ATBD

Algorithm Theoretical Basis Document

MapBiomas "Handbook"

Collection 6.0

Version 1.0

General Coordinator

Banchero, Santiago

Team

Santiago Verón

Ana Eljall

Sofia Sarrailhé

Andrea Moreno

María Jesus Mosciaro

Magdalena Bozzola

Gabriela Barraza

Octubre 2025

Table of Contents

EXECUTIVE SUMMARY

MapBiomas Chaco for Argentina is a collaboration network made up of governmental and non-governmental organizations, research institutes, universities, and companies from Argentina. This initiative uses advanced remote sensing technologies to produce annual series of land cover and land use maps. The MapBiomas Chaco network combines the efforts of the MapBiomas Brazil initiative and the inter-institutional arrangement between the National Institute of Agricultural Technology (INTA), Guyra Paraguay, The Nature Conservancy, Fundación Amigos de la Naturaleza (FAN), and Fundación Vida Silvestre Argentina.

This document describes the theoretical basis, relevance, and methods applied to produce annual land cover and land use (LCLU) maps in the Gran Chaco Americano region from 1985 to 2024, representing the MapBiomas Chaco Collection 6. All the MapBiomas Chaco maps and datasets are freely available at the project website (http://chaco.mapbiomas.org/).

1. Introduction

1.1. Scope and content of the document

The scope of this document encompasses the entire products' processing chain including the theoretical basis, justification, and methods applied to produce annual maps of land use and land cover (LCLU) in the Argentine Chaco from 1985 to 2024 of the MapBiomas Collection 6.

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 Chaco Collection 6 products. The classification algorithms are available on MapBiomas Github (https://github.com/mapbiomas-brazil).

The MapBiomas Chaco initiative was launched in July 2017, aiming to contribute to understanding LCLU dynamics in Chaco. The LCLU annual maps produced in this project were based on the Landsat archive available in the Google Earth Engine platform, encompassing the years from 1985 to the present. Since then, the MapBiomas mapping evolved year by year and was divided into Collections.

- Collection 1: 2000 - 2017 (released in December 2018).
- Collection 2: 2000 - 2019 (released in December 2020).
- Collection 3: 2000 - 2021 (released in September 2022).
- Collection 4: 1985 - 2022 (released in July 2023).
- Collection 5: 1985 - 2023 (released in September 2024).
- Collection 6: 1985 - 2024 (released in September 2025).

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 LCLU annual maps. All data, classification maps, codes, statistics, and further analyses are openly accessible through the MapBiomas Platform (http://chaco.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; and iii) MapBiomas collaborative network of organizations and experts that share knowledge and mapping tools.

The products of the MapBiomas Chaco Collection 6 are the following:

- Annual maps with land use and land cover.
- Pre-Processed feature mosaics generated from Landsat archive collections (Landsat 5, Landsat 7, and Landsat 8).
- Image processing infrastructure and algorithms (scripts in Google Earth Engine and source code).
- LCLU transitions' statistics and spatial analysis within administrative units, watersheds, protected areas, and other land tenure categorical maps.
- Quality assessment of the Landsat mosaics. Thus, each pixel in a given year was characterized according to the number of available cloud and aerosol free observations (varying from 0 to 23 observations per year).
- Temporal analysis (stable areas and number of classes).

1.3. Region of Interest

The "Gran Chaco Americano" is a forest ecoregion of exceptional environmental and social diversity. With 1,100,000 km^2, it is the second-largest woodland ecoregion in South America after the Amazon and includes territories of Argentina (62.19%), Paraguay (25.43%), Bolivia (11.61%), and Brazil (0.77%). In this region, deforestation for agriculture or cattle ranching is the dominant land-cover change. Potential ecological consequences include forest fragmentation, changes in primary productivity, carbon balance, and loss of biodiversity among others.

1.4. Key Science and Application

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

- Mapping and quantifying LCLU transitions.
- Quantification of gross and net forest cover loss and gain.
- Monitoring of regeneration and secondary growth forests.
- Monitoring agriculture and pasture expansion.
- Regional planning.

2. Overview and Background information

New features of Collection 6 include i) the extension of the time series now spanning from 1985 to 2024, and ii) an improved and reorganized legend comprising 22 classes (compared to 18 from the previous collection).

Algorithm Descriptions, Assumptions, and Approaches

3.1. Legend

The legend is described in three organizational levels and includes 22 land cover and land use classes (Table 1). The main differences between Collection 5 and Collection 6 legends are twofold: a) discrimination of the former woodlands into forests and shrublands, and b) the inclusion of 5 new perennial crops. Thus, in the new legend the former classes “open woodland” and “close woodland” are now separated into “open forest” and “closed forest”, “open shrubland” and “close shrubland”. Similarly, collection 5 class “Forest Plantations” has now been disaggregated into the following classes: “Citrus”, “Olives”, “Blueberry crops”, “Walnut crops”, “Wine crop - perennial”, “Other perennial crops”, and “Forest Plantations". In addition, some classes from collection 5 appear merged in collection 6. These are: a) “single crop” and “multiple crops” are now presented as “Temporary crops”, b) the new class “shrub and herbaceous mosaic” consists of the previous “sparse grassland” and “sparse woodland”, and c) “grassland” merges collection 5 classes “close grassland” and “open grassland”. Finally, in collection 6 the class “urban area” has been assigned to “other non-vegetated areas” as it will be mapped throughout Argentina using a specific methodology (see Argentina Urban Areas ATBD).

Details of the description of each class can be found in the Legend document (see Annex 2).

Table 1. Land cover and land use categories considered for digital classification of the Landsat mosaics for Gran Chaco in Collection 6

COLLECTION 6 - CLASSES	ID	Hexadecimal code	Color
1.Forests		**#1F8D49**	
1.1.Closed forests	3	#1F8D49	
1.2.Open forests	4	#7DC975	
1.3.Flooded forests	6	#026975	
2.Natural herbaceous and shrub vegetation		**#D6BC74**	
2.1.Closed shrublands	66	#A89358	
2.2.Open shrublands	77	#86B074	
2.3.Shrubs and herbaceous mosaic	63	#C7E0AB	
2.4.Grasslands	12	#D6BC74	
2.5.Flooded grasslands	11	#519799	
3.Agricultural and livestock area		**#FFEFC3**	
3.1.Agriculture	18	#E974ED	
3.1.1.Temporary crops	19	#C27BA0	
3.4.Perennial crops	36	#D082DE	
3.2.Pastures	15	#EDDE8E	
3.3.Forest plantations	9	#7A5900	
3.4.Citrus	47	#9631C8	
3.5.Other Perennial Crops	48	#E6CCFF	
3.6.Wine crop - Perennial	83	#690A8F	
3.7.Olives	84	#8E9F53	
3.8.Walnut trees crops	85	#ADA48E	

3.9.Blueberries crops	86	#3B33AB	
4. Non-vegetated area		**#D4271E**	
4.1.Beach, Dune and Sand Spot	23	#FFA07A	
4.2.Other non-vegetated areas	25	#DB4D4F	
4.3.Salt Flats	61	#F5D5D5	
5. Water bodies		**#2532E4**	
5.2.Rivers, lakes or ocean	33	#2532E4	
6. Not observed	**27**	**#FFFFFF**	

3.2. Landsat Mosaics

As usual in MapBiomas Chaco, Collection 6 relied on Landsat images. Thus, we processed all available images from Landsat 5 (L5), Landsat 7 (L7) and Landsat 8 (L8) from 1985 to 2024, to produce the mosaics from which the feature space was generated. In years where available Landsat scenes were not enough to generate the mosaics we applied a cubic interpolation method to model the 23 Landsat scenes using all per pixel available observations from up to 2 years after the target year (see section 3.4).

