

AN ANALYSIS OF THE INFLUENCE OF THE NUMBER OF OBSERVATIONS IN A RANDOM FOREST TIME SERIES CLASSIFICATION TO MAP THE FOREST AND DEFORESTATION IN THE BRAZILIAN AMAZON

L. de S. Vieira^{a,*}, G. R. Queiroz^a, E. H. Shiguemori^b

^a Earth Observation and Geoinformatics Division, National Institute for Space Research, INPE,
São José dos Campos 12227-010, Brazil - (leonardo.vieira, gilberto.queiroz)@inpe.br

^b Surveillance and Reconnaissance Division, Institute for Advanced Studies, IEAv,
São José dos Campos 12228-001, Brazil - elcio@ieav.cta.br

Commission III, WG III/7

KEY WORDS: Classification, Random Forest, Landsat, Time series, Land cover, Brazilian Amazon

ABSTRACT:

Remote sensing has been an essential tool in combating deforestation. However, the ever-rising deforestation rates require new remote sensing techniques. This paper presents a study to determine the effects on the accuracy of the data analysis of varying the number of satellite observations, using a Random Forest classification algorithm. We carried out experiments on the Landsat-8 data cube with 22 images and developed an automatic sampling system based on PRODES to generate the labeled time series. We split the time series dataset to build data subsets with different number of observations. The results showed that a fewer number of observations negatively effects the accuracy of the RF algorithm when analyzing deforested areas, but not forest areas. The RF classifiers were compared using a random test data set, where all classifiers presented an Overall Accuracy (OA), Balanced Accuracy (BA), and f1-score (F1) above 97%. In the first evaluation, the variation in the number of observations appears to cause little influence on the classification accuracy. The analysis used the reference map to contrast the RF classifier's results. The results showed that the best results in OA occurred with fewer observations. The best performance of 96% happened with four observations. We evaluated the performance of the classes, deforestation, and forest individually. The results showed that a fewer number of observations had negative effects on the accuracy of the RF algorithm when analyzing deforested areas, but not forest areas. Finally, we evaluated the visual quality of the land cover maps produced.

1. INTRODUCTION

The Amazon Forest is the largest tropical rainforest in the world, which plays a leading role in the Earth's climate. However, changes in the forest structure threaten its role as a carbon sink (Yang et al., 2018). Deforestation and degradation are the principal causes of these changes (Matricardi et al., 2020). Therefore, the development of mechanisms to curb deforestation is essential to maintain the stability of the Amazon ecosystem.

About 60 percent of the Amazon biome lies within Brazilian territory, with an area of 5,015,067.75 km² (IBGE, 2020). The most effective way of monitoring deforestation such immense forested areas has been through remote sensing. Forest monitoring using remote sensing discourages deforestation and forest degradation as shown in Brazil by the PRODES during the first decade of the 21st century (Arima et al., 2014).

Despite the success of the Brazilian forest monitoring program, deforestation in the Amazon Forest is still a concern of authorities (Escobar, 2020). The dynamics of deforestation are complex and change over time. Thus, the monitoring methods must evolve to maintain the effectiveness of information produced. A faster mapping of changes is required. An agile approach requires automated techniques to process massive Earth observation data promptly.

Luckily, the open data policies changed the guidelines for the

distribution and availability of satellite images, which resulted in an abundance of free of charge Earth observation data (Showstack, 2014, Sá and Grieco, 2016, Zhu et al., 2019). The open access to satellite data benefits the development of remote sensing methods that exploit temporal information, like time series analysis and classification (Kuenzer et al., 2015). A wealth of free data, added to the advance in hardware and software has led to automatic forest mapping based on spatio-temporal information.

Recent studies have explored the use of spatio-temporal, and time series information to monitor the Amazon Forest (Wang et al., 2019, Fortin et al., 2020, Maretto et al., 2020). In these papers, the time series approach is based on at least a year of satellite observations, resulting in time series that show annual and seasonal changes in the target area. This kind of approach is important when the study pattern is defined as time-series changes over time, such as is used in crop mapping. The Amazon Forest has a dense vegetation structure resulting in a land with high green density. Vegetation indices are designed to accentuate vegetation properties. In general, these indices present higher values in land covered by Amazon Forest when compared with others types of cover. Thus, a time series forest pattern in Amazon can be represented by a homogeneous time series, which does not change over time. A break in a forest time-series pattern may be a sign of deforestation. After the break, the deforestation pattern changes rapidly because, usually, the region is then used for agricultural activity. Thus, there is little time to define a time series pattern before the pattern changes. Therefore, we consider that it is possible to character-

* Corresponding author.

ize the forest time series as a homogeneous pattern with high vegetation indices values and the deforestation time series as a brief deforestation pattern with low vegetation indices values. Therefore, we want to verify if we can generate an accurate map with a minimum number of observations as possible to produce constant information benefiting from the time component of time series classification.

This paper presents the impact of the number of observations used in the time series to create classifiers based on machine learning algorithms. In this study, we analyze the impact on the perspective of the Random Forest (RF) algorithm. We generated distinct RF classifiers from the same training sample with a varied length of time series to compare their, using standard metrics. The main objective was to evaluate the quality of these classifiers and to verify the possibility of using samples with fewer observations in a time series classification to obtain more granulated information.

