

MAPBIOMAS
COLOMBIA

Appendix - Urban Infrastructure

Collection 4.0

Version 1

Technical lead
Karen Andrea Huertas

Responsible team
Valentina Castellanos

November 2026

Table of contents

1. Introduction **3**

2. Satellite image mosaics 4

2.1 Mosaic composition 4

2.2 Selection of classification variables 5

3. Classification **7**

3.1 Definition of the Urban Infrastructure class (ID: 24) 9

3.2 Classification algorithms 10

3.3 Reference data 10

4. Post-classification **12**

4.1 Merging of Collection 3.0 and Collection 4.0 years 12

4.2 Gap filling - Gap fill filter 12

4.3 Temporal filter 13

4.4 Adjusted frequency filter 14

4.5 Spatial filter 15

5. References **17**

List of tables

Table 1. Landsat data collections used in Collection 4.0. 4

Table 2. Feature space: variables for the detection of Urban Infrastructure surfaces with the associated reducer. 5

Table 3. Statistical reducers used in the image composition for Urban Infrastructure, Collection 4.0. 6

Table 4. Detailed definition of Urban Infrastructure, Collection 4.0. 8

Table 5. Colombian data used as global reference in the urban infrastructure classification process. 10

Table 6. Colombian data used as national reference in the urban infrastructure classification process. 11

List of figures

Figure 1. Study area for the detection of Urban Infrastructure cover in MapBiomas Colombia Collection 4.0. 3

Figure 2. Annual mosaic generation method, Collection 4.0 MapBiomas Colombia. 5

Figure 3. Methodological workflow for the Urban Infrastructure classification, Collection 4.0 8

Figure 4. Urban Infrastructure classification regions, Collection 4.0. 8

Figure 5. Urban Infrastructure detection, Collection 4.0: (A) Landsat mosaic, (B) urban surface classification ID: 24, (C) Google Earth high-resolution image (Barranquilla - Colombia). 10

Figure 6. Reference sources available for the MapBiomas Colombia study area. 11

Figure 7. Gap fill filter. 13

Figure 8. Temporal filter. 14

Figure 9. Frequency filter. 15

Figure 10. Spatial filter. 15

Figure 11. Multitemporal classification of Urban Infrastructure, Collection 4.0 - 1985 to 2025 (case Bogotá - Colombia). 16

1. Introduction

Collection 4.0 of the MapBiomas Colombia annual maps incorporates the fourth version of the urban infrastructure mapping as a cross-cutting binary layer covering the entire territory of Colombia over the period from 1985 to 2025 (Figure 1). This cover class consists of human settlement areas associated with large and small urban centers (towns) with built-environment infrastructure, such as road and railway networks and associated land, as well as other artificial areas such as hydrocarbon extraction facilities, hydroelectric plants, military bases, airports, port areas, and non-agricultural green areas such as recreational facilities in urban centers, urban lawns, road medians, and unconventional airstrips in rural areas. Peripheral areas undergoing a gradual process of urbanization for residential and/or industrial purposes are also considered.

The urban landscapes in the region are spatially organized into numerous hubs of infrastructure concentration distributed throughout the territory. Some urban areas have emerged mainly from rural population concentration and, over time, have become integrated into the network of cities at different hierarchical levels according to their territorial status. This mapping includes intermediate cities, twin cities, border cities, and gradually expanding rural centers. Intermediate cities predominate in the region and are defined as centers whose economy is based on the surrounding territory. These nodes have transnational characteristics that create proximity with their surroundings to diversify their socioeconomic mechanisms (UCLG, 2018).

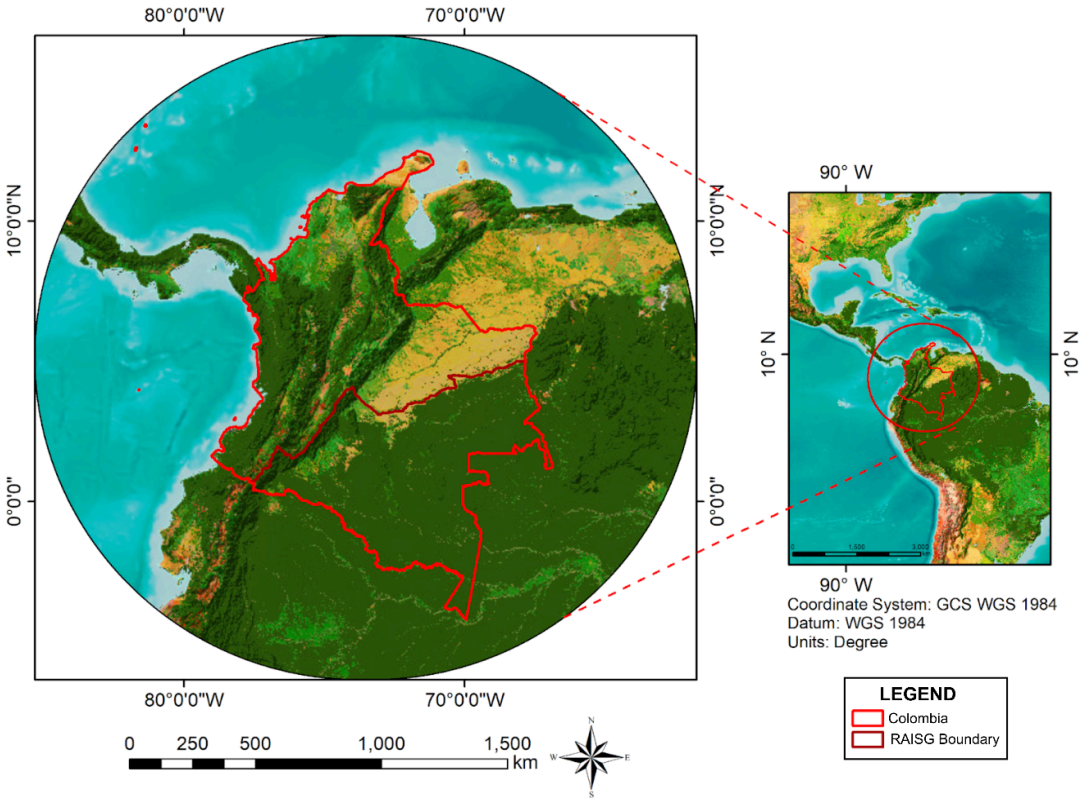


Figure 1. Study area for the detection of Urban Infrastructure cover in MapBiomas Colombia Collection 4.0.