3.3. Image Processing

The Gran Chaco Americano encompassed 76 path-row combinations (Figure 1), potentially representing 62,920 Landsat scenes during the study period. For all sensors, we used Collection 2, Tier 1, Level 2 atmospherically corrected surface reflectance image collections available at Google Earth Engine. Each scene was cropped to avoid known artifacts at the borders that hampered seamless mosaicking. Additionally, the QA index was used to filter cloudy pixels. Finally, we generated annual image stacks by mosaicking all remaining scenes - including all 30 m resolution spectral bands - and considering the calendar year (i.e. January to December).

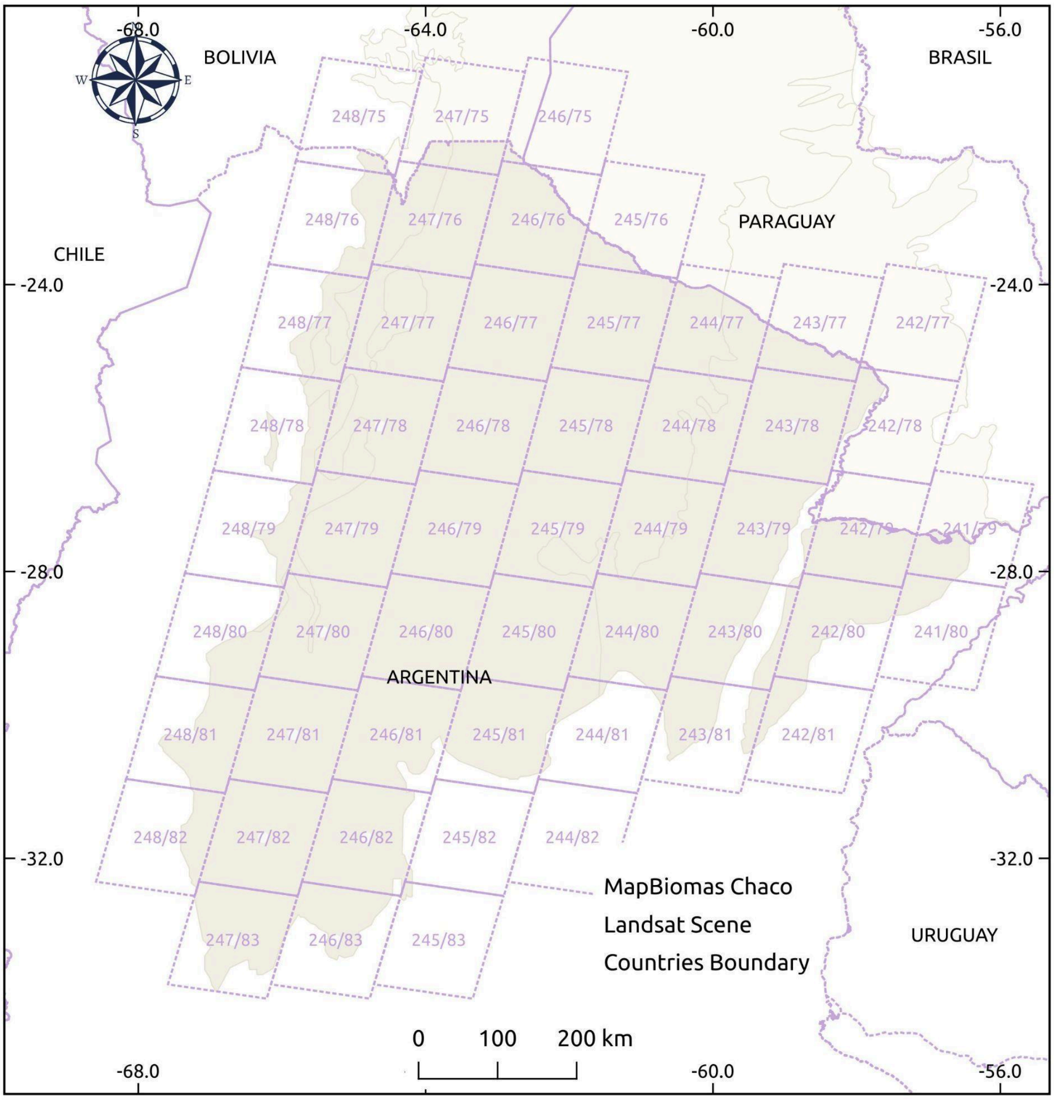


Figure 1. Distribution of Landsat path-rows for MapBiomas Chaco and Yungas.

3.4. Synthetic Landsat scenes

To avoid large areas with missing information due to lack of cloud free observations -in particular between 1985 and 1987- we synthesized Landsat scenes. Synthetic scenes were generated by applying a cubic interpolation (Descals et al. 2020) to all cloud free observations from the target year and the following two years. Thus, an "artificial" calendar year was created by subtracting 365 days to each observation of the following year and 730 days to every observation of the target year +2 years. The time lag -i.e. subperiods over which the cubic model was applied- used was 40 days. The resulting synthesized Landsat scenes had six bands (blue, green, red, NIR, SWIR1, SWIR2) with a time interval of 16 days (23 scenes per year). We acknowledge and make explicit that filling the target year missing data with observations from up to 2 years later does introduce errors due to land use and land cover changes that might have occurred within that period. However, we judged this was an unavoidable outcome if we were to produce maps for areas that were not observed -or observed very few times- during the target year.

3.5. Feature space

From the annual mosaics we calculated several variables that were used in the feature space to characterize each pixel. These variables described different aspects of the spectral behavior from a central tendency metric (i.e. median) of a particular period to variability metrics (i.e. range, standard deviation and coefficient of variation) from the full period or calendar year (Table 2).

In turn, we used bands, indexes, fractions, slopes and accumulated values as proxies of different features of the spectral behavior. While some of these variables were already available (e.g. spectral bands) others (e.g. fractions) were obtained from spectral unmixing methods (Souza *et al.,* 2003), involved arithmetic operations with different spectral bands (e.g. indexes), or were calculated from linear models fitted to the relationship between a given index (GCVI, NDVI and NDFI) and time (e.g. slope) or as the summation of each index value along a two-year timeframe .

Table 2. Selection of variables, calculation formula and measures of central tendency and dispersion according to time extension that were initially taken into account for the generation of the attribute space. This attribute space allows characterizing the spectral and phenological behavior of each entity/pixel and assigning it to one of the classes for which training data is available through the Random Forest classification algorithm.

			Key Period	Full Period	
Variables Types	**Names**	**Fórmula**	**Median**	**Range**	**Standard Deviation**
Bands	Blue	B1 (L5 and L7); B2 (L8)	x		
	Green	B2 (L5 and L7); B3 (L8)	x		
	Red	B3 (L5 and L7); B4 (L8)	x		
	Near Infrared (NIR)	B4 (L5 and L7); B5 (L8)	x		
	Shortwave Infrared 1 (SWIR1)	B5 (L5 and L7); B6 (L8)	x		
	Shortwave Infrared 2 (SWIR2)	B7 (L5), B8 (L7); B7 (L8)	x		
Fractions	GVS		x	x	x
	NPV		x	x	x

