

MAPBIOMAS

Proyecto de Mapeo Anual  
de Uso y Cobertura del Suelo

# MAPBIOMAS CHACO GENERAL “HANDBOOK”

ALGORITHM THEORETICAL BASIS DOCUMENT  
(ATBD)

COLLECTION 1.0

VERSION 1.0

## **Team**

de Abelleira, Diego; Banchemo, Santiago; Verón, Santiago; Volante, José; Ortiz,  
Edder; Elena, Hernán; Mosciaro, María Jesús

---

# Contents

<b>1</b>	<b>Executive Summary</b>	<b>2</b>
<b>2</b>	<b>Scope and content of the document</b>	<b>2</b>
<b>3</b>	<b>Overview and Background Information</b>	<b>2</b>
3.1	Context and Key Information . . . . .	2
3.1.1	Remote Sensing Data . . . . .	2
3.1.2	Google Earth Engine and MapBiomass Computer Applications	2
3.2	Historical Perspective: Existent Maps and Mapping Initiatives . . . .	3
3.2.1	International land cover data . . . . .	3
3.2.2	National land cover data . . . . .	4
3.2.3	Regional land cover data . . . . .	5
<b>4</b>	<b>Methodological description</b>	<b>6</b>
4.0.1	Agreement of LULC class definition . . . . .	6
4.0.2	Generation of Landsat based Feature Space . . . . .	7
4.0.3	Generation of training samples . . . . .	10
4.0.4	Random Forest Classification . . . . .	12
4.0.5	Temporal and Spatial filtering . . . . .	12
4.0.6	Validation Process . . . . .	12
<b>5</b>	<b>Map Collections and Analysis</b>	<b>13</b>
5.0.1	Map Collections . . . . .	13
5.0.2	Transition analysis . . . . .	13
5.1	Map Collections and Analysis . . . . .	13

---

# 1 Executive Summary

MapBiomias Chaco is the initiative of a collaboration network made up of non-governmental organizations, universities and companies from Argentina, Brazil and Paraguay organized with the aim of using quality technology and lower cost to produce annual series of maps (from 2010 to 2017). The working network of MapBiomias Chaco is formed by the developers of the initiative in Brazil, by the National Institute of Agricultural Technology (INTA) in Argentina and by Guyra Paraguay.

## 2 Scope and content of the document

The objective of this document is to describe the theoretical basis, relevance and methods applied to produce annual land use and land cover (LULC) maps in the Gran Chaco Americano region from 2000 to 2017, representing the MapBiomias Chaco Collection 1.

## 3 Overview and Background Information

### 3.1 Context and Key Information

This section addresses complementary contextual and key information relevant to the understanding of the MapBiomias products and methods used to generate the map collections.

#### 3.1.1 Remote Sensing Data

#### 3.1.2 Google Earth Engine and MapBiomias Computer Applications

This project follows the MapBiomias image processing chain that is based on Google technology, which includes image processing in cloud computing infrastructure, programming with Javascript and Python via Google Earth Engine, and data storage using Google Cloud Storage. Google Earth Engine is defined by Google as: “a platform for petabyte-scale scientific analysis and visualization of geospatial datasets, both for public benefit and for business and government users.”

The MapBiomias project has developed the following computer applications based on Google Earth Engine:

- 
- Javascript scripts - these computer codes were written directly in the Google Earth Engine Code Editor and were used to prototype new image processing algorithms and test large-scale image processing. Most of the pre-processing and image classification of Collection 1 were written in Javascript.
  - Python scripts – This category of code was used to optimize image processing of large datasets in Google Earth Engine. In addition, the post processing of Collection 1 with the the spatial and temporal filters were applied through the Earth Engine Platform with the Python API scripts.
  - Samples Validation Collect (SVC): a front-end web application based on the Code Editor was developed to allow image analysts to collect and interpret sample points (i.e.pixels) by visual interpretation of Landsat color composite images. The main application of the SVC is to derive reference LULC classes for accuracy assessment.
  - *chaco.mapbiomas.org* (Dashboard). The web-platform from the MapBiomias initiative that presents the Landsat image mosaics and its quality, land cover and the annual maps of land use belonging to Collection 1, transitions analysis, statistics, and all the methodological information (i.e. the ATBD), tools, scripts, and accuracy analysis. All the maps and Landsat mosaics of the MapBiomias Collections are publicly available to download at the MapBiomias website.