Other global initiatives have mapped urban land using different remote sensing techniques and data sources (Schneider *et al*, 2010). Global products are available, such as FROM-GLC by Gong *et al*. (2013) and GlobeLand30 by Chen *et al*. (2015); however, most of this information covers a limited range of years, which has made it difficult to monitor urban dynamics over a long historical period.

The Landsat data series has made it possible to improve the availability of annual information for MapBiomas Colombia at the regional level, serving as the starting point for the present collection through automated cloud processing in Google Earth Engine. The methodological basis for the multitemporal mapping of Urban Infrastructure (ID: 24) over the last 41 years is described below, following this structure:

A. Composition of satellite image mosaics.
B. Cross-cutting multitemporal binary classification.
C. Post-classification.

2. Satellite image mosaics

2.1 Mosaic composition

The mosaic composition for the cross-cutting classification of "Urban Infrastructure" used the image mosaics generated for the general land cover classification of MapBiomas Colombia Collection 3.0. The mosaics were composed from atmospherically corrected surface reflectance satellite images from Landsat 4, 5, 7, 8 and 9 (Collection 2) covering the area within the boundary of Colombia, generated annually over a 41-year time series from 1985 to 2025[1]. Table 1 lists the image collections hosted in the Google Earth Engine repositories.

Table 1. Landsat data collections used in Collection 4.0.

Sensor	GEE Collection ID
Landsat 4 TM	LANDSAT/LT04/C02
Landsat 5 TM	LANDSAT/LT05/C02
Landsat 7 ETM+	LANDSAT/LE07/C02
Landsat 8 OLI /TIRS	LANDSAT/LC08/C02
Landsat 9 OLI 2 /TIRS 2	LANDSAT/LC09/C02

Figure 2 describes the general methodology for mosaic production. First, images were composited according to the availability of Landsat image collections for each year from 1985 to 2025 in the Google Earth Engine repositories. Image compositing used the International Map of the World (1:1,000,000) chart grid, for which the optimal parameters were defined to produce each mosaic according to criteria related to temporal range (start and end date), sensor, cloud cover, and excluded scenes that interfere with the quality of the median image.

In addition, the reference layer available in Google Earth Engine, DEM SRTM (Shuttle Radar Topography Mission - V03), was used.

[1] The mosaics were generated using the median pixel reducer at a spatial resolution of 30 meters.

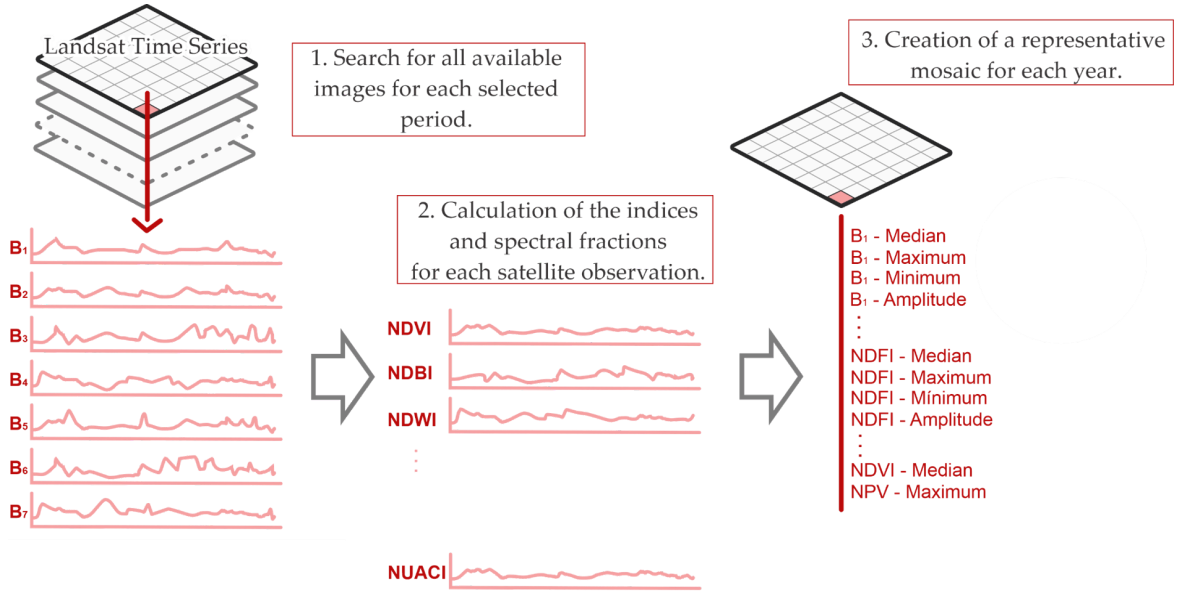


Figure 2. Annual mosaic generation method, Collection 4.0 MapBiomas Colombia.

Based on this parameterization, preprocessing included the application of the CFMask and CloudScore algorithms, which for Landsat 4, 5 and 7 (1985–2012 mosaics) used two bands: ***pixel_qa*** (pixel quality attributes generated by the CFMASK algorithm) and ***sr_atmos_opacity*** (atmospheric opacity; <0.1 = clear; 0.1–0.3 = average; >0.3 = hazy). If the cloud bit (5) is set and cloud confidence (7) is high, or if the cloud shadow bit (3) is set, the pixel is considered invalid; edge pixels that do not occur in all bands are therefore removed. For Landsat 8 and 9 (2013–2025 mosaics), the ***pixel_qa*** band was used, with bits 3 and 5 identifying clouds and shadows. After removal, the median of the cloud-free image collection was computed to generate the annual mosaic.