	Soil		x	x	x
	Cloud		x	x	x
	Shade	100 - (gv + npv + soil + nubes)	x	x	x
Indexes	NDVI	(NIR - red)/(NIR + red)	x	x	x
	EVI2	(2,5 * (NIR - red)/(NIR + 2,4 * red + 1)	x	x	x
	CAI	(IM2 / IM1)	x	x	x
	NDWI	(NIR - IM1)/(NIR + IM1)	x	x	x
	GCVI	(NIR / green - 1)	x	x	x
	HALL Index	(-red * 0,0017 - NIR * 0,007 - IM2 * 0,079 + 5,22)	x	x	x
	PRI	(blue - green)/(blue + green)	x	x	x
	SAVI	(1+L) * (NIR - red)/(NIR + red + 0,5)	x	x	x
	GVS	gv / (gv + npv + soil + cloud)	x	x	x
	NDFI	(gvs - (npv + soil))/(gvp + (npv + soil))	x	x	x
	SEFI	(gv+nvp_s - soil)/(gv+npv_s + soil)	x	x	x
	WEFI	((gv + npv) - (soil + shade)) / ((gv + nov) + (soil + shade))	x	x	x
	FNS	((gv + shade) - soil) - ((gv + shade) + soil)	x	x	x
Coefficient of variation	CV(GCVI)	(stdDev_gcvi / median_gcvi)	x		x
	CV(NDVI)	(stdDev_gcvi / median_gcvi)	x		x

Elevation based	Terrain slope	The same for all years			

3.6. Definition of the key period

The selection of periods determines the temporal extension over which a subset of the variables in the feature space are calculated. The definition of the period represented a trade-off between the probability of maximizing differences in spectral behavior of classes and the availability of cloud-free images. Since photosynthetic activity varies over time in different ways according to vegetation characteristics, the ability to discriminate between classes will depend on the period of the year under consideration.

To characterize the spectral behavior of Woody Vegetation, Agriculture, Pastures and Grasslands, 300 polygons were randomly sampled from each zone (see Section 3.7.5). NDVI for the year 2018 was used to calculate the euclidean distance as a measure of the separability between classes on a monthly basis. Quarterly euclidean distances were then calculated as the average distances of the three corresponding months. Additionally, we quantified the amount of good quality observations for each polygon for the year 2018 available monthly.

3.7. Classification

3.7.1. Classification Scheme

The classification of the Gran Chaco Americano results from an iterative process represented in Figure 2. Most of the steps were implemented in the Google Earth Engine platform. Preliminary supervised classifications were generated annually using training samples obtained from the maps of MapBiomas Chaco Collection 5 and annually generated Landsat mosaics. For these preliminary classifications, a set of stable pixels - i.e. pixels that remained in a single class throughout the study period - were selected using a stratified random sampling. These samples were filtered using reference information (previously generated vectors of land cover). At each zone (see section 3.7.5), a second classification was performed with these stable samples wherein new complementary samples were added in an iterative way. Complementary samples were selected from hotspot areas - i.e. places where there was an evident mismatch between classification output and actual land cover/land use. Post-classification temporal, spatial and frequency filters were applied sequentially. The validation was performed with independent random samples obtained annually using visual interpretation of satellite images using an ad-hoc Google Engine app.

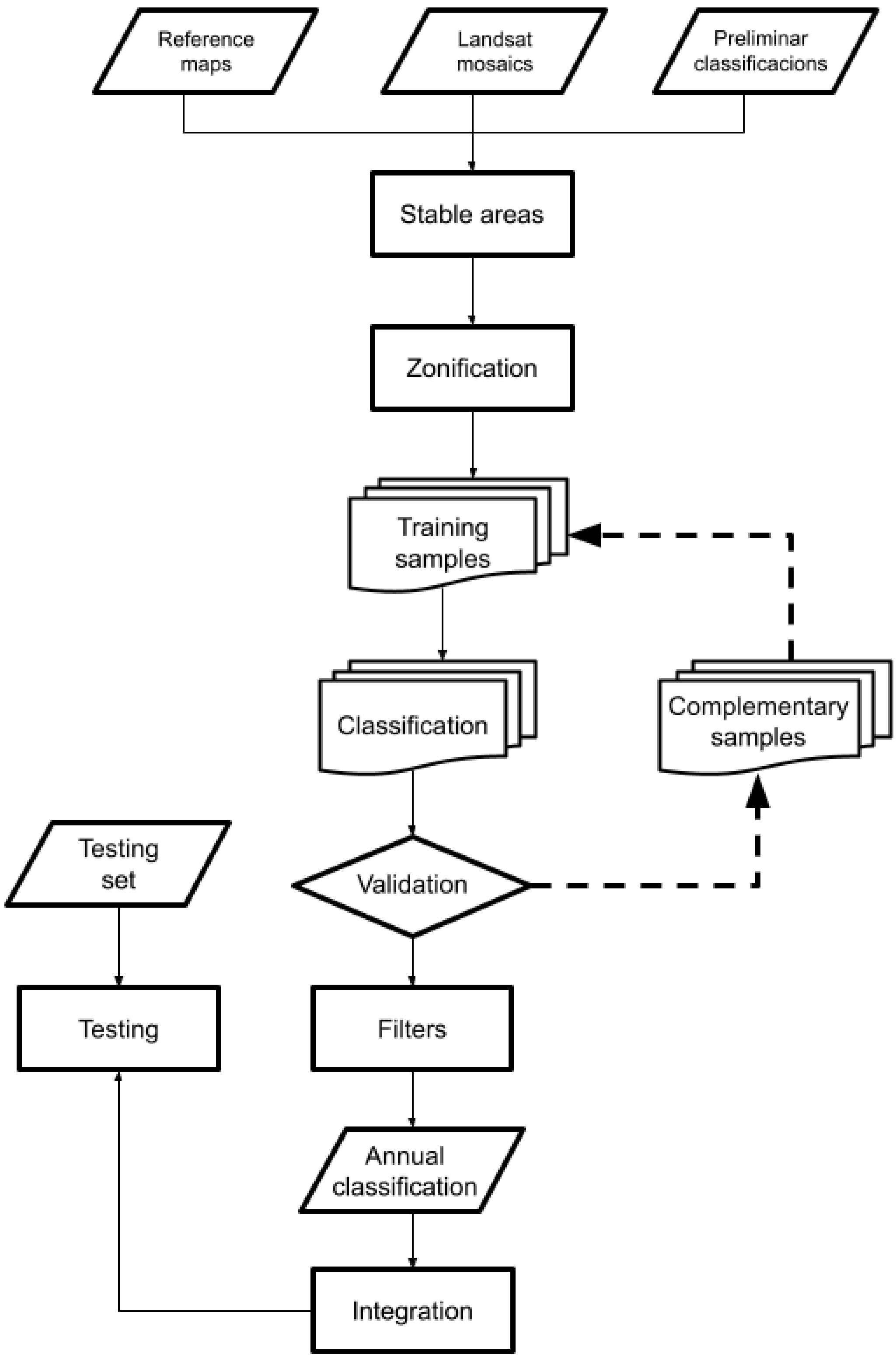


Figure 2. Classification scheme for MapBiomas Collection 6 in the Gran Chaco Americano.

3.7.2. Stable samples

From the preliminary classifications from 1985 to 2023, stable pixels -i.e. those that were classified as the same class over the years in Collection 5- were masked. A stratified random sampling process was performed over these masks to generate a fixed number of training samples for each class and zone to be used on the 1985 to 2024 set of annual classifications. Given the known problems that imbalanced data pose on random forests classification algorithms we followed MapBiomas Pampas method to balance training samples. Briefly, only a random subset of the initial fixed number was selected based on the class area proportion within each zone in each year to be classified. Therefore, linear simple functions were adjusted to estimate the proportion of each zone occupied by a given class for each year from 1985 to 2023, based on the annual class area observed along the Collection 5 dataset. These functions were used to estimate, for each year, the proportion of each class to train the classifier. Then, these annual proportions for each class were set to extract a subset of the available samples for the correspondent classification in each year. Whenever the classification resulted in overestimation or underestimation of the class after comparing with supplemental information (e.g.: Collection 5 maps) this proportion was adjusted changing the bias (intercept of linear regression model) accordingly. In any case, a minimum number of 100 samples per class was set for each region and year, to ensure the correct detection of the less frequent categories.