2. MATERIALS AND METHODS

2.1 A Random Forest Based Algorithm

RF is a supervised learning approach, where an independent set of decision trees is built, based on a random training sample subset, where each tree output is used in an ensemble process to define the value or classification label (Breiman, 2001). RF can be considered a state-of-the-art algorithm in remote sensing. And It has been applied with success in land cover and land use time series classification (Zhu and Woodcock, 2014, Wang et al., 2019, Pelletier et al., 2019, Fortin et al., 2020). These studies were based on at least a year of satellite data to map a region, resulting in a large of number of observations over time. This means each sample has the number of steps equal to the number of images in the period. Each time-series step results in one feature in a machine learning classifier. So, to predict new data, the same number of observations must be used as in the classifier training. In a scenario where the objective is to produce fast and constant information, a classification approach that uses many images is too slow to be effective, because the latency to deduce new information is directly linked to the number of new observations necessary to provide the classifier. For example, if the classifier needs 10 images to generate a new classification and the temporal resolution of the images is 16 days, the latency to produce new information will be at least 160 days not counting the processing time. Therefore, testing a state-of-the-art classification algorithm like RF is the first step in to developing an approach that uses fewer time steps in time series classification to produce information with lower latency and, evaluate its accuracy.

2.2 Region of Interest

The Region of Interest (ROI) crosses five indigenous reserves and two conservation units. Intensive agriculture surrounds these areas, making them a focus of attention for government authorities concerned with deforestation, as shown in Figure 1. The location of the ROI is in the Brazilian state of Rondônia within the Amazon Forest, between 10°58'5''S and 9°56'50''S (latitude), and 65°16'31''W and 63°40'49''W (longitude); it covers an area of $\approx 18463 km^2$. The indigenous and reserve land areas combined cover $\approx 6871 km^2$, corresponding to approximately 0.37% of the ROI.

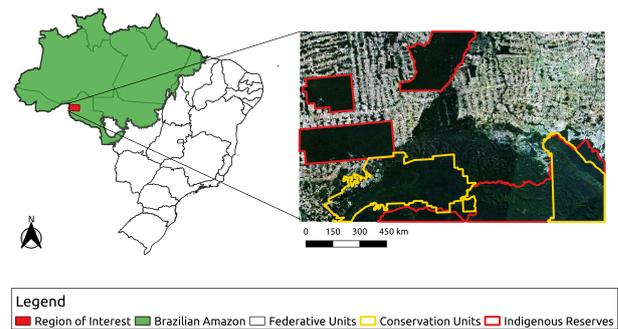


Figure 1. Left: Region of Interest. Right: Sentinel-2 image from Brazil Data Cube on 07-28-2018, indigenous reserves, and conservation units of the ROI.

2.3 Satellite Data

The study was carried out by using the Landsat-8/OLI image data cube available as open data in the Brazil Data Cube (BDC) platform (Ferreira et al., 2020). Data cubes are multidimensional arrays, using a regular grid of spatial coordinates and a temporal dimension made of a sequence of time intervals. Data cubes are built using analysis-ready data, which are images that have been processed according to a common set of reference standards¹. The *LC8_30_16D_STK-1* data cube had 30m of spatial resolution and a temporal composition of 16-days as the result of the selection of the pixels with less cloud interference within this period. This data cube can be retrieved using the BDC Data Cube Explorer or by using the BDC STAC service².

2.4 Reference data

PRODES is a project maintained by Brazil's National Institute for Space Research (INPE), that produces information about the annual rate of deforestation in the Brazilian Amazon. The PRODES data is respected worldwide given its high assertiveness rate of accuracy in mapping, with around 95% of accuracy (INPE, 2021). Thus, the training and validation samples used in this study were based on the datasets available in PRODES. Figure 2 shows a 2018 PRODES vector map of the ROI and presents part of the PRODES information. The "no data" class in the image corresponds to previous mapping, like a river, rock formation, or deforestation map made before 2018.

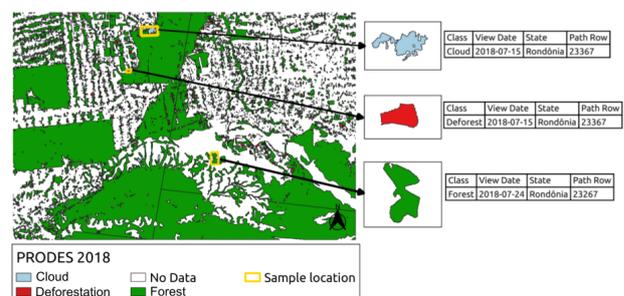


Figure 2. Left: PRODES map of ROI in 2018. Right: Example of polygons and attributes that can be visualized.

¹ <https://ceos.org/ard/>

² <https://brazildatacube.dpi.inpe.br/portal/>

2.5 Defining the Time Series Period Based on PRODES

PRODES is based on a visual analysis of a remote sensing image, in which specialists identify clear-cut deforestation. It can identify deforestation within an area greater than 6.25 ha. The system maps areas of deforestation in primary forests incrementally. Once deforestation is identified in a region, that location is not assessed in mappings in upcoming years. The period of mapping begins on August first of the current year to July 31st of the subsequent year (INPE, 2020). The “view date” of the polygon is the date of the image (best image of PRODES year) used to map the region. The polygons over ROI generally have a “view date” during the Amazon dry season at the end of the PRODES year. These considerations informed our sampling approach.

