Australian Journal of Basic and Applied Sciences

ISSN :1991-8178
EISSN: 2309-8414

2022 August; 16(8): pages 1-14
https://doi.org/10.22587/ajbas.2022.16.8.1

Research Paper AENSI Publications

Sentinel-1 and Sentinel-2 data fusion by Principal Components Analysis applied to the vegetation classification around power transmission lines

Hypolito Suarez Fernandez[1], Francisco Henrique de Oliveira[2], Jessica Gerente[3], Francisco Caruso Gomes Junior[4], Lucas Antonio Providelo[5], Guilherme Marchiori[6], Yang Liu[7]

[1]MSc Student, Territorial Planning and Socio-environmental Development Graduate Program, Santa Catarina State University – UDESC, Florianópolis-SC, Brazil
[2]Associate Professor, PhD, Territorial Planning and Socio-environmental Development Graduate Program, Santa Catarina State University – UDESC, Florianópolis-SC, Brazil
[3]Research and Development (R&D) Project Manager, CARUSO Environmental and Technological Solutions, Florianópolis, Santa Catarina, Brazil
[4]Director, CARUSO Environmental and Technological Solutions, Florianópolis, Santa Catarina, Brazil
[5]GIS Specialist, CPFL Renewables, Jundiaí, São Paulo, Brazil
[6]Innovation Analyst, CPFL Renewables, Campinas, São Paulo, Brazil
[7]Deputy CEO of CPFL-T, CPFL Transmission, Porto Alegre, Rio Grande do Sul, Brazil

Correspondence Author: Hypolito Suarez Fernandez. MSc Student, Territorial Planning and Socio-environmental Development Graduate Program, Santa Catarina State University – UDESC. 2007, Av. Madre Benvenuta, Florianópolis-SC, Brazil
E-mail address: 04666168990@edu.udesc.br

Received date: 12 April 2022, Accepted date: 15 June 2022

Citation: H. S. Fernandez., F. H. de Oliveira., J. Gerente., F. C. Gomes Junior., L. A. Providelo., G.Marchiori., Y. Liu., 2022. Sentinel-1 and Sentinel-2 data fusion by Principal Components Analysis applied to the vegetation classification around power transmission lines. Australian Journal of Basic and Applied Sciences, 16(8): 1-14. DOI: 10.22587/ajbas.2022.16.8.1.

ABSTRACT: Brazil's continental dimensions combined with an overwhelmingly hydroelectric energy generation matrix puts the country's energy transmission system among the more extensive in the world. A considerable portion of this extension passes through forested areas, which represent a risk to transmission since the growth of these forest formations can interrupt the energy transmission in cases of contact or fall on the asset. Monitoring these assets by means of remote sensing is an efficient alternative to observing vegetation development. Therefore, the fusion of satellite images associated with machine learning mapping techniques can be considered an especially important resource for Electrical Energy Transmission companies in Brazil regarding energy transmission asset monitoring. Considering this universe of electrical sector action, this project investigates the potential of combining radar and multispectral images to demarcate areas with consolidated forest physiognomy. By the methodological principle, the images provided by the Sentinel-1 and Sentinel-2 platforms went through an experimental fusion process by the appliance of the Principal Component Analysis method (PCA), and classified by the Random Forest algorithm (RF). The result was then compared to the product classified by RF which uses only Sentinel-2 data. In conclusion, it was observed that the fusion between images allied to the classification method allowed for a significant improvement in the demarcation of vegetation targets (objects of interest). However, it must still be considered that the results can be improved in future studies – requiring a deeper scientific research in each step of the method's execution established in the research.

Keywords: Principal Components Analysis, Random Forest, Vegetation monitoring, Power transmission

INTRODUCTION

The institutional scene of the Brazilian electrical sector is organized hierarchically by a regulatory agency, a national secretariat, some public companies and administrative councils. The National Electrical Energy Agency (ANEEL) is configured in an autarchy tied to the Ministry of Mines and Energy (MME) and has the incumbency of regulating and monitoring the activities related to the Brazilian electrical sector. In addition, it acts as a mediator between interested parties: companies of the industry, state and consumer (Law #9.427, 1996). The National Operator of the System (ONS) is the secretariat responsible for the National Linked System (SIN), which articulates the operation of electrical energy generation power plants and transmission networks following its production of projections on the supply and demand for energy (Law #9.648, 1998).

