MAPBIOMAS

CHILE

ATBD R

Algorithm Theoretical Base Document & Results

MapBiomas "Handbook"

Octubre 2024

Executive Committee MapBiomas Chile

Jaime Hernández (U. de Chile)
Cristian Echeverria (U. de Concepción)
Adison Altamirano (U. de la Frontera)

General Coordination MapBiomas Chile

Jaime Hernández
Danniela Espinoza

Supervisory team

Alejandro Miranda
Rodrigo Fuentes
Jaime Hernández

Developers

Valentina González
Lissette Cortés
Carlos Hormazabal

Technical team

Diego Muñoz
Felipe Lepín
Valeria Castro

MapBiomas Brasil

Tasso Azevedo
Julia Shimbo
Marcos Rosa
Joao Siquiera
Mayra Milkovic

Table of Contents

EXECUTIVE SUMMARY

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) in Chile from 1999 to 2024 (Collection 2). The document presents a general description of the satellite image processing, the feature inputs and the process step by step applied to obtain the annual classifications.

1.2. Overview

Details about the classification methods are provided in order to assist the user to gain a general understanding of the technical considerations involved, the definition of intermediate inputs and outputs as well as scientific references supporting each decision. In addition, this document presents a historical context and background information, a general description of the satellite imagery datasets, feature inputs, and the accuracy assessment method applied. This information is intended to inform users about the strengths and limitations of MapBiomas Chile Collection 1 product.

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 LULC annual maps. All data, classification maps, codes, statistics, and further analyses are openly accessible through the MapBiomas Chile platform (http://chile.mapbiomas.org/). This is possible thanks to:

i. Google Earth Engine platform, which provides access to data, image processing, standard algorithms, and cloud computing facilities.
ii. Freely available Landsat time-series dataset.
iii. MapBiomas collaborative network of organizations and experts that share knowledge and mapping tools.

1.3. Region of Interest

Chile is a long, narrow country located along the western edge of South America, stretching over 4,300 kilometers (2,670 miles) from North (17°30' S) to South (55°59' S). Notably, a high rate (about 45%) of the species are endemic, which can be attributed to its isolation (Squeo et al. 2012). Chile boasts diverse geography, including the Atacama Desert, one of the driest places on Earth, in the North, the Andes Mountains running along its Eastern border, and the Pacific Ocean to the west. The Central valley, in the mid latitudes of the country, is an agricultural and heartland known for its fertile soils and Mediterranean climate. It is where most of Chile's population resides and is home to major cities like Santiago, the capital. In Southern Chile, the Lake District is known for its stunning landscapes of lakes, forests, and snow-capped mountains. Further South, the northern part of Chilean Patagonia is characterized by rugged terrain, fjords, and glaciers. Near the end of the continent, Patagonia becomes even more remote and wild. It features vast expanses of untouched wilderness, including the Southern Patagonian Ice Field, the third-largest ice mass in the world.

1.4. Key Science and Applications

The scientific applications derived from an annual time-series history of **LULC** maps in Chile include:

- Mapping and quantifying LULC transitions.
- Quantification of gross and net forest cover loss and gain.
- Monitoring agriculture, forest plantations and pasture expansion.
- Land planning and assessment.
- Assessing urban expansion and planning.
- Water surface and glaciers loss and gain.
- Protected areas management.
- Biodiversity assessment and distribution modelling.
- Climate change adaptation and mitigation strategies.

2. Overview and Background Information

2.1. Context and Key Information

Land cover is defined as the biological or physical cover observed on the earth's surface (FAO, 2016). Land cover data are used at both a global and local scale in the analysis of climate change, carbon stock assessment, monitoring of forestry and agricultural activities, disaster management, territorial planning, urban planning, biodiversity conservation and in many other public and private spheres. On the other hand, due to the transformations inherent to growth and development, land cover is highly dynamic over time and depends on a set of factors and human decisions that determine the use given to each portion of the territory. To meet the demands of sustainable development and international commitments, the country requires multiple sources of data for land cover monitoring that can satisfy current and future demands both domestically and internationally. However, to do this, a land cover monitoring system should have the following characteristics:

a. To be based on consistent, unique and systematically applied classification principles, replicable in space and time.
b. To be able to describe the full range of possibilities and their details through different hierarchical levels.
c. To be complete, covering the whole target territory.
d. To consider unique, mutually exclusive and unambiguous classes and categories.

2.2. Historical Perspective in Chile: Existent Maps and Mapping Initiatives

In Chile, there currently exists no comprehensive Land Cover monitoring system covering the national territory. The most analogous initiative, albeit only partially comparable, is the "Catastro and Evaluation of Native Vegetational Resources," established in the 1990s and presently managed by the National Forestry Corporation (CONAF). However, this initiative predominantly focuses on forested areas, often neglecting agricultural or urban land covers. Its inadequacy as a true monitoring system stems from its lack of systematic application over time. Furthermore, its complexity arises from the amalgamation of three distinct concepts: plant formations (utilizing a land occupation mapping method), biophysical coverage, and land use (socio-economic designation). Consequently, this framework remains largely incompatible with international land cover products, and that needs to be addressed in the near future.

Moreover, methodological alterations, such as changes in the minimum mappable area and revisions in the definition of certain forest types, further undermine its reliability as a monitoring tool. Compounding these challenges is the absence of standardized or routine updates at the national level, with variations observed across different regions. This persistent stagnation is highlighted by several scientific publications openly disputing the accuracy of data derived from the cadastre (Miranda et al., 2018).

While the Natural Resources Information Center (CIREN) serves as another repository of cartographic data, focusing on the management of natural resources and production in Chile, its primary emphasis lies on providing information pertinent to the economic utilization of land. This includes data on hydrogeological zones, well formations, hydraulic infrastructure, agroclimatic parameters, orchard mapping, land ownership records, and land use capacity. However, akin to the Cadastre, CIREN does not offer regular or systematic updates conducive to effective land cover monitoring.

The only preexistent “actual” land cover map was produced by Zhao et al. (2016). They generated a conclusive land cover map by integrating results from multi-seasonal mapping, utilizing primarily Landsat 8 imagery acquired predominantly in 2013 and 2014. Supplementary data sources included MODIS Enhanced Vegetation Index data, high-resolution imagery from Google Earth, and Shuttle Radar Topography Mission DEM data. The integrated map achieved an overall accuracy of 80% at level 1 and 73% at level 2, as confirmed by independent validation data. Accuracy assessments for seasonal map series yielded approximately 70% for each season, with significant enhancements observed through the integrated use of seasonal information. However, this product wasn't conceived as a monitoring system providing only a 2014 LULC map.

3. Algorithm Descriptions, Assumptions, and Approaches

The Collection 2 general methodological steps are presented in Figure 1.

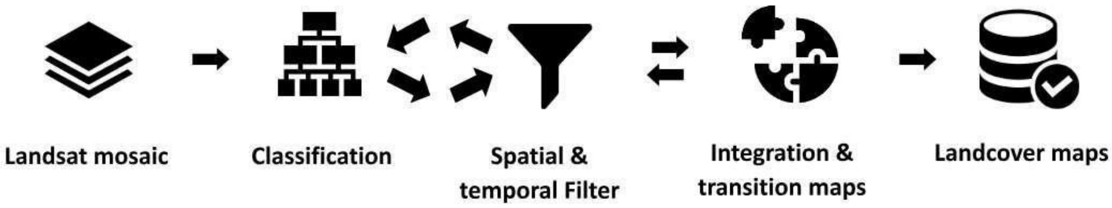

Figure 1: General methodological steps of Collection 2 of land cover and land use annual maps in Chile.