3.7.3 Reference maps

To capitalize spatially explicit information, we surveyed available land use land cover maps from official, academic, and ONG sources. The maps used are listed in Annex 1. Reference maps assisted in the identification of stable areas where stable samples were generated.

3.7.4. Zonation

The rationale underlying the zonation process is that supervised land use land cover classifications perform better on smaller and more homogeneous areas wherein training samples are more representative of class characteristics than over larger and more heterogeneous areas. Collection 6 improved previous zonation by including additional data sources. For Argentina, collection 4 zonation - done using the SNIC algorithm (Achanta & Susstrunk, 2017) fed with biophysical data layers- was fused with Oyarzabal et al (2018) and Morello et al (2012) zonations based on expert opinions to join or divide zones. Thus, collection 6 zonation (Figure 3) totalled 14 zones.

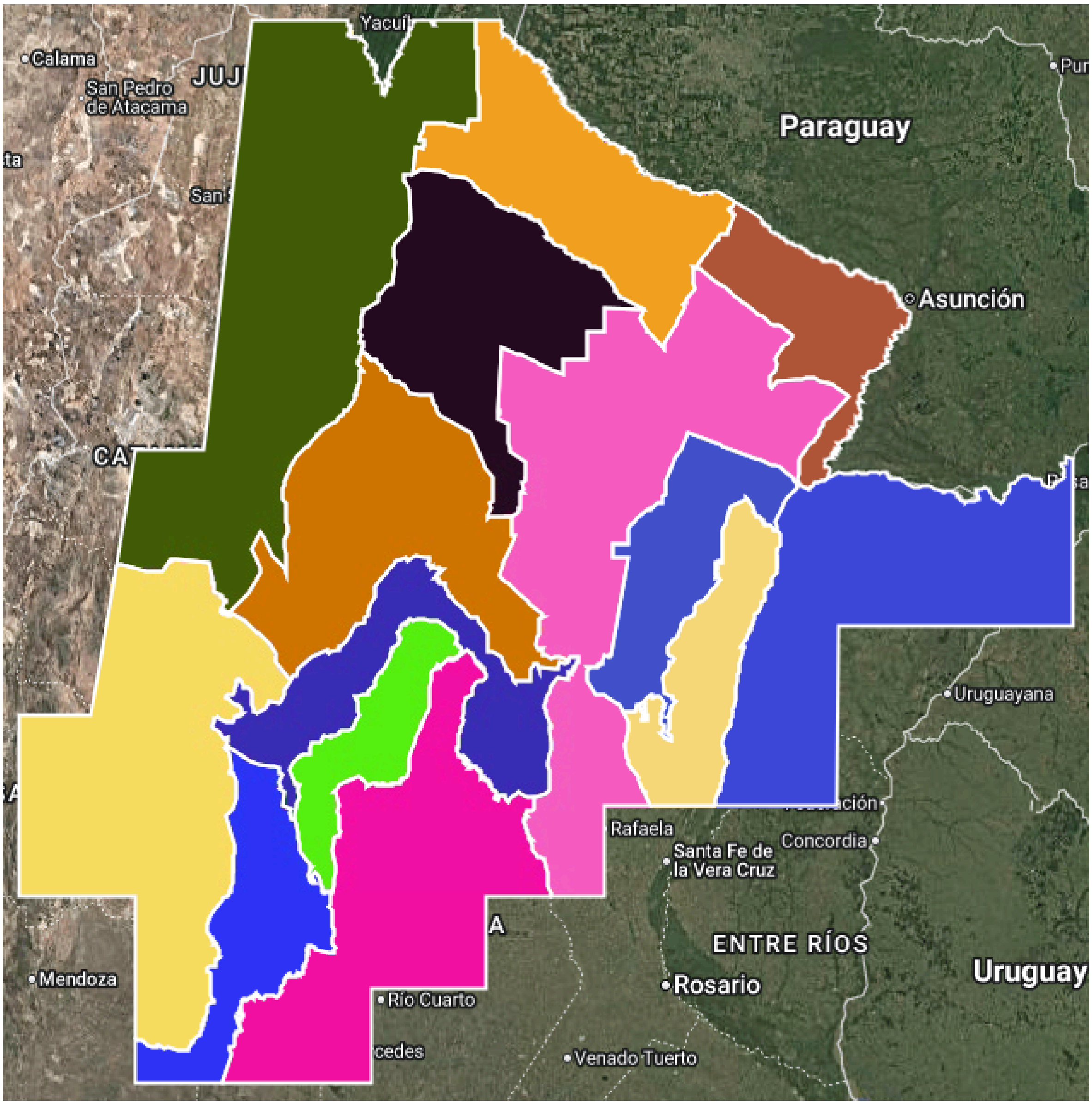


Figure 3. Zoning from biophysical variables that make up 14 homogeneous zones used for the classification of MapBiomas Chaco Collection 6.

3.7.5. Iterative Classification process

After the classification obtained from the stable training samples, complementary samples were added in hotspot areas after a revision process. New classifications were generated with stable and complementary samples. This process was carried out in an iterative manner until a stable map was obtained without unexpected widespread changes in classes between years. Collection 6 added a set of ad-hoc tools to improve the classification. These were: a) geotiff visualizers to inspect classification snapshots of 4 years periods selected by the user (as an alternative to the complete 1985-2024 period) and b) a pixel-based tool to plot the classes assigned to a given pixel on a yearly basis, and c) a general improvement in the scripts coding to speed up intermediate process.

3.7.6. Discrimination between forests and shrublands from previous woodland classes.

The discrimination of forests and shrublands was performed based on the assumption that trees are taller than shrubs and it was restricted to the areas assigned to woodlands in collection 5. Thus, we trained a Random Forest Classifier using samples selected from open and close woodlands that were tall or short based on a canopy height map and the same feature space as the other classes. From this preliminary classification we selected stable samples (those that remained unchanged throughout 10 years periods) and produced a new classification that was visually interpreted and whenever classification errors were detected new complementary samples were added -and the classification was performed again.

After evaluating different sources of canopy height information, we selected Potapov et al (2021) map to sample the height distribution of pixels previously (in collection 5) assigned to open or close woodland in 2021. Thus, for each class and zone, an interpreter calculated a height threshold value in order to assign each pixel to a “short” or a “tall” group assuming that the height frequency histogram would be bimodal and that samples from the “short” group would be representative of shrublands and from the “tall” group of forests. Whenever the histogram failed to show a bimodal distribution the interpreter would set the height threshold value based on trial and inspecting the resulting maps in GEE. In this way, we accommodated for spatial differences in trees and shrubs height differences. Finally we randomly selected 10000 samples per zone and deleted anomalous samples using the Isolation Forest score as well as shrublands samples located in areas not mapped as shrublands in auxiliary maps (see Annex 1).

3.7.7. Discrimination among perennial crops

In MapBiomas Chaco Collection 6, the class “Forest Plantations” was disaggregated into several classes: olives, blueberries, wine crops, citrus, walnuts, other perennial crops and forest plantations. We performed this disaggregation by means of auxiliary maps (see Annex 1) that assisted the identification of training samples. Thus, training samples were randomly selected from the spatial subset that satisfied two conditions: had been classified as forest plantation in collection 5 and to overlap with an area assigned to any of the new categories -olives, blueberries, wine crops, citrus, walnuts, and forest plantations-. For each zone, the training samples selection was performed over a 10 year period (i.e. 1985-1994,1995-2004,2005-2015, and 2016_2024). For zones and periods for which there were not auxiliary maps, interpreters selected training samples based on visual inspection of Landsat images and NDVI time series. As usual, anomalous samples were removed based on the Isolation Forest score. The resulting samples trained the Random Forest Classifiers using the same feature space as any other class. Once the preliminary classification was ready, errors were identified and complementary samples were added to minimized missclasified pixels.

