Monte, Puna and High Andes

Collection 2
Version 1

General coordinator
Ana Eljall

Technical coordinator
Hernán Dieguez

Technical support
Paula Torre Zaffaroni
Leticia Castro Sardiña

Team

Abril Barrios
Carolina Ramos
Eduardo Jorge Ocher
Gonzalo Dieguez Gaviola
Jazmin Suarez
Joaquin Asad

Jose Volante
Luna Schteingart
Mariana Guerra Lara
María Luisina Torre Azzaroni
Silvina Alday
Sofia Antonena

September 2025

1. Introduction

1.1. Scope and content of the document

The objective of this document is to describe the theoretical basis, justification and methods applied to produce annual maps of land use and land cover (LULC) of the *Monte, Puna and High Andes* MapBiomas initiative from 1985 to 2024 (Collection 2). This collection was integrated into Collection 2 of Argentina to generate a characterization of the whole country. The document presents a general description of the satellite image processing, the feature inputs, and the step-by-step process carried out to obtain the annual classifications.

1.2. Overview

The MapBiomas initiative combines people, algorithms, satellite information and large-scale processing in a methodology that has revolutionized the operational large-scale generation of LULC maps. MapBiomas provides an ideal environment to enhance and share skills and abilities by collaborators from different locations and perspectives but with the mutual core value of learning by doing. Thanks to Google Earth Engine and open source technology it was possible to access and process large-scale datasets of satellite imagery such as the one generated by the MapBiomas project.

MapBiomas collections aim to contribute to developing a fast, reliable, collaborative, and low-cost method to process large-scale datasets and generate historical time series of land use and land cover annual maps. All data, classification maps, codes, statistics, and further analyses are openly accessible through the MapBiomas Platform. This is possible thanks to i) the Google Earth Engine platform, which provides access to data, image processing, standard algorithms, and the cloud computing facilities; ii) the freely available Landsat time-series dataset; and iii) the MapBiomas collaborative network of organizations and experts that share knowledge and mapping tools. MapBiomas maps are an evolving product and subsequent collections will be available in the future.

1.3. Region of Interest

The *Monte, Puna and High Andes* initiative was created to produce annual land use and land cover (LULC) maps for the Argentine territories west of Mapbiomas Chaco and Pampas and north of Patagonia. The region of interest encompasses a total of 495,123 km², including most of the Puna, Low and High Monte, and High Andes ecoregions, as well as the northern part of the Patagonian steppes (Burkart et al., 1999; Olson et al., 2001).

The study area was divided into 11 subregions to minimize confusion between samples and classes, and to ensure a more balanced distribution of samples and results (Figure 1).

Figure 1. Region of interest of *Monte, Puna and High Andes* initiative.

1.3.1. Monte

East of the Andes lies the Monte ecoregion, encompassing arid and semi-arid steppes. This vast ecoregion is dominated by shrubs and perennial grasses, well-adapted to the region's low rainfall and hot summers. A network of rivers and streams originating in the Andes provides vital water sources for the Monte's socio ecosystems. The Monte's arid steppes have traditionally been used for extensive livestock ranching, primarily sheep and goats.

However, overgrazing can lead to desertification, posing a threat to the region's delicate ecological balance. Monte's aridity generally restricts intensive agriculture. However, pockets of intensive agriculture exist particularly near the foothills of the Andes where rivers provide a reliable source of water for irrigation. These areas might support vineyards, fruit orchards, alfalfa production, horticulture, pecan groves, walnut orchards, or other woody cultivated areas. Most towns and cities rely on oases or rivers for their water needs.

1.3.2. Puna

High up in the Andes, nestled between its peaks and foothills, lies the Puna, a vast plateau characterized by a cold, dry climate. It experiences frigid temperatures, strong winds, and intense solar radiation. Precipitation is scarce. The Puna's unique ecological conditions have fostered the development of a specialized ecosystem dominated by low-growing grasses, cushion plants, and dwarf shrubs. The Puna is an important source of freshwater, feeding rivers that flow eastward. The Puna's high-altitude steppes have long supported communities of indigenous herders who raise llamas, alpacas, and vicuñas for their wool and meat. The Puna is also rich in mineral resources, leading to mining activity. Some communities boast a long history of cultivation of small plots using water from rivers fed by glacial meltwater. Large-scale urbanization is also restricted by water scarcity. Existing towns and cities in the Puna rely on rivers or underground water sources for their needs.

