MAPBIOMAS

CHILE

ATBD_R

Algorithm Theoretical Base Document & Results

MapBiomas "Handbook"

Coordinación General MapBiomas Chile

Jaime Hernández

Equipo Técnico y Desarrolladores

Valentina González
Lissette Cortés
Rodrigo Fuentes
Diego Muñoz
Valentina Retamal
Alejandro Miranda
Carlos Hormazábal
Miquel Nieto

Coordinadores Institucionales

Jaime Hernández (U. de Chile)
Cristian Echeverría (U. de Concepción)
Adison Altamirano (U. de la Frontera)
Carlos Cárdenas (U. de Magallanes)
Natalya Hernández (WWF Chile)

MapBiomas Brasil

Tasso Acevedo
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 2000 to 2022 (Collection 1). 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. The classification algorithms are available on MapBiomas Github:

https://github.com/mapbiomas-brazil

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 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 the 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.

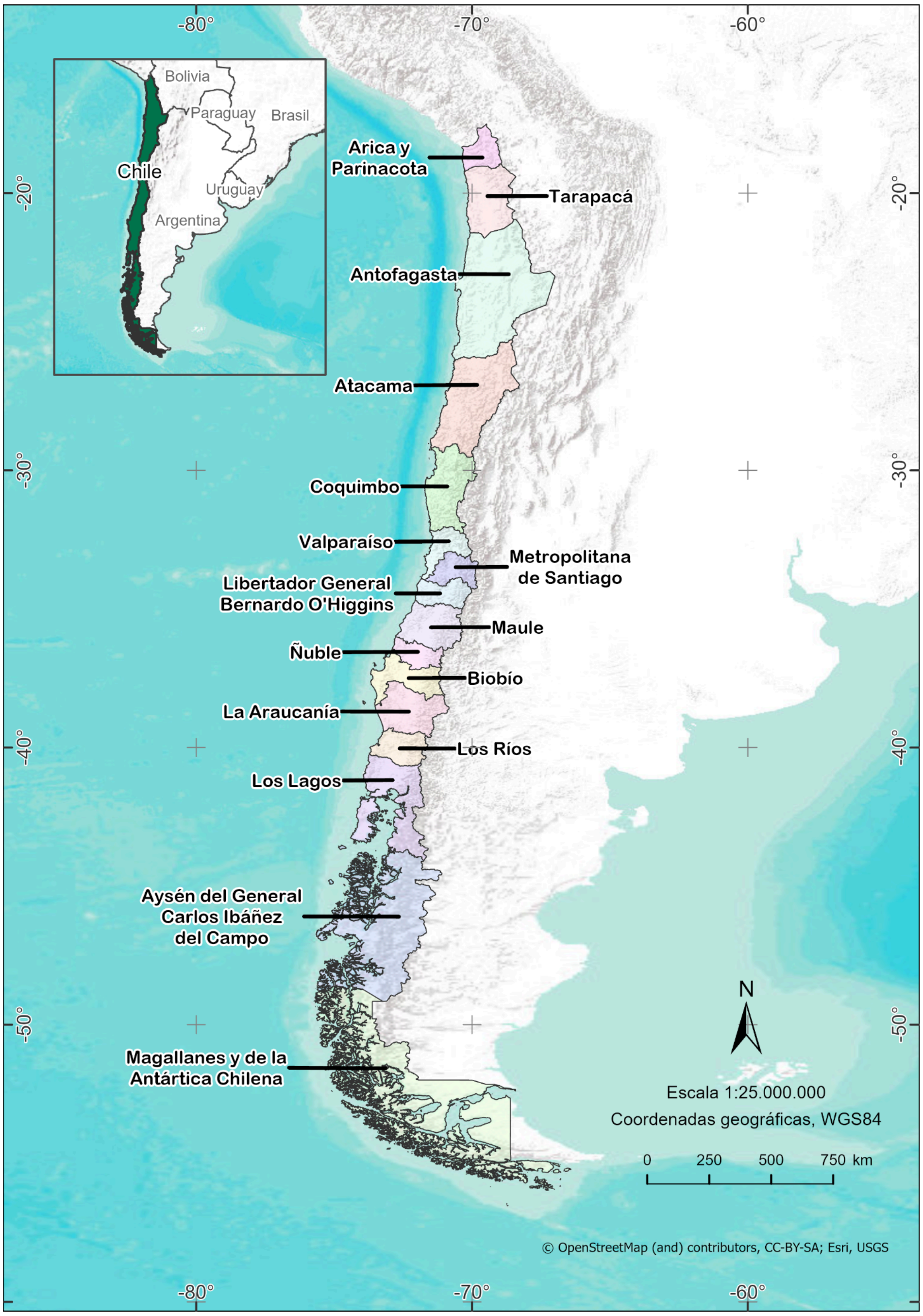

Figure 1: Chile and its Regions.