In addition, in Collection 4.0, a topographic correction was applied to minimize the effects of terrain relief on the reflectance recorded by the sensor. This process adjusts the reflectance values of the satellite images to compensate for the effects of terrain topography, such as slope and aspect. Using the solar incidence angle, the reflectance value of each pixel is adjusted so that it better represents the actual properties of the surface.

2.2 Selection of classification variables

The next step was to calculate spectral indices and fractions for each satellite observation (Table 2), including the Landsat spectral bands, spectral indices, and fractional and texture information derived from these indices and fractions. Static variables were also used: *HAND*, *shademask2*, *slppost*, *altitude*, *slope*, *latitude* and *longitude*; these support the classification of classes that are spectrally very similar but can be distinguished by their topographic conditions.

Table 2. Feature space: variables for the detection of Urban Infrastructure surfaces with the associated reducer.

Variable	Description	Reducer
blue	Landsat BLUE band	Median
green	Landsat GREEN band	Median
red	Landsat RED band	Median
nir	Landsat NIR band	Median
swir1	Landsat SWIR1 band	Median
swir2	Landsat SWIR2 band	Median
soil	Soil Fraction	Median
snow	Snow Fraction	Median
cloud	Cloud Fraction	Median
slope	Slope	Static topographic variable
ndvi	Normalized Difference Vegetation Index	Median
ndwi_mcfeeters	Normalized Difference Water Index (Mcfeeters)	Median
ndbi	Normalized Difference Built-up Index	Median
nuaci	Normalized Urban Areas Composite Index	Median

Finally, a representative mosaic was created for each year, consisting of 156 bands in total,[2] based on the calculation of statistical reducers to generate the value of each pixel (Table 3).

Table 3 Statistical reducers used in the image composition for Urban Infrastructure, Collection 4.0.

Reducer	Description
Median	Median of all values available in the annual mosaic for that location (pixel).
Dry season median	Statistical median applied to the pixels in the 25th percentile (lowest values NDVI (dry season proxy).
Wet season median	Statistical median applied to the pixels in the 75th percentile (highest values) of NDVI (wet season proxy).
Amplitude	Range of variation among all pixels available in the annual mosaic.
Standard deviation	Standard deviation of the values of all pixels available in the annual mosaic for a given location.
Minimum	Lowest value of all pixels available in the annual mosaic at a given location.
Maximum	Highest value of all pixels available in the annual mosaic at a given location.

[2] See the full description of bands and variables in the MapBiomas Colombia Collection 4.0 General ATBD.

Dry season minimum	Lowest value of all available pixels from the images in the quartile with the lowest NDVI values (dry season proxy).
Wet season minimum	Lowest value of all available pixels from the images in the quartile with the highest NDVI values (wet season proxy).
Dry season maximum	Highest value of all available pixels from the images in the quartile with the lowest NDVI values (dry season proxy).
Wet season maximum	Highest value of all available pixels from the images in the quartile with the highest NDVI values (wet season proxy).
Dry season QMO	Band value at the date of maximum EVI2 during the dry season.
Wet season QMO	Band value at the date of maximum EVI2 during the wet season.

3. Classification

The objective of the classification was to produce a binary layer distinguishing urban infrastructure from non-urban areas, considering only two possible values for each pixel (24: presence of Urban Infrastructure; 27: absence of Urban Infrastructure). The classification, based on training samples over the Landsat mosaics, was performed using the Random Forest (RF) classification algorithm (Breiman, 2001).

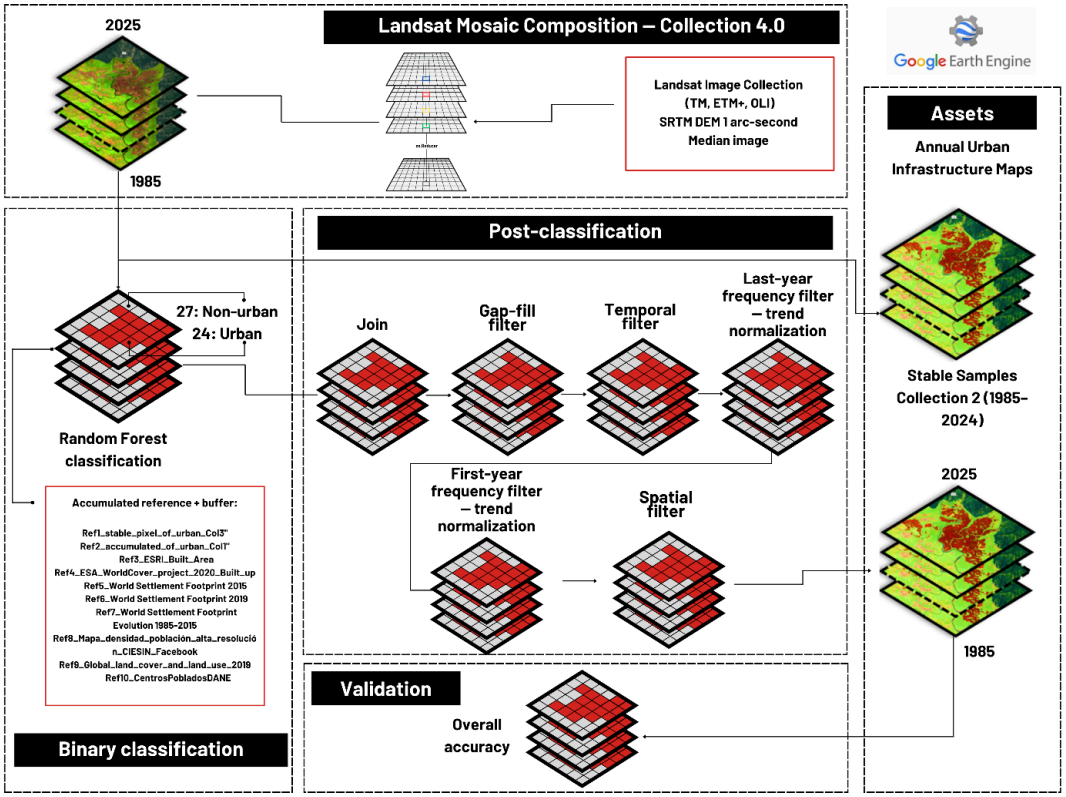


Figure 3. Methodological workflow for the Urban Infrastructure classification, Collection 4.0

