Agriculture and Forest Plantation - Appendix

Collection 7

General coordinator
Bernardo Rudorff

Team
Adriana Moreira
Joelen Silva
Kênia Santos
Luciana Oliveira
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 the classification of 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 Random Forest algorithm. Specifically for MapBiomas Collection 5 and 6, the most important methodological change was the use of 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 another 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) classes, both as beta version, as well as improvement of Other Perennial Crops maps.

In Collection 7 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). Figure 1 presents the evolution history of MapBiomas Collections of ‘Agriculture’ and ‘Forest Plantation’ (Collections 1-7).

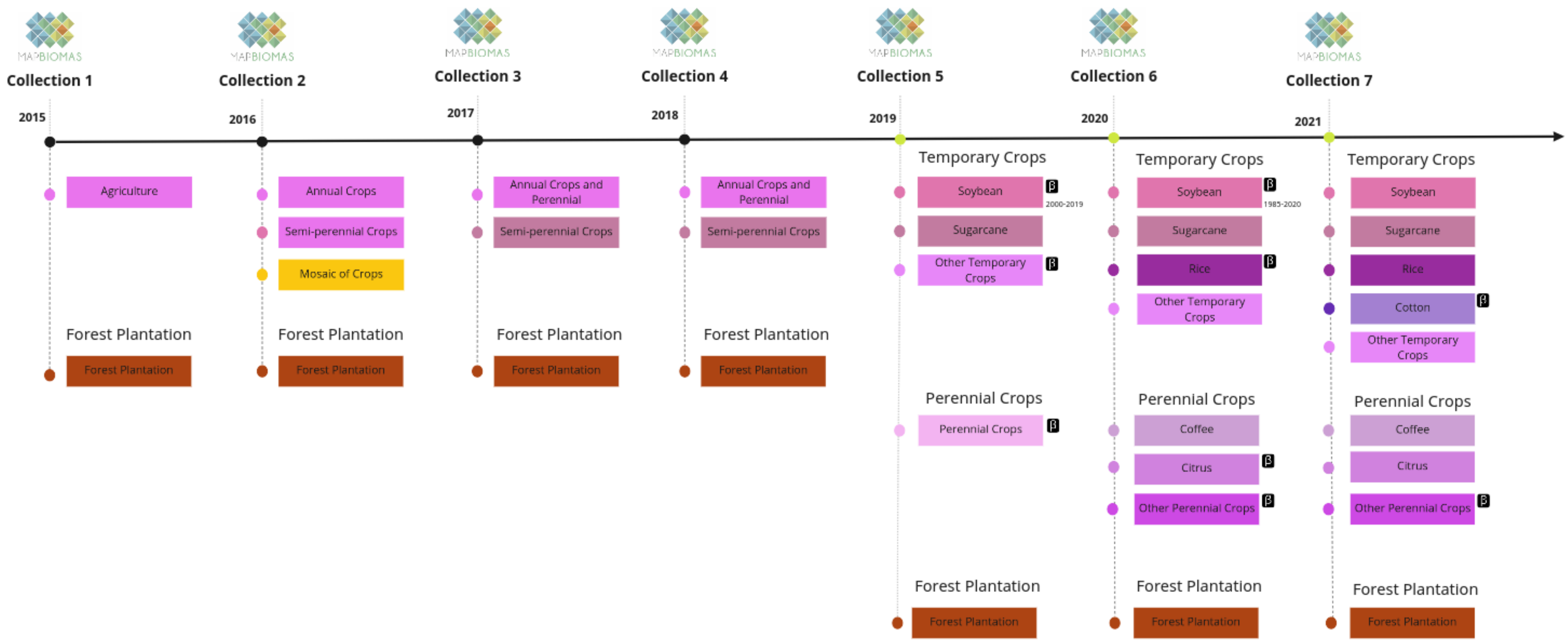


Figure 1: Evolution history of MapBiomas Collections of ‘Agriculture’ and ‘Forest Plantation’ (Collections 1-7).

Overall, ‘Agriculture’ classes comprises ‘Temporary Crops’ and ‘Perennial Crops’. In ‘Temporary Crops’ are included important agricultural classes for the Brazilian economy, such as soybeans, sugar cane, rice, cotton (new class), and other temporary crops, a class corresponding to the other temporary crops that are not map individually yet by the project. Besides the addition of the cotton mapping, in the ‘Temporary Crops’, it was also developed a method based on the time series of vegetation index from the Moderate Resolution Imaging Spectroradiometer (MODIS) to automatize the period definition of growing season and off-season to compose the Landsat mosaics. Furthermore, another approach was used to select the mosaic bands according to a metric of importance, and the sampling was revisited and improved.

Regarding the ‘Perennial Crops’, we made a great effort to seek improvements, especially in the citrus class, with the collection of new reference maps, obtained via visual interpretation through the use of Planet Scope images, for the states of Paraná and Minas Gerais, in addition to the map already produced for the state of São Paulo. For the coffee class, a new sampling approach was also used, in addition to obtaining a new reference map for the Espírito Santo state, in order to improve the mapping quality, reducing omission areas.

In ‘Forest Plantation’ class we made a huge effort to improve the map, mainly related to the type of sampling used, adopting the same sampling type that was used to the ‘Temporary Crops’ and to coffee classes, however, despite reduction of omission of this class (‘Forest Plantation’), errors of inclusion from other classes were verified, thus, the new sampling approach was used only to regions where was verified an increase in mapping quality. ‘Forest Plantation’ mapping was conducted based on a regional approach, in order to focus on important regions, increase the quality of classification and avoid misclassifications.

2 Classification

The MapBiomas-brazil account in GitHub has all the scripts used to classify ‘Agriculture’ and ‘Forest Plantation’ classes in MapBiomas Collection 7. The repository links are:

- Agriculture:

  https://github.com/mapbiomas-brazil/agriculture/tree/mapbiomas70

- Forest Plantation:

  https://github.com/mapbiomas-brazil/forest-plantation/tree/mapbiomas70

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

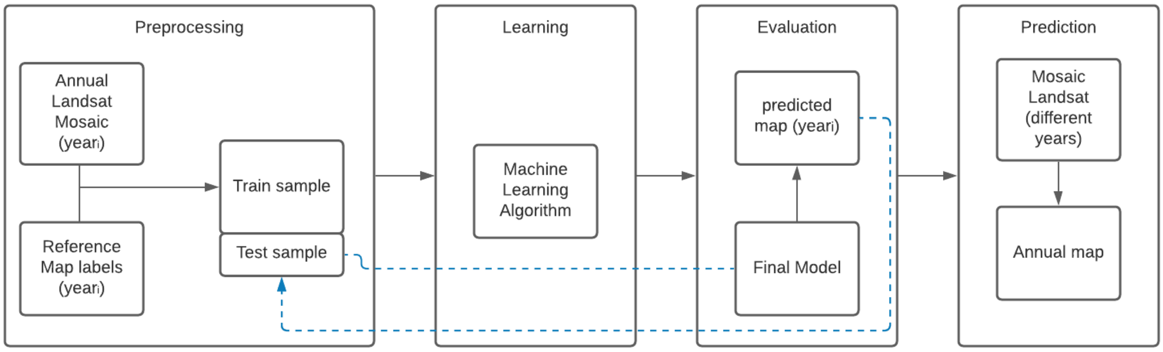


Figure 2. Supervised learning workflow in context of image classification.

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

The Landsat images availability for the Collection 7 period (1985 to 2021) varies among years. Throughout this period, Landsat 5 Collection 1 (1985 to 2012), Landsat 7 Collection 1 (1999 to present), and Landsat 8 Collection 1 (2013 to present) provided the images for the mosaics compositions. Figure 3 shows the variability of available Landsat images for Collection 7 period.

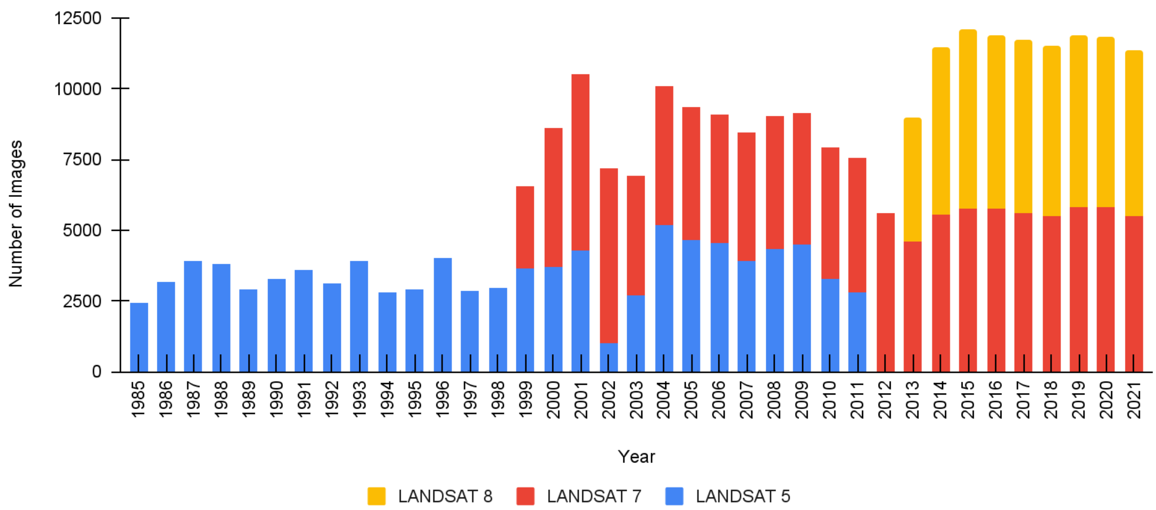


Figure 3. Number of available TOA Landsat Collection 1 images covering the Brazilian territory from 1985 to 2021.

Although Landsat 9 imagery has been available since October 2021, we have chosen not to use this data in the MapBiomas Collection 7.

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 types of land cover (for more information, see Potapov *et. al.* (2020)). Both collections of Landsat data were used in MapBiomas Collection 7. Table 1 shows the collection used for each class of 'Agriculture' and 'Forest Plantation' classification.

Table 1: Landsat collection used for each class of agriculture and forest plantation classification.

Level 1	Level 2	Level 3	Level 4	Landsat Collection
Farming	Agriculture	Temporary Crop	Cotton	Normalized Landsat Collection
			Soybean	Normalized Landsat Collection
			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 taken into account to better distinguish the class of interest from the remaining land cover and land use classes. For instance, for different types of agriculture crops and for different regions in Brazil, the growing season can cover different periods of the year, predominantly during the wet period, in most Brazil 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.

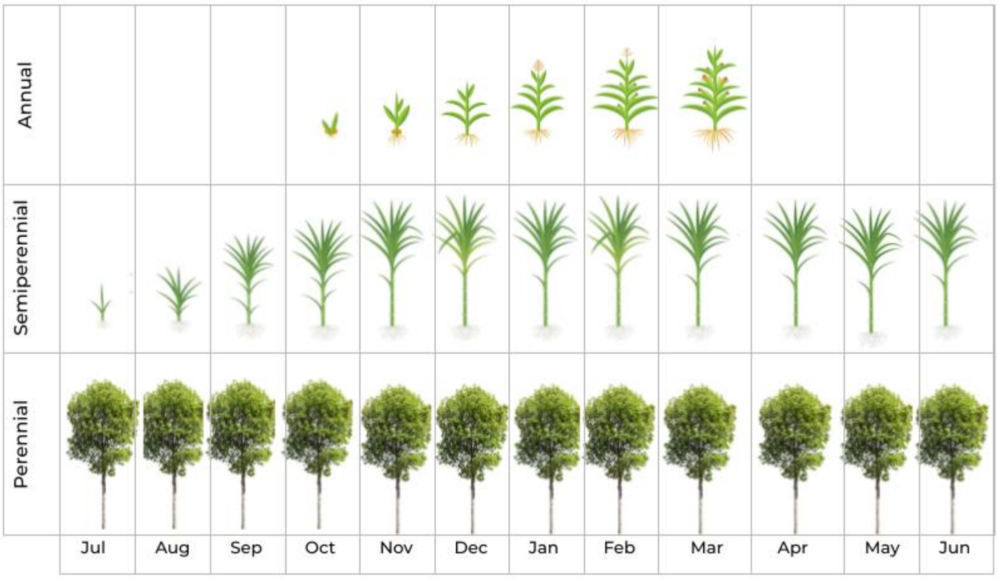


Figure 4. Temporal behavior of agricultural crops according to their cycle.

