# Agriculture and Forest Plantation - Appendix

## Collection 5

## Version 1.1
## (12.2020)


**General coordinator**

Bernardo Rudorff

**Team**

Djonathan Goulart
Kênia Santos
Luciana Oliveira
Marciano Saraiva
Moisés Salgado

# 1 Overview of the classification method

Mapping 'Agriculture' and 'Forest Plantation' emerged as one of the challenges of 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 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 EVI2 and CEI vegetation indexes 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 and 5. Specifically for MapBiomas Collection 5, 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.

Generally, the process to classify the cross-cut themes of Agriculture and Forest Plantation is the following: initially, an annual Landsat mosaic is created from the collection of Landsat images (normalized or TOA) and according to the period of the year (season and off-season). After selecting the periods (specific for each class, as shown below), bands are built with specific metrics for each class. From the annual mosaic created, simple random sampling is performed based on the reference map. These samples are used to train the classifier and classify the classes of interest. 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. The classification step by step is illustrated in Figure 1.

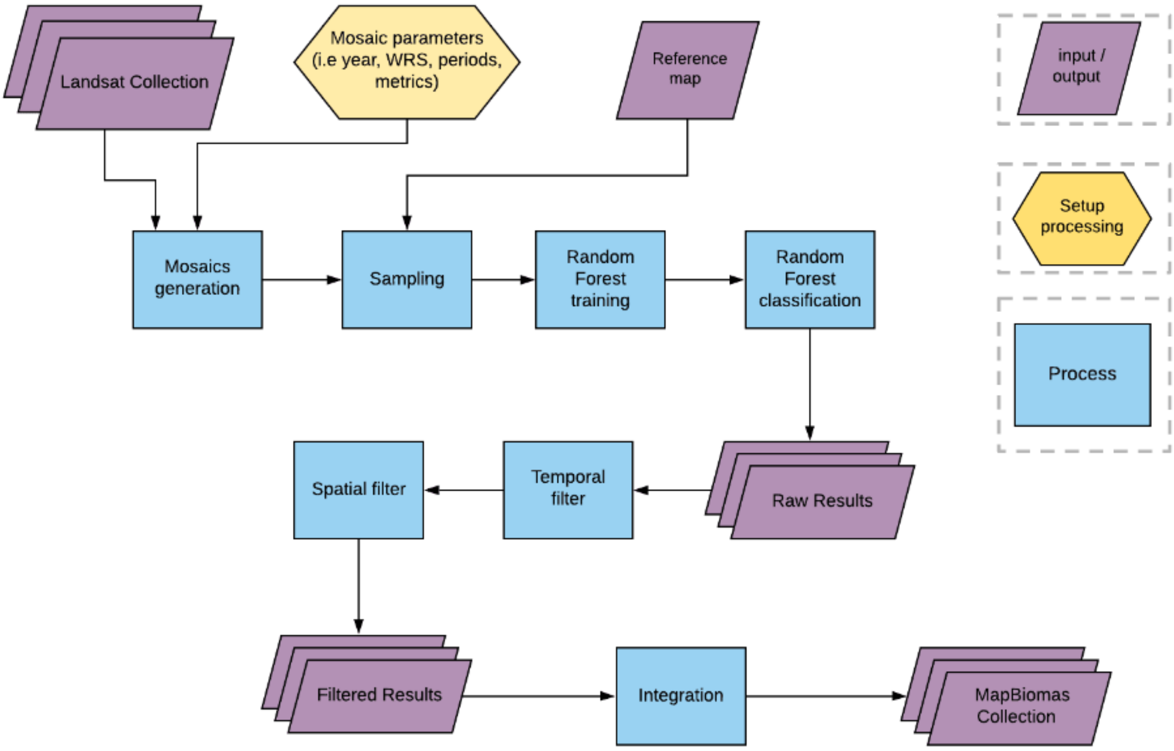



**Figure1.** Fluxogram of agriculture and forest plantation classification.

The production of MapBiomas Collection 5 for the cross-cutting themes 'Agriculture' and 'Forest Plantation' in the Brazilian territory from 1985 to 2019 followed a sequence of steps similar to those used in Collection 4. However, some improvements were added, especially the addition of new classes in the Collection 5, such as soybean class (from 2000 to 2019) and sugar cane class, as well as the distinction between temporary and perennial crops (Figure 2). The perennial crop class is, in MapBiomas Collection 5, a beta version due to the first effort to map this kind of crop, which is a challenge due to this confusion with other types of vegetation, such as native forest and forest plantation.

Due to these improvements in Collection 5, the 'Agriculture' mapping was restructured to include these new classes, as shown in Figure 2. The reformulation of the 'Agriculture' classification was done mainly to differentiate extensive areas of specific cultivation cycles and estimate emissions and removals of greenhouse gases associated with land use changes. This classification is aligned with the land use and land cover classification by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística* - IBGE).

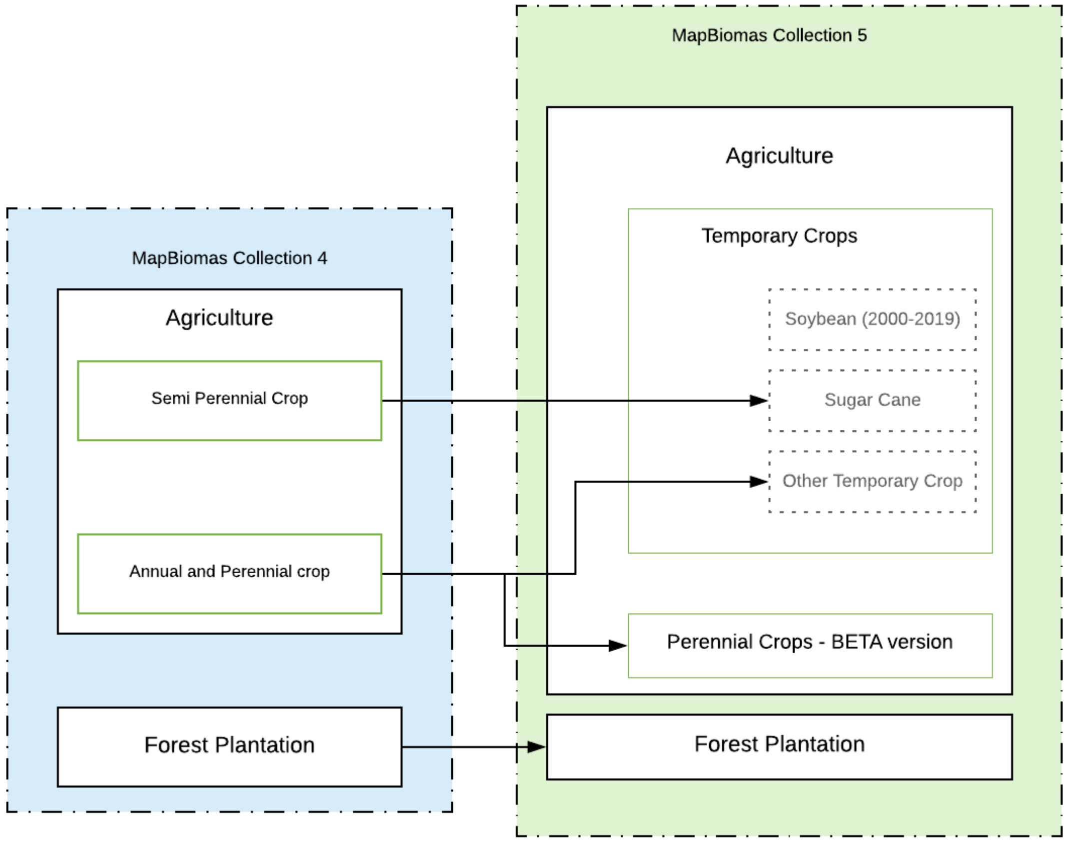



**Figure 2**. Agriculture and Forest Plantation classes mapped in the MapBiomas Collections 4 and 5.

Initially, the 'Soybean' class was mapped only for the period from 2000 to 2019, and the classes 'Other Temporary Crop' and 'Perennial Crop Beta version' emerged from the previous class 'Annual and Perennial Crop' in MapBiomas Collection 4. It is important to highlight that the class 'Annual and Perennial Crop' was mainly composed of annual crops, therefore, in this first moment, an effort was made to separate annual cycle crops from perennial cycle crops on the already existing maps in the previous collections.

## 2 Landsat image mosaics

To classify the cross-cutting theme of 'Agriculture' and 'Forest Plantation' in Collection 5, in addition to the Landsat images used in previous collections (available on the Google Earth Engine platform), a Landsat normalized time series was created based on the reflectance data from Moderate-Resolution Imaging Spectroradiometer (MODIS). The normalization of reflectance is an important step to guarantee the spectral similarity of the same types of land cover. It is the first time that normalized Landsat series are used, which was used to classify only three classes: 'Soybean', 'Other Temporary Crop', and 'Perennial Crop'. The maps of the cross-cutting theme 'Forest Plantation' and 'Sugar cane' class were generated with the TOA Landsat time series used in previous Landsat collections.

### 2.1 TOA Landsat Time Serie

For the classification of sugar cane and forest plantation the mosaics were created from Landsat top-of-atmosphere collections images available on the Google Earth Engine platform:

- USGS Landsat 5 TM Collection 1 Tier 1 (TOA) Reflectance;
- USGS Landsat 7 Collection 1 Tier 1 TOA Reflectance;
- USGS Landsat 8 Collection 1 Tier 1 TOA Reflectance.

### 2.2 Normalized Landsat Time Series

The maps of soybean, other temporary crops, and perennial crops were generated from normalized Landsat time series. The first step was to combine the TOA image collections of the different Landsat sensors (*i.e.* TM, ETM+, OLI, and TIRS) to harmonize their differences. With the combined collections, a reflectance normalization algorithm based on MODIS data was applied, similar to the methodology described by Potapov *et. al.* (2020). This step was necessary to ensure the spectral similarity of the same types of land cover, allowing the extrapolation in time and space of the image classification models. The normalized Landsat time series corrects some factors that affect the measurement of the surface reflectance, such as the atmospheric dispersion and attenuation of the electromagnetic radiation, the scattering of the radiation caused by surface anisotropy, and the sensor degradation over time.

The normalization process starts with the generation of a normalization target based on the MODIS MOD09A1 product. All products from 2000/01/01 to 2019/12/31 were used. First, the 75th percentile of the Normalized Difference Vegetation Index (NDVI) per pixel was calculated. Then, all pixels with NDVI below this percentile were masked and a mosaic was generated with the average of the other bands (*i.e*. blue, green, red, NIR, SWIR1, and SWIR2). Finally, the result was resampled to 30 meters (as Landsat spatial resolution) and saved as an asset (Figure 3).

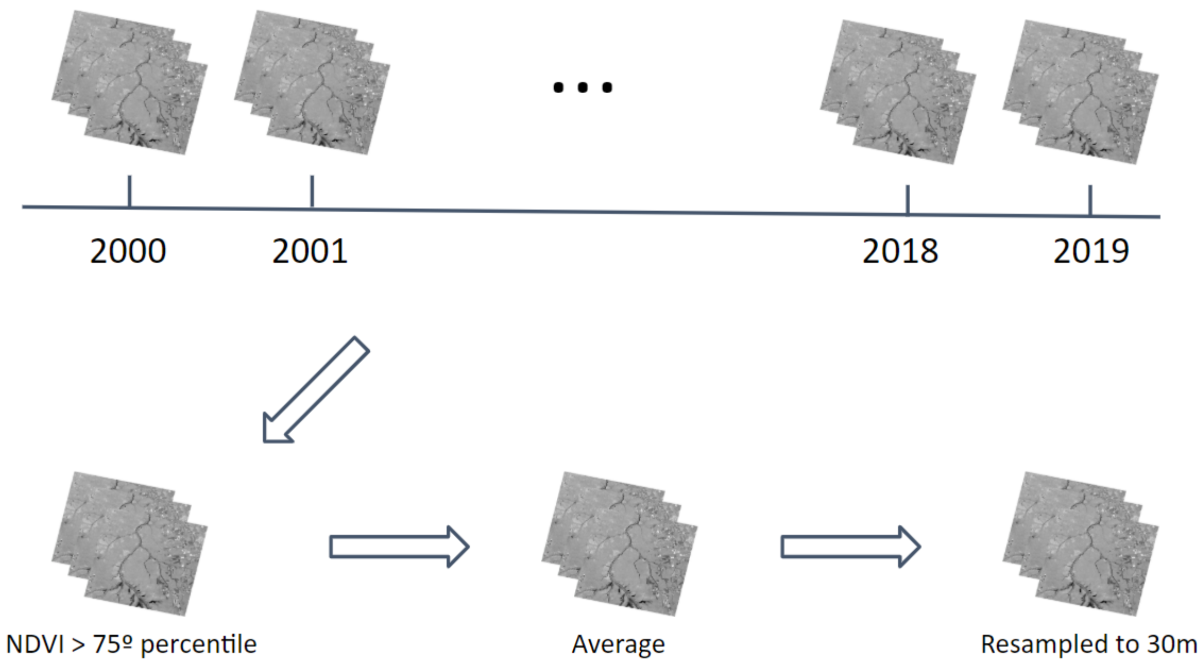



**Figure 3**. Procedure to generate the normalization target: with the MODIS product MOD09A1, all pixels with the NDVI below the 75th percentile were masked and a single mosaic was generated with the average reflectance of each spectral band. Then, the result was resampled to the spatial resolution of 30 meters.

Next, each Landsat image was compared with the normalization target and a mask was applied on the pixels with the red band above 0.5 (called 'bright objects' by Potapov *et al.*, 2020), and on the pixels where the absolute difference of the red and shortwave infrared 1 (SWIR1) bands with the normalization target greater than 0.05. The remaining pixels, called pseudo-invariant pixels, were counted, and images with less than 5,000 of these pixels were excluded from the normalization process. The average and standard deviation of these pseudo-invariant pixels were calculated per image, and these values were utilized for the normalization process in all pixels (not only in the pseudo-invariants) with a technique of histogram matching.

Figure 4 demonstrates two Landsat scenes before (A) and after (B) the reflectance normalization. After the normalization, the NDVI distribution from agriculture samples collected in both Landsat scenes is closer to the values collected in MODIS (product MOD09A1) (Figure 4). This similarity in the reflectance enables to obtain a consistent Landsat temporal series, with spectral harmonization of the same land cover through space and time.

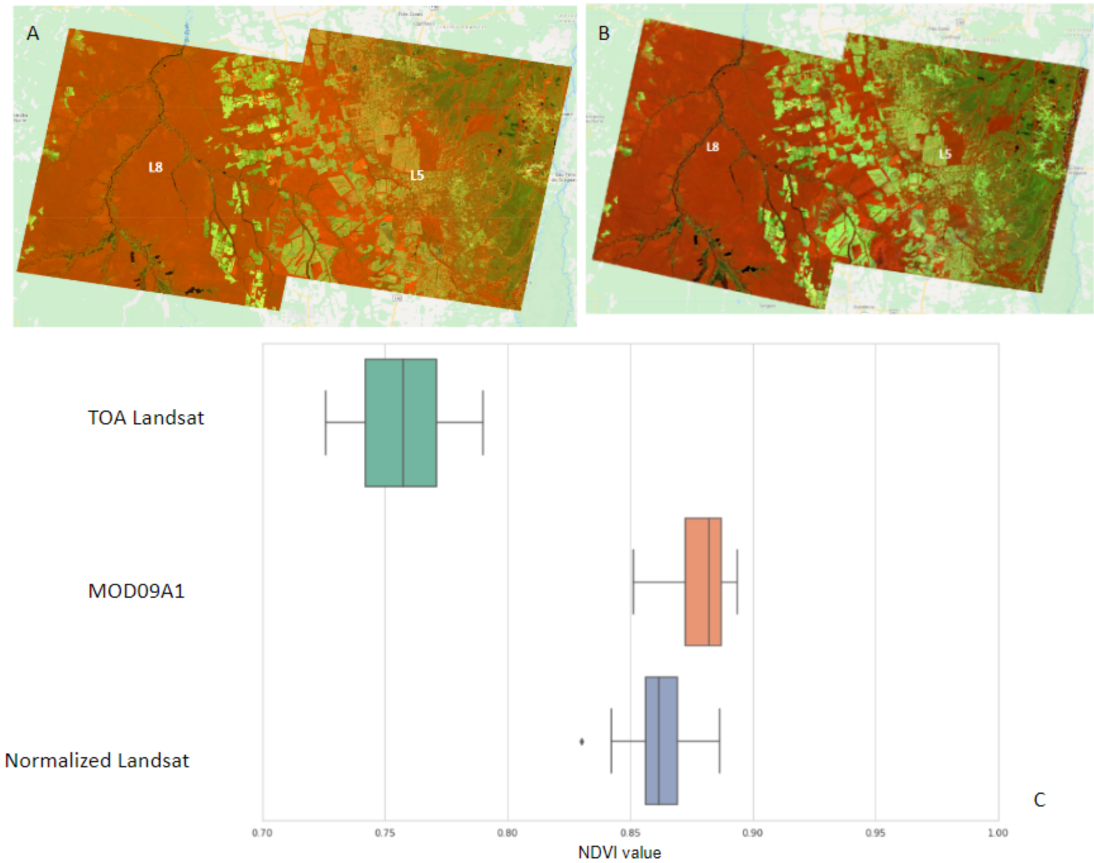



**Figure 4.** Reflectance normalization based on MODIS data applied to Landsat scenes. A) Scenes from Landsat 8 and 5 TOA collections before reflectance normalization. B) The same scenes after the normalization. C) NDVI boxplot from agriculture samples collected in both scenes before and after the normalization, and from MODIS. Spacecraft/WRS/year: L8/225068/2019 and L5/224068/2005.

