Multi-source remote sensing data improves the classification accuracy of natural forests and eucalyptus plantations

Dados de sensoriamento remoto de múltiplas fontes eleva a exatidão de classificação de florestas naturais e plantadas

Gustavo Fluminense Carneiro [1], Matheus Pinheiro Ferreira [2] e Carlos Frederico de Sá Volotão [3]

1 Instituto Militar de Engenharia, Engenharia Cartográfica, Rio de Janeiro, Brasil. gustavocarneiro@ime.eb.br.
  ORCID: https://orcid.org/0000-0001-5595-5785
2 Instituto Militar de Engenharia, Engenharia Cartográfica, Rio de Janeiro, Brasil. matheus@ime.eb.br.
  ORCID: https://orcid.org/0000-0003-0687-4422
3 Instituto Militar de Engenharia, Engenharia Cartográfica, Rio de Janeiro, Brasil. volotao@ime.eb.br.
  ORCID: https://orcid.org/0000-0001-7200-8093

Abstract: It is challenging to map the spatial distribution of natural and planted forests based on satellite images because of the high correlation among them. This investigation aims to increase accuracies in classifications of natural forests and eucalyptus plantations by combining remote sensing data from multiple sources. We defined four vegetation classes: natural forest (NF), planted eucalyptus forest (PF), agriculture (A) and pasture (P), and sampled 410,251 pixels from 100 polygons of each class. Classification experiments were performed by using a random forest algorithm with images from Landsat-8, Sentinel-1, and SRTM. We considered four texture features (energy, contrast, correlation, and entropy) and NDVI. We used F1-score, overall accuracy and total disagreement metrics, to assess the classification performance, and Jeffries–Matusita (JM) distance to measure the spectral separability. Overall accuracy for Landsat-8 bands alone was 88.29%. A combination of Landsat-8 with Sentinel-1 bands resulted in a 3% overall accuracy increase and this band combination also improved the F1-score of NF, PF, P and A in 2.22%, 2.9%, 3.71%, and 8.01%, respectively. The total disagreement decreased from 11.71% to 8.71%. The increase in the statistical separability corroborates such improvement and is mainly observed between NF-PF (11.98%) and A-P (45.12%). We conclude that combining optical and radar remote sensing data increased the classification accuracy of natural and planted forests and may serve as a basis for large-scale semi-automatic mapping of forest resources.
Keywords: Random Forest. Landsat-8. Sentinel-1.

Resumo: É um desafio mapear a distribuição espacial de florestas naturais e plantadas com base em imagens de satélite, devido à alta correlação entre elas. Esta investigação visa aumentar a precisão nas classificações de florestas naturais e plantações de eucalipto, combinando dados de sensoriamento remoto de várias fontes. Definimos quatro classes de vegetação: floresta natural (NF), floresta de eucalipto plantada (PF), agricultura (A) e pastagem (P), e amostramos 410.251 pixels de 100 polígonos de cada classe. Os experimentos de classificação foram realizados usando um algoritmo florestal aleatório com imagens de Landsat-8, Sentinel-1 e SRTM. Foram consideradas quatro características de textura (energia, contraste, correlação e entropia) e NDVI. Utilizamos o escore F1, a precisão geral e as métricas de desacordo total, para avaliar o desempenho da classificação e a distância de Jeffries – Matusita (JM) para medir a separabilidade espectral. A precisão geral apenas para as bandas Landsat-8 foi de 88,29%. Uma combinação de bandas Landsat-8 e Sentinel-1 resultou em um aumento de precisão geral de 3% e essa combinação de bandas também melhorou o escore F1 de NF, PF, P e A em 2,22%, 2,9%, 3,71% e 8,01% , respectivamente. O desacordo total diminuiu de 11,71% para 8,71%. O aumento da separabilidade estatística corrobora essa melhora e é observado principalmente entre NF-PF (11,98%) e A-P (45,12%). Concluímos que a combinação de dados de sensoriamento remoto óptico e por radar aumentou a precisão da classificação de florestas naturais e plantadas e pode servir de base para o mapeamento semi-automático em larga escala dos recursos florestais.
Palavras-chave: Random Forest. Landsat-8. Sentinel-1.

1   INTRODUCTION

In Brazil, tropical forests cover some 5 mi km², which corresponds to approximately 62% of its territory (MAPBIOMAS, 2019). Tropical forests provide key ecosystem services for mankind such as soil conservation, carbon sequestration and habitat protection. They harbor the largest portion of global tree diversity, with up to 53,000 tree species or more (SLIK et al., 2015). Moreover, tropical forests store half of the global forest carbon (C) stock (PAN et al., 2011) and are responsible to fix about 70% of the terrestrial nitrogen (WANG; HOULTON, 2009). The provisioning of these important ecosystem services relies on the maintenance and conservation of forested areas. However, deforestation and forest depletion have reached alarming levels. In Brazil, for example, approximately 20,000 km² of Amazonian forests were cleared yearly between 2000 and 2017 (QIN et al., 2019), and only 12,40% of the original extent (ca. 1.3m km²) of the Atlantic Forest (AF) remains (SOS/INPE, 2019). The situation is even more critical in the state of Espírito Santo (ES) that lost 90% of its natural forest areas (SOS/INPE, 2019).

The growing demand for forest products, coupled with the need to conserve tropical forests, has contributed to the global extent of planted forests in more than 65% since the 1990s (FAO, 2015). In Brazil, there was an increase of approximately 47 mi ha of planted forests between 1985 and 2017 (MAPBIOMAS, 2019). Currently, the most cultivated genus is Eucalyptus spp., occupying an area of about 5.7 mi ha (IBA, 2017). According to the Brazilian Tree Industry (IBA), the planted tree sector has a significant impact on the economy, representing 6.1% of the industrial gross domestic product and showed in 2017 a surplus of US$ 9.0 billion. Espírito Santo stands out as one of the most important states in planted forests, occupying fifth place in the national scene with a planted area of 233,760 ha. Economically, this industry represents 6.8% of the state's gross industrial product (CNI, 2019).

Traditionally, information regarding tropical forests and forest plantations is obtained with field inventories, which require the establishment of sample plots and involves high costs and complex logistics, especially in regions with difficult access. For example, for the AF only 0.01% of the total number of fragments were sampled (LIMA et al., 2015).

By contrast, remote sensing data can be used to map the distribution of natural and planted forest areas over large spatial extents (JOSHI et al., 2016). A common problem reported in several studies and land cover mapping initiatives is the misclassification rate between forest classes or agricultural areas. For example, the MapBiomas project (MAPBIOMAS, 2019), which maps the Brazilian land cover and use change in an annual basis with Landsat images, reported in 2016 a classification accuracy of 69.21% for tree plantations, and a misclassification rate (omission error) between planted and natural forests of 18.63% Moreover, 9.38% of planted forest areas were classified as mosaics of agriculture and pasture. Furthermore, 1.87 % of commercial forest were labeled as pasture and 0.92% was considered as agriculture. Natural and planted forests exhibit similar spectral behavior particularly in medium resolution (pixel $\geq$ 20 m), which challenges their discrimination using automatic approaches.