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 compose the mosaics.

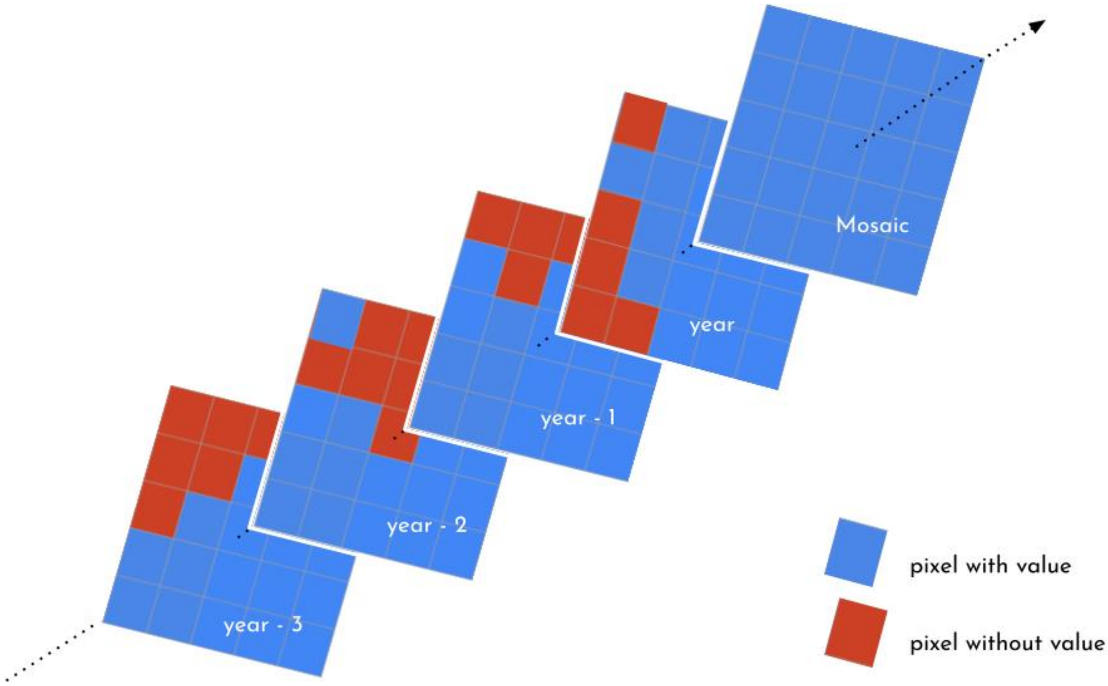


Figure 5. Scheme to compose Landsat mosaic. Pixels in red color do not have a valid value (due to cloud and/or shadow incidence), and pixels in blue color have a valid value. This approach seeks to fill pixels in red replacing them by pixels in blue from the next previous year, until to compose a Landsat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

2.1.3.1 Cotton, Soybean and Other Temporary Crops

Until Collection 6, we used a static growing season and off season periods to compose the mosaics for the all classes mapped. However, a temporal static calendar (static dates for beginning and end of growing season for all years) can often fail to follow variations in growing season, according to weather conditions or other factors related to management. For the Collection 7, we developed a methodology to automatize the periods of growing season and off-season, according to the phenological development of the culture, that was based on the EVI2 (Jiang et al., 2008). Through this method it is possible to obtain one calendar year by year from 2000.

To obtain year by year growing season and off-season calendars, a time series of EVI2 data calculated from MODIS was smoothed using the Fourier Transform in order to minimize variations. In the annual curves for each pixel, the vegetative peak value in agricultural regions was identified. The peak month was identified, and a reduction by Landsat scene was done in order to identify the mode of this month in each scene. Then, the peak vegetative month for each year in each scene was obtained, which corresponds approximately 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 to 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 from the period between 2000 until 2021.

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

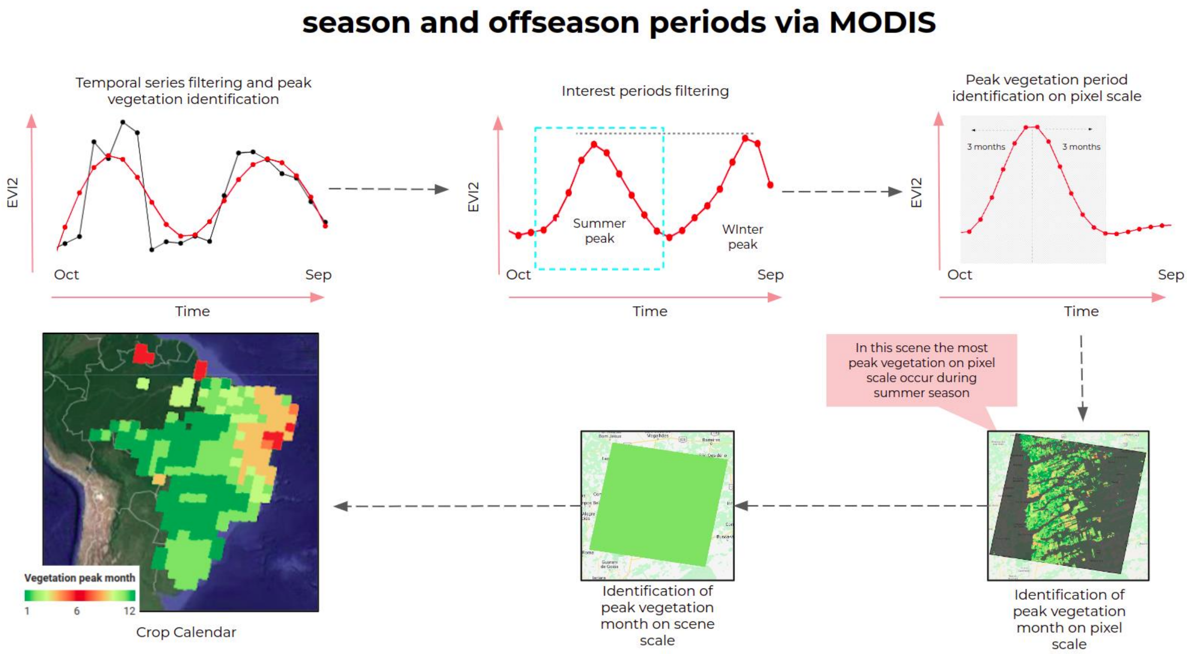


Figure 6. Scheme to obtain vegetation peak month, year by year, to Landsat scene.

An example of the results of the annual vegetative peaks per Landsat scene is shown in Figure 7.

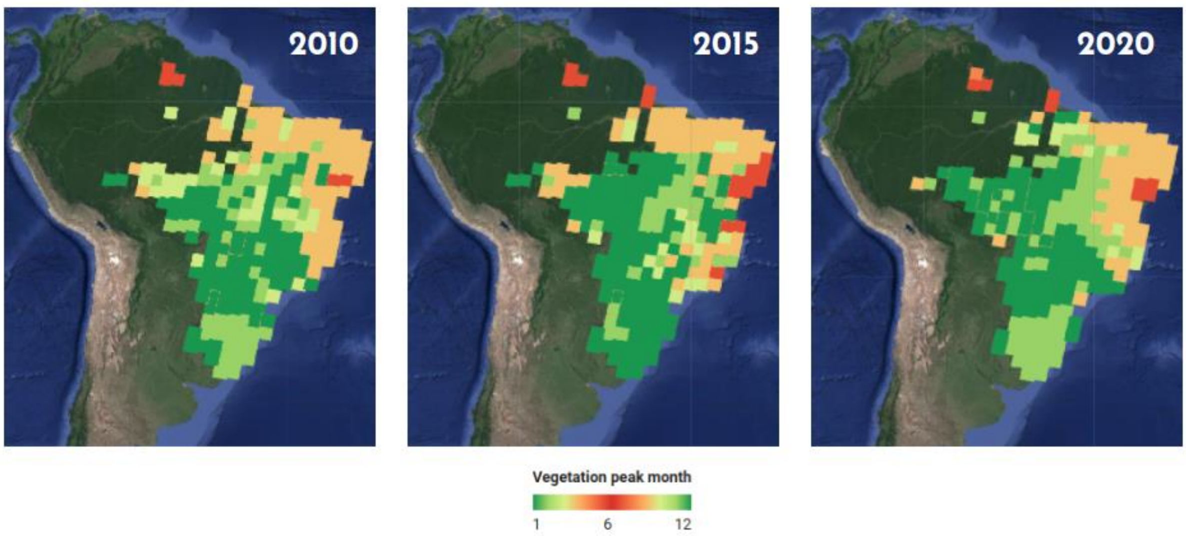


Figure 7. Vegetation peak month, obtained using MODIS EVI2 series smoothed for each Landsat scene for 2010, 2015 e 2020 years.

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 month 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 for the selection of mosaic images of cotton, soybean and other temporary crops in Collection 7.

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 the selection of mosaic images of sugar cane in Collection 7.

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

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 7.

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 (e.g. 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. Therefore, at the first moment, the 'Perennial Crops' classes were divided into three subclasses: coffee, citrus and Other Perennial Crop. The last one doesn't distinguish between types of crops. For the coffee, citrus and Other Temporary Crop classifications, a median of annual mosaic (i.e. 01-01-year to 12-31-year) was obtained (Table 5).

Table 5. Periods used for the selection of mosaic images of Perennial Crop in Collection 7.

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

2.1.3.5 Forest Plantation

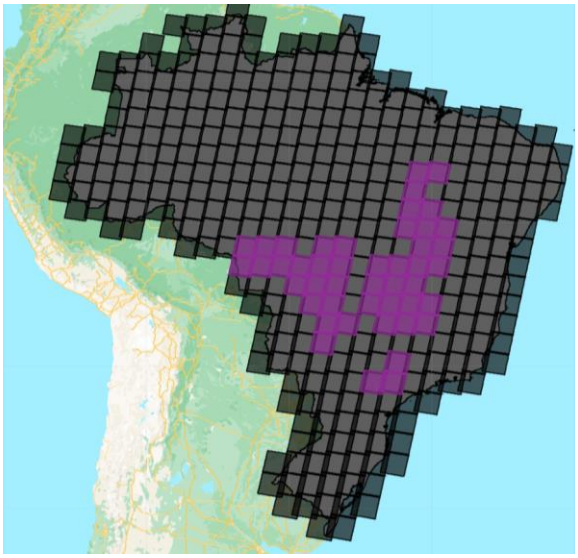
Two periods were defined to compose the Landsat mosaics to classify forest plantation, covering from January from the current year until January of the next year (Table 6).

Table 6. Periods used for the selection of mosaic images of forest plantation in Collection 7.