3.8. Post-Classification

Collection 6 post-classification activities involved the application of 15 types of filters grouped into 3 broad categories (see annex 3 for detailed schemes of each filter). These filters are basically decision rules to reduce the errors of the classification using spatial, temporal, frequency or incidence information. The following sections provide a brief description of how these filters work. These filters were applied at three different spatial extents: overall region, zones (see Section 3.7.5), and areas within zones as well as over specific classes when needed.

3.8.1. Gap filter

The gap filter is intended to fill missing labels in any annual classification. Thus it is implemented at the beginning of the post-classification process and assigns a label to any pixel with missing label based on the previous or posterior classification. In general missing label pixels correspond to very bright areas where the cloud filter erroneously assumes cloud presence all along the year.

3.8.2. Spatial filters

The spatial filters were applied to a mask of continuously connected pixels, patches up to six pixels were processed with a morphological operations filter (focal_mode) with a kernel of 1 pixel. This reduces the salt and pepper effect by modifying isolated pixels.

3.8.3. Temporal filters

Temporal filters were divided into two broad categories according to their functioning: i) extremes (beginning and ending), and ii) regular. Both filters use a 3-year window to apply the decision rule. Thus, the rule involved in the extreme filters states that if in 1985 (2024) 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 1985 (2024) was reclassified to match the following (antecedent) class. On the other hand, regular filters were applied between 1986 and 2022 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

3.8.4. Frequency - incidence filters

The frequency filters used for Chaco were applied to correct noise problems over different natural and anthropic covers. These ad-hoc developed filters allow reclassifying the occurrence of a class that to a large extent remains constant but suffers random changes as a result of some false positives. Incidence filters were applied to minimize false positives that were not corrected by the frequency filters. Incidence is defined as the number of times a pixel changes its class over time. Thus, incidence filters complemented the frequency filters particularly when 2 given classes are alternated between years -as for example borders between patches-. In these cases the decision rule applies the mode of the pixel classes (or the mode of a subset of classes, see annex 3).

3.8.5. Transitions maps

Transitions are defined as a land use/land cover change occurring in a given period. For each pixel we calculated transitions for the following the periods: (A) any consecutive years (e.g. 2001-2002); (B) five-year periods (e.g. 2000-2005); (C) ten-year periods (e.g. 2000-2010); and (D) complete period (1985-2024). The places and the amount of area experiencing transitions are available as maps and Sankey diagrams respectively in the MapBiomas Chaco web-platform. Classes were aggregated in order to calculate the following 6 possibilities:

- Transitions from farming classes or non-vegetated areas to forest cover or non-forest natural areas.
- Transitions that add water surface.
- Transitions that reduce water surface.
- Transitions with gain of tree plantations.
- Forest cover or non-forest natural areas transitions to farming classes or non-vegetated areas.
- Areas without transitions or transitions involving unobserved areas or transitions involving native vegetation, or transitions between anthropic classes except gain of trees plantation

3.9. Statistics

The area (in has) for each of the different classes in the legend (level 1, level 2, level 3, and level 4) was calculated for different spatial units: biomes, countries, provinces (or departments in Paraguay) and districts as well as protected areas. These data are available to download as an MS Excel spreadsheet in the MapBiomas Chaco web-platform.

4. Validation strategy

Validation was performed through the generation of geolocated points that were assigned to one of the 22 classes by experts' visual interpretation of satellite images. Validation was performed for the 39 years between 1985 and 2024. A random sampling method was used to generate the validation samples taking the classification from 2004 as the stratification layer -i.e from where the proportions of the area occupied by each class were calculated. The number of validation samples was determined using equation 5.1 taken Olofsson *et al.* (2014) where n is the number of the validation samples, S(O) is the standard error of the estimated overall accuracy that would like to be achieved, W_i is the mapped proportion of area of class i, and S_i is the standard deviation of stratum i. As N -the number of pixels in the study area- is a very large number, we ignored the second term in the denominator. The resulting number of overall samples from the 22 classes thus calculated was 3130.

$$n = \frac{(\sum Wi\, Si)^{2}}{[S(O)]^{2} + (1/N)\sum WiSi^{2}} \quad (5.1)$$

For this purpose, a tool was developed in the Google Earth Engine platform. This tool allows for each sample point, to assign each class for each year and visualize Landsat images for each year, Landsat time series for the period (1985-2023), and links for visualization of Bing and Google Earth maps.

5. Concluding remarks and Perspectives

The MapBiomas initiative combines people, algorithms, satellite information and large-scale processing in a methodology that has revolutionized the operational large-scale generation of LCLU 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.

6. References

Achanta, R., & Susstrunk, S. (2017). Superpixels and polygons using simple non-iterative clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4651-4660).

Clinton, N. (2019). L7 gap filling snippet. Stackoverflow https://stackoverflow.com/questions/55256739/slc-code-not-filling-all-landsat-7-sr-gaps

Descals, A., Verger, A., Yin, G. & Peñuelas, J. (2020). "A Threshold Method for Robust and Fast Estimation of Land-Surface Phenology Using Google Earth Engine," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 601-606, doi: 10.1109/JSTARS.2020.3039554.

Dinerstein, E., Olson, D. M., Graham, D. J., Webster, A. L., Primm, S. A., Bookbinder, M. P., & Ledec, G. (1995). Una evaluación del estado de conservación de las ecorregiones terrestres de América latina y el Caribe. Washington D.C.: WWF - World Bank. Pp.: 135.

Ibisch, P. L., Beck, S. G., Gerkmann, B., & Carretero, A. (2003). La diversidad biológica: Ecorregiones y ecosistemas. In P. L. Ibisch, & G. Mérida (Eds.), Biodiversidad: La riqueza de Bolivia. Estado de conocimiento y conservación (pp. 47–88). Santa Cruz, Bolivia: Fundación Amigos de la Naturaleza.

Liu F. T., Ting K. M., Zhou H.: Isolation-based Anomaly Detection. ACM Transactions on Knowledge Discovery from Data 6(1), 1556-4681 (2012)

Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Hughes, M.J., Laue, B. (2017). Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sensing of Environment, 194, 379-390.

Morello, J., S.D. Matteucci, A.F. Rodríguez & M. Silva. 2012. Ecorregiones y complejos ecosistémicos argentinos. Orientación Gráfica Editora S.R.L., Buenos Aires. 752 pp y un CD con mapas. (ISBN 978-9-871922-00-0).

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-57.

Oyarzabal, M., Clavijo, J., Oakley, L., Biganzoli, F., Tognetti, P., Barberis, I., Maturo, H. M., Aragón, R., Campanello, P. I., Prado, D., Oesterheld, M., & León, R. J. (2018). Unidades de vegetación de la Argentina. *Ecología Austral*, *28*(1), 040–063. https://doi.org/10.25260/EA.18.28.1.0.399

Roy, D. P., Zhang, H. K., Ju, J., Gomez-Dans, J. L., Lewis, P. E., Schaaf, C. B., ... & Kovalskyy, V. (2016). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, *176*, 255-271.

Souza Jr, C., Firestone, L., Silva, L. M., & Roberts, D. (2003). Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models. Remote sensing of environment, 87(4), 494-506.

Annex 1

Reference maps used in Map Biomas Chaco Collection 4. Some of these maps do not yet have public access because they are in the process of being published. When they are available, its corresponding link will be included to allow the download of the data directly from the original source.