## 3.2 Historical Perspective: Existent Maps and Mapping Initiatives

### 3.2.1 International land cover data

There are several global LULC initiatives with different temporal and spatial scales. For example, Globeland30 (Chen et al. (2015)) is a Landsat based (30 m resolution) global map generated for specific times (year 2000 and year 2010). In turn, GlobCover (Arino et al. 2012) generated a land cover map at 300 m resolution. Nevertheless these global initiatives frequently show inconsistencies when used at local or regional level. Therefore high temporal and spatial resolution maps based on local information are still needed.

---

### 3.2.2 National land cover data

In Argentina, the information related to land use and land cover is characterized by its scarcity, diversity of scale, and poor data quality. This limits the provision of updated information to decision makers for designing environmental policies and generating knowledge about the factors that condition the sustainability of agro-ecosystems. Agricultural production data (acreage and yield) comes from subjective surveys with unknown accuracy level. This information is reported at department level and therefore hampers detailed and updated territorial synopsis on land use dynamics. Conversely, the environmental diversity of Argentina and its territorial extension are factors that hinder the possibilities to get updated and reliable information frequently enough.

The National Institute of Agricultural Technology (INTA) published the first map on land cover in Argentina using exploratory scale (Volante et al. (2010)), being the first national map of this type. This cartography information was produced from field data and visual interpretation of Landsat and MODIS images, which involved great effort and many hours of work. In addition, an important effort was made to achieve a national legend compatible with other initiatives in the region (e.g. Uruguay). For this purpose, the FAO legend building system called Land Cover Classification System (Di Gregorio (2005)) was used. In 2013 an attempt was made to update it. Unfortunately though, this intention was unsuccessful due to fieldwork costs and interpretation time demand.

In Paraguay, efforts have been made to generate maps of land use and land cover in different time periods. Within the framework of the Project "Development of methodologies for monitoring carbon stored in forests for REDD+ in Paraguay", executed jointly by the Forest and Forest Products Research Institute of Japan (FFPRI) and the Forestry Engineering Career (CIF/FCA/UNA), a coverage map of the Republic of Paraguay was prepared for 2011. The methodology was based on the analysis and interpretation of Landsat 5 images, carrying out supervised classification and geoprocessing techniques. The minimum mapping unit was 100 hectares, due to existing information on the study area, time and available resources (FCA/CIF/FFPRI (2013)).

Within the framework of the Global Land Cover Facility (GLCF) research "Changes in Paraguay's subtropical forests during the 1990s", maps of changes in forest cover were generated through the classification of Landsat ETM and TM images. This product was developed by professionals from the Department of Geography and the Institute of Advanced Computer Studies at the University of Maryland in collaboration with Guyra Paraguay's technical team.

---

In order to obtain national and official deforestation data needed to generate a reference level of forest emissions, the Joint National Programme on "Reducing Emissions from Deforestation and Forest Degradation" (UN-REDD+ Paraguay NCP) prepared historical deforestation maps for the period 2000-2005-2011-2015 through supervised classification of Landsat images, identifying forest cover (forest, not forest) (REDD (2016)). Subsequently, in order to continue the process of generating cartographic information on the state of forest cover, the National Forestry Institute (INFONA) has launched the Land Use Change Map 2016-2017 in Paraguay.

### 3.2.3 Regional land cover data

For the Chaco region, there is an initiative developed by Faculty of Agronomy (University of Buenos Aires), INTA and Red Agroforestal Chaco Argentina<sup>1</sup> that generates annual maps of deforestation at 30 m resolution based on visual observation of Landsat imagery. In addition, Guyra generates deforestation maps for the Chaco region. Regional maps of land cover changes were developed also through the ONU-REDD project for land use planning in the Argentine Chaco.

