**Urban Areas mapping**

**Algorithm Theoretical Basis Document (ATBD)**

**Collection 10.0 - ATBD Version 1**


## Team authors

| | |
|---|---|
| Breno Malheiros de Melo | USP/EESC |
| Edimilson Rodrigues dos Santos Junior | USP/EESC |
| Eduardo Felix Justiniano | USP/FFLCH |
| Julia Cansado | USP/FAU |
| Julio Cesar Pedrassoli | UFBA/POLI |
| Mayumi Hirye | USP/FAU |
| Talita Micheleti | USP/FAU |

## Collaborators

| | |
|---|---|
| Fabio Mariz Gonçalves | USP/FAU |
| João Meyer | USP/FAU |
| Marcelo Montaño | USP/EESC |
| Marcos Roberto Martines | UFSCAR |


São Paulo, 2025

## Summary

**Summary** 1
**1. Overview** 2
**2. Landsat image mosaics** 3
**3. Classification based on Random Forest algorithm** 4
3.1. Spatial scope 4
3.2. Samples collection 5
3.3. Classification algorithm 8
3.3.1. Samples selection 9
3.3.2. Feature Space 9
3.3.3. Random Forest training and classification 11
3.3.4. Temporal smoothing 11
3.3.5. Urban areas binary classification 12
**4. Post-classification procedures** 13
4.1. Gap fill 13
4.2. Temporal filter 13
4.3. Spatial filter 14
**5. Comparison with previous collections** 15
**6. Validation Strategies** 17
6.1. Accuracy Analysis 17
6.2. Comparison with reference maps 19
**7. References** 20
**Supplemental material** 23

## 1. Overview

This document presents the methodology developed to map urban areas across the Brazilian territory from 1985 to 2024, as part of Collection 10 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.

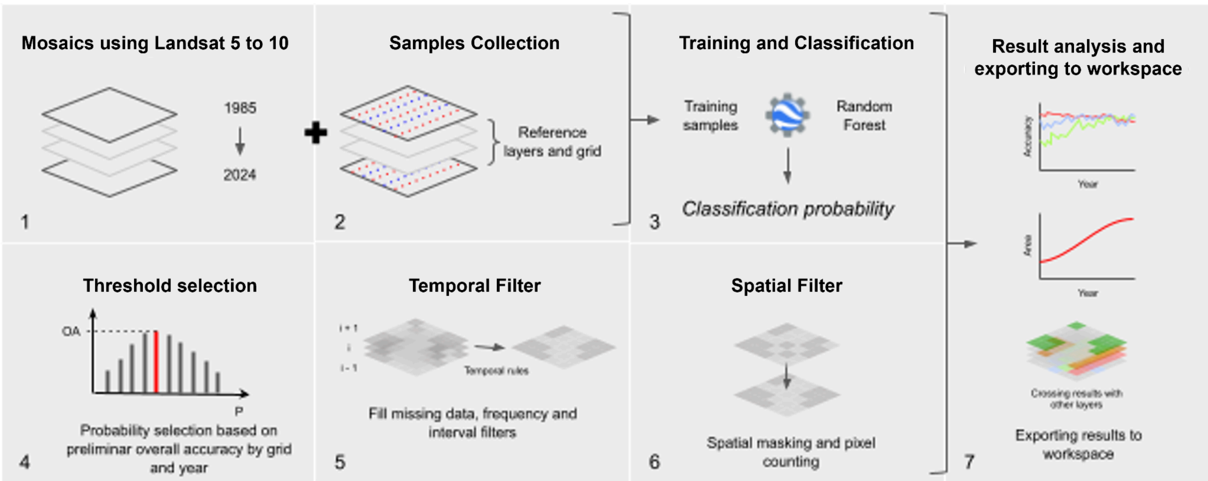



**Figure 1.** Basic scheme of urban areas classification.

Over successive collections, the urban classification method has been continuously refined through conceptual and methodological improvements. For Collection 10, updates include procedures for training sample selection, and probability thresholding for urban classification. Additional enhancements were made to improve the temporal consistency of the results and reduce classification noise, including the use of a temporal smoothing, probability threshold optimization.

Post-processing steps, including temporal and spatial filtering, were revised in this collection. The main changes compared to previous versions include (i) the reordering of the filter application sequence, with the temporal filter applied prior to the spatial filter; (ii) a simplified and less aggressive temporal filter design to better preserve legitimate temporal dynamics; and (iii) an updated spatial filtering approach, enabling control over the size of spatial artifacts, such as isolated misclassified pixels and small internal gaps while incorporating new auxiliary datasets to enhance spatial consistency.

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 supplementary table **ST1**). 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 supplementary **ST2**).

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. 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 2024) 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 (see supplementary **ST2**).

5. Apply appropriate reducers to each band and index (see supplementary **ST2**).

6. Calculate differences between reduced indices to capture intra-annual variability (see supplementary **ST2**).

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 Section “Classification algorithm”).

---

[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 .

## 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 2024.

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

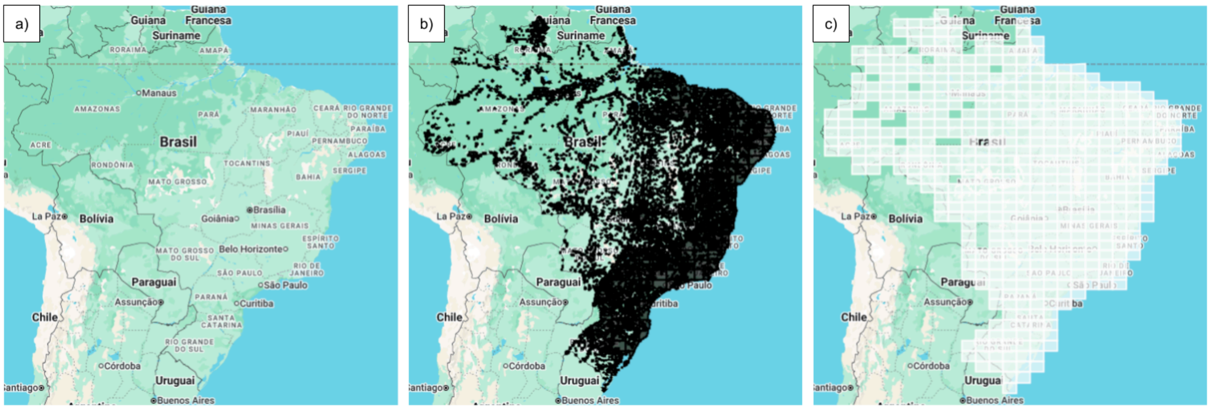



**Figure 2.** Search area and regular sheets (processing tiles).

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*

Training samples were obtained from the OpenStreetMap database (OSM, 2021), combined with nightlight imagery from NOAA, land cover and land use maps from the Third National Inventory (MCTI, 2015), and built-up area maps from the Global Human Settlement Layer (GHSL), provided by the Joint Research Centre (JRC) (Corbane *et al.*, 2018).

First, a preliminary urban mask was generated based on polylines from OpenStreetMap, representing roads, streets, sidewalks, and unclassified routes contributed by users. Pathways located within urban patches or specific categories (such as residential, service, path, and living street) were selected. To refine this initial mask, pathways outside urban areas were removed using nightlight imagery (NOAA) (**Figure 3**). For selected years, additional filtering was applied using GHSL built-up maps for 1985 and the urban area mappings from the Third National Inventory (MCTI, 2015) for 1994, 2002, and 2010. Each selected pathway was then buffered by approximately 100 meters to define urban candidate areas.

Next, an exploratory classification was conducted using normalized difference indices for vegetation (NDVI) and water (NDWI) to mask out vegetated and aquatic regions (**Figure 4**).

The final urban mask was derived from the intersection of the filtered OpenStreetMap-based mask and the results from the exploratory classification (**Figure 5** and **Figure 6**). The non-urban mask was defined as the symmetrical difference of this final urban mask.