The first step was to generate annual Landsat mosaics comprising specific temporal windows to optimize the spectral contrast and better discriminate the LULC classes across the country. The second step was to derive all feature space attributes from the Landsat bands to train one random forest classifier (feature space definition) for each year (Breiman, 2001). In the third step, spatial-temporal filters were applied over the classified data for noise removal and temporal stabilization.

Subsequently, the filtered LULC maps of each working zone were hierarchically merged (integrated) based on a set of prevalence rules. Spatial and temporal filters, as well as post-processing remapping algorithms, were once again applied on the integrated maps to create the final Collection 2 product. For Collection 2, ecoregions were used to stratify the national territory. This allowed for the collection of sampling points that explicitly considered the biophysical differences of the country's various geographic zones. Ecoregions are large geographic units that group natural communities with distinctive species compositions, ecological dynamics, and shared environmental conditions (Olson et al., 2001). We consider eight ecoregions: Dry Puna, Atacama Desert, Andean Steppe, Chilean Matorral, Valdivian temperate forests, Magellanic subpolar forests, Patagonian steppe and Ice and Rocks (Figure 2).

Figure 2: Chilean Ecoregions.

3.1. Landsat Mosaics and Feature Space

Landsat cloud free composites obtained from images distributed along the whole year were considered (Figure 3) The cloud/shadow removal script takes advantage of the 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 conjunction, GEE can be instructed to pick the median pixel value in a stack of images. By doing so, the engine 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. Because for a large part of the national territory, approximately from the latitude of Santiago to the South, the availability of images from the Landsat program for years prior to 1998 is very low, it was decided to cover the period 1999 to 2024 for the second collection. For future collections, it is expected to extend the length of the time window used and to improve the quality of the mosaics.

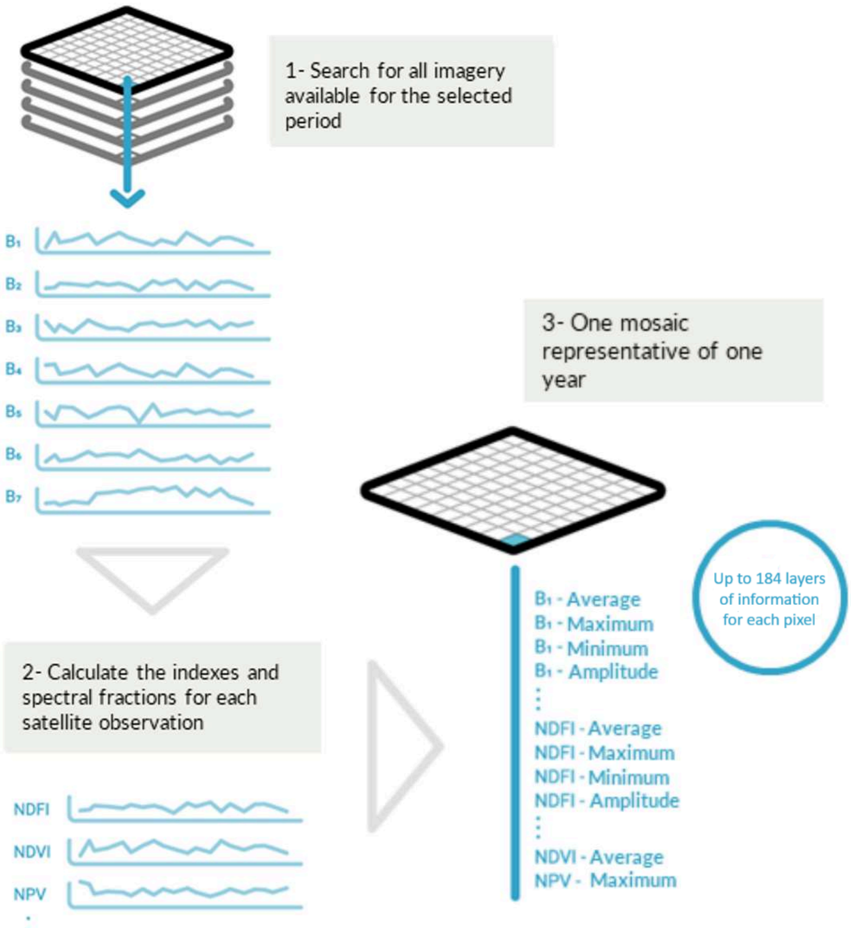

Figure 3: General workflow to build a year's mosaic.

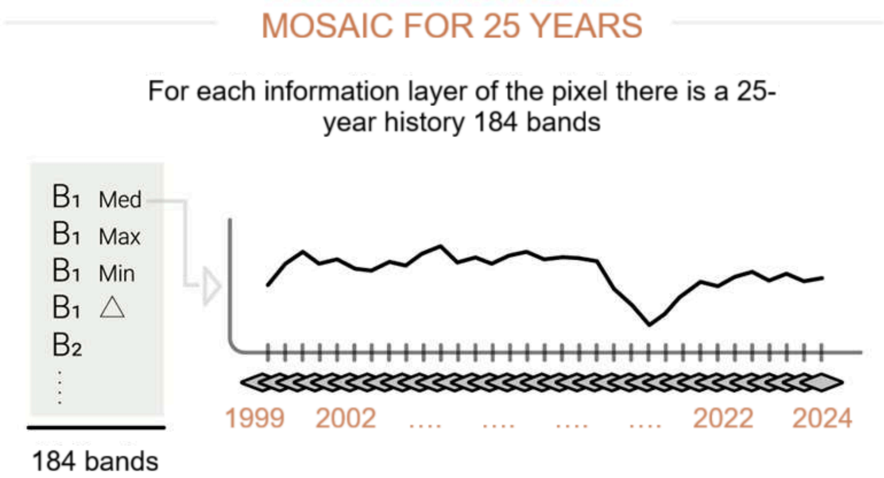

Figure 4: Each pixel of the mosaic contains 184 metrics or layers of information.

The feature space for LCLU classification is composed of 184 input variables per year, including the original Landsat bands and fractional and textural information derived from those bands (Table 1, Figure 4). Table 1 presents the formula or the description to obtain these variables, as well as highlights in grey all the bands, indices, and fractions available in the feature space. In addition, statistical reducers were used to generate temporal features such as:

- Median: median of the pixel values within the defined stack of images.
- Median_dry: median of the quartile of the lowest pixel NDVI values.
- Median_wet: median of the quartile of the highest pixel NDVI values.
- Amplitude: amplitude of variation of the index considering all the year's images.
- stdDev: stdDev of the pixel values within the defined stack of images.
- Min: the lower annual value of the pixels of each band.
- Max: the higher annual value of the pixels of each band.

Some classes, due to their specific methodologies, used a subselection of these bands for classification.

Table 1: List of variables used as predictors for LULC classification.