For the Collection 4.0 cross-cutting theme, the classification was performed over 82 of the 86 regions defined for the country; 4 regions were excluded because no urban infrastructure is present in them, as verified using reference sources and high-resolution imagery.

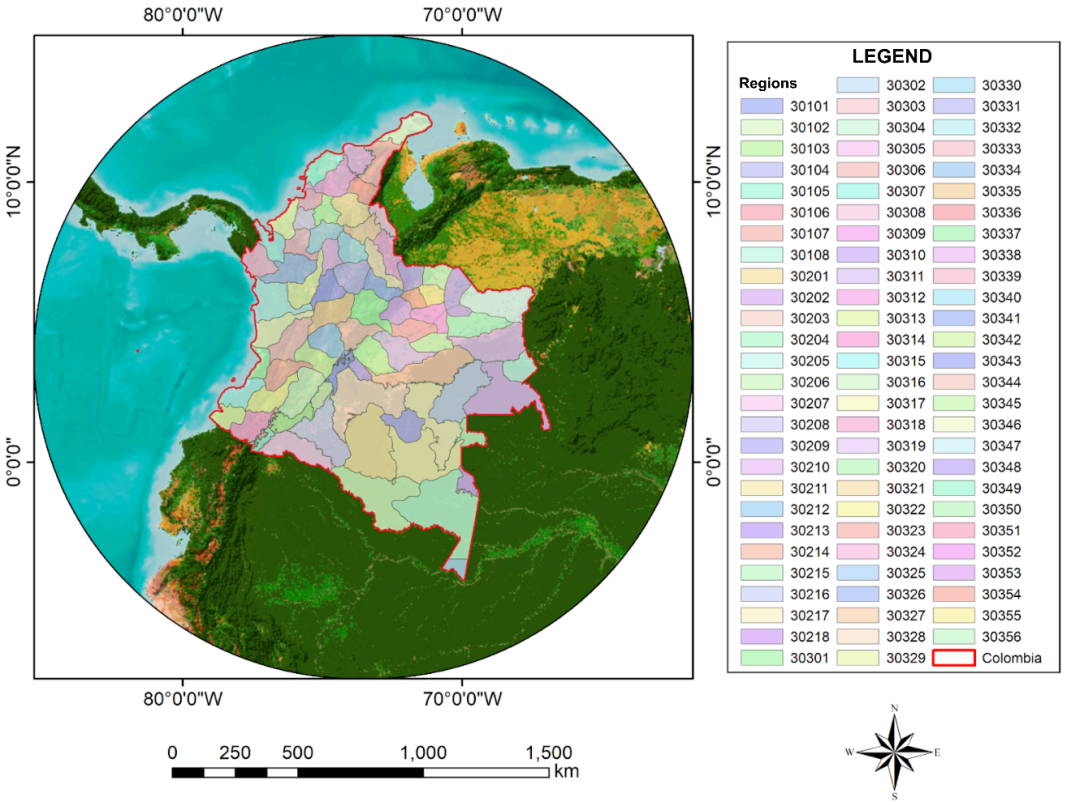


Figure 4. Urban Infrastructure classification regions, Collection 4.0.

3.1 Definition of the Urban Infrastructure class (ID: 24)

This cover class is defined as human settlement areas associated with large and small urban centers (towns) with built-environment infrastructure, such as road and railway networks and associated land, as well as other artificial areas such as hydrocarbon extraction facilities, hydroelectric plants, military bases, airports, port areas, and non-agricultural green areas such as recreational facilities in urban centers, urban lawns, road medians, and unconventional airstrips in rural areas. Peripheral areas undergoing a gradual process of urbanization for residential and/or industrial purposes are also considered (Table 4).

Table 4. Detailed definition of Urban Infrastructure, Collection 4.0.

Includes	Excludes
1. Urbanized areas 1.1 Continuous urban fabric (large population centers) 1.2 Discontinuous urban fabric (towns)	Unpaved roads and tracks in rural areas
2. Road networks	Indigenous communities

2.1 Road and railway networks and associated land

3. **Other artificial areas**

3.1 Hydrocarbon extraction facilities (oil blocks)

3.2 Hydroelectric plants

3.3 Military bases

3.4 Airports (airstrips in urban centers)

3.5 Port areas

3.6 Other industrial-scale facilities

4. **Green areas (non-agricultural)**

4.1 Recreational facilities in urban centers

4.2 Urban lawns and road medians

Unconventional airstrips in rural areas



A **B** **C**

Figure 5. Urban Infrastructure detection, Collection 4.0: (A) Landsat mosaic, (B) urban surface classification ID: 24, (C) Google Earth high-resolution image (Barranquilla - Colombia).

3.2 Classification algorithms

The Random Forest machine learning algorithm (Breiman, 2001) was implemented on the Google Earth Engine platform. Training samples were generated and the classification was then performed based on the behavior of the indices used, supported by reference maps available at the regional level, such as OpenStreetMap, as well as national and global reference data.