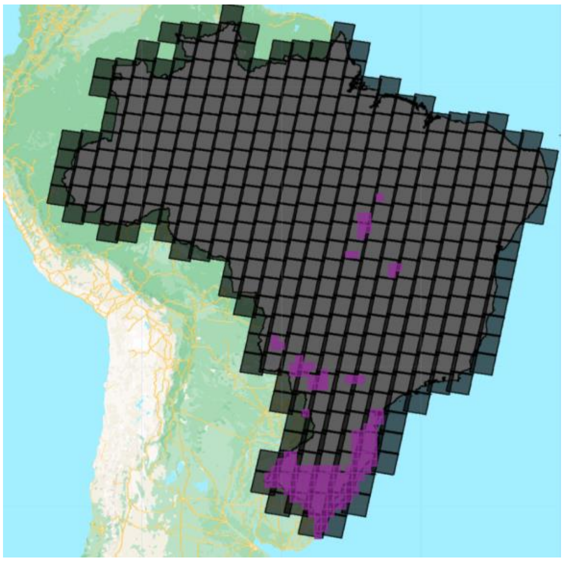
Period	Start	End
P1	01/01/year	07/01/year
P2	07/01/year	01/01/year+1

2.1.4 Definition of regions for classification

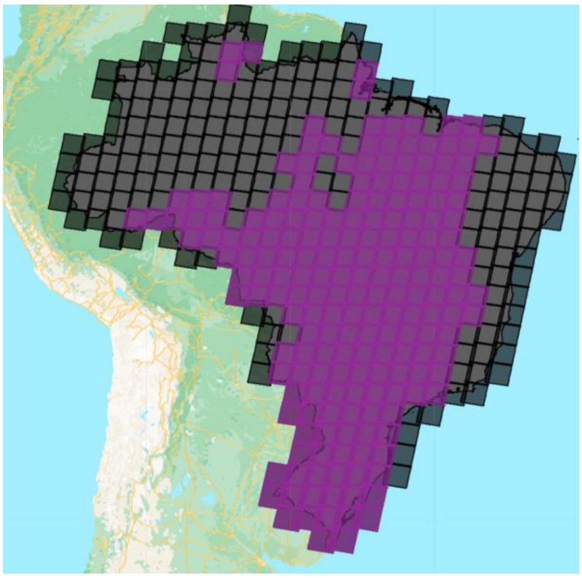
The ‘Agriculture’ and ‘Forest Plantation’ are heterogeneously distributed in the Brazilian biomes. Therefore Landsat scenes were selected in regions with the highest occurrence of each class according to the reference maps. Figure 8 illustrates the scenes chosen for each land use class.



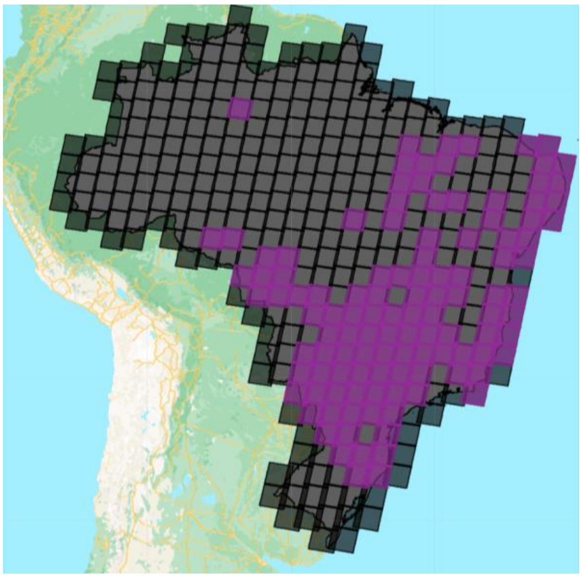
Cotton



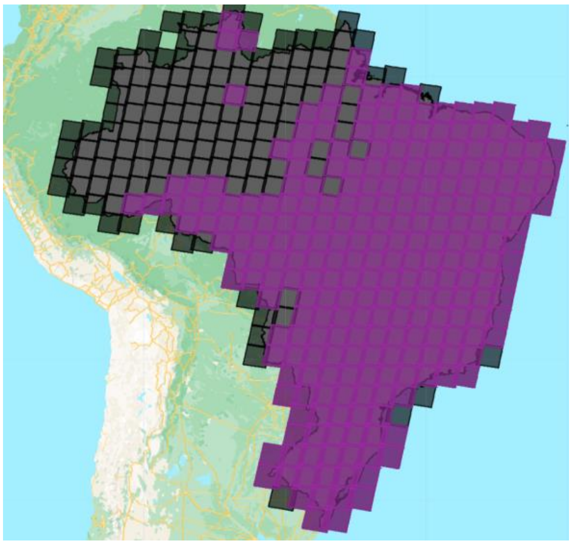
Rice



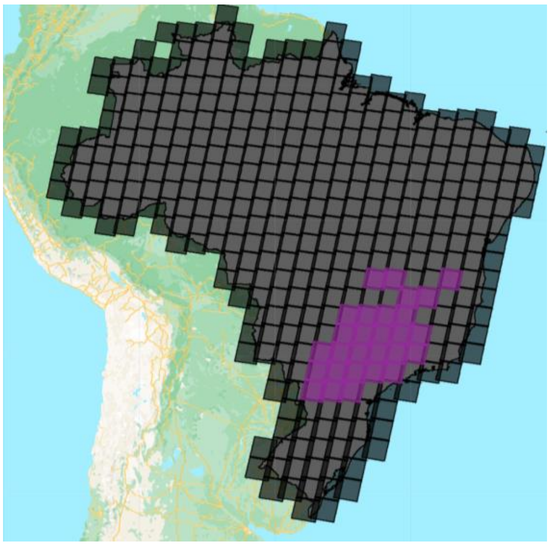
Soybean



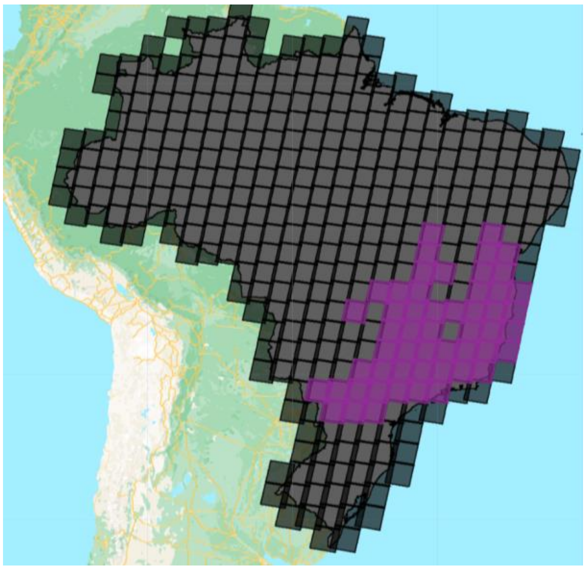
Sugar cane



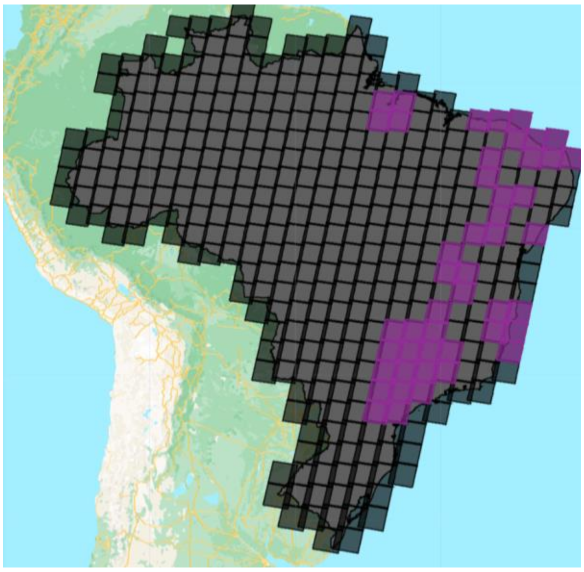
Other Temporary Crop



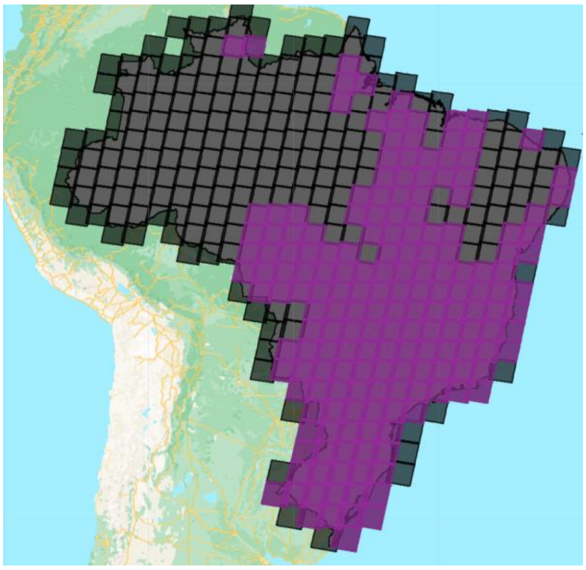
Citrus



Coffee



Other Perennial Crop



Forest Plantation

Figure 8. Selected scenes of Landsat series to the classification of 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), in Collection 7. This limitation is due to two main reasons: *i*) reference maps coverage, a challenge to map areas where there are not reference maps, and *ii*) spatial scale, a challenge for mapping smaller scale crops using landsat images.

2.1.5 Feature space

For Collection 7 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.

2.1.5.1 Cotton, Soybean and Other Temporary Crop

The total band of the mosaic consists of metrics calculated from Landsat bands and indices over three time periods, with a total of 234 bands. The visible, near-infrared and shortwave bands, as well as statistical metrics and vegetation indices, 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 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, 80th percentile, 20th percentile, and CEI (NDWI, NIR, EVI2)

The dimensionality reduction process corresponds to the selection of only the most important 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 there is a step that aims to identify a metric of 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 have to be cautious about the effectiveness of this index, especially due to the possibility of overestimating the importance of correlated predictors. This method was used at this time because of the potential to reduce 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 (e.g. other types of agricultural 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	Paraná	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 7.

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 7.

Bands	RED, NIR, SWIR1
Metrics	median

2.1.5.6 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.

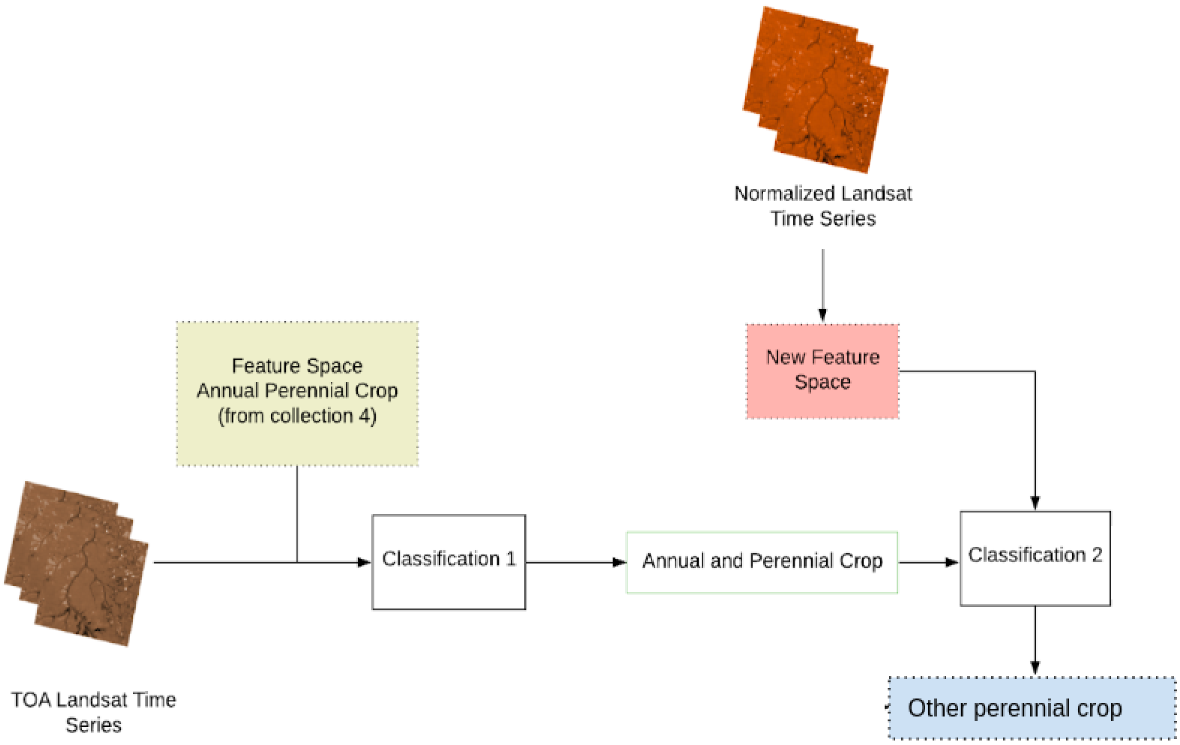


Figure 9. Steps to separate Other Perennial Crop from the previous class 'Annual and Perennial Crop' of the Collection 4.