1.4. Key Science and Applications

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

- Mapping and quantifying LULC transitions.
- Quantification of gross and net forest cover loss and gain.
- Monitoring agriculture, forest plantations and pasture expansion.
- Assessing urban expansion.

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: 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 1 general methodological steps are presented in Figure 2.

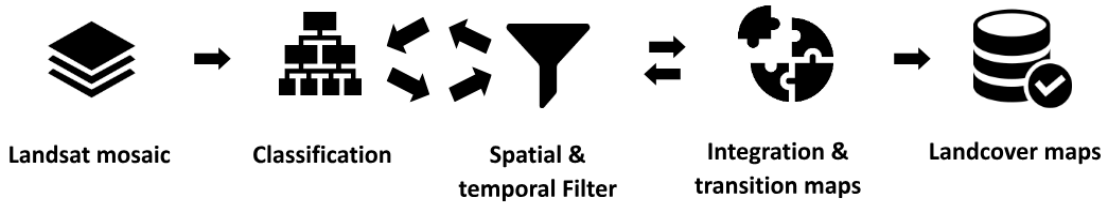

Figure 2: Collection 1 general methodological steps.

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). Then, yearly training samples were acquired using four working zones (WZ), each one assigned to University teams as follows:

- GEP Lab (U. de Chile): From Arica y Parinacota to Coquimbo Regions (WZ1).
- LEP Lab (U. de Concepción): From Valparaíso to Biobío Regions (WZ2).
- LEPCON Lab (U. de la Frontera): From Araucanía to Los Lagos Regions (WZ3).
- Universidad de Magallanes: Aysén and Magallanes Regions (WZ4).

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 1 product.

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 2000 to 2022 for the first collection. For future collections, it is expected to extend the length of the time window used.