3.3 Reference data

For Colombia, nine reference sources were used in the Random Forest algorithm. They were compiled from projects providing global geographic data for different periods, and also included the accumulated and stable samples of MapBiomas Colombia Collection 3.0 based on annual data from 1985 to 2024. The data were filtered according to the presence of urban infrastructure surfaces or urban clusters with residential, commercial, service, industrial, park, and other attributes.

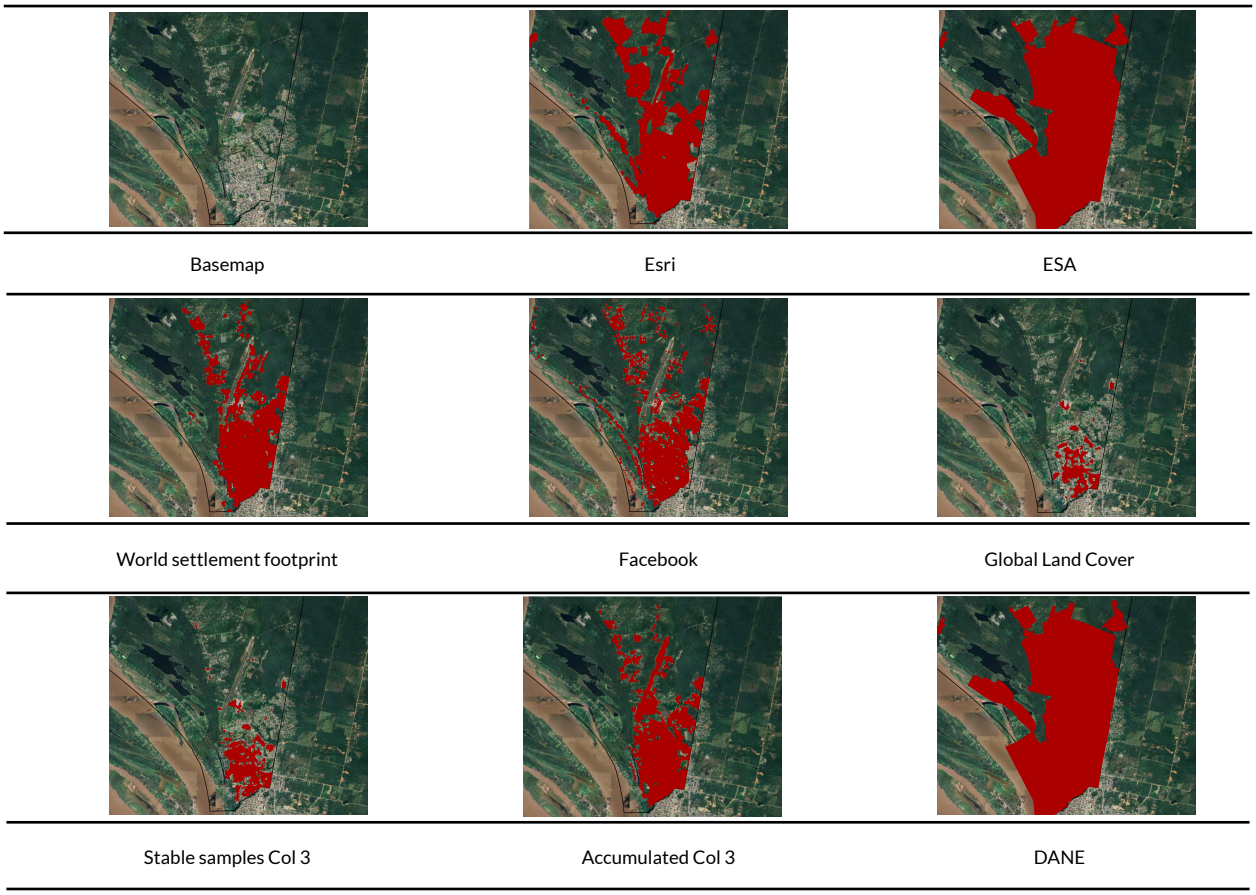


Figure 6. Reference sources available for the MapBiomas Colombia study area.

Tables 5 and 6 describe the secondary data used in Colombia as reference sources in the urban infrastructure classification process.

Table 5. Colombian data used as global reference in the urban infrastructure classification process.

Global references	
Global references	**Colombia**
Ref1_stable_pixel_of_urban_Col3	🔴
Ref2_accumulated_of_urban_Col3	🔴
Ref3_ESRI_Built_Area	
Ref4_ESA_WorldCover_project_2020_Built_up	🔴
Ref5_World Settlement Footprint 2015	
Ref6_World Settlement Footprint 2019	🔴
Ref7_World Settlement Footprint Evolution 1985-2015	
Ref8_Mapa_densidad_población_alta_resolución_CIESIN_Facebook	🔴
Ref9_Global_land_cover_and_land_use_2019	🔴

Table 6. Colombian data used as national reference in the urban infrastructure classification process.

National references	Source
Urbanized areas, industrial or commercial areas and communication networks, artificial green areas (non-agricultural), roads, residential areas, buildings.	Land Cover of Colombia, CORINE Land Cover methodology adapted for Colombia - IDEAM (2024).
Integrated urban area	Map of continental, coastal and marine ecosystems of Colombia - IDEAM (2017).
Oil wells	OpenStreetMap (2021) Populated places layer for Colombia -DANE (2018). Petroleum information bank, Wells - Colombian Geological Survey (2022)

4. Post-classification

Due to the pixel-based nature of the classification method and the long time series, a post-classification process was carried out, applying filters (gap fill, temporal, spatial, and frequency) to the binary Urban Infrastructure classification in order to reduce temporal inconsistencies, remove classification noise smaller than the minimum mapping unit, and fill data gaps. In addition, a "join" step was included to merge the years classified or reclassified in Collection 4.0 with the years already processed in Collection 3.0.