There have been several regional mapping efforts the Paraguayan Chaco. Among the first initiatives is the map of Vegetation and Land Use of the Western Region of Paraguay in 1986/1987. It contains the description of the categories of Chaco vegetation defined according to the classification system used. This work was carried out in cooperation with CIF/FCA/UNA and the German Technical Cooperation (GTZ) (FCA/CIF/FFPRI (2013)).

Shortly thereafter, within the framework of the "Chaco Environmental Systems" Project, the Vegetation Map of the Paraguayan Chaco was prepared on a scale of 1:250,000 using Landsat TM images as inputs. This initiative was carried out in cooperation with the former Environmental Management Directorate (SSERMA/DOA) and the German Federal Institute for Geosciences and Natural Resources (BGR).

More recently, research led by the Department of Geography of Kansas State University generated annual coverage maps with the objective of analyzing the Land-cover change in the Paraguayan Chaco between 2000 and 2011, this study was carried out using MODIS images with a spatial resolution of 250m (Caldas et al. (2015)).

Likewise, Guyra Paraguay has published the Reports of deforestation in the Gran Chaco Americano area (which includes territories of Bolivia, Argentina and Paraguay) since 2010. The report includes information on deforestation in the Paraguayan Chaco, in terms of land use change, fires and floods. Documents containing tabular

---

<sup>1</sup><http://www.monitoreodesmonte.com.ar/>

data as well as files containing spatial information (kmz format, compatible with Google Earth) are available online. These products are generated on a monthly basis in order to monitor and compare the loss of forest cover and the behavioural pattern of the activity (REDD (2016)).

## 4 Methodological description

We here used satellite imagery from Landsat 5, 7, and 8 together with ground truth data determined by visual inspection of high-resolution imagery to train a random forest classifier and generate annual land cover maps of the Gran Chaco Americano from 2010 to 2017. This required: 1) Class definition; 2) Feature Space generation; 3) Collection of training samples; 4) Random Forest Classification; 5) Temporal and Spatial Filtering; 6) Validation (Figure 1).

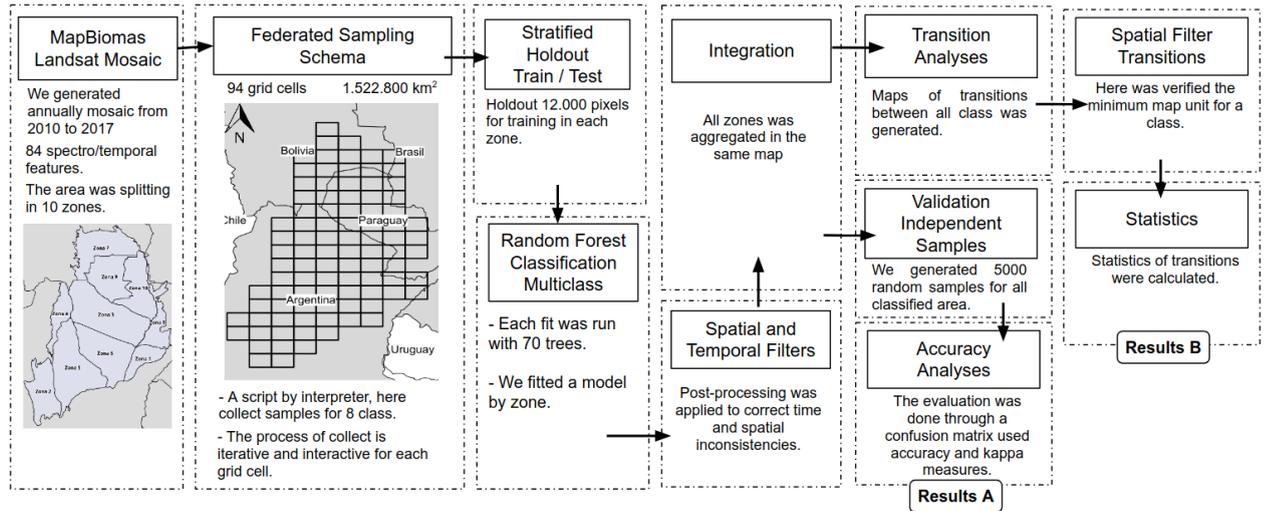


Figure 1: Methodological steps of Collection 1 to implement MapBiomass algorithms in the Google Earth Engine.