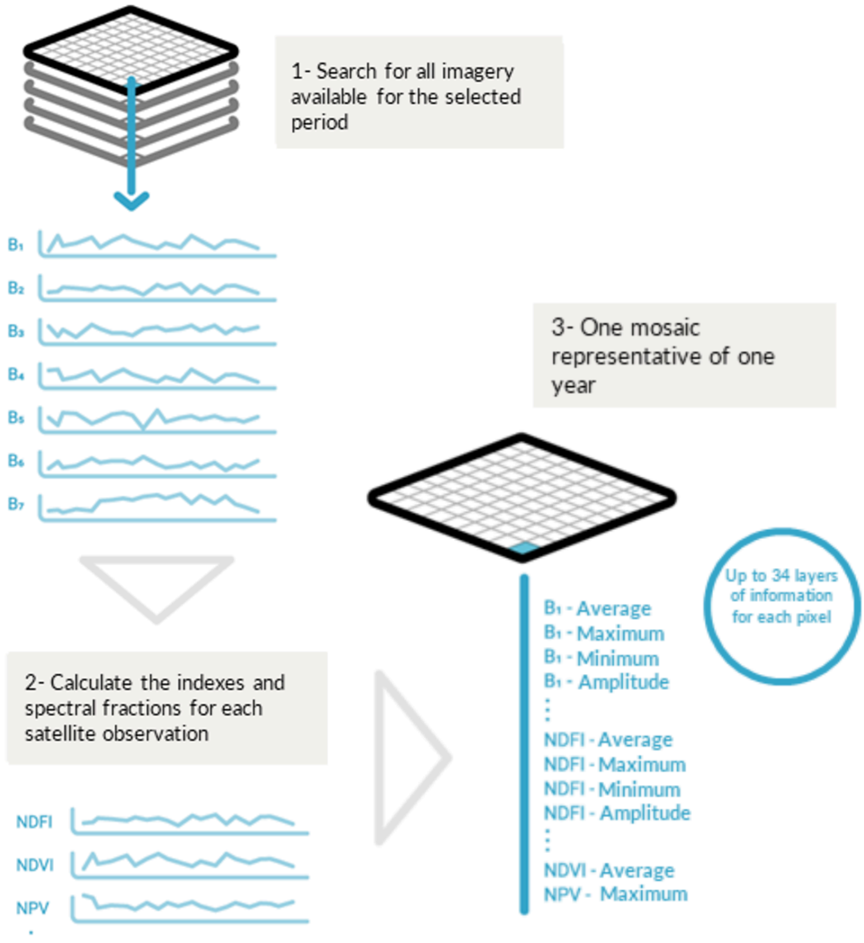

Figure 3: General workflow to build a year's mosaic.

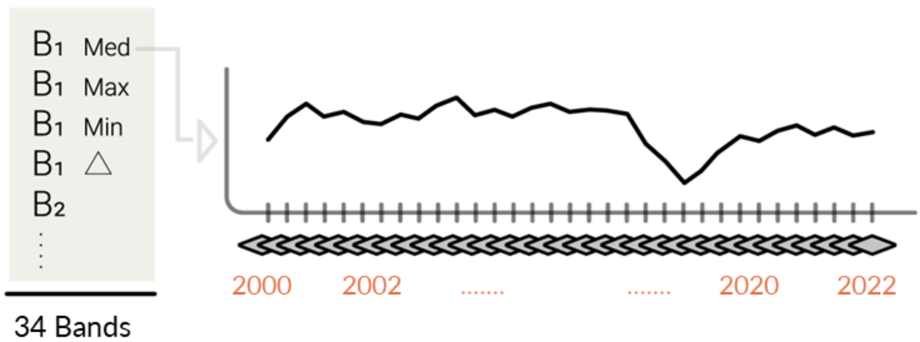

Figure 4: For each year a mosaic every pixel contains through 34 metrics or layers of information.

The feature space for digital classification of the categories of interest for the MapBiomas Chile Collection 1 comprises a subset of 34 variables (Table 1, Figure 4).

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

Variable	Description
Slope	In radians derived from the SRTM digital elevation model (Farr et al. 2007).
Aspect	In radians derived from the SRTM digital elevation model (Farr et al. 2007).
Elevation	Elevation in meters obtained from digital elevation model SRTM (Farr et al. 2007)
TPI	Topographic position index (Janness 2006).
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 Improved Vegetation Index (Huete et al. 2002).
Green median	Annual median of the green band.
Red median	Annual red band median.
Nir median	Annual median of the near-infrared band.
Swir 1 median	Annual median shortwave infrared band 1.
Swir 2 median	Annual median shortwave infrared band 2.
Gv median	Green Vegetation (Souza et al. 2021).
Gvs median	Median of the Green Vegetation Shade index (Souza et al. 2021).
Npv median	Median de Non-Photosynthetic Vegetation (Souza et al. 2021).
Soil median	Median Soil Fraction.
Shade median	Median shade fraction.
Ndfi median	Median of Normalized Difference Fraction Index (Souza et al. 2021).
Ndfi median wet	Median of the Wet period of Normalized Difference Fraction Index (Souza et al. 2021).
Ndvi median	Median normalized difference vegetation index (Rouse et al. 1973).
Ndvi median dry	Median normalized difference vegetation index dry period (Rouse et al. 1973).
Ndvi median wet	Median normalized difference vegetation index wet period (Rouse et al. 1973).
Ndwi median	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.
Sefi median	Median of Savanna Ecosystem Fraction Index.
Ndfi stdDev	Standard deviation of the Normalized Difference Fraction Index (Souza et al. 2021).
Sefi stdDev	Standard deviation of Savanna Ecosystem Fraction Index.
Soil stdDev	Standard deviation of Soil fraction (Souza et al, 2021).
Npv stdDev	Standard deviation of Non-Photosynthetic Vegetation (Souza et al. 2021).
Ndwi amp	Amplitude of Normalized Difference Water Index.

3.3. LULC scheme and Classification

3.3.1. Legend

The general MapBiomas classification scheme is a hierarchical system comprising four categorical levels. For the Chilean Collection 1 we only considered the first two (Table 2). 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 has 13 classes across the six classes of the first categorical level.

Table 2: Classes of land cover and land use of MapBiomas Collection 1 in Chile. Numbers in parentheses represent the categorical identifier in the LULC collection.