1.3.3. High Andes

Nestled along the western edge of Argentina, the High Andes mountain range stretches for thousands of kilometers, forming a formidable natural barrier. Many of its imposing peaks exceed 6,000 m in elevation. The region is characterized by a harsh, high-altitude climate with frigid temperatures, strong winds, and significant variations in temperature throughout the day and the year. Precipitation is scarce, falling mainly as snow that accumulates on the peaks, feeding glaciers and contributing to the region's unique hydrological systems. The thin air and intense solar radiation pose challenges for most life forms. However, the High Andes ecosystem has adapted, harboring a variety of specialized plant and animal species, many of which are endemic to the region. Despite the harsh conditions, the High Andes have a long history of human use. Indigenous communities have adapted to the high altitude, practicing a transhumance lifestyle where they move herds of llamas and alpacas between seasonal pastures at lower and higher elevations. Mining for precious metals like gold and silver has a long history in the High Andes, but its environmental consequences are now raising growing concerns and require careful

consideration. Some oasis pockets at lower elevations benefit from meltwater from glaciers and snowpack, allowing for small-scale agriculture. Urbanization is also limited due to the extreme conditions. Small settlements exist primarily to support mining activity or tourism.

1.4. Enhancements Introduced in Collection 2 Compared to Collection 1

Collection 2 improves upon Collection 1 through several key enhancements. First, a thorough qualitative revision of the training samples was conducted, ensuring that their thematic accuracy and spatial representativeness meet updated mapping standards. This was complemented by the application of a quantitative outlier filter to the original training samples from Collection 1, removing anomalous or inconsistent data samples and improving the overall reliability of the classification process. The legend was expanded to allow for a separation of woody vegetation classes, explicitly distinguishing forests from shrublands to provide greater thematic detail. Additionally, an urban class was incorporated, developed by a dedicated transversal working group that integrated information from multiple sources. Finally, the temporal coverage was extended to include data up to the year 2024, allowing for a more up-to-date representation of land cover and land use dynamics.

2. Overview of methodological process

The methodological steps of *Monte, Puna and High Andes* initiative Collection 2 are presented in Figure 2 and detailed below. Annual Landsat mosaics based on yearly periods and the spectral feature inputs derived from the Landsat bands (i.e., feature space) to run a random forest-based classification were generated by the MapBiomas Brazil team. The *Monte, Puna and High Andes* team defined the classes of land use or cover to be mapped (hereafter, map legend) and the subregions of the region of interest. Through visual interpretation and phenological signatures analysis by the whole team, temporally stable training samples were acquired for each class and subregion. The training data set was derived from the spectral feature inputs over the samples, and a pre-classification was generated with a random forest classifier. The pre-classification process was used to identify stable areas. Within these areas, stable samples were randomly selected through stratification, ensuring representativeness based on the relative area of each subregion. A new random forest classifier was run based on random, stratified, stable samples. Interpreters further collected complementary samples in areas or for classes where errors were evident, which were added to the stable samples for a new classification. Land