Random points were then generated within the search area of each of the 522 tiles. These points were labeled as urban or non-urban using PostGIS, based on the final masks (**Figure 7**), resulting in a labeled dataset of training samples for the years 1985, 1994, 2002, 2010, and 2018.

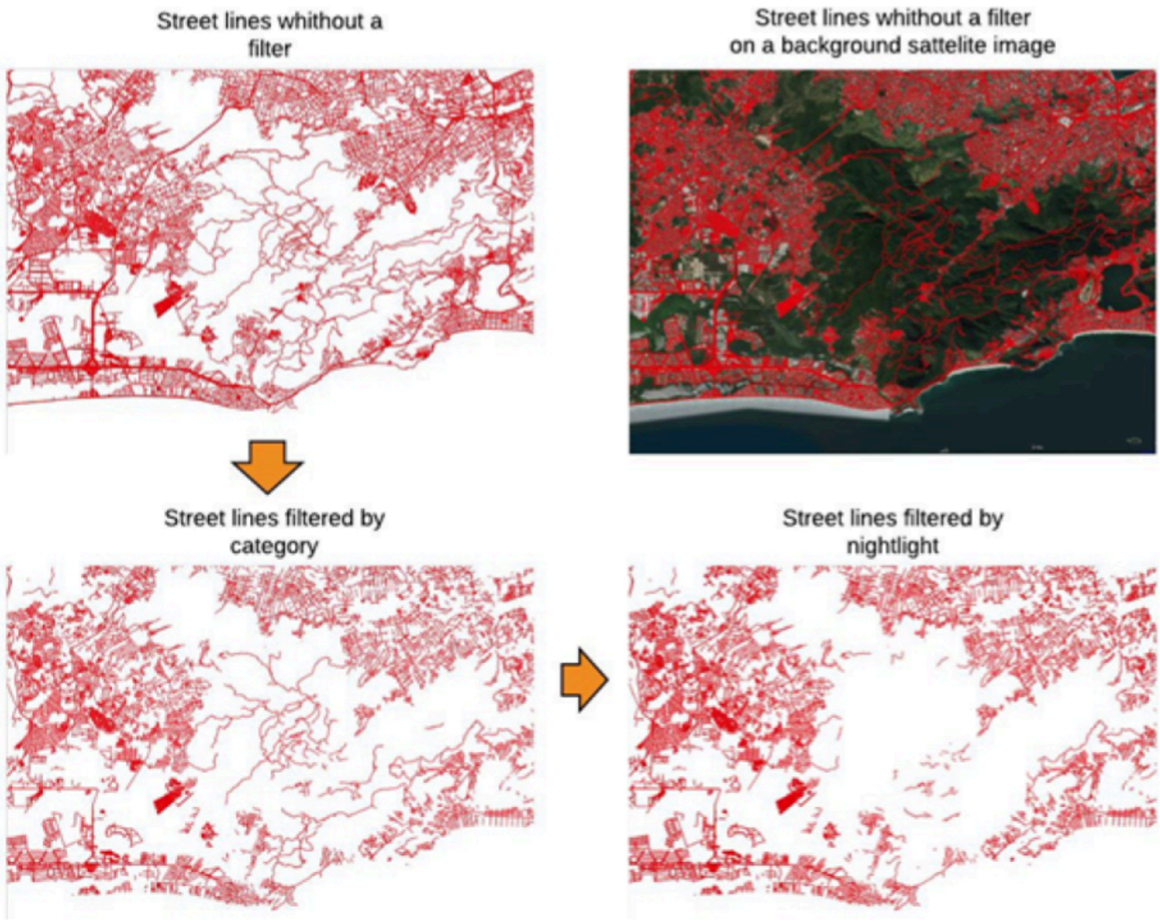



**Figure 3.** Example of filters used on the vector layer of OpenstreetMap in Rio de Janeiro - RJ Brazil.

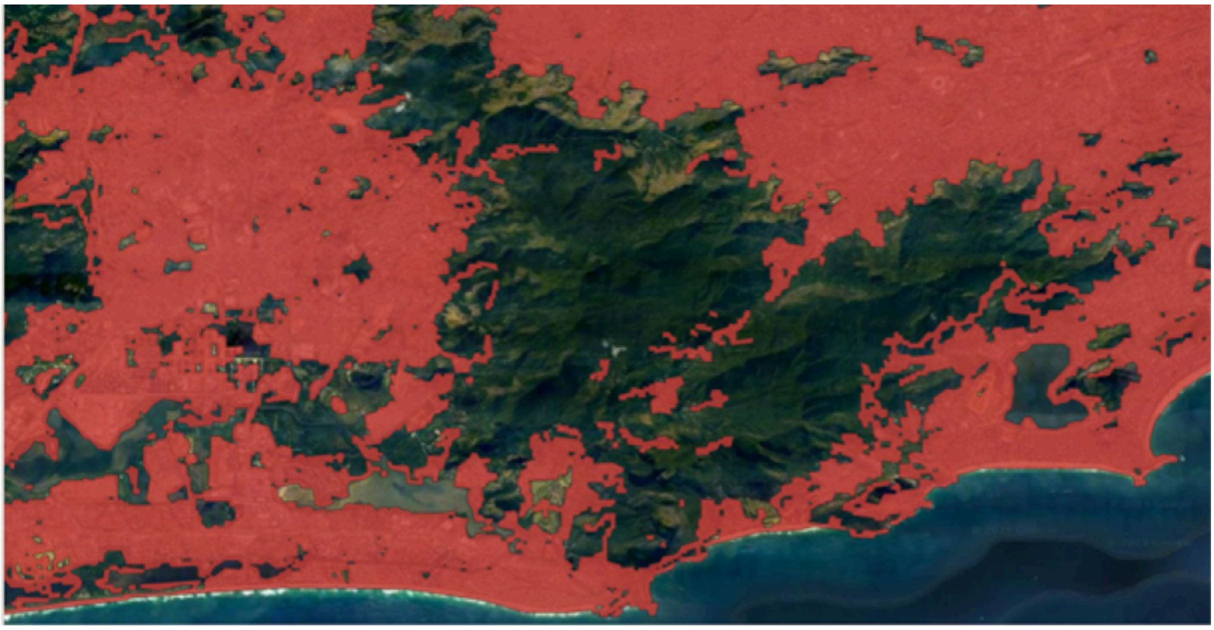


**Figure 4.** Exploratory classification results for Rio de Janeiro - RJ, Brazil.

**Figure 5.** Final urban mask for  Rio de Janeiro - RJ, Brazil.

**Figure 6.** Final non urban mask (orange color) for Rio de Janeiro - RJ, Brazil.

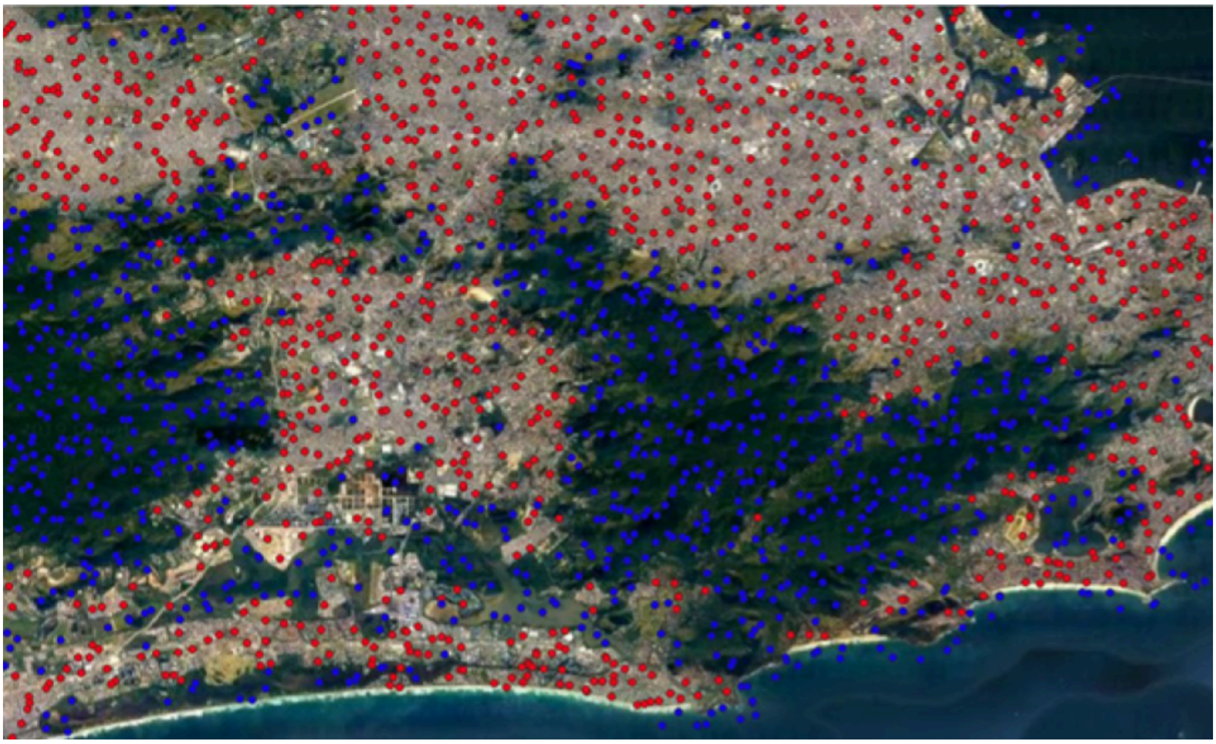


**Figure 7.** Random points divided by urban areas (red) and non-urban areas (blue).

This method enabled the automatic generation of a large and comprehensive sample dataset, totaling 891,427 non-urban and 532,520 urban samples for 1985; 883,101 non-urban and 453,857 urban samples for 1994; 977,644 non-urban and 546,407 urban samples for 2010; and 971,507 non-urban and 614,208 urban samples for 2018.

### 3.3. *Classification algorithm*

The Random Forest algorithm implemented in Google Earth Engine (smileRandomForest) was applied to map urban areas in MapBiomas Collection 10.0 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 probability, resulting in a classification image assigning to each pixel its probability of being urban.