### 4.0.1 Agreement of LULC class definition

Eight classes were defined among INTA and Guyra Paraguay participants following the criteria of Land Cover Classification System LCSS-FAO Di Gregorio (2005) (Table 2). A protocol for the definition of the legend and criteria for visual interpretation was generated and discussed among interpreters of the project.

---

Code	Level	Legend Chaco
1	1	1. Woody Formations
2	2	1.1. Woody Natural Forestry
9	2	1.2. Woody crops
10	1	2. Natural non-woody formation
12	2	2.1. Herbaceous
13	2	3. Scatter vegetation
14	1	4. Agricultural Use
15	2	4.1. Pasture
18	2	4.2. Agriculture
22	1	5. Non-vegetated area
26	1	6. Water bodies
27	1	7. Not Observed

Figure 2: MapBiomass Chaco mapped classes follow LCSS–FAO.

#### 4.0.2 Generation of Landsat based Feature Space

We used Google Earth Engine to process Top of Atmosphere (TOA) reflectance images from USGS Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI sensors intersecting the boundaries of the Chaco region between 2010 and 2017. Image availability with cloud covers less than 60 percent differed between WRS tiles as well as between years. We additionally removed cloud, shadow, water or dark outliers pixels from each images using pixelQA information and the methodology proposed by Carson Stam and adapted by Ian Housman. Thus, the Feature Space included a) reflectance and thermal bands, b) spectral indexes, c) fractional cover obtained by spectral unmixing Souza Jr and Barreto (2000) and d) indexes derived from fractional cover (Table 1).

For each of the above mentioned features we calculated 3 different composites based on 2 different types of image grouping. The first type was based on annual (january 1st - december 31st) information while the second on NDVI values (high: NDVI above 50 percentile, low: below 50 percentile). Thus, we reduced each of the 3 groups of images into a median, amplitude and standard deviation 30 m resolution composite on a per-pixel basis. From the resulting 225 composites -i.e. 25 variables calculated over 3 image sets by means of 3 metrics- we selected 84 (Table 1) as inputs for the classifications.