	Variable	**Descripción**	**median**	**dry**	**wet**	**min**	**max**	**amp**	**stdDesv**
Bands	*blue*	*B1(L5 L7); B2 (L8)*							
	red	*B2(L5 L7); B3 (L8)*							
	green	*B3(L5 L7); B4 (L8)*							
	nir	*B4(L5 L7); B5 (L8)*							
	swir1	*B5(L5 L7); B6 (L8)*							
	swir2	*B7(L5); B8 (L7); B7 (L8)*							
Fractions	*gv*	fractional abundance of green vegetation within the pixel (0-1)							
	npv	fractional abundance of non-photosynthetic vegetation within the pixel (0-1)							
	soil	fractional abundance of soil within the pixel (0-1)							
	cloud	fractional abundance of cloud within the pixel (0-1)							
	shade	abundancia fraccional de sombras (0-1)							
Fraction Indexs	*fns*	Fracción no sombreada (1-shade)							
	gvs	Green fraction in shadow [gv / (gv + npv + soil + cloud)]							
	ndfi	vegetation structure index [(gvs - (npv + soil))/(gvs + (npv + soil))]							
	sefi	Enhanced soil fraction index [(gv+npv_s - soil)/(gv+npv_s + soil)]							
	wefi	Enhanced wet fraction index [((gv+npv) - (soil+shade)) / ((gv+npv) + (soil+shade))]							
Bands Index	*cai*	Cellulose Absorption Index [swir2/swir1]							
	evi2	Enchanced Vegetation Index [(2.5 $\times((NIR - RED)$							

		$(NIR + 2.4 \times RED + 1))]$							
	gcvi	Green chlorophyll vegetation index $[NIR * (GREEN - 1)]$							
	hallcover	Estimation of vegetation cover [(-red*0.017 - nir*0.007 - swir2*0.079+ 5.22)]							
	hallheigth	Estimation of vegetation height cove[a×Red+b×NIR+c×SWIR2+d]							
	ndvi	Normalized difference vegetation index [(nir - red)/(nir + red)]							
	ndwi	Normalized difference wet index [(nir - swir1)/(nir +swir1)]							
	pri	Photoquimical reflectance index $[(BLUE - GREEN)/(BLUE + GREEN)]$							
	savi	Soil-adjusted vegetation index $[(1 + 0.5) \times (NIR - RED)/(NIR + RED + 0.5)]$							
	ndmi	Normalized Difference Moisture Index $[(NIR - (SWIR1 - SWIR2))/(NIR + (SWIR1 - SWIR2))]$							
	ndbi	Normalized Difference Built-Up Index $[(SWIR1 - NIR)/(SWIR1 + NIR)]$							
	mbi	Modified Bare Soil Index $[(SWIR1 - SWIR2 - NIR)/(SWIR1 + SWIR2 + NIR) + 0{,}5]$							
	ndsi	Normalized Difference Snow Index [(GREEN-SWIR1)/(GREEN+SWIR1)]							
Otras variables	*slope*	Extracted from ALOS DSM: Global 30m with Terrain Analysis							
	Elevation	Extracted from SRTM 30m							
	Aspect	Extracted from SRTM 30m with Tearrain Analysis							
	TPI	elevation - mean elevation (from square neighborhood of 180 meters)							
	green_median_texture	Median of the spatial pattern of the green band in a defined neighborhood.							

3.3. LULC scheme and Classification

3.3.1. Legend

The MapBiomas classification scheme is a hierarchical system comprising four categorical levels (Table 2). For the Chilean Collection 2, we considered the first three levels. At Level 1, there are six classes: 1) Forest formation, 2) Natural non-forest formation, 3) Farming and silviculture, 4) Non-vegetated area, 5) Water bodies, and 6) Not observed.

Level 2 comprises 13 classes distributed across the six Level 1 categories. Among them, class 1.1 (Forest) is the only one subdivided into a third categorical level, which includes three subclasses that differentiate types of forest formations.

Table 2: Classes of land cover and land use of MapBiomas Collection 2 in Chile.

Class Level 1	Class Level 2	Class Level 3	Description	FAO Class*
1. Forest formation	1.1. Forest		Vegetation formations dominated by native tree species over 2 meters tall. This includes primary, secondary (regrowth), and dwarf native forests. Forest is defined as having a minimum canopy cover of 25%.	FEP, FEM, FEY, FDP, FDM, FDY
	1.1. Forest	1.1.1 Primary Forest	Forests with diverse vertical and horizontal structure. They contain large individuals, according to their ecoregion, that remain stable during the study period. In the imagery, this is reflected as textural diversity due to the influence of structural heterogeneity and the presence of shadows and gaps.	FEP, FDP, FSP
		1.1.2 Secondary Forest	Forests with homogeneous vertical and horizontal structure. They contain individuals of similar size and age, which may or may not remain stable during the study period. In the imagery, this is reflected as lower textural diversity compared to primary forests. In general, they are represented by a single dominant species.	FEM, FEY, FDM, FDY, FSM, FSY
		1.1.3 Dwarf forest	Forest ecosystems characterized by the presence of low-stature trees and very dense forest. In general, they show interlaced crowns and are dominated by a single species. These forest communities, mainly of the genus Nothofagus, develop under restrictive environmental conditions such as poor soils, wind exposure, high radiation, low water availability, or high elevations, which limit the vertical growth of the species. In the imagery, this is reflected as lower textural diversity compared to secondary forests.	N/A

2. Natural non forest formation	2.1. Wetland	Ecosystems associated with water-saturated substrates, either temporarily or permanently. For delimitation purposes, the presence of hydrophilic vegetation is considered. They are generally found in topographic depressions or poorly drained soils. They may correspond to marshes, swamps, high-Andean wetlands (bofedales), and peatlands of natural or artificial origin, with stagnant or flowing water, fresh or saline.	OM, WW
	2.2. Grassland	Areas dominated by natural herbaceous vegetation.	OG
	2.3. Steppe	This class is subdivided into two types based on their geographic distribution: (1) High Andean steppe: Areas dominated by natural formations of herbaceous vegetation of the pajonal type. In this class, the presence of low woody species of the tolar type with variable cover is allowed. This class is mainly found above 3,200 m a.s.l., which may vary according to latitude. (2) Patagonian steppe: Areas dominated by natural formations of herbaceous vegetation of the pajonal type. In this class, the presence of low woody species with variable cover is allowed.	N/A
	2.4. Shrubland	Vegetation formations dominated by shrub or low-stature tree species (less than 2 meters tall), whether native or exotic, with a minimum canopy cover of 25%. Shrublands may have different densities and can mix with other classes.	WS, WG
	2.5. Rocky Outcrop	Areas with exposed rocks, generally found in steep slope zones or severely eroded areas.	OX
3. Farming and silviculture	3.1. Silviculture	Areas covered by exotic species of silvicultural interest, mainly from the genera Pinus and Eucalyptus. This class includes harvested stands and areas invaded by natural regeneration of these species. No distinction is made between different regimes of ownership or land use (silvicultural, agricultural, or other).	FPB, FPC, FPM
	3.2. Agriculture	Areas covered by annual, temporary, and/or perennial crops, either active or in fallow. Orchards, fruit trees, vineyards, and greenhouses are included.	OCA, OCP, OCM, OF
	3.3. Pasture	Areas covered by herbaceous vegetation, associated with livestock activity.	OP
4. Non-vegetated area	4.1. Infrastructure	Urban and industrial areas, main roads, and other artificial structures such as seaports and airports. Mining and solar energy areas are not included.	OB
	4.2. Beach, Dune and Sand Spot	Areas covered by different types of sand with little or no vegetation.	OX

	4.3. Salt Flat	Areas characterized by large expanses of flat land covered by a layer of salt or mineral salts. Water bodies within salt flats are classified as water.	OX
	4.4. Other non-vegetated area	Areas characterized by large expanses of flat land covered by a layer of salt or mineral salts. Water bodies within salt flats are classified as water.	OX
5. Water bodies	5.1. River, lake or ocean	Areas permanently covered by water, of natural or anthropic origin.	IRP, IRS, IL, ID, IP, XO
	5.2. Ice and snow	Areas with permanent and superficial snow or ice cover, generally located in the Andean peaks. Debris-covered or rock glaciers are not included.	N/A
6. Not observed	6. Not observed	Unclassified areas due to the presence of permanent clouds and shadows, atmospheric noise, low quality of satellite imagery, or lack of information.	90

3.3.2. Samples and stable Samples

The first step of the classification process was created in each of the eight ecoregions a set of sampling points for training the classification algorithm reaching a total of 311.442 points (see table 3 for details) obtained using random pixels selected from manually drawn polygons made by photo interpretation for all the classes for each ecoregion. For this, false-color composites of the Landsat mosaics for all the 25 years as backdrop and graphs with the temporal behavior of spectral indices per pixel were used to establish the LULC class. 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.