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 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 section Samples selection) and mosaic compositions (see section Feature Space) were analyzed through a Random Forest classification (see 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 2024, annually, see section Temporal smoothing), and a cutoff threshold was estimated per grid to produce a binary classification of urban areas (see 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 supplementary table **ST3**). 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 supplementary figure **SF. 1**), 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 supplementary table **ST4** 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 **ST2** and **ST4**).
- Mix 2 - Additional spectral mixture bands (see table **ST2** and **ST4**).
- Bands used in MapBiomas Collection 9 - complete mosaic as defined in supplementary table **ST1** 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 supplementary figure **SF. 2**). 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 10.0 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 8** ).

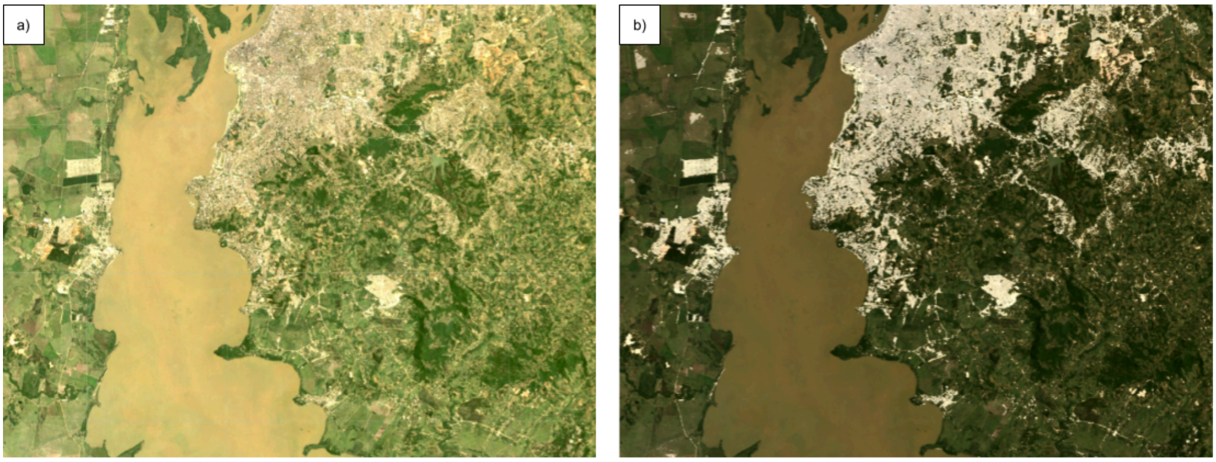



**Figure 8.** Classification example using the selected mosaic. Porto Alegre - RS, Brazil.

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 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 9**). For evaluation purposes, the final validation was conducted using MapBiomas validation samples (see MapBiomas website).