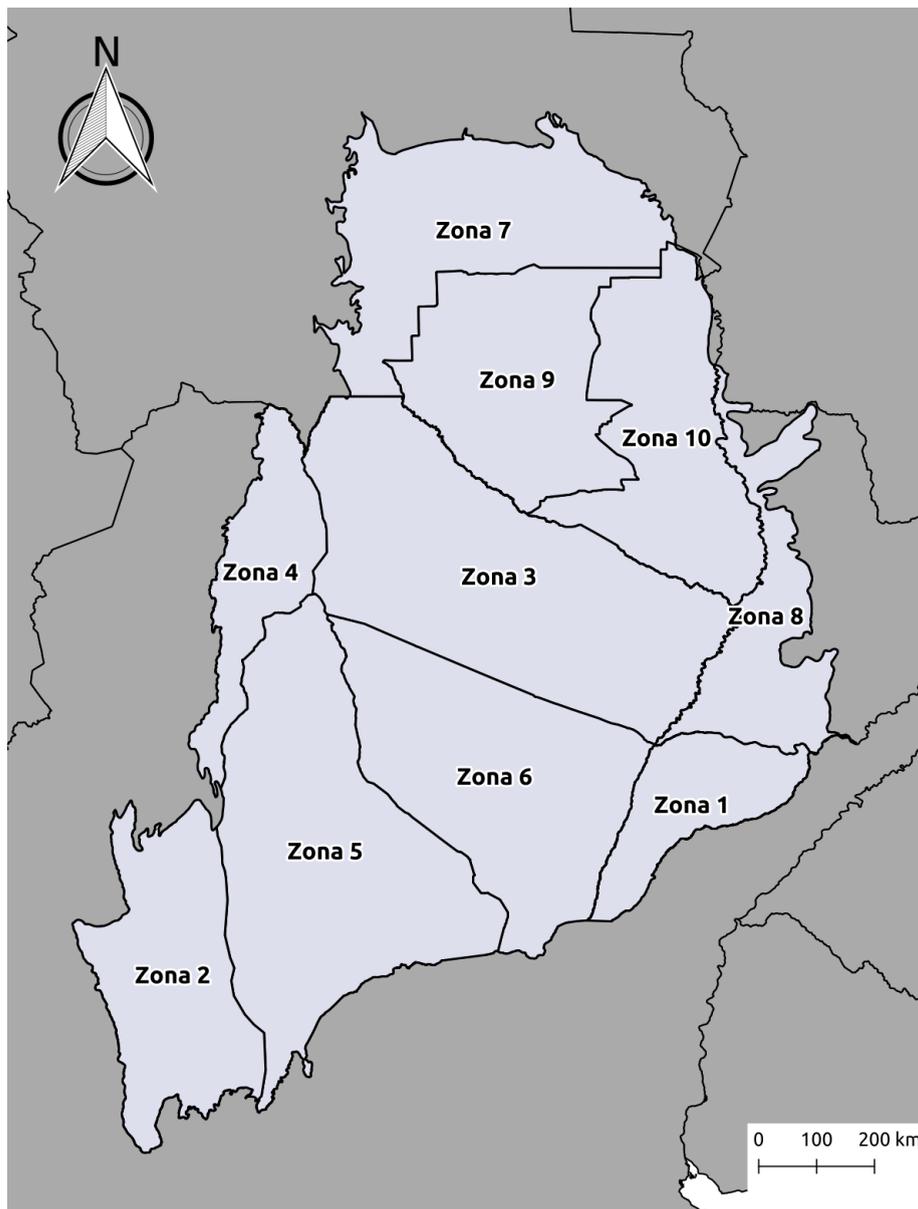


Figure 3: Mapbiomas Chaco subregions used for Feature Space generation and training of Random Forest Classification algorithms.

Table 1: List, description and reference of bands, fractions and indices available in the feature space.

		Reducer					
Band or Index Name	Formula	Median year	Median dry	Median wet	Amplitude	Std. Dev	
Bands	blue	B1 (L5 e L7); B2 (L8)	X				
	green	B2 (L5 e L7); B3 (L8)	X				
	red	B3 (L5 e L7); B4 (L8)	X				
	nir	B4 (L5 e L7); B5 (L8)	X				
	swir1	B5 (L5 e L7); B6 (L8)	X				
	swir2	B7 (L5); B8 (L7); B7 (L8)	X				
	temp	B6 (L5 e L7); B10 (L8)	X				
Spectral Indexes	NDVI	$(nir - red)/(nir + red)$	X	X	X	X	X
	EVI2	$(2.5 * (nir - red)/(nir + 2.4 * red + 1))$	X	X	X	X	X
	CAI	$(swir2/swir1)$	X		X		
	NDWI	$(nir - swir1)/(nir + swir1)$	X	X	X	X	X
	GCVI	$(nir/green - 1)$	X	X	X	X	X
	Hall_cover	$(-red * 0.017 - nir * 0.007 - swir2 * 0.079 + 5.22)$	X	X	X	X	X
	PRI	$(blue - green)/(blue + green)$	X		X		
	SAVI	$(1 + L) * (nir - red)/(nir + red + 0,5)$	X		X		
Fractions	gv	fractional abundance of green vegetation within the pixel	X	X	X	X	X
	npv	fractional abundance of non-photosynthetic vegetation within the pixel	X	X	X	X	X
	soil	fractional abundance of soil within the pixel.	X	X	X	X	X
	cloud	fractional abundance of cloud within the pixel	X	X	X	X	X
	shade	$100 - (gv + npv + soil + cloud)$	X	X	X	X	X
Frac. Indexes	gvs	$gv/(gv + npv + soil + cloud)$	X				
	ndfi	$(gvs - (npv + soil))/(gvs + (npv + soil))$	X	X	X	X	X
	sefi	$(gv + npv_s - soil)/(gv + npv_s + soil)$	X	X	X	X	X
	wefi	$((gv + npv) - (soil + shade))/((gv + npv) + (soil + shade))$	X	X	X	X	X
	fns	$((gv + shade) - soil)/((gv + shade) + soil)$	X	X	X	X	X