The National Linked System (SIN) has an installed generation capacity of 166.76 GW, of which 85% are derived from renewable sources. In 2018, 601.396 GWh were generated, of which 64.68% were from hydroelectric power plants, 9.08% from natural gas thermal power plants, 8.69% from biomass thermal power plants and 8.06% from wind farms. With 141.756 km of transmission lines, the basic Brazilian network is one of the largest in the world, spanning the entirety of the national territory, and presenting smaller distribution segments in northern Brazil, as can be observed in Fig. 1.

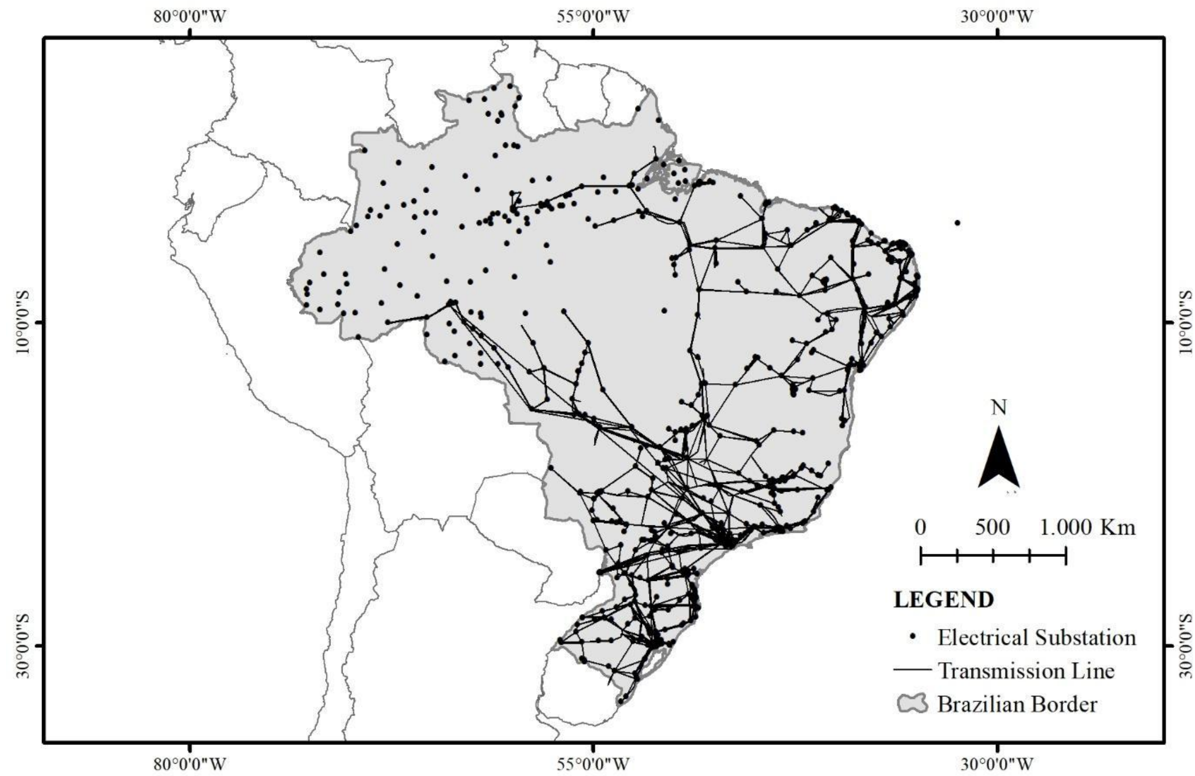

Figure 1: Brazilian National Linked System (SIN).

The existing integration in the hydro-thermo-wind system (with a predominance of the hydro-electrical matrix), constituted of the South, Southeast/Centre-West, Northeast, and North subsystems, allows for the interconnectivity between the regions through the transmission grid. It is also opening the way to transfer energy between subsystems, generating synergetic gains by using the diversity of hydrological regimes of the different hydrographic basins (Institute for Applied Economic Research, 2020).

The Brazilian electrical energy transmission system's dimensions and the role of transmission as an essential component of this system generate a vast maintenance cost to adapt itself to the nation's technical norms established for this type of operation. For example, among the numerous sections of the technical standards related to transmission installations, one pertains to the presence of vegetation on the security band, which extends on both sides of the transmission line's axis (Institute for Applied Economic Research, 2020).