4.1 Merging of Collection 3.0 and Collection 4.0 years

Since Collection 2.0, a step prior to the post-classification filters has been implemented. This step merges the years classified for the current collection (e.g., 2025) with those already classified and processed in the previous collection (e.g., 1985–2024). The result is a single image stack, which is then processed with the standard filters. This step also made it possible to modify the classification of specific areas in the previously processed years (before 2025), in order to improve classification quality across the entire time series.

4.2 Gap filling - Gapfill filter

The filter sequence for the urban infrastructure cross-cutting theme begins with gap filling (*gap fill*), since the area covered by MapBiomas Colombia and its urban surfaces are subject to atmospheric and climatic conditions that result in annual mosaics with some pixels lacking information or satellite observations ("gaps"). This filter reduces these residual gaps by replacing them with the temporally closest value. When a "future" pixel has no value (no data), the gap fill filter assigns it the value of the nearest year. The filter reviews the series by first filling gaps in a "back-to-front" sweep, in which gaps are filled with data from the nearest preceding years. Any remaining gaps are filled with data from the nearest following year. For each pixel whose value has been filled by this filter, the change is recorded in a metadata file that stores the pixel history. Several years may be used to fill gaps. Therefore, data gaps persist only if a given pixel has been permanently classified as having no data throughout the entire time series.

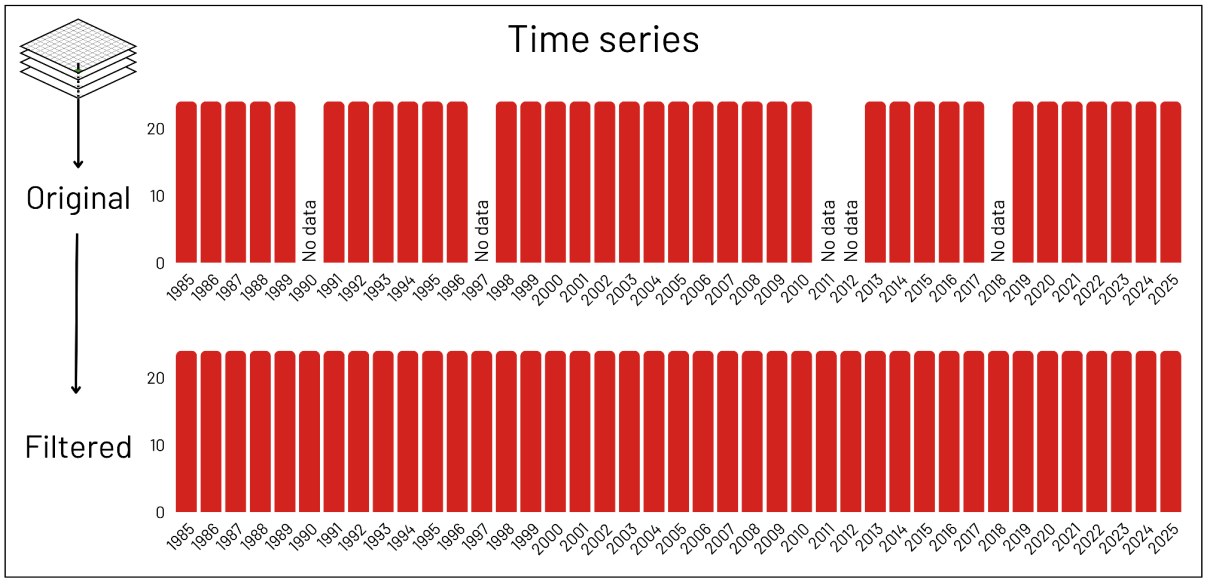


Figure 7. Gapfill filter.

4.3 Temporal filter

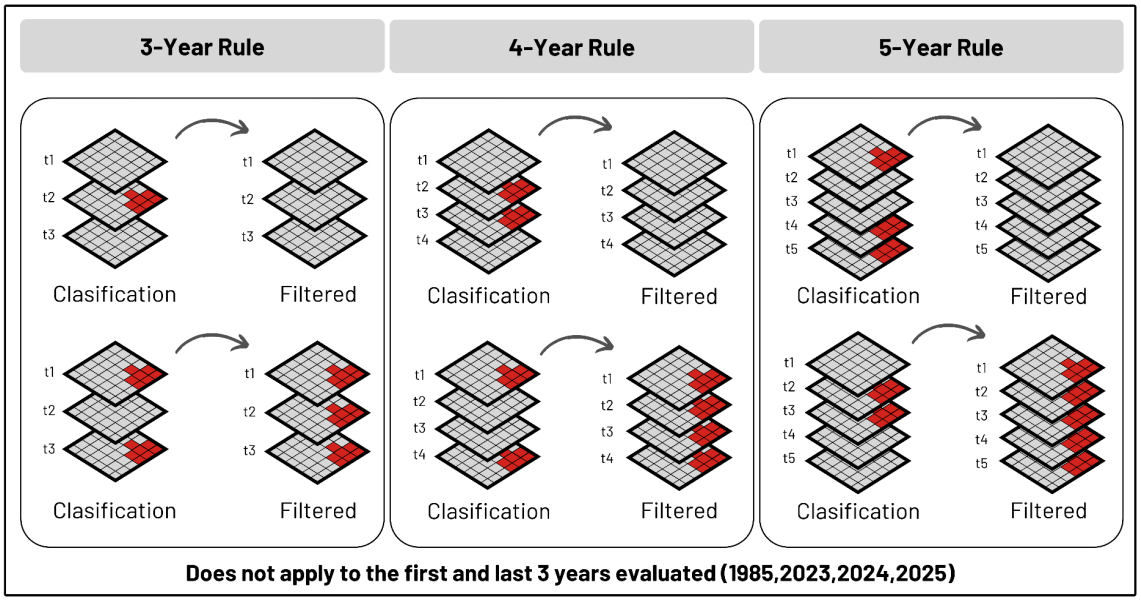
A temporal filter was applied using 3-, 4- and 5-year windows over the entire time series. The filter examines the value of each classified pixel in relation to its value in temporally consecutive classifications. To do so, it uses a unidirectional moving window that considers sequences of 3 to 5 years of classifications and identifies non-permitted temporal transitions. The temporal filter is applied to every pixel in all years of the collection. Depending on the year to be modified, there are three types of rules:

General rules (GR). Applied to pixels of years in intermediate positions within 3- to 5-year sequences. This rule is applied only where there is a temporal inconsistency, for example, when sequences of consecutive years have identical values except for the pixel in the central position. In these cases, the filter changes the value of the central pixel so that it is consistent with the preceding and following pixels. For 3-year sequences, there is only one possible central position or intermediate year. For 4- or 5-year sequences, there are two or three possible central positions. This rule modifies the classification values for the years 1986 to 2023.

