Agriculture and Forest Plantation - Appendix

Collection 8

General coordinator

Bernardo Rudorff

Team

Adriana Moreira

Joelen Silva

Kênia Santos

Luciana Oliveira

Mateus Pinto da Silva

Moisés Salgado

Paulo Teixeira

1 Overview of the classification method

Mapping ‘Agriculture’ and ‘Forest Plantation’ emerged as one of the challenges of the MapBiomas project. The first challenge was in Collection 1, mapping ‘Agriculture’ and ‘Forest Plantation’ from 2008 to 2015 in a short period to prove the innovative concept of the project: the production of cheaper, faster, and updated annual maps of coverage and land use for Brazil’s territory compared to the methods and practices applied so far. Based on the results from Collection 1, Agrosatelite’s team adopted a more appropriate approach for the classification of agriculture. The algorithm developed for classifying annual and semi-perennial agriculture in MapBiomas Collection 2 (2000 - 2016) incorporated each region’s growing season and off-season periods in Brazil. This algorithm selects the Landsat images available in each scene’s specific season period and creates a mosaic from these images. In addition, Collection 2 used the Enhanced Vegetation Index 2 (EVI2) and Crop Enhancement Index (CEI) to train the Random Forest classifier (Breiman, 2001).

In Collection 3, the methodology was reformulated. A new approach to obtain metrics was adopted: the use of reducers (minimum, maximum, median, standard deviation, and quality mosaic) applied to the vegetation indexes and spectral bands. A total of 178 bands were created for each annual mosaic. From these bands, we selected those that presented the classifier’s best response for each class (more details on the selection of the bands are shown in the topics below). This approach has been used in Collections 4, 5, and 6 for the classes mapped by the Random Forest algorithm. Specifically, for MapBiomas Collection 5 and 6, the most important methodological change was using a normalized Landsat series based on Modis data. The normalization of the images provides a series with similar spectral characteristics, thus allowing the use of samples of only one year for training the model and improving the final quality of the classification.

In Collection 6, other improvements were added, especially the addition of new classes, such as soybean class for all MapBiomas temporal series (from 1985 to 2020), rice (irrigated only), coffee and citrus (São Paulo state only), both as beta version, as well as improvement of Other Perennial Crops maps.

In Collection 7 (and 7.1) for the cross-cutting themes ‘Agriculture’ and ‘Forest Plantation’ in the Brazilian territory from 1985 to 2021, were added more improvements, mainly related to the improvements of methods, becoming more robust, as well as the addition of a new class (cotton - beta version).

And in Collection 8, an effort was made to seek the stabilization of the methodologies developed and improved throughout the MapBiomas Collections. The stabilization of methodologies is important to avoid sudden differences in the mapping of classes between the Collections, reducing sources of errors and uncertainties, in addition to supporting the spatio-temporal land use and land cover analysis that is carried out with the MapBiomas Collections.

Thus, with the stabilization of the methodologies of the consolidated classes, it is possible to advance in other important points for the constant improvement of the mappings, such as the detailing of the classes and the expansion of the mapped area. In this sense, in Collection 8, an effort was made to detail another class contained in the Other Perennial Crops class, mapping the Palm Crop class individually. In Figure 1 is presented the evolution of MapBiomas agricultural classes throughout the eight Collections.

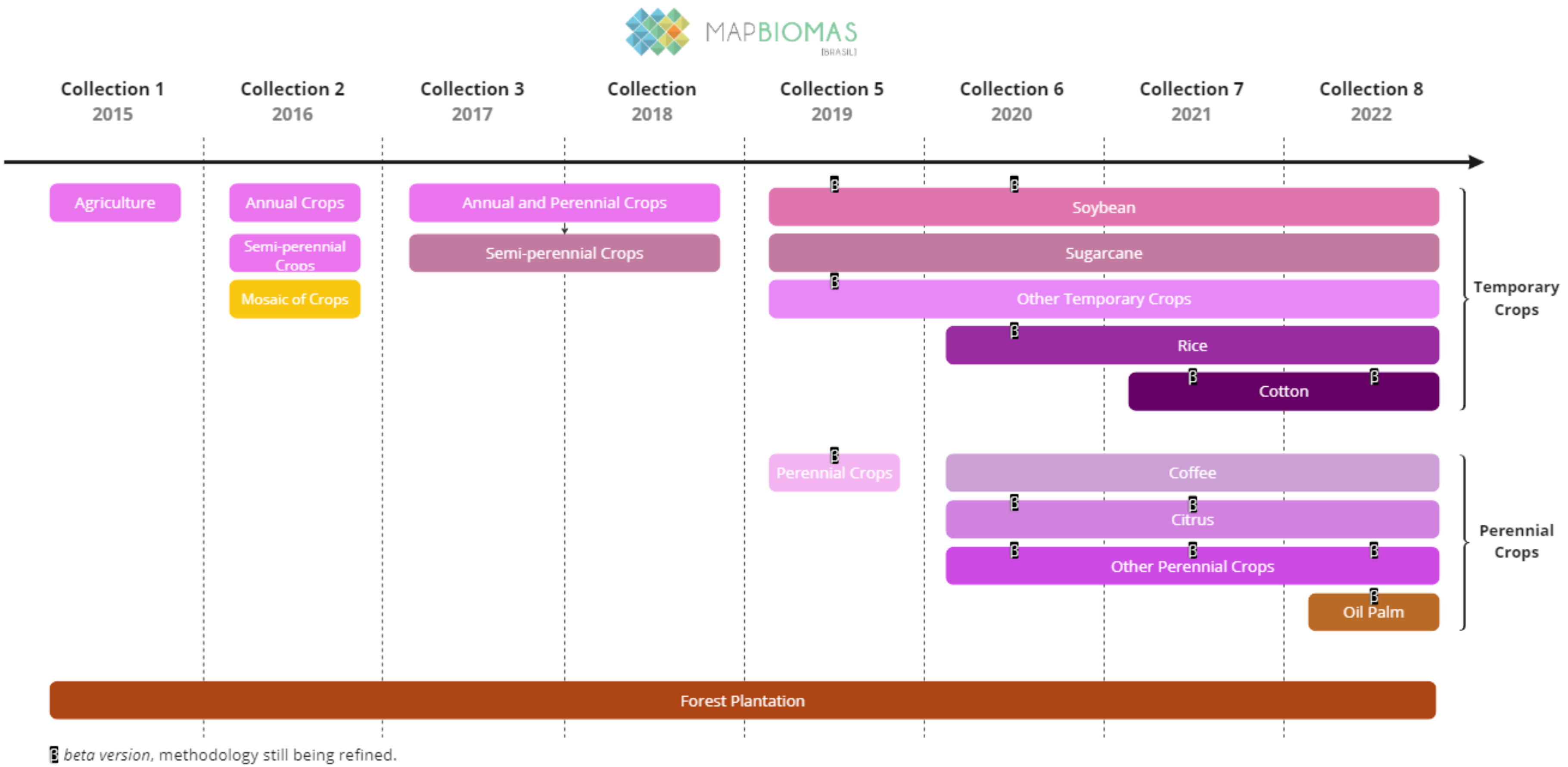


Figure 1: Evolution history of MapBiomas Collections of 'Agriculture' and 'Forest Plantation' (Collections 1-8).

Overall, ‘Agriculture’ classes comprise ‘Temporary Crops’ and ‘Perennial Crops’. In ‘Temporary Crops’ are included important agricultural classes for the Brazilian economy, such as soybeans, sugar cane, rice, cotton, and other temporary crops, a class corresponding to the other temporary crops that are not yet mapped individually by the project.

Regarding the ‘Perennial Crops’, since Collection 6, a great effort was made to seek improvements, especially in the citrus class, with the collection of new reference maps, obtained by visual interpretation using Planet Scope images for Paraná and Minas Gerais states, in addition to the map already produced for São Paulo state. For the coffee class, in Collection 7, the mapping was expanded to Espírito Santo state, to improve the mapping quality, reduce omission areas, and improvements in the sampling approach. In addition, in Collection 8, a new class of ‘Perennial Crops’ was added to the MapBiomas agricultural classes, the oil palm, providing another advance for the detailing of information on Brazilian agriculture.

In the ‘Forest Plantation’ class, since Collection 7.1, improvements to the map were mainly related to a new sampling approach, however, errors of inclusion from other classes were verified, thus, the new sampling approach has been used only to regions where was verified an increase in mapping quality. Thus, in this Collection, ‘Forest Plantation’ mapping was conducted based on a regional approach, to focus on important regions, increase the quality of classification, and avoid misclassifications, following the same methodology applied in Collection 7.1.

2 Classification

The MapBiomas-brazil account in GitHub has all the scripts used to classify ‘Agriculture’ and ‘Forest Plantation’ classes in MapBiomas Collection 8. The repository links are:

- Agriculture:

  https://github.com/mapbiomas-brazil/agriculture/tree/mapbiomas80

- Forest Plantation:

  https://github.com/mapbiomas-brazil/forest-plantation/tree/mapbiomas80

In general, the use of supervised classification via machine learning algorithms has adopted the procedure illustrated in Figure 2.

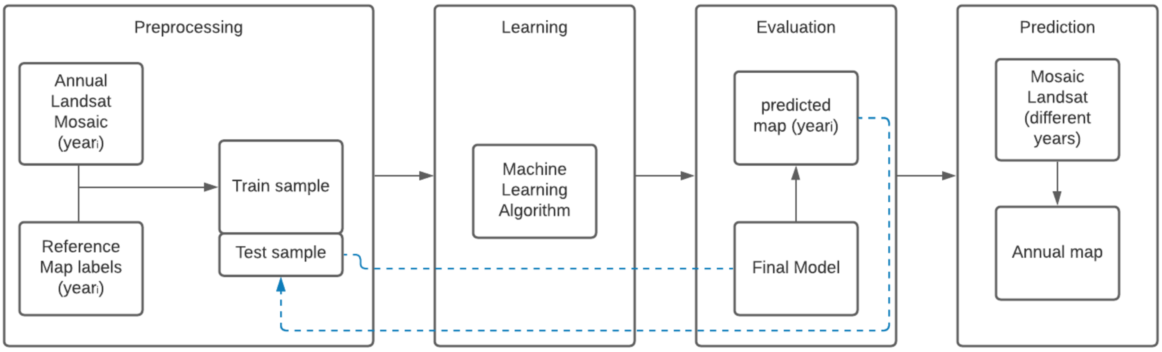


Figure 2. Supervised learning workflow in context of image classification.

The preprocessing step and prediction were the same for both algorithms used in 'Agriculture' and 'Forest Plantation' mapping (*i.e*. Random Forest and Convolutional Neural Network). The learning and evaluation steps were specific according to each of the algorithms. The annual rice and citrus maps were generated using a convolutional neural network (*i.e*. U-Net) and the other classes were obtained using Random Forest.

2.1 Landsat image mosaics

2.1.1 Landsat Images availability

Since in Collection 8 a stabilization of the methodologies was sought, for all consolidated classes, only the last year of the series, *i.e*., 2022, was processed, and only the Palm Crop class was processed for the entire time series. In this sense, for all the classes already mapped in the previous Collections, the images used for the construction of the mosaics come from the Landsat Collection 1. In contrast, only for the oil palm class and the last year of the series of classes, images from the Landsat Collection 2 were used.

Figure 3 shows the availability of images from the Collection 2 Landsat series between 1985 and 2022. It is worth mentioning the greater number of images available from the launch of Landsat 8 and Landsat 9, which provides a greater probability of obtaining cloud-free mosaics.

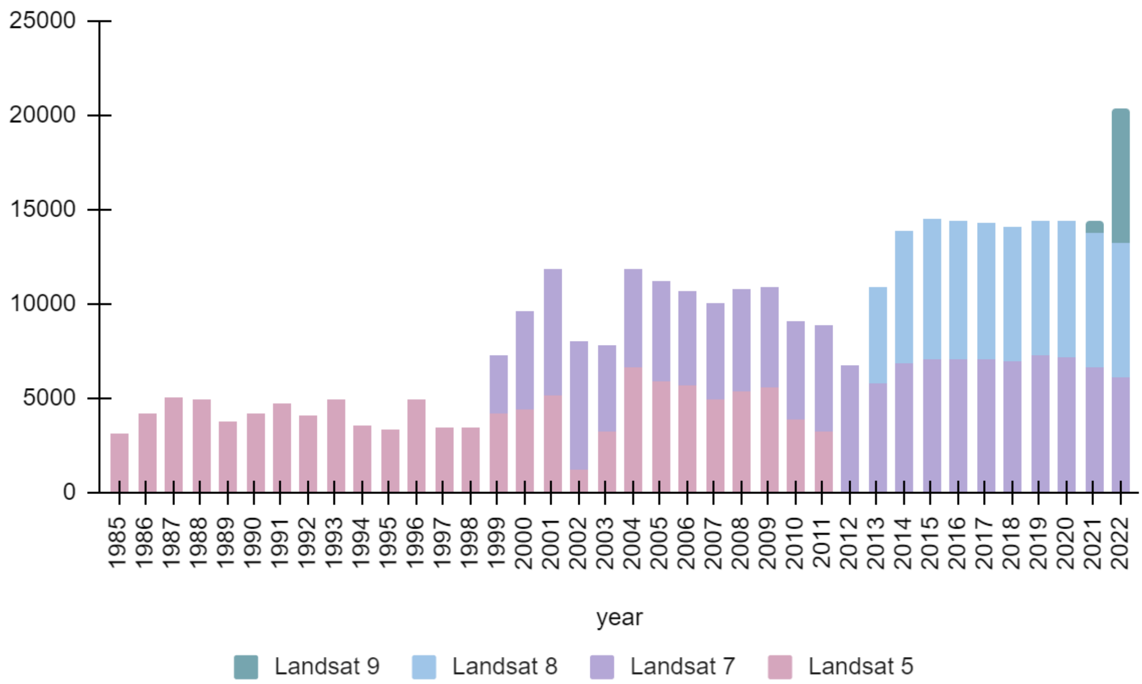


Figure 3. The number of available TOA Landsat Collection 2 images covering the Brazilian territory from 1985 to 2022.**Figure 3.** The number of available TOA Landsat Collection 2 images covering the Brazilian territory from 1985 to 2022.

2.1.2 Image selection

Since Collection 5, a Landsat normalized time series is used, created based on the reflectance data from MODIS, in addition to the Landsat images used in previous collections (available on the Google Earth Engine platform). The normalization of reflectance is an important step to guarantee the spectral similarity of the same land cover types (for more information, see Potapov *et. al.* (2020)). Table 1 shows the collection type used for each class of 'Agriculture' and 'Forest Plantation' classification.

Table 1: Landsat collection type used for each class of 'Agriculture' and 'Forest Plantation' classification.

Level 1	Level 2	Level 3	Level 4	Landsat Collection
			Cotton	Normalized Landsat Collection
			Soybean	Normalized Landsat Collection
Farming	Agriculture	Temporary Crop	Sugar Cane	Landsat ToA Collection
			Rice	Landsat ToA Collection

		Other Temporary Crops	Normalized Landsat Collection
	Perennial Crop	Coffee	Normalized Landsat Collection
		Citrus	Normalized Landsat Collection
		Other Perennial Crops	Landsat ToA Collection
Forest Plantation		Forest Plantation	Normalized Landsat Collection

2.1.3 Definition of the temporal period

To define the best period to compose the mosaics used in the supervised classification of 'Agriculture' and 'Forest Plantation', the seasonal characteristics of each agriculture class were considered to better distinguish the class of interest from the remaining land cover and land use classes. For instance, for different types of agricultural crops and for different regions in Brazil, the growing season can cover different periods of the year, predominantly during the wet period, in most Brazilian regions. Figure 4 presents the development stages of different types of agricultural crops, for a given region where the wet period extends from October to March. According to this example, we can note that for mapping annual crops, the Landsat mosaics require images that cover the period from October to March, while for semi-perennial and perennial crops, we can use images collected throughout most of the year or all year.

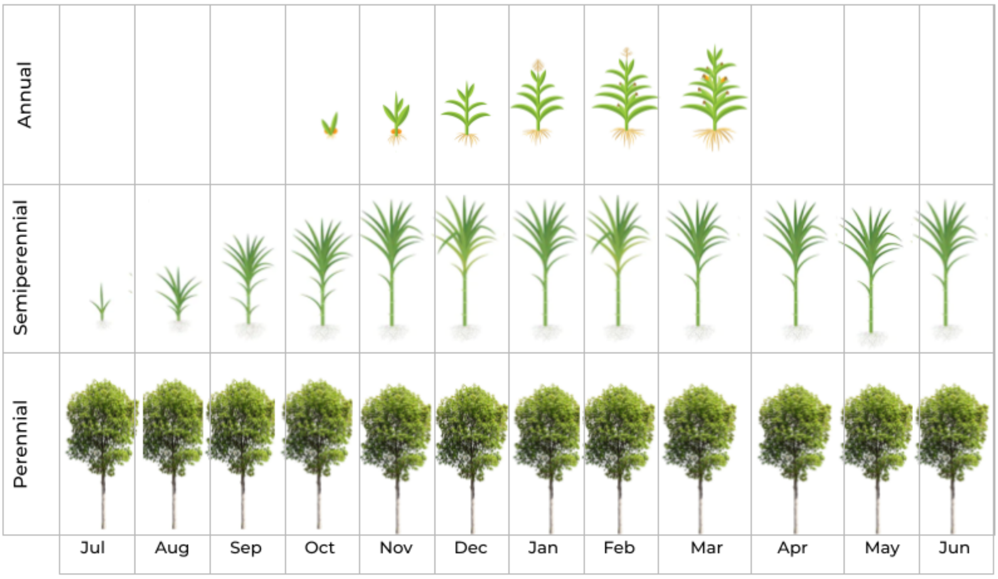


Figure 4. Temporal behavior of crops according to their cycle.