Table 3: Sampling points per class and ecoregion.

CLASSES		ECOREGIONS							
Name	Id	Dry Puna	Atacama Desert	Chilan Matorral	Andean Steppe	Valdivian Temperate Forest	Patagonian Steppe	Magellanic Subpolar Forest	Ice and Rock
Primary Forest	59	4723	5000	1628	64	11091	4977	1468	735
Secondary Forest	60	0	0	966	49	4082	316	1023	181
Dwarf forest	67	0	0	0	0	1122	263	502	53
Shrubland	66	231	608	1665	240	3519	205	270	
Wetland	11	5000	850	3783	2945	13526		5000	
Silviculture	9	0	0	4528		14777		222	
River, Lake or Ocean	33		410	5532	712	20590	5000	5000	157
Beach, Dune and Sand	23		1955	5585		5597	1059	380	
Pasture	15			692		2693	176	137	
Agriculture	18	129	192	3304	258	2223	137		
Grassland	12	0		407	542	1027			
Steppe	63	2623	817			0	894	948	
Non-vegetated area	25	5000	2000	5000	5000	15000			
Ice and Snow	34	5000			5000	15000	5000	5000	5000
Rocky Outcrop	29	974	886	6468	5000	3066	212	688	
Infrastructure	24	329	5000	9000	558	19590	850	5000	
Salar	61	5000	5994	5000	1039				
Number classes		14	14	15	13	16	12	13	5
Total Points	311.442								

These were the starting points for building the preliminary classification.

3.3.3. Classification

Digital classification was performed in each working zone, year by year, using a Random Forest algorithm (Breiman, 2001) available in Google Earth Engine, running 100 iterations (random forest trees). Final classification was performed for all working zones and years with stable and complementary samples (Figure 5). All years used the same subset of samples and it was trained in the same mosaic of the year that was classified. The class "Infrastructure" was left out of this classification step (for details, see Appendix 1). It was classified in a separate process and integrated in the post-processing step. In addition, due to mosaic issues, the ice and snow class was also processed independently and overridden in the final classification (for details, see 3.4.7).

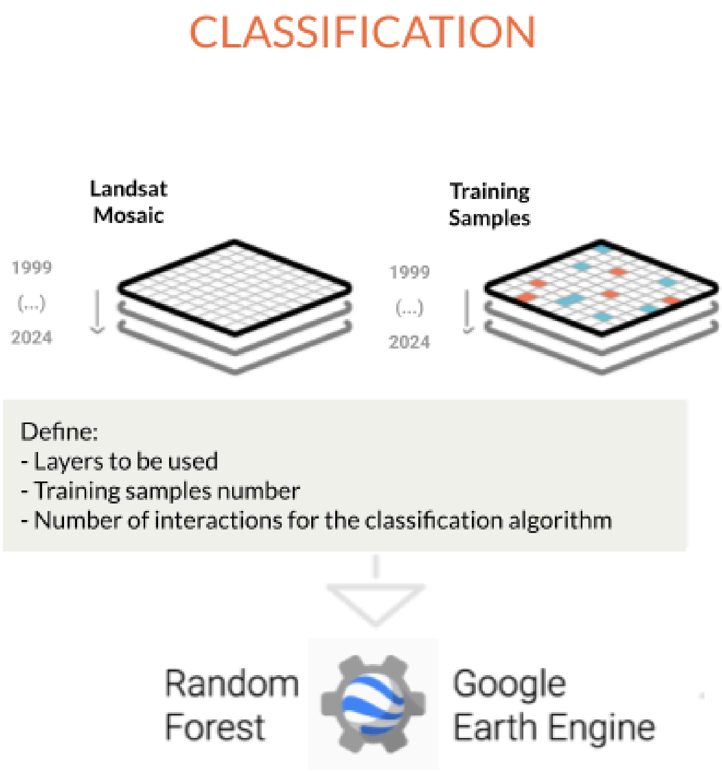

Figure 5: Representation of Landsat mosaic and training samples with two classes (red and blue}.

3.4. Post-Processing

The results of the final classification in each working zone were improved by applying a set of filters. In the first place, a serie of temporal filters were done with the aim to generate a more stable classification pattern over time, avoiding unexpected class variation during consecutive years or a short period of time. In second place, filters to correct missing data were applied in order to recover areas without information with pixels from previous or subsequent years. In third place, spatial filters were applied to reduce the "salt-and-pepper" effect.

3.4.1. Temporal filters

The temporal filters are applied over sequential annual classifications using a unidirectional moving window of three, four, or five years, in order to identify temporally non-permitted transitions (Figure 6). These filters evaluate the temporal consistency of class assignments for each pixel, correcting inconsistencies that are interpreted as classification errors.

The temporal filter inspects the central position of the moving window, and if a pattern inconsistent with temporal stability is detected, the central pixel(s) are reclassified to match the dominant neighboring class.

Temporal filters were divided into two main categories according to their function: regular and extremes (beginning and ending).

(i) Regular filters (3-, 4-, and 5-year windows)

Regular filters are applied to the central part of the time series and aim to correct short-term temporal inconsistencies.

- The **3-year filter** (ternary) operates on three consecutive years (*t−1, t, t+1*). If the classes of the first and third years are identical and the central year differs, the central pixel (*t*) is reclassified to match its neighbors.

- The **4-year filter** operates on a four-year window (*t−1, t, t+1, t+2*). If the first and fourth years share the same class and the two central years are identical but different from the outer years, then both central pixels (*t* and *t+1*) are reclassified to match the stable outer class.

- The **5-year filter** extends this logic to a five-year window (*t−2, t−1, t, t+1, t+2*). When the three central years share the same class that differs from the outer years, and those outer years are identical to each other, all three central pixels are reclassified to match the temporally stable class at the extremes.