Another step was the temporal aggregation of Landsat images into 16-days compositions (Figure 5). In the periods with two operating Landsat satellites (therefore one image in each 8 days), these mosaics were generated with up to two images. This interval corresponds to the Landsat orbital cycle and is compatible with the 16-days MODIS products (e.g. MOD13Q1). The last composition of each year covers an interval of 13 days (14 days for leap years).

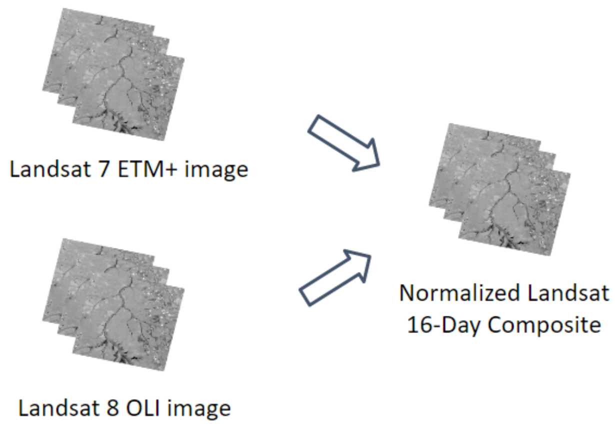



**Figure 5.** 16-days compositions of normalized Landsat time series.

## 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 theme were taken into account to better distinguish the class of interest from the remaining land cover and land use classes.

### 3.1 Agriculture

Similar to the method used in the classification of Agriculture in previous MapBiomas collections, the acquisition of Landsat images (TOA and normalized) to compose the mosaics for classification of Agriculture was carried out according to the crop season calendar in six regions in Brazil (Figure 6). The Landsat mosaics used to classify agriculture were built to highlight the seasonal change observed between the WET and DRY seasons.

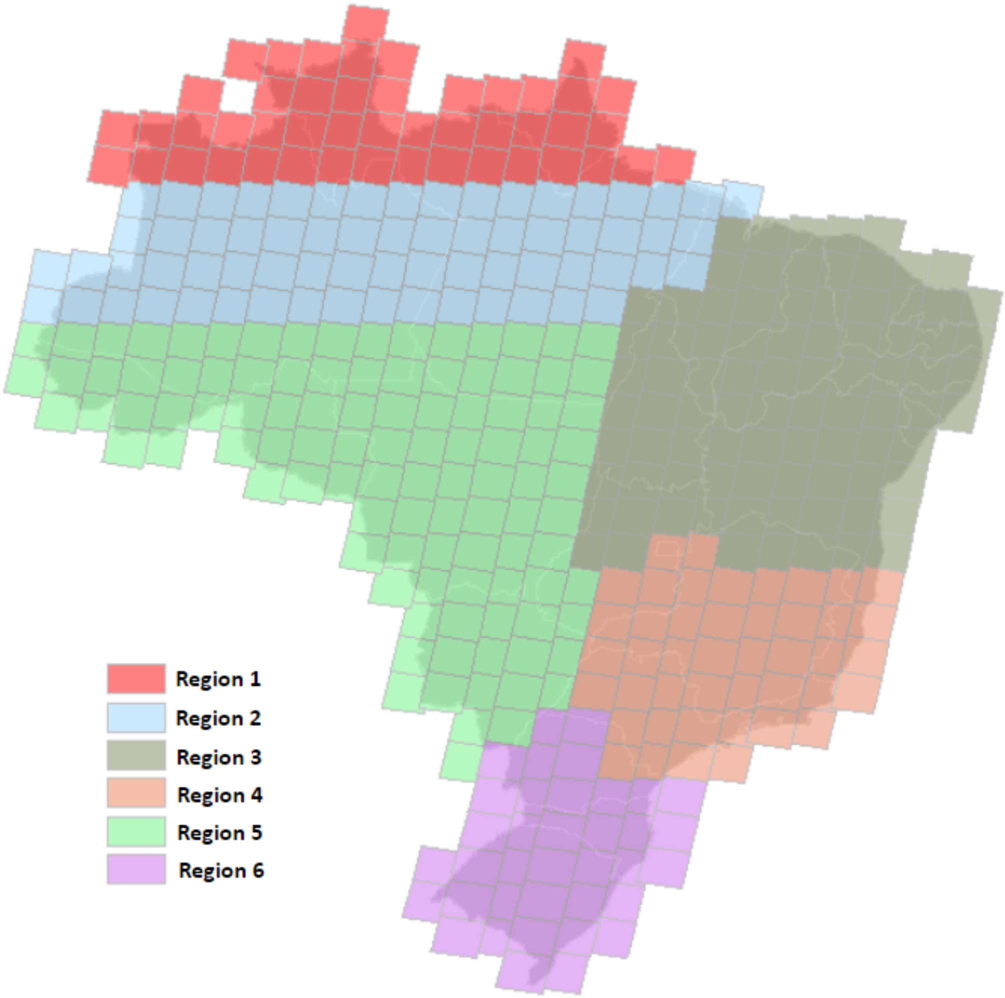



**Figure 6.** Regional crop calendar differences in Brazil considered to build the Landsat mosaics in the classification of the seasonal crops in the MapBiomas Collection 5.

The off-season also has important characteristics that help to distinguish crops. Season and off-season crop periods used to filter Landsat image collections to classify agriculture in the Collection 5 are shown in Table 1.

**Table 1.** Periods for the temporal composition of the Landsat mosaics used for the regional crop classification in MapBiomas Collection 5.

| Region | WET start | WET end | DRY start | DRY end |
|---|---|---|---|---|
| **Region 1** | 04/01/Year | 08/31/Year | 11/01/Year-1 | 03/31/Year |
| **Region 2** | 02/01/Year | 06/30/Year | 09/01/Year-1 | 02/15/Year |
| **Region 3** | 11/15/Year-1 | 06/15/Year | 10/15/Year-1 | 12/15/Year-1 |
| **Region 4** | 10/15/Year-1 | 06/15/Year | 05/15/Year-1 | 11/15/Year-1 |
| **Region 5** | 11/01/Year-1 | 06/30/Year | 05/01/Year-1 | 10/31/Year-1 |
| **Region 6** | 10/01/Year-1 | 11/15/Year | 01/01/Year-1 | 04/30/Year-1 |

Table 1 specifies the periods per region for the Landsat images selection in each year to classify seasonal crops. Also, images from previous years were used to increase the likelihood of cloud-free images acquisition during the seasonal crop to better map the crop fields and, therefore, improve the quality of the classification and final maps.

### 3.1.1 Temporary Crops

#### 3.1.1.1 Soybean

The life cycle of plants is an important factor for monitoring and mapping agricultural crops via Remote Sensing, mainly the short-cycle crops (e.g. two to six months) (FORMAGGIO and SANCHES, 2017), like soybeans crops. The main challenge of this class was the definition of metrics to distinguish this crop from others short-cycle crops. As there are specific months in which soy is usually cultivated, temporal metrics were used to differentiate soy from other short-cycle crops. These metrics were obtained at regular intervals and adapted to the agricultural calendar of each region (Table 2).

**Table 2.** Period of temporal metrics according to the agricultural calendar of each region.

| Period | Region 1 | Region 2 | Region 3 | Region 4 | Region 5 | Region 6 |
|---|---|---|---|---|---|---|
| **Month 1** | 04/01/Year to 05/01/Year | 02/01/Yea to 03/01/Year | 11/15/Year-1 to 12/15/Year-1 | 10/15/Year-1 to 11/15/Year-1 | 11/01/Year-1 to 12/01/Year-1 | 10/15/Year-1 to 11/15/Year-1 |
| **Month 2** | 05/01/Year to 06/01/Year | 03/01/Year to 04/01/Year | 12/15/Year-1 to 01/15/Year | 11/15/Year-1 to 12/15/Year-1 | 12/01/Year-1 to 01/01/Year | 11/15/Year-1 to 12/15/Year-1 |
| **Month 3** | 06/01/Year to 07/01/Year | 04/01/Year to 05/01/Year | 01/15/Year to 02/15/Year | 12/15/Year-1 to 01/15/Year | 01/01/Year to 02/01/Year | 12/15/Year-1 to 01/15/Year |
| **Month 4** | 07/01/Year to 08/01/Year | 05/01/Year to 06/01/Year | 02/15/Year to 03/15/Year | 01/15/Year to 02/15/Year | 02/01/Year to 03/01/Year | 01/15/Year to 02/15/Year |
| **Month 5** | 08/01/Year to 09/01/Year | 06/01/Year to 07/01/Year | 03/15/Year to 04/15/Year | 02/15/Year to 03/15/Year | 03/01/Year to 04/01/Year | 02/15/Year to 03/15/Year |
| **First three observations** | 04/01/Year to 05/19/Year | 02/01/Year to 03/21/Year | 11/15/Year-1 to 01/02/Year | 10/15/Year-1 to 12/02/Year-1 | 11/01/Year-1 to 12/19/Year-1 | 10/15/Year-1 to 12/02/Year-1 |
| **Last three observations** | 07/15/Year to 09/01/Year | 05/14/Year to 07/01/Year | 02/26/Year to 04/15/Year | 01/26/Year to 03/15/Year | 02/12/Year to 04/01/Year | 01/26/Year to 03/15/Year |

The normalized Landsat images of the periods described in Table 2 were used to generate image mosaics for each year. Additional images from the same periods from two previous years were used to provide better results.

The classification of soybean crops was carried out throughout the Brazilian territory from 2000 to 2019. Before 2000, soybean crops were included as Other Temporary Crop legend.

#### 3.1.1.2 Sugar cane

The classes 'Sugar cane' and 'Forest Plantation' used Landsat mosaics created to highlight intra-annual variations based on bimonthly compositions for 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 and forest plantation in Collection 5.

| Period | Start | End |
|---|---|---|
| **WET1** | 12/01/year-1 | 01/31/year |
| **WET2** | 02/01/year | 03/31/year |
| **DRY1** | 04/01/year | 05/31/year |
| **DRY2** | 06/01/year | 07/31/year |
| **DRY3** | 08/01/year | 09/30/year |
| **WET3** | 10/01/year | 11/30/year |

In Collection 3, 4 and 5, we used bimonthly compositions and biannual mosaics to classify semi perennial crops and forest plantation. The Landsat images of the periods described in Table 3 were used to generate image mosaics for each year in addition to images from the same periods from two previous years. Using three years provided better results when compared to the biannual mosaics. Thus, in Collection 5, we used three years of bimonthly image mosaics to classify sugar cane.

#### 3.1.1.3 Other Temporary Crop

The period defined for the classification of the 'Other Temporary Crop' class was the same period of the seasonal crop shown in Table 1.

### 3.1.2 Perennial Crop - *beta version*

The period defined for the classification of the 'Perennial Crop - *beta version*' class was the same period of the seasonal crop shown in Table 1.

### 3.2 Forest Plantation

The period defined for the classification of the theme 'Forest Plantation' was the same used for the classification of sugar cane shown in Table 3.

## 4 Image selection

### 4.1 TOA Landsat Time Series

For the classification of sugar cane and forest plantation classes, all available images were used to compose the Landsat mosaics. A cloud filter and a cloud shadow filter were applied to the Landsat images before the composition of the mosaic. These filters were developed based on the quality assessment band available in the Google Earth Engine image collections.

### 4.2 Normalized Landsat Time Series

As mentioned earlier, for the normalized Landsat time series generation, all images in the product MOD09A1 between 2000 and 2019 were used, as well as the Landsat time series images from the same period. For the 16-day normalized products, Landsat images with more than 90% of cloud/shadow coverage were not considered, and a pseudo-invariant pixel filter was applied to exclude Landsat images with less than 5,000 pseudo-invariant pixels.

### 4.3 Final quality

The Landsat images availability in Collection 5 period (1985 to 2019) varies among years. Throughout this period, Landsat 5 (1985 to 2012), Landsat 7 (1999 to present), and Landsat 8 (2013 to present) provided the images for the mosaics compositions. Figure 7 shows the variability of available Landsat images for Collection 5 period. The annual number of normalized 16-day products varies according to the availability of images, limited to 23 products per year.

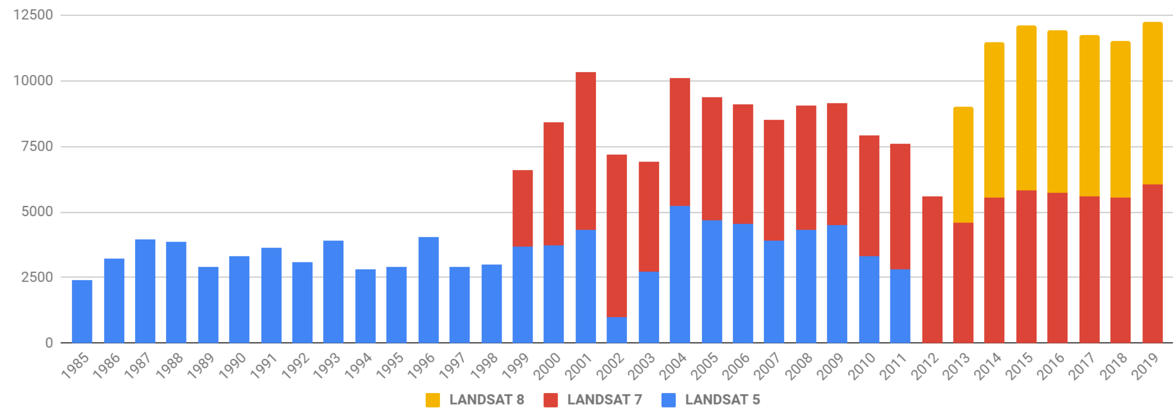
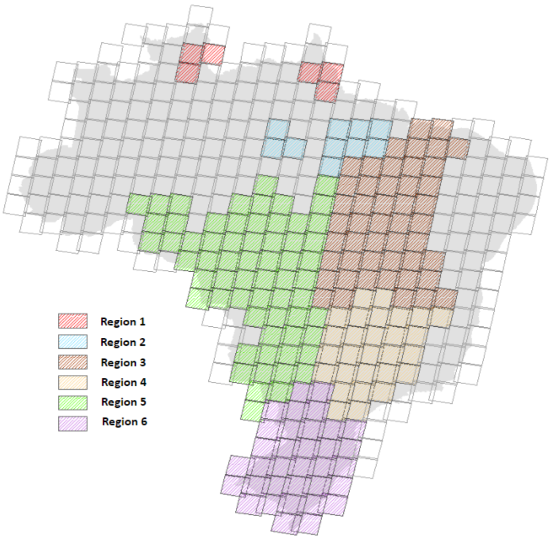



**Figure 7.** Number of available TOA Landsat images covering the Brazilian territory from 1985 to 2019.

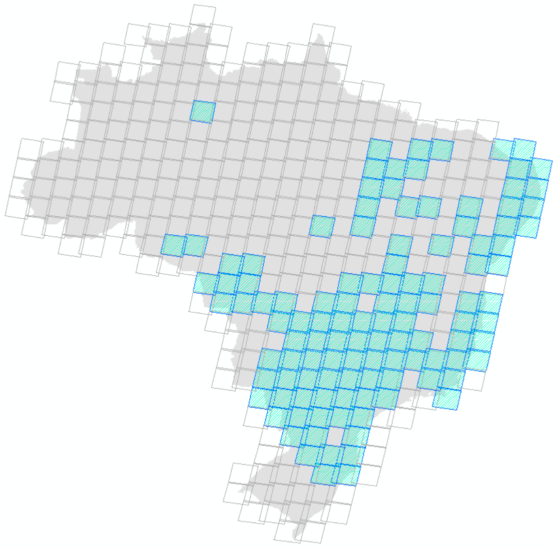
## 5 Definition of regions for classification

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.