Furthermore, since the Landsat mosaics comprise the growing season, which consequently (especially for annual crops), comprise the wet period, in most Brazil regions, while has a high cloud incidence, it is necessary to use images from the same period from previous years to overcome the challenge of missing images throughout the time series. Figure 5 presents an example of this approach to composing the mosaics.

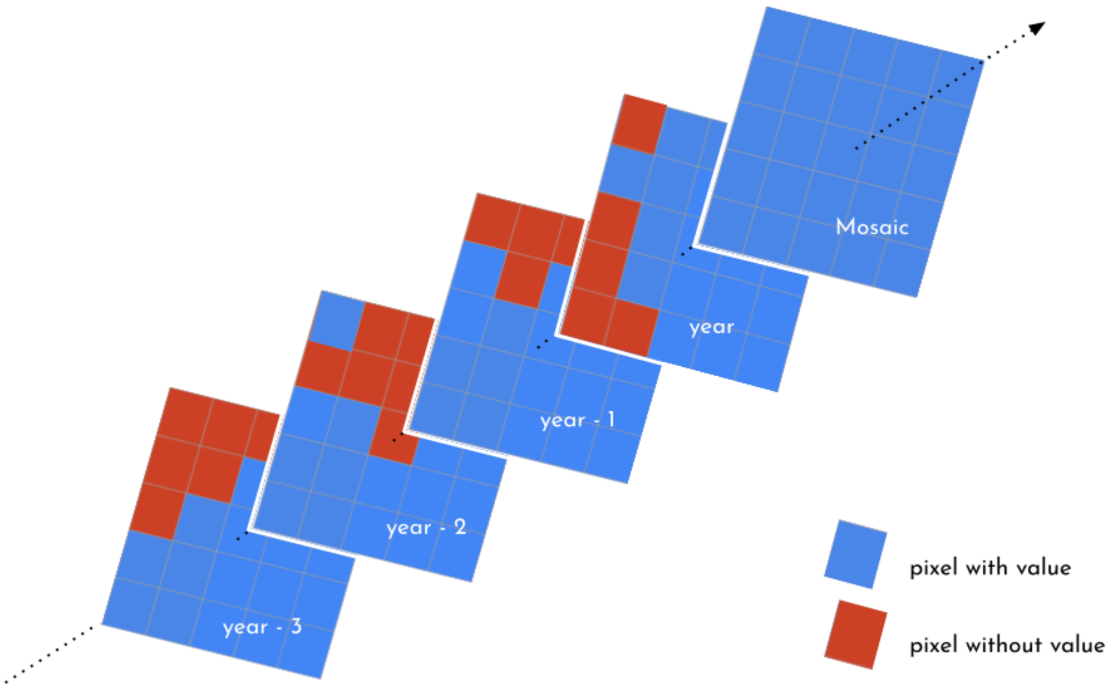


Figure 5. Scheme to compose Landsat mosaic. Pixels in red do not have a valid value (due to cloud and/or shadow incidence), and pixels in blue color have a valid value. This approach seeks to fill pixels in red, replacing them with pixels in blue from the next previous year, to compose a Landsat mosaic with only valid pixel values.

2.1.3.1 Cotton, Soybean, and Other Temporary Crops

In the MapBiomas project, the 'Temporary Crops' correspond to those cultivated in the period called 'first crop' (generally occurring over the summer period). Consequently, the mapping methodologies must consider this period to collect the images to build the mosaic. In addition, it is important to note that among the crop types mapped by the MapBiomas project, the cotton class (*beta version*), mainly occurs in the period called 'second crop', in the largest cotton-producing region in Brazil (Mato Grosso) (ABRAPA, 2020). Thus, as MapBiomas only maps agriculture classes of the first crop period, the cotton mapping from MapBiomas Collection 8 will only reflect the planted area of the less cotton-producing regions.

Until Collection 6, we used a static growing season and off-season periods to compose the mosaics for all classes mapped. However, a temporal static calendar (static dates for the beginning and end of the growing season for all years) can often fail to follow variations in the growing season, according to weather conditions or other factors related to management. Since Collection 7.1, a methodology based on the phenological development of the culture, has been used to obtain the periods of growing season and off-season.

To obtain a one-by-one-year growing season calendar, a time series of EVI2 data calculated from MODIS was smoothed using the Fourier Transform to minimize variations. In the annual curves for each pixel, the vegetative peak value in agricultural regions was identified. However, as the same area can be cultivated up to three times over the year (*i.e*., soybean, corn, and winter wheat), there are cases where the crop with the most evident spectral response occurs in a winter crop period. As the main objective is to map the crops that occur in the summer crop period, the vegetation peaks that occur only in this period will be considered. After the peak month was identified, a reduction by Landsat scene was made to identify the mode of this month in each scene. Then, the peak vegetative month for each year in each scene was obtained, corresponding to the maximum point of the main crop in that scene. As we used MODIS data, the methodology to obtain crop calendar year by year was only possible for the period after 2000. Thus, to obtain the crop calendar to years before 2000, we defined the vegetation peak month as the mode of months obtained between 2000 and 2022.

An example of the methodology to obtain the annual vegetative peaks per Landsat scene is shown in Figure 6.

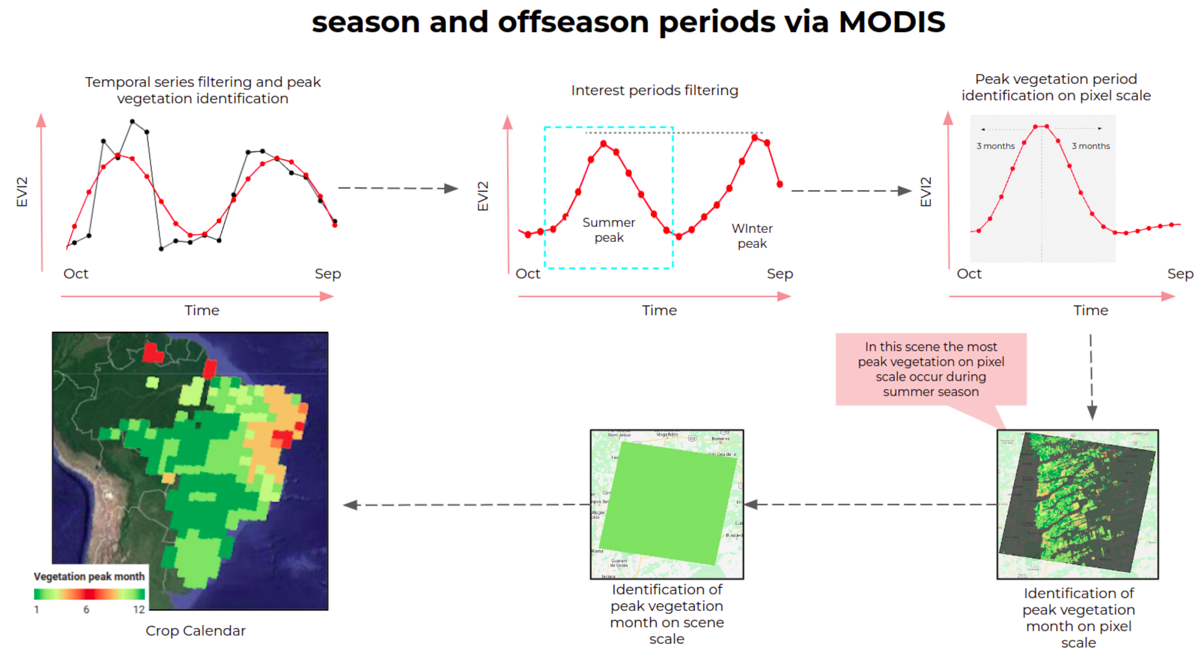


Figure 6. Scheme to obtain vegetation peak month, year by year, to Landsat scene.

Figure 7 presents a growing-season calendar obtained through this methodology, for each Landsat scene, and the comparison with the agriculture calendar from *Companhia Nacional de Abastecimento* (CONAB) to the 2019/2020 crop-year.

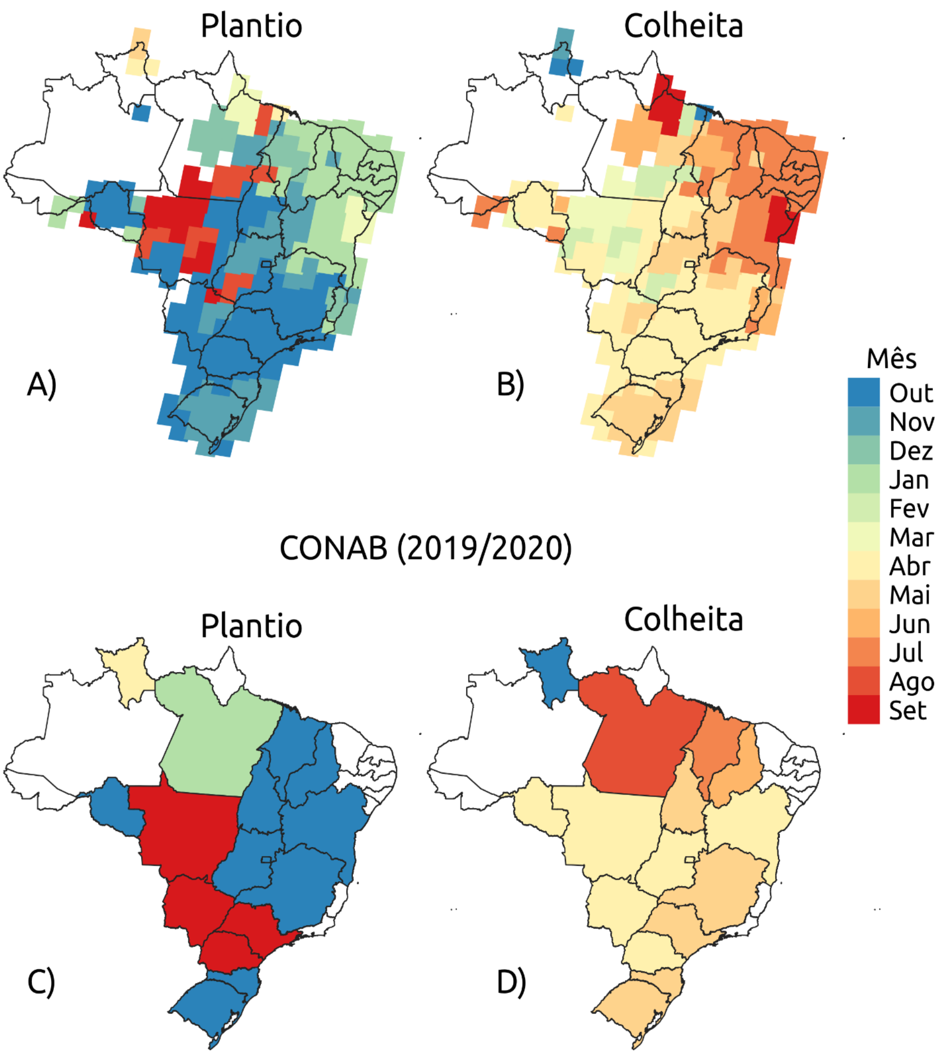


Figure 7. Agricultural calendar (planting and harvest months), for Brazil. A-B) Automated method based on the EVI2 MODIS vegetative peak, C-D) *Companhia Nacional de Abastecimento* (CONAB), for the 2019/2020 harvest.

Thus, the seasonal mosaics for soybean, cotton, and other temporary crops were based on the peak vegetation agricultural crop rotation month information, according to Table 2. We defined the 'growing season' as between 3 months before and 3 months after (+3/-3) the peak vegetation month, 'off-season' as between 5 months before and 3 months before (-5/-3), and 'annual' as between the peak month and 12 months after (+0/+12).

Table 2. Periods used to select mosaic images of cotton, soybean, and other temporary crops in Collection 8.

Period	Start	End
growing season	vegetation peak month -3 months	vegetation peak month -+3 months
off-season	vegetation peak month - 3 months	vegetation peak month - 5 months
annual	vegetation peak month	vegetation peak month + 12 months

2.1.3.2 Sugar cane

For the sugar cane class, we used Landsat mosaics created to highlight intra-annual variations based on bimonthly compositions for the entire country, which were used to select the images according to the periods presented in Table 3.

Table 3. Periods used for selecting mosaic images of sugar cane in Collection 8.

Period	Start	End
growing season 1	12/01/year-1	01/31/year
growing season 2	02/01/year	03/31/year
growing season 3	10/01/year	11/30/year
off-season 1	04/01/year	05/31/year
off-season 2	06/01/year	07/31/year
off-season 3	08/01/year	09/30/year

2.1.3.3 Rice

For rice class, the selection of images was made based on the growing season period according to the year of mapping carried out in each state (Table 4).

Table 4. Periods used for the selection of mosaic images of rice in Collection 8.

State	Start growing season	End growing season	Start off-season	End off-season
Tocantins – TO	04/01/year	07/30/year	08/01/year-1	11/01/year-1
Rio Grande do Sul – RS	10/01/year-1	04/01/year	01/10/year-1	01/01/year
Santa Catarina – SC **Paraná – PR**	10/01/year-1	04/30/year	01/01/year	07/30/year

2.1.3.4 Perennial Crop

Due to the quantity and complexity of perennial crops existing in Brazil (*i.e*., coffee, orange, banana, oil palm), since Collection 6 there is an effort to map each type of 'Perennial Crops' separately. Thus, an effort was made to train the classifier to specific classes, and for Collection 8, it was possible to include a new class in the 'Perennial Crops' classes, the oil palm. In this sense, in Collection 8, 'Perennial Crops' classes were divided into four subclasses: coffee, citrus, oil palm, and Other Perennial Crop. The last one doesn't distinguish between types of crops. For all the 'Perennial Crops' classes, a median of annual mosaic (*i.e*., 01-01-year to 12-31-year) was obtained (Table 5).

Table 5. Periods used to select mosaic images of Perennial Crop in Collection 8.

Period	Start	End
Annual	01/01/year	12/31/year

2.1.3.5 Forest Plantation

For the Forest Plantation class, two periods were defined to compose the Landsat mosaics. These periods cover from January to January of the next year and are presented in Table 6.

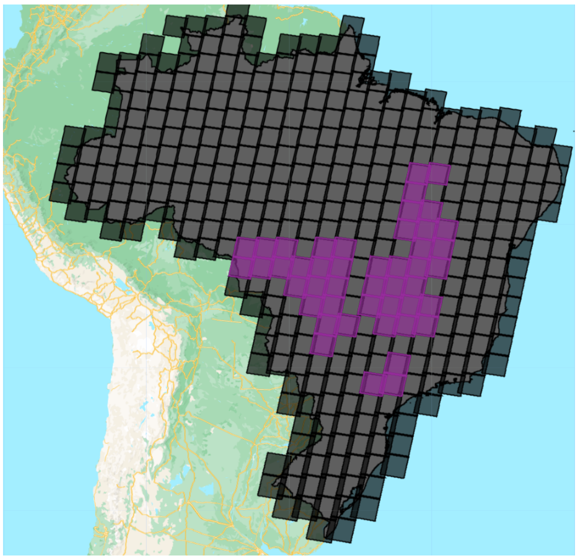
Table 6. Periods used to select mosaic images of "Forest Plantation' in Collection 8.