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 12).

Table 12. Bands, indexes and metrics used to classify other 'Perennial Crops' in MapBiomas Collection 7.

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

2.1.5.7 Forest Plantation

The bands, indexes and metrics used to classify forest plantation in the Collection 7 are presented in Table 13.

Table 13. Bands, indexes and metrics used to classify forest plantation in MapBiomas Collection 7.

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

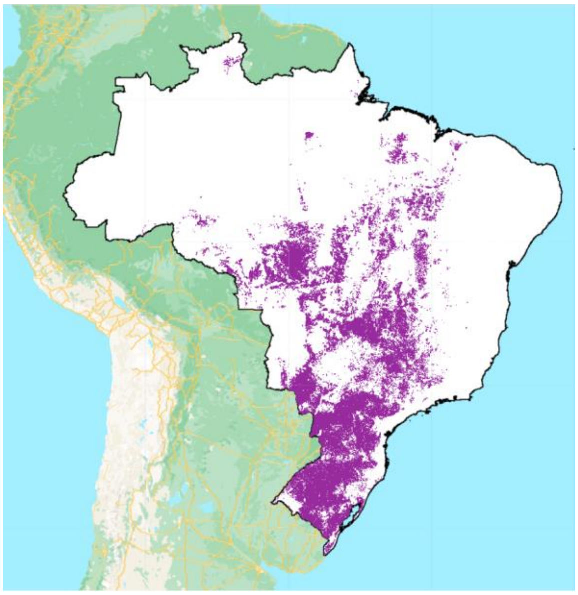
The reference maps used to obtain samples to train the classifier are shown in Table 14.

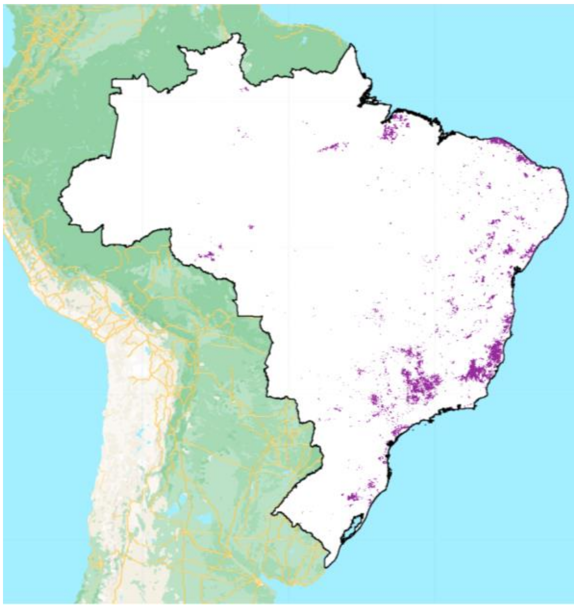
Table 14. Reference maps used in the Random Forest classification for the classes 'Agriculture' and 'Forest Plantation' in Collection 7.

Class	Landsat time series	Number of training samples	Sampling Approach	Rule	Type	Year of acquisition	Reference
Soybean (2000 - 2021)	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
Other Perennial Crop	Normalized	5,000		Minimum of 20% for the interests 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

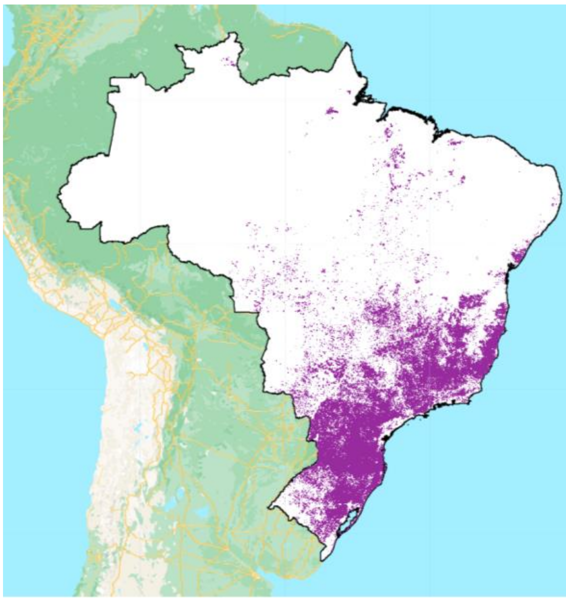
(Agrosatélite, 2020b)

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)

Figure 10. Reference maps representing the areas with training samples for the classification of ‘Agriculture’ and ‘Forest Plantation’ in Brazil in Collection 7.

2.1.6.2 Random Forest

For the classes mapped by 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. In order 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.

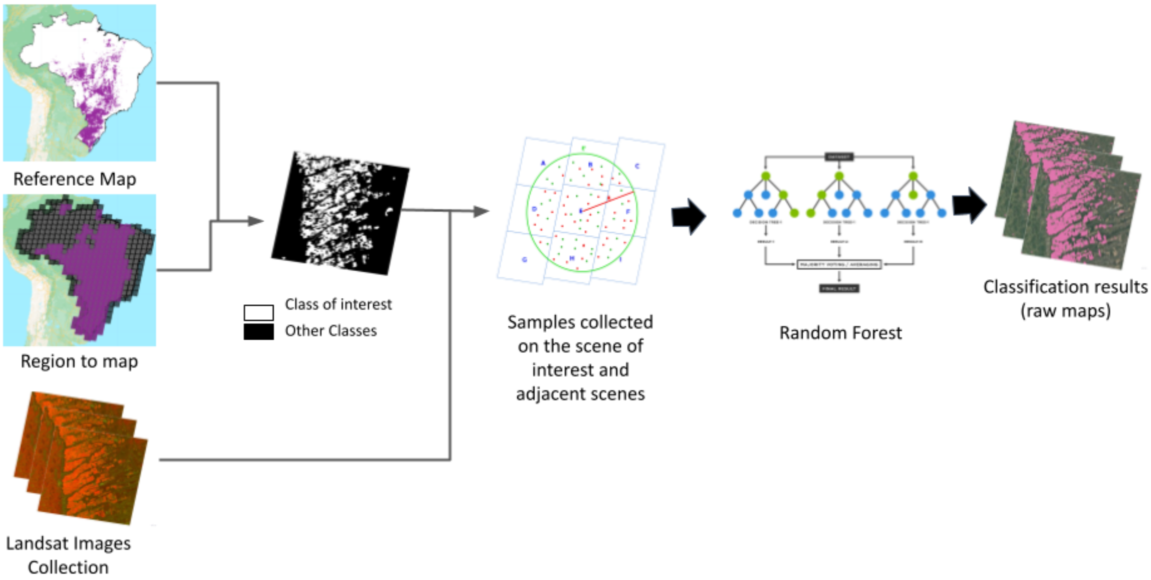


Figure 11. Fluxogram of agriculture and Forest Plantation classification using Random Forest algorithm.

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.

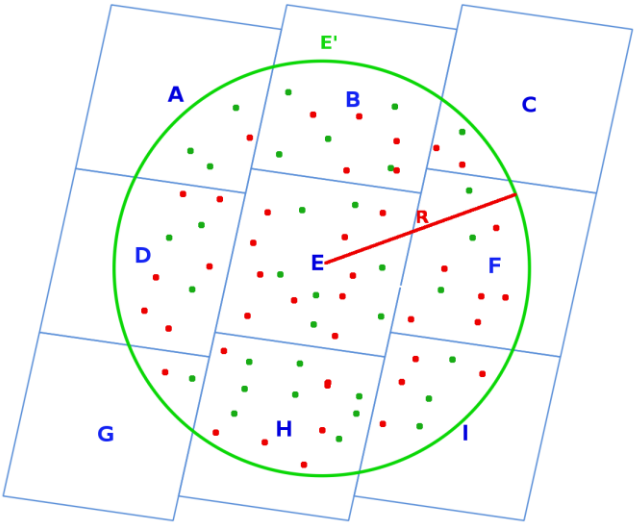


Figure 12. Scheme for sample acquisition for the regionalized training of the Random Forest classifier in Agriculture and Forest Plantation.

As there are no reference maps available for all classes in all years of the time series (1985 to 2021), stable samples were created. However, these samples were only obtained in classes which used the normalized Landsat series, due to the characteristics of this time series mentioned above. For the 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 algorithms 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 take into account 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.

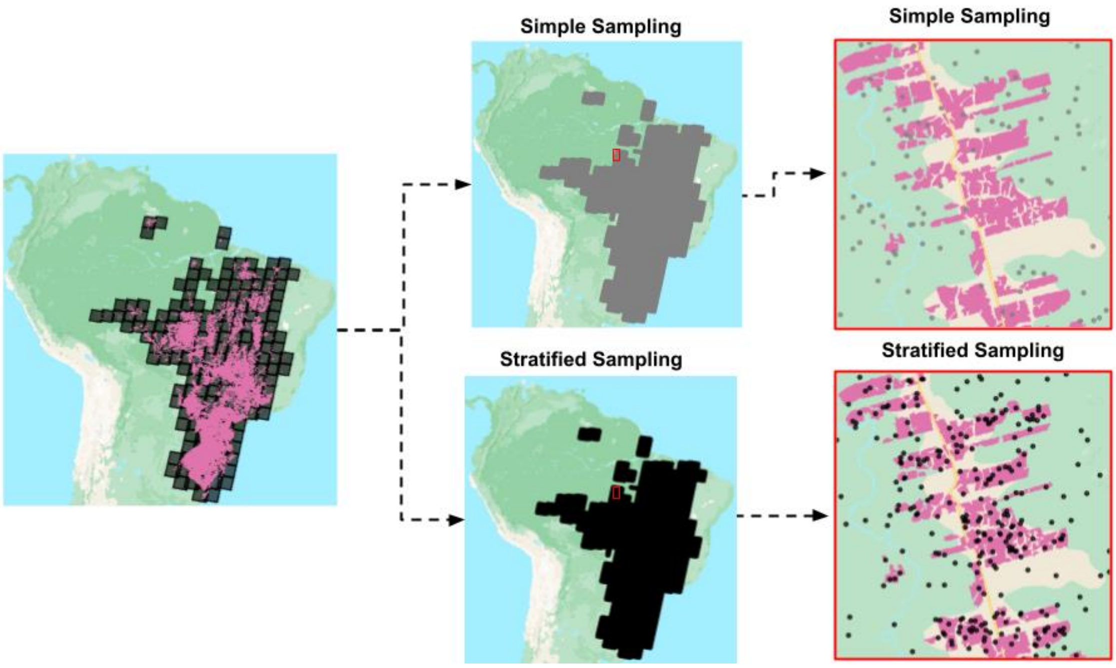


Figure 13 - Difference between sampling approaches. a) simple sampling, and b) stratified sampling.

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 just the spectral response for each pixel, this architecture uses the context in which the pixels are. This architecture is illustrated in Figure 14.

