# MapBiomas Argentina - Urban areas “Handbook” Algorithm Theoretical Basis Document (ATBD)

## Collection 2

## Version 1

AUTHORS

Natasha PICONE (CONICET/UNCPBA)
Santiago LINARES (CONICET/UNCPBA)
Martina ÁLVAREZ (UNCPBA)
Luna SCHTEINGART (Fundación Vida Silvestre)
Gonzalo DIEGUEZ GAVIOLA (Fundación Vida Silvestre)
Ana Virginia ELJALL QÜESTA (Fundación Vida Silvestre)
Pablo RIZZO (UNCuyo)
Laura GOMEZ (CONICET/UNNE)
Osvaldo CARDOZO (CONICET/UNNE)
Pamela FLORES (UNGS)

## Table of Contents

## 1. Introduction

This document explains all the processes implemented to produce the second collection of urbanized areas for Argentina for the period 1985-2025. Urbanized areas are defined as areas with a high density of roads and buildings and a predominance of non-vegetated surfaces, including infrastructure and open spaces. This report focuses on the methodological approach and decisions made by the team to produce this product.

The methodology draws on the MapBiomas Brasil Urban Areas Collection 10 experience and lessons learned from MapBiomas Argentina Collection 1. The most significant evolution for Collection 2 is the adoption of a geographic delimitation based on the 16 ecoregions of Argentina defined by Matteucci et al. (2005). This shift from macro-regions to ecoregions allows for the calibration of classification models and post-classification rules according to the local environmental characteristics of each geophysical unit, solving growth rate biases in Northern Argentina and spectral confusions in complex transitions (e.g., transition from Patagonian Forest to Patagonian Steppe).

## 2. MapBiomas Argentina - Urban areas

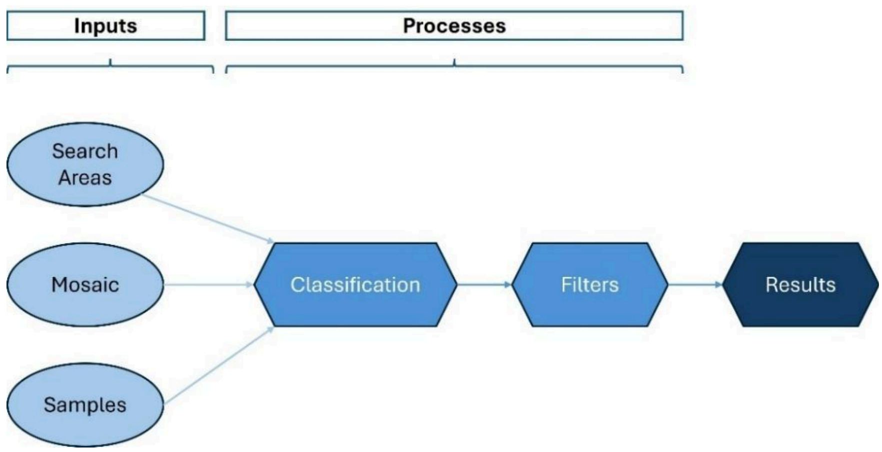
MapBiomas Argentina - Urban areas is a new initiative in the project of MapBiomas Argentina. The working group started its activities during 2024 and was formalized at the beginning of 2025. Currently, it is a group of academics and NGOs focusing on the analysis of how urbanized areas in Argentina have been changing.

The group is coordinated by members of Universidad Nacional del Centro de la Provincia de Buenos Aires (Natasha Picone and Santiago Linares) and Fundación Vida Silvestre Argentina (Ana Virginia Eljall Qüesta, Gonzalo Dieguez Gaviola and Luna Schteingart). There are also three other national universities that actively participate in the generation of this first collection: Universidad Nacional de Cuyo (Pablo Andrés Rizzo), Universidad Nacional del Nordeste (Osvaldo Cardozo and Laura Fabiana Gomez) and Universidad Nacional de General Sarmiento (Andrea Pamela Flores).