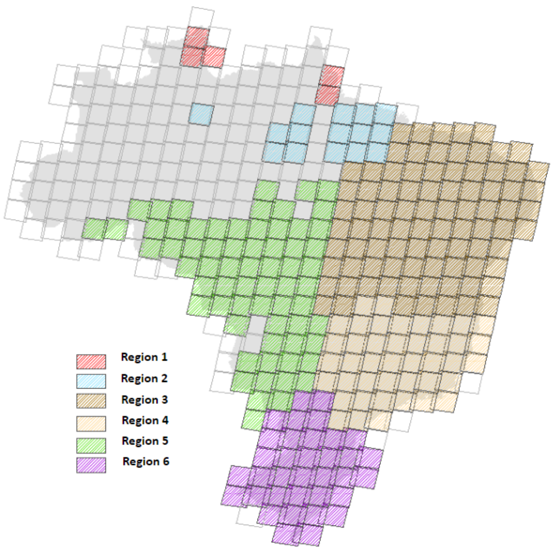
**Agriculture**



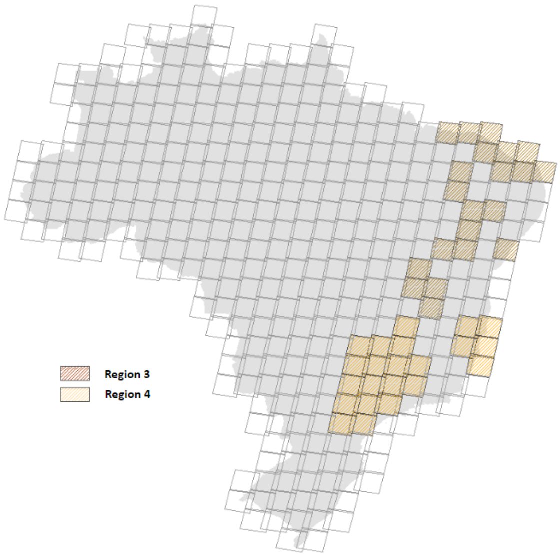

Soybean



Sugar cane




Other Temporary Crop




Perennial Crop

**Forest Plantation**

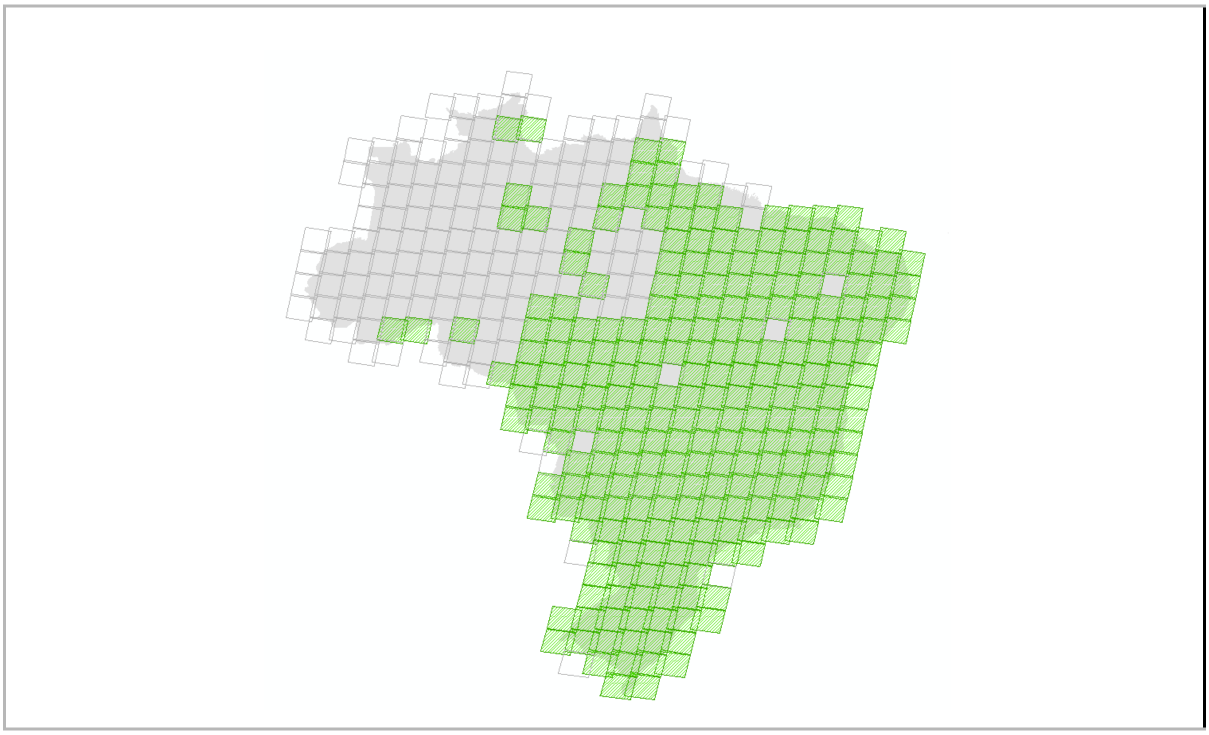


**Figure 8.** Selected scenes of Landsat series to the classification of maps by land use class.

## 6 Classification

The automatic classification procedure was fully performed on Google Earth Engine platform, using the Random Forest classifier (BREIMAN, 2001). The following sections present the steps used to classify the 'Agriculture' and 'Forest Plantation' classes.

### 6.1 Classification scheme

Each land-use class of interest was mapped separately. Therefore, five independent classifications were performed: 1) soybean; 2) sugar cane; 3) other temporary crops; 4) perennial crop; 5) forest plantation. The classification process considered only two possible classes for each pixel, the class of interest (soybean, sugar cane, other temporary crops, perennial crop or forest plantation) and the background (all other land uses).

For the supervised classification of Landsat mosaics, training points were selected based on reference maps. The Random Forest parameter, reference maps, and the feature space produced for each mosaic are presented in the following sections.

#### 6.1.1 Training sample

The acquisition of training samples of the classifier for all classes of Collection 5 was performed based on the Landsat scenes which contained data of the class of interest. 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 9.

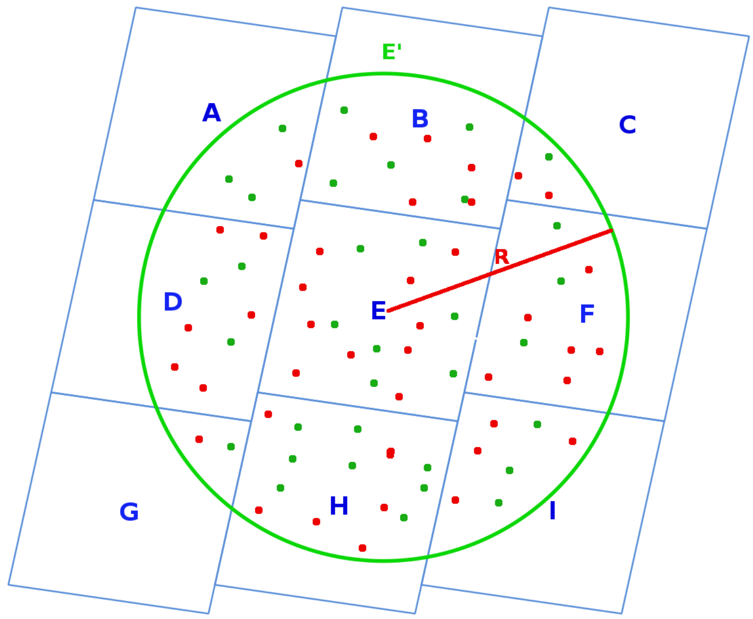



**Figure 9.** Scheme for sample acquisition for the regionalized training of the Random Forest classifier in Agriculture and Forest Plantation.

### 6.2 Feature space

The feature space for 'Agriculture' in the Collection 5 was the same of Collection 4, which was composed of Landsat bands, and reducers (minimum, maximum, median, standard deviation and quality mosaic) calculated for each spectral indices presented in Table 4, and each band, resulting in 178 variables.

**Table 4.** Relation of spectral indices used to classify Agriculture and Forest Plantation classes in Collections 4 and 5.

| Index | Expression | Reference |
|---|---|---|
| **EVI2** | 2.5 * ((NIR - RED) / (NIR + 2.4*RED + 1)) | Jiang et al, 2008 |
| **NDVI** | (NIR - RED) / (NIR + RED) | Deering, 1978 |
| **NDWI** | (NIR - SWIR1) / (NIR + SWIR1) | Gao, 1996 |
| **CAI** | SWIR2 / SWIR1 | Nagler et al, 2003 |
| **LAI** | 0.3977 * exp(2.5556 * (NIR - RED)/(NIR + RED)) | Chen et al, 2012 |
| **CEI** | (WET_EVI2_max - DRY_EVI2_min) / (WET_EVI2_max + DRY_EVI2_min) | Rizzi et al., 2009 |

In Collection 4, for the 'Annual and Perennial Crop' class, following the acquisition of the feature space with 178 variables, the method proposed by Kursa *et al*. (2010) was applied to select all relevant data and reduce the number of variables used in the classification model for crop types (annual and perennial or semi-perennial crops) and forest plantation. The metrics with greater relevance for mapping annual and perennial crops are shown in Table 5.

**Table 5.** Metrics used to classify 'Annual and Perennial Crop' in Collection 4.

| Id | Variable | Description | Statistics | Period |
|---|---|---|---|---|
| 1 | WET_GREEN_min | Landsat Green band minimum value | minimum | WET |
| 2 | WET_GREEN_median | Landsat Green band median value | median | WET |
| 3 | WET_RED_max | Landsat Red band maximum value | maximum | WET |
| 4 | WET_RED_median | Landsat Red band median value | median | WET |
| 5 | WET_RED_stdDev | Landsat Red band standard deviation value | standard deviation | WET |
| 6 | WET_NIR_qmo | Landsat NIR band selected based on maximum EVI2 | maximum | WET |
| 7 | WET_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | WET |
| 8 | WET_SWIR1_max | Landsat SWIR1 band maximum value | maximum | WET |
| 9 | WET_SWIR1_min | Landsat SWIR1 band minimum value | minimum | WET |
| 10 | WET_SWIR1_median | Landsat SWIR1 band median value | median | WET |
| 11 | WET_SWIR1_stdDev | Landsat SWIR1 band standard deviation value | standard deviation | WET |
| 12 | WET_SWIR2_max | Landsat SWIR2 band maximum value | maximum | WET |
| 13 | WET_SWIR2_median | Landsat SWIR2 band median value | median | WET |
| 14 | WET_SWIR2_stdDev | Landsat SWIR2 band standard deviation value | standard deviation | WET |
| 15 | WET_TIR1_max | Landsat TIR1 band maximum value | maximum | WET |
| 16 | WET_TIR1_stdDev | Landsat TIR1 band standard deviation value | standard deviation | WET |
| 17 | WET_EVI2_max | Spectral index EVI2 maximum value | maximum | WET |
| 18 | WET_EVI2_min | Spectral index EVI2 minimum value | minimum | WET |
| 19 | WET_EVI2_stdDev | Spectral index EVI2 standard deviation value | standard deviation | WET |
| 20 | WET_NDWI_max | Spectral index NDWI maximum value | maximum | WET |
| 21 | WET_NDWI_min | Spectral index NDWI minimum value | minimum | WET |
| 22 | WET_NDWI_median | Spectral index NDWI median value | median | WET |
| 23 | WET_NDWI_stdDev | Spectral index NDWI standard deviation value | standard deviation | WET |
| 24 | WET_CAI_max | Spectral index CAI maximum value | maximum | WET |
| 25 | WET_CAI_min | Spectral index CAI minimum value | minimum | WET |
| 26 | WET_CAI_median | Spectral index CAI median value | median | WET |
| 27 | WET_CAI_stdDev | Spectral index CAI standard deviation value | standard deviation | WET |
| 28 | DRY_GREEN_median | Landsat Green band median value | median | DRY |

| | | | | |
|---|---|---|---|---|
| 29 | DRY_RED_max | Landsat Red band maximum value | maximum | DRY |
| 30 | DRY_RED_min | Landsat Red band minimum value | minimum | DRY |
| 31 | DRY_RED_median | Landsat Red band median value | median | DRY |
| 32 | DRY_SWIR1_max | Landsat SWIR1 band maximum value | maximum | DRY |
| 33 | DRY_SWIR1_median | Landsat SWIR1 band median value | median | DRY |
| 34 | DRY_SWIR1_stdDev | Landsat SWIR1 band standard deviation value | standard deviation | DRY |
| 35 | DRY_SWIR2_max | Landsat SWIR2 band maximum value | maximum | DRY |
| 36 | DRY_SWIR2_min | Landsat SWIR2 band minimum value | minimum | DRY |
| 37 | DRY_SWIR2_median | Landsat SWIR2 band median value | median | DRY |
| 38 | DRY_EVI2_min | Spectral index EVI2 minimum value | minimum | DRY |
| 39 | DRY_EVI2_median | Spectral index EVI2 median value | median | DRY |
| 40 | DRY_NDWI_qmo | Spectral index NDWI selected based on maximum EVI2 | maximum | DRY |
| 41 | DRY_NDWI_max | Spectral index NDWI maximum value | maximum | DRY |
| 42 | DRY_NDWI_min | Spectral index NDWI minimum value | minimum | DRY |
| 43 | DRY_NDWI_median | Spectral index NDWI median value | median | DRY |
| 44 | DRY_NDWI_stdDev | Spectral index NDWI standard deviation value | standard deviation | DRY |
| 45 | DRY_CAI_qmo | Spectral index CAI selected based on maximum EVI2 | maximum | DRY |
| 46 | DRY_CAI_max | Spectral index CAI maximum value | maximum | DRY |
| 47 | DRY_CAI_min | Spectral index CAI minimum value | minimum | DRY |
| 48 | ANNUAL_NIR_cei | Spectral index CEI using NIR band instead EVI2 | normalized difference | ANNUAL |
| 49 | ANNUAL_EVI2_cei | Spectral index CEI value | normalized difference | ANNUAL |
| 50 | ANNUAL_NDWI_cei | Spectral index CEI using NDWI band instead EVI2 | normalized difference | ANNUAL |

### 6.2.1 Soybean

In addition to the variables shown in table 5, temporal metrics for soybean classification were added. These time metrics were based on the work of King et al. (2017) and are shown in Table 6.

**Table 6.** Metrics added to the classification of soybean crops in MapBiomas Collection 5.

| Period | Bands and Spectral Indexes | Statistical Metrics |
|---|---|---|
| Month 1 | RED, NIR, SWIR1, SWIR2, NDVI, NDWI, EVI2 | Average |
| Month 2 | | |
| Month 3 | | |
| Month 4 | | |
| Month 5 | | |
| First three observations | | Median and Average |

| | | |
|---|---|---|
| Last three observations | | |

These metrics aimed to obtain spectral characteristics in the phenological period of the crops, which is a relevant characteristic for the distinction between them.

A total of 104 layers were used to compose the mosaics for the classification of soybean crops (50 metrics selected for the classification of annual and perennial crops, which were used for mapping temporary crops in Collection 5, and 54 metrics added for the classification of soybean).

### 6.2.2 Sugar cane

The 'Sugar cane' class corresponds to the 'Semi-Perennial Crop' class in the previous MapBiomas collections. Therefore, the same variables were adopted for the classification of this crop in Collection 5 (Table 7).

**Table 7.** Metrics used to classify sugar cane in MapBiomas Collection 5.

| Id | Variable | Description | Statistics | Period |
|---|---|---|---|---|
| 1 | WET1_GREEN_median | Landsat Green band median value | median | WET1 |
| 2 | WET1_RED_median | Landsat Red band median value | median | WET1 |
| 3 | WET1_NIR_median | Landsat NIR band median value | median | WET1 |
| 4 | WET1_SWIR1_median | Landsat SWIR1 band median value | median | WET1 |
| 5 | WET1_SWIR2_median | Landsat SWIR2 band median value | median | WET1 |
| 6 | WET1_NDVI_median | Spectral index NDVI median value | median | WET1 |
| 7 | WET1_NDWI_median | Spectral index NDWI median value | median | WET1 |
| 8 | WET2_GREEN_median | Landsat Green band median value | median | WET2 |
| 9 | WET2_RED_median | Landsat Red band median value | median | WET2 |
| 10 | WET2_NIR_median | Landsat NIR band median value | median | WET2 |
| 11 | WET2_SWIR1_median | Landsat SWIR1 band median value | median | WET2 |
| 12 | WET2_SWIR2_median | Landsat SWIR2 band median value | median | WET2 |
| 13 | WET2_NDVI_median | Spectral index NDVI median value | median | WET2 |
| 14 | WET2_NDWI_median | Spectral index NDWI median value | median | WET2 |
| 15 | DRY1_GREEN_median | Landsat Green band median value | median | DRY1 |
| 16 | DRY1_RED_median | Landsat Red band median value | median | DRY1 |
| 17 | DRY1_NIR_median | Landsat NIR band median value | median | DRY1 |
| 18 | DRY1_SWIR1_median | Landsat SWIR1 band median value | median | DRY1 |
| 19 | DRY1_SWIR2_median | Landsat SWIR2 band median value | median | DRY1 |
| 20 | DRY1_NDVI_median | Spectral index NDVI median value | median | DRY1 |

| | | | | |
|---|---|---|---|---|
| 21 | DRY1_NDWI_median | Spectral index NDWI median value | median | DRY1 |
| 22 | DRY2_GREEN_median | Landsat Green band median value | median | DRY2 |
| 23 | DRY2_RED_median | Landsat Red band median value | median | DRY2 |
| 24 | DRY2_NIR_median | Landsat NIR band median value | median | DRY2 |
| 25 | DRY2_SWIR1_median | Landsat SWIR1 band median value | median | DRY2 |
| 26 | DRY2_SWIR2_median | Landsat SWIR2 band median value | median | DRY2 |
| 27 | DRY2_NDVI_median | Spectral index NDVI median value | median | DRY2 |
| 28 | DRY2_NDWI_median | Spectral index NDWI median value | median | DRY2 |
| 29 | DRY3_GREEN_median | Landsat Green band median value | median | DRY3 |
| 30 | DRY3_RED_median | Landsat Red band median value | median | DRY3 |
| 31 | DRY3_NIR_median | Landsat NIR band median value | median | DRY3 |
| 32 | DRY3_SWIR1_median | Landsat SWIR1 band median value | median | DRY3 |
| 33 | DRY3_SWIR2_median | Landsat SWIR2 band median value | median | DRY3 |
| 34 | DRY3_NDVI_median | Spectral index NDVI median value | median | DRY3 |
| 35 | DRY3_NDWI_median | Spectral index NDWI median value | median | DRY3 |
| 36 | WET3_GREEN_median | Landsat Green band median value | median | WET3 |
| 37 | WET3_RED_median | Landsat Red band median value | median | WET3 |
| 38 | WET3_NIR_median | Landsat NIR band median value | median | WET3 |
| 39 | WET3_SWIR1_median | Landsat SWIR1 band median value | median | WET3 |
| 40 | WET3_SWIR2_median | Landsat SWIR2 band median value | median | WET3 |
| 41 | WET3_NDVI_median | Spectral index NDVI median value | median | WET3 |
| 42 | WET3_NDWI_median | Spectral index NDWI median value | median | WET3 |

### 6.2.3 Other Temporary Crop and Perennial Crop - *beta version*

The ‘Other Temporary Crop’ class came from the separation between annual crops and perennial crops from the class ‘Annual and Perennial Crop’ of MapBiomas Collection 4. Therefore, the other temporary crop maps were created from two classification processes: 1) the classifier was trained to separate the areas of ‘Annual and Perennial Crop, using the same method of Collection 4; 2) 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. Figure 10 illustrates the processes performed to generate the maps ‘Other Temporary Crop’ and ‘Perennial Crop’.