Active remote sensing, such as radio detection and ranging (RADAR) or light detection and ranging (LiDAR) have been successfully employed for forest cover mapping (YU et al., 2015; MARTONE et al., 2018; FAGAN et al., 2018). However, the integration between active and passive remote sensing technologies has been poorly investigated. Tree plantations have distinct spectral and structural characteristics that can be explored for their detection.

In this study, we explored the combination of passive optical and RADAR data to discriminate between planted, natural forest, pasture and agricultural areas in the state of Espírito Santo, southeast of Brazil. Specifically, we tested the following hypotheses: (i) The combination of Landsat-8 and Sentinel-1 data improves the classification accuracy of natural and planted forests. (ii) The spectral separability between natural forests and tree plantations increases by combining passive optical and RADAR data.

2   MATERIAL

2.1   Study area

The study area comprises the entire state of Espírito Santo (ES), with an area of 46,074.44 km² (Figure 1). The predominant climate of ES is characterized by a wet tropical summer and a dry winter. Annual precipitation ranges from 1045 to 1375 mm and the average temperature ranges from 17.9ºC to 25ºC, depending on the region (CLIMATE, 2019).

Figure 1 – Study area location.
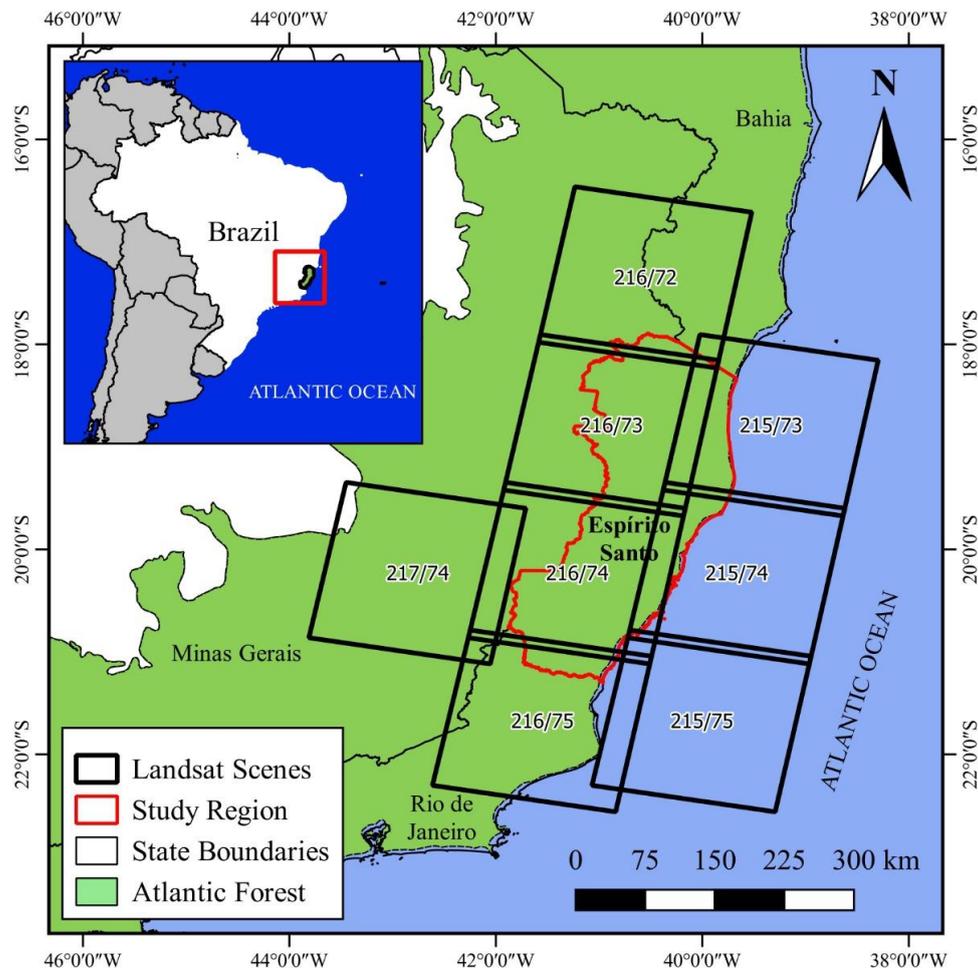
Source: The authors (2020).

According to Garbin et al. (2017), vegetation of ES is divided into six types of phytoecological regions: dense ombrophylous forest, open ombrophilous forest, seasonal semideciduous forest, savannas, new formations and ecological refuge. Dense ombrophilous forests are the state's lushest vegetation and concentrate the largest number of endemic taxa.

The natural forest cover present in Espírito Santo, as in other states of Brazil, experienced forest degradation, mainly due to the exploitation of native trees by timber industries. According to data from the Atlantic Forest Remnants Atlas' Technical Report, developed by the SOS Atlantic Forest Foundation in partnership with the National Institute for Space Research (INPE) for the 2017-2018 period, Espírito Santo has 483,087 ha of forest remnants. This corresponds to 10.5% of the original forest cover extent.

With the industrial growth that occurred in the country especially in the second half of the twentieth century, ES received new industrial centers, most of which deal with oil, steel, pulp and wood (THOMAZ, 2010). Pulp and wood have considerably accelerated deforestation in the region. Between 1985 and 1990 there was a loss of native forest area of 536,480 ha, and from 1990 to 1995 of a loss of 500,317 ha (SOS/INPE,

2019).

According to data from the 2016 Brazilian Forestry and Timber Yearbook (MENDES, 2016), Espírito Santo is among the five largest Brazilian states exporting timber products and among the ten largest in terms of planted forest area. Data from the Brazilian Tree Industry (IBA, 2017) indicate that the area planted with eucalyptus trees in 2016 was 233,760 ha. As a reference to understand the importance of eucalyptus in the economy, it is estimated that the forestry sector in ES is around R$ 5 billion, which corresponds to 25% of the state's agribusiness gross domestic product (CEDAGRO, 2011). Therefore, due to the strong presence of native forest (Atlantic Forest) and planted forest (for commercial exploitation by the logging industry), Espírito Santo state was chosen as the area of study for this research.

2.2   Remote Sensing Data

In this section, we describe the remote sensing and ancillary data used to discriminate among planted forests, natural forests, pasture and agriculture within the study area.

2.2.1   OPTICAL IMAGES

Landsat-8 (USGS, 2019) images acquired by the Operational Land Imager (OLI) sensor available for the year 2017 (January 1st to December 31st) were used. In total, 137 scenes distributed in the following path/row were employed: 215/073, 215/074, 215/075, 216/072, 216/073, 216/074, 216/075, 217/074 (Figure 1). For this study, we used six OLI bands: Blue (0.45-0.51 μm), Green (0.53-0.59 μm), Red (0.64-0.67 μm), Near-infrared (0.85-0.88 μm), Short-wave infrared 1 (1.57-1.65 μm) and Short-wave infrared 2 (2.11-2.29 μm) with a spatial resolution of 30 m. The OLI images were obtained through the Google Earth Engine (GEE) platform. The images employed met the geometric and radiometric quality requirements (Tier 1) and were retrieved in top of atmosphere (TOA) reflectance.