Period	Start	End
P1	01/01/year	07/01/year

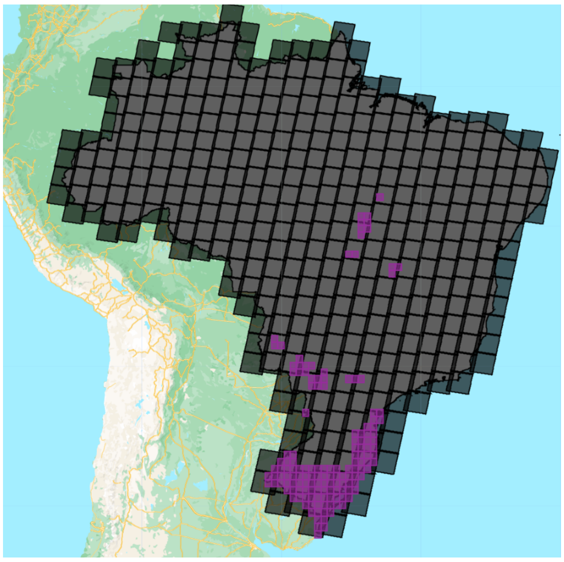
P2	07/01/year	01/01/year+1

2.1.4 Definition of regions for classification

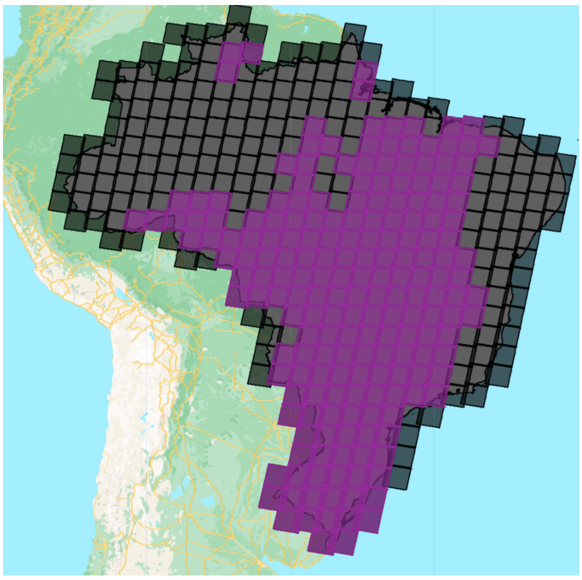
The ‘Agriculture’ and ‘Forest Plantation’ are heterogeneously distributed in the Brazilian biomes. Therefore, Landsat scenes were selected in regions with the highest occurrence of each class according to the reference maps. Figure 8 illustrates the scenes chosen for each land use class.



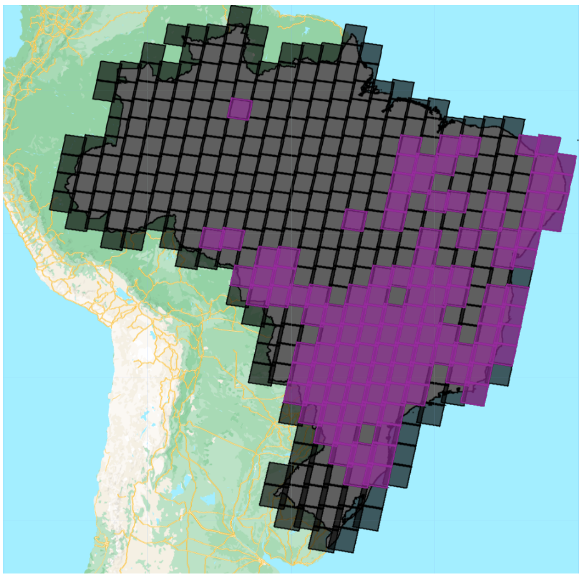
Cotton



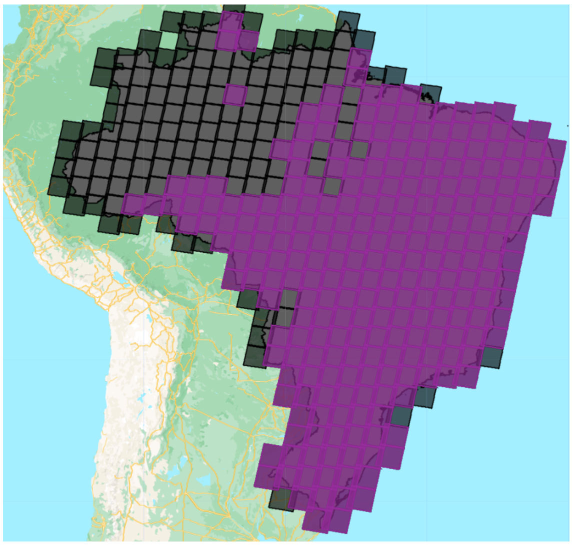
Rice



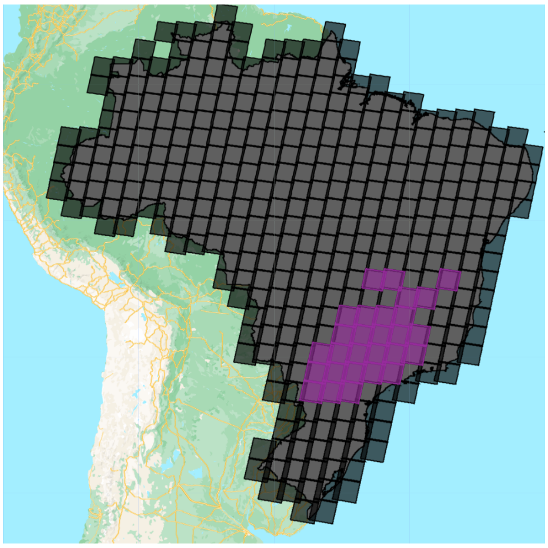
Soybean



Sugar Cane



Other Temporary Crop



Citrus

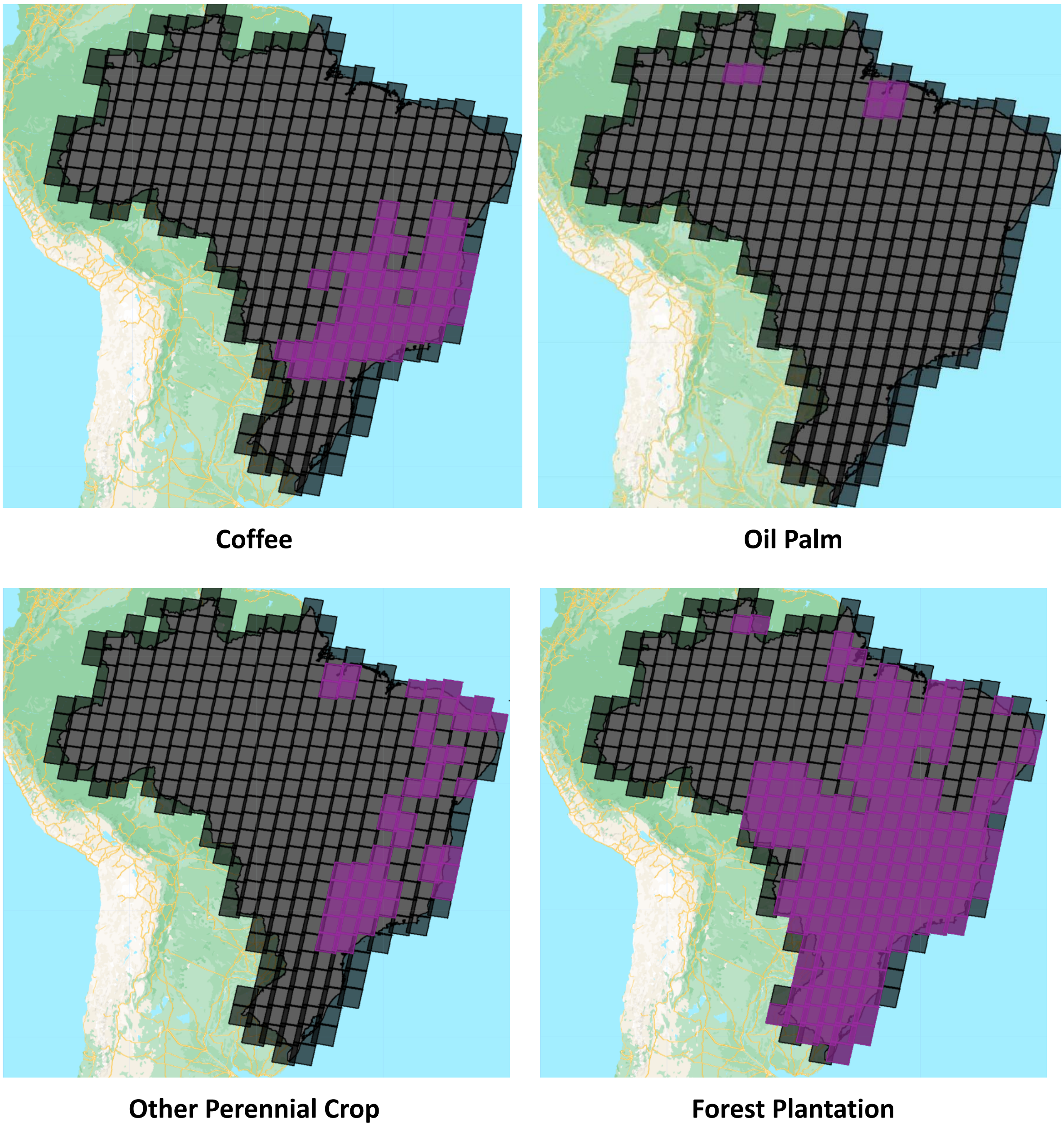


Figure 8. Selected scenes of Landsat series to classify maps by land use class.

The rice, coffee, and citrus maps do not cover the total spatial distribution of their crops in Brazil, despite the expansion to new areas to coffee (Espirito Santo state), and to citrus (Minas Gerais and Paraná states), since Collection 7.1. This limitation is due to two main reasons: *i*) reference maps coverage, a challenge to map areas without reference maps, and *ii*) spatial scale, a challenge for mapping smaller scale crops using Landsat images.

2.1.5 Feature space

In Collection 7.1, the selection of the mosaic composition bands was revised for the 'Temporary Crops' and for 'Forest Plantation' classes, while for the other classes, the selected bands remained the same as in Collection 6. Thus, in Collection 8, the feature space used was based on the improvements made in Collection 7.1 and in those adopted since Collection 6.

2.1.5.1 Cotton, Soybean and Other Temporary Crop

The total band of the mosaic consists of metrics calculated from Landsat bands and indexes over three time periods, with a total of 234 bands. The visible, near-infrared, and shortwave bands, as well as statistical metrics and vegetation indexes, such as Normalized Difference Water Index (NDWI) (GAO, 1996) and EVI2, were already used for classification of the 'Temporary Crops' class. However, other vegetation indices, such as Soil Adjusted Vegetation Index (SAVI) and Leaf Area Index (LAI), were added, due to the importance of these indexes to agricultural classification. Bands, indexes, and metrics used are presented in Table 7.

Table 7: Bands, indexes, and metrics used to compose the Landsat mosaics.

Bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indexes	SAVI, CAI, EVI2, NDWI, LAI
Metrics	median, mean, max, min, stdDev, 80^{th} percentile, 20^{th} percentile, and CEI (NDWI, NIR, EVI2)

The dimensionality reduction process corresponds to the selection of only the most essential bands of the mosaic to be used in the classification. This approach can reduce the processing time and can increase classification efficiency. Thus, in the classification methodology, a step aims to identify a metric of the importance of each band to the classifier. The implementation of *smileRandomForest()* in Google Earth Engine already provides this information, which can be extracted from the trained classifier by the *explain()* function. The importance values extracted from this function are calculated by the Gini Importance (Breiman, 1984), and are by-products of the Random Forest classifier, having the advantage of an almost zero additional computational cost. However, we must be cautious about the effectiveness of this index, primarily due to the possibility of overestimating the importance of correlated predictors. This method was used at this time because of the potential to reduce the computational cost, in addition to previous results that have not indicated notable changes in the classification.

From the importance information, 30% of the bands with the highest importance value for classification were identified for each scene in each year, and this list was used as band selection for the classifier.

2.1.5.2 Sugar Cane

The bands, indexes, and metrics used to classify sugar cane are presented in Table 8.

Table 8: Bands, indexes, and metrics used to compose the Landsat mosaics to classify sugar cane.

Bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indexes	NDVI, NDWI
Metrics	median

2.1.5.3 Rice

The bands for rice mapping using the U-Net were selected to ensure the greatest highlight between rice crops and other land uses (*i.e*. other types of crops). The variables were selected according to the state to be mapped, as shown in Table 9.

Table 9: Bands, indexes, and metrics used to compose the Landsat mosaics to classify rice.

State	Tocantins	Santa Catarina	Parana	Rio Grande do Sul
Bands	SWIR1, SWIR2	SWIR2	SWIR1, SWIR2	SWIR1, SWIR2, TIR1
Indexes	EVI2, NDWI	EVI2, NDWI	EVI2, NDWI	EVI2
Metrics	CEI (EVI2), CEI (NDWI)	CEI (EVI2), CEI (NDWI)	CEI (EVI2)	CEI (EVI2)
Period	Bands – off season CEI – Annual	Bands – off season CEI – Annual	Bands – off season CEI – Annual	Bands – growing season CEI – Annual

2.1.5.4 Coffee

The annual Landsat mosaics were composed based on bands, indexes, and metrics presented in Table 10.

Table 10. Bands, indexes, and metrics used to classify coffee in MapBiomas Collection 8.

Bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indexes	EVI2, NDWI
Metrics	median, mean, max, min, stdDev, 80th percentile, 20th percentile, and quality mosaic (qmo)

2.1.5.5 Citrus

The bands used for training and classification of citrus were annual compositions generated from the median of the five images with less cloud cover in each point orbit. The bands used are shown in Table 11.

Table 11. Bands and metrics used to classify citrus in MapBiomas Collection 8.

Bands	RED, NIR, SWIR1
Metrics	median

2.1.5.6 Oil Palm

A Unet architecture neural network was employed in the methodology to map oil palm. The selected bands were carefully chosen to enhance the satellite image's features, thus enabling effective training and classification of this class. This approach, incorporating the Unet neural network architecture, proved to be particularly effective in remote sensing data analysis for the detection and identification of oil palm trees. Table 12 presents the bands and metrics used.

Table 12. Bands and metrics used to classify oil palms in MapBiomas Collection 8.

Bands	RED, NIR, SWIR1
Metrics	median

2.1.5.7 Other Perennial Crop

Part of the Other Perennial Crop map came from the separation of that class from the class ‘Annual and Perennial Crop’ of MapBiomas Collection 4. Therefore, this map was created from two classification processes: 1) in Collection 4, the classifier was trained to classify ‘Annual and Perennial Crop’ without distinction; 2) in this collection, these maps resulting from the first classification were submitted to a second classification, in which the classifier was trained with new feature spaces to distinguish the pixels of short-cycle crops and long-cycle crops. The resulting map of ‘Perennial Crops‘ became part of the class Other

Perennial Crop, while the resulting map of 'Temporary Crops' wasn't used in this collection (it was processed again using the methodology described before). Figure 9 illustrates the processes performed to generate the class Other Perennial Crop.

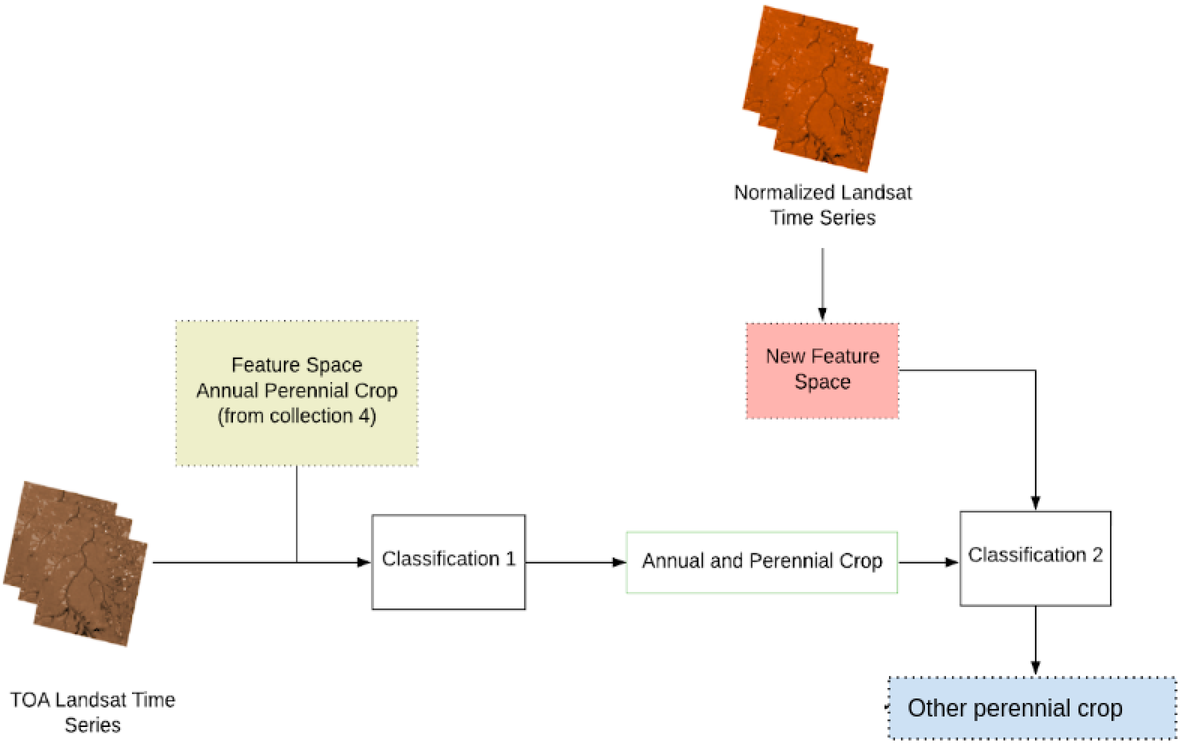


Figure 9. Steps to separate Other Perennial Crop from the previous class 'Annual and Perennial Crop' of the Collection 4.