use/land cover patterns were evaluated and corrected using reclassification rules designed specifically for classes and regions where errors were detected. Following that, spatial and temporal filters were applied to remove classification noise and stabilize the classifications.
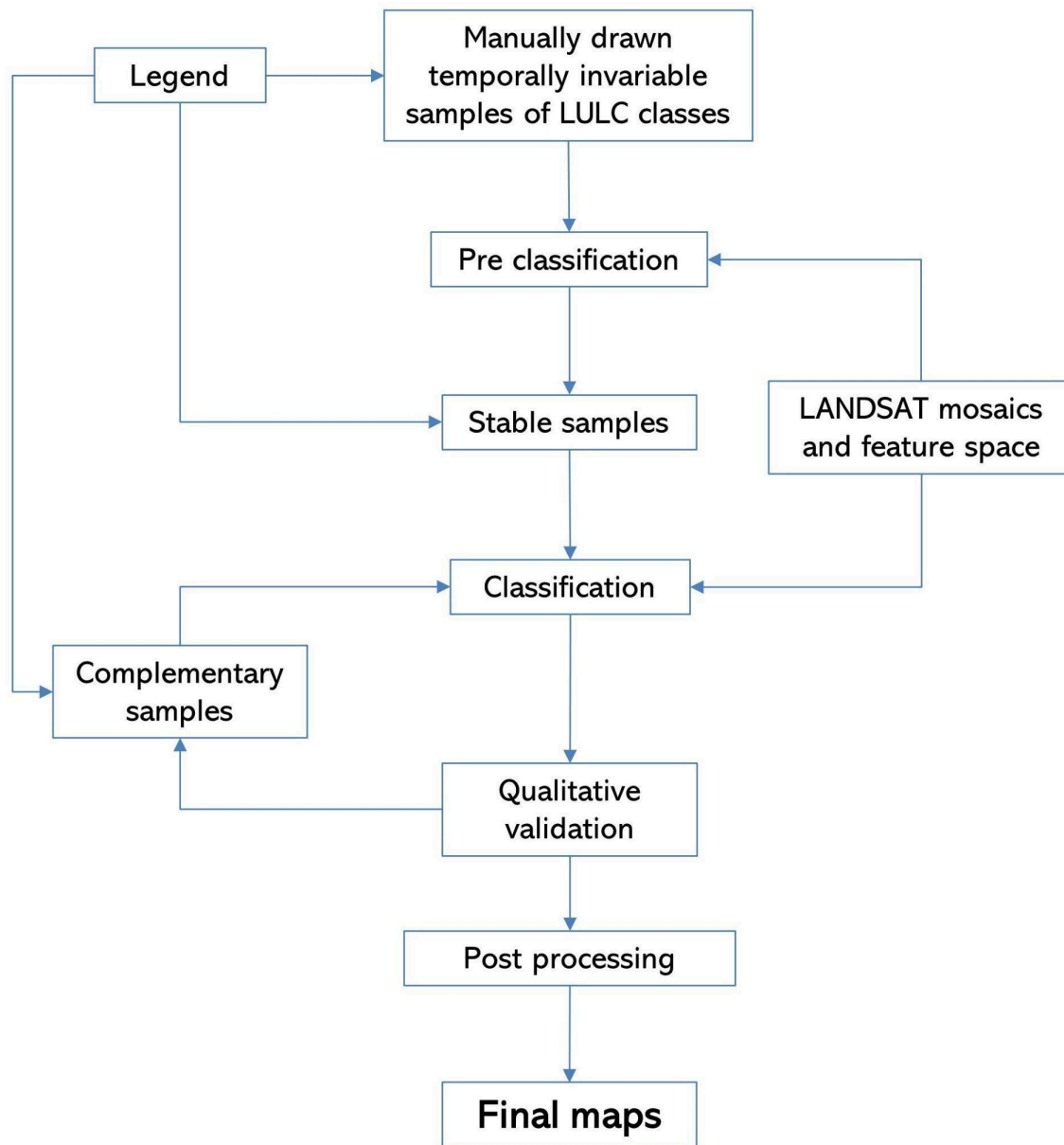

Figure 2. Methodological scheme of *Monte, Puna and High Andes* initiative Collection 2.