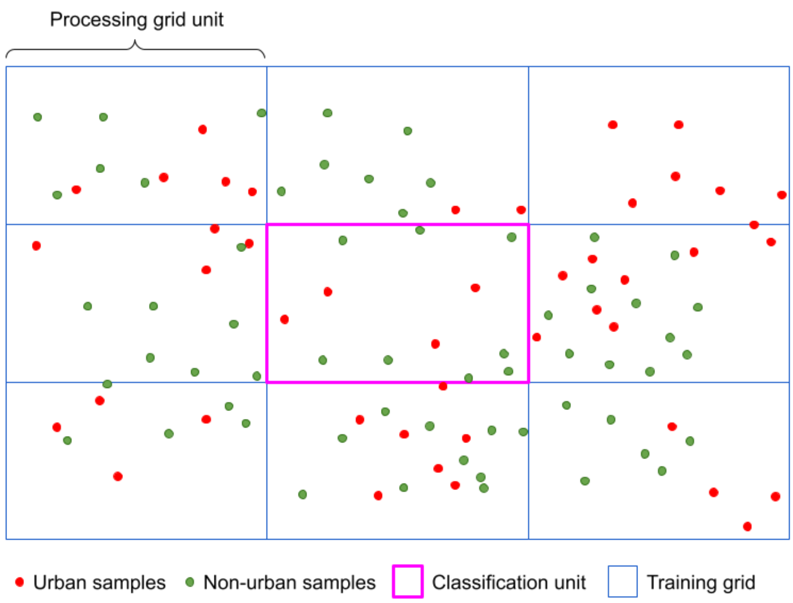



**Figure 9.** Processing grid for classification scheme.

### *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 10**.

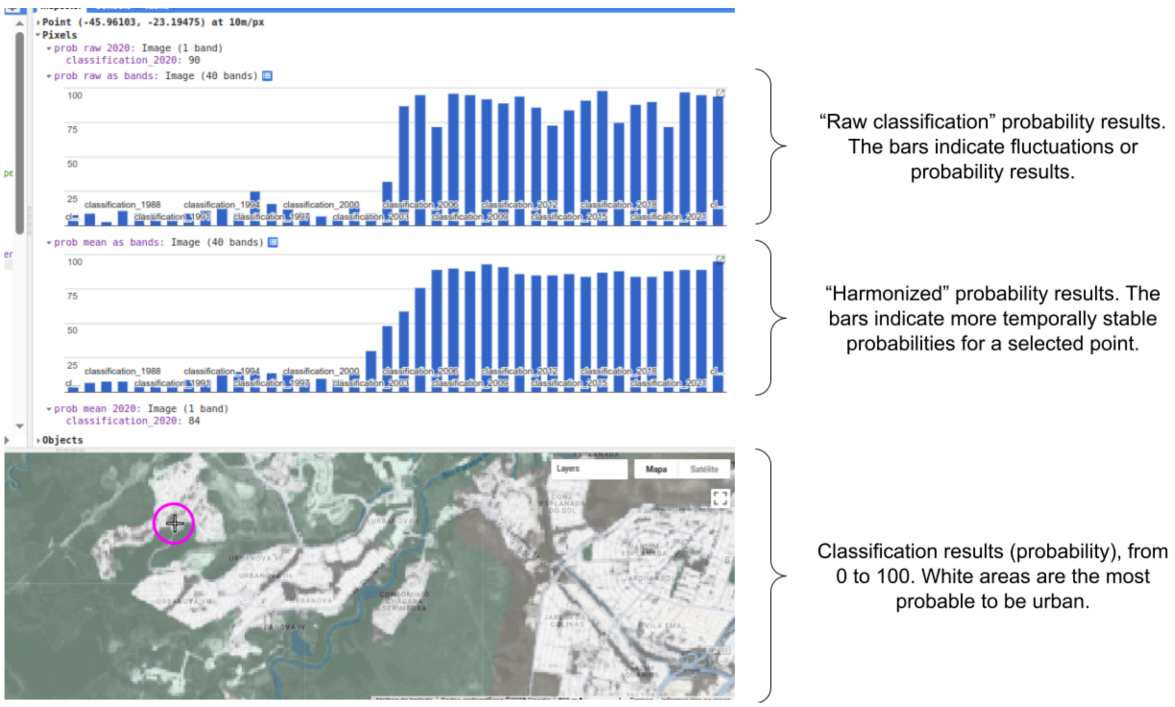



**Figure 10.** Example of temporal harmonization impact. São José dos Campos - SP, Brazil.

### 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 distributions 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. The only exception was the case of Corumbá municipality, an isolated urban area within the Pantanal biome where the cutoff probability value was adopted as 50%. 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 11**).

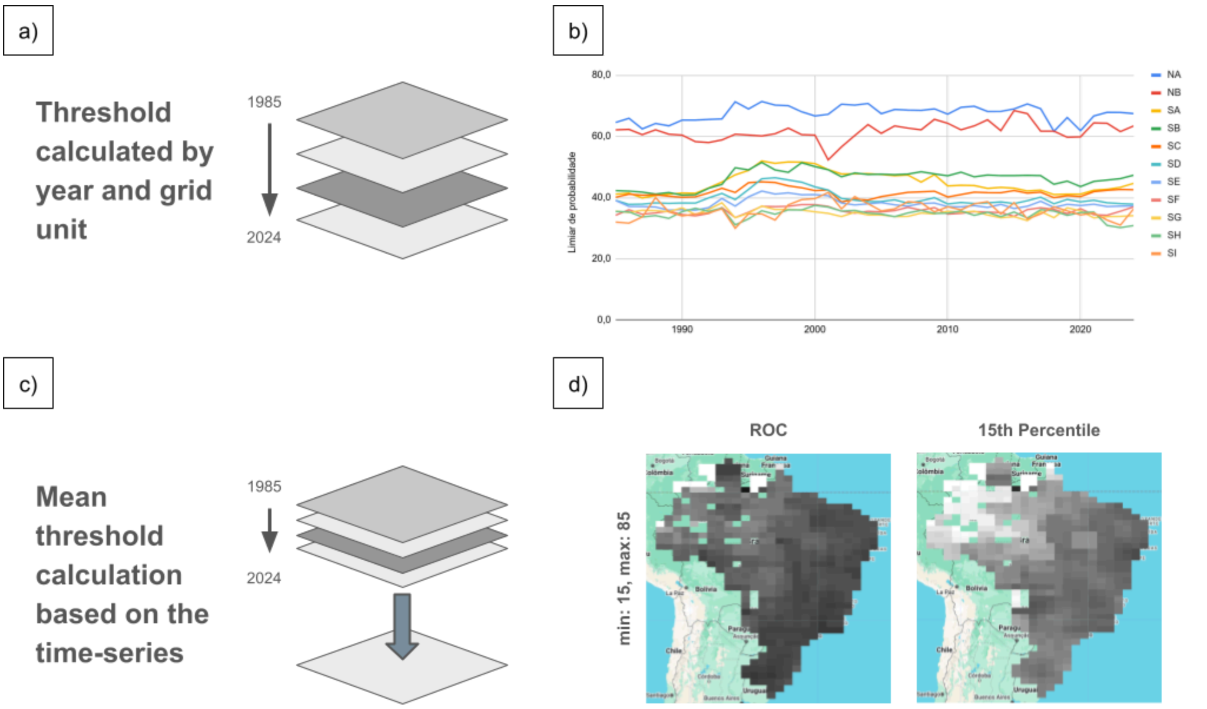



**Figure 11.** Summary of binary classification procedures.

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; Liu; Clarke, 2003; Lu; 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 was applied to correct no-data pixels—originating from cloud and cloud shadow masking—by replacing them with the most frequent (mode) class value from neighboring years within a 3-year temporal window. Assuming classification quality improves toward recent years, this filtering is applied backward through time. For boundary years (1985 and 2024), the window was adapted to include all available years: for 1985, the years 1986 and 1987 were used; for 2024, the years 2022 and 2023 were used to fill no-data values and maintain temporal consistency.

### 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 pixels with values differing from their neighbors within a 5-year window (function TempFilter_wMask_5yearsunique): Unique pixel values—urban or non-urban—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 years (1985, 1986, 2023, and 2024), the 5-year window is adjusted to include all available neighboring years.

(ii) Identify breakpoints (function getBreakpoints): 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 2024. 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).

(iii) Select the first breakpoint and accumulate forward (function accumulateForward): 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.

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— clusters of non-urban pixels typically corresponding to vegetation patches or vacant urban spaces surrounded by urban pixels— a threshold of 280 pixels (approximately 25 hectares) is 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—is 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 2024, mapping urban areas across the Brazilian territory. This dataset was integrated with other thematic maps to compose MapBiomas Collection 10 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 12**. For Collection 10, the total mapped urbanized area starts at approximately 1.80 million hectares in 1985 and increases over time, reaching around 4.55 million hectares by 2024.