---

### 4.0.3 Generation of training samples

We selected ground polygons over six different cover types across the region: 1) woody natural forests, 2) woody crops, 3) herbaceous vegetation, 4) scatter vegetation, 5) pastures, 6) agriculture, 7) non-vegetated areas and 8) water bodies. We based our selection on the visual inspection of high-resolution imagery from Google Earth and Landsat images and MODIS NDVI time series. For this purpose we used the Code Editor and visual tools from GEE platform and the Google Earth software. False color LANDSAT median composites were generated specially for visual interpretation considering three periods along the growing season a) dry season composite (from Jun 1st to October 30th), b) wet season composite (from November 1st last year to April 1st current year) and c) annual composite (from January 1st to December 30th). The bands considered for visualization were Near infrared, Shortwave infrared 1 and Red. Thus, we manually picked locations where cover types were relatively homogeneous and remained unchanged over period considered (i.e. 2010-2017). To assure a representative sample, the region was gridded into 94 cells of  $1^\circ \times 1.5^\circ$  (approx. 100 x 150 km) wherein each interpreter had to select a minimum of 10 polygons per class with an average size of 200 pixels, if possible. Finally, all the pixels located within the selected polygons were used as samples to train the classifier.

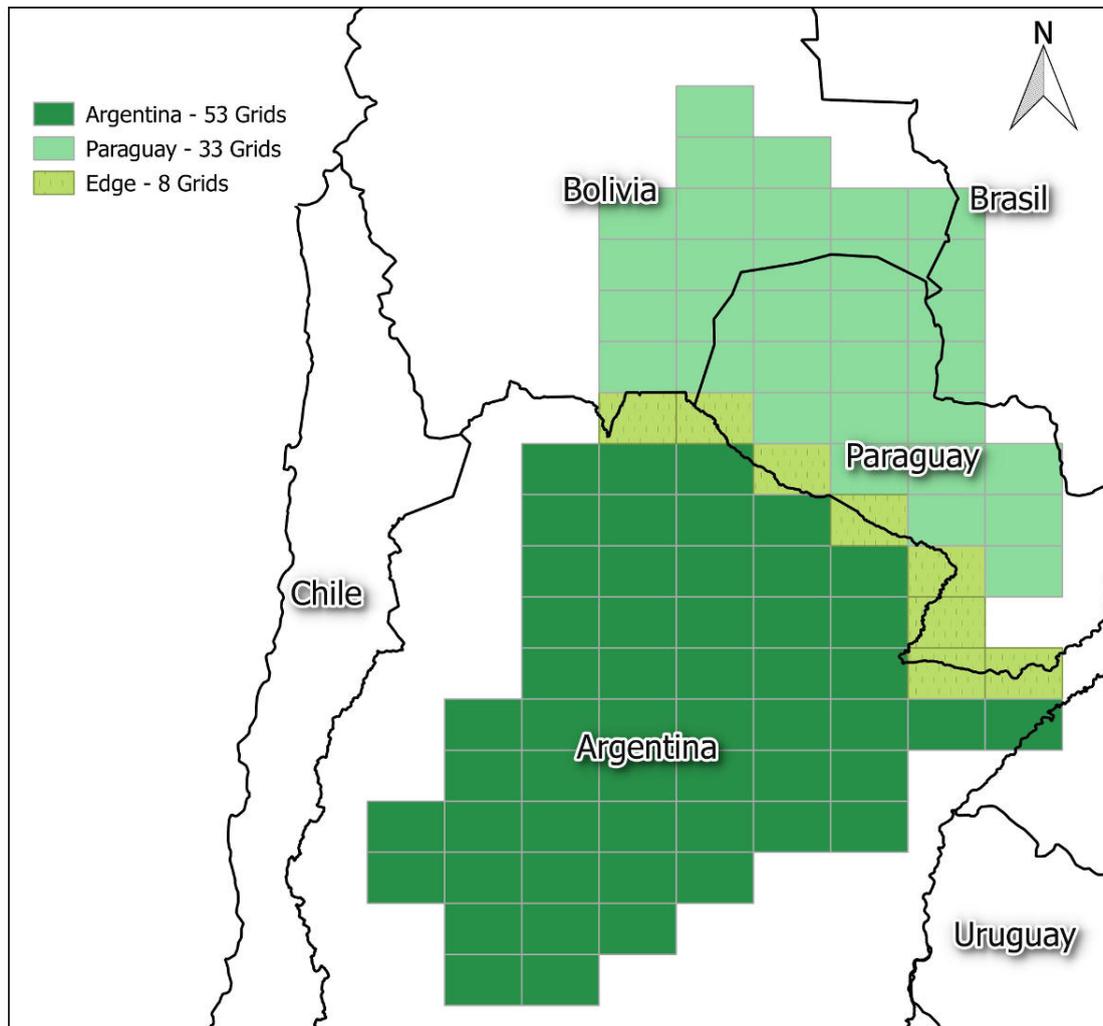


Figure 4: Distribution and number of grid cells along the Gran Chaco Americano, an area of ca. 1.5 Mkm<sup>2</sup>. Colors identify the cells where interpreters from Argentina, Paraguay and both selected the training samples.

Sample selection represented a challenge in terms of: 1) number of people involved, 2) location of people (Buenos Aires, Salta and Mendoza cities in Argentina; Asunción city in Paraguay which were thousands of km apart) and 3) defining a consistent criterion to select the samples. A common set of recommendations was agreed among interpreters based on several rounds of virtual and face to face meetings. Recommendations related to: spatial distribution and balanced number of samples for each class, scale of viewing (i.e. zoom positioning on Google Earth), explicitly assess class homogeneity of pixels and temporal stability of classes. In every case at

---

least two interpreters selected samples at each grid cell. In addition, samples were reviewed by a third -or fourth in some cases- person .