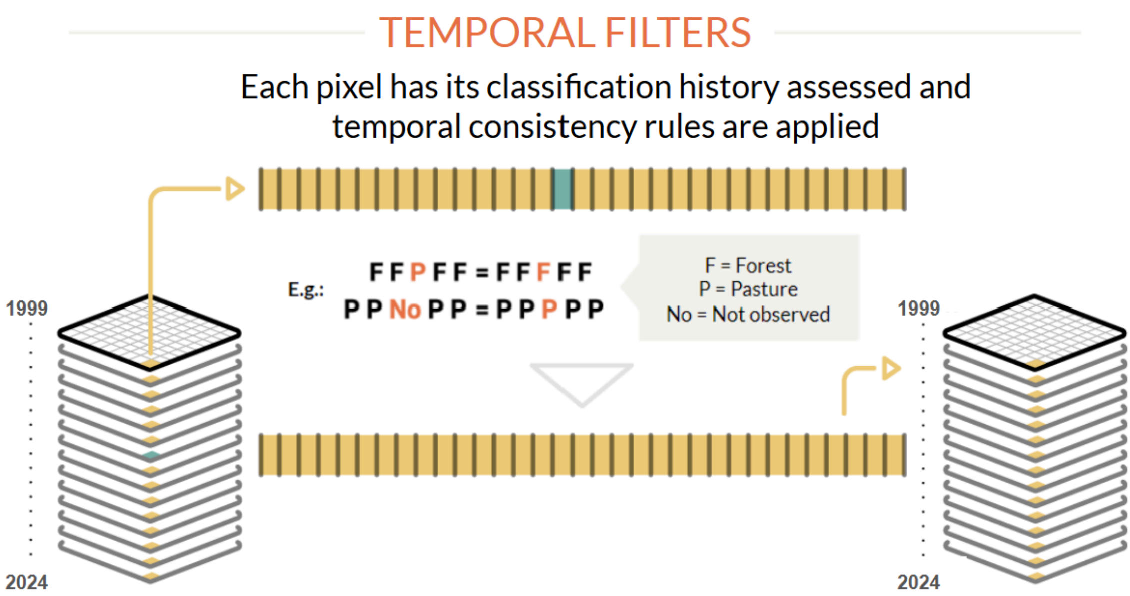

Figure 6: Temporal filter concept.

(ii) Extreme filters (beginning and ending)

Extreme filters are applied exclusively to the temporal boundaries of the time series — that is, the first and last years. For the first year, each pixel is compared with its class in the following year (x+1); for the last year, the comparison is made with the preceding year (x–1).

The decision rule states that if a pixel's class in year x differs from that of its adjacent year, belonging to different thematic groups (i.e., anthropic vs. natural), the pixel in year x is reclassified to match the class of its neighboring year (Figure 7).

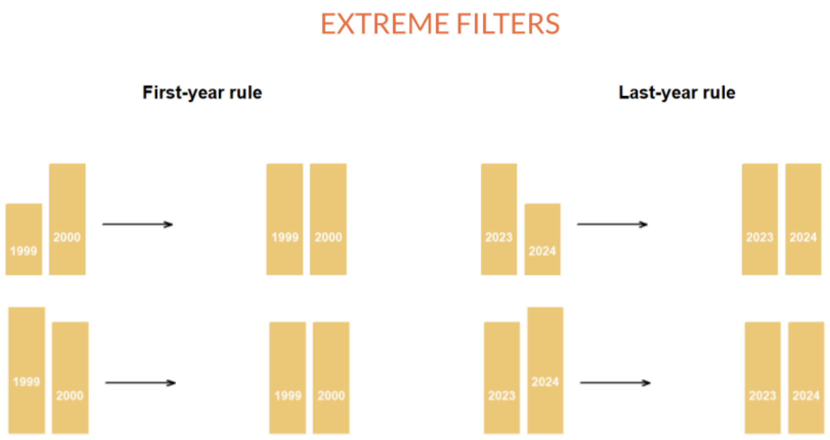

Figure 7: Extreme filter concept

3.4.2. Gap fill filters.

A filter was applied to fill *no-data* pixels, or "gaps," within the temporal series. Since, in principle, *no-data* values are not permitted, they were replaced with the temporally nearest valid classification.

In this procedure, if no valid *future* observation was available, the *no-data* pixel was replaced by its most recent valid class. Consequently, gaps should only persist in cases where a given pixel remained classified as *no-data* throughout the entire temporal domain.

3.4.3. Spatial filters

The spatial filter avoids unwanted modifications to the edges of the pixel groups, a spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value (Figure 8). Thus, only pixels that did not share connections to a predefined number of identical neighbors were considered isolated. In this filter, at least six 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).

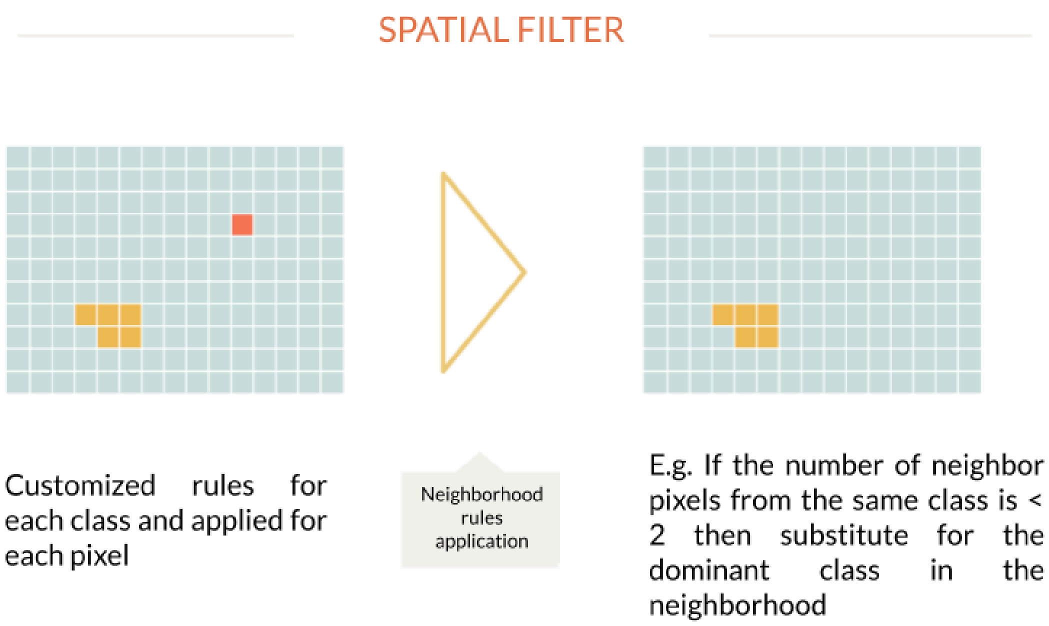

Figure 8: Spatial filter concept.

3.4.4. Additional filters and remapping.

In addition to the general filters, a set of additional rule-based filters was applied to address specific classification inconsistencies observed in particular coverage types. These filters consist of targeted remapping procedures designed to improve thematic coherence and ensure a more realistic temporal evolution of the classes.

- **Post-fire plantation remapping**

One of these procedures corresponds to the post-fire plantation remapping, which adjusts the classification of pixels affected by fire events, ensuring that areas reforested or reestablished as *Plantations* after a burn are consistently represented through time. This correction is based on a set of predefined logical rules that evaluate the class sequence before and after the detected fire event, remapping affected pixels to the appropriate stable class when necessary.

Such additional rule-based corrections were applied selectively and only in regions or land-cover types where systematic misclassifications were identified, improving the overall temporal and thematic consistency of the MapBiomas Chile Collection 2 classification series

- **Frequency filters.**

The frequency filter was applied after the temporal filtering stage and was specifically designed for some classes (i.e. Silviculture, Agriculture). The purpose of this filter is to reinforce temporal stability in land-cover types that are expected to remain consistent over time, correcting residual fluctuations that may persist after the application of the temporal filters.

This filter operates by analyzing the frequency of class occurrence for each pixel within the final years of the time series. For every pixel, the filter identifies the first year from which a given class becomes dominant or recurrent in the subsequent years. Once this year is detected, all following years are reassigned to the most frequent class within that period. In practice, this approach replaces sporadic or inconsistent class changes with the modal (most frequent) class observed in the stable portion of the series (Figure 9).

By consolidating the most recent and stable periods of classification, the frequency filter enhances the temporal coherence of the dataset and minimizes isolated classification noise.The result is a more stable and realistic representation of these coverage types across the most recent years of the MapBiomas Chile time series.