Name of the map	Source/Author	Description	Map data	Link
Land cover of the Argentine Republic (LCCS- FAO)	INTA	Land cover and land use of the Argentine Republic at exploratory scale (E 1:500.000) using the Land Cover Classification System (LCCS-FAO).	2006-2007	Report https://inta.gob.ar/documentos/cobertura-del-suelo-de-la-republica-argentina.-ano-2006-2007-lccs-fao Cartography http://geoportal.idesa.gob.ar/layers/?limit=100&offset=0&title__icontains=Cobertura%20del%20suelo%20de%20la%20Republica%20Argentina.%20A%C3%B1o%202006-2007%20(LCCS-FAO)%20
Maps of agricultural campaigns	INTA	Annual survey of summer and winter campaigns of extensive crops in Northwest Argentina.	2001-2017	https://inta.gob.ar/documentos/monitoreo-de-cultivos-del-noroeste-argentino-a-traves-de-sensores-remotos
Management Unit of the Forest Evaluation System (UMSEF)	Ministerio de Ambiente y Desarrollo Sostenible	Monitoring of native forests in Argentina for the detection, quantification and follow-up over time of natural and/or anthropogenic processes that modify the structure and/or	2002-2018	https://www.argentina.gob.ar/ambiente/bosques/umsef

		extent of these ecosystems.		
First inventory of native forests in Argentina	Secretaría de Ambiente y Desarrollo Sustentable (SAyDS)	Vegetation Unit Monitoring for the native forest sector (E 1:250.000)	2005	
Land cover of Salta and Jujuy (LCCS-FAO)	Infraestructuras de datos espaciales Salta (IDESA).	2013 update of the Land Cover and Current Land Use Map at exploratory scale (E 1: 500.000), using the Land Cover Classification System (LCCS) (Di Gregorio et al., 1998)	2013	http://geoportal.idesa.gob.ar/layers/geonode%3Alccs_2013_niii_final_03_2017
Citrus Crops Production Information (SENASA).	Red de información para el Desarrollo Productivo (RIDES). Ministerio de Desarrollo productivo. Gobierno de Tucumán.	Survey of production units by the National Agri-Food Health and Quality Service (SENASA).	2015	http://rides.producciontucuman.gov.ar/visor/visor/index.html
Blueberry Crop Production Information	Red de información para el Desarrollo Productivo (RIDES). Ministerio de Desarrollo productivo. Gobierno de Tucumán.	Survey of production units by the National Agri-Food Health and Quality Service (SENASA).	2016	http://rides.producciontucuman.gov.ar/visor/visor/index.html
MapBiomas Chaco Colección 5	Proyecto MapBiomas Chaco	MapBiomas Chaco Collection 5 includes annual land use and land cover data for the period 2000 to 2023.	1985-2023	http://plataforma.chaco.mapbiomas.org/map

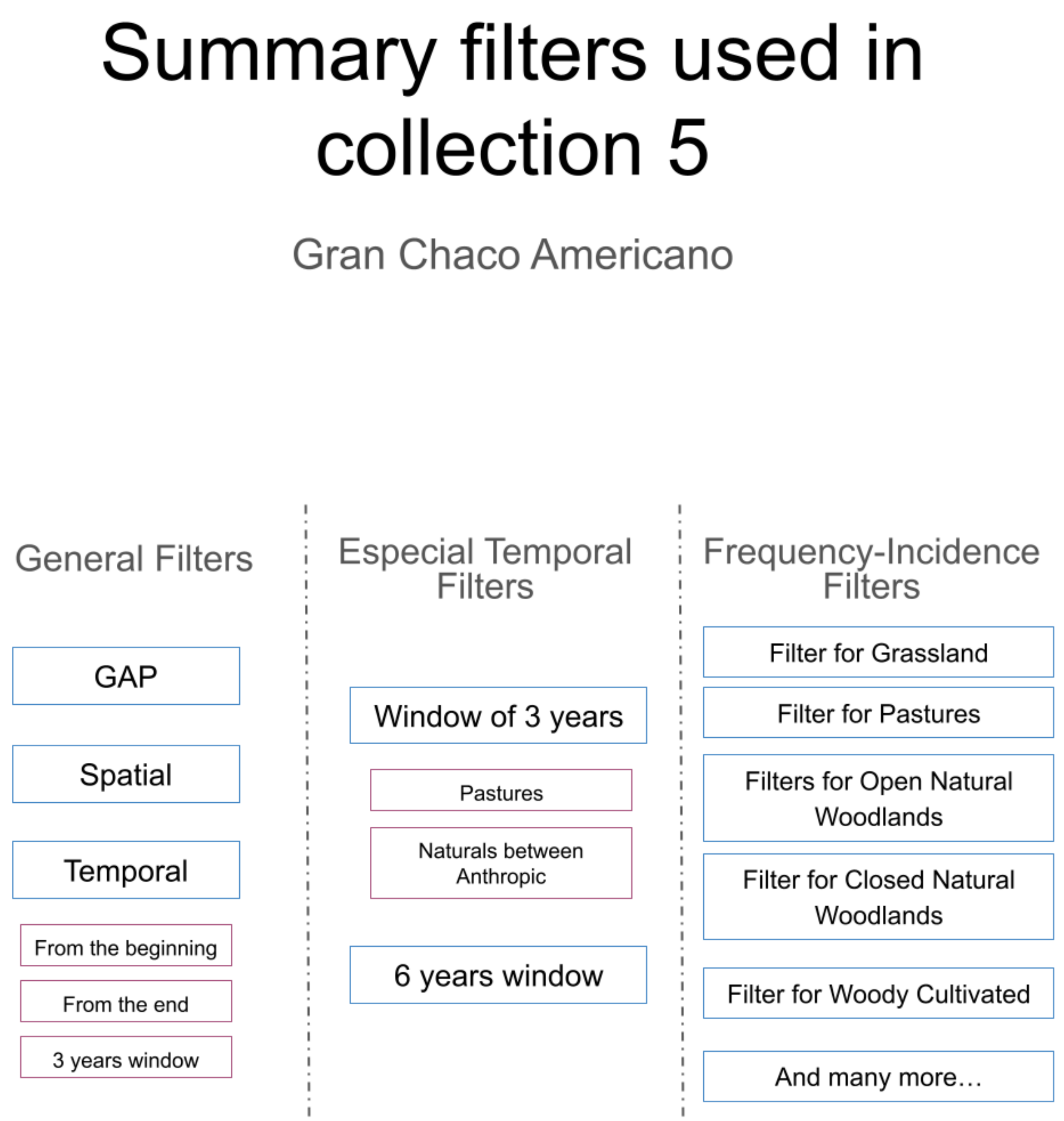
Citrus vector information	RIDES/EEAOC	Shapefile with polygons indicating the location of citrus crops	2008, 2010 2012, 2014, 2016, 2018, 2019, 2020, 2023, 2011, 2012, 2013, 2014, 2015, 2018	https://rides.producciontucuman.gov.ar/visor/visor/index.html
Citrus vector information	INTA	Kml with polygons indicating the location of citrus crops in Salta province	2023	https://geo-nodo01.inta.gob.ar/catalogue/#/dataset/214
Wine crops	SEC	Road survey along Tucuman and Salta provinces	2023	
Walnut	IDECAT	Catamarca province. Points taken from app		https://nodoide.catamarca.gob.ar/visor/index.html?zoom=13&lat=-28.6465&lng=-65.8776&layers=sector-vitivinicola,esri_imagery,sector-organico,Productores-citr%C3%ADcolas-set-2022-NODO
Walnut	RIDES	Polygons in KML format	2019	https://rides.producciontucuman.gov.ar/visor/visor/index.html
Olives	SEC	Points and polygons in shp format	2004-2023	La Bella SRL
Shurblands	IDECOR	LULC from Cordoba province	2021	https://www.mapascordoba.gob.ar/#/descargas

Annex 2

Legend description.