Our main goal was to select the most homogeneous sample pattern within the time-series period, where the land cover samples did not change over time. For example, if a coordinate represented a forest cover in the first image, the coordinate should also represent forest during all subsequent periods of observation. In addition, we wanted to define a time series pattern where PRODES already confirmed deforestation change to avoid a time series pattern where the sample still contained forest. Thus, to specify the range of satellite data, we used the “view date” information to create our sampling approach.

The *deforestation* “view date” does not represent when the forest was cut, since the deforestation event may have occurred days, weeks, or months before mapping. Thus, it is not possible to determine the ground truth before the deforestation “view date”. After the “view date”, the location is no longer considered a primary forest. The area within the polygon can regenerate or be used in farming, but this takes some time. Thus, the period after the deforestation “view date” is the safest period to sample deforestation samples. In the case of the *forest* “view date”, however, the polygon represents an area covered by forest up to the forest “view date”. After that, in these polygon regions, a deforestation event can occur. So, it is not possible to deduce the ground truth after the forest “view date”. For *forest* polygon, therefore, the safest period to sample a *forest* time series is before the “view date”. So, the time period that is most likely to produce accurate, or safe samples begins with the last “view date” showing forest cover and ends with the first “view date” indicating deforestation for each ROI. This corresponds to the intersection of the forest and deforestation graphs where the known ground truth data overlaps.

Figure 3 summarizes how we used the “view date” of a polygon to define a safe period to extract a time series. The safe period means the one where we can deduce the ground truth over the time series periods. Figure 3 shows a hypothetical example, using a NDVI time series representation to clarify how we combined the “view date” of *forest* and *deforestation* polygon and defined a time series period. In the example, a couple of days were used, but the strategy can be applied using a longer time interval, which will be presented in Section 2.6.

2.6 The Sampling Approach

The sampling approach selected samples of two classes, *forest* and *deforestation*. The previous notes about PRODES were the basis for the sampling approach, where our objective was to minimize appointment errors in sampling process, define coordinates, define the coordinates, and define a time series period of around one year of observations. Figure 4 summarizes the

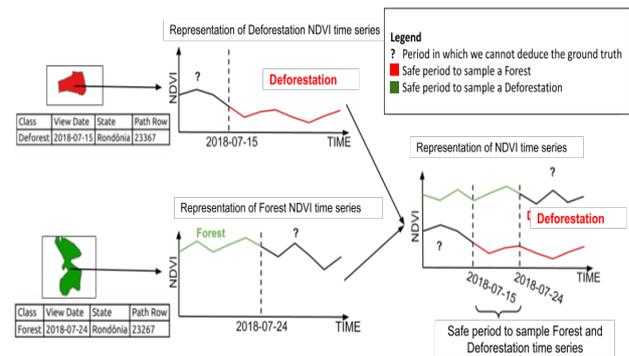


Figure 3. Example of how to define a safe period in the sampling approach.

sampling approach. The boxes with black lettering show the operations applied in the PRODES maps. The boxes in red are the main outputs used in the time series extraction.

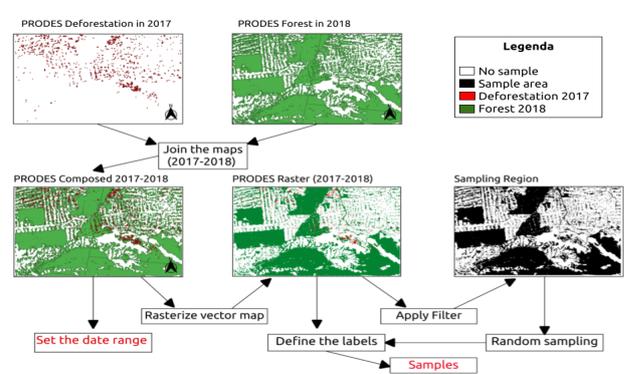


Figure 4. Sampling Approach.

The approach used two consecutive years of the PRODES polygons, (2017-2018), mapped over the ROI to generate a composite map, where the *deforestation* polygons came from the 2017 map and the *forest* polygons were taken from the 2018 map. The data range was defined from the composite map, where the start date of the sample period was the last date among the *deforestation* polygons and the end date was the first of the *forest* polygons.

In order to facilitate the sampling procedure, the PRODES dataset was rasterized with the same coordinate reference system and spatial resolution used in the BDC Landsat-8 data cube (*LC8_30_16D_STK-1*). The raster map made it possible to choose a coordinate by evaluating the neighborhood of the pixel. Then, the points were randomly selected by considering their neighbourhoods to avoid transition areas. Only pixels with all neighbours in a 3×3 grid belonging to the same class were selected. The sampling approach resulted in a time series period starting on July 28, 2017 and ending on July 12, 2018, thus providing about a year of observations. The random sampling resulted in 32,402 samples (18,680 of *forest* and 13,722 of *deforestation*).

2.7 Time Series Extraction

The time series period defined through the sampling approach was comprised of 22 Landsat-8 images in the data cube *LC8_30_16D_STK-1* taken at 16-day interval, for a total time series of

349 days. In this study, only vegetation indices were used due to their variation over a predetermined range. This controlled variation also limited the variety of classification patterns. The development was based on three vegetation indices to define the time series data set: Normalized Difference Vegetation Index (NDVI)(Rouse et al., 1974), Visible Atmospherically Resistant Index (VARI)(Stow et al., 2005), and Enhanced Vegetation Index (EVI)(Jiang et al., 2008). Equations 1, 2, and 3 show how to calculate the indices:

$$NDVI = \frac{NIR + Red}{NIR - Red} \quad (1)$$

$$EVI = 2.5 * \left(\frac{NIR - Red}{NIR + 6 * Red - 7.5 * Blue + 1} \right) \quad (2)$$

$$VARI = \frac{Green - Red}{Green + Red - Blue} \quad (3)$$

The BDC Landsat-8 data cube provided a layer, based on Fmask cloud detection (Qiu et al., 2019). This information was used to fill in the missing values with a linear interpolation of the time-series.