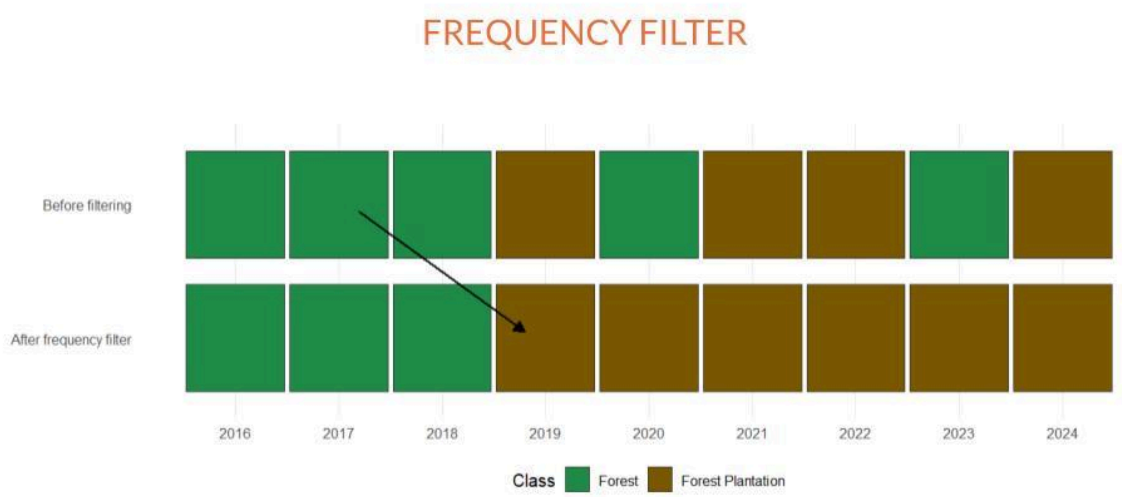

Figure 9: Frequency filter concept

- **Mode Filter**

The temporal mode filter is applied to improve the year-to-year consistency of specific land-cover classes and to reduce residual high-frequency changes in the series after applying other temporal filters. It relies on a multiband classification, a target class (e.g., Wetlands), a set of alternative classes for controlled remapping, and a spatial mask that limits the intervention to areas where inconsistency is detected. For each year, the image is remapped to the selected alternative classes. The correction is applied only where the pixel belongs to the target class in that year. This method stabilizes false changes induced by classification noise, shadows, or spectral confusion, while the remapping constrains substitutions to ecologically coherent classes.

- **Morphological filter.**

The morphological filter is a spatial post-processing operation applied to the classification maps to improve the geometric consistency of the forest plantation class. It aims to remove small holes, isolated pixels, and internal irregularities that appear within plantation patches, ensuring a more homogeneous and realistic spatial representation. The procedure is based on a morphological closing operation, a common technique in image processing used to fill small internal gaps while preserving the general shape of the features.

The annual classification is first converted into a binary mask identifying all pixels corresponding to plantations, over this binary mask, the closing operation is performed by applying a dilation followed by an erosion. The dilation temporarily expands the extent of the plantation areas, allowing small gaps between neighboring pixels to be filled. Subsequently, the erosion step contracts the expanded areas, restoring the external shape of the polygons while maintaining the newly filled internal pixels. Both operations are applied using a diamond-shaped structuring element with a radius of one pixel, equivalent to 30 meters at the working resolution of the MapBiomas Chile classification.

The effect of this operation is subtle but significant: it eliminates micro-fragmentation within homogeneous plantation blocks and reduces pixel-level noise that can occur along internal boundaries. Importantly, the morphological filter acts only within the existing plantation mask and does not expand the plantations beyond their local boundaries, preserving the integrity of adjacent classes and the original spatial structure of the map.
This morphological approach provides a controlled, geometry-based correction that improves the internal coherence and visual quality of plantation areas, particularly in regions where topography or image mosaicking may introduce small classification discontinuities. The resulting product achieves greater cartographic smoothness without compromising thematic precision.

3.4.5. Ice and snow classification

Because of gaps in the mosaics used for Landcover classification, the ice and snow class (class 34) was developed separately. To accomplish this, we used the Collection 1 (COL1) classification for the period 2002–2020, and produced new classifications for the years 1999, 2000, 2001, 2021, 2022, 2023, and 2024.

For each year, a mosaic was generated using all available images from that year, after filtering out pixels affected by clouds, cloud shadows, and cirrus. Subsequently, a classification was generated by applying a decision tree, based on a minimum NDSI threshold and NIR and red band values representative of the dry season, defined as the first quartile (25th percentile) of NDSI values. Pixels meeting these criteria were classified as ice and snow (Figure 10).

Finally, a temporal filter with 3-, 4-, and 5-year windows—as described for the main land cover classification— was applied to enhance temporal consistency.

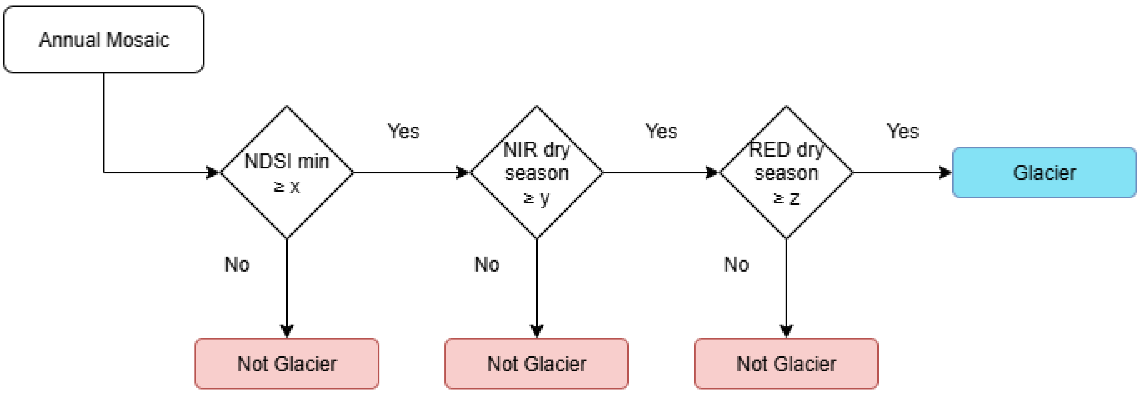

Figure 10: Decision tree applied

3.5. Integration and Transitions Maps

3.5.1. Integration

The integration step involved a process that compiled/overlapped classifications prepared by the different teams for the four working zones. This process resulted in a raster map with all classes from 1999 to 2024. Maps of each working zone were integrated on a pixel-by-pixel basis through the hierarchical overlap of each mapped class, following prevalence rules defined by experts.

The "Infrastructure" (id 24) class was classified independently. This was necessary due to frequent confusion with non-vegetated use classes such as "other areas without vegetation" (id 25), "rocky soils" (29) and "sand, beach and dunes" (id 23). Before the overlay was performed, the infrastructure map was further processed. The vector map of main regional and national roads of Chile[1] was converted to raster and joined year by year with the infrastructure classification. Finally, the infrastructure layer was integrated into the general map following a series of rules that were applied pixel by pixel to different areas of the country. A similar process occurred with Ice and Snow class (id 34), which was also generated independently. Once it is fully processed, it is superimposed on the general map.

[1] https://mapas.mop.gov.cl/red-vial/Red_Vial_Chile.zip

3.5.2. Transition maps

The pixel-by-pixel class differences between any two maps are now computed based on user-defined temporal ranges (Figure 11). Unlike the previous approach — which produced predefined transition maps for fixed intervals such as consecutive years, five-year, and ten-year periods — the current methodology allows the user to freely select the start and end years for the analysis. This flexible configuration enables the generation of customized transition maps that better reflect the temporal scale and dynamics of interest.