The cycle of 'Temporary Crops' tends to have greater annual variation in the spectral response than 'Perennial Crops', which are more stable over time. Therefore, metrics were selected to highlight this difference between 'Temporary' and 'Perennial Crops' (Table 13).

Table 13. Bands, indexes, and metrics used to classify Other Perennial Crops in MapBiomas Collection 8.

Bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indexes	NDVI
Metrics	median, max, min, stdDev, 20th percentile, and quality mosaic (NDVI)

2.1.5.8 Forest Plantation

The bands, indexes, and metrics used to classify 'Forest Plantation' in Collection 8 are presented in Table 14.

Table 14. Bands, indexes, and metrics used to classify 'Forest Plantation' in MapBiomas Collection 8.

Bands	BLUE, GREEN, RED, NIR, SWIR1, SWIR2
Indexes	EVI2, MNDWI, LAI
Metrics	median, mean, max, min, stdDev, 80th percentile, and quality mosaic (qmo)

2.1.6 Classification algorithm, training samples and parameters

2.1.6.1 Reference Maps

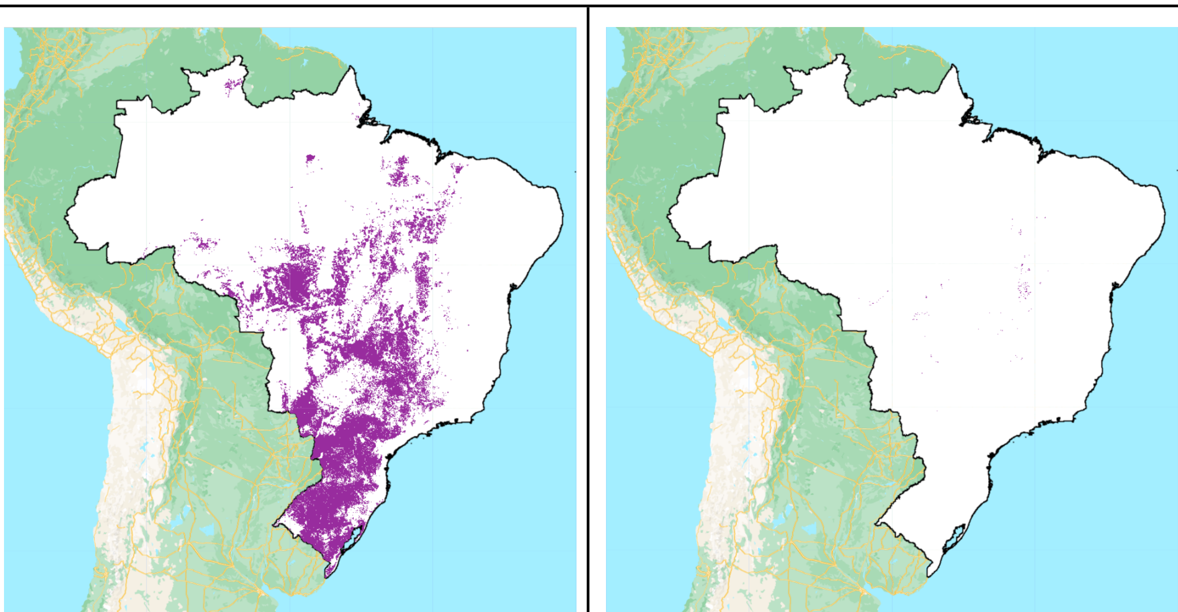
The reference maps used to obtain samples to train the classifier are shown in Table 15.

Table 15. Reference maps used in the Random Forest classification for the classes 'Agriculture' and 'Forest Plantation' in Collection 8.

Class	Landsat time series	Number of training samples	Sampling Approach	Rule	Type	Year of acquisition	Reference
Soybean (2000 - 2022)	Normalized	10,000	Stratified	-	stable samples	2021	Agrosatélite Song et al. (2021)
Soybean (1985 - 1999)	L5 TOA	10,000	Simple	-	stable samples	2000	Agrosatélite (2020A) Agrosatélite (2020B) Song et al. (2021)
Sugar cane	TOA	10,000		-	annual samples	2003 - 2019	Rudorff et al. (2010)
Rice	TOA	-	-	-	chips	2017-2020	Agência Nacional de Águas (ANA) and Companhia Nacional de Abastecimento (Conab)

Cotton	Normalized	10,000	Stratified	-	stable samples	2021	Agrosatélite
Other Temporary Crop	Normalized	10,000	Stratified	-	stable samples	2021	Agrosatélite
Coffee	Normalized	10,000	Stratified		stable samples	2015, 2016, 2017, 2018, 2019	Companhia Nacional de Abastecimento (Conab)
Citrus	TOA	-	-	-	chips	2020	Agrosatélite
Oil Palm	TOA	-	-	-	chips	2020	Agrosatélite
Other Perennial Crop	Normalized	5,000	-	Minimum of 20% for the interest class	stable samples	2016	Agrosatélite
Forest Plantation	Normalized	10,000	Stratified	-	stable samples	2012 - 2014	Global Forest Watch, Transparent World (2015)

The reference maps used are shown in Figure 10.

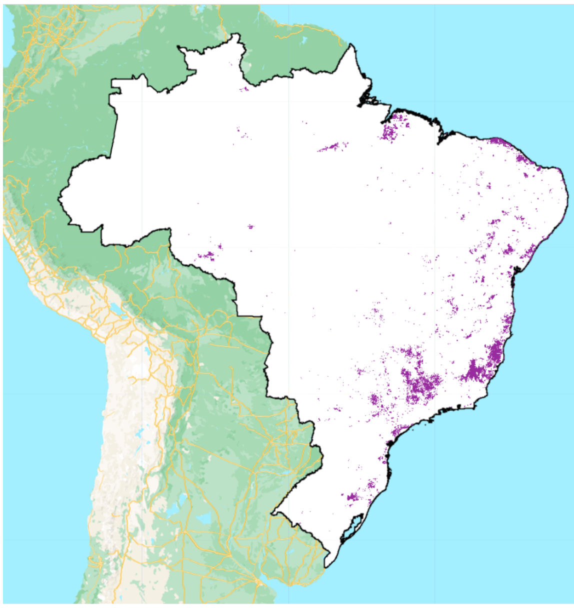


Soybean reference map - Agrosatélite (2021)

Cotton reference map - Agrosatélite (2021)

Canasat project (RUDORFF *et al.*, 2010) map of 2018/2019

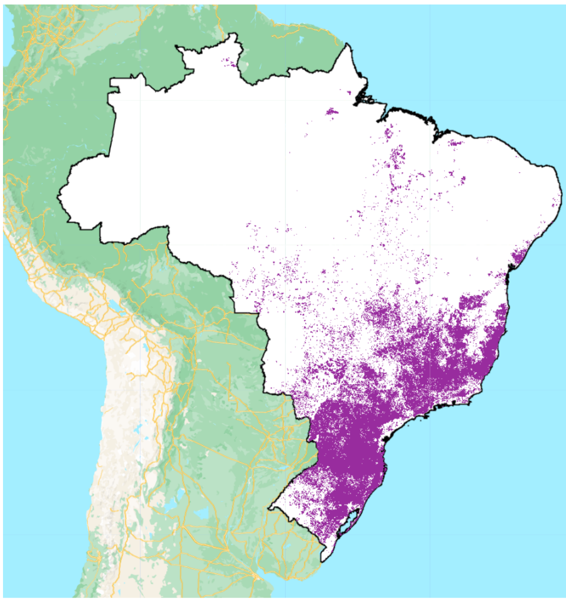
Rice reference map - Conab/ANA (2020)



Other Perennial Crop reference map - Agrosatélite (2020b)

Coffee reference map - Conab (2015, 2016, 2017, 2018, 2019)

Citrus reference map - Agrosatélite



Global Forest Watch, APRE, Klabin, Fórum Florestal da Bahia - TRANSPARENT WORLD, 2015

Oil Palm reference map

Figure 10. Reference maps representing the areas with training samples for the classification of 'Agriculture' and 'Forest Plantation' in Brazil.

2.1.6.2 Random Forest

For the classes mapped by the Random Forest algorithm (Breiman, 2001), the process steps are: a) Initially, an annual Landsat mosaic is created, according to the period of the year (i.e. growing season and off-season), specific for each class; b) bands are compose with specific metrics for each class; c) simple or stratified random sampling is performed based on the reference map; d) the samples are used to train the classifier; e) classify the classes of interest. The results of the process are annual maps of interest classes. To reduce the amount of noise and inconsistencies, the maps obtained after the classification undergo spatial and temporal post-processing and then are integrated into the other themes of MapBiomas. An important observation is that the annual mosaic used in the training process must be from the same year as the reference map used. An example of Random Forest classification is presented in Figure 11.

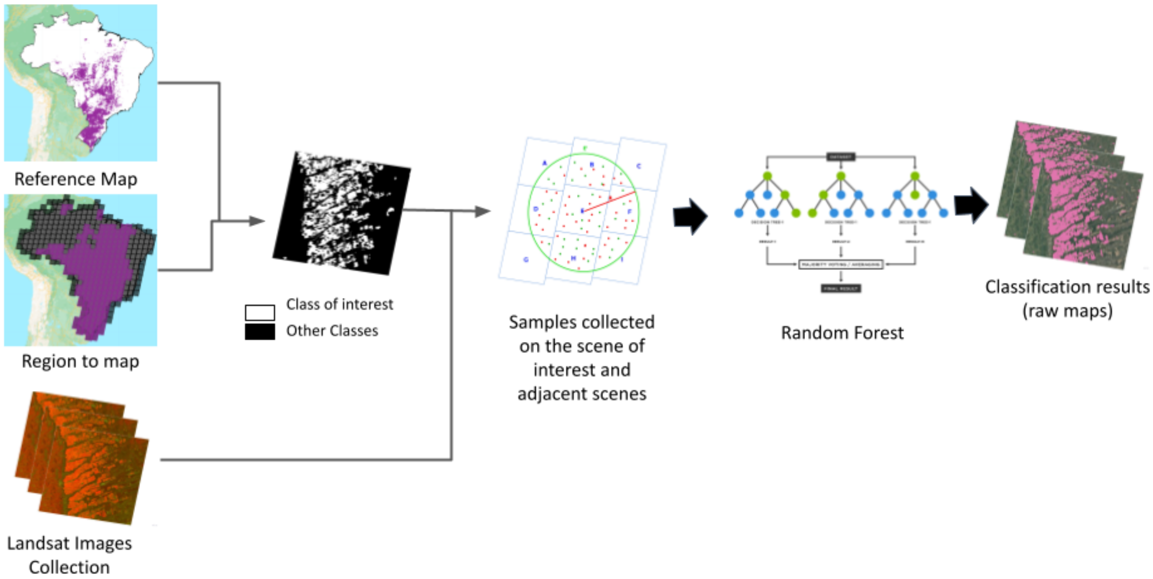


Figure 11. Flowchart of agriculture and Forest Plantation classification using Random Forest algorithm.

The classes mapped with Random Forest algorithm were soybean, sugar cane, cotton, Other Temporary Crops, coffee, Other Perennial Crops and 'Forest Plantation'. All classes were trained with 100 trees, with default values for other parameters.

2.1.6.3 Simple Sampling

The acquisition of training samples was performed by each Landsat scene. In addition to the samples collected in the target scenes, samples collected in adjacent scenes were included inside an E' buffer of radius R, in which the center of that radius corresponds to the center of the target scene (E), as shown in Figure 12.

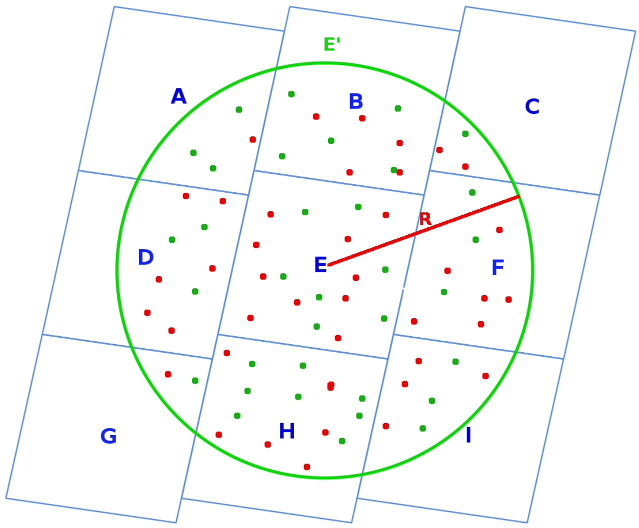


Figure 12. Scheme for sample acquisition for the regionalized training of the Random Forest classifier in 'Agriculture' and 'Forest Plantation'.

As no reference maps are available for all classes in all years of the time series (1985 to 2022), stable samples were created. However, these samples were only obtained in classes that used the normalized Landsat series, due to the characteristics of this time series mentioned above. For another group of classes obtained from the TOA Landsat time series, as a reference map was not available for each year to be classified, annual samples were used on the available reference maps for training and classification only for those years with available reference maps. The classification result based on a reference map was used to support the subsequent training and classification procedure of previous years up to the year with the available reference map. The Random Forest training scheme was used to classify the subsequent years in which a reference map was not available.

2.1.6.4 Stratified Sampling

The quality of training samples has been related as one of the ways to increase the accuracy of remote sensing image classifications, as well as the algorithm's performance and accurate input data (LI et al., 2021; ZHU et al., 2016). The sampling methods commonly used for supervised classification (such as simple sampling), may often not consider the spatial distribution of the targets of interest in the scene, resulting in unbalanced samples between classes. Thus, a stratified sampling approach aimed to balance the sample distribution between the interest and non-interest targets (LI et al., 2021).

The difference between stratified and simple sampling is that, in the traditional method (simple sampling), the training samples are distributed randomly, considering the whole Landsat scene boundary, while in the stratified sampling method, the distribution of the number of training samples is weighted by the percentage area of the class of interest, obtaining a balanced distribution of samples as is shown by Figure 13.

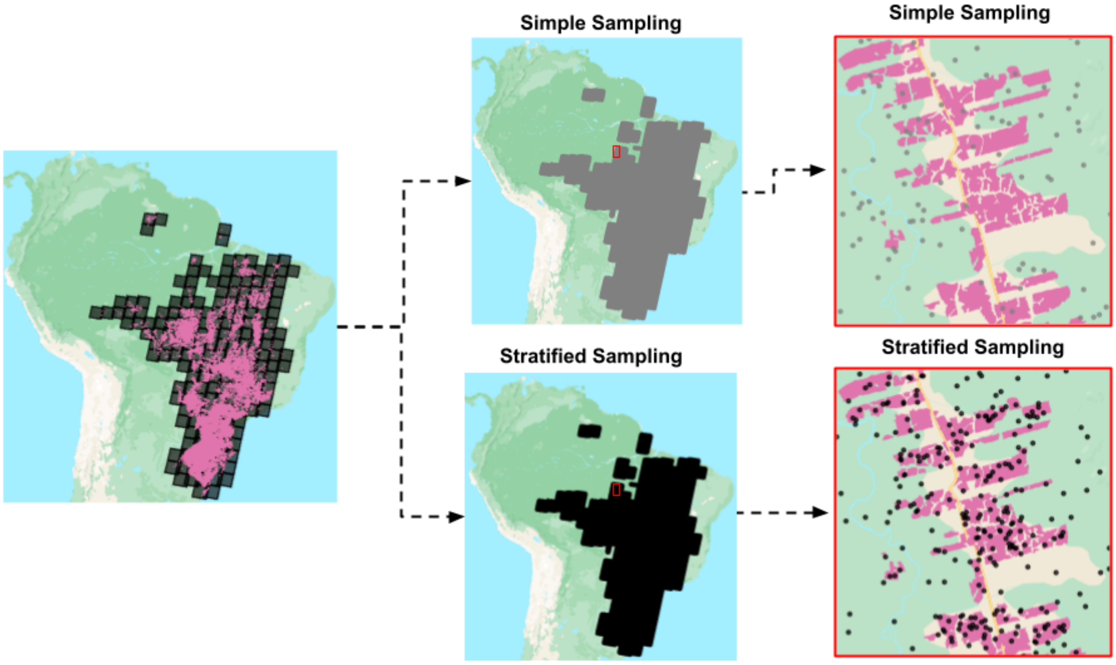


Figure 13 - Difference between sampling approaches. a) simple sampling, and b) stratified sampling.

2.1.6.5 Deep Learning

For the mapping of rice and citrus, an adaptation of the U-Net convolutional neural network (RONNEBERGER et al., 2015) was used. Unlike machine learning algorithms that classify each pixel considering the spectral response for each pixel, this architecture uses the context in which the pixels are. This architecture is illustrated in Figure 14.