3. Remote Sensing Data

3.1. Landsat Collection

The imagery dataset used in *Monte, Puna and High Andes* initiative Collection 2 was obtained from the Landsat sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and the Operational Land Imager and Thermal Infrared Sensor (OLI-TIRS), on board of Landsat 5, Landsat 7 and Landsat 8, respectively. The Landsat imagery collections with 30-pixel resolution were accessible via Google Earth Engine, and sourced by NASA and USGS. *Monte, Puna and High Andes* initiative Collection 2 has used Tier 1 from USGS and surface reflectance (SR), which underwent radiometric calibration and orthorectification correction based on ground control points and a digital elevation model to account for pixel co-registration and correction of displacement errors. The scenes were used for covering the entire region, where each of them is totally or partially within the area. For each year we used images from the best Landsat mission available:

- 1985 to 1999 – Landsat 5
- 2000 to 2002 – Landsat 7
- 2003 to 2011 – Landsat 5
- 2012 – Landsat 7
- 2013 to 2024 – Landsat 8

3.2. Landsat Mosaics

Landsat cloud-free composites were obtained from images distributed throughout the whole year. The cloud/shadow removal script applied leverages Landsat's own quality assessment (QA) band and the GEE median reducer. When used, QA values can improve data integrity by indicating which pixels might be affected by artifacts or subject to cloud contamination (USGS, 2017). In simple terms, the algorithm rejects values that are too bright (e.g., clouds) or too dark (e.g., shadows) and picks the median pixel value in each band for a specific year.

3.3. Feature space

The feature space for classifying the classes of interest for the *Monte, Puna and High Andes* initiative Collection 2 comprises a set of 31 variables including the original Landsat bands, spectral indices, and fractional and textural information derived from these bands (Table 1). The variables of the feature space were the same for all subregions.

Reducers (i.e., temporal aggregation algorithms) were used to generate temporal features such as:

● Median - Median of the NDVI pixel values considering all the images of each year.

- Median_dry = median of the lower quartile of the NDVI pixel values.

- Median_wet = median of the higher quartile of the NDVI pixel values.

- Amplitude = amplitude of variation of the index over each year.

- stdDev = standard deviation of all pixel values over each year.

Table 1. List of spectral bands, fractions and indices comprising the feature space.

Variable	Description
Slope	Derived from the SRTM digital elevation model.
Green median texture	Average green texture extracted from the monthly mosaic.
GCVI median wet	Median of the wet period of the green chlorophyll vegetation index.
GCVI median	Annual median green chlorophyll vegetation index.
GCVI median dry	Median dry period green chlorophyll vegetation index.
Blue median	Annual median of the blue band.
EVI2 median	Annual median of the Enhanced Vegetation Index.
Green median	Annual median of the green band.
Red median	Annual red band median.
NIR median	Annual median of the near-infrared band.
SWIR1 median	Annual median shortwave infrared band 1.
SWIR2 median	Annual median shortwave infrared band 2.
GV medlan	Median of Green Vegetation.
GVS median	Median of the Green Vegetation Shade index.
NPV median	Median of Non-Photosynthetic Vegetation.
Soil median	Median Soil Fraction.
Shade median	Median shade fraction.
NDFI median	Annual median of Normalized Difference Fraction Index.
NDFI median wet	Median of the Wet period of NDFI.
NDVI median	Annual median of Normalized Difference Vegetation Index.
NDVI median dry	Median of Normalized Difference Vegetation Index dry period.
NDVI median wet	Median of Normalized Difference Vegetation Index wet period.
NDWI median	Annual median of Normalized Difference Water Index.
NDWI median wet	Median of Normalized Difference Water Index for wet season.
SAVI median	Median of Soil-adjusted Vegetation Index.
SETI median	Median of Savanna Ecosystem Fraction Index.
NDFI stdDev	Annual standard deviation of the Normalized Difference Fraction Index.
SEFI stdDev	Annual standard deviation of Savanna Ecosystem Fraction Index.
Soil stdDev	Annual standard deviation of Soil fraction.
NPV stdDev	Annual standard deviation of Non-Photosynthetic Vegetation.
NDWI amp	Annual amplitude of Normalized Difference Water Index.

4. Classification of LULC

The production of the Collection 2, with land use and land cover annual maps for the period 1985-2025 in each subregion, included a) manually drawn polygons of LULC classes temporally non-variable, based on the visual interpretation of annual Landsat composite images and temporal behavior of spectral indices, b) stable samples generated from LULC preliminary classifications, c) balancing of samples based on areal representativeness of each class, d) collection of complementary samples, e) annual LULC classifications, f) application of temporal and spatial post classification filters.