## 3. Methodological Description

Most of the methodology's processes are developed on the Google Earth Engine (GEE) platform. The procedure is divided into input generation, probability classification, score harmonization, threshold determination, and spatial-temporal filtering. Figure N° 1 shows the overview of the steps that have been taken in Argentina to obtain the urbanized areas from 1985 to 2025.

Figure N°1. Methodological overview



Source: MapBiomas Argentina - Urban areas (2025)

Table 1 shows the complete series of data sources used in this project and in which part of the process they were used.

Table 1. Sources and data used

| Source | Used Information | Methodological Step | Cover |
|---|---|---|---|
| **INDEC** | Census urban areas (2022) | Search areas | National |
| **INDEC** | Population per census urban area (2022) | Search areas | National |
| **RENABAP** | Slums | Search areas | National |
| **10 m Annual LULC** | Class 7: “Built area” (2024 y 2025) | Search area / Spatial filter | Global |
| **Global Urban Settlement** | GHS-S Urbanized area (1985/2005/2020) | Sample | Global |
| **MapBiomas Argentina Collection 2** | Stable classes for: agriculture, bushes, water, and forest | Sample | National |
| **LANDSAT 5, 7, 8 and 9 Collections** | Mosaic on the fly per year (1985 – 2024) | Classification | Global |

Source: MapBiomas Argentina - Urban areas (2025)

## 3.1. Input preparation

### 3.1.1. Search Area

Based on the preliminary analysis, two complementary sources were used to define the search areas of the algorithm to be applied for the Argentine territory. The search areas were constructed from an intersection between Census Urban Areas provided by INDEC (2022), the Urban Class (7) of the Land Cover Land Use (LCLU) 2024 provided by Impact Observatory, and the Slums Census from RENABAP (2023).

The population registered by the 2022 National Census of Population, Households, and Housing (INDEC, 2022) was annexed to these search areas. From this input, and to include transitional or expansion areas in cities, variable buffers were constructed based on the urban category (INDEC, 1998), defined as follows: Agglomerated rural population (less than 2000 inhabitants), 300 meters; Large Villages (2000 to 19999 people), 600 meters; Small Towns (20000 to 49999), 900 meters; Minor ATIS (50000 to 99999 people), 1200 meters; Intermediate ATIS (100000 to 399999), 1500 meters; Great ATIS (400000 to 999999 inhabitants), up to 2000 meters, Large Cities (more than 1 million people), 2500 meters buffer and the Buenos Aires Agglomeration, 3000 meters.

### 3.1.2. Mosaic and ecoregions

Mosaics are generated "on the fly" in Google Earth Engine using the Landsat repositories covering the 1985-2025 period (sensors TM from Landsat 5, ETM+ from Landsat 7, and OLI from Landsat 8 and 9). The images were corrected for clouds and shadows using the CFMASK algorithm.

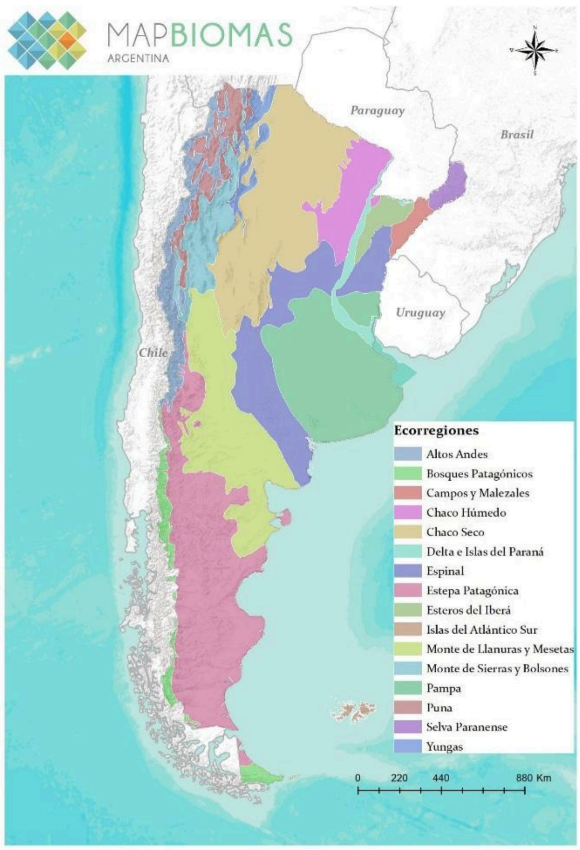
Table 2: List of Bands, Indexes, and Fractions that are created to generate the annual Mosaic.