Only the red, blue, and green bands, along with the EVI and NDVI spectral indices, were extracted from the BDC data cube. The first three bands were used to compute the VARI index. The feature set of the samples was composed from the NDVI, VARI, and EVI time series. Thus, considering a time series length with 22 images at regular intervals, called time steps, each sample had a total of 66 features.

2.8 Generating the RF classifiers

The training strategy used the 22 LANDSAT images, or observations, in the cube to define the number of time steps. From the first observation, the same training and test data set was split to generate data subsets with different time steps. Figure 5 summarizes the strategy used to generate and calibrate the RF classifiers based on the number of observations time steps included. The first observation was the start point of all datasets, so that all classifiers benefitted from the more cloud-free images that occurred at the beginning of the observation period, during the Amazon's dry season. Then, using a cumulative approach, new data subsets were created that contained from 2 to all 22 time steps, by increasing the number of time steps included, one by one. The first data subset had 2 time steps, t1 and t2. The next was composed of three time steps, t1, t2, t3 and so on, until all 22 time steps were included in the dataset. Each of these test runs generated one RF classifier.

To select the hyper-parameters for each classifier, a random search was applied to a list of predefined parameters. The range of parameter settings tested is presented below.

- Number of Estimators (NE): [200; 400; 600; 800; 1,000; 1,200; 1,400; 1,600; 1,800; 2,000]
- Maximum depth (MD): [10; 20; 30; 40; 50; 60; 70; 80; 90; 100; 110; None]
- Minimum of split samples (MS): [2; 5; 10]

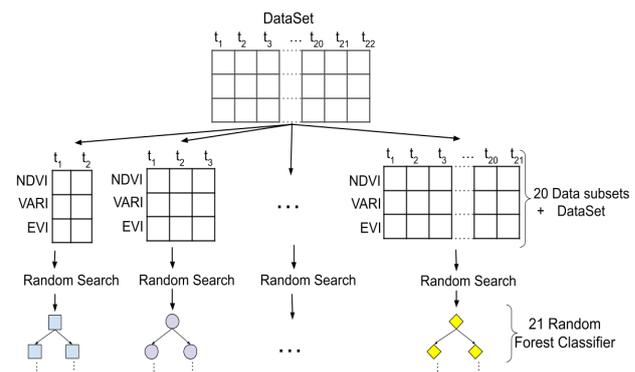


Figure 5. Strategy to split the dataset based on the number of observations included to generate RF classifiers.

- Minimum of leaf (ML): [1; 2; 4]
- Bootstrap (BS): [True; False]

For each classifier, we tested 30 random parameter settings, where each classifier was optimized by 3-fold cross-validation. The best training result for each sub-dataset was selected from the resulting classifiers. We adopted the random strategy to generate and calibrate the classifiers to avoid benefitting one subset over another and to maintain impartiality in terms of the number of time-steps.

2.9 Validation and accuracy metrics

The RF classifiers were trained using a random search over a list of predetermined parameters. We split the dataset obtained in the sampling approach into 70% training and 30% testing samples, using the training dataset in the random search and the test data to evaluate the resulting classifiers. Each set of random parameters generated an RF classifier and then fit and score them based on k -fold cross-validation. Cross-validation is a statistical method used to estimate the performance of a machine learning method with new data (Refaeilzadeh et al., 2009). The k -fold cross-validation technique is a computational technique that trains the classifier k times, using a $1/k$ fraction of the data set as test material and the rest as training material. This fraction varies in each interaction testing all data (Rodriguez et al., 2010). Using this approach, we expected a lower variance between each random forest classifier that was generated from the different data sub-sets with a distinct number of time steps in each.

We used the first metric to evaluate the classifiers globally, using distinct accuracy metrics: Overall Accuracy (OA), Balanced Accuracy (BA), and, F1-score (F1) (Fatourechhi et al., 2008). In addition to this review, we evaluated all classifiers, using the PRODES map for the ground truth. We classified the ROI to apply all metrics to the composite map created in the sampling approach. We evaluated the accuracy of the approach by comparing the *forest* and *deforestation* classes to PRODES, using: User's Accuracy (UA), and Producer's Accuracy (PA) (Congalton, 1991).

3. RESULTS

3.1 RF Training, calibration, and test classifiers

The training approach generated 21 RF classifiers with an increasing number of observations. All classifiers were generated

from the same source, but for each RF classifier, a different number of observations in the time series was used. Table 1 shows the best parameters found in the random search and validation metrics for the test samples.