First-year rules (FR). Applied only to the first year of the time series. This rule modifies the classification values for 1985.

Last-year rules (LR). Applied to the last year of the classification. This rule modifies the classification values for 2025.

In this way, the temporal filters reduce data gaps and temporal inconsistencies, as well as changes that are not possible or not permitted.



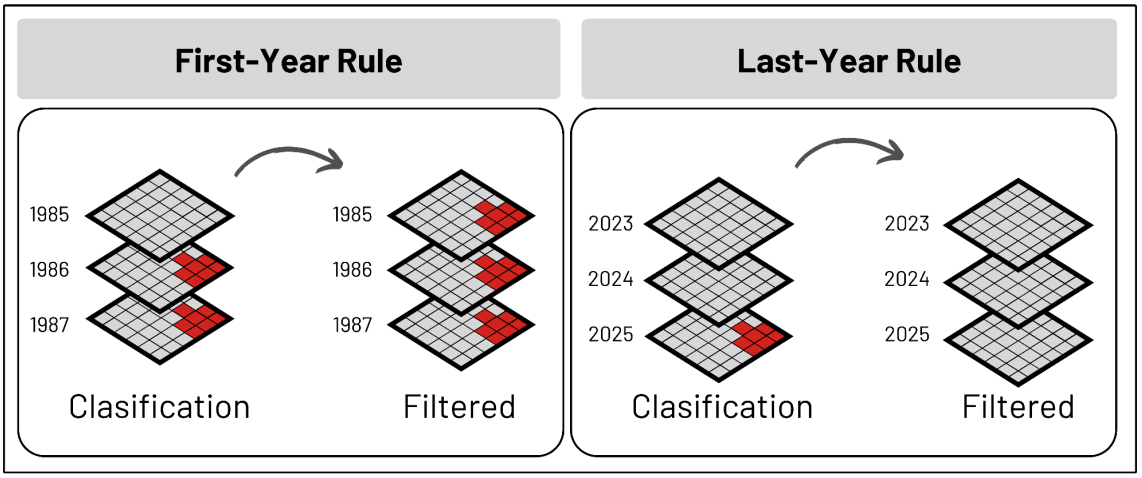


Figure 8. Temporal filter.

4.4 Adjusted frequency filter

This filter considers the occurrence of the class throughout the time series in order to normalize its historical trajectory. Given that urban infrastructure shows an increasing, expansive trend over the series, two filters (first and last year) were applied to regulate the consistent growth of urban pixels and prevent anomalous fluctuations from one year to the next.

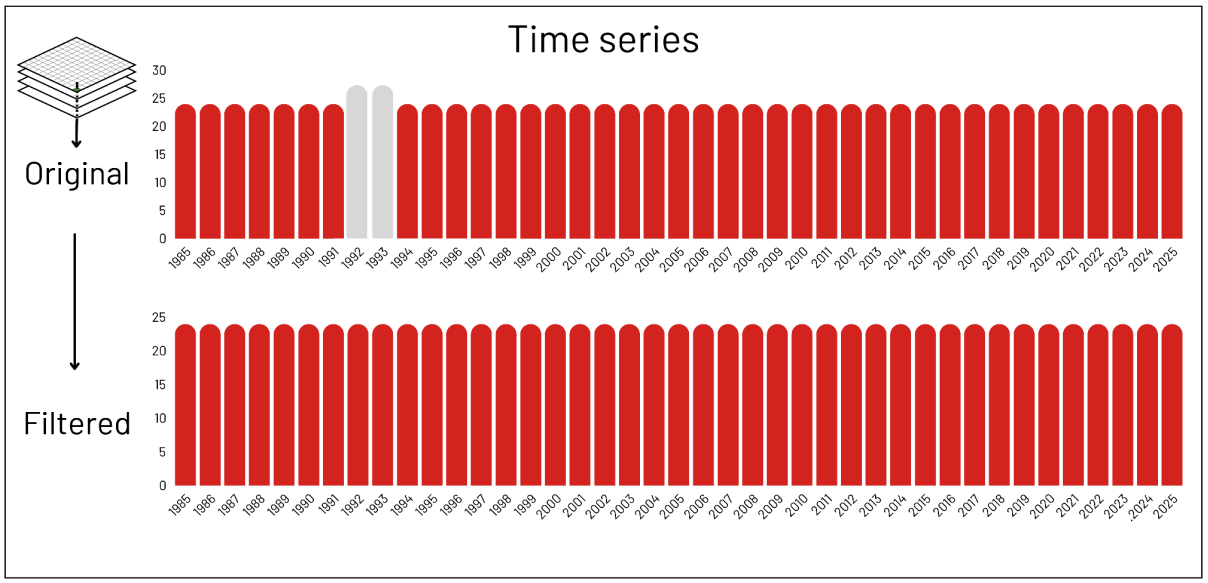


Figure 9. Frequency filter.