2.2.2   RADAR IMAGES

Sentinel-1 (ESA, 2019) images acquired by the C-SAR sensor available for 2017 (January 1st to December 31st) were employed. In total, 175 images were used, covering the entire study area (Figure 1). These images were obtained on the GEE platform in Interferometric Wide Swath Mode (IW) acquisition, VV and VH polarization, in Ground Range Detected (GRD) with pixel spacing of 10 m.

2.2.3   DIGITAL ELEVATION MODEL

The Shuttle Radar Topography Mission (SRTM) (NASA, 2019). produced a complete and Digital Elevation Model (DEM) from the surface of the Earth. The project was a joint effort of the National Aeronautics and Space Administration (NASA) with the German and Italian space agencies. It flew in February 2000, using dual radar antennas to acquire interferometric radar data. In this research, we used SRTM V3 product (SRTM Plus), provided by NASA at a resolution of 1 arc-second (approximately 30m). Data were accessed through the GEE platform.

2.2.4   ANCILLARY DATA

To help in identifying samples of natural forests, we used the data of the SOS Atlantic Forest Project for the year 2017 (SOS/INPE, 2019). The SOS Atlantic Forest Foundation is a private non-profit organization that aims to map deforestation in the AF Biome. The methodology adopted is based on visual inspections of forest fragments using satellite images.

To assist in obtaining the samples of planted forests, pastures and agriculture we used data from the MapBiomas project. MapBiomas uses machine learning, cloud computing, and Landsat images to annually map 27 land use classes in Brazil (MAPBIOMAS, 2019).

Additionally, because no field data was not available, high-resolution images from the Google Earth Pro platform were used to define polygons representative of natural forests, planted forests, pasture and agriculture. Table 1 provides a summary of the remote sensing data used for classification.

Table 1 – Material used for classification.

Data	Source	Band	Images	Observation
Image	Landsat 8 (https://code.earthengine.google.com)	R - G - B - NIR SWIR1 - SWIR2	137	Optical images
	Sentinel 1 (https://code.earthengine.google.com)	VV - VH	175	RADAR images
DEM	SRTM (https://code.earthengine.google.com)	elevation	1	Elevation
Ancillary	IBGE	-	-	Administrative Territorial Division
	MapBiomas	-	-	"Planted Forest", "Pasture" and "Agriculture" reference
	SOS Mata Atlântica	-	-	"Natural Forest" reference
	Google Earth Pro	-	-	High Resolution Images

Source: The authors (2020).

3 METHODS

First, we acquired all the inputs used for the work. Second, we pre-processed all remote sensing images to create a set of feature bands. Then, we manually draw polygons in high-resolution images to collect training and testing samples of the classes. All samples were selected at random within each polygon and their pixel values were normalized. We created ten different combinations of bands (data models) that were used to train the random forest classifier. We performed the classification procedure 100 times, changing at each realization the samples used to train and test the classifier. Finally, the results were evaluated to verify which is the best possible combination of bands to discriminate the studied classes. Figure 2 shows the flowchart of the study.

Figure 2 – Research flowchart.
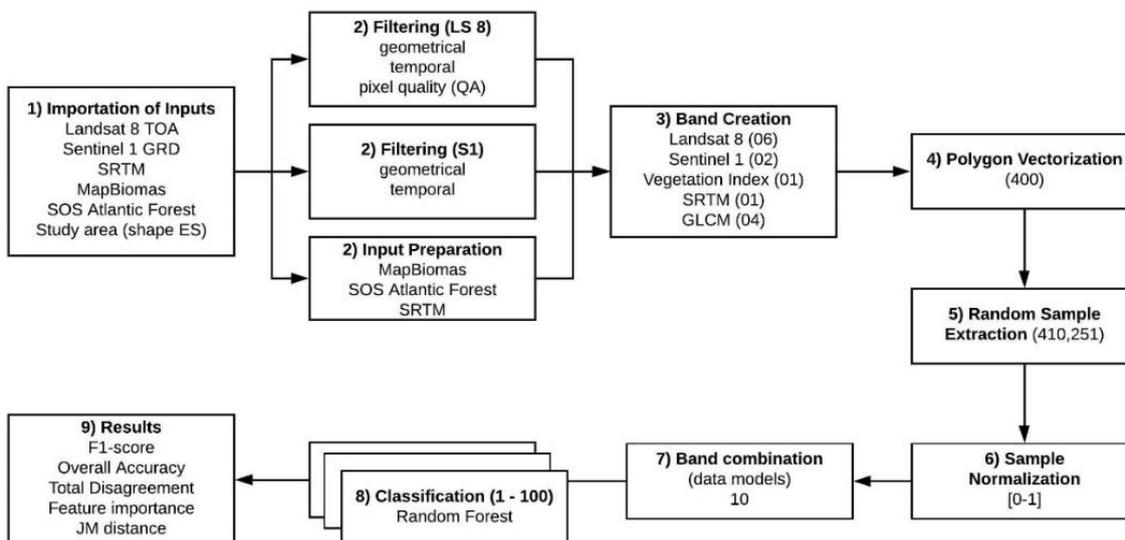
Source: The authors (2020).

3.1 Image pre-processing

3.1.1 LANDSAT-8 IMAGES

Landsat-8 images were filtered by geometric, temporal and pixel quality criteria. Initially, the geometric filter leaves only the images within the study area (Espírito Santo, Figure 1). As a result, less

computational effort is needed and there is a significant reduction in processing time. Then, time filtering chooses all available images for the year 2017, the period being analyzed in the study. Finally, a pixel quality filter is applied. This filter works on the Quality Assessment (QA) Landsat-8 Band provided by the United States Geological Survey (USGS, 2019). The QA band pre-classifies all pixels in the image. Only pixels that have been pre-classified as "CLEAR TERRAIN" are accepted, i.e. those that do not show clouds, shadows, ice, snow or other attributes that may make a terrain classification difficult. Then, it was defined that the value of each pixel will be the median of all values of the period (January 1st to December 31st, 2017). Finally, the selected images were cropped with the studied area shapefile (red polygon shown in Figure 1).

3.1.2   SENTINEL-1 IMAGES

Like the Landsat-8 images, the first pre-processing step for Sentinel-1 images refers to the geometric filter, so that only images within the studied area are chosen. Then, all available 2017 images were separated. Then, images with VV and VH polarization were selected in the Interferometric Wide Swath Mode (IW) with High Resolution (H) in Ground Range Detected (GRD). At the end of this step, the value of each pixel was defined as the median of all values for the period. Finally, the selected images were cropped with the studied area shapefile (red polygon shown in Figure 1).