Ts	NE	MF	MD	MS	ML	BS	OA	BA	F1
2	2000	auto	20	2	1	TRUE	0.973	0.973	0.973
3	600	auto	70	2	1	TRUE	0.977	0.977	0.977
4	1000	auto	40	2	2	FALSE	0.979	0.979	0.979
5	800	sqrt	20	2	1	FALSE	0.981	0.981	0.980
6	800	sqrt	20	2	1	FALSE	0.982	0.982	0.982
7	800	sqrt	20	2	1	FALSE	0.981	0.982	0.981
8	800	sqrt	20	2	1	FALSE	0.981	0.982	0.981
9	800	sqrt	20	2	1	FALSE	0.981	0.981	0.981
10	800	sqrt	20	2	1	FALSE	0.981	0.981	0.981
11	800	sqrt	20	2	1	FALSE	0.981	0.982	0.981
12	800	sqrt	20	2	1	FALSE	0.982	0.982	0.981
13	800	sqrt	20	2	1	FALSE	0.982	0.982	0.981
14	800	sqrt	20	2	1	FALSE	0.982	0.982	0.981
15	800	sqrt	20	2	1	FALSE	0.982	0.982	0.981
16	800	sqrt	20	2	1	FALSE	0.982	0.982	0.982
17	800	sqrt	20	2	1	FALSE	0.982	0.982	0.981
18	800	sqrt	20	2	1	FALSE	0.983	0.983	0.983
19	800	sqrt	20	2	1	FALSE	0.984	0.984	0.983
20	800	sqrt	20	2	1	FALSE	0.984	0.985	0.984
21	1000	auto	40	2	2	FALSE	0.984	0.984	0.984
22	800	sqrt	20	2	1	FALSE	0.985	0.985	0.985

Table 1. Random Forest parameter settings and accuracy metrics. Table header: Time series steps (Ts), Number of Estimators (NE), Maximum depth (MD), Minimum of samples split (MS), Minimum of samples leaf (ML), Bootstrap (BS), Overall Accuracy (OA), Balanced Accuracy (BA), and F1-Score (F1)

The calibration of the parameter settings resulted in classifiers with high indices for all metrics. The results showed that the algorithm performed well irrespective of the number of time steps included in the sample. To visualize differences among the results, it is necessary to display the values with three decimal places to highlight the similarity of the accuracy rates for all numbers of observations in the time series. There was an improvement when samples included more time steps, but the improvement value was only around 0.012 from the smaller size to the larger size tested. This value is not enough to affirm that the classifier with fewer observations would generate the worst results in an all-area classification. Therefore, this result demonstrates that a good selection in parameter settings generated accurate classifiers for all numbers of time steps. This would suggest that, at least for the training data set, the number of time steps used in the data subset did not interfere drastically with the test sample rate of accuracy.

3.2 Comparison with PRODES reference data: a global analysis

The 21 RF classifiers were used to produce a map of all the ROI. The next comparison used these maps against the PRODES map used in the sampling approach, disregarding the "no data" region in the map. We used metrics to compare each produced map globally, using the PRODES map as ground truth. Figure 6 shows a chart with OA, BA, and F1 for all the numbers of observations tested.

The OA initially presented a crescent behavior for the first 4 observations totals, reaching the highest value of 0.963. After that, the overall accuracy declined, until achieving the lowest value (0.924) for 21 total observations. In the BA, there was an

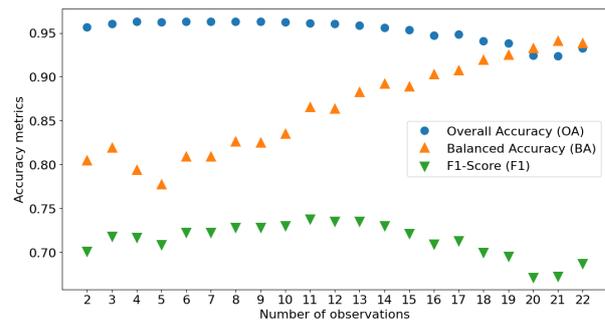


Figure 6. OA, BA, and F1 vs Total Number of Observations.

increase in metric value from 2 to 3 total observations, followed by a drop over the next two steps. After that, the BA presented a crescent behavior for the next 16 observations totals, reaching its highest value at 0.94 for 21 total observations. In the F1-Score, the best result occurred with 11 observations, scoring around 0.74; the worst F1-score happened with 20 total observations, with a score of about 0.67.

The OA shows that, increasing the number of observations does not result in better classifier performance. But the OA does not consider the imbalance between the classes, which is caused by the fact that the forest area is much greater than the deforested area in the ROI. Therefore, to analyze the performance regarding the class imbalance, we analyzed the BA and F1 metrics. The BA showed a significant improvement when increasing the total number of observations, indicating that the *deforestation* class score can benefit from that increase. On the other hand, the F1 value did not exhibit the same behavior; while the *deforestation* class improved with the number of observations, the *forest* class performed worse. The progressively worse performance for the forest class as the number of observations increased occurred mainly because the classifier became more sensitive to the edge regions, as shown in the next section 3.4. Therefore, the question of which is the better classification will depend on which class is the priority in classification rate accuracy.

3.3 Compare with PRODES reference data: analysis per class

To analyse the behaviour of the accuracy per class, we calculated user and producer accuracy for *forest*, and *deforestation* classes. Figure 7 presents the results obtained from the metrics against the total number of observations for the total *forest* class.

The *forest* PA was approximately 99% for all numbers of observations tested; so, it may be considered invariant for the number of observations. The worst percentage of UA occurred with 21 total observations at around 92%, and the best result occurred with 5 observations at around 97%. As for the UA, there was a significant decrease in accuracy when the number of observations increased. In short, the accuracy for the *forest* class was high for all numbers of observations, but the increase in observation totals worsened accuracy.