4.5 Spatial filter

The last filter in the post-classification sequence was the spatial filter based on the Google Earth Engine "connectedPixelCount" function. This function identifies connected (neighboring) pixels that share the same value using a moving window. Only pixels that are not connected to a predefined number of identical neighbors are considered isolated pixels. For MapBiomas Colombia, the minimum mapping unit was set at 0.5 ha (approximately 5 pixels). Consequently, at least five connected pixels were required to meet the minimum connectivity criterion. In this way, the spatial filter smooths local differences by removing isolated or edge pixels smaller than 0.5 ha, increasing the spatial consistency of the class.

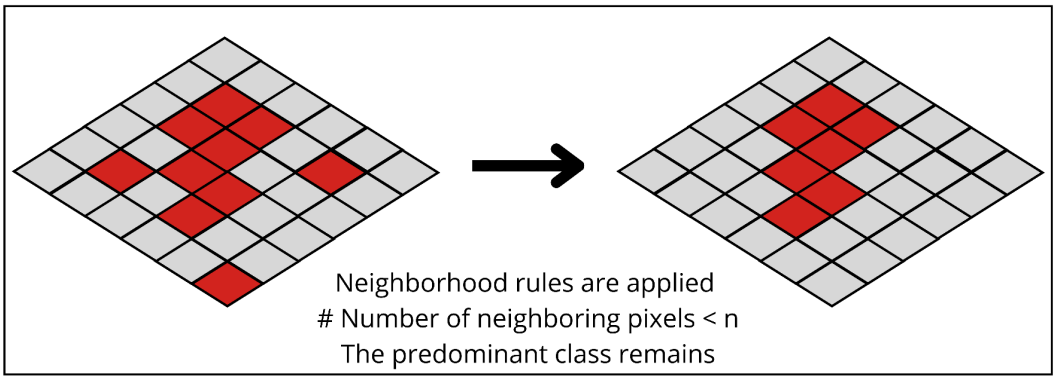


Figure 10. Spatial filter.

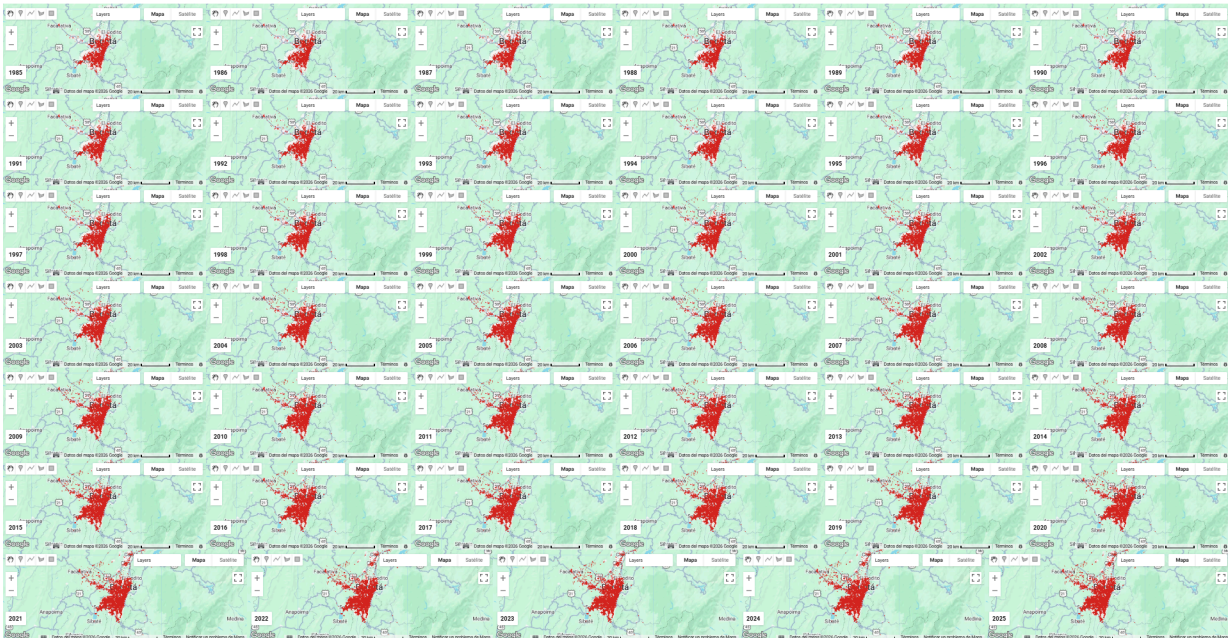


Figure 11. Multitemporal classification of Urban Infrastructure, Collection 4.0 - 1985 to 2025 (case Bogotá - Colombia).

5. References

Breiman, L. (2001). Random Forest. Statistics Department University of California Berkeley. 1-30.

Chen, J., Chen J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S. & Lu, M. (2015). Global land cover mapping at 30 m resolution: a POK-based operational approach. ISPRS J. Photogramm. Remote Sens. 103, 7-27.

Gong, P., Wang, J., Yu, L., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H. & Liu, S. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat ™ and ETM+ data. Int. J. Remote Sens. 34, 2607-2654.

Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. (2014) Good Practices for Estimating Area and Assessing Accuracy of Land Change. Remote Sens. Environ. 148, 42–57, doi:10.1016/j.rse.2014.02.015.

Schneider, A., Friedl, M.A., Potere, D. (2010). Mapping global urban areas using MODIS 500-m data: new methods and datasets based on 'urban ecoregions'. Remote Sens. Environ. 114, 1773-1746.

Souza, C., Monteiro, D., Shimbo, J. & Azevedo, T. (2019). Urban Infrastructure Appendix. Version 1. Retrieved July 2, 2021, from https://mapbiomas-br-site.s3.amazonaws.com/Infraestrutura_Urbana__Appendix_-_ATBD_Cole%C3%A7%C3%A3o_4_dez_2019.pdf

UCLG (2018). Amazonian Cities. Peer Learning about the sustainable use of terrestrial ecosystems #23. United Cities and Local Governments. Bolivia. 36 p.