2 (L8/9), SR_B1 (L5/7)	Median
	GREEN	Green	SR_B3 (L8/9), SR_B2 (L5/7)	Median
	RED	Red	SR_B4 (L8/9), SR_B3 (L5/7)	Median
	NIR	Near Infrared	SR_B5 (L8/9), SR_B4 (L5/7)	Median
	SWIR1	Shortwave Infrared 1	SR_B6 (L8/9), SR_B5 (L5/7)	Median
	SWIR2	Shortwave Infrared 2	SR_B7 (L8/9), SR_B7 (L5/7)	Median
Urban and Bare Soil indices	NDBI	Normalized Difference Built-up Index (Zha et al., 2003)	(SWIR1 - NIR) / (SWIR1 + NIR)	Median
	EBBI	Enhanced Built-up and Bareness Index (As-syakur et al., 2012)	((SWIR1 - NIR) / (SWIR1 + NIR + RED)^0.5)	Median, p25, p75, p75-p25
	UI	Urban Index (Kawamura et al., 1997)	(SWIR2 - NIR) / ((SWIR2 + NIR) +v1)	Median
	NDRI	Normalized difference roof index (Santos et al., 2022)	(RED - BLUE) / (RED + BLUE)	Median
	BAI	Bare soil area index (Santos et al., 2022)	(BLUE - NIR) / (BLUE + NIR)	Median
	BU	Built-up Index (ZHA et al., 2003)	NDBI - NDVI	Median
Vegetation indices	NDVI	Normalized Difference Vegetation Index (Rouse et al., 1974)	(NIR - RED) / (NIR + RED)	Median
	EVI	Enhanced Vegetation Index (Huete et al., 2002)	2.5 * ((NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1))	Median, p10, p90, p90-p10
	EVI2	Enhanced Vegetation Index modified	2.5 * ((NIR - RED) / (NIR + 2.4 * RED + 1))	Median, p10, p90, p90-p10
	SAVI*	Soil Adjusted Vegetation Index (Huete, 1988)	((NIR - RED) / (NIR + RED + 0.5)) * 1.5	Median
Water indices	MNDWI	Modified Normalized Difference Water Index (Xu, 2006)	(GREEN - SWIR1) / (GREEN + SWIR1)	Median
	NDWIm	NDWI Modified (McFEETERS, 1996)	(GREEN - NIR) / (GREEN + NIR)	Median
	AWEIsh*	Automated Water Extraction Index – Shadow (Feyisa et al., 2014)	BLUE + 2.5 * GREEN - 1.5 * (NIR + SWIR1) - 0.25 * SWIR2	Median
Bare Soil indices	BSI	Bare Soil index (Rikimaru et al., 2002)	((SWIR1 + RED)-(NIR + BLUE))/((SWIR1 + RED)+(NIR + BLUE))	Median
	NBR	Normalized Burn Ratio (Key and Benson, 2006)	(NIR - SWIR2) / (NIR + SWIR2)	Median
	NDMI	Normalized Difference Moisture Index, also referred as MNWI or NDUI (Gao, 1996)	(NIR - SWIR1) / (NIR + SWIR1)	Median
Spectral Mixture Analysis (SMA) calculated from bands ['BLUE', 'GREEN',	GV	Green Vegetation	Endmembers [0.0119, 0.0475, 0.0169, 0.6250, 0.2399, 0.0675] respective to each band	Median
	NPV	Non-Photosynthetic Vegetation	Endmembers [0.1514, 0.1597, 0.1421, 0.3053, 0.7707, 0.1975] respective to each band	Median

Type	Name	Description	Equations (if applicable)	Statistics
'RED', 'NIR', 'SWIR1', 'SWIR2'] and derived indices (Souza et al., 2005)	SOIL	Bare Soil	Endmembers [0.1799, 0.2479, 0.3158, 0.5437, 0.7707, 0.6646] respective to each band	Median
	CLOUD	Cloud	Endmembers [0.4031, 0.8714, 0.7900, 0.8989, 0.7002, 0.6607] respective to each band	Median
	GVS	Shadowed GV	GV / (GV + NPV + SOIL)	Median
	SHADE	Shade	abs((GV + NPV + SOIL) - 100)	Median
	NDFI	Normalized Difference Fraction Index (mixing components)	(GV - (NPV + SOIL + CLOUD)) / (GV + NPV + SOIL + CLOUD)	Median
Spectral Mixture Analysis (SMA) calculated from bands ['BLUE', 'GREEN', 'RED', 'NIR', 'SWIR1', 'SWIR2'] with Global Endmemembers Components (Small and Milesi, 2013)	SUBS	Substrate (Soil + Built-up)	Endmembers [0.178,0.337,0.458,0.559,0.683,0.645] respective to each band	Median
	VEG	Vegetation	Endmembers [0.030,0.060,0.031,0.669,0.240,0.096] respective to each band	Median
	DARK	Water + Shade	Endmembers [0.019,0.010,0.005,0.007,0.003,0.002] respective to each band	Median