| Dataset Type | Category | Parameters |
|---|---|---|
| Bands | Landsat bands | BLUE, GREEN, RED, NIR, SWIR1, SWIR2 |
| Indexes | Vegetation | NDVI, EVI, EVI2, SAVI |
| Indexes | Water | MNDWI, NDWIm, AWEish |
| Indexes | Soil, Urban | NDBI, UI, NDUI, BSI, NBR, NDRI, BAI, BU, NDFI |
| Spectral Mixtures | Mixtures | GV, NPV, SOIL, CLOUD, GVS, SHADE, SUBS, VEG, DARK |
| Fractions | Index Fractions | NDVI_75, NDWIm_75, NDBI_75, EVI_10, EVI_90, EVI2_10, EVI2_90, EBBI_raw |
| Fractions | Band Fractions | SR_10, SR_90, SR_dif1090, RAW_25, RAW_75, RAW_90, RAW_dif2575 |

Source: MapBiomas Brasil Urban areas (2024)

For Collection 2 of the Urbanized Areas, the processing framework was applied across 16 distinct ecoregions (Figure 2). This spatial subdivision reduces regional heterogeneity compared to previous methods, allowing the algorithm to more accurately learn the local environmental context surrounding each search area.

Figure 2: Argentine ecoregions.




Source: MapBiomas Argentina - Urban areas (2026)

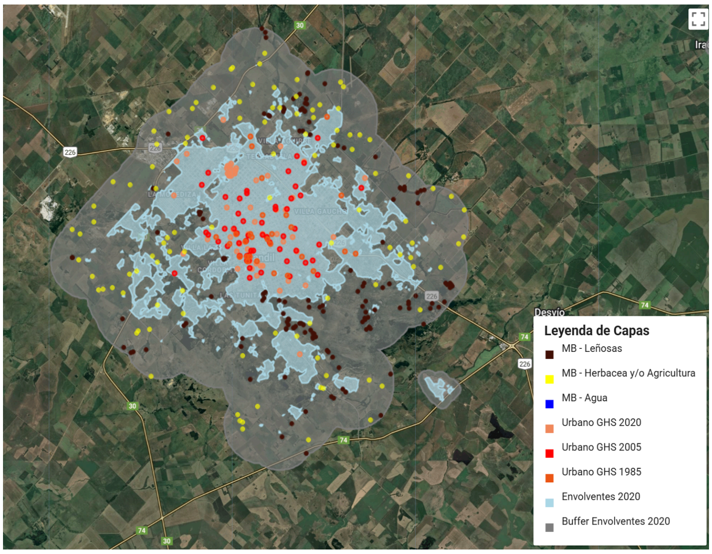
### 3.1.3. Samples

A statistically calculated total of 16,641 samples (5,756 urban and 10,885 non-urban) across all ecoregions was required to train the algorithm.

- **Urban Samples:** Extracted from the built-up surface product (R2023) developed by the JRC of the European Commission. Data from 1985, 2005, and 2020 were used, selecting pixels with at least a 10% built-up surface (>= 1000 square meters).
- **Non-Urban Samples:** Extracted from the stable samples of MapBiomas Argentina Collection 2 (water, herbaceous, crops, and forests). Only samples located between the urban search perimeter and the buffer area were considered (Figure 3).

In several ecoregions, it was determined that balancing urban and non-urban samples optimized the classification. Additionally, complementary samples were generated via visual analysis in areas where the initial classification was confusing uncovered/non-vegetated (*e.g.,* bare ground) areas with urbanized zones.