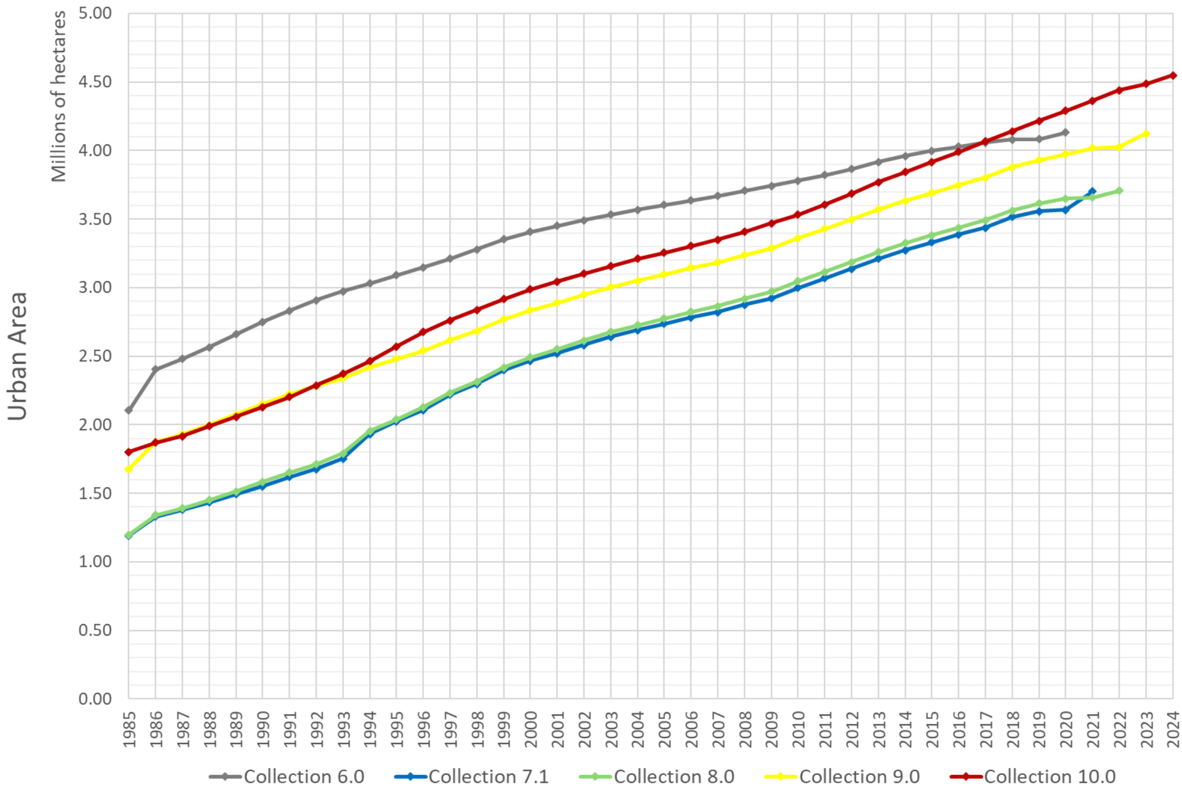



**Figure 12.** Total urban area (in millions of hectares) mapped by year in MapBiomas Collections 6, 7.1, 8, 9, and 10

**Figure 12** represents differences between collections, reflecting the evolution of classification methodologies over time. Collection 6 stands out as the most distinct, reporting consistently higher urbanized areas throughout the time series, especially in the early years, reflecting earlier classification methods that tended to overestimate urban areas compared to subsequent collections. Collections 7.1 and 8 display very similar trajectories throughout the time series, especially in the early years, which aligns with their methodological similarity. Collections 9 and 10 also follow closely aligned trends in the initial period (1985–1994), but begin to diverge after 1994, with Collection 10 detecting systematically larger urbanized areas, particularly in more recent years. It is worth noting that the sharp increase from 1985 to 1986 observed in all previous collections was corrected in Collection 10, which shows a smoother and more gradual transition between these years.

A qualitative comparison between Collections 9 and 10 highlights clear improvements in spatial detail. As shown in Figure 10, Collection 10 presents a more refined delineation of urban areas, particularly along the urban-rural interface. Compared to Collection 9, Collection 10 more accurately captures the built-up footprint, reducing commission errors in adjacent vegetated areas (**Figure 13** [A] and [B]). Notably, it also better identifies linear urbanization patterns, such as small settlements along roadways, which appear more fragmented or underrepresented in Collection 9 (**Figure 13** [C] and [D]).

[A]

[B]

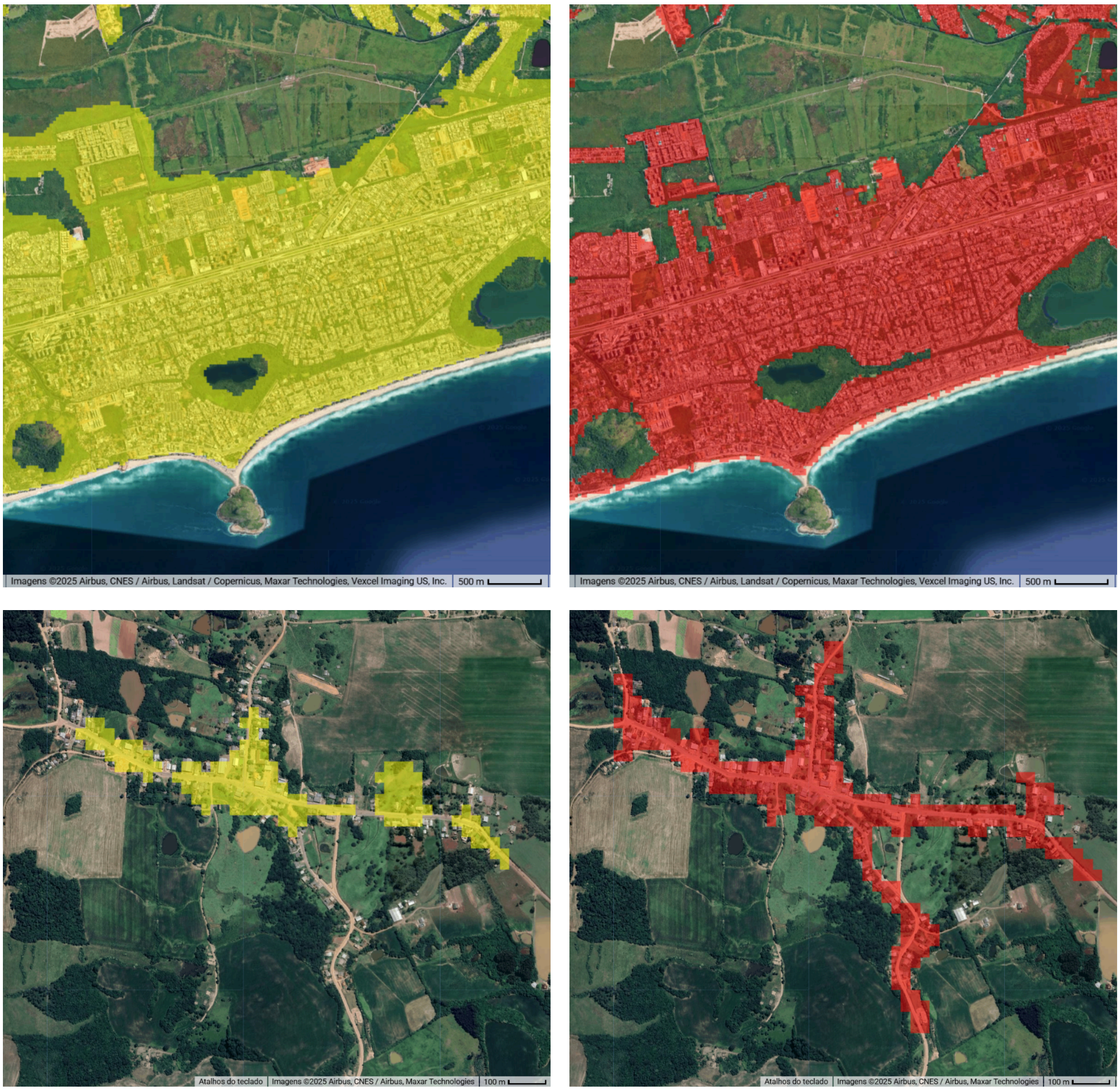



**Figure 13.** Comparison of urban area mapping between Collections 9 and 10 in two locations: [A] Rio de Janeiro (RJ) and [B] Agudo (RS). In each location, yellow represents the urban area mapped in Collection 9, and red represents Collection 10.