* Additional indices introduced in Collection 10; all other indices and bands were previously used in Collection 9.

Table S3. Standard mosaic used for sample's quantity analysis.

Type	Band, index name
Landsat bands	BLUE
	GREEN
	RED
	NIR
	SWIR1
	SWIR2
Vegetation indices	NDVI
	EVI
	EVI2
	SAVI
Water	MNDWI
	NDWIm
	AWEIsh
Urban areas	NDBI
	UI
	BSI
	NDRI
	BAI
	EBBI

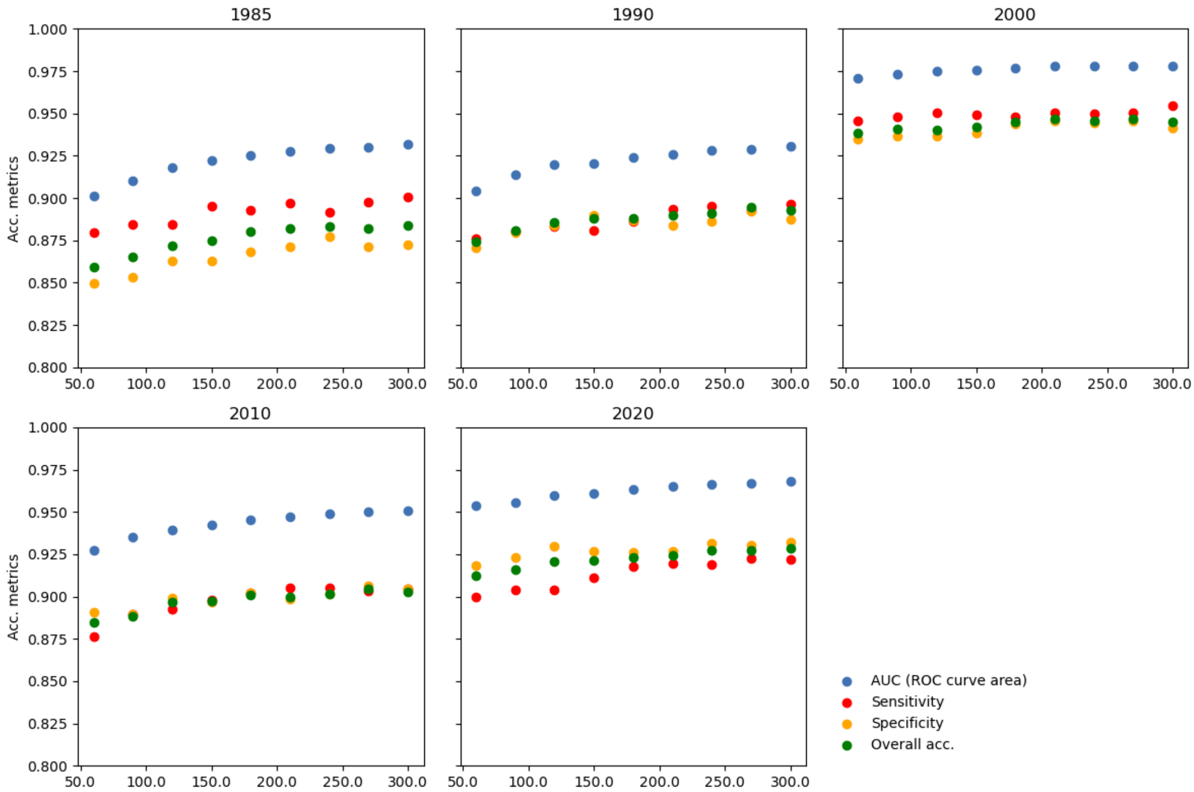


Figure S1. Samples quantity analysis summary. The graphs show sample quantities (x-axis) and selected metrics (y-axis) for each studied year. The results refer to the set of grids analyzed for Brazilian capitals.