The technical norm NBR 5422, titled "Transmission Lines" figures as the first national guiding document focused on the physical structure of Transmission Lines (TL) and was published for the first time in 1985 (Brazilian Association of Technical Norms, 1985). It calculates the safety clearance width based on the central axis of the transmission line according to its tension. Regarding the vegetation in the safety clearance, it states that its deforestation must be reduced to the minimum necessary to guarantee the TL's building, operational and safe maintenance conditions. Furthermore, undergrowth must be preserved to avoid erosion and for the same reason, deforestation and cuts in the terrain which cause soil erosion must be avoided. After TL is installed, the clearance may be used for cultivation as long as the top of the plants and the conductor in the recommended temperature of 50° (windless), remain at least equal to the distance H defined in item 13.2 of norm NBR 5422 and represented by Fig. 2.

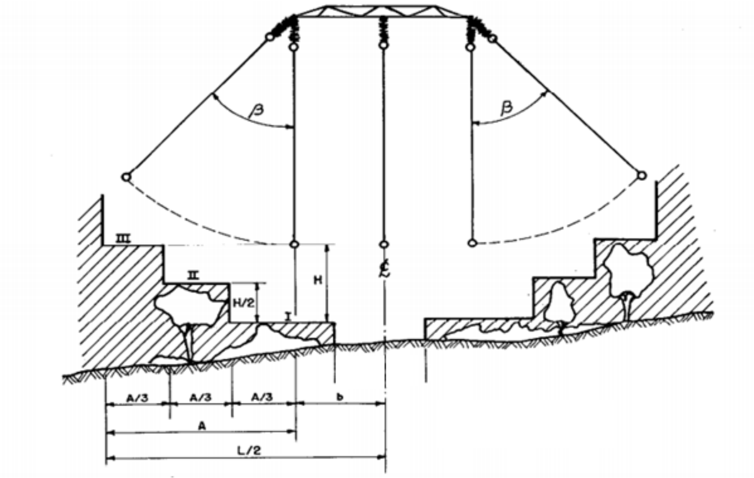

Figure 2: Transverse section of a TL and the safety clearance.

The security parameter H is due to the risk of branches or tree trunks colliding with the Electrical System. In case the observance of the good maintenance practices fails and there is an interruption of service, ANEEL published the Normative Resolution #846, dated 11 July 2019, which has the goal of establishing procedures, parameters and criteria for the imposition of penalties on sector agents, as well as explaining about general monitoring directives (National Electrical Energy Agency, 2019).

Remote sensing techniques allow images of the Earth's surface in multiple electromagnetic spectrum wavelengths. One of the most critical characteristics of remote sensing products is the electromagnetic band in which they are obtained. In the visible and near-infrared wavelengths, the sensor's wavelengths may be solar light reflected by the Earth's surface. Others may be the measuring of energy emitted by the Earth, such as in thermal infrared. Both cases are called passive multispectral sensors because they collect only the signal which exists in nature. The sensors emit signals toward the surface and capture its reflection, such as synthetic aperture radars – which work in the microwave wavelengths – are known as active sensors (Dyring, 1973).

By measuring the energy reflected by the targets situated on the Earth's surface in a variety of wavelengths, a spectral signature of each object can be created. The comparison between answering patterns of different characteristics of targets allows for them to be distinguished, which in some situations may not be possible in only one wavelength. A classic example of this is the confusion caused by the similarity between water surfaces and some kinds of vegetation on the visible wavelengths but never in the infrared spectrum. As a means of aggregating even more potential to the remote sensing data, synthetic aperture radar data may be used, which offers information on surface texture (Dyring, 1973).

Forest formation mapping projects combining radar data, especially Synthetic Aperture Radars (SARs), and multispectral data are abundant. In general, studies are tied to estimating the canopy height and the biomass covering the soil. However, combining data from different sensors is more widely made by stacking multispectral bands, backscatter coefficient channels and indexes.