3.1.3   SRTM

The elevation band of the SRTM image served as the basis for the terrain slope calculation. GEE's native function (ee.Terrain.slope) calculates the slope in degrees of terrain. The local gradient is calculated using the 04 neighboring pixels (GORELICK et al., 2017).

3.1.4   VEGETATION INDEX

The Normalized Difference Vegetation Index (NDVI) is intended for monitoring the health of vegetation and biomass and is a commonly used technique for detecting changes in forests (SADER, BERTRAND and WILSON 2003). NDVI corresponds to the difference between the Near Infrared (NIR) band and the Red (R) band divided by the sum of the two bands (ROUSE et al., 1973), as shown in Equation 1:

$$NDVI = \frac{NIR - R}{NIR + R}$$

(1)

The NDVI band was added to the image to aid in class separation.

3.1.5   GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM)

Pattern recognition has been successfully employed in classification tasks focused on different application domains, including image classification (MARTINS et al., 2011). Haralick (1973) first proposed a Gray Level Co-Occurrence Matrix (GLCM), which now is the most common form of texture extraction today. Also, GLCM was one of the first methods for texture extraction in images and has been widely employed.

Co-occurrence matrix is a tabulation of how many different combinations of pixel intensity values (gray levels) occur in an image. Fourteen textural features can be extracted from GLCM according to Haralick (1973). Zhang et al. (2017) state that four of them stand out in their use for remote sensing applications: energy, contrast, correlation, and entropy. Therefore, four texture bands were calculated from the Landsat-8 NIR band. The NIR band was chosen to calculate the texture metrics due to the high reflectance of vegetation in this range of the spectrum, which may enhance structural differences between the vegetation classes. Table 2 shows the features that will be used in this study.

Table 2 – All bands employed in the study.

Source	Bands
Landsat 8	R
	G
	B
	NIR
	SWIR1
	SWIR2
Sentinel 1	VV
	VH
Vegetation Index	NDVI
SRTM	slope
GLCM	entropy
	contrast
	correlation
	energy

Source: The authors (2020).

3.2 Training sample collection

For this study, four classes were defined: natural forest (NF), planted forest (PF), pasture (P) and agriculture (A). Information from the SOS Atlantic Forest Project (NF class) and the MapBiomas Project (PF, P and A classes) were used as ancillary data (Section 1.2.4) to assist in the collection of training samples.

A shapefile from the SOS Atlantic Forest Project was used with the location of the AF remnants in the state of Espírito Santo dated from 2017. This shapefile file was converted to a Keyhole Markup Language (KML) file and imported to the Google Earth (GE) platform. In GE, 100 polygons of the "Natural Forest" class were traced manually using information from the SOS Atlantic Forest as reference. This vectorization occurred using high-resolution images from the year 2017 (Figure 3). It should be noted that the polygons obtained with high-resolution GE images were verified in the 2017 Landsat image afterward. The importance of this procedure is to ensure that the samples were in fact from native forests. We sought to obtain these polygons with a considerable distance from each other, aiming to add variability. The 100 polygons of natural forests were then loaded into the GEE.

Figure 3 –Polygon acquisition: on the left, the yellow lines represent the reference (SOS Atlantic Forest or MapBiomas) and in green, the polygon that was vectorized with high-resolution images in GE. On the right, the same polygon exported to the GEE using Landsat-8 images.
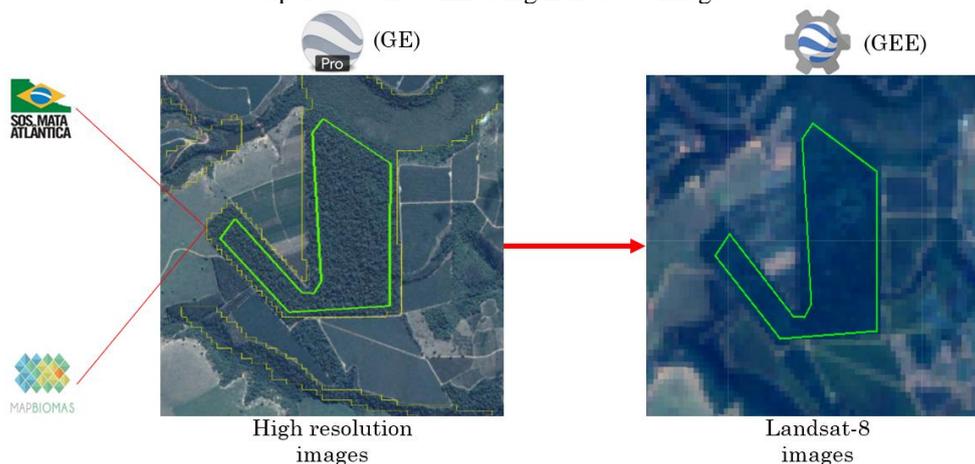
Source: The authors (2020).

The last available mapping collection from the MapBiomas Project (Collection - v3.1) was used, which contains data from 2017. Similarly, to the "Natural Forest" class, 100 polygons were manually delimited for each of the other classes ("Planted Forest", "Pastures" and "Agriculture"). This vectorization was carried out with high-resolution images from the year 2017 in GE and using as reference the information obtained from MapBiomas. Finally, the resulting 300 polygons were imported to GEE. Figure 4 shows the spatial location of all vectorized polygons within the studied area.

Figure 4 – Distribution of vectorized polygons.
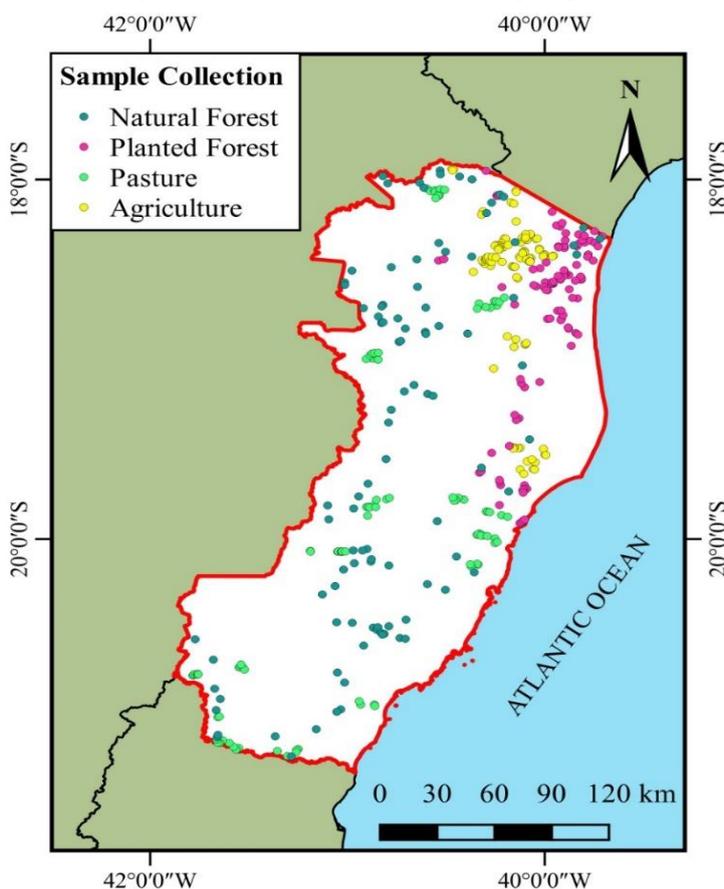
Source: The authors (2020).