Figure 8 displays the chart of UA, and PA against the number of observations obtained for the deforestation class. For the *deforestation* PA, the increase of observations did not appear to influence the PA very much, with the PA's best result at around

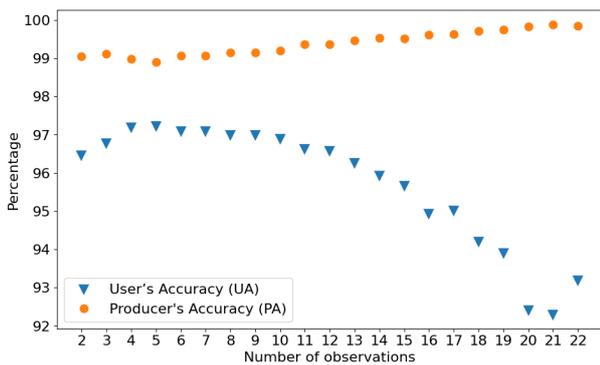


Figure 7. *forest* User's accuracy, Producer's accuracy, Errors of commission and Errors of omission vs Number of Observations.

37%. But, in the case of the UA, there was considerable improvement in the accuracy as the total number of observations increased, with about a 30% improvement between worst and best results.

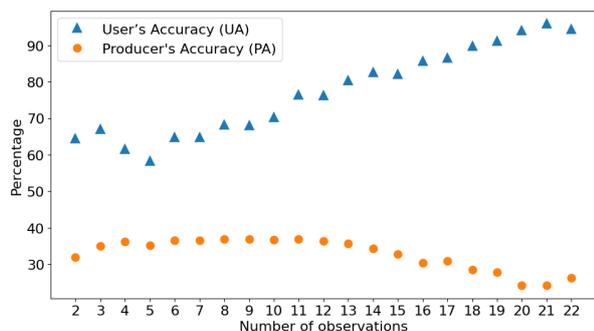


Figure 8. *Deforestation* User's accuracy, Producer's accuracy, Errors of commission and Errors of omission vs Number of Observations.

The results for the *deforestation* class were different from those obtained for the *forest* class. While the increase in the number of observations decreased the accuracy of the classifier for the *forest* class, there was a UA increase for the *deforestation* class.

3.4 Comparison with PRODES reference data: Visual analysis

The last comparison was a visual analysis of the ROI classifications, using PRODES as a reference and a true-color image at the beginning of the time series period. Figure 9 shows a small ROI area inside the classified area to exemplify the visual analysis. All the considerations for the ROI were observed in all the classified areas. Figure 9 presents the boundaries of the small ROI, the PRODES ground truth reference, ROI image at the beginning of the period, and ROI in the RF classifications that show the number of observations used in the classification.

Note that the *forest* pattern is more homogeneous for all forest areas identified by PRODES. In the *deforestation* class, the deforestation pattern can have different characteristics, depending on the time the deforestation event occurred, with three patterns for the *deforestation* class: recent, ancient, and in-progress deforestation.

The forest areas presented a relatively homogeneous rigid green inside the boundaries. But there were some points in the forest, where some small areas showed a different pattern. Some isolated regions in forest areas differ from their surroundings.

Where there had been recent deforestation, the exposed soil showed up sharply on the true-color image, which presented a sharp red color in the deforestation limits. But, we can also see remnants of forest areas inside the boundaries of deforestation. The ancient deforestation can be identified easily in visual analysis, but the red tone did not appear in the true-color image. We visualize a soft shade of green in this kind of deforestation, which contrasts with the rigid green of its forest surroundings. In the in-progress deforestation, the boundaries of deforestation were not visible. Along with the recent deforestation, there was also forest formation inside the deforestation boundaries.

The classifications show that there was an improvement in the *deforestation* class, when the number of observations increased, according to the results of Section 3.2. The greatest improvement was obtained in the recent deforestation and deforestation in-progress, where the classifiers with fewer observations did not detect the class changes, as seen in d2 and d3 examples in Figure 9. Now, for the ancient deforestation, all classifications obtained good results, as shown in d1. All classifications presented boundary errors, but increasing the number of observations increased errors substantially regard less of the number of observations (see d4 in 9). An error in the edges may have caused this result. as noted in Sections 3.2, and 3.3.

In general, the results obtained were satisfactory for all numbers of observations. There was an improvement in the *deforestation* class when the number of observations increased. In contrast, the same increase of time steps in time series samples resulted in a deterioration in the accuracy of the *forest* class scores. Therefore, we can consider that the best result depends on the objective of the classification. If we want a conservative classification for the forest areas, we can use fewer observations to classify the region. But, if we want a classification where the priority highlights the changes not dealing with the *forest* errors, we can use more observations.

3.5 Improving the methodology and future work

The analysis of the results suggests some directions to improve the methodology. We can improve the sampling approach, given that the quality of machine learning classifiers depends directly on the samples presented to the algorithm. One way is to avoid appointment errors in boundary areas. In this regard, we can improve the filter by using a larger grid to define the sampling region. Another possibility would be to use an edge detector, such as Canny (Canny, 1986) or Sobel (Sobel and Feldman, 1968) to identify the transition areas and avoid the sample selecting the areas pointed out by the detector edges. Another way to improve the sampling approach would be to avoid appointment to pixels, where the pattern inside the PRODES polygon can be different from the general pattern as shown in Section 3.4. To do that, we could use an NDVI image at the beginning of the period to define a threshold to identify a vegetation pixel inside the *deforestation* polygon.