Class Level 1	Class Level 2 (Id)	Description
1. Forest formation	1.1 Forest (3)	Plant formations dominated by native tree species, with an average height greater than 2 m. Collection 1 includes primary and secondary (renewal) native forests. In addition, stunted forests, or Krumholz, of ñirre or lenga are included in this class.
2. Natural non forest formation	2.1. Wetland (11)	Vegetative cover dominated by herbaceous vegetation subject to periodic flooding by fresh and/or salt water.
	2.2. Grassland (12)	Plant formations with dominant herbaceous species. Those areas where the dominant vegetation was herbaceous, whether natural grasslands or northern grasslands. According to the complementary instruments, the classifications could be permanent or non-permanent natural vegetation.
	2.3. Shrubland (66)	Plant formations dominated by woody species with an average height of less than 2 m and a diversity of densities, which include from very sparse thorny thickets to dense sclerophyllous thickets.
3. Farming and silviculture	3.1. Forest Plantation (9)	Forest plantations of exotic tree species of commercial interest. This category includes plantations of *Pinus radiata* and several species of eucalyptus (i.e. *E. globulus* and *E.nitens*).
	3.2. Mosaic of agriculture and pasture (21)	Set of agricultural and livestock production areas. It includes annual crops, rice fields, fruit orchards, vineyards, fallow lands and meadows for animal production.
4. Non-Vegetated areas	4.1. Infrastructure (24)	Urban areas with a predominance of impermeable surfaces. Highways and primary paved roads are included in this category.
	4.2. Beach, Dune and Sand Spot (23)	Areas dominated by sand with little or no vegetation.
	4.3. Salt flat (61)	Areas characterized by large expanses of flat land covered by a layer of salt or mineral salts.
	4.4. Rocky outcrop (29)	Naturally exposed rocks without soil cover, often with partial presence of rock vegetation and on steep slopes.
	4.5. Other non-vegetated area (25)	Areas of bare soil or scarce vegetation, with less than 5% coverage, of natural origin or the product of anthropogenic activities.
5. Water Bodies	5.1. River, Lake and Ocean (33)	Areas permanently covered with water, of natural or artificial origin.
	5.2. Ice and snow (34)	Ice or snow covered areas.
6. Not observed (27)		Areas not classified due to the presence of clouds, shadows, atmospheric noise or low quality of satellite images.

3.3.2. Samples and Stable Samples

Training samples for each working zone were defined following a strategy of using random pixels for which the LULC remained the same throughout the entire period (23 years), so named “stable samples”. The stable areas were identified through annual preliminary classification made using random pixels selected from manually drawn polygons made by photo interpretation. For this, false-color composites of the Landsat mosaics for all the 23 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.

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. 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. It was classified in a separate process and integrated in the post-processing step.

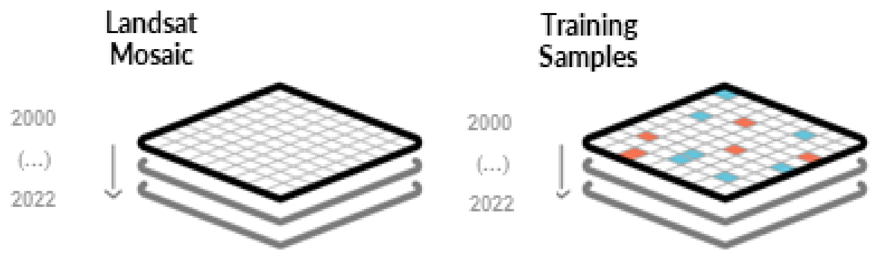

Figure 5: Representation of Landsat mosaic and training samples with two classes (red and blue).

3.4. Post-Processing

The results of the final classification in each working zone were improved by applying a set of filters, to correct missing data, “salt-and-pepper” classification errors and, specially, cases of misclassification. Furthermore, 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**.**

3.4.1. Gap fill filters

A filter to fill no-data pixels, or “gaps”, was applied. Because theoretically the no-data values are not allowed, they are 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.

3.4.2. 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 6). 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).