4.1. Legend

The supervised classification of the Landsat mosaics for *Monte, Puna and High Andes* initiative Collection 2 aimed to individualize a set of twelve land use and land cover classes (Table 2). The classes were carefully considered to ensure the accurate representation of the region's land cover types and involved the knowledge of the local experts for their selection. Urban areas were integrated via superposition, based on a product developed by a dedicated transversal working group.

Table 2. Land use and land cover (LULC) categories considered for the classification of Landsat mosaics for the *Monte, Puna and High Andes* initiative.

Legend	ID	Description
Closed forests	3	Areas with natural vegetation dominated by trees, with a coverage greater than 65%.
Open forests	4	Areas with natural vegetation dominated by trees, with a coverage between 20 and 65%.
Closed shrublands	66	Areas with natural vegetation dominated by shrubs, with a coverage greater than 65%.
Open shrublands	77	Areas with natural vegetation dominated by shrubs, with a coverage between 20 and 65%.
Sparse wooded shrublands	45	Areas with natural vegetation dominated by shrubs, with a coverage between 5 and 20%.
Grasslands	12	Areas with natural vegetation dominated by herbaceous plants, where woody plant coverage is less than 20% or nonexistent.
Flooded Grassland or Non-woody floodable vegetation	11	Areas with natural vegetation influenced by a shallow water table or dependent on flooding.

Mosaic of agriculture and pasture	21	Areas with anthropogenic vegetation composed of a heterogeneous association, varying over time and space, including agriculture, pastures, vegetables, and fruit trees. This cover is typically small-scale and commonly found in irrigated areas, although not exclusively so.
Woody Cultivated	9	Areas covered by woody (trees or shrubs) anthropogenic vegetation.
Non vegetated areas	25	Areas with vegetation coverage less than 5%.
Superficial water	33	Areas permanently covered by water. It includes streams, rivers, ponds, and natural or artificial lakes.
Ice and snow on surface	34	Areas with permanent and superficial snow or ice coverage. It does not include debris-covered glaciers.
Not observed	27	Attributed to missing values, shadows, or other errors that may arise during the classification process.

4.2. Classification algorithm, training samples and parameters

Classifications were performed yearly, using a *Random Forest* algorithm (Breiman, 2001) available in Google Earth Engine, running 70 iterations (random forest trees). Stable training samples for each region were defined following a strategy of using random pixels for which the land use and land cover remained the same for at least 35 of the 40 years, so-named "stable samples". The stable areas were identified through annual preliminary classifications trained with random pixels selected from manually drawn polygons. For this, false-color composites of the Landsat mosaics for all 40 years as backdrop and graphs with the temporal behavior of spectral indices per pixel were used to establish the LULC class.

4.3. Preliminary classification

From manually drawn polygons, a subset between 200 and 2000 pixels per class was randomly selected and used as training areas to classify each of the 40 years with the Random Forest algorithm, running 70 iterations. A total of 40 yearly preliminary classifications were obtained and the frequency with which a pixel was classified with the same LULC class was calculated to define the stable areas.

4.4. Sample outlier filtering

To identify the presence of outlier samples, we quantified the rarity of each one using the Isolation Forest algorithm on the attribute space defined by the 33 synthetic bands used to train samples. The algorithm progressively isolates observations in random decision trees, and those that require fewer splits to become isolated obtain a lower score, indicating greater rarity. Since the class distribution is not homogeneous across zones—particularly in areas where certain covers are very poorly represented—we grouped the data into three broad regions (metaregions) that combine several subregions: High Monte (6–8), Low Monte (9–11), and Puna and Andes (1–5). In this way, each sample is evaluated against a more representative regional distribution, avoiding biases related to the local scarcity of classes. Based on the scores, we applied a statistical rule to define the threshold that separates "typical" from "atypical" samples: using the skewness of the score distribution for each region–year–class combination, thresholds were set at 2%, 5%, and 10% of the samples for sets ranging from more symmetric (skewness > –0.5) to less symmetric (skewness < –1), respectively.