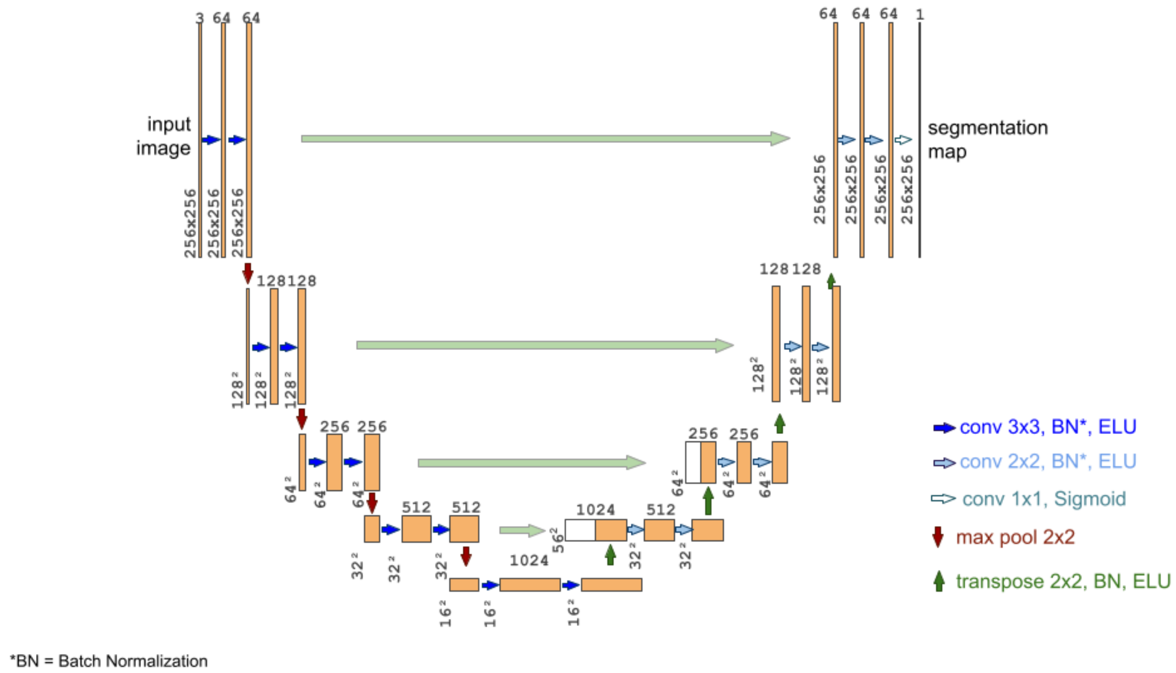


Figure 14 - Adapted U-Net convolutional neural network, with its layers and connections, used for the mapping of rice and citrus.

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 and with the states of interest of Collection 7, as illustrated in Figure 15.

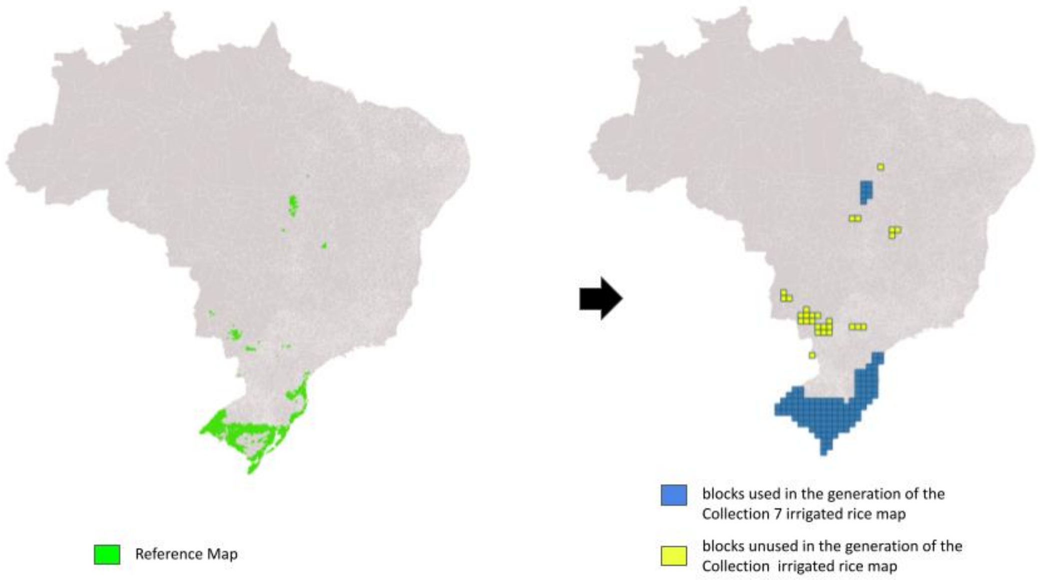


Figure 15. Study area used for the mapping of irrigated rice in the MapBiomas Project.

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.

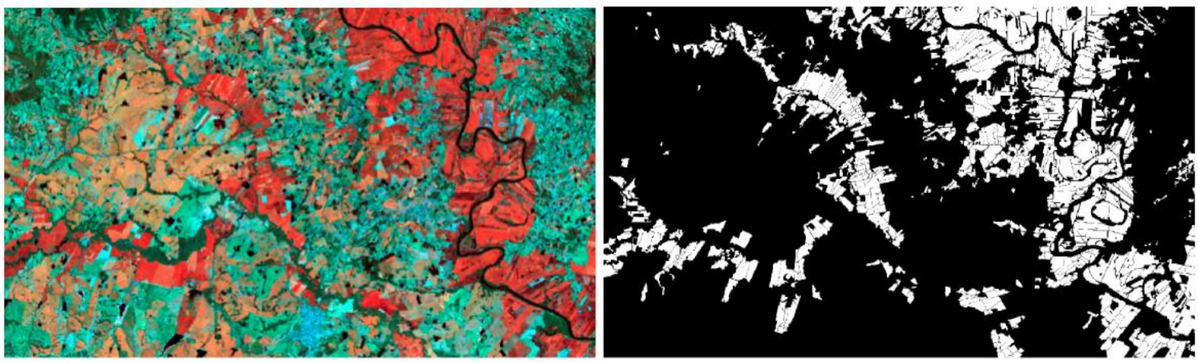


Figure 16. Example of U-Net samples to mapping rice.

The test data were used for 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 (1985-2021).

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.

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 less pixels of the interest class or the "others" class (Figure 17).

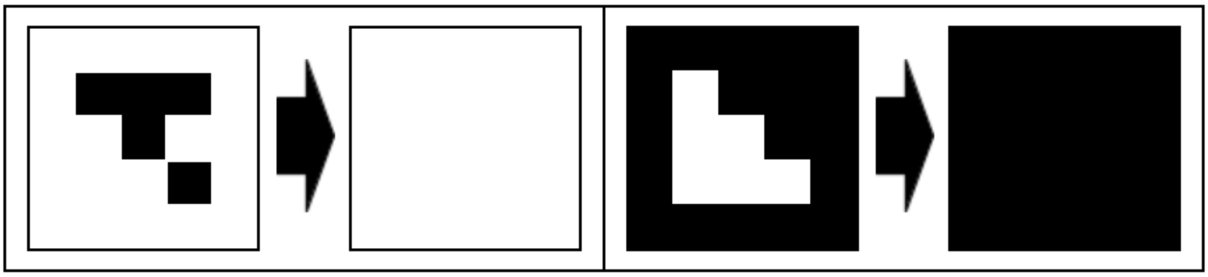


Figure 17. Example of the minimum connected pixels spatial filter. The image on the left shows an exclusion of pixels of the interest class (in black). The image on the right shows an inclusion of pixels of "other classe" (in white) to the interest class.

3.2 Temporal filter

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

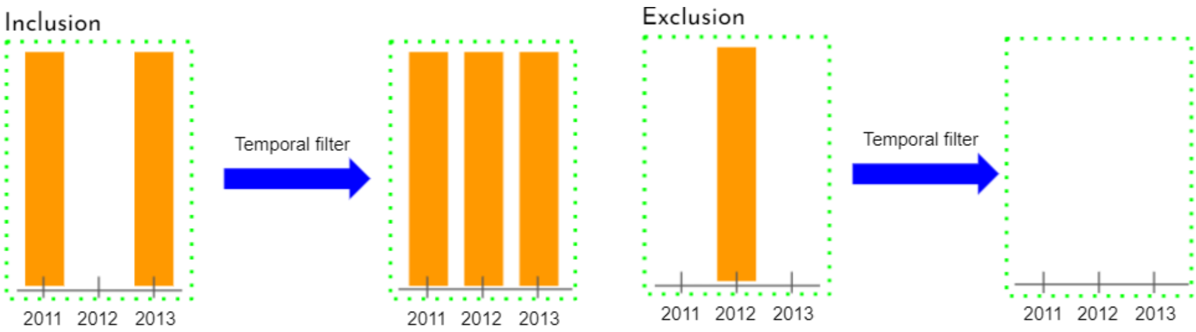


Figure 18. 3-years temporal window filter: The orange bars represent pixels of the mapped class (interest class). The exclusion filter changes a pixel to "others" class when the same pixel was not of the interest class in the adjacent years. The inclusion filter changes a pixel to the interest class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

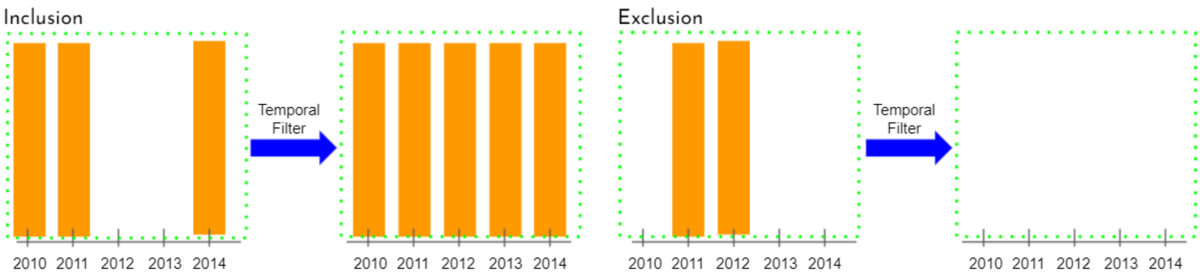


Figure 19. 5-years temporal window filter: The orange bars represent pixels of the mapped class (interest class). The exclusion filter changes a pixel to "others" class when no more than 1 another year is of the interest class. The inclusion filter changes a pixel to the interest class when at least 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.* 2021), no temporal filter was applied.

3.2.1 Soybean, Cotton and Other Temporary Crops

For MapBiomas Collection 7, the soybean, cotton and other temporary crops classes followed an unified process, represented in Figure 20.

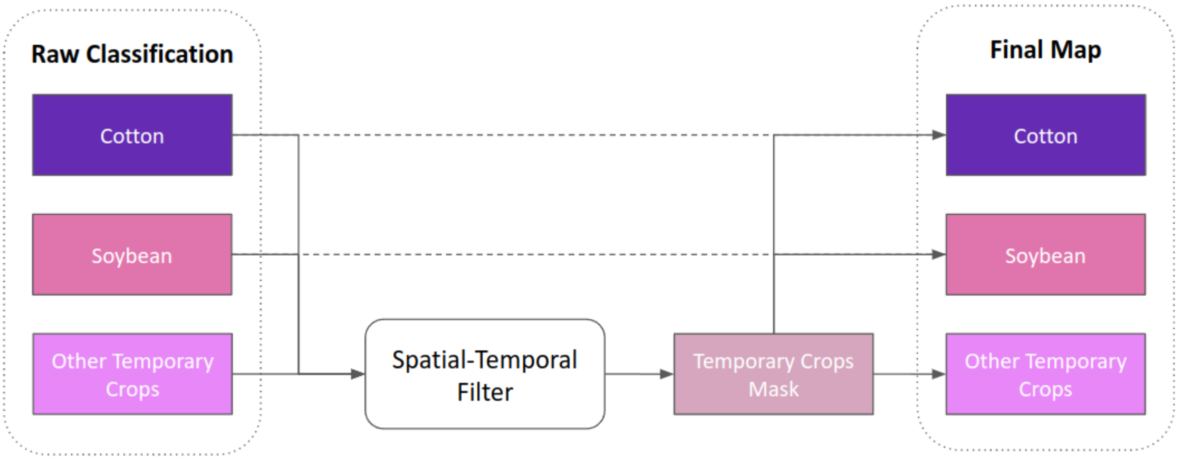


Figure 20: Soybean, cotton and other temporary crops classification and post-classification fluxogram.