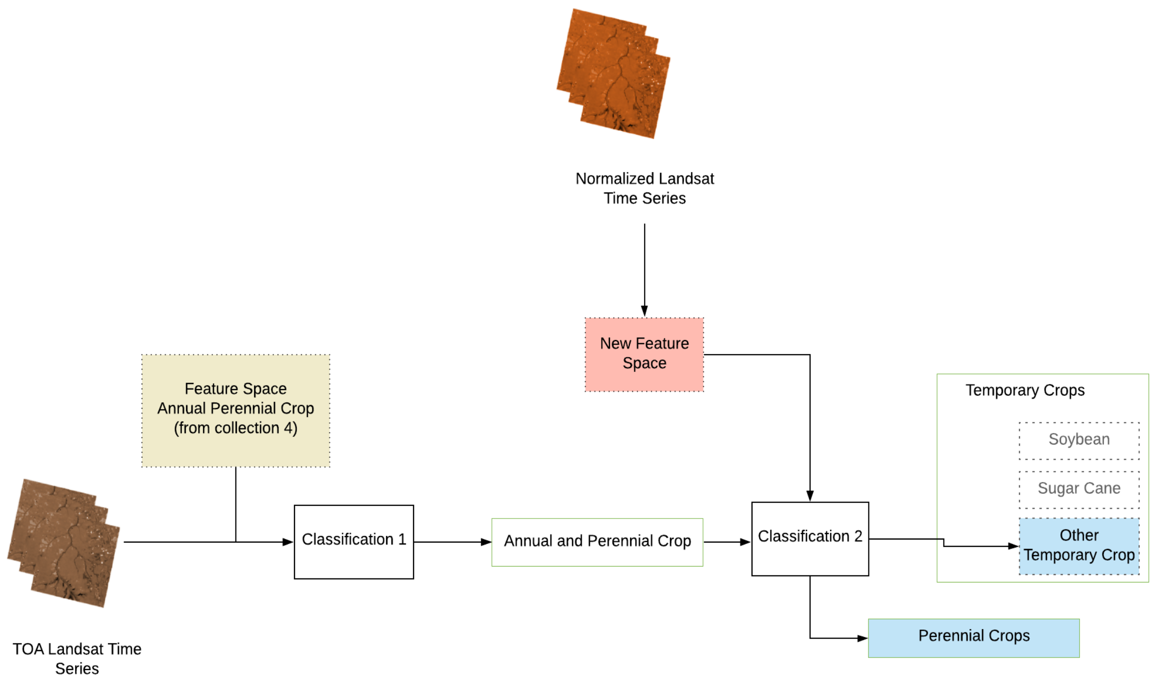



**Figure 10.** Steps to separate 'Perennial Crop - *beta version*' and 'Temporary Crop' in the MapBiomas Collection 5 from the previous class 'Annual and Perennial Crop' in the Collection 4.

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 8).

**Table 8.** Metrics used to separate perennial and other temporary crops in Collection 5.

| Id | Variable | Description | Statistics | Period |
|---|---|---|---|---|
| 1 | WET_NDWI_stdDev | Spectral index NDWI standard deviation value | standard deviation | WET |
| 2 | WET_NDWI_min | Spectral index NDWI minimum value | minimum | WET |
| 3 | WET_EVI2_stdDev | Spectral index EVI2 standard deviation value | standard deviation | WET |
| 4 | DRY_EVI2_min | Spectral index EVI2 minimum value | minimum | DRY |
| 5 | ANNUAL_EVI2_amplitude | Spectral index EVI2 amplitude value | amplitude | ANNUAL |
| 6 | ANNUAL_EVI2_stdDev | Spectral index EVI2 standard deviation value | standard deviation | ANNUAL |
| 7 | ANNUAL_EVI2_min | Spectral index EVI2 minimum value | minimum | ANNUAL |
| 8 | ANNUAL_EVI2_p10 | Spectral index EVI2 10th percentile value | 10th percentile | ANNUAL |
| 9 | ANNUAL_EVI2_median | Spectral index EVI2 median value | median | ANNUAL |
| 10 | ANNUAL_EVI2_mean | Spectral index EVI2 average value | average | ANNUAL |
| 11 | ANNUAL_NIR_cei | Landsat NIR band normalized difference value | normalized difference | ANNUAL |
| 12 | ANNUAL_EVI2_cei | Spectral index EVI2 normalized difference value | normalized difference | ANNUAL |
| 13 | ANNUAL_NDWI_cei | Spectral index NDWI normalized difference value | normalized difference | ANNUAL |

The result of this separation was a map with two classes: temporary crop and perennial crop. The pixels classified as temporary crops originated the class ‘Other Temporary Crop’. Knowing that there are soybean crops (and eventually sugar cane) in this class, a rule was created: the pixels of the new ‘Other Temporary Crop’ class which overlapped the pixels of the ‘Soybean’ or ‘Sugar cane’ classes lose priority in the internal integration and therefore are classified as ‘Soybean’ or ‘Sugar cane’.

### 6.2.4 Forest Plantation

The metrics used to classify forest plantation were the same as in Collection 4 (Table 9).

**Table 9.** Metrics used to classify forest plantation in Collection 5.

| Id | Variable | Description | Statistics | Period |
|---|---|---|---|---|
| 1 | WET1_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | WET1 |
| 2 | WET1_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | WET1 |
| 3 | WET1_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | WET1 |
| 4 | WET1_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | WET1 |
| 5 | WET1_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | WET1 |
| 6 | WET1_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | WET1 |
| 7 | WET1_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | WET1 |
| 8 | WET2_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | WET2 |
| 9 | WET2_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | WET2 |
| 10 | WET2_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | WET2 |
| 11 | WET2_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | WET2 |
| 12 | WET2_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | WET2 |
| 13 | WET2_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | WET2 |
| 14 | WET2_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | WET2 |
| 15 | DRY1_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | DRY1 |
| 16 | DRY1_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | DRY1 |
| 17 | DRY1_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | DRY1 |
| 18 | DRY1_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | DRY1 |
| 19 | DRY1_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | DRY1 |
| 20 | DRY1_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | DRY1 |
| 21 | DRY1_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | DRY1 |
| 22 | DRY2_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | DRY2 |
| 23 | DRY2_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | DRY2 |
| 24 | DRY2_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | DRY2 |

| | | | | |
|---|---|---|---|---|
| 25 | DRY2_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | DRY2 |
| 26 | DRY2_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | DRY2 |
| 27 | DRY2_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | DRY2 |
| 28 | DRY2_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | DRY2 |
| 29 | DRY3_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | DRY3 |
| 30 | DRY3_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | DRY3 |
| 31 | DRY3_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | DRY3 |
| 32 | DRY3_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | DRY3 |
| 33 | DRY3_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | DRY3 |
| 34 | DRY3_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | DRY3 |
| 35 | DRY3_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | DRY3 |
| 36 | WET3_GREEN_qmo | Landsat Green band selected based on maximum EVI2 | maximum | WET3 |
| 37 | WET3_RED_qmo | Landsat Red band selected based on maximum EVI2 | maximum | WET3 |
| 38 | WET3_NIR_qmo | Landsat Nir band selected based on maximum EVI2 | maximum | WET3 |
| 39 | WET3_SWIR1_qmo | Landsat SWIR1 band selected based on maximum EVI2 | maximum | WET3 |
| 40 | WET3_SWIR2_qmo | Landsat SWIR2 band selected based on maximum EVI2 | maximum | WET3 |
| 41 | WET3_NDVI_qmo | Spectral index NDVI selected based on maximum EVI2 | maximum | WET3 |
| 42 | WET3_LAI_qmo | Spectral index LAI selected based on maximum EVI2 | maximum | WET3 |

## 6.3 Classification algorithm, training samples and parameters

Knowing that there are no reference maps available for all classes in all years of the time series (1985 to 2019), 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. The fluxogram (Figure 11) illustrates the classification process with Normalized Landsat Time Series.

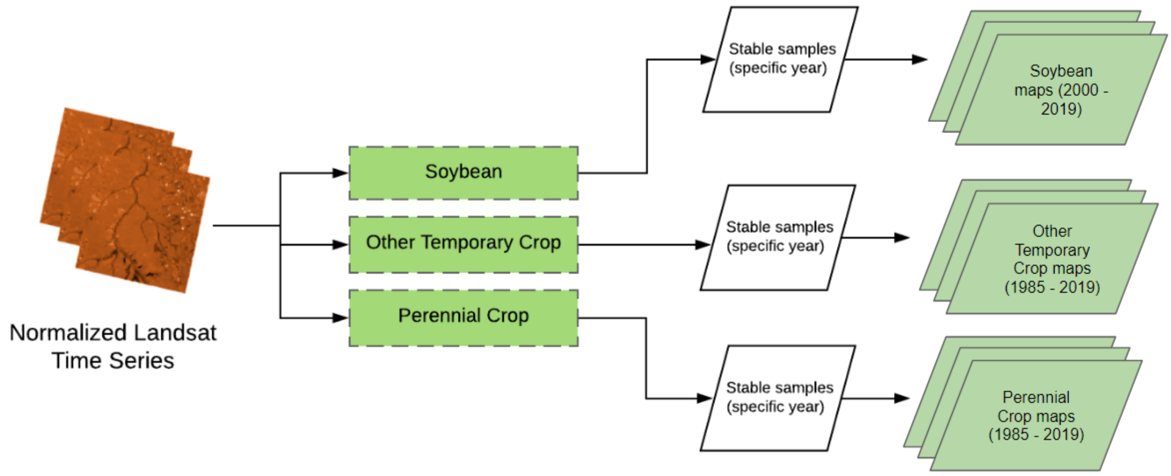



**Figure 11.** The use of stable samples in classes classified with the Normalized Landsat Time Series.

For the group of classes obtained from TOA Landsat time series, as a reference map was not available for each year to be classified, were used annual samples 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 (as illustrated in Figure 12).

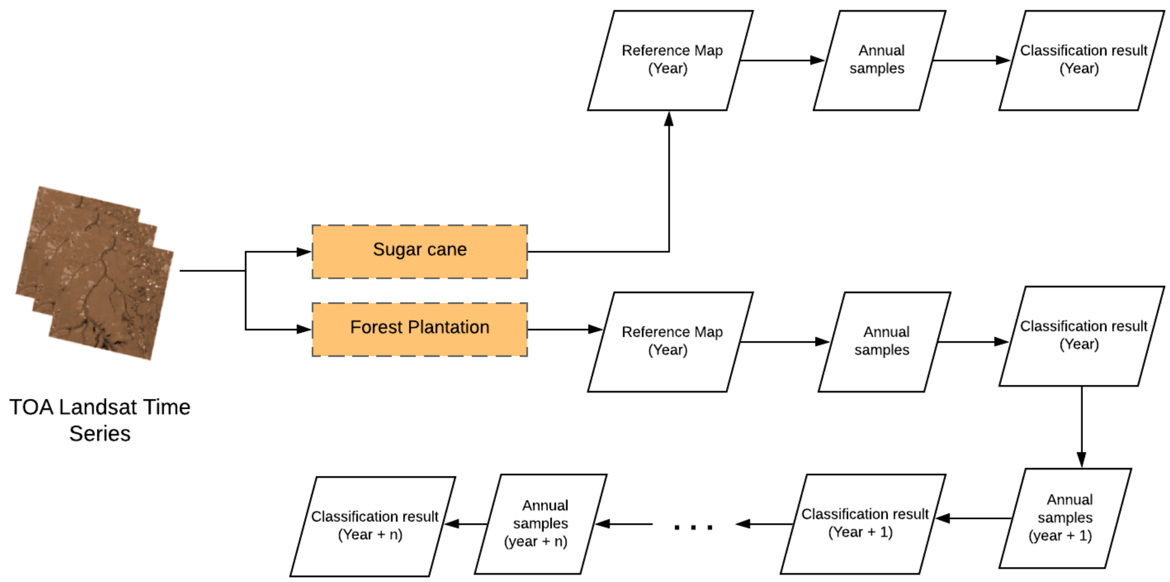



**Figure 12.** The use of annual samples in classes classified with the TOA Landsat Time Series.

The reference maps used to obtain samples to train the classifier are shown in Table 10.

**Table 10.** Reference maps used in the Random Forest classification for the classes Agriculture and Forest Plantation in Collection 5.

| Class | Landsat time series | Number of points | Rule | Type | Year of acquisition | Reference |
|---|---|---|---|---|---|---|
| **Soybean** | Normalized | 10,000 | Minimum of 20% for the interests class | stable samples | 2016 | Agrosatélite (2020A)<br>Agrosatélite (2020B)<br>Song (2020, no prelo) |
| **Sugar cane** | TOA | 10,000 | - | annual samples | 2003 - 2019 | Rudorff et al. (2010) |
| **Other Temporary Crop** | Normalized | 5,000 | Minimum of 20% for the interests class | stable samples | 2016 | Agrosatélite (2020A)<br>Agrosatélite (2020B)<br>Agrosatélite |
| **Perennial Crop** | Normalized | 5,000 | Minimum of 20% for the interests class | stable samples | 2016 | Agrosatélite |
| **Forest Plantation** | TOA | 10,000 | - | annual samples | 2012 - 2014 | Global Forest Watch, Transparent World (2015) |

The Random Forest classifier was configured to training with 100 trees for all classes.

### 6.3.1 Soybean

The classification of soybean crops incorporated two reference maps for the acquisition of training samples for the classifier: the maps produced by Agrosatélite, which consist of maps from the visual interpretation of annual crops (soybeans, corn, and cotton) in the Cerrado and Amazon biomes; and for the other regions where there are an expressive soy area, Song (2020, no prelo) maps for soy in South America were used (Figure 13).

Soybean reference map - Agrosatélite

Soybean reference map - Agrosatélite and Song (2020, no prelo) maps

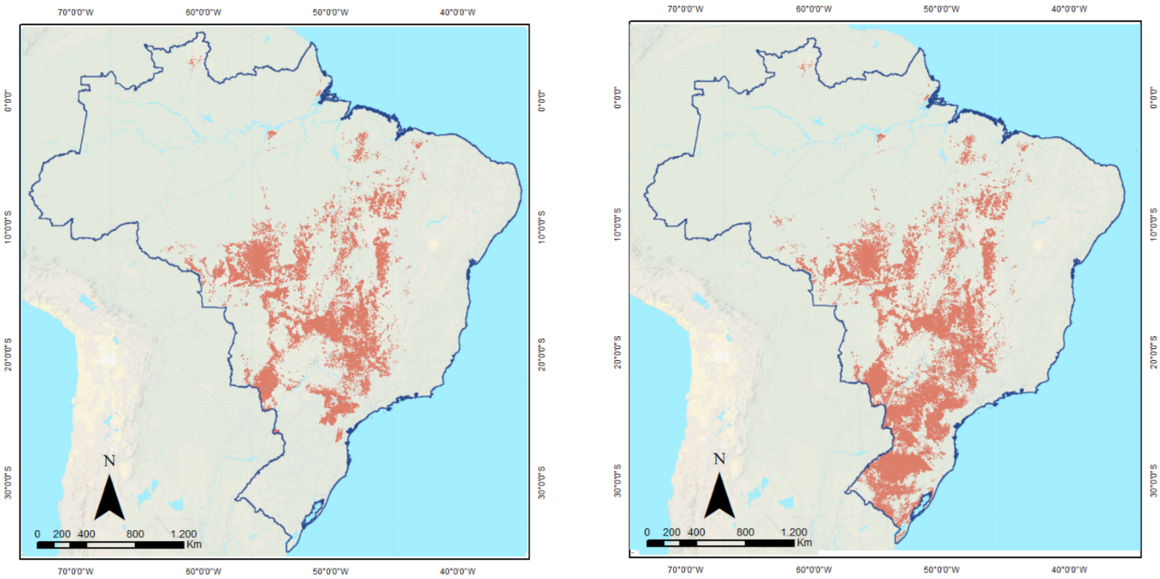



**Figure 13.** Reference maps representing the areas with training samples for the classification of temporary crops in Brazil in Collection 5.

### 6.3.2 Sugar cane

Reference maps developed by the Canasat project (RUDORFF *et al.*, 2010) from 2003 to 2019 were used in the sugar cane classification (Figure 14).

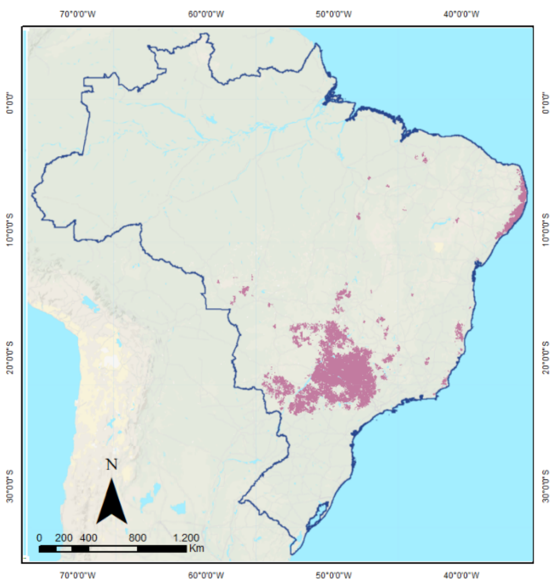



**Figure 14.** Canasat project (RUDORFF *et al*., 2010) map of 2018/2019 representing the areas with training samples for the classification of sugar cane in Brazil in Collection 5.

### 6.3.3 Other Temporary Crop and Perennial Crop

The 'Annual and Perennial Crop' class of previous MapBiomas collections was separated into two classes: 'Temporary Crop' and 'Perennial Crop'. This separation was made in the regions where the reference map showed an expressive quantity of these two kinds of crops, as explained previously (Figure 15).