Figure 3. Example of urban and non-urban samples




Source: MapBiomas Argentina - Urban areas (2025)

## 3.2. Processing

### 3.2.1 Classification with Random Forest

The classification was executed independently by ecoregion using the Random Forest algorithm in Google Earth Engine. The model was built using 100 decision trees (*numberOfTrees: 100*) with a minimum leaf population of 5 (*minLeafPopulation: 5*). Utilizing the on-the-fly mosaics alongside the search areas and optimized samples, the model calculates the score (0-100) for each pixel representing an urbanized area for every year between 1985 and 2025.

### 3.2.2. Score Harmonization

Given the 40-year independent classification timeline, interannual variations in probability scores emerge due to changes in soil moisture, adjacent vegetation phenology, atmospheric artifacts, and illumination. To eliminate this spectral noise, a continuous temporal window of 5 years (t-2, t-1, t, t+1, t+2) is applied to the probability scores. This weighted moving average stabilizes the historical series by attenuating abnormal peaks and drops.

### 3.2.3. Determination of Thresholds

Transforming the continuous score map into a binary classification (Urban / Non-Urban) requires adaptive thresholding. The optimal cutoff point is defined per ecoregion and classification period (1985-2004, 2005-2019, 2020-2025) using two primary statistical criteria:

- **ROC Curve (Receiver Operating Characteristic):** Identifies the point that maximizes the detection of urban areas while minimizing confusion with other land covers.

- **15th Percentile (P15):** Based on MapBiomas standards, defining the operative threshold through the 15th percentile of the probability values of true urban training samples.

### 3.2.4. Spatial and Temporal Filters

To transition from annual binary layers into a coherent cartographic sequence, a structured succession of post-classification filters is applied.

**Spatial Filter 1:** Uses a 3x3 pixel mobile window (kernel) to evaluate geographical contiguity. It eliminates salt-and-pepper noise by discarding isolated pixels and fills in logical gaps within consolidated urban tissues.

**Temporal Filters 1 to 3:** Assess the pixel trajectory utilizing a 5-year moving window (t-2, t-1, t, t+1, t+2). The urban class is validated for a given year only if it persists in at least 3 out of the 5 years. It smooths interannual inconsistencies and interpolates temporal data absences (e.g., caused by clouds).

**Spatial Filter 2:** Implements a supermask using external high-resolution data (Copernicus LULC Class 7 derived from Sentinel). It removes persistent false positives (such as quarries, dry rivers, or salt flats) in peri-urban or rural contexts that survived earlier filters due to spectral similarities.

**Temporal Filter 4 (Irreversibility Rule):** Assumes that the transformation of natural or agricultural land into paved/constructed surfaces is a cumulative, irreversible process. Once a pixel is validated and consolidated as urban, it cannot revert to a non-urban condition in subsequent years, ensuring progressive growth trajectories.

## 4. References


- Chuvieco, E. (2016). Teledetección Ambiental. La observación de la Tierra desde el Espacio. 3rd ed. Ariel, Barcelona.
- Impact Observatory (2025). *10 m Annual Land Use Land Cover (LULC) map* [Dataset]. Copernicus / Sentinel Hub Collections. Retrieved September 11, 2025, from https://collections.sentinel-hub.com/impact-observatory-lulc-map/
- European Commission (2024). *Global Human Settlement Layer (GHSL) – Download Portal*. Copernicus Emergency Management Service. Retrieved from https://human-settlement.emergency.copernicus.eu/download.php
- INDEC (1998). CENSO 91. El concepto de localidad: definición, estudios de casos y fundamentos teóricos- metodológicos. Serie D N° 4, Buenos Aires.
- MapBiomas Brasil (2025). Urban Areas - Algorithm's Theoretical Basis Document (ATBD) (Collection 10, Version 1).
- MapBiomas Argentina (2025). MapBiomas Argentina - Colección 2 de los Mapas Anuales de Cobertura y Uso del Suelo en Argentina.
- Matteucci, S. D. (2005). Ecorregiones y complejos ecosistémicos de la Argentina. Buenos Aires: Orientación Gráfica Editora.