For our next steps we intend to evaluate how the classification using fewer observations behaves over time, producing consecutive classifications. For example, if we have 24 satellite observations in a year, we want to classify all the period considering

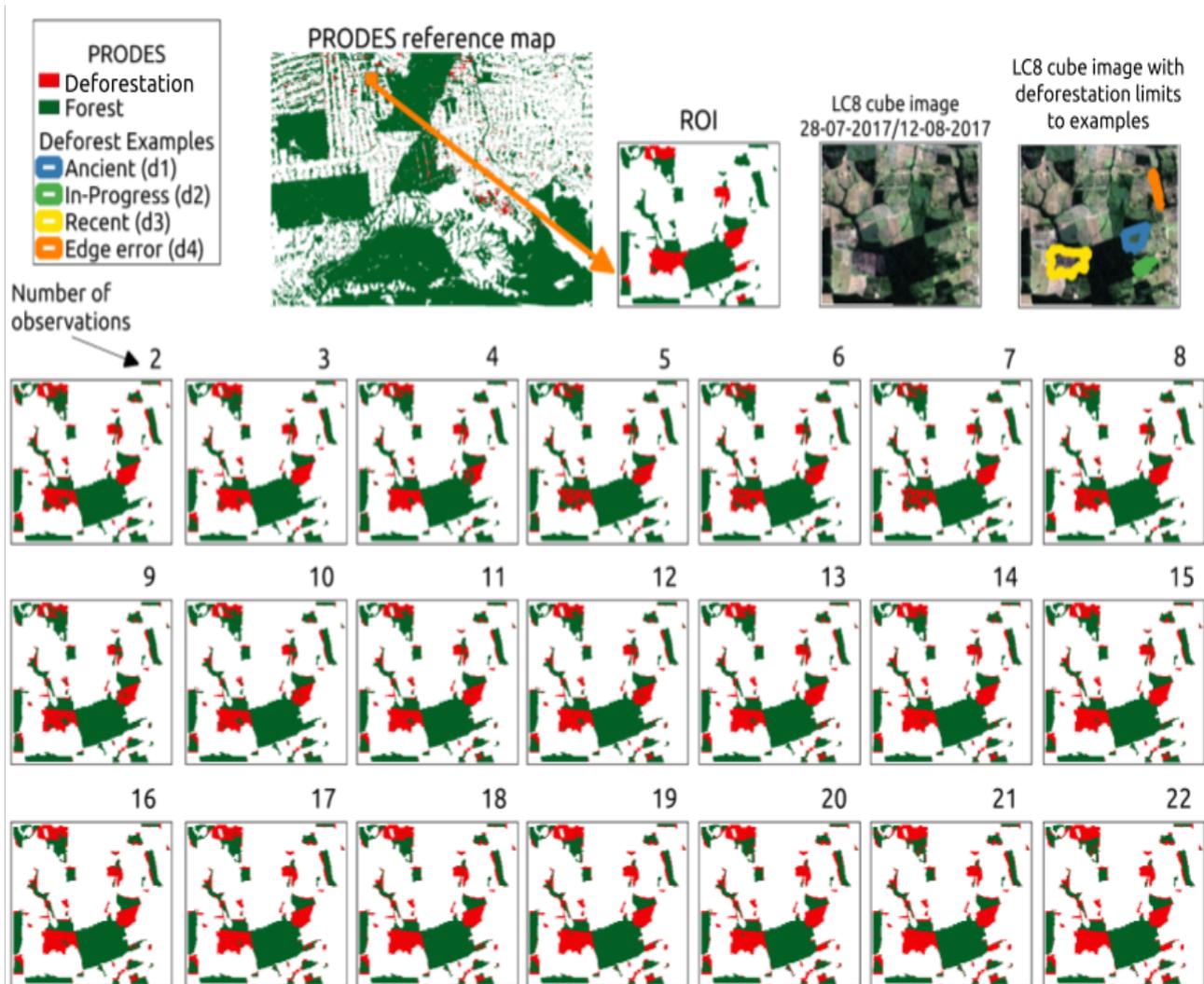


Figure 9. Visual analysis over a ROI in the all the mapped region.

a specific number of observations, like 3, producing 8 classifications in a year. Sequentially will combine multiple classifications, like 8 classifications, and compare them with one classification using all 24 observations.

4. CONCLUSION

This paper analyzes the behavior of RF classifiers when varying the number of observations in land cover mapping through accuracy metrics and visual analysis of a distinct time-series data set. The data set was composed of the same appointments. The difference between them is the number of observations used and consequently the number of features of the samples.

The results obtained show little variation in overall accuracy against the number of observations. When comparing the classified maps with the reference map, overall accuracy was greater than 94% for all RF classifications. But analyzing the accuracy and errors for the classes individually, it is possible to observe differences between the classifications. The increase in the number of observations shows improved accuracy in the *deforestation* class and worse accuracy in the Forest class performance. Therefore, the choice of which classifier is the most suitable will depend on the user's priority class in mapping.

The proposed analysis tested the impact on the performance of the RF against the number of observations to consider the use fewer observations in *deforestation* mapping. In this context, the results show a relevant impact in performance in terms of *deforestation* detection with an increase in the number of observations. But considering a scenario where the focus is producing information quickly, the results demonstrated that it is possible to use fewer observations to gain valuable time in forest conservation.

ACKNOWLEDGEMENTS

Acknowledgements of support for the Environmental Monitoring of Brazilian Biomes project (*Brazil Data Cube*), funded by the Amazon Fund through the financial collaboration of the Brazilian Development Bank (BNDES), the Foundation for Science, Technology and Space Applications (FUNCATE), Process 17.2.0536.1, and the "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" - Brasil (CAPES) - Finance Code 001.