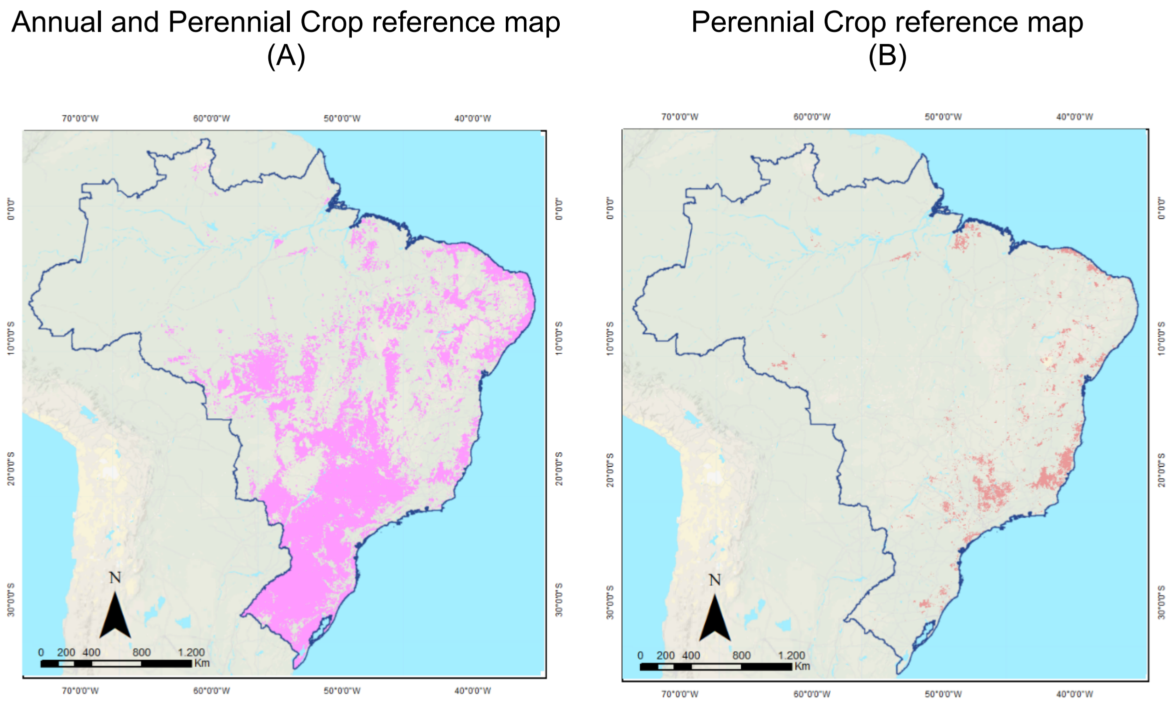



**Figure 15.** Reference maps representing the areas with training samples for the classification of temporary crops and perennial crops in Brazil in Collection 5. (A) Annual and Perennial Crop reference map (Agrosatélite, 2020a); (B) Perennial Crop reference map (Agrosatélite, 2020b).

### 6.3.4 Forest Plantation

The reference map of tree plantations in 2014 provided by Global Forest Watch (TRANSPARENT WORLD, 2015) (Figure 16) was used for the Random Forest training classification of the forest plantation class.


**Figure 16.** Tree plantations map in 2014 provided by Global Forest Watch (TRANSPARENT WORLD, 2015) used for the Random Forest classification of the forest plantation in Collection 5.

## 6.4 Post-classification

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

### 6.4.1 Spatial filter

A convolutional spatial filter, with a 5 x 5 kernel weighted, was applied in the agriculture classes (Figure 17). This spatial filter removes or adds the filtered pixel (central pixel) to the class of interest. The rule for assigning the class of interest to the filtered pixel requires that the sum resulting from the multiplication of the filter values with the map values (1 for the interest class and 0 for the non-interest class) was greater than or equal to 15. If the resulting sum is less than 15 and the filtered pixel belongs to the class of interest, it will be removed from the class of interest and added to the class of non-interest.

| | | | | |
|---|---|---|---|---|
| **1** | **1** | **1** | **1** | **1** |
| **1** | ***2*** | ***2*** | ***2*** | **1** |
| **1** | ***2*** | ***2*** | ***2*** | **1** |
| **1** | ***2*** | ***2*** | ***2*** | **1** |
| **1** | **1** | **1** | **1** | **1** |

**Figure 17.** Representation of the kernel with values for the spatial filter used in the post-

classification process of the Agriculture cross-cutting theme in the Collection 5.

Another spatial filter was used in the forest plantation classification. This spatial filter removes groups of pixels with 6 or less pixels of the interest class. The same was done for the opposite: groups with less than 'other' class were converted to forest plantation (Figure 18).

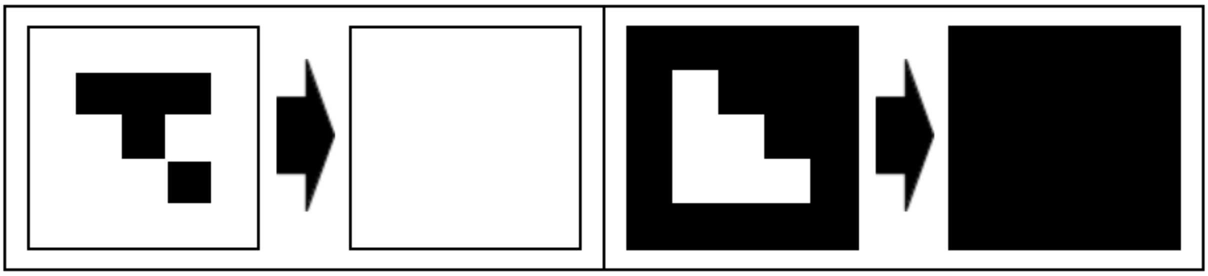


**Figure 18.** Example of the spatial filter used in forest plantation post-processing. The image on the left shows an exclusion of pixels of the interest class (in black). The image on the right shows an inclusion of pixels of "other classe" (in white) to the interest class.

### 6.4.2 Temporal filter

#### 6.4.2.1 Soybean

The temporal filtering was carried out from 2001 to 2018, using a 3-year window: the current year in process of post-processing, the previous year and the following year. The first and last years (*i.e.* 2000 and 2019) were post-processed separately after the other years, as the lack of previous or later data restricted the use of the same filter rule.

For the years 2001 to 2018, the temporal filter rule excluded soy pixels when they were not classified as soy in the previous and following year, and included soy pixels when they were classified in the previous and following year as soy. For the previous year, the post-processed data was considered. These results were then used to process the first and last year of the time series. For the first year of the series (*i.e.* 2000), pixels were excluded when, in the following year, they were not classified as soy. For the last year of the time series (*i.e.* 2019), pixels were included when these, in the previous year, were classified as soybeans (Figure 19).

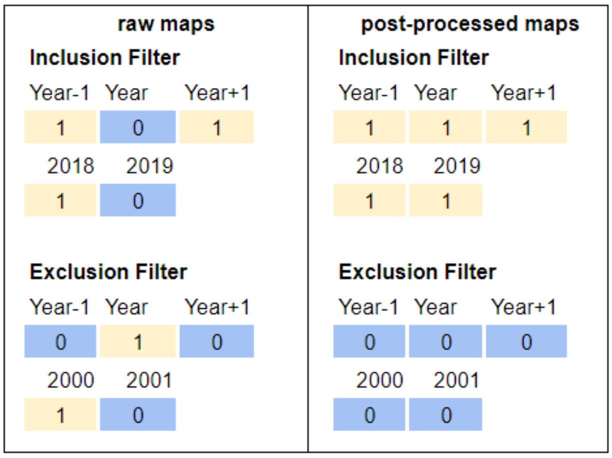



**Figure 19.** Temporal filter rules used in soy post-processing. The inclusion filter changed a pixel to soybean when the same pixel was soybean in the adjacent years or, in the last year of the time series, when the pixel was soybean in the prior year. The exclusion filter changed a pixel to “other class” when the same pixel was not soybean in the adjacent years or, in the first year of the series, when the pixel was not soy in the following year.

#### 6.4.2.2 Sugar cane

The temporal filter of the sugar cane maps was the same as the method of the ‘Semi-Perennial Crop’ class of Collection 4, which used  a 5-year window (the year of interest and 2 year before and 2 years after) to remove noise from the classification. The temporal filter changes a pixel of the year evaluated according to the years of the established window (i.e. 2 years). For example, if the evaluated pixel was in the same class as at least two other pixels (from the previous year, ahead or both) it will remain in the same class. However, if the evaluated pixels do not belong to the same class of at least two pixels (the year of interest, 2 years before and 2 years after), their class will be changed (Figure 20).

| raw maps | | | | | post-processed maps | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Inclusion filter** | | | | | **Inclusion filter** | | | | |
| year-2 | year-1 | year | year+1 | year+2 | year-2 | year-1 | year | year+1 | year+2 |
| 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1985 | 1986 | 1987 | 1988 | | 1985 | 1986 | 1987 | 1988 | |
| 0 | 1 | 1 | | | 1 | 1 | 1 | | |
| 1 | 0 | 1 | 1 | | 1 | 1 | 1 | 1 | |
| 2016 | 2017 | 2018 | 2019 | | 2016 | 2017 | 2018 | 2019 | |
| 1 | 1 | 0 | 1 | | 1 | 1 | 1 | 1 | |
| | 1 | 1 | 0 | | | 1 | 1 | 1 | |
| **Exclusion Filter** | | | | | **Exclusion Filter** | | | | |
| year-2 | year-1 | year | year+1 | year+2 | year-2 | year-1 | year | year+1 | year+2 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1985 | 1986 | 1987 | 1988 | | 1985 | 1986 | 1987 | 1988 | |
| 1 | 0 | 0 | | | 0 | 0 | 0 | | |
| 0 | 1 | 0 | 0 | | 0 | 0 | 0 | 0 | |
| 2016 | 2017 | 2018 | 2019 | | 2016 | 2017 | 2018 | 2019 | |
| 0 | 0 | 1 | 0 | | 0 | 0 | 0 | 0 | |
| | 0 | 0 | 1 | | | 0 | 0 | 0 | |

**Figure 20.** Temporal filter rules used in sugar cane post-processing. A 5-year window was used. Pixels of sugar cane were included when at least 2 of the 4 adjacent years were also sugar cane. Pixels of sugar cane were excluded when none of the 4 adjacent years were sugar cane.

#### 6.4.2.3 Other Temporary Crop

For the Landsat scenes that underwent temporal filter, the following rule was applied: in a 3-year window, pixels classified as temporary crop in the previous and next year were updated for the same class; pixels classified as temporary crops, but that did not receive this classification in the previous and following year, were excluded from the class of interest. The exceptions were the first and last year of the time series. In the first year (i.e. 1985), those pixels that did not belong to it were excluded from the class of interest. In the last year (i.e. 2019), pixels that belonged to the interest class were included in the interest class (Figure 21).

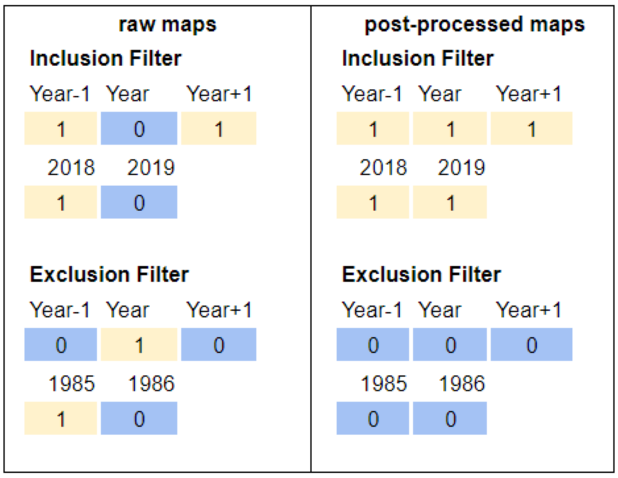



**Figure 21.** Temporal filter rules applied in other temporary crop post-processing. The inclusion filter changed a pixel to temporary crop when the same pixel was temporary crop in the adjacent years or, in the last year of the time series, when the pixel was temporary crop in the prior year. The exclusion filter changed a pixel to "other class" when the same pixel was not a temporary crop in the adjacent years or, in the first year of the series, when the pixel was not a temporary crop in the following year.

#### 6.4.2.4 Perennial Crop- *beta version*

The temporal filter of inclusion for the class 'Perennial Crop' used a 5-year window (year of interest, two years earlier, and two years ahead) to remove noise from the classification. The time filter changes a pixel of the year evaluated according to the years of the established window (*i.e.* 5 years). If the pixel does not belong to the same class as the other pixels within the window (*i.e.* if the pixel of the year evaluated belongs to the class 'Temporary Crop', but in the other years of the window it was classified as 'Perennial Crop') it will be changed to the class of the other pixels within the window (*i.e.* 'Perennial Crop') (Figure 22).

The temporal exclusion filter removes 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. If the pixel in the previous year (year - 1), which has already passed through the exclusion filter, was perennial crop, it means that it is within a minimum interval of 5-year of this class and, therefore, if the pixel of the current year is also perennial crop, it is within the same 5-years interval and should not be excluded. If the pixel in the previous year was not perennial crop, but it is in the current year, the next 4 years are verified: if all of them are perennial crop, this means that the pixel in the current year is within a minimum interval of 5 years and should not be excluded; otherwise, the pixel is not within a minimum interval of 5 years and should be excluded. The exceptions to these rules were for intervals with less than 5 years at the beginning or at the end of the series (i.e. 1985 and 2019). For the beginning of the series, only the pixels in an interval of less than 3 years were excluded; in the last year, no pixels were excluded, whatever the intervals they were in. These filters are illustrated in Figure 22.

**raw maps**

**Inclusion Filter**

| Year-2 | Year-1 | Year | Year+1 | Year+2 |
|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 1 |

**Exclusion Filter**

| Year-1 | Year | Year+1 | Year+2 | Year+3 | Year+4 |
|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 0 |

| 1985 | 1986 | 1987 | 1988 | 1989 |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 |

| 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 1 |

**post-processed maps**

**Inclusion Filter**

| Year-2 | Year-1 | Year | Year+1 | Year+2 |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |

**Exclusion Filter**

| Year-1 | Year | Year+1 | Year+2 | Year+3 | Year+4 |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |

| 1985 | 1986 | 1987 | 1988 | 1989 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 |

| 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 1 | 1 | 1 |

**Figure 22.** Temporal filter rules used in perennial crop post-processing. These filters are explained in the paragraph above.

#### 6.4.2.5 Forest Plantation

A temporal filter of 5 years (the year of interest, 2 years before and 2 years after) was used in the post-classification process of the time series of forest plantation maps resulting from the supervised classification of the image mosaics from the period of 1985 to 2018. The rule applied by the temporal filter took into account the persistence of the forest plantation.

The temporal filter can change a pixel of the assessed year depending on the class (class of interest or class of non-interest) of the pixels in the two previous and the two following years. For example, if the assessed pixel was from the same class of at least two other pixels (previous, ahead, or both) it will remain in the same class. However, if the assessed pixel is not from the same class of at least two pixels (previous, ahead, or both), its class will be changed (Figure 23).

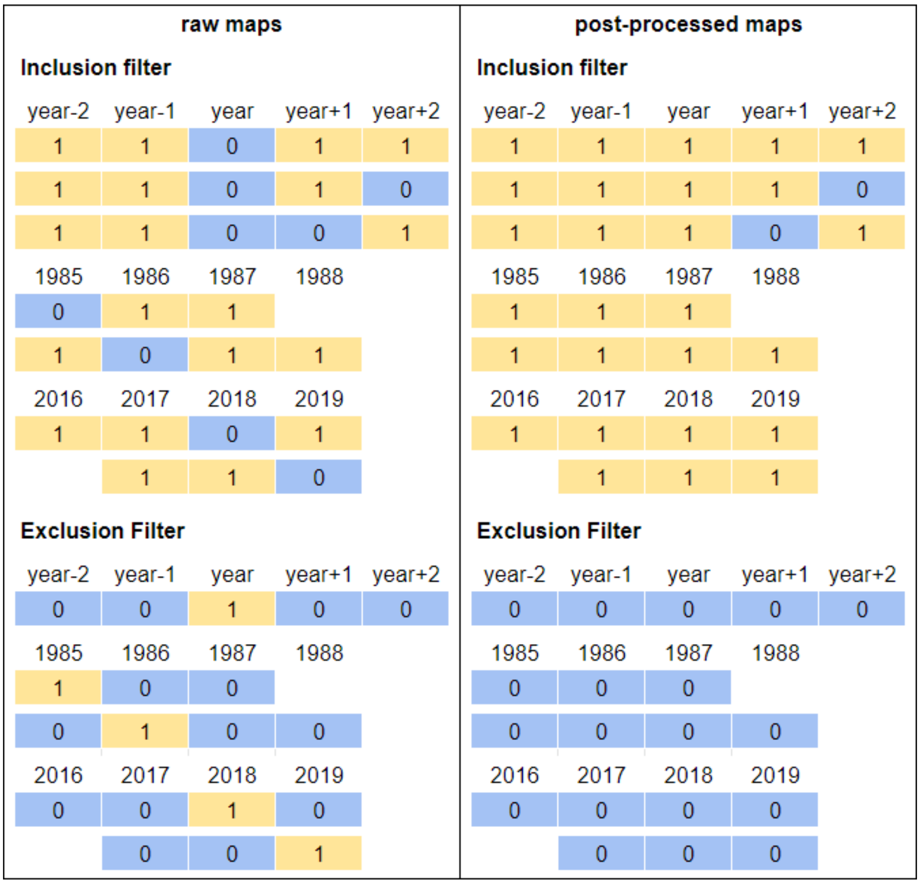



**Figure 23.** Temporal filter rules used in forest plantation post-processing. A 5-year window was used. Pixels of forest plantation were included when at least 2 of the 4 adjacent years were also forest plantation. Pixels of forest plantation were excluded when none of the 4 adjacent years were forest plantation.