#### **4.0.4 Random Forest Classification**

We applied a Random Forest classifier with 70 trees on each subregion (Figure 3). Random Forest is extensively used algorithm to classify remote sensing images and has been shown to provide higher accuracies and demand less processing time than other techniques. The algorithm was trained on data obtained by extracting the 84 values corresponding to the Feature Space from each of the training samples (i.e. pixels). To that end we first uploaded the Feature Space to GEE Assets.

#### **4.0.5 Temporal and Spatial filtering**

Temporal and spatial filters based on MapBiomass procedure (Souza and Azevedo (2017)) were applied to remove classification inconsistencies. In particular transitions that are very infrequent were removed (for example Natural Woody Vegetation followed by pasture and then followed by Natural Woody Vegetation was changed to three years of Natural Woody Vegetation). The LULC maps of each subregion were integrated based on prevalence rules in order to generate the final Collection 1 map .

#### **4.0.6 Validation Process**

We generated a completely independent set of validation samples. This validation set was based on statistical techniques to define sample points based on the extent and on the number of classes of all Chaco biome. A preliminary analysis was performed based on 5,000 pixel samples to each one of the years based on visual interpretation of Landsat data. Each sample was assign to and inspected by only one interpreter. This activity was based on a web application in Google Earth Engine over Code Editor, developed by INTA Team. The application allowed to classify a given pixel into one of the eight classes defined (Table 1), and an additional one for non-observed cases (i.e. lack of certainty). With this validation data we created a confusion matrix. The global accuracy was calculated for each year. Accuracy assessment analysis was conducted in the map Collection following the practices proposed by Olofsson et al. (2014)

---

## 5 Map Collections and Analysis

### 5.0.1 Map Collections

The MapBiomass Collections produced so far are listed and summarized below:

- Collection 1 - comprised the period of 2010 to 2017 and was based on random forest classification for all themes.

### 5.0.2 Transition analysis

LULC change transition analysis (with spatial filter application) statistics, and spatial analysis with political, watershed, protected areas, and other categorical maps, were generated to understand LULC within these boundaries.