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 years 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, at the same time that the annual crop variation on 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). The initial year (1985) and final year (2021) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2019.

3) Temporal filter using 5 years with 2 years threshold applied only on the final edge year (2020).

4) Temporal filter using 3 years with 2 years threshold applied to all series, except to the edge years (1986-2020) 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 2020) 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 2020). The initial year (1985) and final year (2021) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2019.

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 2020). The initial year (1985) and final year (2021) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2019.

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 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.

3.2.9 Forest Plantation

For the forest plantation class, it was applied the same temporal filters applied to coffee class:

1) Temporal filter using 3 years with 2 years threshold applied only on the edge years (1986 and 2020). The initial year (1985) and final year (2021) no temporal filter was applied.

2) Temporal filter using 5 years with 3 years threshold for the period 1987 to 2019.

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 plantations 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.

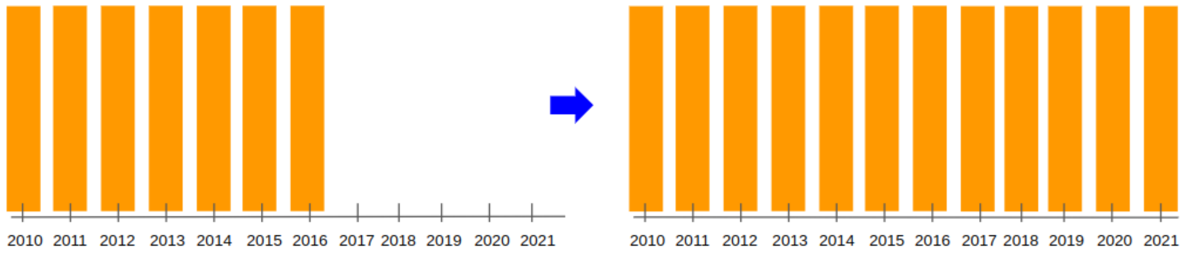


Figure 21. Temporal filter applied in the last years of the forest plantation series.

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.

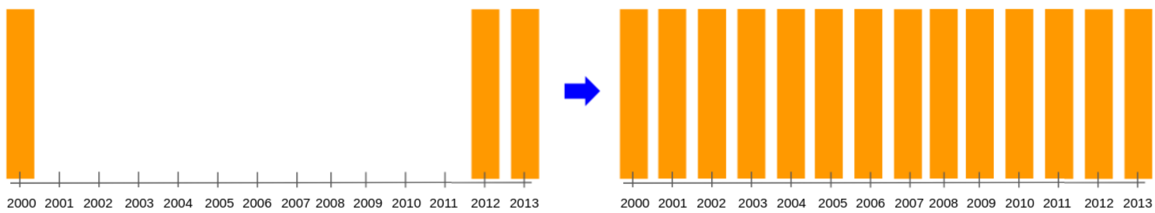


Figure 22. Temporal filter that converted longer intervals into forest plantation when it was in the past and became again years after.

4 Integration with biomes and themes

After the classification of the 'Agriculture' and 'Forest Plantation' themes, they were integrated to the other land use and land cover classes to compose the MapBiomas Collection 7 final maps. This integration process was based on the overlap 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 Crop' 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 Crop' 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.

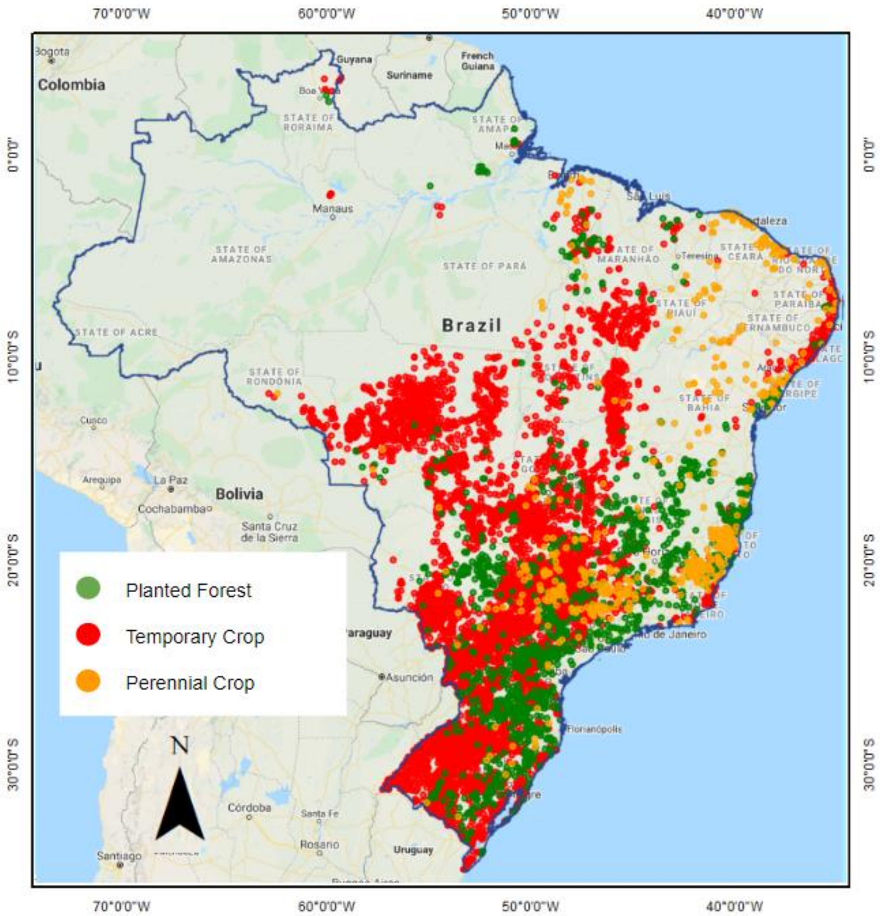


Figure 23. LAPIG points that were used for the accuracy assessment of 'Temporary Crops', 'Perennial Crops' and 'Forest Plantation' classes.

5.1.1 Temporary Crops

The results of accuracy assessment of the ‘Temporary Crops’ class are presented by Figures 24 and 25, covering the period from 1985 until 2018 (final date of validation collection).

Figure 24 presents a comparison between the producer’s accuracy of Collection 6 and 7. We verified that the Collection 7 the producer’s accuracy was similar or slightly increased throughout the years when compared to the Collection 6.

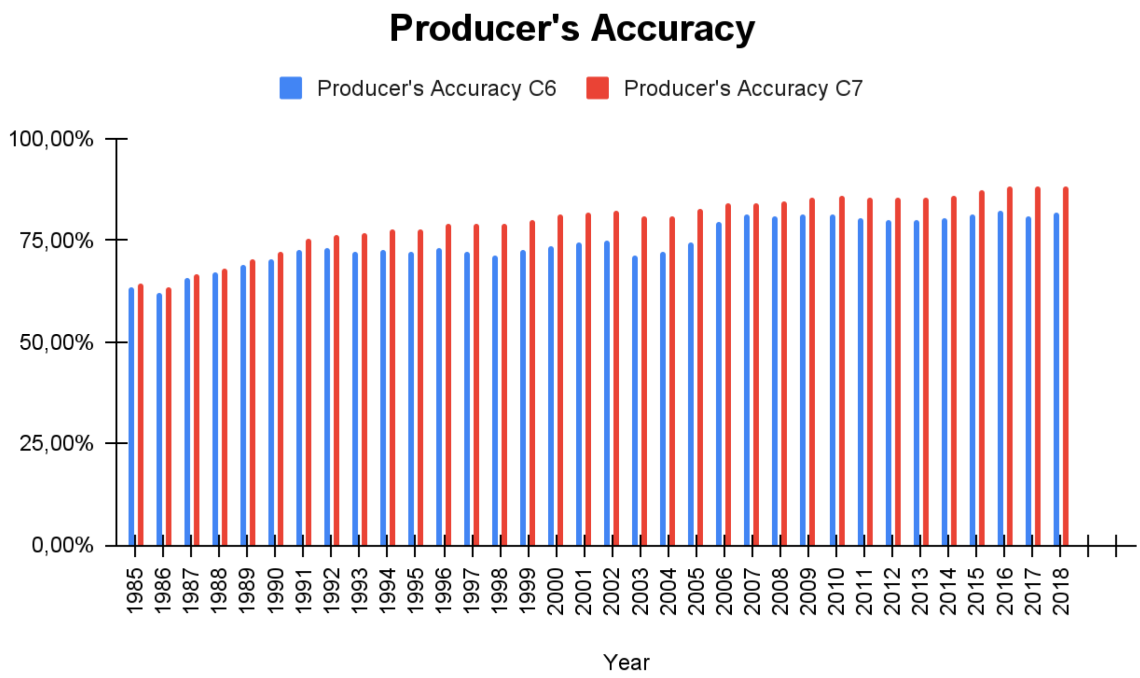


Figure 24. Comparison of producer’s accuracy of the ‘Temporary Crop’ class in Collections 6 and 7.

When we evaluate the user's accuracy we can note a clear increase of the accuracy in Collection 7 compared to Collection 6 (Figure 25). Through all years of the temporal series, user's accuracy in Collection 7 was higher than in Collection 6, demonstrating an average increase of approximately 9.5% in accuracy. These results reinforced that the improvements of the Collection 7 provide maps with lower or similar omission errors than Collection 6 and lesser inclusion of misclassified pixels in the ‘Temporary Crop’ map than the last collection.

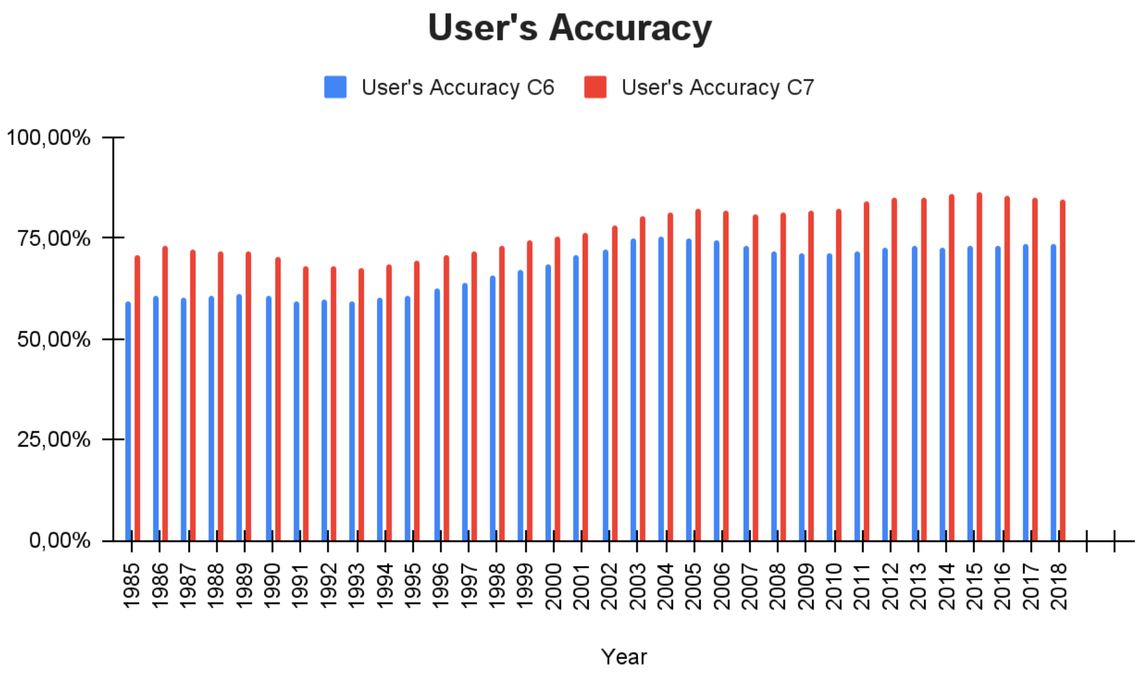


Figure 25. Comparison of user's accuracy of the 'Temporary Crop' class in Collections 6 and 7.