Class Level 1	Class Level 2	Class Level 3	Region	Description
Forests	Closed Forests		Chaco y Yungas	Areas with natural vegetation formed by trees with canopy cover equal to or greater than 65%.
	Open Forests		Chaco y Yungas	Areas with natural vegetation formed by trees with coverage equal to or greater than 20% and less than 65%.
	Flooded Forests		Chaco y Yungas	Transitional areas between pure terrestrial and aquatic systems, where the water table is generally at or near the surface (flooded areas). The natural vegetation cover formed by trees, shrubs or a mixture of both is significantly influenced by water and/or depends on flooding.
Natural Herbaceous and Shrub Vegetation	Closed Shrublands		Chaco y Yungas	Areas with natural vegetation dominated by tall shrubs (greater than one meter and less than three), with coverage equal to or greater than 65%.
	Shrubs and Herbaceous Mosaic		Chaco y Yungas	Areas with natural vegetation formed by trees, shrubs or a mixture of both, with coverage equal to or greater than 5% and less than 20%.
	Grasslands		Chaco y Yungas	Areas with natural vegetation formed by herbaceous plants with cover equal to or greater than 5%. In this category, the presence of woody plants is admitted, but they must have a cover of between 1-5% and 20%.
	Flooded Grasslands		Chaco y Yungas	Transitional areas between pure terrestrial and aquatic systems, where the water table is generally at or near the surface (flooded areas). Natural plant cover formed by herbaceous plants is significantly influenced by water and/or depends on flooding (e.g.: estuaries, wetlands, ravines, swamps and aquatic beds).
Agricultural and Livestock Area	Agriculture	Temporary Crops	Chaco y Yungas	This class includes crops of annual, biennial or perennial herbaceous species for the production of grains or fibers. It encompasses all modalities or sequences of crops in time (rotations or alternations) and space (combined or intercropped crops).
		Perennial Crops	Chaco y Yungas	Areas with perennial crops, mainly blueberries, olive, vine, walnut and citrus.
	Pastures		Chaco y Yungas	Areas with herbaceous crop cultivation for forage purposes (animal production).
	Forest Plantations		Chaco y Yungas	Areas with tree crops (e.g.: pine or poplar plantations).

Annex 3


Summary filters used in collection 5
Gran Chaco Americano
General Filters
GAP
Spatial
Temporal
From the beginning
From the end
3 years window
Especial Temporal Filters
Window of 3 years
Pastures
Naturals between Anthropic
6 years window
Frequency-Incidence Filters
Filter for Grassland
Filter for Pastures
Filters for Open Natural Woodlands
Filter for Closed Natural Woodlands
Filter for Woody Cultivated
And many more…

Especial Temporal Filters

With 3 year window

1 <u>Pasture Special</u>

It only applies between two years of pasture, if the intermediate class is NOT agriculture

It is also applied as an end filter

Pe-Filter	P	P	A	P	P	W	P	P	P		P: Pasture
Post-Filter	P	P	A	P	P	P	P	P	P		A: ACiculture
Years	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	W: Natural Woodland

Especial Temporal Filters

With 3 year window

2 <u>Natural Classes in herbaceous anthropic zones</u>

It only applies between two years of pasture or agriculture, if the intermediate class is natural

Natural Classes: Closed Natural Woodlands, Open Natural Woodlands, Closed Grassland or Flooded Grassland

Pe-Filter	P	P	W	P	A	FG	P	P	P		P: Pasture
Post-Filter	P	P	P	P	A	P	P	P	P		A: ACiculture
Years	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	W: Natural Woodland
											FG: Flooded Grassland

Especial Temporal Filters

With 3 year window

3 Inverted Special

It only applies between two years that are not Woody Cultivated, if the intermediate class is Woody Cultivated

Pe-Filter	NotWC	NotWC	WC	NotWC	NotWC	WC	NotWC	NotWC	NotWC		WC: Woody Cultivated
Post-Filter	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC		
Años	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	

Especial Temporal Filters

With 6 year window

4 Special of six years

The filter applies to two core years, instead of one. It only applies when the two years before and after correspond to the same class

Applied classes: Woody Cultivated

Pe-Filter	WC	WC	AC	P	WC	WC	WC	WC	WC		AC: Agriculture
Post-Filter	WC	WC	WC	WC	WC	WC	WC	WC	WC		P: Pasture
Años	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	WC: Woody Cultivated

Especial Temporal Filters

With 6 year window

5 Especial of six years *2 intermediate*

The filter is applied to two core years, instead of one. It only applies when the two years before and after correspond to a different class than the target class.

Applied classes: Woody Cultivated

Pe-Filter	NotWC	NotWC	WC	WC	NotWC	NotWC	NotWC	NotWC	NotWC		
Post-Filter	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC		
Años	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	WC: Woody Cultivated

Especial Temporal Filters

With 6 year window

6 Especial of six years *3 intermediate*

The filter is applied to three core years, instead of one. Only applies when the two previous years and the following year correspond to a different class than the target class

Applied classes: Woody Cultivated

Pe-Filter	NotWC	NotWC	WC	WC	WC	NotWC	NotWC	NotWC	NotWC		
Post-Filter	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC	NotWC		
Años	1985	1986	1987	1988	1989	1990	1991	1992	1993	... 2021	WC: Woody Cultivated

Filters on Polygons for specifics years and others

#	Filter	Conditions			Result
7	Filter for Pasture 2003	Only **AC**	2002 = P	2004 = P	**P** en 2003
8	Filter for Woody Cultivated	Filter Especial W6 2 intermediate	Filter Especial W6 3 intermediate	Filter W3 inverted	**Mode**
9	Filter for Water in Urban areas	2021 = Ur	2022 = Ur	2023 = W	**Urban**
10	Filter for Woody Cultivated	POLYGONS	Only years with **WC**		**Mode**
11	Filter for Woody Cultivated in agriculture areas	POLYGONS	Only years with **WC**		**Mode AC-P**

Filters of Polygons year and class - specifics*

12 Filter for z17 Bolivia

This filter changes within a polygon only pixels of a particular class from a particular set of years. 4 polygons were made, each one with the class to change and the years

4 x 1

Conditions			Result
Polygons	Only years with **W**	Between 1987 y 1999	**B&D**
Polygons	Only years with **P**	Between 1995 y 1999	**B&D**
Polygons	Only years with **AC**	Between 1995 y 1999	**B&D**
Polygons	Only years with **AC**	Between 1995 y 1999	**B&D**

* Developed for Marcelo Char

Filters on Polygons for specifics years

Filter		Conditions	Result
13 Filter for 8 zone	2 x 1	POLYGON 1 + Only years with **FW** and **P**	**Flood Mode**
		POLYGON 2 + Only years with **P**	**CG**
14 Filter for 5 zone	2 x 1	POLYGON 1 + Only years with **WC**	**CW**
		POLYGON 2 + Only years with **FW** y **CG**	**Flood Mode**
15 Filter for 1 and 6 zones		POLYGONS + Only years with **SC**	**OG**
16 Filter for 1 zone		POLYGONS + Only years with **OW**	**CW**
17 Filter for 23 zone		Mask NOT Urban + Only years with **Urban**	**P**

Filters on Polygons for specifics years

Filter		Conditions	Result
18 Filter for z8	4 x 1	POLYGON 1 + Only years with **SC, MC, P y WC**	**Flood Mode**
		POLYGON 1 + Only years with **P**	**FG**
		POLYGON 1 + Only years with **WC**	**CW**
		POLYGON 2 + Only years with **SC** y **MC**	**Flood Mode**
19 Filter for 3 zone		POLYGONS + Only years with **SC**	**Mode**
20 Filter for 6 zone		POLYGONS + Only years with **CG**	**OW**
21 Filter for 5 zone		POLYGONS + Only years with **W**	**OW**

Frequency-Incidence Filters

Frequency: number of years that a pixel takes the value of a certain class

Incidence: number of times a pixel changes its class over time (39 years) before adding filters