In total, 400 polygons were delineated and imported to GEE for sample extraction (pixels), which represented 410,251 pixels that were used for training and testing the classifier. Table 3 shows in detail the distribution of polygons and samples (pixels) in each class.

Table 3 – Distribution of polygons and samples by class.

Class	Polygons	Samples
Natural Forest	100	245,353
Planted Forest	100	79,285
Pasture	100	45,659
Agriculture	100	39,954
	400	**410,251**

Source: The authors (2020).

3.3   Classification experiments and accuracy assessment

First, we extracted the pixels within the polygons obtained in Section 2.2, resulting in a dataset composed of 14 features (Table 2) and 410,251 samples (Table 3). The samples were then randomly assigned for training or testing, in the proportion of 60% to 40%, respectively. To avoid imbalance between classes and consequently biased training of the classifier, it was established that the classifier will receive 5,000 training samples for each class. There was no quantitative restriction whatsoever for test samples.

The classifier chosen for this research was the Random Forest (RF) (BREIMAN, 2001). According to Diniz et al. (2019), the RF algorithm was selected as the classifier of 15,000 articles in the last 5 years, thus being one of the most used classifiers in scientific studies. RF consists of a decision tree combination, where the input data is entered simultaneously in all created trees, obtaining a sequence of results. The winning class is the one that appears most frequently (majority voting).

According to Belgiu and Drăgu (2016), two parameters must be defined for the use of RF: the number

117

of trees to be generated (number of trees) and the number of variables to be selected and tested for the best tree division (variables per split). Preliminary tests showed that there was no significant improvement in accuracy after using more than 500 trees. The number of variables in each split was chosen as the square root of the number of variables because this is the default value for RF. Prior classification, data were normalized to the [0-1] range using the minimum and maximum values of each band.

The experiments were divided into groups called data models. Each data model is a combination of bands from the various sources cited (Table 4). The first data model is a combination of the six Landsat-8 (LS 8) bands. It is the basis for the other band combinations and will be used as a reference. Then, the other bands are added, forming different data models. For this study, 10 different data models were created in order to observe which one yields the best classification results.

Table 4 –Band combinations (data models) used in classification experiments: Landsat-8 visible to shortwave-infrared (LS8), Normalized Difference Vegetation Index (NDVI) and polarization of Sentinel-1 satellite (VV and VH).

Data model	Bands
1	LS 8
2	LS 8 + NDVI
3	LS 8 + VV
4	LS 8 + VH
5	LS 8 + VV + VH
6	LS 8 + VV + VH + slope
7	LS 8 + VV + VH + contrast
8	LS 8 + VV + VH + correlation
9	LS 8 + VV + VH + energy
10	LS 8 + VV + VH + entropy

Source: The authors (2020).

To ensure greater statistical robustness of the results, it was decided that the classification procedure is repeated 100 times, changing at each realization the samples used to train and test the classifier. It is worth noting that during this process the polygon identity was respected, that is, training and testing pixels came from different polygons. Table 5 shows the classifier parameters and image features used in the experiments. The classification accuracy was computed at each realization by using the testing samples. The F1-score (DERCZYNSKI, 2016), the overall accuracy and the total disagreement (PONTIUS & MILLONES, 2011) were calculated.

Table 5 – Classifier attributes and classification parameters.

Parameters	Values
classifier	Random Forest
number of trees (ntree)	500
variables per split (mtry)	square root of the number of variables
iterations	100
training samples (class)	60 % (limit: 5000)
testing samples (class)	40%
attributes	14 (R, G, B, NIR, SWIR1, SWIR2, NDVI, VV, VH, slope, entropy, contrast, correlation, energy)
classes	4 (Natural Forest, Planted Forest, Pasture and Agriculture)
metrics	F1-score Overall Accuracy Total Disagreement

Source: The authors (2020).

3.4　Feature importance

The quantification of the feature importance (or variable importance) is an important issue in many applied problems complementing variable selection by interpretation issues, especially in remote sensing. (GENUER, POGGI and TULEAU-MALOT, 2010). Here, we computed the variable importance in prediction

of the band combination that yields the best classification results, aiming to verify which variables most contributed to classification. For this, it was used the function *randomForest* in R to calculate it.

Breiman (2001) outlines an explanation of the good performance of random forests related to the quality of each tree along with a small correlation between forest trees, where the correlation between trees is defined as the ordinary correlation of predictions in out-of-bag (OOB) samples. The OOB sample is the set of observations that are not used to build the current tree. It is used to estimate the prediction error and then to evaluate the variable importance. The purpose of variable importance is to quantify the relevance of each attribute (band) in the classification procedures. At each realization, when training and testing samples are chosen, the variable importance is calculated. In the end, the average of the importance values is computed.

3.5   Class separability

The computation of class separability provides valuable insights to explain the classification results. The Jeffries–Matusita (JM) distance is a widely used statistical separability criterion. For the separability measurement, the normal distribution is usually considered. Aiming to verify if the separability of the classes increases by combining multi-source remote sensing data, we compute the JM distance according to Richards and Jia (1999), as seen in the Equation 2:

$$JM_{ij} = 2\left(1 - e^{-B}\right) \tag{2}$$

in which B is the Bhattacharyya distance (KAILATH, 1967) calculated as follows in Equation 3:

$$B = \frac{1}{8}\left(m_i - m_j\right)^T \left\{\frac{\Sigma_i + \Sigma_j}{2}\right\}^{-1}\left(m_i - m_j\right) + \frac{1}{2}\,ln\left\{\frac{\left|(\Sigma_i + \Sigma_j)/2\right|}{|\Sigma_i|^{1/2}\,|\Sigma_j|^{1/2}}\right\} \tag{3}$$

where mi and mj are the mean value and classes i and j, $\sum$i and $\sum$j corresponds to their covariance matrices, while |Σi |and |Σj | are the determinants of Σi and Σj, respectively. T is the transposition function. The JM distance varies from 0 to 2 and was calculated for all pairwise combinations between classes using the ENVI 5.5 software.

4   RESULTS

Table 6 shows the F1-score, overall accuracy and total disagreement results (values in percentage) of the ten band combinations that were evaluated in the study. Concerning the baseline scenario (using only LS 8 bands), the improvement observed by combining LS 8 bands (data model 1) with RADAR data (data model 5) in F1-score was 2.22%, 2.9%, 3.71% and 8.01% for the NF, PF, P and A classes, respectively (Table 6). The improvement in terms of overall accuracy was 3.0% (Table 6). The evolution in the Total Disagreement was 3.0% (Table 6). It is worth noting that, similar results were obtained with data model 9, which incorporated the energy feature in the classification process.