Table S4. Mosaic compositions evaluated (mosaic refinement).

Set	Type	Indices list
Bands	Landsat bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indices 1	Vegetation	NDVI, EVI, EVI2
	Water	MNDWI, NDWIm, AWEIsh
	Soil, urban	NDBI, UI, BSI,
Indices 2	Soil, urban	EBBI
Mix 1	Spectral mixture 1	GV, NPV, SOIL, CLOUD, GVS, SHADE
Mix 2	Spectral mixture 2	SUBS, VEG, DARK
Indices Col. 9	All indices used in MapBiomas Collection 9	All indices used in MapBiomas Collection 9 added with AWEIsh

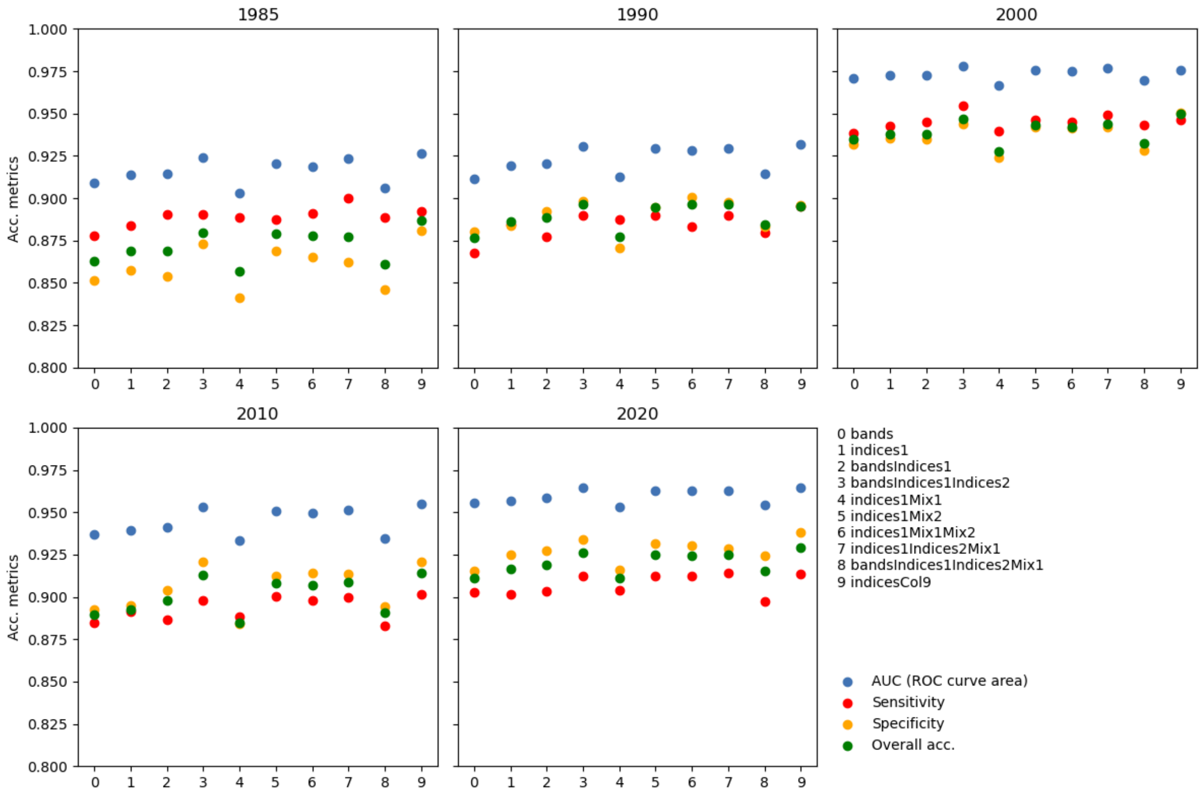


Figure S2**.** Mosaic evaluation. The graphs show that different mosaics performed similarly, mainly for the cases 5, 6, 7, and 9. The mosaic 9 was selected for consistency with previous MapBiomas collections.