The resulting transition maps represent land-cover and land-use changes (LULC) between the selected years, highlighting the main processes of transformation in the landscape, including:

- Transitions from agricultural classes or non-vegetated areas to forest cover or non-forest natural areas.
- Areas corresponding to primary forests that, due to natural or anthropic disturbances, have been degraded and converted into secondary forests.
- Transitions that add water surface.
- Transitions that reduce water surface
- Transitions with gain in forest plantation.
- Transitions from forest cover or non-forest natural areas to agricultural or non-vegetated areas.
- Areas under a rotational system alternating from agriculture to pasture from year to year.
- Areas under a rotational system alternating from pasture to agriculture from year to year.
- Areas without transition or transitions involving unobserved areas or transitions between classes within level 1 of the legend.

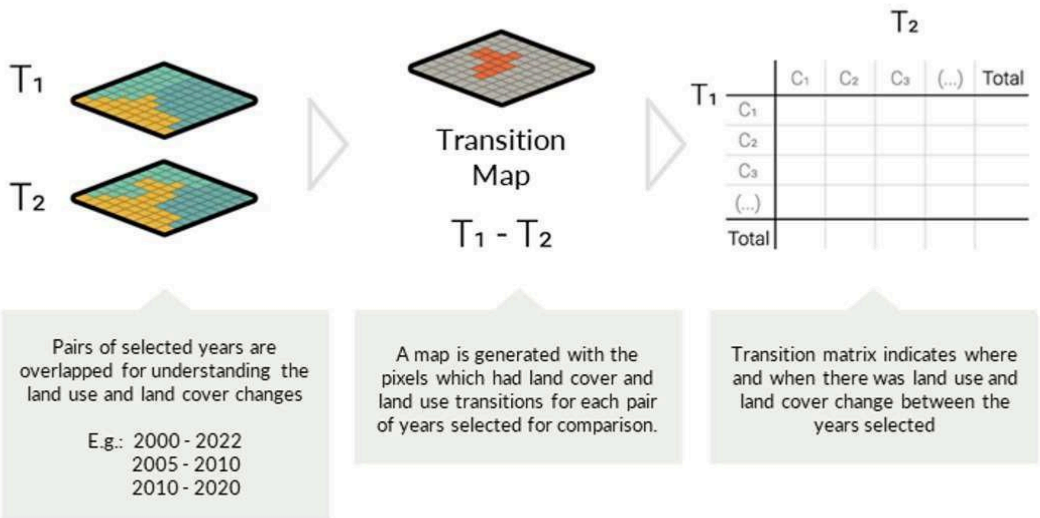

Figure 11: Transition map concept.

Additional spatial filters were applied to the transition maps to remove isolated pixels or linear artifacts located along class boundaries, which may arise during the creation of pixel-based transition layers. The main rules for this filter include: (i) the elimination of pixels with only one neighboring pixel belonging to the same transition class, and (ii) the removal of linear features or small groups of up to five pixels with one or two neighboring pixels of the same class. These filters improve spatial coherence and reduce noise in the final transition products.

3.6. Post-processing

3.6.1. Slope-shadow filters

Due to the steep topography of the Chilean Andes, deep shadows in the mosaic led to many pixels being erroneously classified as “River, lake or ocean” (class 33). To correct this, we first built a mask of pixels whose slope (from SRTM) exceeds 15°. Within this mask, we applied (i) a spatial filter using a 4×4 neighborhood: when a pixel met the slope condition and was labeled 33, its value was remapped to the neighborhood mode. Next, (ii) we applied a temporal mode filter under the same conditions, excluding class 33 as a possible target class (i.e., the mode was computed among the other classes). This approach reduced false water detections in shadowed areas.

3.6.2. Rules-based remapping

The Steppe class (id 63) in northern Chile was restricted to areas with elevations greater or equal to 3,200 m a.s.l., in accordance with the published literature. All vegetation below this threshold was reclassified as Grassland (id 12). To implement this rule, we built an elevation mask from the SRTM DEM. The restriction was applied in all Chile, except in the territories corresponding to the Aysén and Magallanes regions.

For the Salt flat class (id 61), an additional remapping was applied due to the frequent confusion with Ice and Snow (class 34), arising from their spectral similarity in the feature space used. We employed the Chilean Salt Flat Layer[1] as a mask; within its extent, all pixels originally classified as id 34 were reclassified to id 61.

3.6.3. Additional post-processing

Some additional remapping and corrections were applied to local errors such misclassifications or problems derived from the quality of the data; all these corrections were included in the scripts of the process.

3.7. Statistics

Zonal statistics of the mapped classes were calculated for spatial units, such as regions. All the spatial units, and the information could be downloaded in the platform of the initiative

[1] https://www.sernageomin.cl/plataforma-publica-de-salares-y-litio/

4. Validation Overall Analysis

Validation was performed for the classifications of the year 2024. We used two databases of 1654 randomly generated points each year with approximately 80 to 100 points per class (Figure 12). After a proper training step, two independent teams classified each point into one the classes via photo interpretation on the Google Earth Pro platform using the available historical images. If images were not found for the year, nearby years before and after the target year were used. The classification was carried out in two rounds. In the first round, each team made up of two members classified a set of points independently. The points where the classifiers' decision did not coincide were sent to the second round, where a third independent subject reclassified each point with a discrepancy. Points where no agreement was reached were eliminated, points for which there were no images close to the target year were also eliminated. With this information, the confusion matrices were generated and the classification accuracies were calculated. All these processes were carried out in Google Earth Engine and R-project. Table 4 presents the final confusion matrix for level 1 classes.

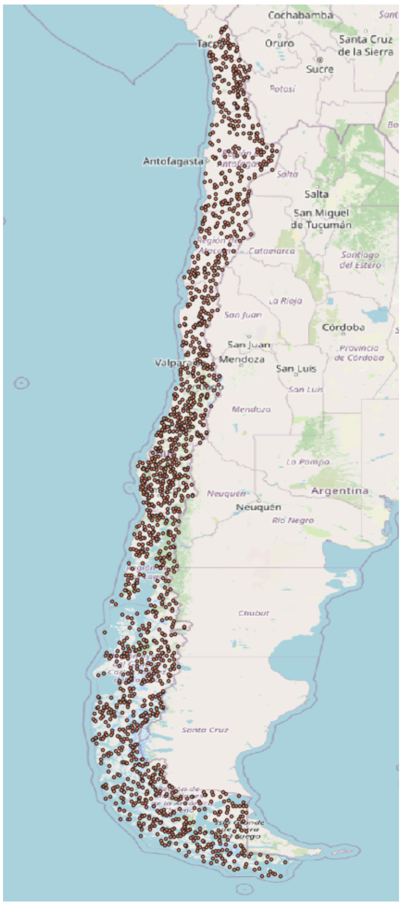

Figure 12: Distribution of validation points.

Table 4: Final confusion matrix for level 1 classes.