Table 6 – Mean of F1-score, overall accuracy and total disagreement resulting from classification (values in percentage). NF stands for Natural Forest, PF denotes a Planted Forest, P means Pasture and A implies Agriculture.

Data model	Bands	Fscore NF	Fscore PF	Fscore P	Fscore A	Overall Accuracy	Total Disagreement
1	**LS 8**	**93.43**	**85.48**	**82.48**	**69.18**	**88.29**	**11.71**
2	LS 8 + NDVI	93.46	85.46	81.99	68.5	88.19	11.81
3	LS 8 + VV	95.36	87.79	83.56	72.91	91.15	8.85
4	LS 8 + VH	93.93	85.01	84.05	74.24	91.06	8.94
5	**LS 8 + VV + VH**	**95.65**	**88.38**	**86.19**	**77.19**	**91.29**	**8.71**
6	LS 8 + VV + VH + slope	95.66	88.3	85.8	76.85	91.18	8.82

Rev. Bras. Cartogr, vol. 72, n. 1, 2020

DOI: http://dx.doi.org/10.14393/rbcv72n1-50477

Data model	Bands	Fscore NF	Fscore PF	Fscore P	Fscore A	Overall Accuracy	Total Disagreement
7	LS 8 + VV + VH + contrast	94.33	84.83	87.16	78.51	89.45	10.55
8	LS 8 + VV + VH + correlation	94.77	85.77	86.81	77.93	89.75	10.25
9	LS 8 + VV + VH + energy	95.69	88.37	86.14	77.18	91.31	8.69
10	LS 8 + VV + VH + entropy	94.5	85.14	87.18	78.52	89.64	10.36

Source: The authors (2020).

Figure 5 shows the confusion matrix (in percentages) of data model 1 (Landsat-8 bands) and data model 5 (Landsat-8 plus Sentinel-1 bands). All classes showed an improvement in their accuracy. The false-negative error (lower diagonal part of the confusion matrix) between planted and natural forest areas, that is, pixels predicted as planted forest that belonged to natural forests, reduced 2.3%. The highest decrease in classification error was observed between pasture and agriculture. For these classes, the false positive rate, that is, pixels belonging to agriculture that were classified as pasture, decreased 5% after the inclusion of Sentinel-1 bands in the classification process.

Figure 5 – Confusion matrices showing the classification accuracy of natural forest, planted forest, pasture and agricultural areas obtained using Landsat-8 bands (left panel) and Landsat-8 bands plus Sentinel-1 VV and VH bands (right panel). The diagonal cells show the classification accuracy of each class. The off-diagonal cells show the percentage of classification error. Classification was performed 100 times with the random forest algorithm. At each realization, training and testing samples were changed.
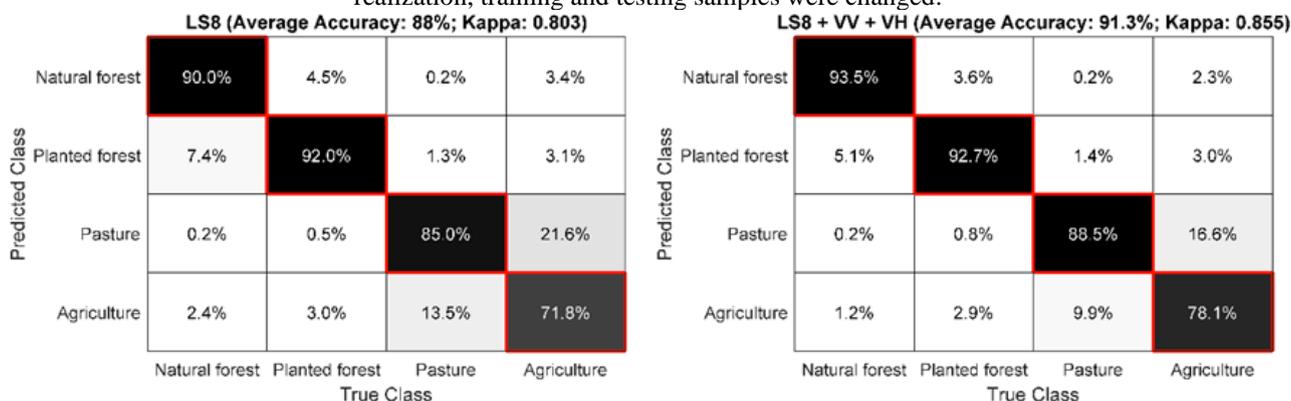
Source: The authors (2020).

The classification results showed above are corroborated by both feature importance and class separability analysis. Figure 6 shows the results of the out-of-bag feature importance. The Sentinel-1 features, along with the NIR bands, were the most important to discriminate among the classes.

Figure 6 – Feature importance for classification of natural forest, planted forest, pasture and agricultural areas with Landsat-8 bands plus VH and VV Sentinel-1 bands. Feature importance was computed 100 times, changing at each realization the samples used to train and test the random forest algorithm.
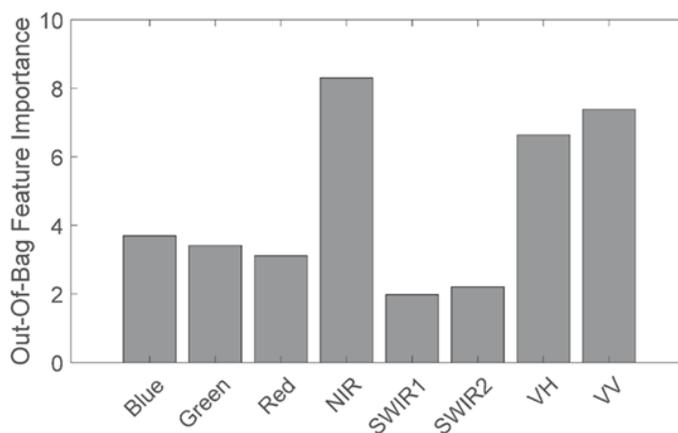
Source: The authors (2020).

The combination of Landsat-8 and Sentinel-1 bands improved the separability of all classes (Figure 7). Specifically, the inclusion of VV and VH features improved the separability between natural and planted forest areas in 11.98% (Figure 7c), while the improvement between agriculture and pasture areas was 45.12%.

Figure 7 – JM Distance: (a) JM Distance using Landsat-8 bands. (b) JM Distance using Landsat-8 plus Sentinel-1 bands. (c) the improvement (%) caused by adding Sentinel-1 bands.

(a) JM Distance (LS8)					(b) JM Distance (LS8 + VV + VH)					(c) JM Distance Improvement (%)				
	NF	PF	A	P		NF	PF	A	P		NF	PF	A	P
NF	-				NF	-				NF	-			
PF	1.67	-			PF	1.87	-			PF	11.98	-		
A	1.89	1.82	-		A	1.94	1.85	-		A	2.65	1.65	-	
P	1.94	1.93	0.82	-	P	1.97	1.95	1.19	-	P	1.55	1.04	45.12	-

Source: The authors (2020).