REFERENCES

- Arima, E. Y., Barreto, P., Araújo, E., Soares-Filho, B., 2014. Public policies can reduce tropical deforestation: Lessons and challenges from Brazil. *Land use policy*, 41, 465–473.
- Breiman, L., 2001. Random Forest, vol. 45. *Mach Learn*, 1.
- Canny, J., 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698.
- Congalton, R. G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1), 35–46.
- Escobar, H., 2020. Deforestation in the Brazilian Amazon is still rising sharply. *Science*, 369(6504), 613-613.
- Fatourech, M., Ward, R. K., Mason, S. G., Huggins, J., Schloegl, A., Birch, G. E., 2008. Comparison of evaluation metrics in classification applications with imbalanced datasets. *2008 seventh international conference on machine learning and applications*, IEEE, 777–782.
- Ferreira, K. R., Queiroz, G. R., Vinhas, L., Marujo, R. F. B., Simoes, R. E. O., Picoli, M. C. A., Camara, G., Cartaxo, R., Gomes, V. C. F., Santos, L. A., Sanchez, A. H., Arcanjo, J. S., Fronza, J. G., Noronha, C. A., Costa, R. W., Zaglia, M. C., Zioti, F., Korting, T. S., Soares, A. R., Chaves, M. E. D., Fonseca, L. M. G., 2020. Earth Observation Data Cubes for Brazil: Requirements, Methodology and Products. *Remote Sensing*, 12(24). <https://www.mdpi.com/2072-4292/12/24/4033>.
- Fortin, J. A., Cardille, J. A., Perez, E., 2020. Multi-sensor detection of forest-cover change across 45 years in Mato Grosso, Brazil. *Remote Sensing of Environment*, 238, 111266. Time Series Analysis with High Spatial Resolution Imagery.
- IBGE, 2020. Amazônia legal brasileira.
- INPE, 2020. Metodologia utilizada nos projetos prodes e deter.
- INPE, 2021. Monitoramento do desmatamento da floresta amazônica brasileira por satélite.
- 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.
- Kuenzer, C., Dech, S., Wagner, W., 2015. Remote sensing time series revealing land surface dynamics: Status quo and the path-way ahead. *Remote Sensing Time Series*, Springer, 1–24.
- Maretto, R. V., Fonseca, L. M. G., Jacobs, N., Korting, T. S., Bendini, H. N., Parente, L. L., 2020. Spatio-Temporal Deep Learning Approach to Map Deforestation in Amazon Rainforest. *IEEE Geoscience and Remote Sensing Letters*, 1-5.
- Matricardi, E. A. T., Skole, D. L., Costa, O. B., Pedlowski, M. A., Samek, J. H., Miguel, E. P., 2020. Long-term forest degradation surpasses deforestation in the Brazilian Amazon. *Science*, 369(6509), 1378-1382.
- Pelletier, C., Webb, G. I., Petitjean, F., 2019. Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing*, 11(5).
- Qiu, S., Lin, Y., Shang, R., Zhang, J., Ma, L., Zhu, Z., 2019. Making Landsat time series consistent: evaluating and improving Landsat analysis ready data. *Remote Sensing*, 11(1), 51.
- Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-validation. *Encyclopedia of database systems*, 5, 532–538.
- Rodriguez, J. D., Perez, A., Lozano, J. A., 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3), 569-575.
- Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W. et al., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351(1974), 309.
- Sá, C., Grieco, J., 2016. Open data for science, policy, and the public good. *Review of Policy Research*, 33(5), 526–543.
- Showstack, R., 2014. Sentinel Satellites Initiate New Era in Earth Observation. *Eos, Transactions American Geophysical Union*, 95(26), 239-240.
- Sobel, I., Feldman, G., 1968. A 3x3 isotropic gradient operator for image processing. R. Duda, P. Hart (eds), *Pattern classification and scene analysis*, John Wiley, New York, 271–272.
- Stow, D., Niphadkar, M., Kaiser, J., 2005. MODIS-derived visible atmospherically resistant index for monitoring chaparral moisture content. *International Journal of Remote Sensing*, 26(17), 3867-3873.
- Wang, Y., Ziv, G., Adami, M., Mitchard, E., Batterman, S. A., Buermann, W., Marimon, B. S., Junior, B. H. M., Reis, S. M., Rodrigues, D., Galbraith, D., 2019. Mapping tropical disturbed forests using multi-decadal 30m optical satellite imagery. *Remote Sensing of Environment*, 221, 474 - 488.
- Yang, Y., Saatchi, S. S., Xu, L., Yu, Y., Choi, S., Phillips, N., Kennedy, R., Keller, M., Knyazikhin, Y., Myneni, R. B., 2018. Post-drought decline of the Amazon carbon sink. *Nature Communications*, 9(1), 1–9.
- Zhu, Z., Woodcock, C. E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144, 152-171.
- Zhu, Z., Wulder, M. A., Roy, D. P., Woodcock, C. E., Hansen, M. C., Radeloff, V. C., Healey, S. P., Schaaf, C., Hostert, P., Strobl, P., Pekel, J.-F., Lyburner, L., Pahlevan, N., Scambos, T. A., 2019. Benefits of the free and open Landsat data policy. *Remote Sensing of Environment*, 224, 382-385.