OW	CW	OW	CW	OW	CW	OW
1985	1986	1987	1988	1989	1990	1991

Incidence: 6
Freq CW: 3/7 (43%)
Freq OW: 4/7 (57%)

OW	OW	OW	OW	CW	CW	CW
1985	1986	1987	1988	1989	1990	1991

Incidence: 1
Freq CW: 3/7 (43%)
Freq OW: 4/7 (57%)

Frequency Filters for specifics years

	Filter	Conditions	Frequencies*	Result
22	Filter for Woodlands and Pasture in Woody Cultivated	Only years with **CW**, **OW** y **P**	WC **>** 80%	**WC**
23	Filter for Water in Natural areas	Only years with **W**	CW + OW + CG + OG + W **>** 75%; W **>** 1%	**Not Flood Mode**
24	Filter for antropics in natural areas	Only years with **WC**, **P** y **AC**	AC + P + WC **<** 6%; AC + P + WC **>** 0%; FG + FW + OW + CG + OG + SG **>** 90%	**Mode**
25	Filter for Water in natural areas 2	Only years with **W**	FG + FW + OW + CW + CG + OG + SG **>** 90%; W **<** 6%; W **>** 0%	**Mode**
26	Filter for Water in natural areas 3 (idem previous)	POLYGONS	Filter for Water in natural areas 2	**Not Flood Mode**

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Frequency Filters on Polygons for specifics years

#	Filter	Frequencies*	Conditions	Result
27	Filter for 1 y 6 zones CW (2 x 1)	CW + CG > 3%	POLYGON 1 + Only years with **OW**	**CW**
		CW + CG > 3%	POLYGON 2 + Only years with **OW**	**CW**
28	Filter for Grassland in Agriculture	CG > 1%	POLYGONS + Only years with **CG**	**AC-P Mode**
29	Filter for 14, 1 y 6 zones	AC + P > 1%	POLYGONS + Only years with **AC y P**	**NOT AC-P Mode**
30	Filter for AC y P in Natural Areas	OW > 6%	POLYGONS + Only years with **CW**	**OW**
31	Filter for OW y Pi in agricultural areas	OW + FG + P + AC > 75%	POLYGONS + Only years with **OW** y **FG**	**AC-P Mode**

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Frequency Filters on Polygons for specifics years

#	Filter	Conditions	Frequencies*	Result
32	Filter for OW y WC Flooded Grassland	POLYGONS + Only years with **OW** y **WC**	OW + FG + WC > 90%	**FG**
33	Filter for FG in Agricultural Areas	POLYGONS + Only years with **FG** y **FW**	FG > 1%	**AC-P Mode**
34	Filter for Spare Naturals in Agricultural Areas	POLYGONS + Only years with **SG** y **SW**	SW + SG + P + AC > 75%	**AC-P Mode**
35	Filter for Salt Flats and Beaches and Dunes (2 x 1)	POLYGON 1 — Only years with **NVA y B&D**	NVA + B&D + Sal > 80%	**Sal**
		POLYGON 2 — Only years with **NVA y Sal**	NVA + B&D + Sal > 80%	**B&D**

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Frequency-Incidence Filters for specifics years

			Conditions				Result
			Frequencies*				
36	Filter for Open Woodland - Closed Grassland	2 x 1	I: 4	Only years with CG	OW + CG > 90%	OW > 70%	OW
			I: 4	Only years with OW	OW + CG > 90%	CG > 70%	CG
37	Filter for Open Woodland - Closed Woodland	2 x 1	I: 4	Only years with CW	OW + CW > 90%	OW > 55%	OW
			I: 4	Only years with OW	OW + CW > 90%	CW > 55%	CW

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Frequency-Incidence Filters for specifics years

			Conditions				Result
			Frequencies*				
38	Filter for Open Woodland - Sparse Woodland	2 x 1	I: 4	Only years with SW	OW + SW > 90%	OW > 55%	OW
			I: 4	Only years with OW	OW + SW > 90%	SW> 55%	SW
39	Filter for Open Woodland in Closed Grassland		I: 3	Only years with **OW**	CG > 60%	CG + OW > 80%	CG
40	Filter Pastizales y Leñosa Dispersa		I: 7	19-23 **NO** AC o P	CG + OG + SG + OW + SW > 90%	AC + P < 15%	Mode

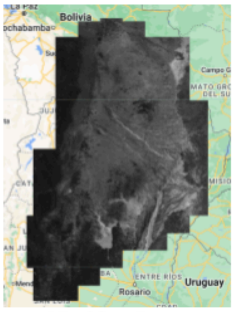
* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Mask of probability of flooding

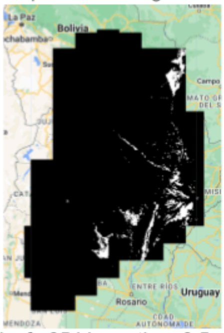
A mask is made that indicates the average probability of a pixel being floodable every year. This includes the Woody Flood and Grassland Flood classes. Different masks can be made, in this case we use the 95th percentile ofl. These layers were calculated from the maps of collection 3.

In turn, a threshold is defined in this type of filters. For example, it could be that the previous mask is greater than 0.5, this means that in the annual average of the series the pixels had values above 50%. In this case we would find sites that half of the years or more were classified as floodable in collection 3, with a 95% probability of being correct. That is to say, it is very likely that these are flood-prone areas.

This threshold can also be used to choose those pixels that have a probability less than a certain percentage, and in general we would be choosing pixels that are unlikely to be floodable.



p95 of flooding



Mask of p95 bigger than 0,5

Frequency Filters + Mask of Probability of flooding + Application in specifics years

41 Filter for Floods in natural areas

Remove pixels from floodable classes on sites with non-floodable natural classes

Conditions

Frequencies*

- CW + OW + SW + CG + OG + SG **>** 60%
- WC + P + AC + CA **<** 5%
- CW + OW + FW **>** 90%

prob p95 Flood **<** 0.45

- Only years WITH FW y FG
- Only years WITHOUT WC, P, AC y CA

Result

Mode

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Frequency Filters + Mask of Probability of flooding + Application in specifics years

42 <u>Filter for Floods in atropics areas</u>

Remove flood class pixels from sites with anthropic classes

Frequencies*	Conditions		Result
CG + OG > 35%	prob p95 Flood < 0.45	1985 NOT applied	AC-P Mode
FG < 30%		Only years WITH FW y FG	
FG > 0%		Only years WITHOUT OW y CW	
WC + P + AC + CA < 40%			
WC + P + AC + CA > 0%			

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.

Filters varius + Mask of Probability of flooding

Filter	Conditions (Frequencies*)				Result
43 <u>Filter for antropics in Flooded areas</u>	WC + P + AC + W + FG > 85%	FG < 75%	W < 15%	prob p95 Flood > 0.50	**FG** the 36 years
44 <u>Filter for FG in natural areas</u>	Only years with FG	CW + OW + CG + OG + FG > 75%	FG < 60%; FG > 1%	prob p95 Flood < 0.45	**NOT Flood Mode**
45 <u>Filter for Woodlands in Flooded areas</u>	I: 4	Only years with **CW**, **OW** y **SW**	CW + OW + SW+ FG + FW > 95%	prob p95 Flood > 0.5	**Flood Mode**
46 <u>Filter for Woodlands in Flooded areas</u>	I: 4	POLYGONS + Only years with **CW**, **OW** y **FW**	CW + OW + SW + FW FG > 95%	prob p95 Flood > 0.5	**Flood Mode**
47 <u>Filter for antropics in FG</u>	POLYGONS	AC + P + WC + FG + Ag > 85%	FG > 75%; AC < 15%	prob p95 Flood > 0.5	**FG** the 36 years

* Frequencies can be worth 0%. That is, one of the classes may not be present in cases where the frequencies are added.