## 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; 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).

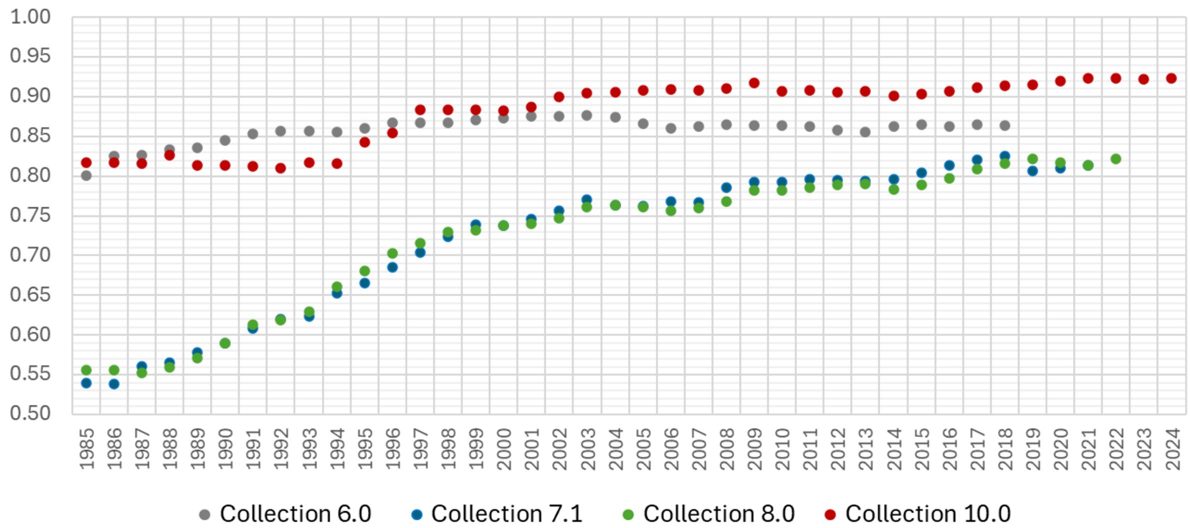



**Figure 14.** Producer's accuracy

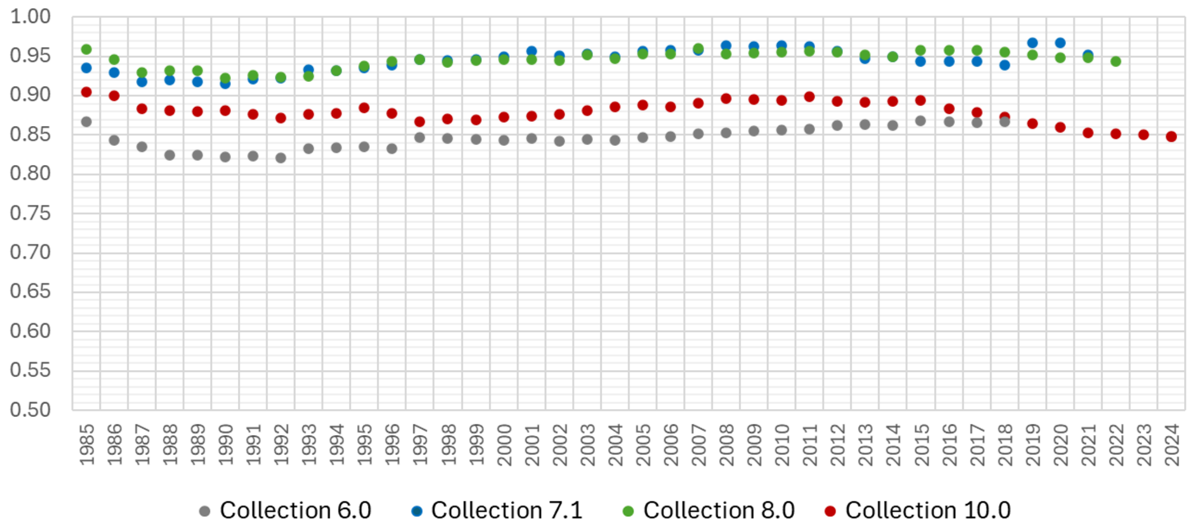



**Figure 15.** User's accuracy

[2] https://brasil.mapbiomas.org/en/estatistica-de-acuracia/colecao-10/

In terms of Producer's Accuracy, Collection 10 consistently shows the highest values across all years, especially from 2000 onwards, maintaining levels above 90%. This reflects a significant improvement in minimizing omission errors, meaning urban areas are more completely captured compared to earlier collections. For User's Accuracy, Collection 10 also shows consistently high performance but with a slight decreasing trend in recent years. It maintains values generally above 85%, which indicates good reliability in minimizing commission errors (i.e., reducing the inclusion of non-urban areas as urban). However, the highest User's Accuracy in the early years (1985–2000) is observed in Collections 7.1 and 8.0, likely due to their more conservative mapping approach. From 2000 onwards, Collection 10 remains stable, balancing a higher Producer's Accuracy with good User's Accuracy.

Overall, Collection 10 outperforms previous collections in terms of Producer's Accuracy while maintaining competitive User's Accuracy, demonstrating the effects of methodological improvements aimed at reducing omission errors without excessively increasing commission errors.

### *6.2. Comparison with reference maps*

MapBiomas Collection 10.0 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 1**) indicates that the urbanized area mapped in 2015 by MapBiomas Collection 10.0 totals 3,916,797 hectares, surpassing the 3,421,975 hectares mapped by the World Settlement Footprint (WSF). The area mapped exclusively by MapBiomas amounts to 1,051,152 hectares, while WSF uniquely maps 556,329 hectares. Despite these differences, both datasets share a concordant area of 2,865,645 hectares. Overall, the overlap between MapBiomas and WSF reaches 83.7% across the Brazilian territory. When disaggregated by biome, the Pantanal exhibits the lowest overlap (73.5%), while all other biomes show overlaps exceeding 80%.

**Table 1.** Comparison of Urban Area Mapping between MapBiomas Collection 10 and the World Settlement Footprint -WSF (Marconcini et al., 2020) for the Year 2015.

| **Year: 2015** Biome | **Total Area Mapped by Col. 10** (in ha) | **Total Area Mapped by WSF** (in ha) | **Area Only Mapped by Col. 10** (in ha) | **Area Only Mapped by WSF** (in ha) | **Overlapping Area** (in ha) | **Overlap Relative to WSF** (in %) |
|---|---|---|---|---|---|---|
| Amazon | 377,062 | 314,743 | 116,752 | 54,433 | 260,310 | 82.7% |
| Caatinga | 521,692 | 337,433 | 243,478 | 59,219 | 278,214 | 82.5% |
| Cerrado | 882,059 | 737,625 | 262,030 | 117,596 | 620,029 | 84.1% |
| Atlantic Forest | 1,997,877 | 1,898,665 | 404,553 | 305,341 | 1,593,324 | 83.9% |

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

| Year: 2015<br><br>Biome | Total Area Mapped by Col. 10 (in ha) | Total Area Mapped by WSF (in ha) | Area Only Mapped by Col. 10 (in ha) | Area Only Mapped by WSF (in ha) | Overlapping Area (in ha) | Overlap Relative to WSF (in %) |
|---|---|---|---|---|---|---|
| Pampa | 133,893 | 128,935 | 23,485 | 18,527 | 110,408 | 85.6% |
| Pantanal | 4,213 | 4,572 | 854 | 1,214 | 3,359 | 73.5% |
| Brazil | 3,916,797 | 3,421,975 | 1,051,152 | 556,329 | 2,865,645 | 83.7% |

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 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 data points to an underestimation of MapBiomas Collection 10.0 urban area (**Table 2**). For the year 2019, the urbanized area mapped by Collection 10 totals 4,215,208 hectares, while the Urbanized Areas dataset identifies a larger extent of 5,321,060 hectares. The area mapped exclusively by Collection 10 amounts to 475,366 hectares, whereas Urbanized Areas exclusively accounts for 1,581,219 hectares. The overlapping area between the two datasets reaches 3,739,841 hectares. Considering the entire Brazilian territory, this represents a 70.3% overlap relative to the total mapped by Urbanized Areas. When analyzed by biome, the overlap varies, with the Cerrado (75.4%) and Atlantic Forest (70.9%) exhibiting the highest levels of agreement, while the Amazon (65.3%) and Caatinga (64.1%) show lower concordance. The Pantanal displays a moderate overlap of 69.4%, and the Pampa reaches 69.7%.