## 5.1 Map Collections and Analysis

The MapBiomass Collections produced so far are listed and summarized below:

- Collection 1 - comprised the period of 2010 to 2017 and was based on random forest classification for all themes.

---

## References

- Caldas, M. M., Goodin, D., Sherwood, S., Campos Krauer, J. M., and Wisely, S. M. (2015). Land-cover change in the paraguayan chaco: 2000–2011. *Journal of Land Use Science*, 10(1):1–18.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., et al. (2015). Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:7–27.
- Di Gregorio, A. (2005). *Land cover classification system: classification concepts and user manual: LCCS*, volume 2. Food & Agriculture Org.
- FCA/CIF/FFPRI (2013). Land cover map paraguay 2011. project "methodology for monitoring carbon stored in forests for redd+ in paraguay". Technical report, FCA/CIF/FFPRI, San Lorenzo, Paraguay.
- REDD, P. O. (2016). Technical document for the elaboration of a map of historical deforestation in paraguay - period 2000 - 2015. Technical report, ONU - REDD Paraguay, Asunción, Paraguay.
- Souza, C. and Azevedo, T. (2017). Mapbiomas general handbook. *MapBiomas: São Paulo, Brazil*, pages 1–23.
- Souza Jr, C. and Barreto, P. (2000). An alternative approach for detecting and monitoring selectively logged forests in the amazon. *International Journal of Remote Sensing*, 21(1):173–179.
- Volante, J. N., Ayesa, J. A., Bubenik, K., Collado, A. D., Ferreyra, E., Lopez, C., Navarro de Rau, M. F., Pezzola, A., and Puente, M. (2010). Cartografía de cobertura del suelo de la república argentina según el sistema de clasificación lccs-fao. año 2007-08. In *Congreso Argentino de la Ciencia del Suelo. 22. 2010 05-06 31-04, 31 de mayo a 4 de junio de 2010. Rosario, Santa Fé. AR.*

---

## Annex

- Vegetation and land use map of the country's Western Region - 1987. CIF/FIA/UNA/GTZ, 1991. Map of vegetation and land use in the Chaco in 1986 and 1987 that contains a categorization of vegetation obtained through classification of Landsat images. 1986/1987
- Vegetation Map of the Paraguayan Chaco. Proyecto SACH/SERMA-DOA/BGR, 1998. Map of the different natural formations of the Paraguayan Chaco obtained through Landsat TM images on a scale of 1:250.000. 1993
- Forest Cover Change Map 1990 - 2000. GLCF, 2009. Maps of change in forest cover in the western region of Paraguay for the period 1985-2001 obtained through classification of Landsat images. 1990 - 2000
- Land-cover change in the Paraguayan Chaco: 2000–2011 . Caldas et al. (2015) Land-cover change in the Paraguayan Chaco: 2000–2011, Journal of Land Use Science, DOI:10.1080/1747423X.2013.807314. Annual maps of land use and land cover using a classification of Modis Terra images (250m) in the years 2000 to 2011 Annual maps of land use and land cover using a classification of Modis Terra images (250m) in the years 2000 to 2011. 2000 – 2011
- Land Cover Map of Paraguay for 2011. FCA/CIF/FFPRI, 2013 . The preparation of the Map has been carried out from the analysis and interpretation of Landsat 5 (TM) satellite images corresponding to the year 2011. Scale 1:250.000. 2011
- Forest cover maps of Paraguay . PNC-ONUREDD+ PY, 2016 . Maps of presence and absence of forest cover in the country. Obtained by Landsat image classification. 2000 - 2005 - 2011 – 2015
- Map of land use change 2016 - 2017 in Paraguay . INFONA 2018 . Map of land use change between 2016 and 2017 obtained by image classification Landsat 8. 2017
- Gran Chaco Americano Land Use Change Change Monitoring . Guyra Paraguay, 2019 . Detection of land use change through monthly monitoring of the forest cover of the Gran Chaco Americano, using Landsat images carried out since 2010. 2010 – 2018