### 6.4.3 Integration with biomes and themes

After the classification of theAgriculture and Forest Plantation themes, they were integrated to the other land use and land cover classes to compose the MapBiomas Collection 5 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.

# 7 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.

## 7.1 Accuracy analysis

The map accuracy analysis was produced using independent validation points provided by the 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. Accuracy analyzes from LAPIG points was performed in the following classes: ‘Forest Plantation’, ‘Perennial Crop’ and ‘Temporary Crop’. The 'Temporary Crop' class contains the classes 'Soy', 'Sugarcane' and ‘Other Temporary Crops’. LAPIG points used for the accuracy analyzes are shown in Figure 24.

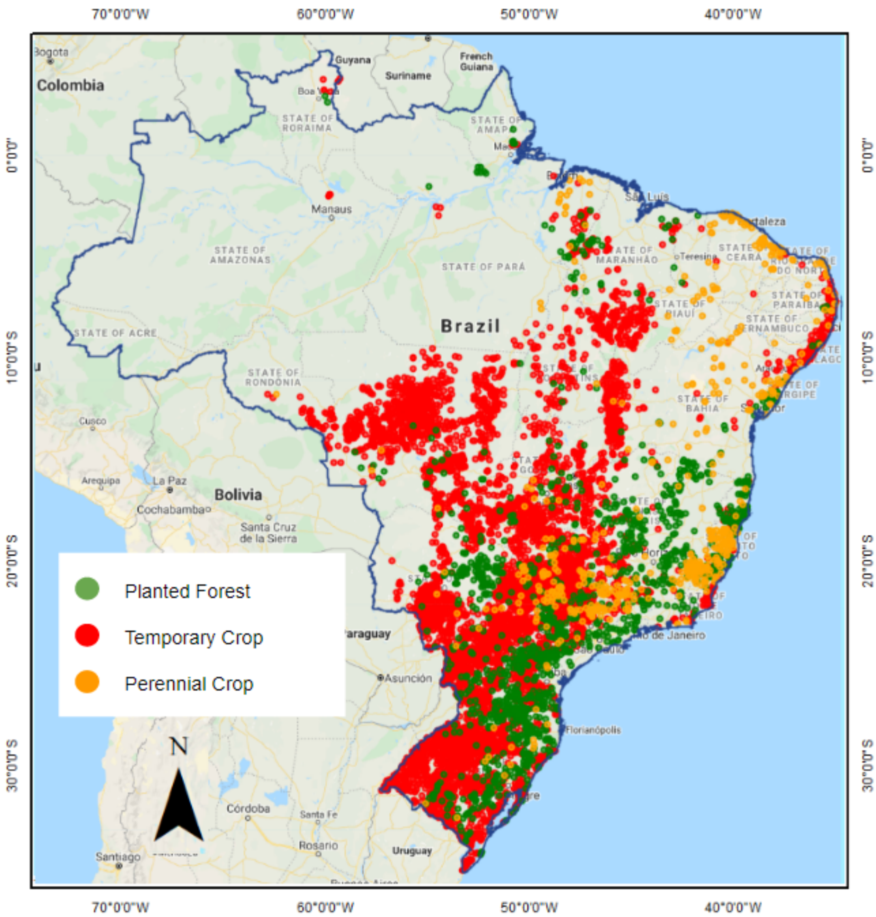



**Figure 24.** LAPIG points used for the accuracy analyzes of ‘Temporary Crops’, ‘Perennial Crops’ and ‘Forest Plantation’ classes.

#### 7.1.1.1 Temporary Crop

The result of the accuracy analysis of the ‘Temporary Crop’ class (Figure 25) showed that the accuracy of this class increases over the years, reaching higher values in the final years of the series. Throughout the series, the maps accuracies were above 60%, reaching the highest values after 2008 (*i.e*. producer and user accuracy above 80%).

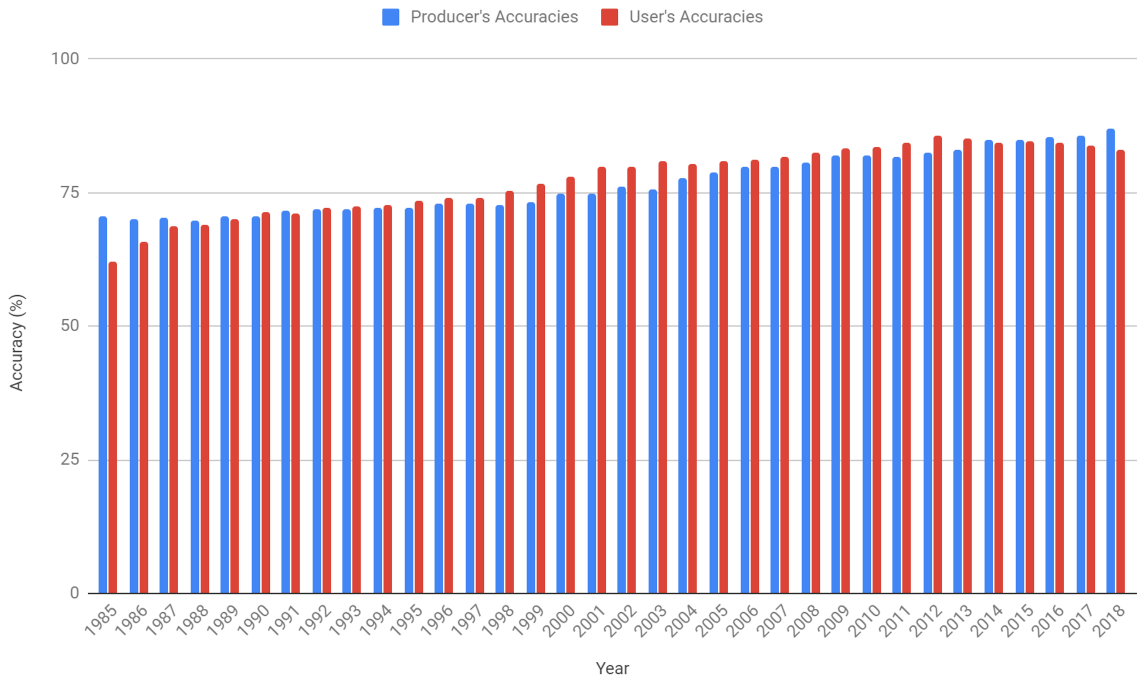



**Figure 25.** Producer and user's accuracy of the temporary crop class in Collections 5.

#### 7.1.1.2 Perennial Crop - *beta version*

The perennial crop accuracy analysis was performed within the mask of the 'annual and perennial crop' map. Thus, only the LAPIG validation points that overlapped within the annual and perennial crop mask were considered.

The accuracy analysis results (Figure 26) showed that the producer's accuracy increases throughout the series, reaching its maximum value (64%) in the last year analyzed in the series (2018). In addition, in the first three years of the series, the separation between 'Temporary Crop' and 'Perennial Crop' was not successful. Regarding the user's accuracy, the highest values were recorded in 1996, 1997 and 1998, with the user's accuracy above 70%. After this period, the user's accuracy remained stable until the end of the series, reaching a value of 32% in the last year analyzed (i.e. 2018).

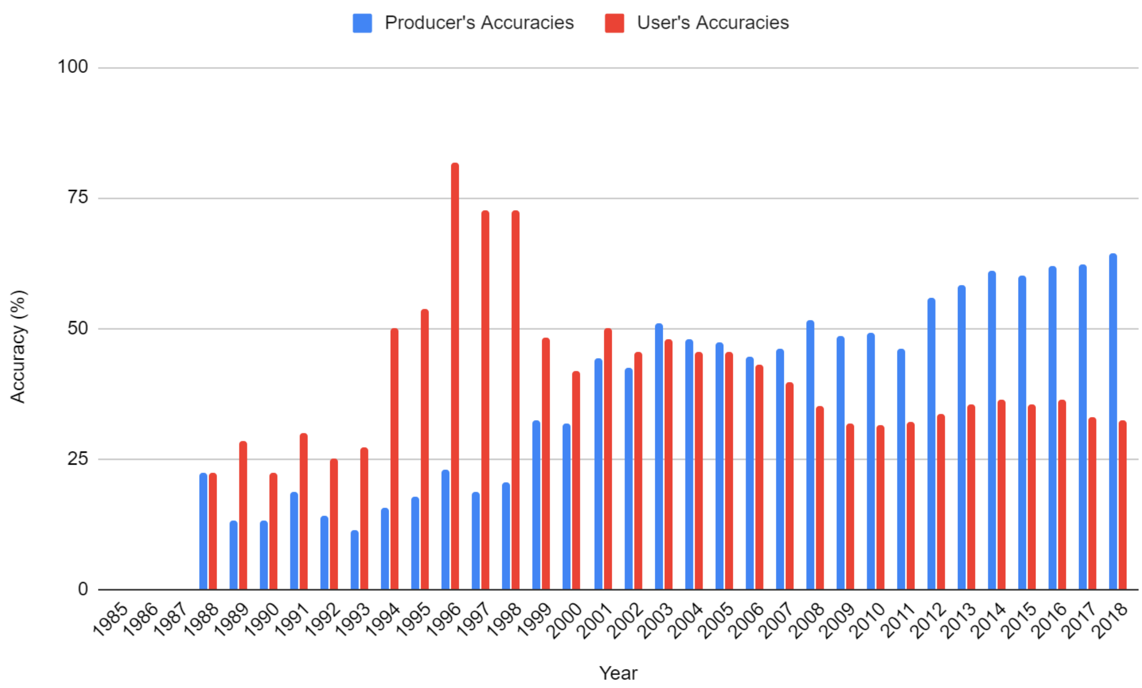



**Figure 26.** Producer' and user's accuracy of the perennial crop class in Collections 5.

### 7.1.2 Forest Plantation

The accuracy analyses of Forest Plantation in Collection 5 were similar to previous MapBiomas collections: high values for user's accuracy (above 75% in all years) and increasing values for the producer's accuracy. The result of the accuracy calculations for the Forest Plantation theme is shown in Figure 27.

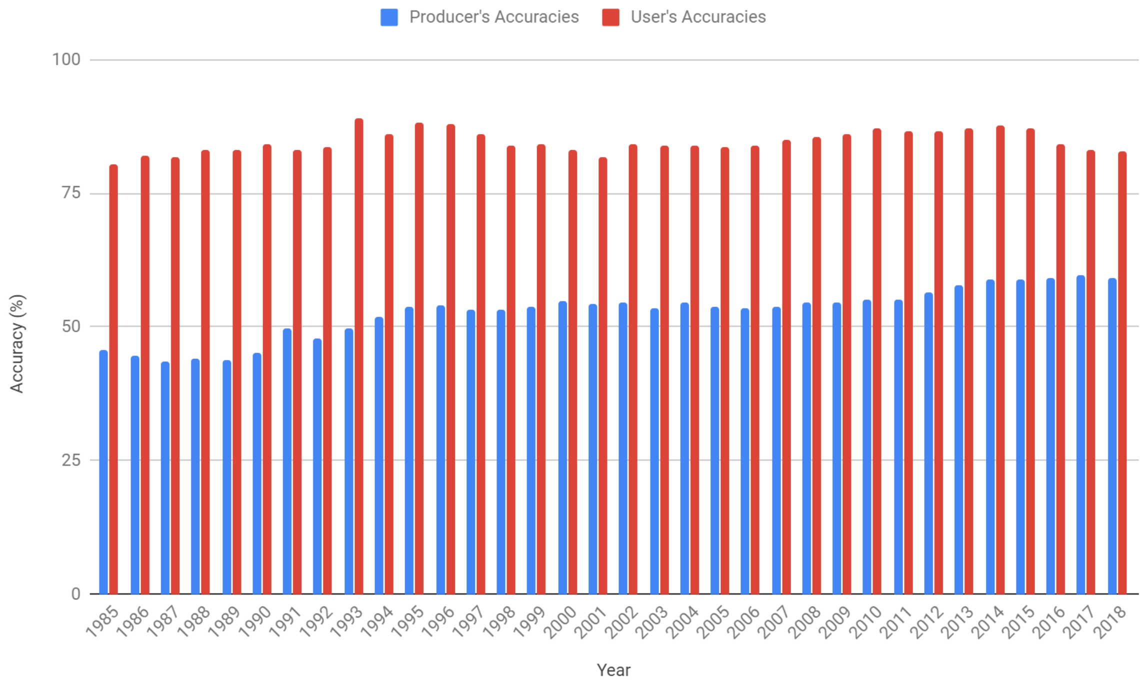



**Figure 27.** Producer and user's accuracy of the forest plantation class in Collections 5.

### 7.2 Comparison with reference maps

The 'Agriculture' and 'Forest Plantation' maps in the MapBiomas Collection 5 were compared to the reference maps used for training the classifier. The comparisons for each class are shown in the following figures.

Figure 28 shows the comparison between the MapBiomas soybean map and the soybean reference map (Agrosatelite, 2017) for the region of Sinop in Mato Grosso state. The region is known for a large amount of land used for soybean production.

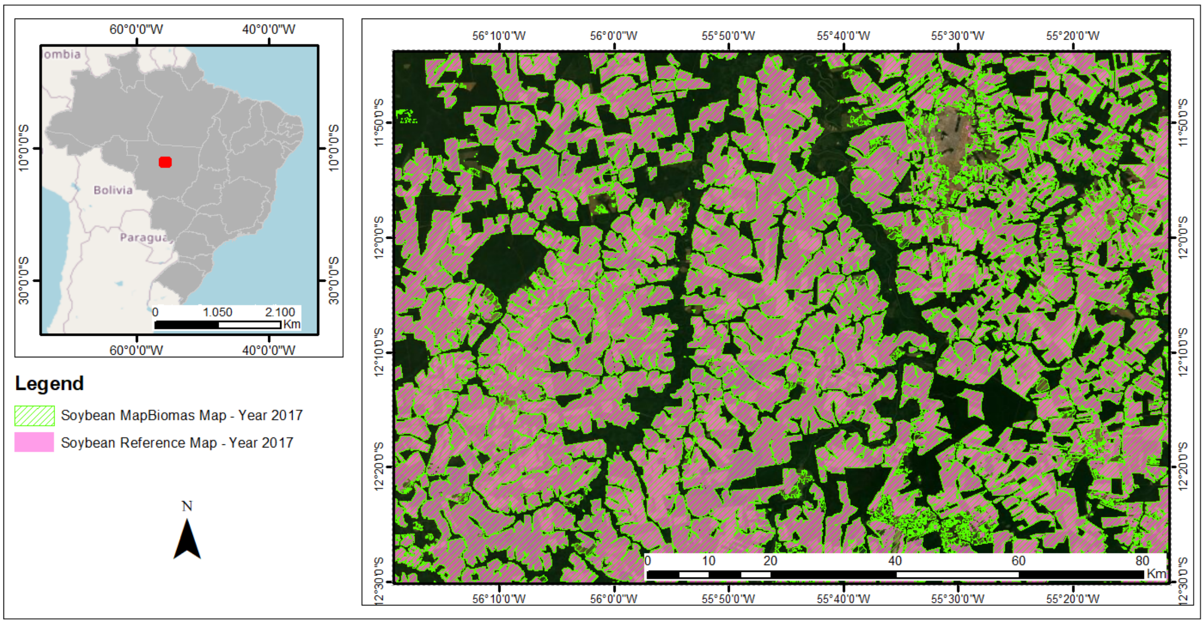


**Figure 28.** Comparison between the MapBiomas soybean map and the soybean reference map (Agrosatelite, 2017) for the region of Sinop in Mato Grosso state.

The map in Figure 29 compares the MapBiomas soybean map and the soybean reference map (Agrosatelite, 2017) for the region of Santarém in Pará state. The western region of the state of Pará has been standing out in soy production in recent years in the Amazon region.

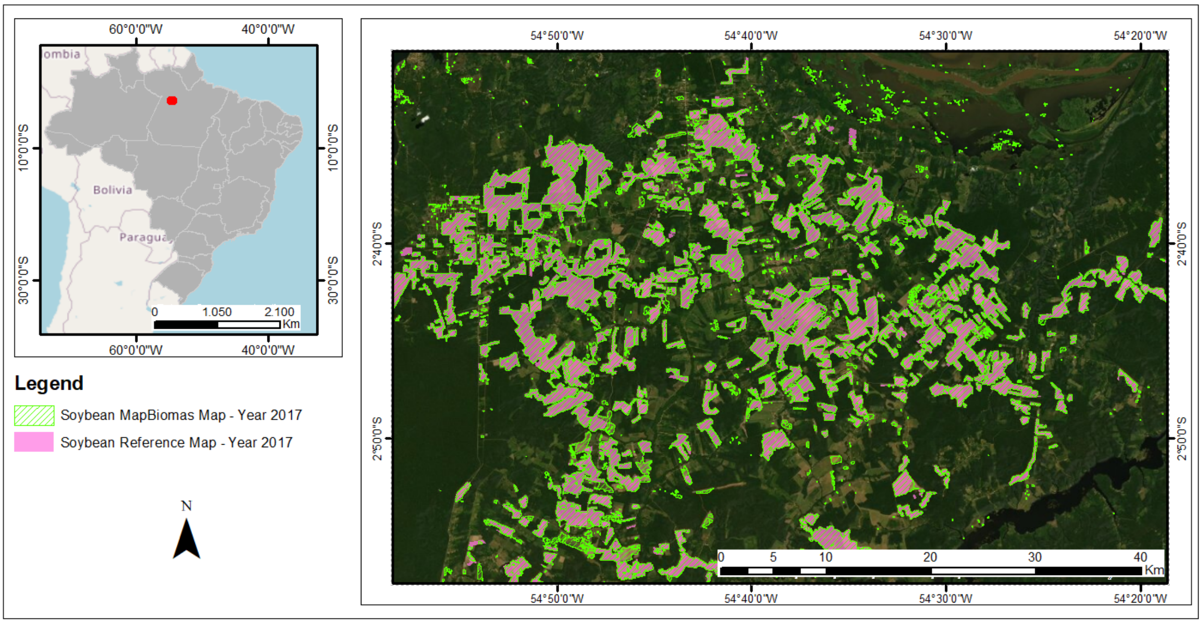



**Figure 29.** Comparison between the MapBiomas soybean map and the soybean reference map (Agrosatelite, 2017) for the region of Santarém in Pará state.