5.1.2 Perennial Crops

Figures 26 and 27 show the differences between the producer and user accuracies for the 'Perennial Crop' class of Collection 7 and their comparison with Collection 6.

Overall, the producer's accuracy comparison between Collection 6 and 7 (Figure 26) demonstrates consistently higher percentages accuracy for the 'Perennial Crops' mapped in Collection 7 than in Collection 6. Furthermore, the producer's accuracy is higher since the 2000s, a period with greater availability of Landsat images. For instance, at the beginning of the series in Collection 7, it was verified producer's accuracy less than 10%, while from 2000 until 2010, this value reached close to 20%, and in the lastest years (from 2010 until 2018), Collection 7 reaches a producer's accuracy of approximately 30% for 'Perennial Crops' class.

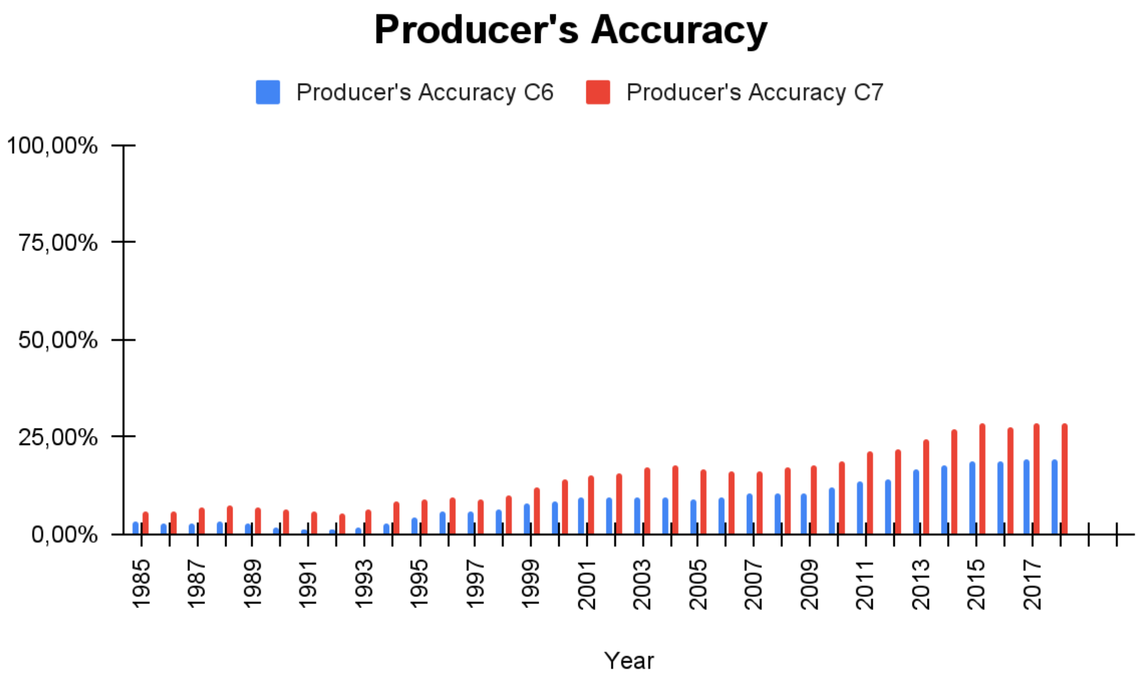


Figure 26. Comparison of producer's accuracy of the 'Perennial Crops' class in Collections 6 and 7.

Different results were found when we evaluated the user's accuracy in Collection 7 compared to Collection 6. For instance, at the beginning of the series, user's accuracy was close to 25% for Collection 7. Higher values were observed in Collection 7 between the period 1985 to 1995, while from 1995 until 1999, there was observed a decrease in user's accuracy in Collection 7 compared to Collection 6. From 2000 until the end of the time series, the user's accuracy is comparable between the Collections. Furthermore, higher percentages of user's accuracy are observed for the period after 2000, where the values are higher than 50% from 2000 until 2018.

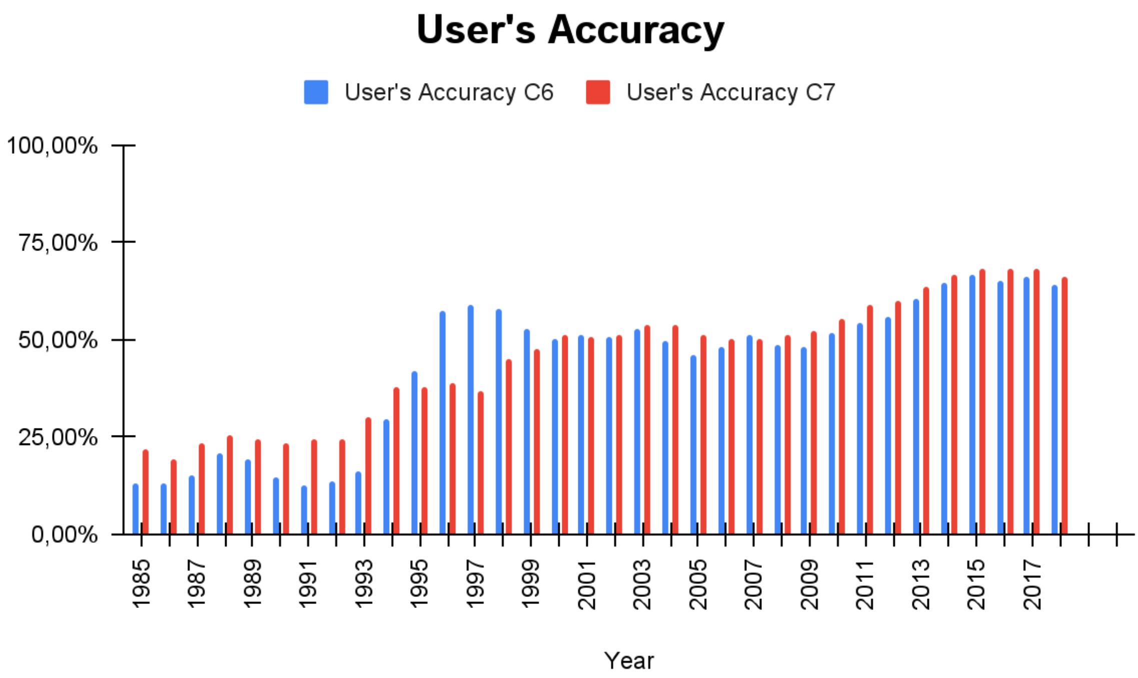


Figure 27. Comparison of user's accuracy of the 'Perennial Crop' class in Collections 6 and 7.

5.1.3 Forest Plantation

Figures 28 and 29 present the results of user and producer accuracies for 'Forest Plantation' class of Collections 6 and 7.

For Collection 7, the 'Forest Plantation' class underwent a methodology reformulation, which impacted positively the final map. The main impact was in the producer's accuracy, which went from an average of 49,7% in Collection 6 to 79,88% in Collection 7. Figure 28 presents this comparison between producer's accuracy over the years in both collections.

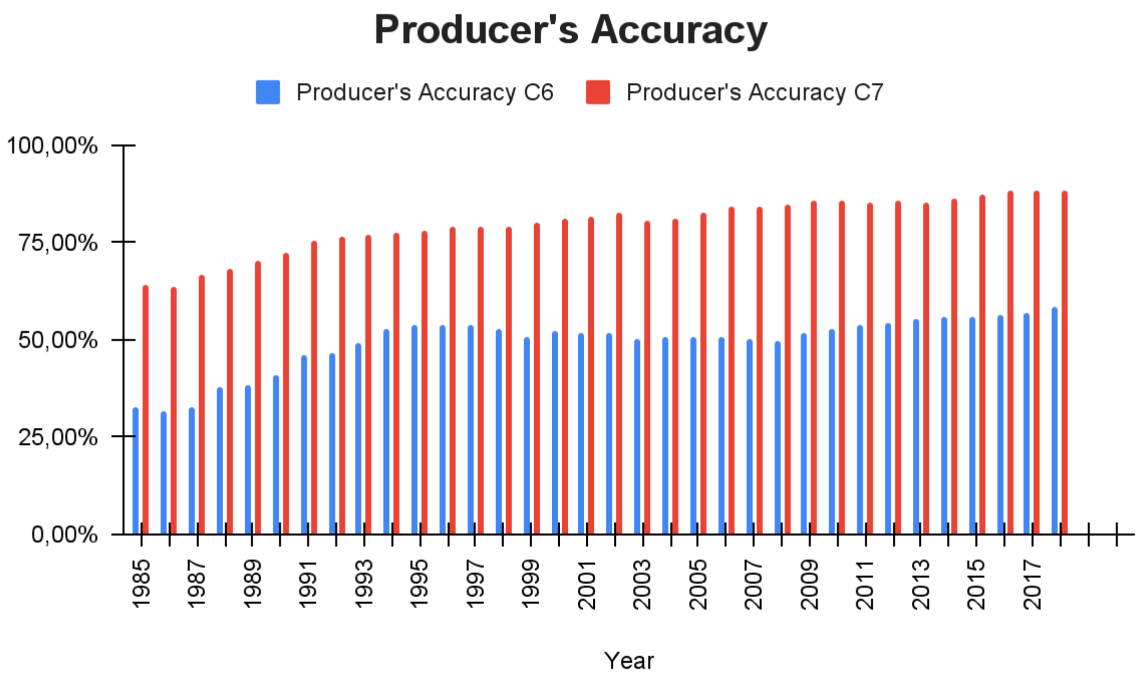


Figure 28. Comparison of producer's accuracy of the 'Forest Plantation' class in Collections 6 and 7.

Regarding the user's accuracy, there was a decrease in Collection 7 compared to Collection 6. In the previous collection, on average, the 'Forest Plantation' map had an user's accuracy of 91,08% over the years, which reduced to 77,27% in Collection 7.