4.5. Stable samples

The identification of stable areas to extract random pixels or "stable samples" was based on a criterion of minimum frequency aiming to ensure their confidence for use as training areas. Each pixel should be classified with the same LULC class at least 35 times in the period 1985-2024 to be considered as stable, i.e. a pixel should remain with that class for a minimum of 35 years to be eligible as a stable sample. A layer of pixels with a stable classification along the 40 years was then generated by applying such a threshold. A minimum of 200 samples was used for rare classes that do not cover at least 10% of the region area and a maximum of 2000 samples for the most abundant classes.

4.6. Complementary samples

The need for complementary samples was evaluated by visual inspection and by comparing the output of the preliminary classification with both Landsat and high-resolution images available in GEE. Complementary sample collection was also done by drawing polygons within the Google Earth Engine Code Editor. The same concept of stable samples was applied, interpreting the false-color composites of the Landsat mosaics for all 40 years for drawing polygons. Based on the knowledge of each region, polygon samples from each class were collected and the number of random points in these polygons were defined to balance the samples.

4.7. Final classification

Final classification was performed for every year, with subregions merged into three metaregions, employing both stable and complementary samples. All years used the same subset of samples and each classifier was run for the same mosaic of the year that was classified.

5. Post-processing

Due to the pixel-based classification method and the long temporal series, a list of post-classification procedures was applied. The post-classification process includes the remapping of certain areas where obvious errors were identified, and the application of gap-filling, temporal and spatial filters.

5.1. Gap fill filter

A no-data values ("gaps") filter was applied. Because theoretically the no-data values are not allowed, it was replaced by the temporally nearest valid classification. In this procedure, if no "future" valid position was available, then the no-data value was replaced by its previous valid class. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain.

5.2 Remapping

In areas where obvious classification errors were identified, manual remapping was conducted. Additionally, pixels classified as water in areas where the slope exceeds 15° were reclassified as "Not observed," as this is a typical outcome of topographical shadows.

5.3. First spatial filter

The spatial filter overcomes the "salt and pepper effect" removing isolated pixels or groups of pixels. A mode filter was applied, by calculating the majority class in the direct neighbors of the focal pixel. It was based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that did not share connections to a predefined number of identical neighbors were considered isolated and subject to the mode filter. In this filter, at least 11 connected pixels were needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, and it was defined as 11 pixels (~1 ha).

5.4. Temporal filters

The temporal filter uses the subsequent years to replace pixels that have spurious transitions. The first process looks across a 3-year moving window to correct any value that changes in the middle year and changes back to the original class at the end year. This process was applied per class sequentially in the following order: [21, 77, 45, 9, 12, 11, 25, 66, 4, 3, 33]. The second and third processes are similar to the first one, but encompass 4- and 5-year moving windows, respectively, correcting all years falling in between. Temporal filters were not applied to the Water and Snow classes, considering their inherent interannual variability. The moving windows were not applied either over 1985 and 2024, for which end-year filters were applied. The end-year filters look for areas that were classified differently in 1985 in comparison to 1986 and 1987 and then correct the 1985 value. Also, these filters look for pixel values in 2024 that differ from its class in 2022 and 2023. In these cases, if the class value in 2024 is not the same as that in 2022 and 2023, it is then converted to the class present in those years.

5.5. Last spatial filter

After applying the filters detailed above, a spatial filter with the same parameters as the first one was applied to remove potential spatial artifacts produced by the temporal filters.

6. References

- Breiman, L. 2001. Random forests. Machine learning, v. 45, n. 1, p. 5-32.
- Burkart, R., N. Bárbaro, R. O. Sánchez, and D. A. Gómez. 1999. Ecorregiones de la Argentina, APN, PRODIA. Pp. 43.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., ... & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience, 51(11), 933-938.
- USGS Landsat (2017). USGS Landsat Collection 1 Level 1 Product Definition; USGS Landsat: Sioux Falls,SD,USA, 2017; Volume 26.