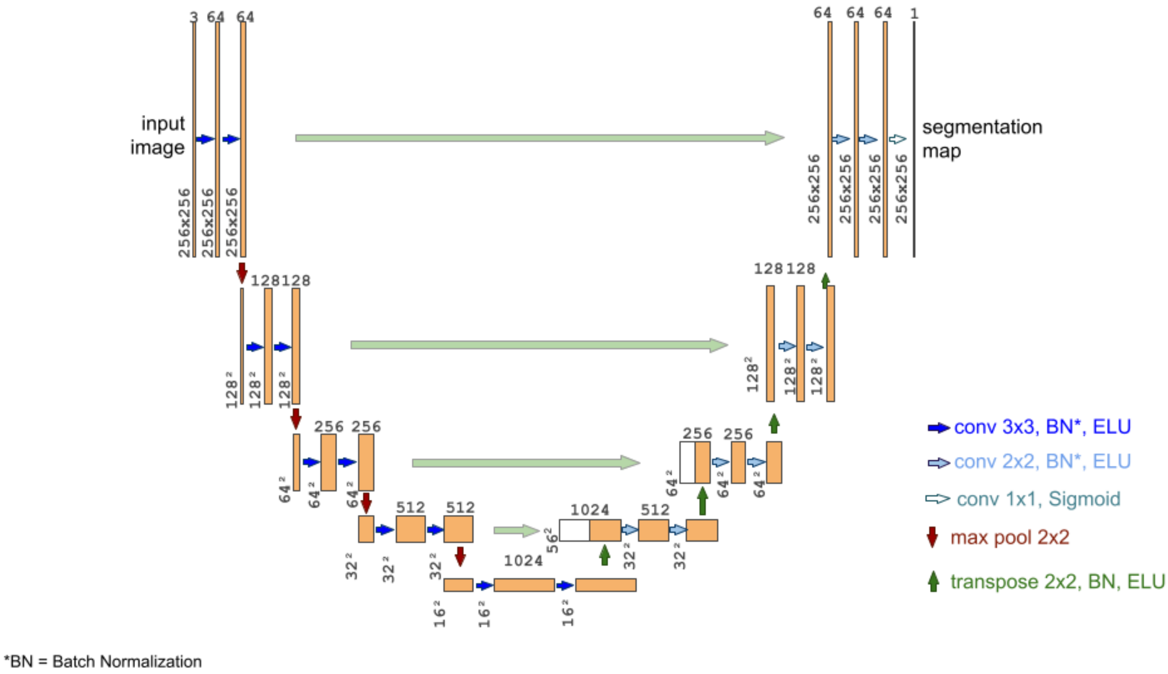


Figure 14 - Adapted U-Net convolutional neural network, with its layers and connections, used for the mapping of rice and citrus.

This architecture was developed in Python, using the TensorFlow 2.0 library. The entire training and mapping process was carried out using the Google Colab platform. To enable the Google Colab platform to have access to satellite images, Google Drive was used to store the images.

To obtain the training and validation sets, each training block was covered to generate chips with 256 x 256 pixels. Then, the chips were divided into 70% for training and 30% for validation for each block. After data separation, the pixel values of each image band were normalized. Normalization scales the numerical values for a given range, making each band have the same weight for the classifier.

2.1.6.5.1 Rice

The delimitation of the mapping area was based on the map of irrigated rice in Brazil published by the National Water Agency (ANA) and the National Supply Company (Conab) in 2020. The selection of images was made based on the growing season according to the year of mapping carried out in each state. The reference map was divided into blocks of 0.5 x 0.5 degrees (~300 thousand ha each). The blocks used for rice mapping and training were those that overlapped the reference map with the states of interest, as illustrated in Figure 15.

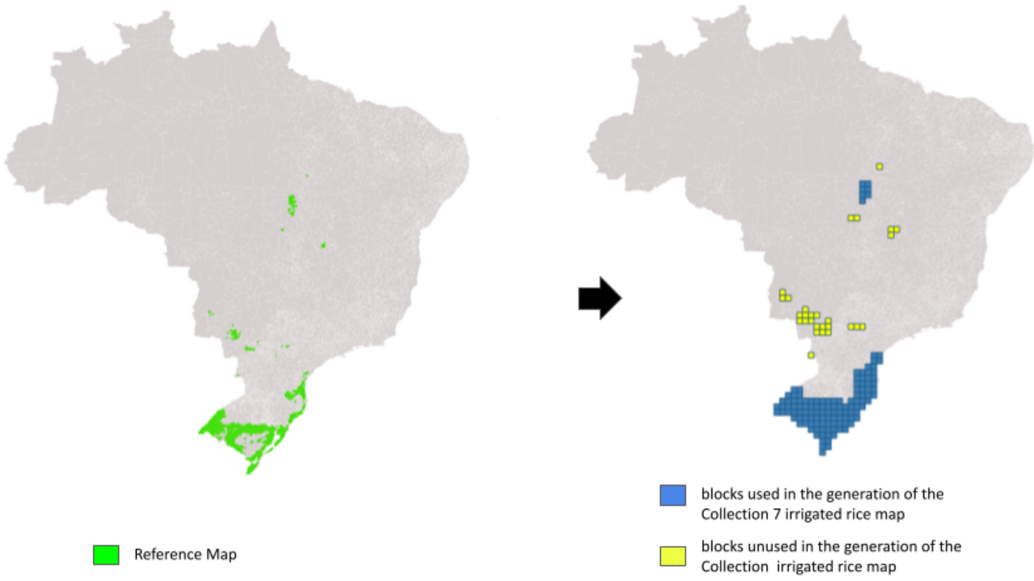


Figure 15. Study area used for the mapping of irrigated rice in the MapBiomas Project.

From the reference map and the annual Landsat mosaics, training samples were created, consisting of pairs of blocks of the annual mosaic (from the reference year) and in the mask of the reference map for this same block. A sample U-Net entry training example is shown in Figure 16.

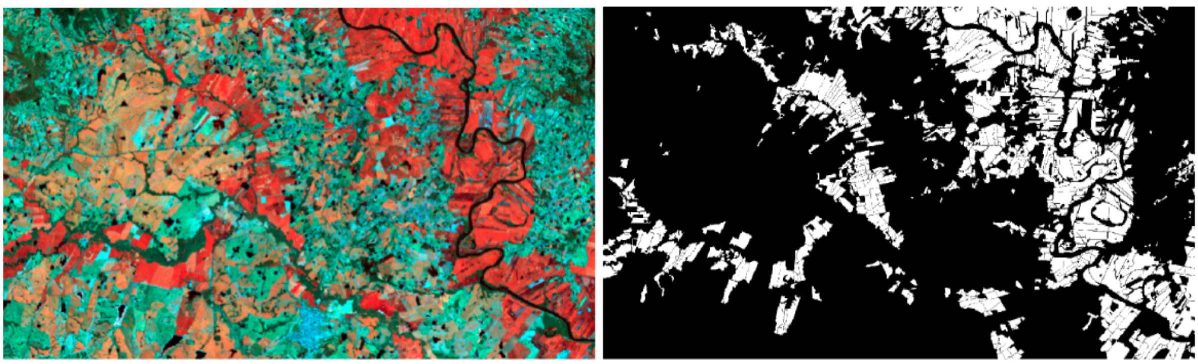


Figure 16. Example of U-Net samples to mapping rice.

The test data was used for the accuracy analysis of the trained model. The final model (*i.e*. the one with the best results) was used in the process of classification of irrigated rice in different states for each year of the series (1985-2022).

2.1.6.5.2 Citrus

The citrus map was performed, similar to rice, using a neural network based on the U-Net architecture. Reference data for training were generated by visual interpretation of Sentinel and Landsat images for the year 2020.

2.1.6.5.3 Oil Palm

The oil palm map was performed, similar to rice, using a neural network based on the U-Net architecture. Reference data for training were generated by visual interpretation of Sentinel and Landsat images for the year 2020.

3 Post-classification

Temporal and spatial filters were applied to remove noise and classification errors.

3.1 Spatial filter

The filter of minimum connected pixels was applied in most classes, except on the classes mapped with U-Net, because the result of the semantic segmentation showed little or no spatial noise. This spatial filter removed groups of pixels with 6 or fewer pixels of the interest class or the "others" class (Figure 17).

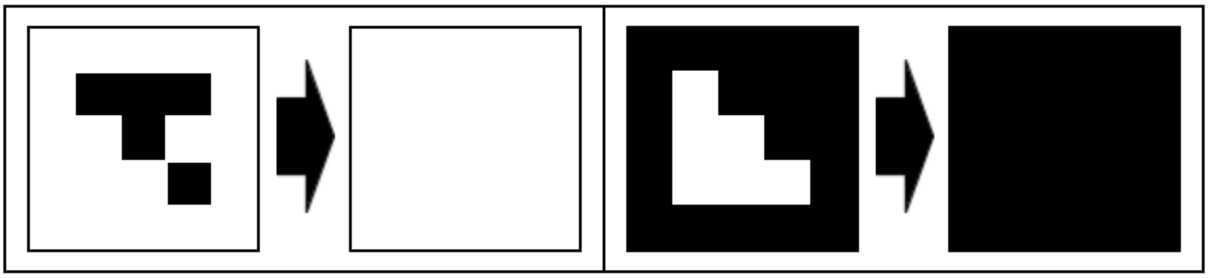


Figure 17. Example of the minimum connected pixels spatial filter. The image on the left shows an exclusion of pixels of the interest class (in black). The image on the right shows an inclusion of pixels of "other class" (in white) to the interest class.

3.2 Temporal filter

In general, two temporal window filters were applied: 3 years with 2 years threshold or 5 years with 3 years threshold. The 3-year window excludes the center year when none of the adjacent years are of interest class and includes the center year when both adjacent years are of interest class (Figure 18). The 5-year window excludes the center year when no more than 2 years are of interest class and includes when at least 3 adjacent years are of interest class (Figure 19).

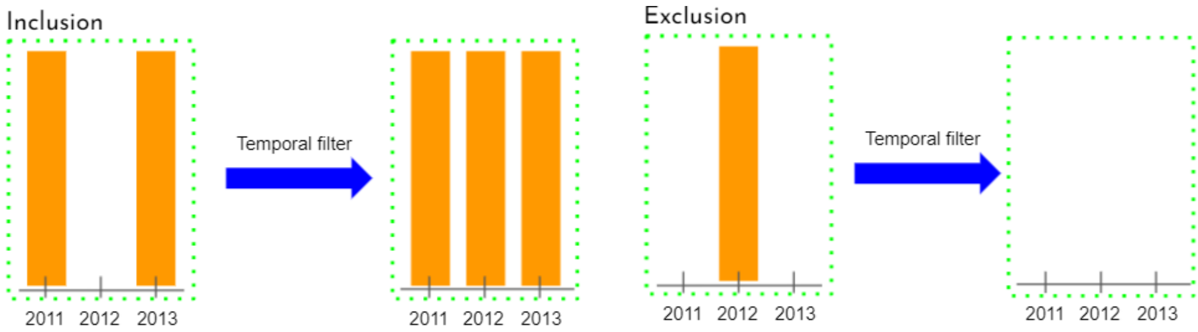


Figure 18. 3-year temporal window filter: The orange bars represent pixels of the mapped class (interest class). The exclusion filter changes a pixel to the “others” class when the same pixel was not of interest class in the adjacent years. The inclusion filter changes a pixel to the interest class when the same pixel was of the interest class in the adjacent year's.

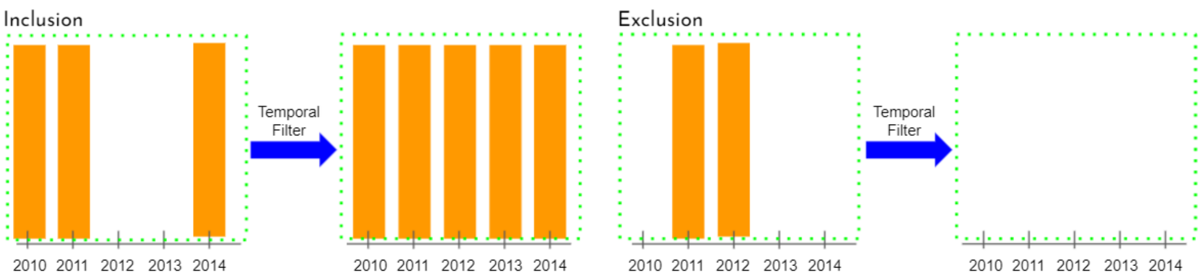


Figure 19. 5-year temporal window filter: The orange bars represent pixels of the mapped class (interest class). The exclusion filter changes a pixel to the “others” class when no more than 1 another year is of interest class. The inclusion filter changes a pixel to the interest class when at least 3 adjacent years are of the interest class.

In addition, for all agriculture classes, the first year of the series (*i.e.* 1985), pixels were excluded when, in the following year, they were not classified, and included when, in the following year, they were. For the last year of the time series (*i.e.* 2022), no temporal filter was applied.

3.2.1 Soybean, Cotton, and Other Temporary Crops

For MapBiomas Collection 8, the soybean, cotton, and Other Temporary Crops classes followed a unified process that was proposed in the last Collection (Figure 20).

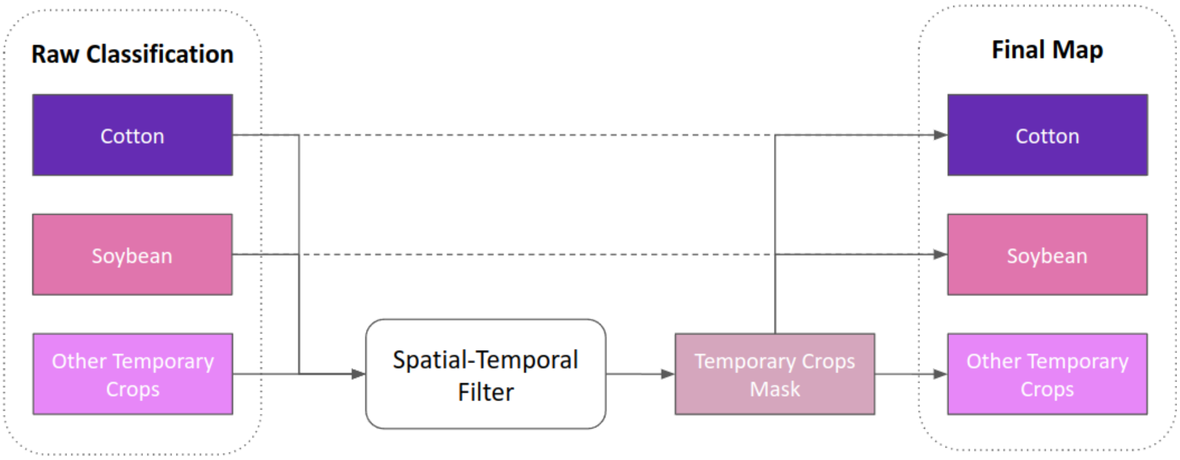


Figure 20: Soybean, cotton, and Other Temporary Crops classification and post-classification Flowchart.

The classification was made using a reference map with class distinction for cotton, soybean, and Other Temporary Crops. However, the raw classification results were not used in the filters as separate classes, but as a unified 'Temporary Crops' class. The temporal filter used a 3-year window with 2 years threshold.

The filtered result was an annual mask that indicated the area used for 'Temporary Crops' in general. As a final step, the raw classifications of cotton and soybean were masked by the 'Temporary Crops' mask, resulting in the final maps for those classes. The remaining area in the 'Temporary Crops' mask was considered as Other Temporary Crops. In this way, it was possible to maintain the temporal stability of the 'Temporary Crops' areas, while the annual crop variation during the growing season period was preserved.

3.2.2 Rice

For the rice class no temporal filter was applied, since this class occurred predominantly in the South of Brazil, in the same area where other crops are cultivated over the year. Thus, the areas are not exclusively for rice crops over the year.

3.2.4 Sugar cane

In sugarcane post-processing it was used four temporal filters:

1) Temporal filter using 3 years with 2 years threshold applied only on the initial edge year (1986). In the initial year (1985) and final year (2022) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2020.

3) Temporal filter using 5 years with 2 years threshold applied only on the final edge year (2021).

4) Temporal filter using 3 years with 2 years threshold applied to all series, except to the edge years (1986-2021) to ensure temporal consistency.

3.2.6 Citrus

As with the coffee class, for the citrus class the same filters were applied for the edge years of the series (1986 and 2021) as for the other years, plus a time consistency filter as follows:

1) Temporal filter using 3 years with 2 years threshold applied only on the edge years (1986 and 2021). In the initial year (1985) and final year (2022) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2020.

3) A fill filter was also applied to convert pixels that were not classified as citrus between a period when these pixels were classified as citrus.

3.2.7 Coffee

For post-processing of the coffee class three temporal filters were used:

1) Temporal filter using 3 years with 2 years threshold applied only on the edge years (1986 and 2021). In the initial year (1985) and final year (2022) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2020.

3) We also defined as pixels of the coffee class, starting in 2016, those pixels that were classified as coffee in any of the mappings of the last 5 years.

3.2.8 Oil Palm

For post-processing of the oil palm class three temporal filters were used:

1) Temporal filter using 7 years with 3 years threshold applied from 1986 to 2000.

2) Temporal filter using 5 years with 3 years threshold for the period 2000 to 2021.

3) And, a Temporal filter using 3 years with 2 years threshold was applied for the last years, from 2019 to 2022.

3.2.9 Other Perennial Crop

For Other Perennial Crop was used a temporal filter using 5 years with 3 years threshold, in addition to a filter to remove intervals of the class of interest with less than 5 consecutive years; therefore, a 6-year window was utilized: the year of interest and 1 year before and 4 years after the year of interest.