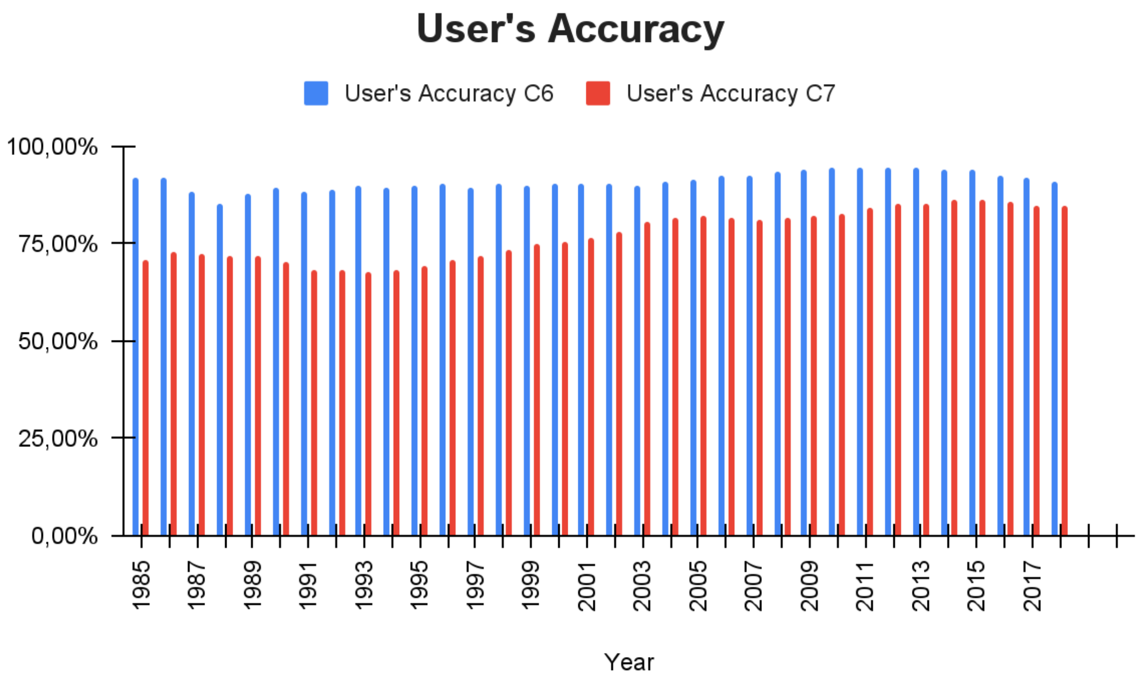


Figure 29. Comparison of user's accuracy of the 'Forest Plantation' class in Collections 6 and 7.

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 7 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) and Ibá - *Indústria brasileira de árvores* (Brazilian Tree Industry), was also made. These are considered official data for estimating agricultural areas in the country.

5.2.1 Comparison of Temporary Crop area

Figure 30 presents a comparison between the area of the class 'Temporary Crops' (which includes sugar cane, rice, soybean, cotton and other temporary crops) with the areas estimated by PAM - IBGE (sugar cane, corn and soybean).

Overall, 'Temporary Crops' mapped by MapBiomas Collection 7 presents a temporal dynamic similar to data from IBGE, showing a pronounced increase over the years.

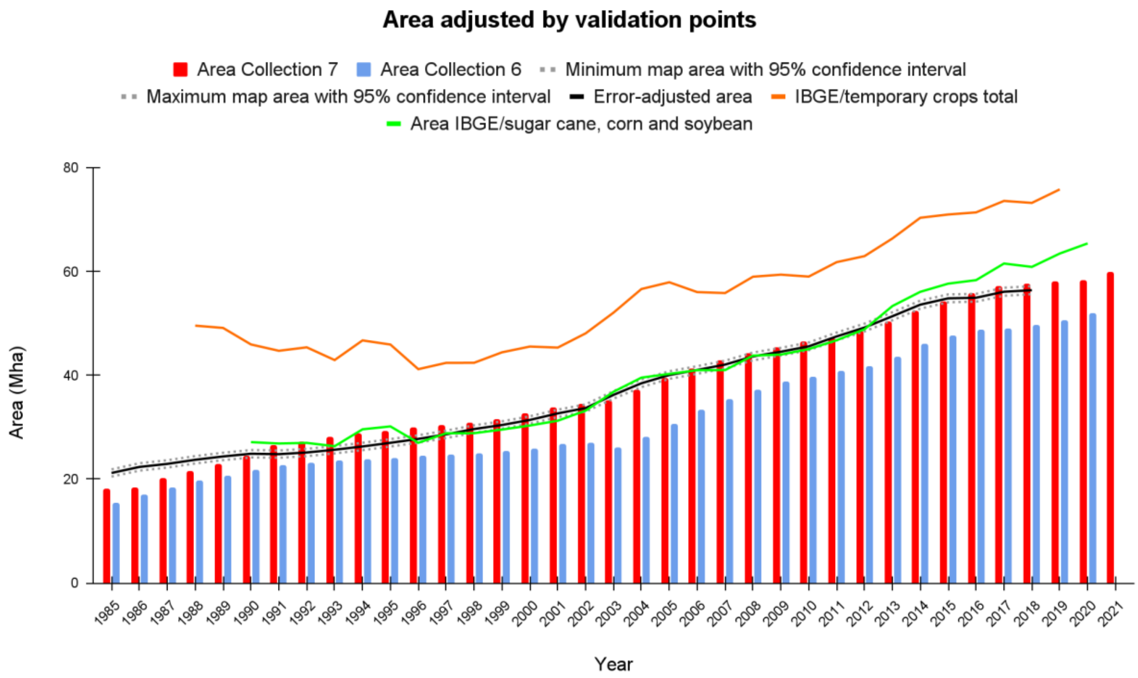


Figure 30. Comparison between MapBiomas 'Temporary Crops' area and LSPA-IBGE Temporary Crop area.

5.2.2 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 used in Collection 7 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 (banana, coffee and orange) that are not totally mapped by MapBiomas (MapBiomas maps citrus, coffee and some concentrations of perennial crops spread throughout the territory). Figure 31 shows the comparison between the MapBiomas Collection 7 perennial area and PAM 'Perennial Crops' area.

The comparison of the 'Perennial Crops' data mapped by Collection 7 and estimated by IBGE demonstrated that MapBiomas still maps about 57% less perennial crop area than IBGE. This difference is greater at the beginning of the time series, about 68% less area mapped by Collection 7 between 1990 and 2010. This difference decreased between 2010 and 2020, when the area of perennial tillage mapped by Collection 7 reached 1 Mha (32% less than the IBGE estimates).

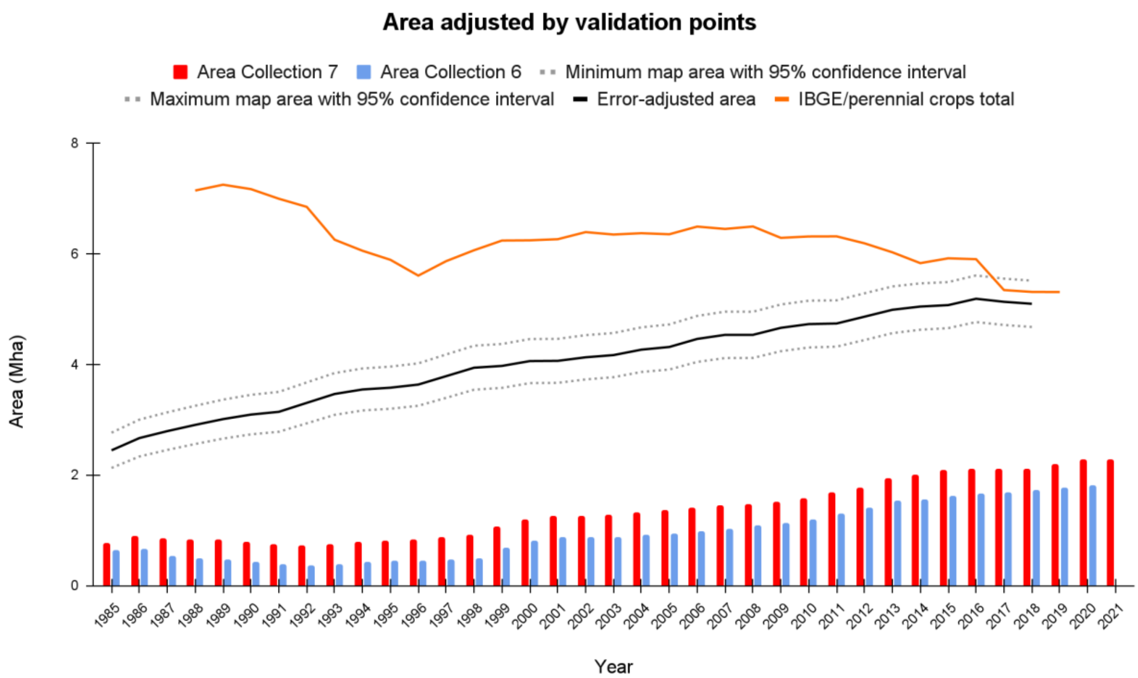


Figure 31. Area comparison between MapBiomas and IBGE for 'Perennial Crop' class.

5.2.3 Comparison of Forest Plantation area

'Forest Plantation' areas obtained from MapBiomas Collection 7 annual maps were also compared with areas from official sources. In order to estimate Brazil's forestry area, a comparison was made between MapBiomas Collection 6, PEVS-IBGE and Ibá. The results are presented by Figure 32.

The results show an increasing 'Forest Plantation' over the years for all source data (MapBiomas, IBGE and Ibá). The estimated forestry area, according to the validation points, is approximately 11.33 million hectares, with a confidence interval of approximately 1 million hectares, plus or minus. The area estimated by Ibá in 2021 is approximately 9.88 million hectares in 2021. The MapBiomas Collection 7 map reached a value of 8.6 million hectares. It is important to note that this area can still vary, due to allocation and quantity disagreements.

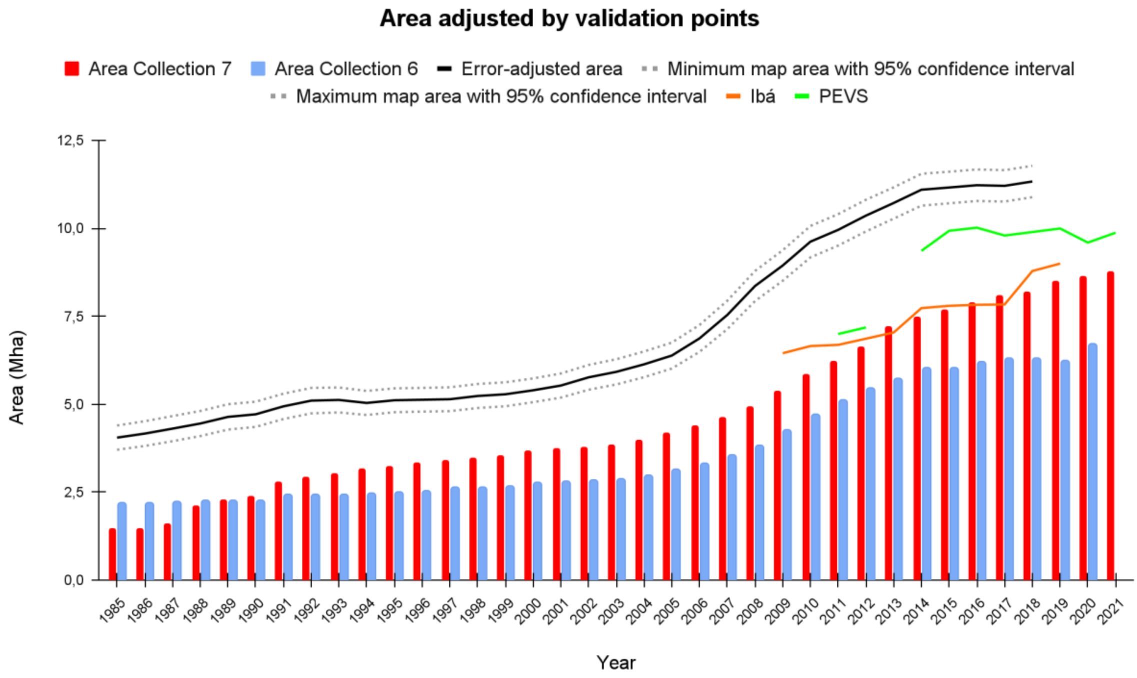


Figure 32. Comparison between MapBiomas and Ibá - *Indústria brasileira de árvores* (Brazilian Tree Industry).

6 References

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.