	Forest Formation	Natural non forest formation	Farming and silviculture	Non-vegetated area	Water bodies	Producer Accuracy	User Accuracy
Forest Formation	258	32	17	1	4	0,82	0.84
Natural non forest formation	50	421	30	89	8	0,81	0.7
Farming and silviculture	10	16	216	2	0	0,86	0.91
Non-vegetated area	2	42	3	188	0	0,57	0.83
Water bodies	6	32	0	20	206	0,93	0.80

5. Map Collections and Analysis

5.1. Collection 2

The MapBiomas Chile Collection 2, including all land cover maps between 1999 and 2024, transitions and methodological documents are available at http://chile.mapbiomas.org

It is important to note that MapBiomas Collection maps are an evolving product and other Collections with improvements will be available in the future. When using the data, be sure to always use the latest version available. MapBiomas maps are best used at scales up to 1:100,000. Although it is possible to view them at a 1:50,000 scale, we do not recommend using them at this scale. Figure 13 shows the mural printable land cover map for 2024.

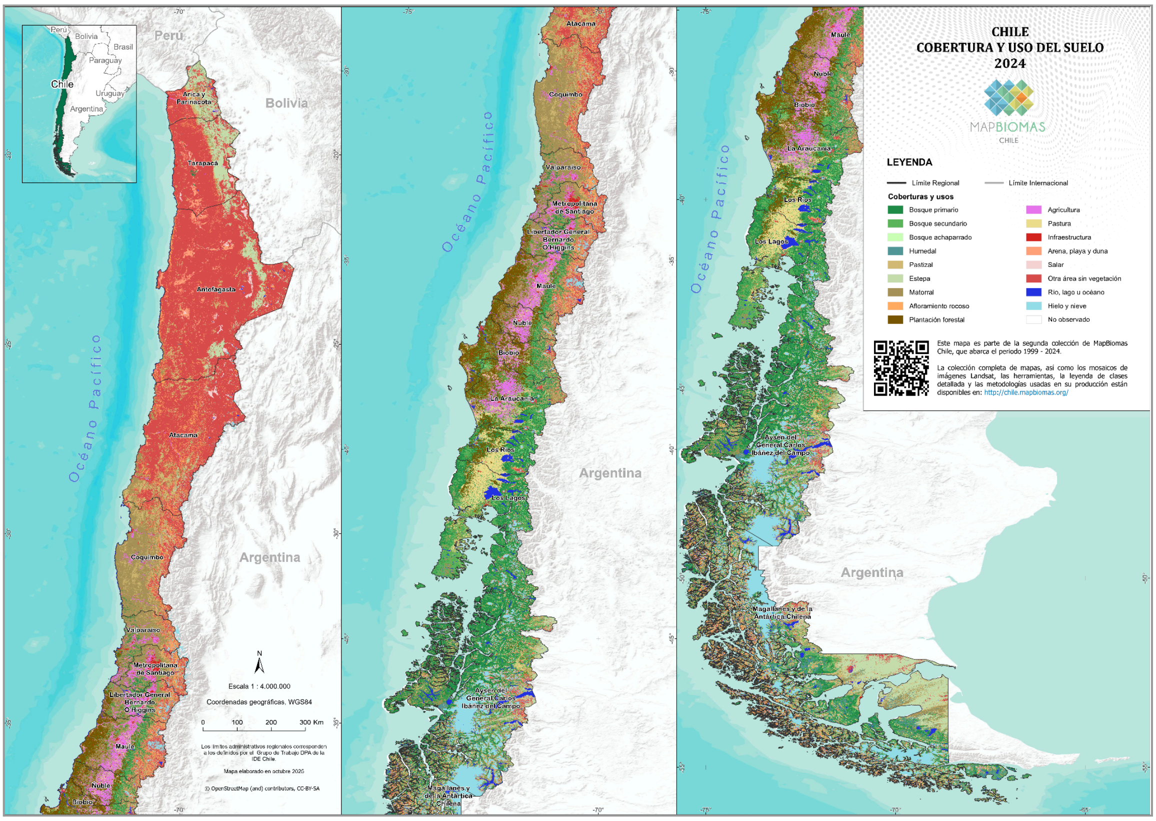

Figure 13: Land cover map for 2024 (MapBiomas Chile Collection 2).

The summary of the main results can be reviewed in the following document:

- **Fact Sheets**: Highlighted statistics for the main classes of Hierarchical Level 2 and differentiated by geographic area (North, Center-Sou and Patagonia): [https://chile.mapbiomas.org/destacados/]

5.2. Concluding Remarks and Perspectives

The MapBiomas Chile 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 provided an ideal environment to enhance and share skills and abilities by collaborators from different countries, cultures, languages but similar values: 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. The next collection of MapBiomas Chile will include an enhanced legend expanding the conceptual resolution of classes.

The change in strategy to using ecoregions for training purposes yielded good results and helped organize the work more efficiently. Collection 2 significantly improved data quality compared to Collection 1. However, several challenges remain for future collections, including:

- Improving the quality of the Landsat mosaics used as a basis for classification, particularly for the early years of the collection.
- Reviewing and eventually modifying the methodological approach for some anthropogenic classes, such as forestry and agriculture.
- Analyzing the need to open some lower-level classes, such as forestry, agriculture, and other non-vegetated areas.
- Because some ecoregions were subdivided due to climatic and biophysical differences, such as the Valdivian Forest Ecoregion, it is necessary to move toward a more formal definition of sub-ecoregion territorial units based on multi-criteria analysis.
- Evaluating the functioning of the governance model, roles, and responsibilities.

Finally, based on this collection, more systematic work will begin to disseminate the results, identify potential collaborators, and expand the local network. Several public and private stakeholders have already been identified as interested in participating, which provides an optimistic outlook for the future of MapBiomas Chile and the international network in general.

6. References

- Breiman, L. (2001). Random forests. Machine learning, v. 45, n. 1, p. 5-32.
- Elvidge, C.D, Zhizhin, M., Ghosh T., Hsu FC, Taneja J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages:2012 to 2019. Remote Sensing 2021, 13(5), p.922, doi:10.3390/rs13050922 doi:10.3390/rs13050922.
- FAQ (2016). Land Cover Classification System Classification concepts Software version 3. Antonio Di Gregorio. http://www.fao.org/3/a-i5232e.pdf
- Farr, T. G., et al. (2007), The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183.
- Huete, A. et al. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sensing of Environment, Volume 83, Issues 1-2.
- Jenness, J. (2006). Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x, v. 1.2. Jenness Enterprises.Available at: http: //www.jennessent.com/ arcview/tpi.htm.
- Miranda, A., Lara, A., Altamirano, A., Zamorano-Elgueta, C., Hernández, H.J., González,
- M. E, Pauchard, A. & Promis, A. (2018). Monitoring Chilean native forest area: a pending challenge. Bosque (Valdivia), 39(2), 265-275.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51(11):933-938.
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. Third Earth Resources Technology Satellite (ERTS) Symposium, 1, 309-317. https://doi.org/citeulike-article-id:12009708
- Souza, L. Oliveira and A. V. Fonseca (2001). Multi-Decadal Annual Land Cover Dynamics and Forest Disturbance in the Brazilian Amazon Biome.IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021, pp. 665-665, doi: 10.1109/IGARSS47720.2021.9554353.
- Squeo, F. A., Estevez, R. A., Stoll, A., Gaymer, C. F., Letelier, L., & Sierralta, L. (2012). Towards the creation of an integrated system of protected areas in Chile: Achievements and challenges. Plant Ecology & Diversity, 5,233-243.
- USGS Landsat (2017). USGS Landsat Collection 1 Level 1 Product Definition; USGS Landsat: Sioux Falls, SD, USA, 2017; Volume 26.
- Zhao, V., Feng, D., Vu, L., Wang, X., Chen, V., Bai, V., Hernandez, H. J., Galleguillos, M., Estades, C., Biging, G. S., Radke, J. D., & Gong, P. (2016). Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. Remote Sensing of Environment, 183, 170-185.