The map in Figure 30 compares the reference map used for the mapping of sugar cane (Canasat) and the map of MapBiomas in the northwest region of São Paulo, a major cane producer in the region.

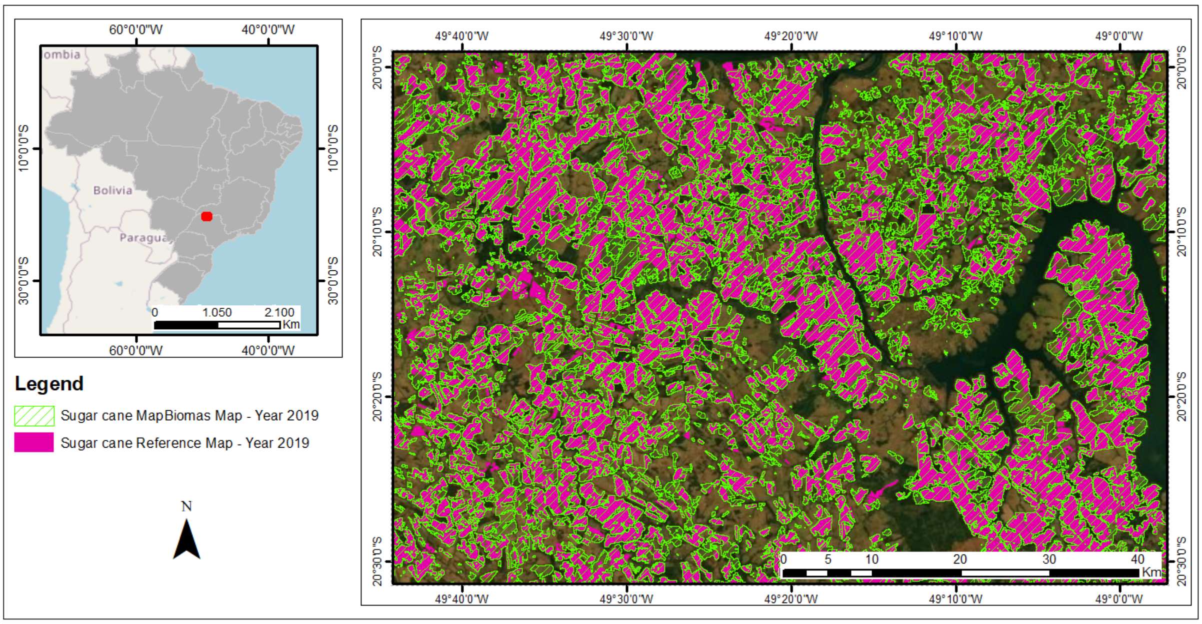


**Figure 30.** Comparison between the MapBiomas sugar cane map and the sugar cane reference map (Canasat, 2019) for the region of São José do Rio Preto in São Paulo state.

The map in Figure 31 shows the comparison between the 'Perennial Crop' reference map (Agrosatellite) and 'Perennial Crop' MapBiomas map on the northeast coast.

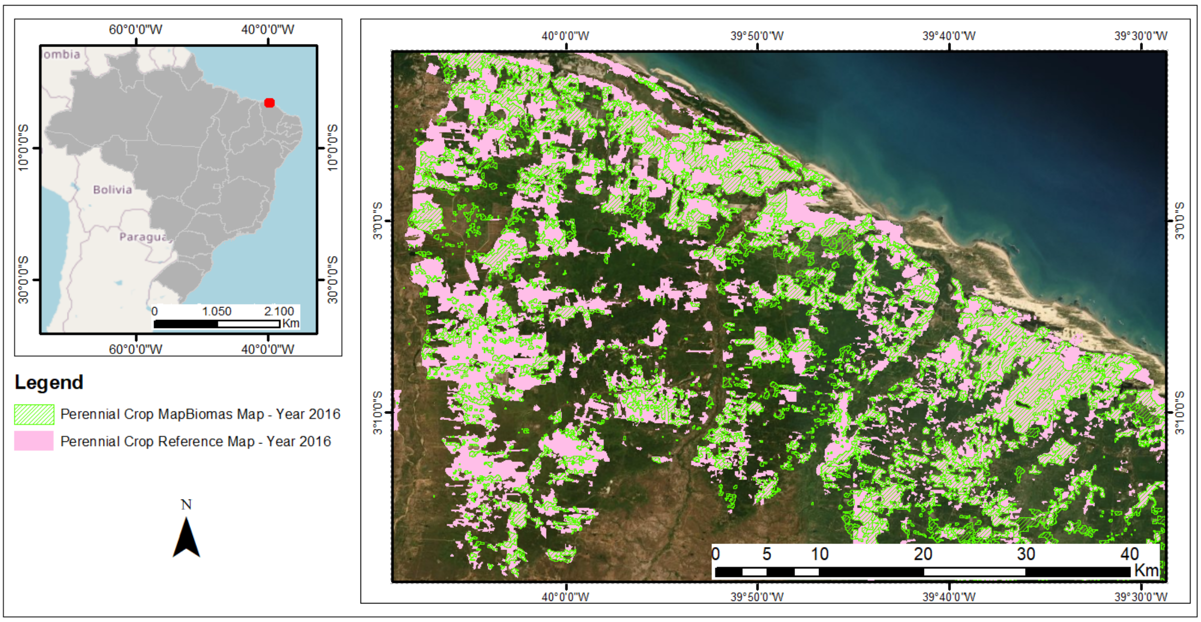



**Figure 31.** Comparison between the MapBiomas sugar cane map and the sugar cane reference map (Canasat) for the region of Ribeirão Preto in São Paulo state.

The map in Figure 31 shows the comparison between the 'Forest Plantation' reference map (TRANSPARENT WORLD, 2015) and 'Forest Plantation' MapBiomas map on the Santa Catarina state.

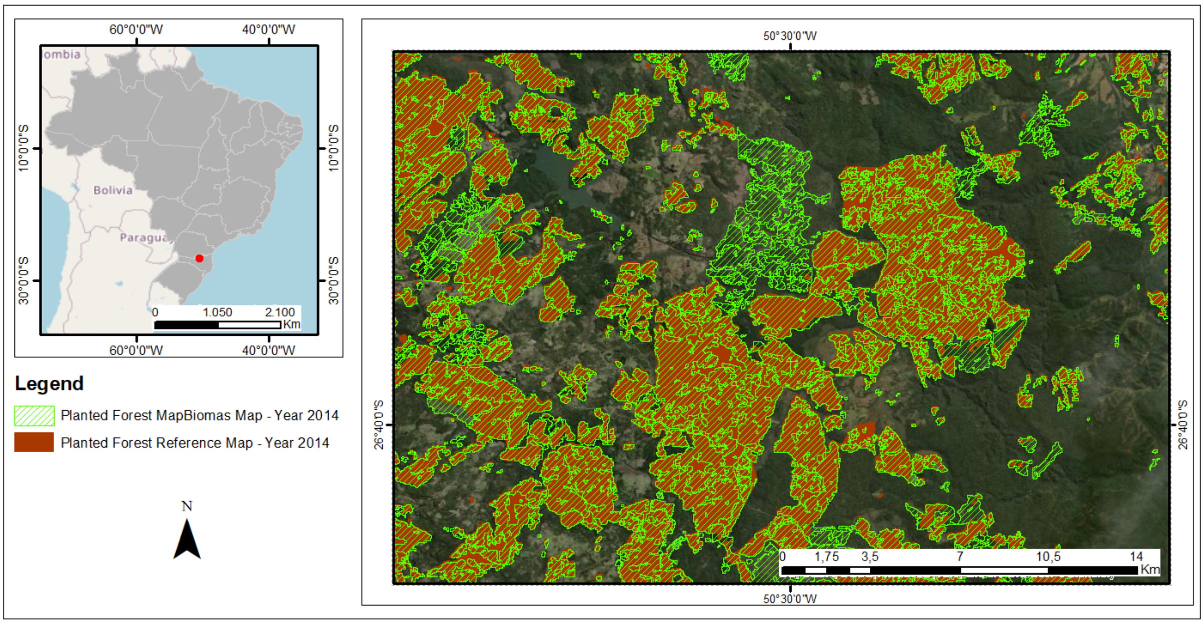


**Figure 32.** Comparison between the MapBiomas forest plantation map and the forest plantation reference map (TRANSPARENT WORLD, 2015) for the region of Três Barras, in Santa Catarina state.

## 7.3 Comparison with reference data

In addition to the comparison with reference maps and validation points, a comparison between the 'Agriculture' and 'Planted Forest' maps of MapBiomas Collection 5 with data from the Systematic Survey of Agricultural Production (LSPA - Levantamento Sistemático da Produção Agrícola), carried out by the Brazilian Institute of Geography and Statistics (IBGE), was also made, and these are considered official data for estimating agricultural area in the country.

### 7.3.1 Comparison of Other Temporary Crop area

The graph in Figure 33 shows the comparison between the area of the class 'Temporary Crops' (which includes sugar cane, soybean and 'Other Temporary Crops') with the areas estimated by LSPA - IBGE.

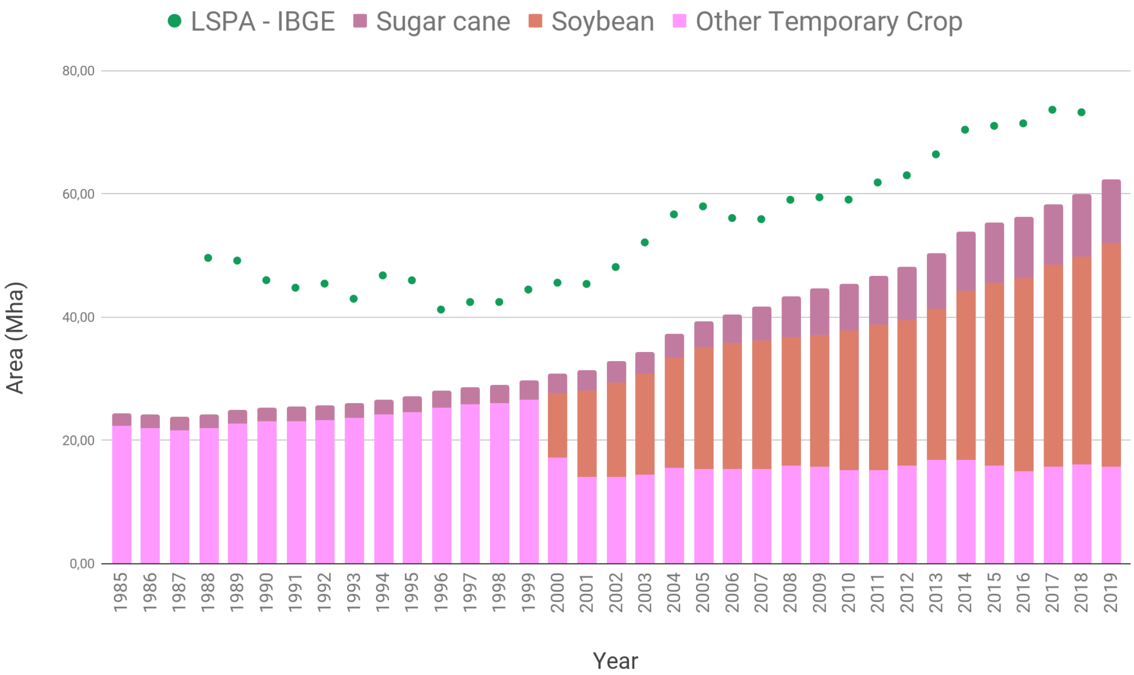



**Figure 33.** Comparison between MapBiomas Temporary Crop area and LSPA-IBGE Temporary crop Area.

### 7.3.2 Comparison of Soybean area

According to the graph above (Figure 33), soybean started to be mapped only after 2000. Since LSPA does not produce a map annually, but only estimates of area by municipality, a class map of the area cultivated by municipalities of the federation was made from 2000 to 2019 to compare the spatial distribution of soybean in Brazil. The results are shown in Figure 34.

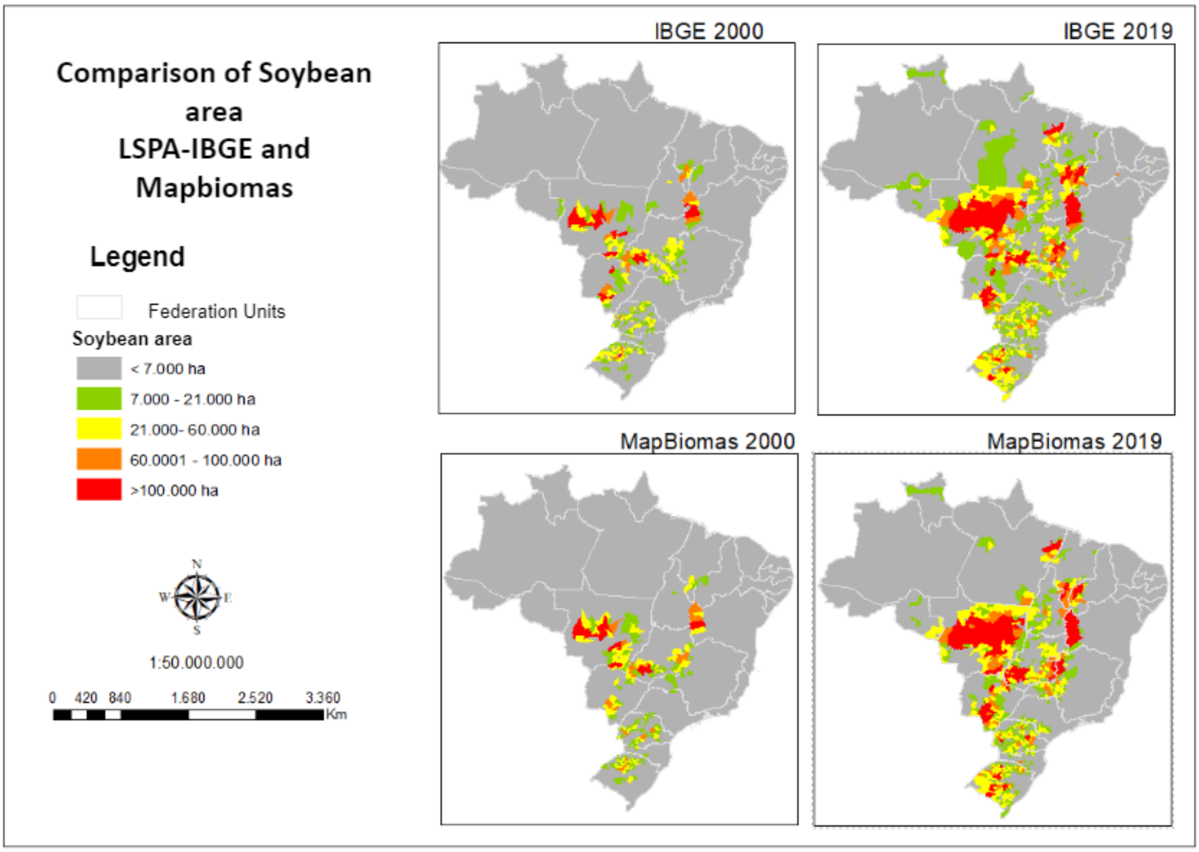



**Figure 34.** Comparison between the MapBiomas Soybean spatial distribution in Brazil and LSPA-IBGE Soybean spatial distribution in the years 2000 and 2019.

The total areas cultivated with soybean in Brazil were compared and shown in Figure 35.

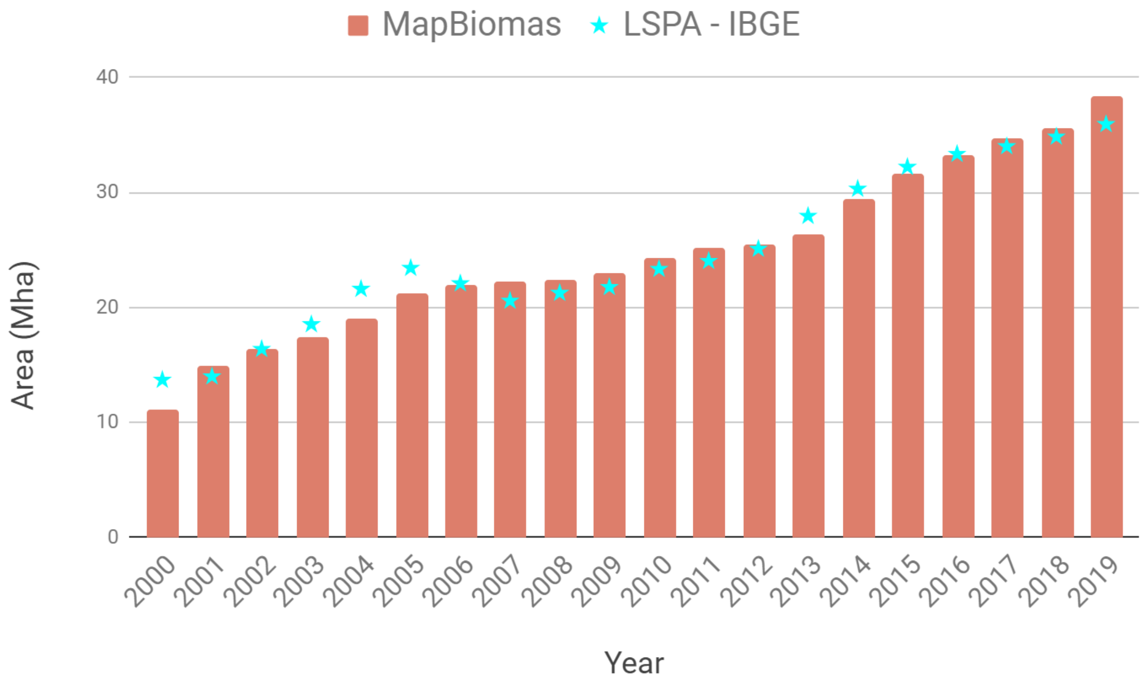



**Figure 35.** Comparison of areas between the 'Soybean' map of MapBiomas Collection 5 and LSPA- IBGE from 2000 to 2019.

### 7.3.3 Comparison of Sugar Cane area

The area cultivated with sugarcane was also compared with the area estimated by LSPA. The results of the comparison between classes of planted area by municipality are shown in Figure 36.