5   DISCUSSION

Previous studies combined multi-sensor data to a variety of purposes. Sonobe et al. (2017) used Sentinel-1A and Sentinel-2A images with the aim of determining the best method to generate crop maps. Among four classification algorithms (SVM, RF, FNN, and KELM), they found that KELM generated the most accurate crop classification map for the study area, with an OA of 96.8%. Carreiras et al. (2017) used ALOS PALSAR and Landsat 5 TM to discriminate mature forest, non-forest and secondary forest. They mapped the Brazilian Amazon in the 2007–2010 period with an overall accuracy of 95–96%. Lucas et al. (2019) retrieved comprehensive information on the biophysical properties of mangroves employing Japanese L-band Synthetic Aperture Radar (SAR) data and WorldView-2 images. The study provided new information on the biophysical properties and growth dynamics of mangrove forests in Malaysa, inputs for future monitoring activities, and methods for facilitating better characterization and mapping of mangrove areas worldwide. Nevertheless, few studies are found in the literature with the objective of differentiating planted forest from native forest, particularly in tropical environments

To advance in this point (topic), our first hypothesis, which states that the combination of Landsat-8 and Sentinel-1 data improves the classification accuracy of natural and planted forests, was confirmed by the results of experiments (Section 2.3). The inclusion of RADAR bands in the classification process increased the overall accuracy by 3.0% and decreased the total disagreement by 3.0% if compared to the use of Landsat-8 bands only (Table 6).

Moreover, it was observed the spectral separability between natural forests and tree plantations increased by combining passive optical and RADAR data, according to the results of experiments (Section 2.5). corroborating our second hypothesis. The addition of RADAR bands in the proceeding incremented the JM distance by 11.98 % (Figure 7). Also, it is important to note that the feature importance experiment (Section 2.4) showed the most important attributes for classification procedure were NIR, VV, VH (Figure 6).

The reason for this improvement, both spatial and spectral resolution, may be the higher spatial resolution of Sentinel-1 than Landsat-8. Another point to be raised is how the trees are arranged on the ground. The interaction between the microwaves of Sentinel-1 with the natural and planted forests may be able to detect the different geometric arrangements between these forest types, which is not possible by using optical Landsat-8 bands. Our results also show that combining Landsat-8 and Sentinel-1 data improved discrimination between pasture and agriculture.

The computation of GLCM bands did not improve the classification results (Table 6). It is believed that the spatial resolution of Landsat-8 (30 m) may have been a key obstacle. The availability of high spatial resolution images would probably allow better results with the use of texture bands.

The challenges are stimulating. Despite the difficulties in differentiating vegetations with such similar spectral responses, the possibility of, soon, use images with better spatial and spectral resolution is encouraging. The results obtained point out to potential applications in on the use of multi-sensor data that are

freely available in the GEE platform for mapping vegetation cover in tropical regions.

6   CONCLUSIONS

In this study, we gather multi-source remote sensing data aiming to improve discrimination between natural forests, planted forests (eucalyptus), pasture and agricultural areas in the state of Espírito Santo, southeast of Brazil.

Combining Landsat-8 and Sentinel-1 was effective for improving the classification accuracy of natural forests and tree plantations, increasing the Overall Accuracy from 89.29% to 91.29% and decreasing the Total Disagreement from 11.71% to 8.71%. Furthermore, F1-score of NF, PF, P and A was improved in 2.22%, 2.9%, 3.71% and 8.01%, respectively. Likewise, spectral separability becomes better not only between planted and native forests (11.98 %) but equally between pasture and agriculture classes (45.12 %), according to JM distance results.

The results of this study verify the validity of this remote sensing method, demonstrate its strong potential for tree discrimination, and suggest that the use of data from both satellites could be expanded in future.

Acknowledgment

Authors' Contributions

The experiments were designed by G.F.C. and M.P.F.; G.F.C. and M.P.F. performed the experiments; G.F.C. and M.P.F. analyzed and discussed the numerical results; G.F.C. drafted the article; C.F.S.V. critically revised the article; All authors approved the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

References

BELGIU, M.; DRĂGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, 2016. pp. 24-31. DOI. 10.1016/j.isprsjprs.2016.01.011

BREIMAN, L. Random Forests. **Machine Learning**, 2001. pp. 5-32. DOI. 10.1023/A:1010933404324

CARREIRAS, J. M. B., JONES, J., LUCAS, R. M., SHIMABUKURO, Y. E. Mapping major land cover types and retrieving the age of secondary forests in the Brazilian Amazon by combining single-date optical and radar remote sensing data. **Remote Sensing of Environment**, v. 194, 2017. pp. 16–32. DOI. 10.1016/j.rse.2017.03.016

CENTRO DE DESENVOLVIMENTO DO AGRONEGÓCIO (CEDAGRO). **Dimensionamento do mercado capixaba de produtos florestais madeiráveis**. Technical report, 2011.

CLIMATE. Available in: <https://en.climate-data.org/south-america/brazil/espirito-santo-210/>. Accessed in: May 2019.

CONFEDERAÇÃO   NACIONAL   DA   INDÚSTRIA   (CNI).   Available   in: <http://perfildaindustria.portaldaindustria.com.br/estado/es>. Accessed in: August 2019.

DERCZYNSKI, L. Complementarity, F-score, and NLP evaluation. **Proceedings of the 10th International**

Conference on Language Resources and Evaluation, LREC 2016. pp. 261-266.

DINIZ, C.; CORTINHAS, L.; NERINO, G.; RODRIGUES, J.; SADECK, L.; ADAMI, M.; SOUZA-FILHO, P. W. M. Brazilian mangrove status: Three decades of satellite data analysis. **Remote Sensing**, v. 11, n. 7, 2019. DOI. 10.3390/rs11070808

EUROPEAN SPACE AGENCY (ESA). Available in: <https://sentinel.esa.int/web/sentinel/missions/sentinel-1>. Accessed in: May 2019.

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO). **Global Forest Resources Assessment 2015**. Desk Reference, Rome, 2015.