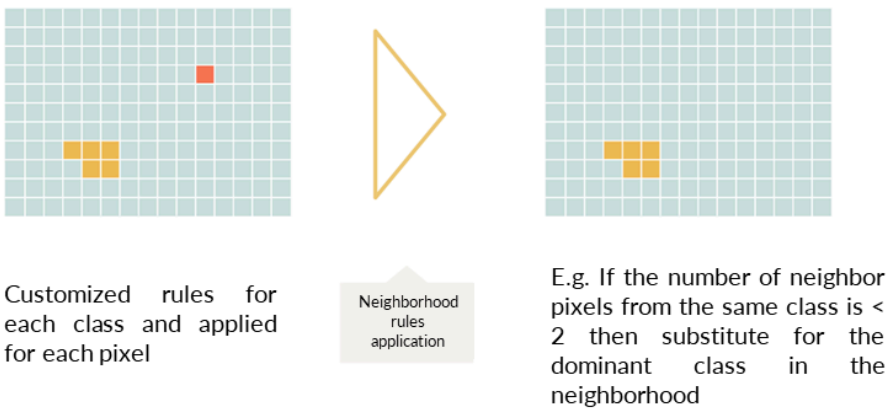

Figure 6: Spatial filter concept.

3.4.3. Temporal filters

The temporal filter uses sequential classifications in a three-year unidirectional moving window to identify temporally non-permitted transitions (Figure 7). The temporal filter inspects the central position of three consecutive years ("ternary"), and if the extremities of the ternary are identical but the center position is not, then the central pixel is reclassified to match its temporal neighbor class. Temporal filters were divided into two broad categories according to their functioning: regular and extremes (beginning and ending). While regular filters only use a 3 year window, extremes can employ a 3 or 5 years window. Thus, the rule involved in the extreme filters states that if in year "x" a given pixel is assigned to a class different from the following (antecedent) two years class - and in that two years the pixel was assigned to the same class - then the pixel in "x" was reclassified to match the following (antecedent) class. On the other hand, regular filters were applied between 2001 and 2021 and are based on the assumption that a class change between consecutive years which is immediately reverted in the third year is due to a classification error. This decision rule is relaxed when the temporal window encompasses five years, wherein the reversion can also occur in the fifth year - that is, the pixel can be misclassified for two consecutive years.

A mode filter was applied in Patagonia of the country where the effect of topography and defects in the mosaics due to high latitudes cause a given pixel to be classified in several years of the time series as 'Not observed' since the information contained in the predictors is altered by the aforementioned factors. In the pixels classified as “Not observed”, a mode filter was applied that replaces with the value of the mode between classifications that the technician defines as potential in a temporal window. In this case given the characteristics of the environment the possible classes were forests, rocky soils and ice and snow.

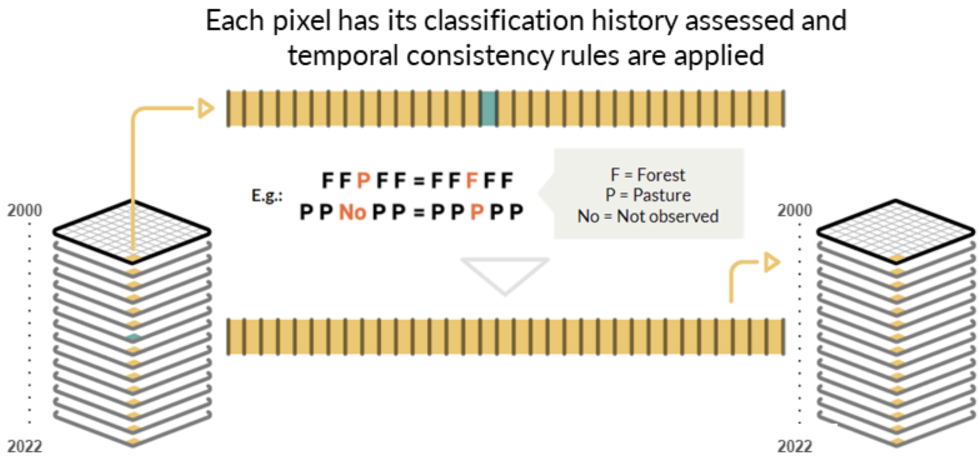

Figure 7: Temporal filter concept.