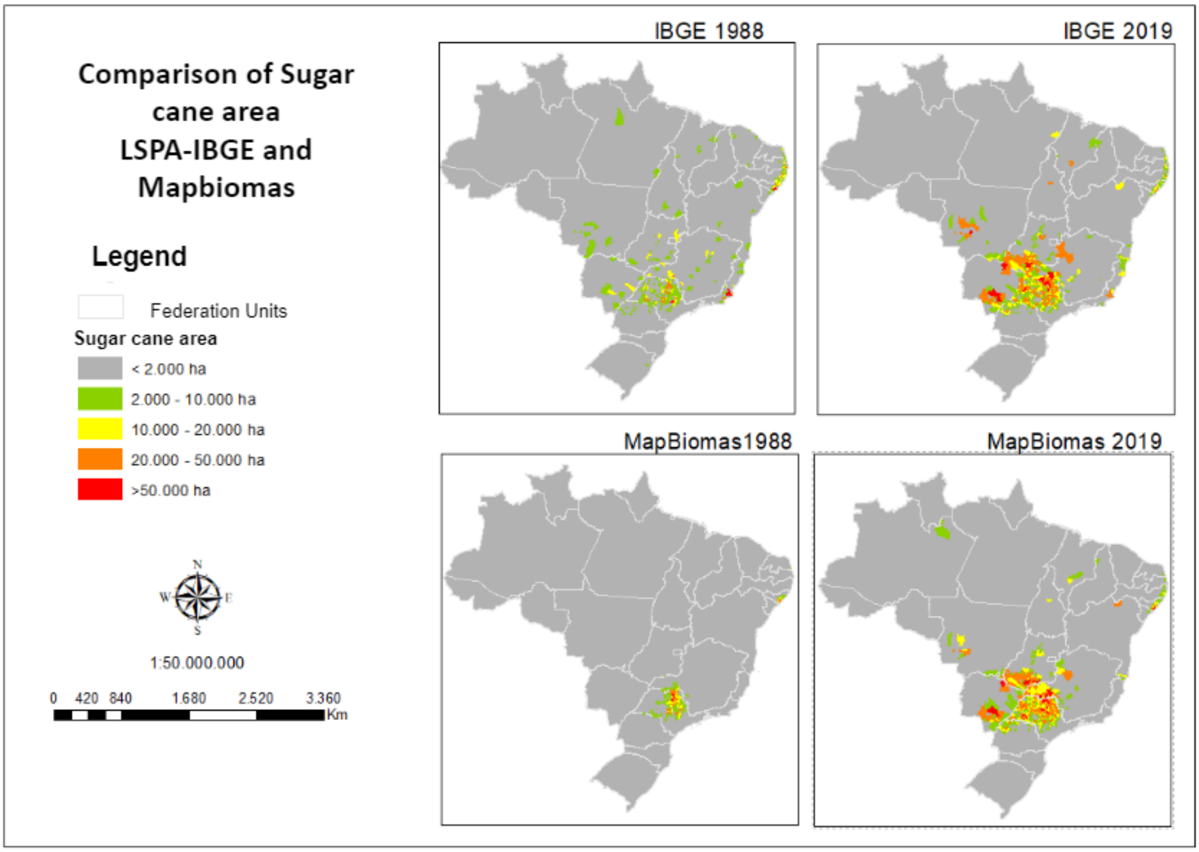



**Figure 36.** Comparison of area classes and spatial distribution by municipality between the 'Sugar cane' map of MapBiomas Collection 5 and PAM- IBGE in the years 1988 and 2019.

The total areas cultivated with sugarcane in Brazil were compared and shown in Figure 37.

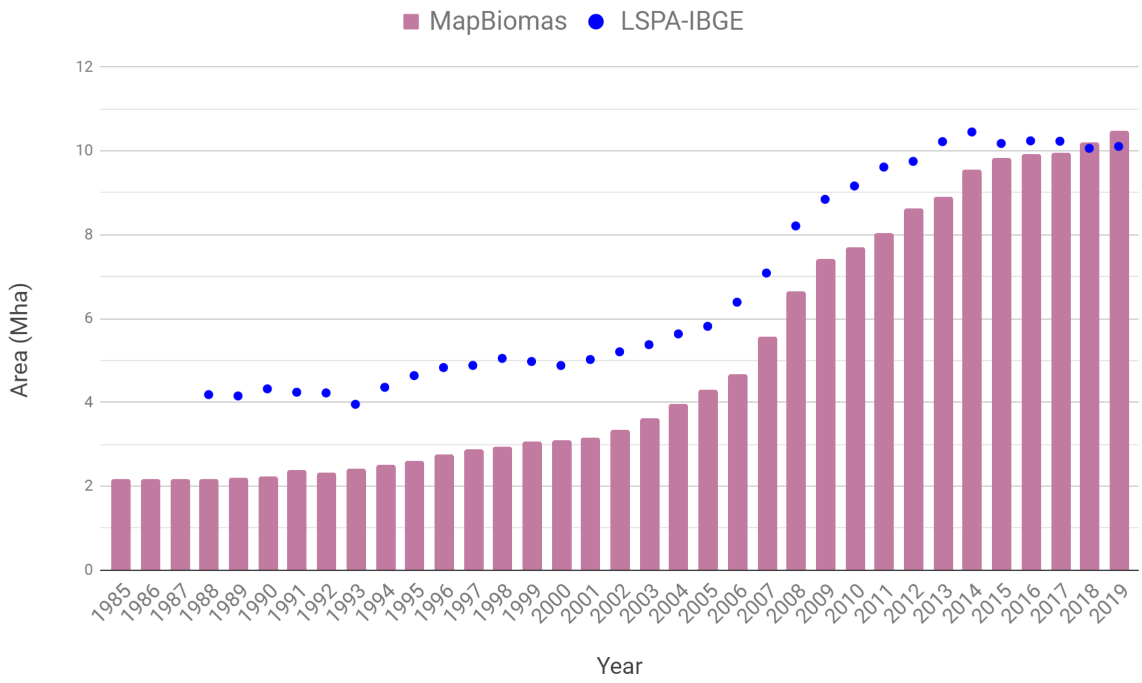



**Figure 37.** Comparison of areas between the 'Sugar cane' map of MapBiomas Collection 5and LSPA - IBGE for the period from 2000 to 2019.

### 7.3.4 Comparison of Forest Plantation area

Planted forest areas obtained from MapBiomas Collection 5 annual maps were also compared with areas from official sources. In order to estimate Brazil's forestry area, a comparison was made between MapBiomas, IBGE's Production of Vegetable Extraction and Silviculture (PEVS-IBGE) and Ibá - *Indústria brasileira de árvores* (Brazilian Tree Industry). The results are shows in Figure 38.

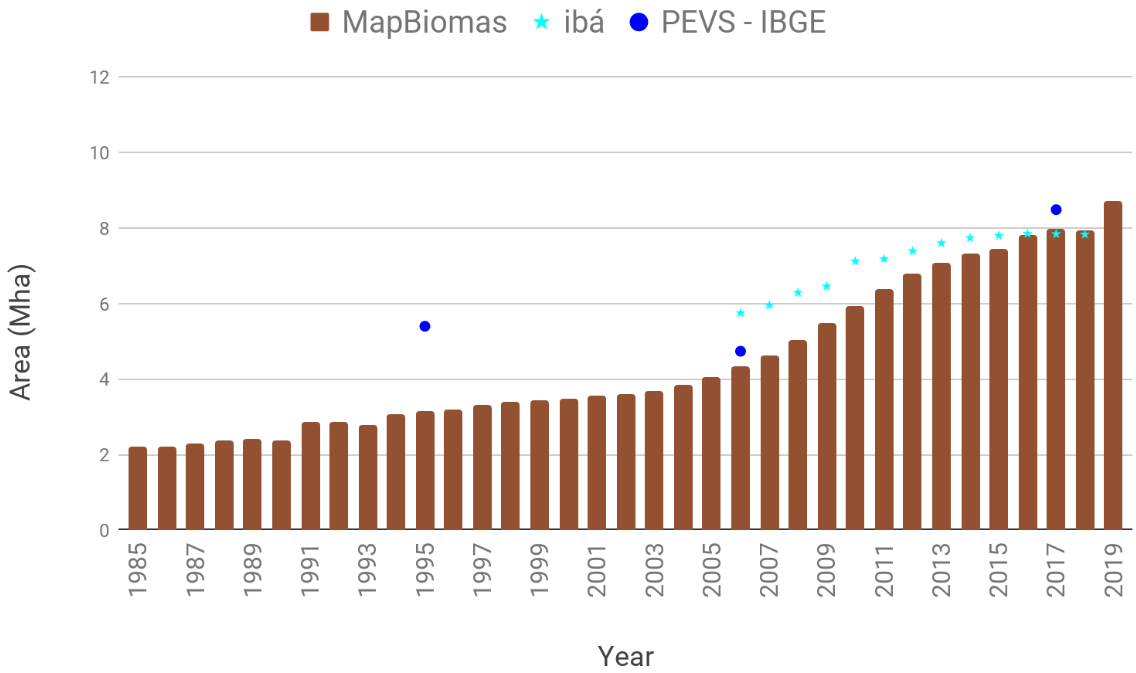



**Figure 38.** Comparison of areas between the 'Forest Plantation' map of MapBiomas Collection 5, PEVS-IBGE and Ibá from 2000 to 2019.

A comparison of the spatial distribution of municipalities with planted forest plantations in Brazil was also made. The comparison between IBGE-PEVS and MapBiomas can be seen in Figure 39.

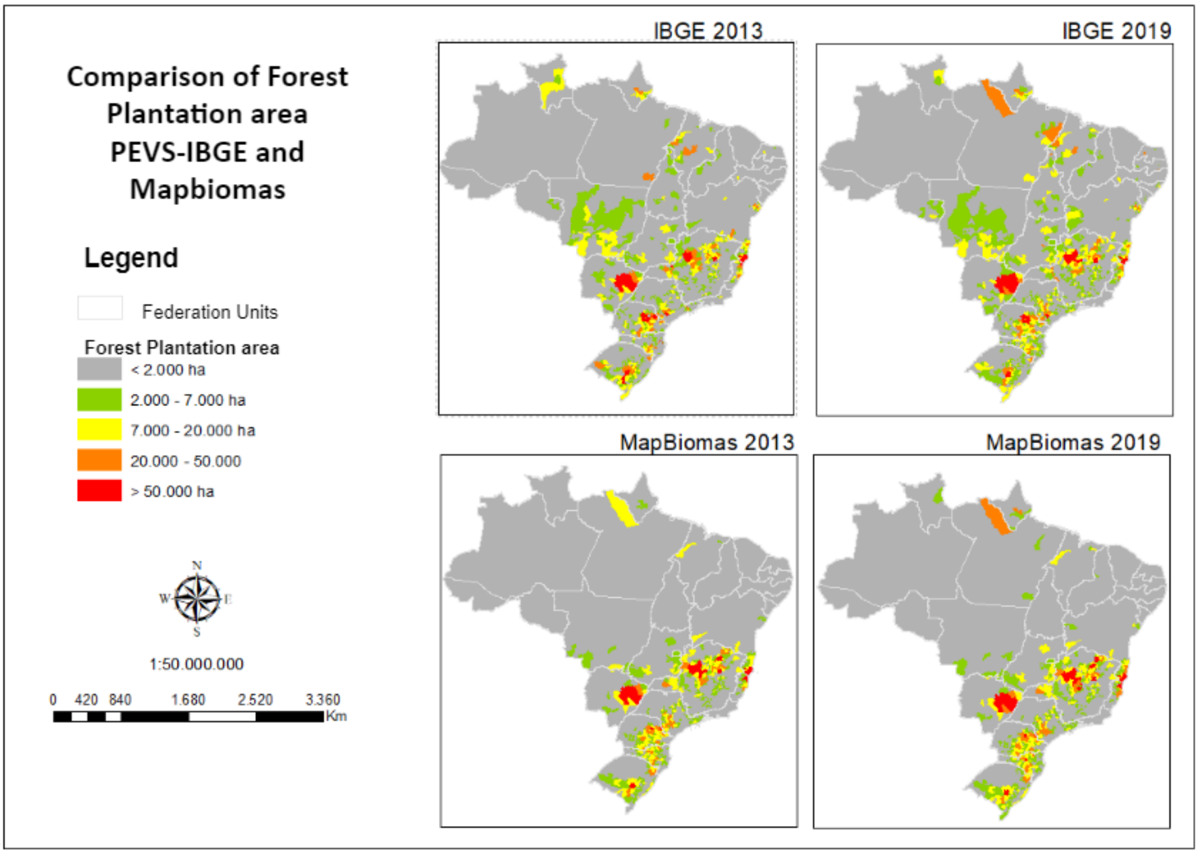



**Figure 39.** Comparison of area classes and spatial distribution by municipality between the 'Forest Plantation' map of MapBiomas Collection 5 and PEVS-IBGE in 2013 and 2019.

# 8 References


AGROSATÉLITE. Agrosatélite Applied Geotechnology Ltd. Scripts used for Agriculture and Forest Plantation classification on Collection 5 of Mapbiomas project. **GitHub repository**, 2020. Available at: <https://github.com/mapbiomas-brazil/agriculture/tree/mapbiomas50> and <https://github.com/mapbiomas-brazil/forest-plantation/tree/mapbiomas50>.

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

CONAB. Companhia Nacional de Abastecimento. Calendário de plantio e colheita de grãos no Brasil. 2019. 75p. ISSN: 2318-6852. Available at: <https://www.conab.gov.br/institucional/publicacoes/outras-publicacoes/item/7694-calendario-agricola-plantio-e-colheita>.

DEERING, Donald Wayne. Rangeland reflectance characteristics measured by aircraft and spacecraft sensors. **Ph. D Dissertation. Texas A&M University**, 1978.

FORMAGGIO, Antonio Roberto, Ieda Del'Arco Sanches | Sensoriamento Remoto em agricultura | São Paulo: Oficina de Textos, 2017. ISBN 978-85-7975-277-3.

GAO, Bo-Cai. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, v. 58, n. 3, p. 257-266, 1996.

IBGE, Instituto Brasileiro de Geografia e Estatística. Produção Agrícola Municipal - PAM. Disponível em: <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, Disponível em: <https://www.iba.org/publicacoes/relatorios>.

JIANG, Zhangyan et al. Development of a two-band enhanced vegetation index without a blue band. **Remote sensing of Environment**, v. 112, n. 10, p. 3833-3845, 2008.

KURSA, Miron B. et al. Feature selection with the Boruta package. **J Stat Softw**, v. 36, n. 11, p. 1-13, 2010.

KING, L. A., Adusei, B., Stehman, S. V., Potapov, P. V., Song, X. P., Krylov, A., Di Bella, C., Loveland, T. R., Johnson, D. M., & Hansen, M. C. (2017). A multi-resolution approach to national-scale cultivated area estimation of soybean. **Remote Sensing of Environment**, 195, 13–29. https://doi.org/10.1016/j.rse.2017.03.047.

LSPA. Levantamento Sistemático da Produção Agrícola. 2019. 75p. ISSN: 2318-6852. Available at: <https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistematico-da-producao-agricola.html?=&t=o-que-e>.

NAGLER, Pamela L. et al. Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. **Remote Sensing of Environment**, v. 87, n. 2-3, p. 310-325, 2003.

PEVS, Produção da Extração Vegetal e da Silvicultura, PEVS. Disponível em: <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.

RIZZI, Rodrigo et al. Estimativa da área de soja no Mato Grosso por meio de imagens MODIS. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 14, p. 387-394, 2009.

SONG, X.-P., HANSEN, M.C., POTAPOV, P.V., ADUSEI, B., PICKERING, J., ADAMI, M., LIMA, A., ZALLES, V., STHEHMAN, S.V., DI BELLA, C.M., CECILIA, C.M., COPATI, E.J., FERNANDES, L.B., HERNANDEZ-SERNA, A., JANTZ, S.M., PICKENS, A.H., TURUBANOVA, S., & TYUKAVINA, A. (In review). Massive soybean expansion in South America since 2000 and implications for conservation. **Nature Sustainability.**

RUDORFF, Bernardo Friedrich Theodor et al. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. **Remote sensing**, v. 2, n. 4, p. 1057-1076, 2010.

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

UNIVERSITY OF MARYLAND. Global Land Analysis Discovery - Commodity Crop Mapping and Monitoring in South America. 2000-2019. Avaiable at:< https://glad.umd.edu/projects/commodity-crop-mapping-and-monitoring-south-america>.