GARBIN, M. L.; SAITER, F. Z.; CARRIJO, T. T.; PEIXOTO, A. L. Breve histórico e classificação da vegetação capixaba. **Rodriguesia**, v. 68, n. 5, 2017. pp. 1883-1894. DOI. 10.1590/2175-7860201768521

GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. Variable selection using random forests. **Pattern Recognition Letters**, v. 31, n. 14, 2010. pp. 2225-2236. DOI. 10.1016/j.patrec.2010.03.014

GORELICK, N., HANCHER, M., DIXON, M., ILYUSHCHENKO, S., THAU, D., & MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, 2017. pp. 18–27. DOI. 10.1016/j.rse.2017.06.031

HARALICK, R. M.; DINSTEIN, I.; SHANMUGAM, K. Textural Features for Image Classification. **IEEE Transactions on Systems, Man and Cybernetics**, v. SMC-3, n. 6, 1973. pp. 610-621. DOI. 10.1109/TSMC.1973.4309314

INDÚSTRIA BRASILEIRA DE ÁRVORES (IBA). Technical report 2017. **Indústria Brasileira de Árvores - IBÁ**, p. 80, 2017. Available in: <http://iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf>. Accessed in: June 2019. DOI. 10.1017/CBO9781107415324.004

JOSHI, N.; BAUMANN, M.; EHAMMER, A.; FENSHOLT, R.; GROGAN, K.; HOSTERT, P.; WASKE, B. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. **Remote Sensing**, v. 8, n. 1, 2016. pp. 1-23. DOI. 10.3390/rs8010070

KAILATH, T. The Divergence and Bhattacharyya Distance Measures in Signal Selection. **IEEE Transactions on Communication Technology**, v. 15, n. 1, 1967. pp. 52-60. DOI. 10.1109/TCOM.1967.1089532

LIMA, C. M. G. **Modelagem de transição florestal na Mata Atlântica**. Dissertação de mestrado. Universidade Federal de Minas Gerais, Instituto de Geociências, Belo Horizonte, 2012.

LUCAS, R., VAN DE KERCHOVE, R., OTERO, V., LAGOMASINO, D., FATOYINBO, L., OMAR, H., … DAHDOUH-GUEBAS, F. Structural characterization of mangrove forests achieved through combining multiple sources of remote sensing data. **Remote Sensing of Environment**, v. 237, 2019. DOI. 10.1016/j.rse.2019.111543

MAPBIOMAS. Available in: <http://mapbiomas.org/>. Accessed in: July 2019.

MARTINS, J. G.; COSTA, Y. M. G.; BERTOLINI, D.; OLIVEIRA, L. S. Uso de descritores de textura extraídos de GLCM para o reconhecimento de padrões em diferentes domínios de aplicação. **XXXVII Conferência Latinoamericana de Informática**, 2011. pp. 637-652.

MARTONE, M.; RIZZOLI, P.; WECKLICH, C.; GONZÁLEZ, C.; BUESO-BELLO, J. L.; VALDO, P.; MOREIRA, A. The global forest/non-forest map from TanDEM-X interferometric SAR data. **Remote Sensing of Environment**, v. 205, 2018. pp. 352-373. DOI. 10.1016/j.rse.2017.12.002

MENDES, L. **Brazilian Forestry and Timber Yearbook 2016**. Technical report, 2016.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). Available in: <https://www2.jpl.nasa.gov/srtm/>. Accessed in: July 2019.

PAN, Y.; BIRDSEY, R. A.; FANG, J.; HOUGHTON, R.; KAUPPI, P. E.; KURZ, W. A.; PHILLIPS, O. L.; CANADELL, J. G.; CIAIS, P.; JACKSON, R. B.; PACALA, S. W.; MCGUIRE, D.; PIAO, S.; RAUTIAINEN, A.; SITCH, S.; HAYES, D. A large and persistent carbon sink in the world's forests. **Science**, v. 333, 2011. pp. 988-993.

PONTIUS JR, ROBERT GILMORE; MILLONES, MARCO. Death to Kappa: birth of quantity disagreement

and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, 2011. pp. 4407-4429.

QIN, Y.; XIAO, X.; DONG, J.; ZHANG, Y.; WU, X.; SHIMABUKURO, Y.; MOORE, B. Improved estimates of forest cover and loss in the Brazilian Amazon in 2000–2017. **Nature Sustainability**, 2019. DOI. 10.1038/s41893-019-0336-9

RICHARDS, J. A.; JIA, X. **Remote Sensing Digital Image Analysis: An Introduction**. Canberra: Springer, 1999. DOI. 10.1007/978-3-662-03978-6

ROUSE, J.W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D.W. Monitoring vegetation systems in the Great Plains with ERTS. **Proceedings of third ERTS-1 symposium**, v. 1, Washington, DC, 10–14 December 1973. NASA, SP-351 pp. 309–317.

SADER, S. A.; BERTRAND, M.; WILSON, E. H. Satellite change detection of forest harvest patterns on an industrial forest landscape. **Forest Science**, v. 49, n. 3, 2003. pp. 341-353.

SLIK, J. F.; ARROYO-RODRÍGUEZ, V.; AIBA, S. I.; ALVAREZ-LOAYZA, P.; ALVES, L. F.; ASHTON, P.; BERNACCI, L. An estimate of the number of tropical tree species. **Proceedings of the National Academy of Sciences**, v. 112, n. 24, 2015. pp. 7472-7477.

SONOBE, R., YAMAYA, Y., TANI, H., WANG, X., KOBAYASHI, N., & MOCHIZUKI, K. ICHIRO. (2017). Assessing the suitability of data from Sentinel-1A and 2A for crop classification. GIScience and **Remote Sensing**, v. 54, n. 6, 2015. pp 918–938. DOI. 10.1080/15481603.2017.1351149

SOS/INPE, FUNDAÇÃO SOS MATA ATLÂNTICA; INPE. **Atlas dos Remanescentes Florestais da Mata Atlântica 2017-2018**. Technical report, 2019.

THOMAZ, L. D. A Mata Atlântica no estado do Espírito Santo, Brasil: de Vasco Fernandes Coutinho ao século 21. **Boletim Do Museu de Biologia Mello Leitão**, n. 27, 2010. pp. 5-20.

UNITED STATES GEOLOGICAL SURVEY (USGS). Available in: <https://www. https://www.usgs.gov/land-resources/nli/landsat/landsat-8?qt-science_support_page_related_con=0#qt-science_support_page_related_con>. Accessed in: July 2019.

UNITED STATES GEOLOGICAL SURVEY (USGS). Landsat Quality Assessment (QA) Tools. Available in: <https://landsat.usgs.gov/sites/default/files/documents/landsat_QA_tools_userguide.pdf> Accessed in: July 2019.

WANG, YP; HOULTON, BZ. Nitrogen constraints on terrestrial carbon uptake: implications for the global carbon-climate feedback. **Geophysical Research Letters**, v. 36, L24403, 2009.

YU, X.; HYYPPÄ, J.; KARJALAINEN, M.; NURMINEN, K.; KARILA, K.; VASTARANTA, M.; KATOH, M. Comparison of laser and stereo optical, SAR and InSAR point clouds from air- and space-borne sources in the retrieval of forest inventory attributes. **Remote Sensing**, v. 7, n. 12, 2015. pp 15933-15954. DOI. 10.3390/rs71215809

ZHANG, X.; CUI, J.; WANG, W.; LIN, C. A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. **Sensors**, v. 17, n. 7, 2017. DOI. 10.3390/s17071474

Lead author's biography

Gustavo Fluminense Carneiro, born in Rio de Janeiro in 1981. Bachelor of Military Science from the Military Academy of Agulhas Negras (2003). Bachelor (2009) and Master (2019) in Cartographic Engineering from the Military Engineering Institute, in the area of Remote Sensing. He served in the 5 Geoinformation Center, acting in various roles, including Deputy Head of the Geoinformation Division. He currently holds the position of Professor at the Military Institute of Engineering.