3.4.4. Classification of infrastructure

We restricted the classification geographical extent to a mask built using a the VIIRS Nightime Day/Night Annual Band Composite V2.1 product which is constructed from an annual time series of nighttime lights using monthly cloud-free radiance averages obtained from luminosity data collected by the sensor NASA/NOAA Visible Infrared Imaging Radiometer Suite (VIIRS) (Elvidge et al., 2021). The “average_masked” band representing the twelve-month average is specifically used to eliminate sporadic lights produced by the Northern Lights, fires or other isolated events. The used mask was built by selecting only the areas that had a pixel value greater than the value 1. This threshold was chosen based on iteration and revision of the area selected by the mask. This area corresponds to the potential infrastructure area, within which, we applied the random forest algorithm to classify the only two classes: Infrastructure and Others. We used the same 34 variables presented in table 1 as predictors. The total samples of both classes were weighted with a weight of 50 and 50% with a maximum of 3000 points per class to later be used as training points for the classifier. The classification was carried out iteratively, increasing the complementary points until an acceptable visual adjustment of the large cities in each area was achieved.

Once the initial classification is obtained, a series of filters were applied. First, a spatial filter, which consists of eliminating areas where the number of pixels connected vertically, diagonally or horizontally is less than 6 pixels (≈ 5000 m2). This process is carried out by adding an auxiliary variable that contains for each pixel the number of connected pixels of the same class, then a moving window of 2x2 pixels is defined in which the mode is calculated and in each pixel that does not meet the condition of connected pixels, the original value is replaced by the mode of the classification in the neighborhood. This process is carried out independently each year.

To eliminate high-frequency variation in classification produced by differences in annual mosaic characteristics, additional spatial filters are applied. The first for a 3-year window that begins with the oldest classification (2001) to the most recent (2021) without modifying the first and last year. In this window, when the target year (t) is not equal to the previous (t-1) and subsequent (t+1), the class of the previous year (t-1) is modified. This filtering is done in order of priority, changing and modifying the class with the biggest classification or noise problems, in this case "24" and then reviewing the most stable class, in this case "68". To the new classification, a temporal filter is applied that considers 4-year moving windows. This filter works the same as the previous one, but the condition is that the year (t) must be equal to the previous year (t-1) and different from the classification from 2 years before (t-2) and different from the year after (t+1). If the condition is met the classification is replaced by t-2. The iteration is carried out in order of priority from 2000 to 2022 without modifying the first and last year. Finally, a 5-year temporal filter was applied, whose operation is similar to the previous ones but the same years must be the 3 central ones. In this filter, a function to manually remap misclassified areas is also considered, but it was not used on this occasion.

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 2000 to 2022. 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 and then overlaid over the integrated mosaic. 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

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

infrastructure classification. Finally, the infrastructure final maps were superimposed with the general classified maps generating the final integrated maps.

3.5.2. Transition maps

The pixel-by-pixel class differences between any two maps were computed for the following periods (Figure 8): (A) any consecutive years (e.g. 2001-2002); (B) five-year periods (e.g. 2000-2005); (C) ten-year period (2000-2010). The class transitions represent LULC changes as follows:

- Transitions from agriculture classes or bare areas to forest cover or non-forest natural areas.
- Transitions that add water bodies surface.
- Transitions that reduce the water surface.
- Transitions with profit in forestry plantation.
- Transitions from forest cover or non-forest natural areas to agricultural or non-vegetated areas.
- Non-transition areas or transitions involving unobserved areas or transitions between classes within level 1 of the legend used.