**Table 2.** Comparison of Urban Area Mapping between MapBiomas Collection 10 and the Urbanized Areas (IBGE, 2022) for the Year 2019.

| Year: 2019<br><br>Biome | Total Area Mapped by Collection 10 (in ha) | Total Area Mapped by Urbanized Areas (in ha) | Area Only Mapped by Collection 10 (in ha) | Area Only Mapped by Urbanized Areas (in ha) | Overlapping Area (in ha) | Overlap Relative to Urbanized Areas (in %) |
|---|---|---|---|---|---|---|
| Amazon | 402,349 | 567,544 | 31,828 | 197,023 | 370,521 | 65.3% |
| Caatinga | 594,173 | 716,514 | 134,666 | 257,007 | 459,507 | 64.1% |
| Cerrado | 950,399 | 1,101,385 | 119,443 | 270,428 | 830,956 | 75.4% |
| Atlantic Forest | 2,124,634 | 2,738,989 | 182,877 | 797,231 | 1,941,758 | 70.9% |
| Pampa | 139,229 | 190,538 | 6,355 | 57,664 | 132,874 | 69.7% |
| Pantanal | 4,423 | 6,091 | 198 | 1,866 | 4,225 | 69.4% |
| Brazil | 4,215,208 | 5,321,060 | 475,366 | 1,581,219 | 3,739,841 | 70.3% |

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

## 7. References


AS-SYAKUR, Abd Rahman *et al.* Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. **Remote Sensing**, v. 4, n. 10, p. 2957–2970, 2012.

BRADLEY, Andrew P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. **Pattern Recognition**, v. 30, n. 7, p. 1145–1159, 1997.

BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

CORBANE, Christina *et al.* GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014). 2018. Available at: http://data.europa.eu/89h/jrc-ghsl-10007.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, ROC Analysis in Pattern Recognition, p. 861–874, 2006.

FEYISA, Gudina L. *et al.* Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. **Remote Sensing of Environment**, v. 140, p. 23–35, 2014.

GAO, Bo-cai. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, v. 58, n. 3, p. 257–266, 1996.

HEROLD, Martin; LIU, XiaoHang; CLARKE, Keith C. Spatial Metrics and Image Texture for Mapping Urban Land Use. **Photogrammetric Engineering & Remote Sensing**, v. 69, n. 9, p. 991–1001, 2003.

HUETE, A.R. A soil-adjusted vegetation index (SAVI). **Remote Sensing of Environment**, v. 25, n. 3, p. 295–309, 1988.

HUETE, A *et al.* Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1, The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring, p. 195–213, 2002.

IBGE. Censo Demográfico 2010 – Aglomerados Subnormais. Rio de Janeiro, RJ: IBGE, 2020.

IBGE, Coordenação de Meio Ambiente (org.). **Áreas urbanizadas do Brasil: 2019**. Rio de Janeiro, RJ: IBGE, 2022.

IBGE. **Malha de Setores Censitários**. Brasil: 2020. Malha censitária.

JUSTINIANO, Eduardo Felix *et al.* Proposal for an index of roads and structures for the mapping of non-vegetated urban surfaces using OSM and Sentinel-2 data. **International Journal of Applied Earth Observation and Geoinformation**, v. 109, p. 102791, 2022.

KAWAMURA, Makoto; JAYAMANNA, Sanath; TSUJIKO, Yuji. Quantitative Evaluation of Urbanization in Developing Countries Using SaTELlite Data. **Doboku Gakkai Ronbunshu**, v. 1997, n. 580, p. 45–54, 1997.

KEY, C. H.; BENSON, N. C. Landscape Assessment: Ground Measure of Severity, the

Composite Burn Index; and Remote Sensing of Severity, the Normalized Burn Ratio. *In*: LUTES, D. C. *et al.* (org.). **FIREMON: Fire Effects Monitoring and Inventory System**. Ogden, UT: USDA Forest Service, Rocky Mountain Research Station, 2006. (Gen. Tech. Rep.).

LU, D.; WENG, Q. A survey of image classification methods and techniques for improving classification performance. **International Journal of Remote Sensing**, v. 28, n. 5, p. 823–870, 2007.

MARCONCINI, Mattia *et al.* Outlining where humans live, the World Settlement Footprint 2015. **Scientific Data**, v. 7, n. 1, p. 242, 2020.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. **International Journal of Remote Sensing**, v. 17, n. 7, p. 1425–1432, 1996.

MCTIC-Ministério da Ciência, Tecnologia, Inovações e Comunicações. **III National Inventory of Greenhouse Gas Emissions (LULUCF sector)**. 2015. Available at: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/sirene/publicacoes/relatorios-de-referencia-setorial/pdf/invetario3/rr_lulucf_mudanca_de_uso_e_floresta.pdf .

OSM. **OpenStreetMap (Standard)**, 2021.

RIKIMARU, A; ROY, P S; MIYATAKE, S. Tropical forest cover density mapping. **International Society for Tropical Ecology**, v. 43, n. 1, p. 39–47, 2002.

ROUSE, J. W. *et al.* Monitoring vegetation systems in the Great Plains with ERTS. **Anais [...]**. Texas: NASA NTRS, 1974. Available at: https://ntrs.nasa.gov/citations/19740022614.

SANTOS, Bruno Dias dos *et al.* Identifying Precarious Settlements and Urban Fabric Typologies Based on GEOBIA and Data Mining in Brazilian Amazon Cities. **Remote Sensing**, v. 14, n. 3, p. 704, 2022.

SIRKO, Wojciech *et al.* **Continental-Scale Building Detection from High Resolution Satellite Imagery**. [*S. l.*]: arXiv, 2021. Available at: http://arxiv.org/abs/2107.12283.

SMALL, Christopher; MILESI, Cristina. Multi-scale standardized spectral mixture models. **Remote Sensing of Environment**, v. 136, p. 442–454, 2013.

SOUZA, Carlos M.; ROBERTS, Dar A.; COCHRANE, Mark A. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. **Remote Sensing of Environment**, v. 98, n. 2, p. 329–343, 2005.

STEHMAN, Stephen V. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. **International Journal of Remote Sensing**, v. 35, n. 13, p. 4923–4939, 2014.

STEHMAN, Stephen V.; FOODY, Giles M. Key issues in rigorous accuracy assessment of land cover products. **Remote Sensing of Environment**, v. 231, p. 111199, 2019.

XU, Hanqiu. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. **International Journal of Remote Sensing**, v. 27, n. 14, p. 3025–3033, 2006.

ZHA, Y.; GAO, J.; NI, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. **International Journal of Remote Sensing**,  v. 24, n. 3, p. 583–594, 2003.

## Supplemental material

**ST. 1.** 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]<br>SR_B2: Green [0.52-0.60 µm]<br>SR_B3: Red [0.63-0.69 µm]<br>SR_B4: Near Infrared [0.77-0.90 µm]<br>SR_B5: Shortwave Infrared 1 [1.55-1.75 µm]<br>SR_B7: Shortwave Infrared 2 [2.08-2.35 µm] |
| | | | Raw Images | LANDSAT/LT05/C02/T1 | B4: Near infrared [0.76 - 0.90 µm]<br>B5: Shortwave infrared 1 [1.55 - 1.75 µm]<br>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]<br>SR_B2: Green [0.52-0.60 µm]<br>SR_B3: Red [0.63-0.69 µm]<br>SR_B4: Near Infrared [0.77-0.90 µm]<br>SR_B5: Shortwave Infrared 1 [1.55-1.75 µm]<br>SR_B7: Shortwave Infrared 2 [2.08-2.35 µm] |
| | | | Raw Images | LANDSAT/LE07/C02/T1 | B4: Near Infrared [0.77 - 0.90 µm]<br>B5: Shortwave Infrared 1 [1.55 - 1.75 µm]<br>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]<br>SR_B3: Green [0.53 - 0.59 µm]<br>SR_B4: Red [0.64 - 0.67 µm]<br>SR_B5: Near Infrared [0.85 - 0.88 µm]<br>SR_B6: Shortwave Infrared 1 [1.57 - 1.65 µm]<br>SR_B7: Shortwave Infrared 2 [2.11 - 2.29 µm] |
| | | | Raw Images | LANDSAT/LC08/C02/T1 | B5: Near infrared [0.85 - 0.88 µm]<br>B6: Shortwave infrared 1 [1.57 - 1.65 µm]<br>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]<br>SR_B3: Green [0.53 - 0.59 µm]<br>SR_B4: Red [0.64 - 0.67 µm]<br>SR_B5: Near Infrared [0.85 - 0.88 µm]<br>SR_B6: Shortwave Infrared 1 [1.57 - 1.65 µm]<br>SR_B7: Shortwave Infrared 2 [2.11 - 2.29 µm] |
| | | | Raw Images | LANDSAT/LC09/C02/T1 | B5: Near infrared [0.85 - 0.88 µm]<br>B6: Shortwave infrared 1 [1.57 - 1.65 µm]<br>B10: Thermal infrared 1 (resampled from 100m to 30m) [10.60 - 11.19 µm] |