3.2.10 Forest Plantation

For the ‘Forest Plantation’ class, it was applied the same temporal filters applied to the coffee class:

1) Temporal filter using 3 years with 2 years threshold applied only on the edge years (1986 and 2021). In the initial year (1985) and final year (2022) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2020.

3) Another consideration was at the end of the series. When the trees are cut, it may take a while for them to grow again and then the classifier can’t identify them as ‘Forest Plantation’. Since it takes 3 to 5 years for ‘Forest Plantation’ to become identifiable again, to solve this situation pixels from 2016 to 2021 were converted to ‘Forest Plantation’ when they were of this class in the 3 years before (2013 to 2015). Figure 21 illustrates this filter.

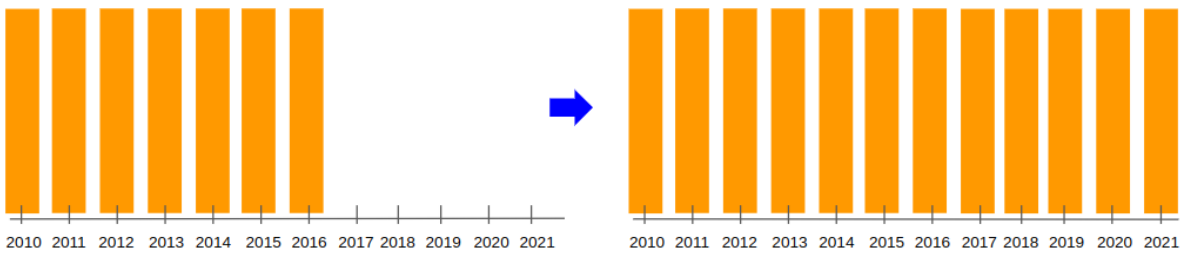


Figure 21. Temporal filter was applied in the last years of the ‘Forest Plantation’ series.

4) Another temporal filter was applied to fill longer intervals of non-occurrence of ‘Forest Plantation’ when it was forest plantation some year in the past and it became again years after, like the example in Figure 22.

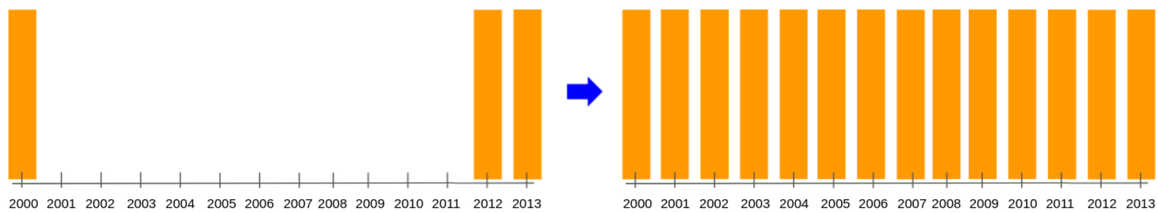


Figure 22. Temporal filter that converted longer intervals into ‘Forest Plantation’ when it was in the past and became again years after.

4 Integration with biomes and themes

After the classification of the 'Agriculture' and 'Forest Plantation' themes, they were integrated into the other land use and land cover classes to compose the MapBiomas Collection 8 final maps. This integration process was based on the overlapping order of the classes. The integration process tends to improve the quality of the 'Agriculture' and 'Forest Plantation' maps as it removes some commission errors.

5 Validation strategies

The independent validation points provided by the LAPIG of the Goias Federal University (UFG) were used to calculate the global accuracy of the mapping and the accuracy for each land use class. The following section also presents some comparisons between the Random Forest classification results and the reference maps.

5.1 Accuracy analysis

The accuracy analysis was produced using independent validation points provided by the *Laboratório de Processamento de Imagens e Geoprocessamento* (LAPIG) of the Goias Federal University (UFG).

We used all points that at least two interpreters considered the same class, resulting in over 12,000 validation points. LAPIG points were collected only for the aggregate classes of 'Forest Plantation', 'Perennial Crops', and 'Temporary Crops', without distinction between the crops that compose these classes. For this reason, we aggregate all perennial classes (coffee, citrus, and Other Perennial Crops) into 'Perennial Crops' and all temporary classes (soybean, sugarcane, rice, cotton, and Other Temporary Crops) into 'Temporary Crops' to evaluate the accuracy using LAPIG points. LAPIG points used for the accuracy assessment are shown in Figure 23.

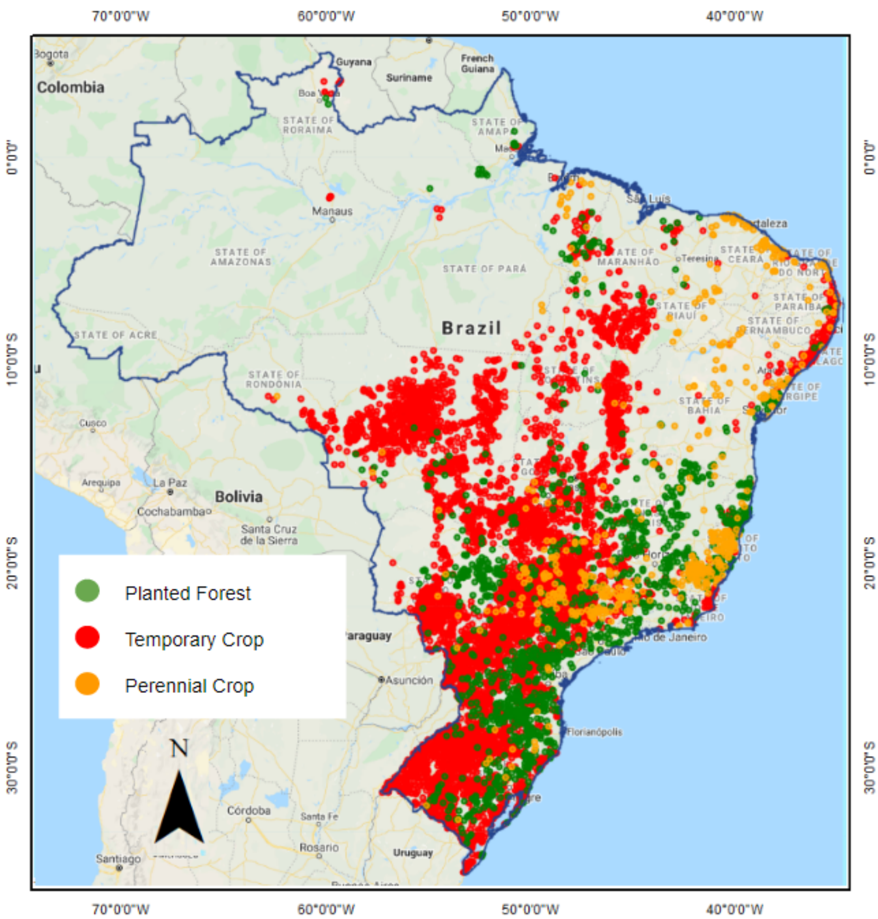


Figure 23. LAPIG points that were used for the accuracy assessment of ‘Temporary Crops’, ‘Perennial Crops’, and ‘Forest Plantation’ classes.**Figure 23.** LAPIG points that were used for the accuracy assessment of ‘Temporary Crops’, ‘Perennial Crops’, and ‘Forest Plantation’ classes.

5.1.1 Temporary Crops

The results of the accuracy assessment of the ‘Temporary Crops’ class are presented in Figures 24 and 25, covering the period from 1985 until 2022.

Figure 24 shows the producer’s accuracy of the ‘Temporary Crop’ class in Collection 8. This statistical metric indicates the portion of the pixels of ‘Temporary Crop’ were correctly classified. In this sense, for ‘Temporary Crop’ the producer’s accuracy informs that over all time-series, more than 80% of ‘Temporary Crop’ areas were classified by MapBiomas as ‘Temporary Crop’.

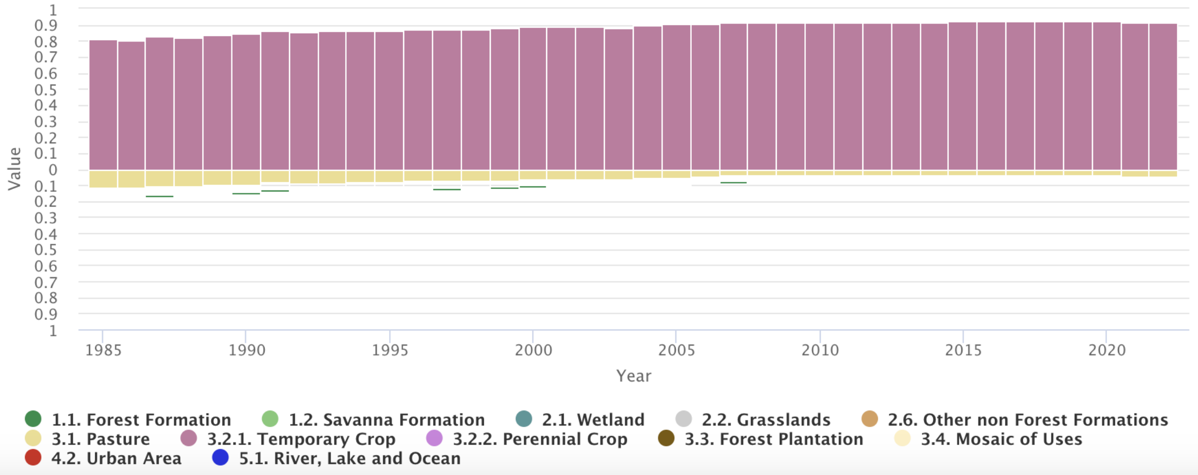


Figure 24. Producer's accuracy of the 'Temporary Crop' class in Collection 8.

Figure 25 presents the user's accuracy of the 'Temporary Crop' class in Collections 8. This statistical metric indicates the portion of the map pixels that were correctly classified. For 'Temporary Crop' in this Collection, the user's accuracy shows that from 2000, between 70-90% of the area mapped as 'Temporary Crop' is in fact 'Temporary Crop'. However, from 1985 until 2000 we can see lower user accuracy values, indicating that an important portion of the area mapped as 'Temporary Crop' is pasture and grassland.

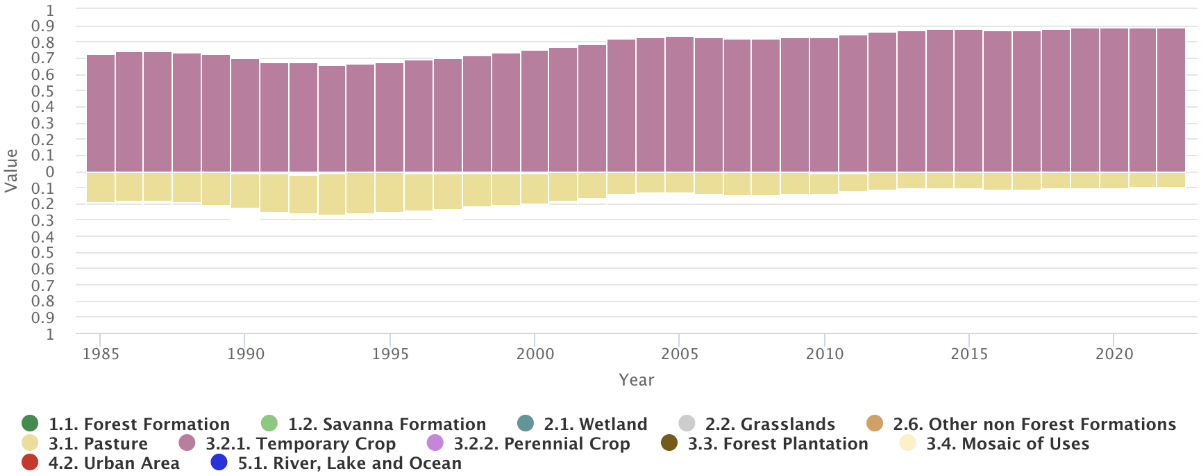


Figure 25. User's accuracy of the 'Temporary Crop' class in Collection 8.

5.1.2 Perennial Crops

The producer's accuracy for 'Perennial Crop' classes, for Collection 8, is presented by Figure 26. This metric indicates the omission of the 'Perennial Crop' classification, where we can observe higher omission at the beginning of the time-series, while from the 2000 until the end of the time-series, the mapping could identify more areas of 'Perennial Crop' classes as 'Perennial Crop'. In addition, it is important to note that the most part of these areas of

‘Perennial Crop’ that were not identified in the MapBiomas as ‘Perennial Crop’ were classified as pasture or natural forests.

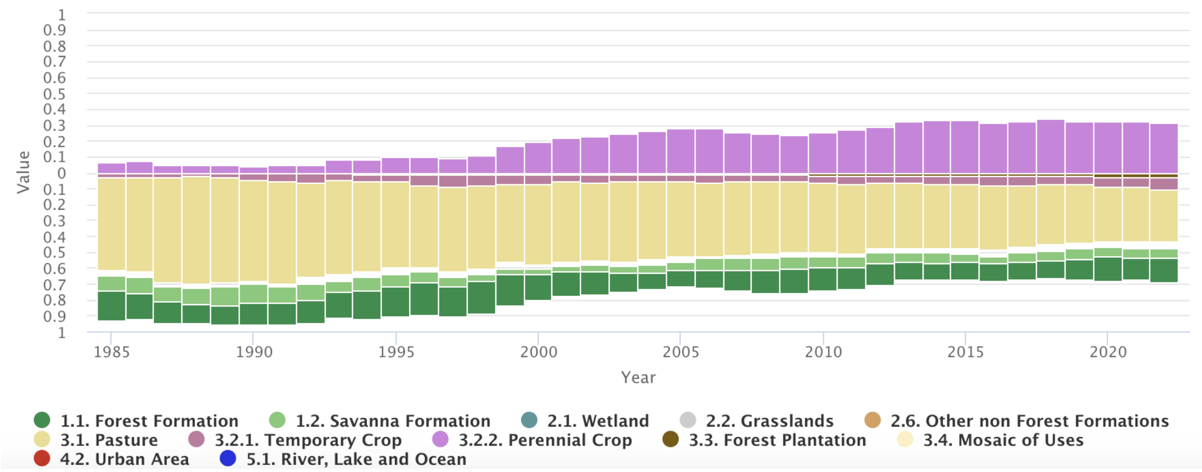


Figure 26. Producer’s accuracy of the ‘Perennial Crops’ class in Collection 8.

Figure 27 presents the user’s accuracy of the ‘Perennial Crop’ class in Collections 8. This metric indicates that at the beginning of the time-series a lower portion of the area mapped as ‘Perennial Crop’ is in fact ‘Perennial Crop’, while a higher portion of the area mapped as ‘Perennial Crop’ is in fact ‘Perennial Crop’ is pasture and natural forests. However, this metric improved over time, showing better agreements between the LAPIG points and the ‘Perennial Crop’.

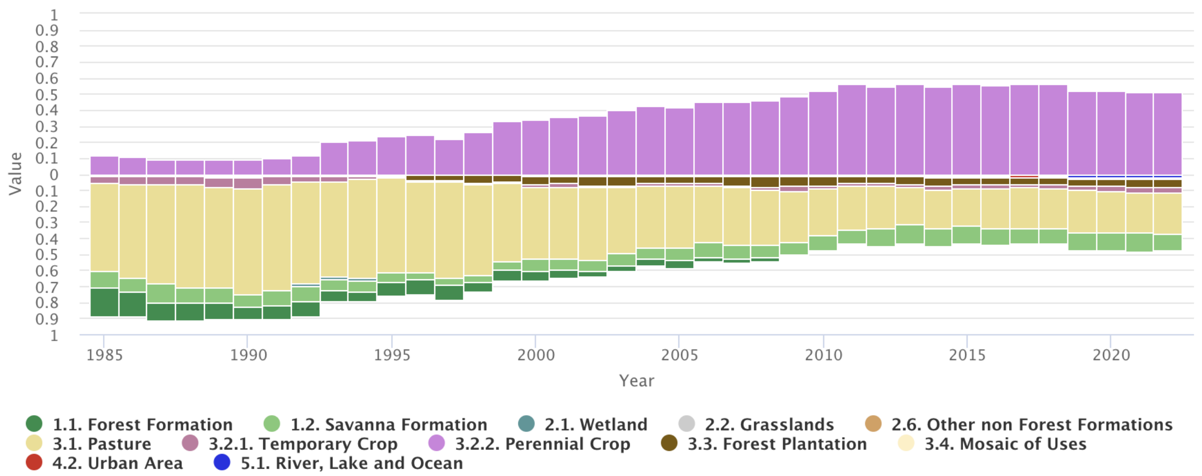


Figure 27. User’s accuracy of the ‘Perennial Crop’ class in Collection 8.

5.1.3 Forest Plantation

Figure 28 presents the results of producer’s accuracy for the ‘Forest Plantation’ class for Collections 8. This metric indicates the amount of the area mapped as ‘Forest Plantation’ is ‘Forest Plantation’ and the amount of the ‘Forest Plantation’ area was not classified as ‘Forest Plantation’. Thus, it is possibly observed that in the beginning of the time-series, from 1985-1993, a higher portion of the ‘Forest Plantation’ area was classified as Mosaic of Uses and natural forests. However, this amount has decreased since 1993, showing that the map improved the identification of the 'Forest Plantation'.