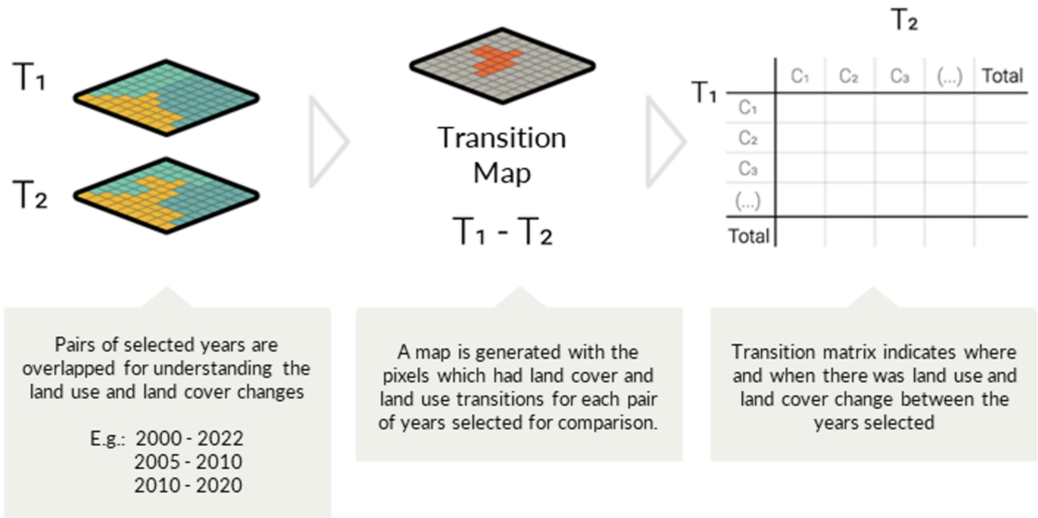

Figure 8: Transition map concept.

Additional spatial filters were applied in the transition maps. The target is to eliminate single pixels or streams of pixels in the border of different classes derived from the creation of transition maps. The general rules for this filter were: (i) pixels with only one neighbor pixel in the same transition class; (ii) streams of up to five pixels with two or one neighbor pixel in the same transition class.

3.6. Post-processing

3.6.1. Remapping pastures clases

For the "Mosaic of agriculture and pasture" class we performed further post-processing. First, a mask was generated using the pixels of the collection where, at least once during the period, they were classified as agriculture. The "NASA SRTM Digital Elevation 30m" product was used to obtain the elevation and slope, which were crossed with the previously generated mask and the areas that had a slope less than 10° and that were less than 4000 meters above sea level were maintained in the mask. In addition, another mask was created based only on slope and elevation, which included areas having a slope less than 10° and an elevation less than 620 meters above sea level. Both masks were then joined and the Southern area of Chile, Aysén and Magallanes regions, were left out as in those areas this postprocessing was not appropriate. A buffer of 100m was generated to the joint buffer. Finally, within the former mask, a reclassification was applied to the pixels of the grassland class (id 12) and agriculture (id 18) class to be remapped as Mosaic of agricultural and pasture uses (id 21). Pixels of these classes that were not found on the mask were classified as grassland (id 12).

3.6.2. Additional post-processing

Some additional post-processing filters were also applied to the final land cover maps. Areas classified as Forest Plantation, that were later years classified as Forest (native) were reclassified to Forest Plantation.

3.7. Statistics

Zonal statistics of the mapped classes were calculated for different spatial units, such as ecoregions, regions, provinces and municipalities, as well as watersheds and protected areas. A toolkit in the Google Earth Engine is available to upload user-defined polygons and download the LULC map.

The used code can be found here:
https://code.earthengine.google.com/?scriptPath=users%2Fmapbiomas%2Fuser-toolkit%3Amapbiomas-user-toolkit-lulc.js

4. Validation Strategies

Validation was performed for the classifications of the years 2002, 2012 and 2022. We used two databases of approximately 1300 randomly generated points each year with approximately 100 points per class. 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. In total, 2044 points were used to validate the 2022 map, 2000 points to validate the 2012 map and 1770 points to validate 2002 map, reaching 5814 points in total. 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. Tables 3 summarize the user and producer accuracies for main classes for 2002, 2012 and 2022 land cover maps.

Table 3: Accuracies for main classes for 2002, 2012 and 2022 land cover maps.
U = User accuracy, **P** = Producer accuracy.

CLASSES	ACCURACIES (%)					
	2002		2012		2022	
	U	P	U	P	U	P
Forest	79	80	83	81	88	70
Forest Plantation	84	70	92	75	83	87
Mosaic of Agriculture and Pasture	74	88	81	91	81	87
Infrastructure	82	82	86	89	79	90
River, Lake or Ocean	98	86	98	89	97	87
Ice and Snow	85	89	76	81	74	83
Salt flat	88	87	86	86	84	89
Shrublands	68	59	62	61	62	58

5. Map Collections and Analysis

5.1. Collection 1

The MapBiomas Chile Collection 1, including all land cover maps between 2000 and 2022, 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 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 3 shows the mural printable land cover map for 2022.