**ST. 2.** 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; Gao; Ni, 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; Jayamanna; Tsujiko, 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; Roy; Miyatake, 2002) | ((SWIR1 + RED)-(NIR + BLUE))/((SWIR1 + RED)+(NIR + BLUE)) | Median |
| | NBR | Normalized Burn Ratio (Key; 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 | 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 |

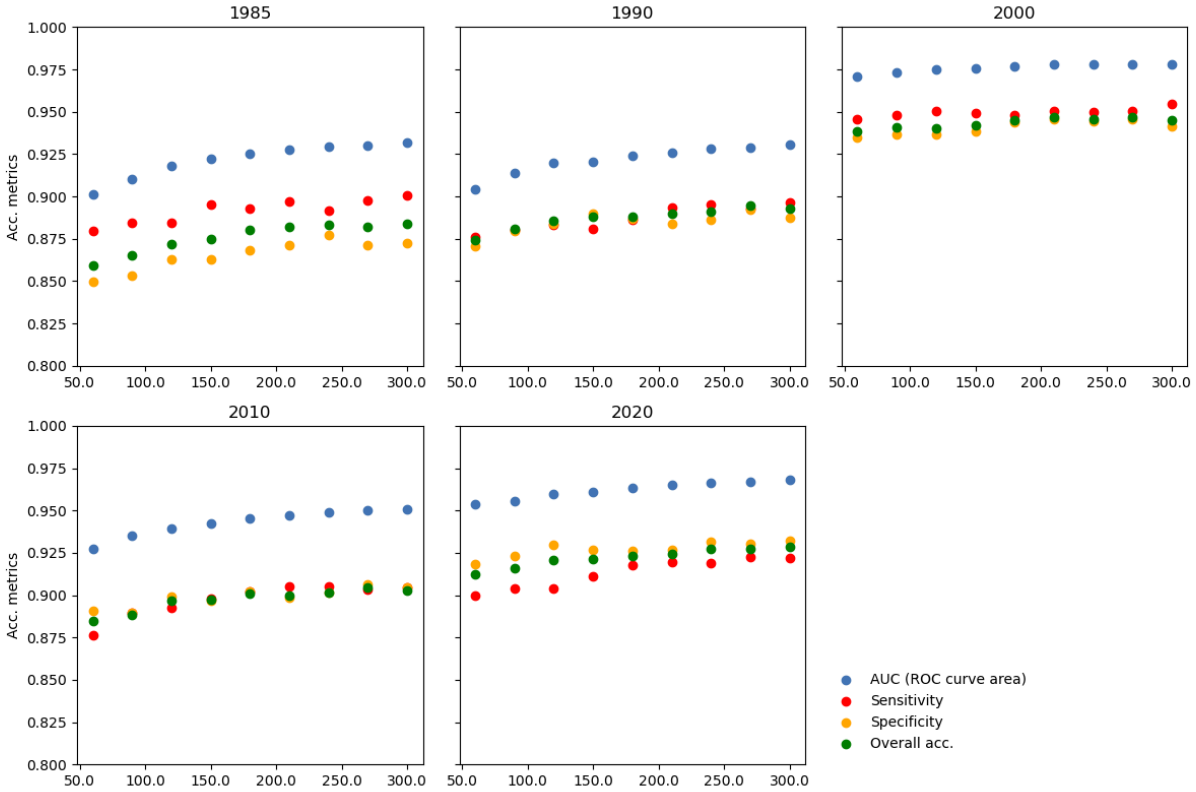
| Type | Name | Description | Equations (if applicable) | Statistics |
|---|---|---|---|---|
| ['BLUE', 'GREEN', 'RED', 'NIR', 'SWIR1', 'SWIR2'] and derived indices (Souza; Roberts; Cochrane, 2005) | SOIL | Bare Soil | Endmembers [0.1799, 0.2479, 0.3158, 0.5437, 0.7707, 0.6646] respective to each band | 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; 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.

**ST. 3.** 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 |

**SF. 1.** Samples quantity analysis summary.

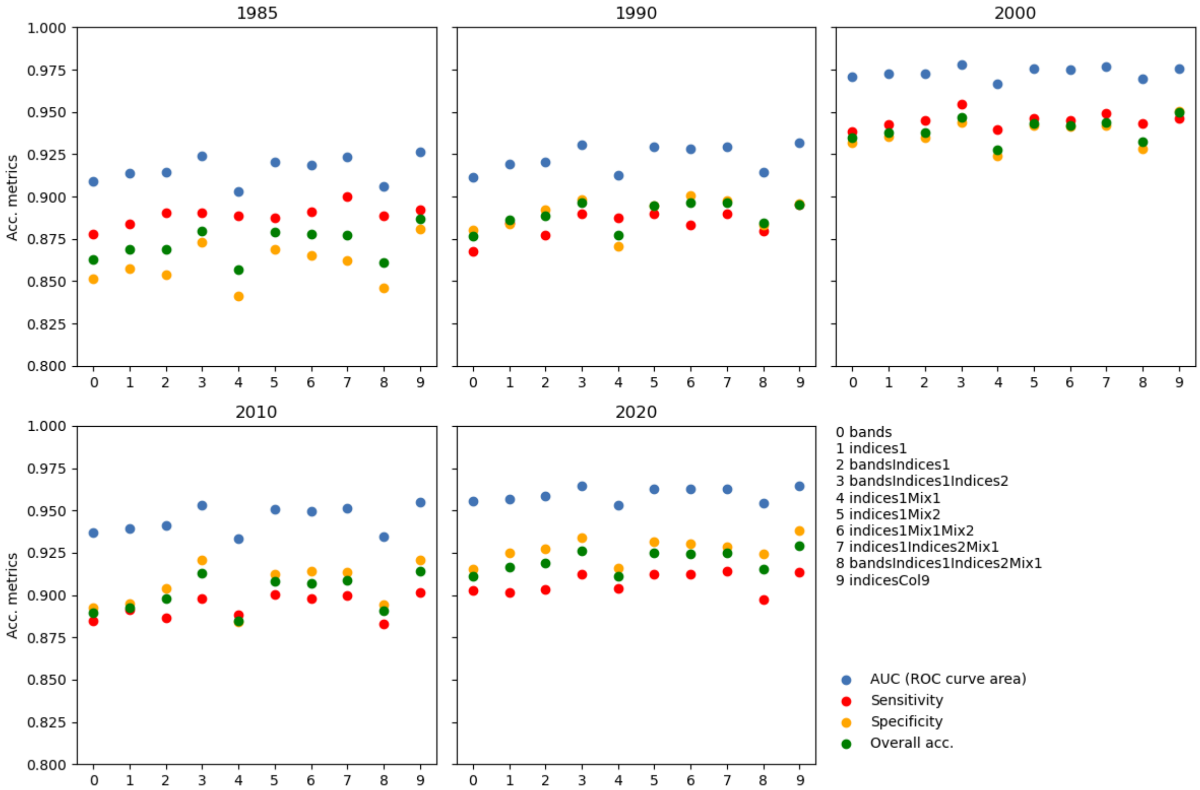



The graphs show sample quantities (x-axis) and selected metrics (y-axis) for each studied year. The results refer to the set of grids analyzed for Brazilian capitals.

**ST. 4.** Mosaic compositions evaluated (mosaic refinement).

| Set | Type | Indices list (defined in table ST2) |
|---|---|---|
| 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 |

**SF. 2.** 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.