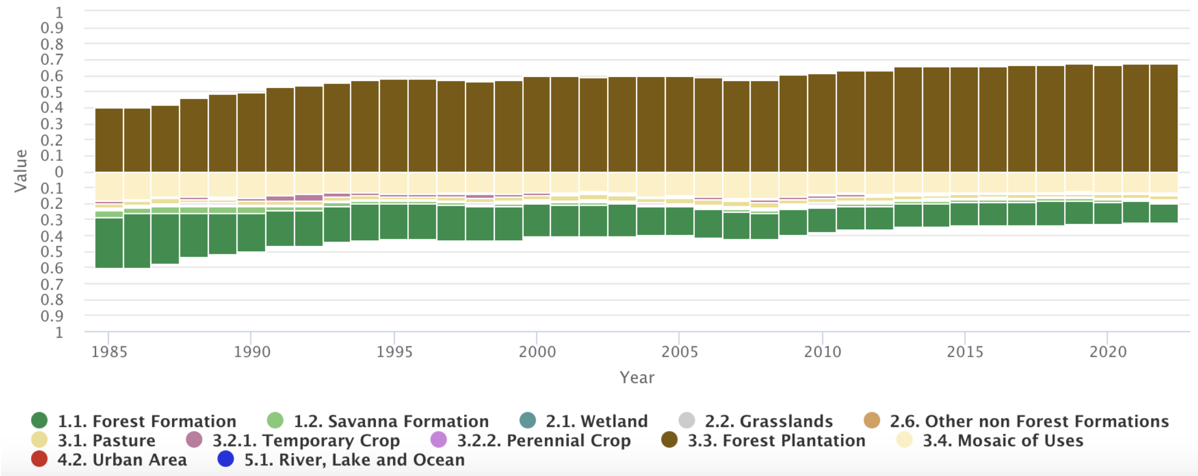


Figure 28. Producer's accuracy of the 'Forest Plantation' class in Collection 8.

Figure 29 presents the user's accuracy, the metric that indicates the inclusion error. Through this metric, for Collection 8, it was possible to observe an inclusion of 10-20% over the middle of the time-series. However, for the last 17 years of the time-series an inclusion lower than 5% was verified.

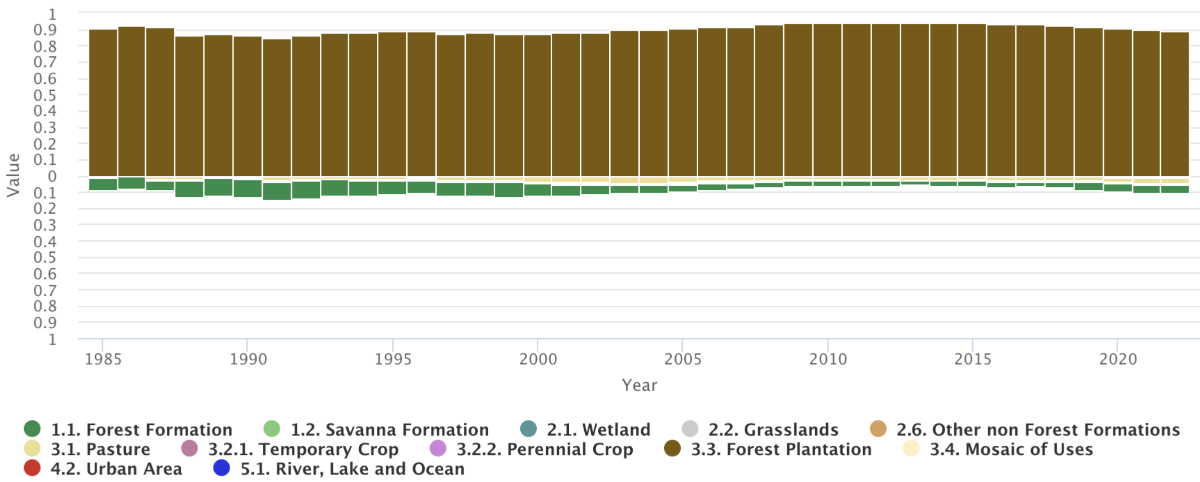


Figure 29. User's accuracy of the 'Forest Plantation' class in Collection 8.

5.2 Comparison with reference data

In addition to the comparison with validation points, a comparison between the 'Agriculture' and 'Forest Plantation' maps of MapBiomas Collection 8 with data from the Municipal Agricultural Production (PAM - *Produção Agrícola Municipal*) and Production of Vegetable Extraction and Forest Plantation (PEVS) , both carried out by the Brazilian Institute of Geography and Statistics (IBGE), Ibá - *Indústria brasileira de árvores* (Brazilian Tree Industry), and the *Companhia Nacional de Abastecimento* (CONAB) was also made for the most of the classes.

Comparison of Temporary Crop Area

The validation points are essential both for accuracy analysis and for promoting the calculation of the map error-adjusted area. Therefore, in addition to accuracy evaluated based on the use of LAPIG points, comparisons were made between the area from Collection 8 with the previous Collection and information about the planting area of IBGE and CONAB, for soybean, cotton, rice, sugar cane, and Other Temporary Crops classes. Figure 30 presents a comparison between the area of the soybean class with the areas estimated by IBGE and CONAB.

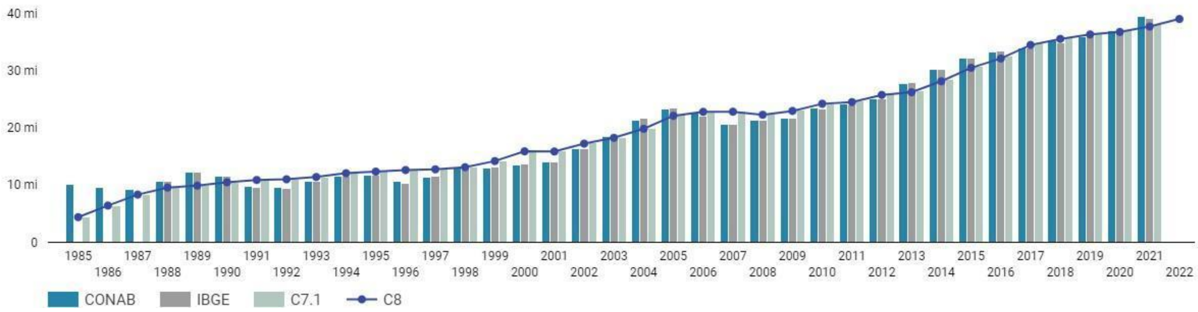


Figure 30. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of soybean area in Brazil.

As mentioned before, in this Collection 8 there was a concern to maintain the methodologies for mapping agricultural classes. Thus, there are no significant differences in terms of area between Collections 8 and 7.1. However, it is essential to emphasize the great adherence, in terms of area, and at the national level, to the amount of soybean area mapped by MapBiomas compared to official sources. It is observed that throughout the time series, the area mapped by MapBiomas shows great adherence to the area informed by IBGE and CONAB, highlighting the importance of stabilizing the methodology, since the approach currently developed has been able to robustly capture soybean plantations throughout the national territory.

In addition, the MapBiomas soybean map also can capture the distribution of this crop over the Brazilian territory, indicating the states from Central and South regions as those with the highest area of soybean in 2022 (Figure 31).

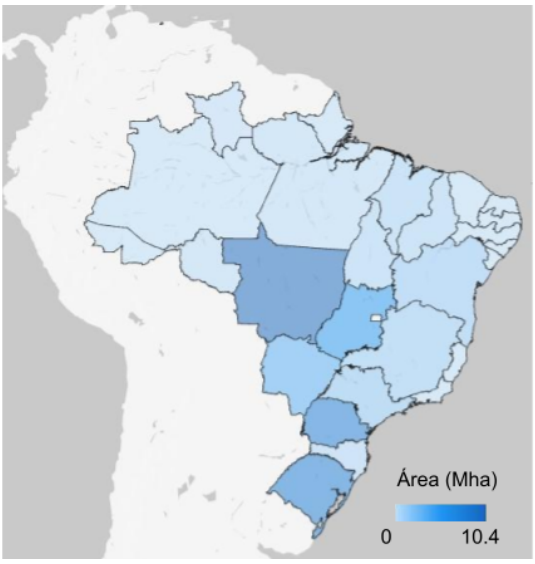


Figure 31. Distribution by Brazilian states of soybean area mapped in Collection 8, for the year 2022.

Regarding the cotton class, as previously mentioned, in the MapBiomas scope only first crop cultivations are mapped. Thus, for the cotton class there is an underestimation of the area when compared to the information from official sources, since the official sources consider the area planted with cotton throughout the first and second crop, while in the MapBiomas scope, only the first crop is considered for mapping (Figure 32).

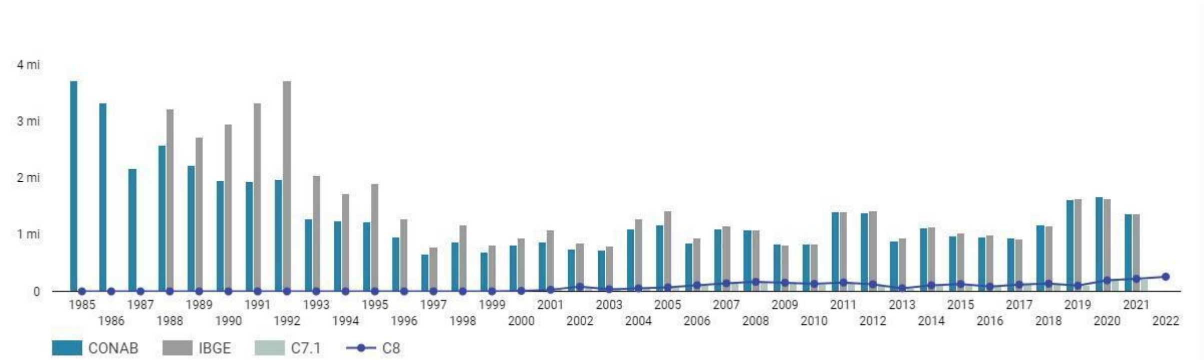


Figure 32. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of cotton area in Brazil.

In addition, the cotton mapping information for the year 2022 shows that the state of Bahia, followed by Mato Grosso, is those with the largest planted areas, considering only the first crop of cotton (Figure 33).

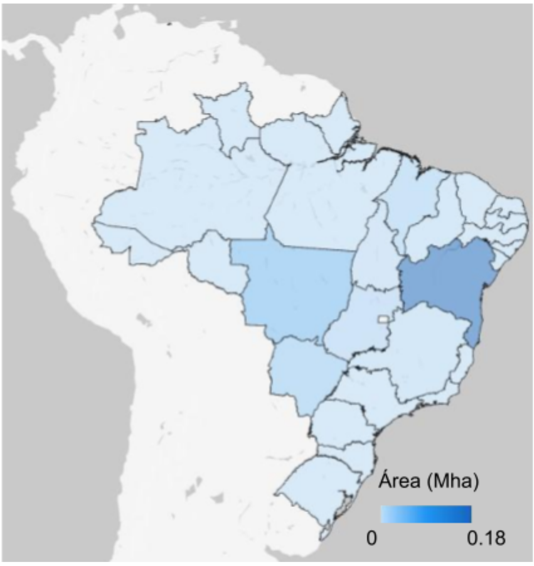


Figure 33. Distribution by Brazilian states of the cotton area mapped in Collection 8, for the year 2022.**Figure 33.** Distribution by Brazilian states of the cotton area mapped in Collection 8, for the year 2022.

Regarding the rice class, the comparison with the area information from IBGE and CONAB shows that there is a greater adherence, in terms of area, in the last year of the MapBiomas series, while at the beginning there is a large underestimation (Figure 34).

This fact is partly explained by the abridgment of the rice mapping methodology, which due to the differences in the spectral response of this crop in the different regions where it is grown in Brazil, the MapBiomas methodology is only employed in the main producing states. Thus, this limitation of coverage is reflected in the underestimation of area over the time series.

Figure 35 shows the distribution of the mapped rice area for the year 2022 in Collection 8, highlighting the state of Rio Grande do Sul, the main rice-producing state in Brazil.

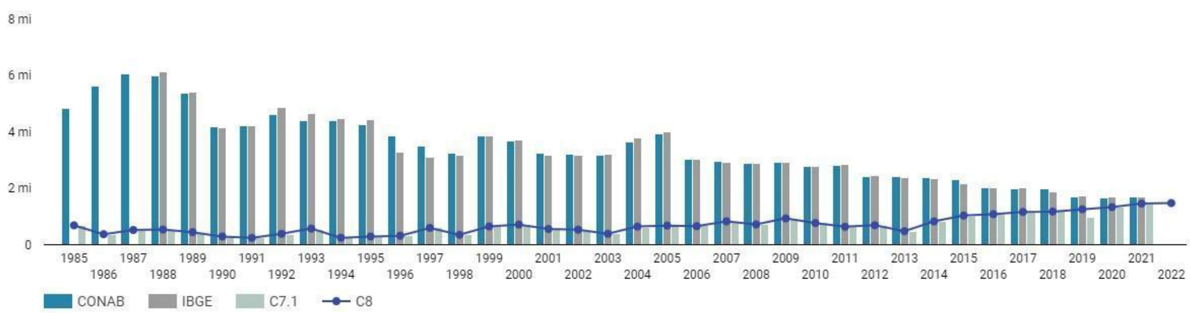


Figure 34. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of the rice area in Brazil.**Figure 34.** Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of the rice area in Brazil.

Figure 35. Distribution by Brazilian states of rice area mapped in Collection 8, for the year 2022.

The comparison of the mapped sugarcane area between 1985 and 2022 by MapBiomas with information from official sources shows a significant agreement in demonstrating how the expansion of the planted area in Brazil has been occurring (Figure 36), with São Paulo state standing out as the largest area with plantations (Figure 37).

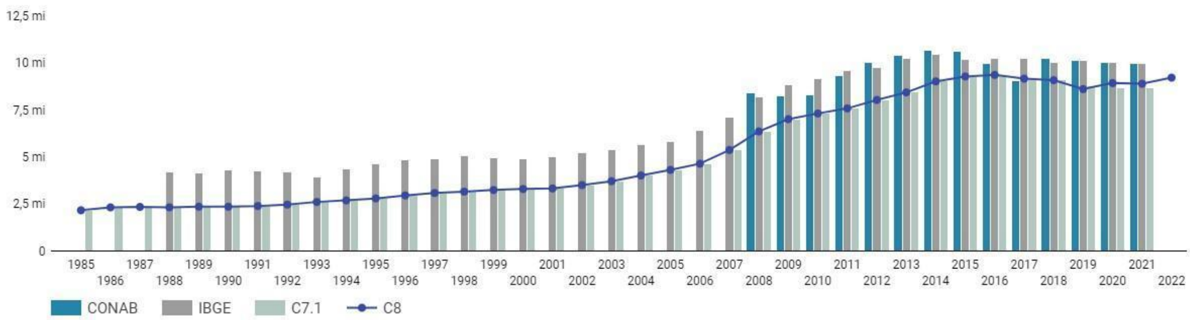


Figure 36. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of sugarcane area in Brazil.

Figure 37. Distribution by Brazilian states of sugar cane area mapped in Collection 8, for the year 2022.

To provide the comparisons of the Other Temporary Crops, it was subtracted from the total area of the Temporary Crops those areas of MapBiomas classes that are mapped individually (soybean, cotton, rice, sugar cane), obtaining the residual area of the other temporary crops from IBGE and CONAB.

The temporal series of the Other Temporary Crops class from MapBiomas, when compared with official data from IBGE and CONAB, highlights an interesting discrepancy (Figure 38). This class within MapBiomas includes various temporary crops that cannot be individually distinguished due to methodological limitations, primarily stemming from the lack of reference maps for model training. The comparison with area data from IBGE and CONAB reveals an underestimation of this class's extent within the MapBiomas dataset. A pivotal factor explaining this is the inherent challenge of accurately mapping all temporary crops due to the absence of a comprehensive reference map. Consequently, the area attributed to this class within MapBiomas tends to be underestimated. Furthermore, an insightful observation emerges when examining the spatial distribution of this class across Brazilian states – a pronounced correlation with major soybean-producing states becomes evident (Figure 39).