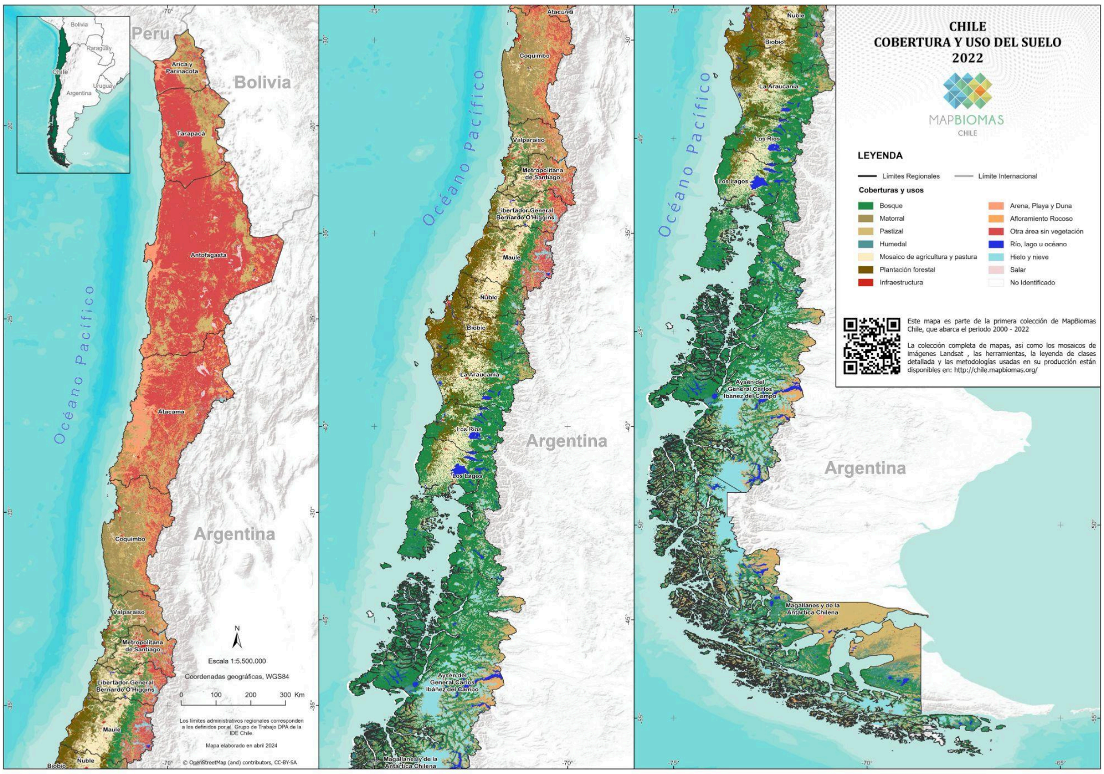

Figure 3: Land cover map for 2022 (MapBiomas Chile Collection 1).

La síntesis de los principales resultados se puede revisar en los siguientes documentos:

- **Infografía nacional**: estadísticas y tendencias resumidas a nivel nacional para el Nivel jerárquico 1. URL = http://xxxxxxxxxxxxxxxxxxxx.
- ***Fact Sheets***: Estadísticas destacadas para las principales clases del Nivel jerárquico 2 y diferenciadas por zona geográfica (Norte, Centro-Sur y Patagonia). URL = http://xxxxxxxxxxxxxxxxxxxxxxxxxxx.

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.

For the next collections it will be necessary to change the strategy of working areas to groups that work on transversal themes across the entire country. This is especially relevant for non-natural classes, such as infrastructure, agricultural mosaic grasslands or forest plantations. On the other hand, due to the great latitudinal variation of the country, it will be necessary to work on natural classes, such as forests, shrublands or grasslands, separately by ecoregions and not by mere administrative limits. The main class divisions that are seen as necessary for the next collections are: i) separation of primary forests from secondary forests, ii) Separation of agriculture from livestock meadows, iii) Separation of pine and eucalyptus plantations.

Finally, the need to have specialized groups, preferably by ecoregions, that collaborate both in the compilation of training areas and validation points has been discussed. During the validation stage, we detected the difficulty of correctly on-screen interpreting the same class throughout the whole country. This is especially relevant for some natural types such as shrublands and grasslands that are often confused with other types of land cover.

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.
- FAO (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, Á. (2018). Monitoring Chilean native forest area: a pending challenge. Bosque (Valdivia), 39(2), 265-275.
- 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
- C. 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., Estévez, 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, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., Hernández, 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.