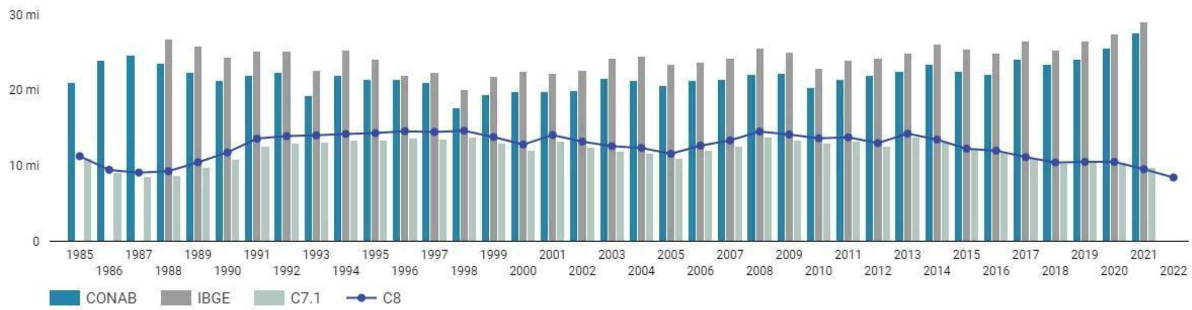


Figure 38. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics

(IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of Other Temporary Crops area in Brazil.

Figure 39. Distribution by Brazilian states of Other Temporary Crops area mapped in Collection 8, for the year 2022.

5.2.1 Comparison of Perennial Crop area

Mapping ‘Perennial crops’ has been a challenge throughout the MapBiomas collections. The approach of mapping each type of crop separately, since Collection 6 has enabled the improvement of the ‘Perennial Crops’ map, however, the mapped area still underestimates the official area provided by IBGE. It is noteworthy that the area provided by IBGE comprises ‘Perennial Crops’ in Brazil that are not totally mapped by MapBiomas (MapBiomas maps citrus, coffee, oil palm (new class), and some concentrations of perennial crops spread throughout the territory).

An analysis of the coffee class within MapBiomas in comparison to official sources (IBGE and CONAB) reveals an interesting pattern. The graph prominently illustrates a considerable underestimation by MapBiomas, particularly at the beginning of the time series, as opposed to the official data (Figure 40). However, in the later years of the series, the coffee area mapped by MapBiomas aligns more closely with the official sources. It's important to emphasize the limitations of the mapping methodology, including the requirement for reference maps and image availability. Furthermore, the distinct characteristics of coffee cultivation in Brazil pose challenges in extrapolating training samples obtained for one region to others. Notably, the mapping of coffee areas in Minas Gerais, the largest coffee producer, and the extension of the mapping to the state of Espírito Santo are worth highlighting (Figure 41).

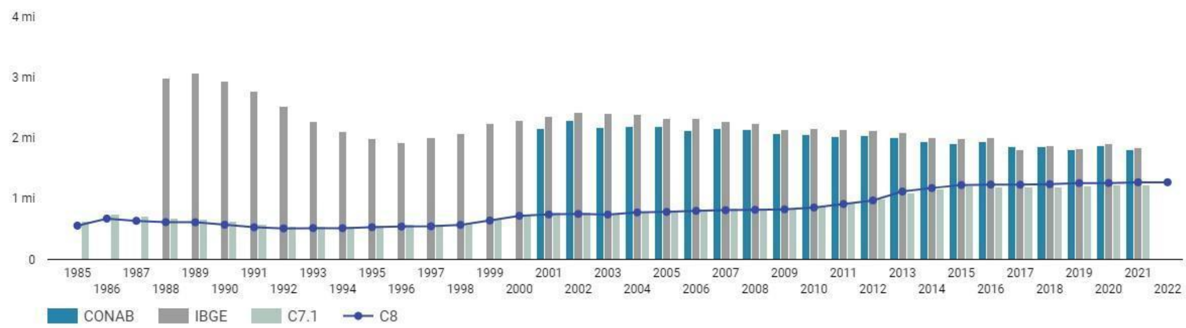


Figure 40. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Companhia Nacional de Abastecimento* (CONAB) for the temporal series (1985-2022) of coffee areas in Brazil.

Figure 41. Distribution by Brazilian states of the coffee area mapped in Collection 8, for the year 2022.

An examination of the citrus class within MapBiomas, juxtaposed with official data from IBGE, reveals a substantial and consistent underestimation of the mapped area over the time span of 1985 to 2022 (Figure 42). This underestimation is particularly evident when compared to IBGE figures. It's crucial to highlight the inherent challenges in mapping this class, given the diverse characteristics of citrus cultivation across various producing regions. This variability often limits the transferability of models trained in one region to perform optimally in another. Moreover, within the realm of MapBiomas, the citrus class has been mapped solely for the states of São Paulo (the largest planted area), Paraná, and Minas Gerais (Figure 43).

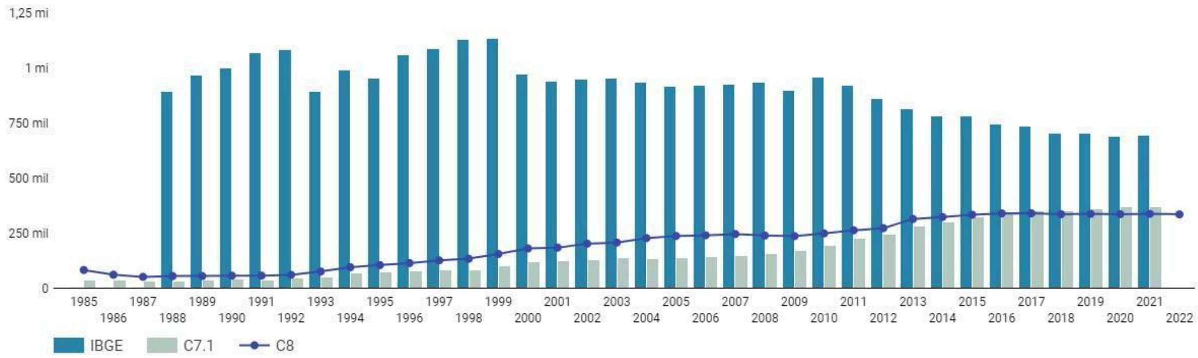


Figure 42. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), for the temporal series (1985-2022) of citrus areas in Brazil.

Figure 43. Distribution by Brazilian states of the citrus area mapped in Collection 8, for the year 2022.

The oil palm class emerges as a novelty in MapBiomas Collection 8. This perennial crop is primarily found in the northern region of the country, notably in the state of Pará (Figure 45). The temporal series provided by MapBiomas unveils a significant surge in the area attributed to this class, beginning around 2012 (Figure 44). To a certain extent, the temporal pattern of this area increase could also be linked to the enhanced availability of Landsat images, facilitating a more robust mapping process.

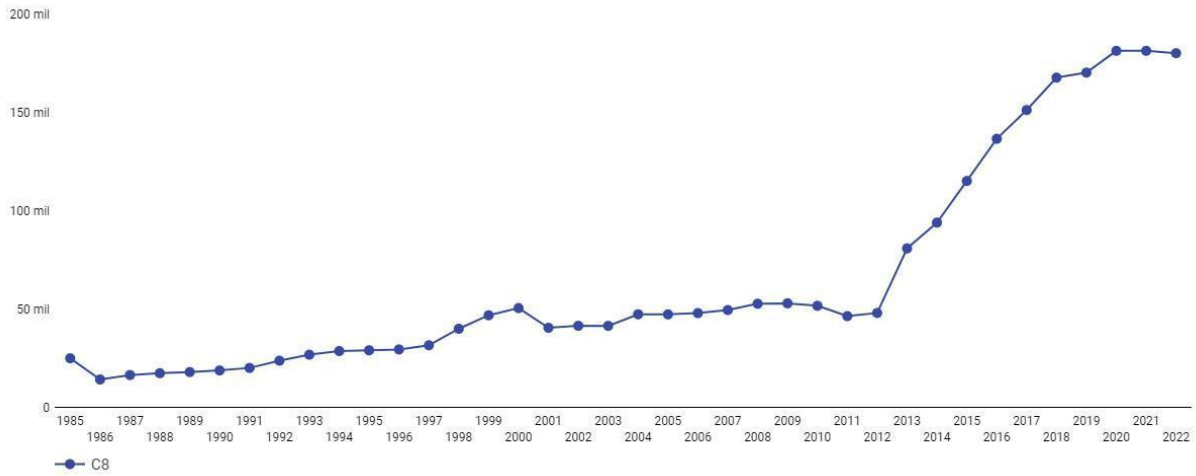


Figure 44. Temporal series (1985-2022) of oil palm areas in Brazil.

Figure 45. Distribution by Brazilian states of the oil palm area mapped in Collection 8, for the year 2022.

The Other Perennial Crops class in MapBiomas denotes various perennial crops without individual distinction. The analysis with the official source (IBGE) emphasizes the important underestimation of the mapped area (Figure 46). This underestimation is a result of intricate factors, including variable image availability and the inherent complexities of distinguishing individual perennial crops within the class. In addition, it is important to recognize that the mapping of this class grapples with challenges arising from the absence of comprehensive reference maps, particularly essential for refining the accuracy of its distinctions. Additionally, the information from MapBiomas informs that the Other Perennial Crops class is predominantly concentrated in the northeastern region of the country, with notable occurrences in the states of Minas Gerais and São Paulo (Figure 47).

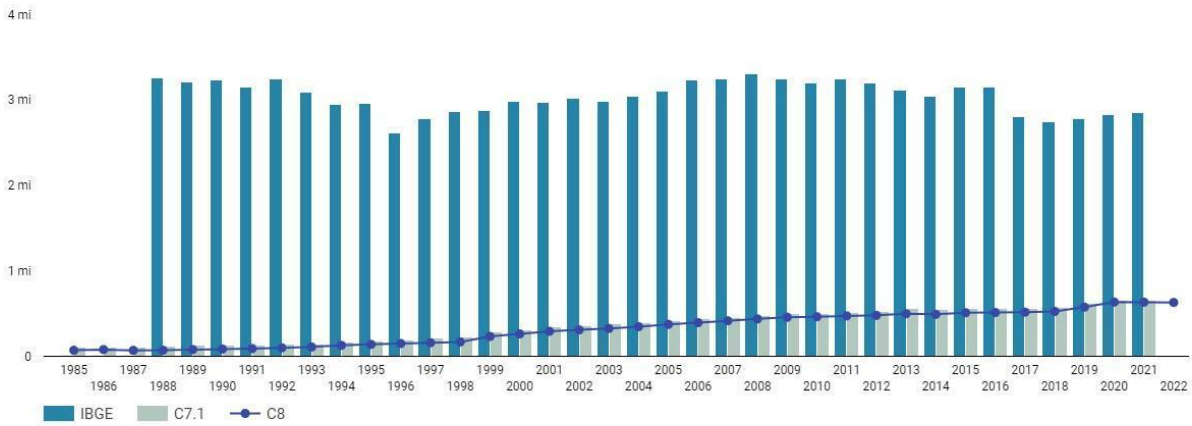


Figure 46. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), for the temporal series (1985-2022) of Other Perennial Crops areas in Brazil.

Figure 47. Distribution by Brazilian states of the Other Perennial Crops area mapped in Collection 8, for the year 2022.

5.2.2 Comparison of Forest Plantation area

'Forest Plantation' areas obtained from MapBiomas Collection 8 annual maps were also compared with areas from official sources and with Collection 7.1. To estimate Brazil's forestry area, a comparison was made between MapBiomas Collection 6, PEVS-IBGE, and Ibá. The results are presented in Figures 48 and 49.

The analysis of the 'Forest Plantation' class presents an insightful comparison between MapBiomas, Ibá, and IBGE data over the temporal span of 1985 to 2022. The findings reveal a noteworthy agreement between MapBiomas and Ibá, illustrating a similar pattern of increasing planted forest area. However, in contrast, the comparison with IBGE demonstrates a notable underestimation (Figure 48). Furthermore, the utility of MapBiomas data extends to examining both temporal and spatial expansion of agricultural classes. In the context of 'Forest Plantation', attention is drawn to the prominent presence of this class in the state of Minas Gerais, where the largest planted forest area is observed (Figure 49). This underscores the significance of MapBiomas in unraveling the growth dynamics of various land cover classes over time.

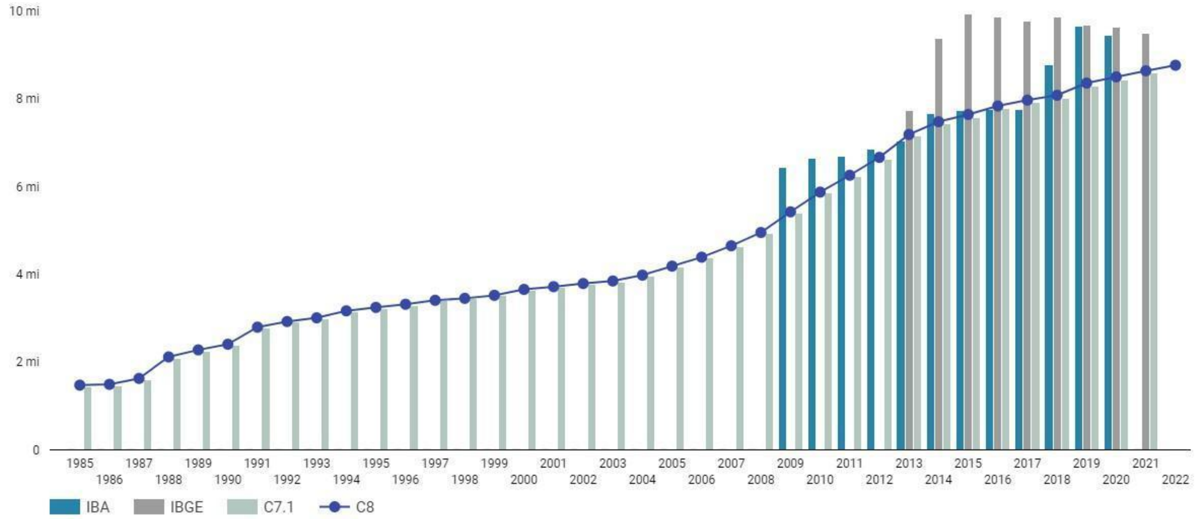


Figure 48. Comparison between MapBiomas, Brazilian Institute of Geography and Statistics (IBGE), and *Indústria Brasileira de Árvores* (Ibá), for the temporal series (1985-2022) of 'Forest Plantation' areas in Brazil.

Figure 49. Distribution by Brazilian states of the 'Forest Plantation' area mapped in Collection 8, for 2022.

6 References

ASSOCIAÇÃO BRASILEIRA DOS PRODUTORES DE ALGODÃO (ABRAPA). Algodão no Brasil. 2020. Available at: <https://www.abrapa.com.br/Paginas/dados/algodao-no-brasil.aspx>

AGÊNCIA NACIONAL DE ÁGUAS (ANA). Mapeamento do arroz irrigado no Brasil. Brasília: ANA & Conab, 2020.

AGROSATÉLITE. Agrosatélite Applied Geotechnology Ltd. Geospatial Analysis of Soy in the Cerrado Biome: Expansion Dynamics | Soy Agricultural Aptitude | Evaluation System for Financial Compensation: 2001 to 2019. - Florianópolis, 2020a 60 p. : il ISBN: 978-65-991465-1-0. Available at: <https://agrosatelite.com.br/wp-content/uploads/2020/06/relatorio_ccm.pdf>.

AGROSATÉLITE. Agrosatélite Applied Geotechnology Ltd. Soy Moratorium|crop season 2018/2019|. Monitoring by satellite images of soybean plantations in non-compliance with the soy moratorium. - Florianópolis, 2020b. 56p. Available at: <https://agrosatelite.com.br/cases/#moratoria>.

Breiman, Leo et al. Classification and regression trees. Routledge, 2017.

Breiman, Leo. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.

CONAB. Companhia Nacional de Abastecimento. Mapeamento Agrícolas. Available at: <https://www.conab.gov.br/info-agro/safras/mapeamentos-agricolas?start=20>.

Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, *58*(3), 257-266.

IBGE. Instituto Brasileiro de Geografia e Estatística. Produção Agrícola Municipal - PAM. Available at: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9117-producao-agricola-municipal-culturas-temporarias-e-permanentes.html?=&t=o-que-e>.

IBÁ. Indústria brasileira de árvores, Available at: <https://www.iba.org/publicacoes/relatorios>.

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote sensing of Environment*, *112*(10), 3833-3845.

Li, C., Ma, Z., Wang, L., Yu, W., Tan, D., Gao, B., ... & Zhao, Y. (2021). Improving the accuracy of land cover mapping by distributing training samples. *Remote Sensing*, *13*(22), 4594.

PEVS. Produção da Extração Vegetal e da Silvicultura, PEVS. Available at: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9105-producao-da-extracao-vegetal-e-da-silvicultura.html?=&t=downloads>

Potapov, P., Hansen, M. C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., ... & Ying, Q. (2020). Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, *12*(3), 426.

Song, X. P., Hansen, M. C., Potapov, P., Adusei, B., Pickering, J., Adami, M., ... & Tyukavina, A. (2021). Massive soybean expansion in South America since 2000 and implications for conservation. *Nature sustainability*, *4*(9), 784-792.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

Rudorff, B. F. T., de Aguiar, D. A., da Silva, W. F., Sugawara, L. M., Adami, M., & Moreira, M. A. (2010). Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. *Remote sensing*, *2*(4), 1057-1076.

Global Forest Watch Transparent World. Tree Plantations. 2015. Accessed through Global Forest Watch in January 2017. Available at: <www.globalforestwatch.org>.

Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., ... & Auch, R. F. (2016). Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, *122*, 206-221.