Erinjery, Singh and Kent (2018), using data provided by the Sentinel-1 and Sentinel-2 platforms, realized the mapping of different vegetation typologies combining channels, spectral bands and textures. This process achieved higher accuracy when classified via the Maximum Likelihood Classification (MLC) algorithm than the Random Forest (RF) classifier. The highest accuracy achieved in the authors' research was obtained by a set of data involving bands 2, 3, 4, and 8; entropies of bands 4 and 8; and NDVI, all sources from the Sentinel-2 platform, as well as the VV and VH channels of platform Sentinel-1 and their respective entropies. In the referred study, according to Erinjery, Singh and Kent (2018), textures improved classification by reducing heterogeneity while preserving the limits between equal land cover typologies.

Dymond et al. (2019) presented a classification made using the SVM (Support Vector Machine) algorithm to map the native physiognomies of the forest, using as predictors data from Sentinel-1, Sentinel-2 and ALOS PALSAR. The authors obtained the best result with Sentinel-2's 2, 3, 4, 8, 11 and 12 bands, combined with the ratio between Sentinel-1's VH and VV channels, achieving 80.5% accuracy.

Spracklen and Spracklen (2021) used Sentinel-1 and Sentinel-2 data to map native and planted forest areas using the RF classifier. They obtained accuracies of 87% for Sentinel-1 data, 92.5% for Sentinel-2 data, and 92.3% from data combined in the process. The authors noted that the ratio between SWIR and red bands is the most effective means to differentiate planted forests when the younger trees are over 6 years old, with an accuracy of 70%.

Chen et al. (2018) applied the backscatter coefficients from Sentinel-1 data combined with multispectral bands from Sentinel-2 and vegetation indexes of the exact origin for estimating the above-ground biomass. The results showed that characteristics of texture and biophysical variables of the vegetation were the most important predictors, obtaining a correlation coefficient (r^2) of 1.

The classifiers used were the Artificial Neural Network (ANN), Support Machine Vector for Regression (SVR) and Random Forest (RF), and SVR achieved the best result.

Other studies, such as Debastiani et al. (2019), tested as predictors: Sentinel-1's VV and VH channels; Sentinel-1's textures; multispectral Sentinel-2 bands; and vegetation indexes to estimate above-ground biomass. In this case, the predictor with the highest correlation was the Soil Adjusted Vegetation Index (SAVI) ($r^2 = 0.44$).

Fagua et al. (2019) integrated multispectral and SAR data to estimate the canopy height in South American tropical forests with a spatial resolution of 30 meters. Effects were examined on three forest types: dry, humid and rainy periods, using five classification algorithms. Among those tested, the one which obtained the highest accuracy was the RF (1.2m +/- 0.05 for dry, 4.9m +/- 0.14 for humid and 5.5m +/- 0.3 for forests in rainy seasons).

Liu et al. (2019) presented a methodological platform to estimate the median height of forest formation and above-ground biomass. Firstly, a product containing the median forest formation height is obtained through Sentinel-1's VV and VH channels and one of Sentinel-2's spectral bands for the respective regions. Then the authors elaborated three models for estimating the above-ground biomass and the different vegetation physiognomies. Field data was associated with the estimated median forest formation height data and biophysical variables of the spectral data, vegetation indexes and topographical variables. The results show that the use of channel VH's backscatter coefficient, when classified with RF, presents higher accuracy. In addition, three over-ground biomass models for different vegetational physiognomies achieved better results when compared to the grouping of all physiognomies, achieving determination coefficients (r^2) of 0.69 to 0.74.

Hirschmugl et al. (2020) used a dense time series of optic Sentinel-2 and Landsat 8 data together with SAR Sentinel-1 data to map disturbances in native forest formations of Peru and Gabon. The accuracies of disturbances mapped individually by optic and SAR data were compared to the accuracies of mappings combined with optic and SAR data. The study confirms a significant improvement in the mappings with combined data, of 13% for areas in Gabon and 25% in Peru.

DATA FUSION: PRINCIPAL COMPONENTS ANALYSIS

The main interest in image fusion is creating a composition with improved interpretability. The image produced by fusion must contain the highest possible amount of spatial information and preserve a good quality of spectral information. The employed fusion process must not distort the spectral characteristics of the original data since it is from them that spectral separation of the imaged targets is made possible. These products allow a more precise delimitation of features, making them useful for various applications (Nikolakopoulos, 2008).

The preservation of spectral characteristics is particularly welcome in the application of vegetation coverage analysis. Some authors suggest that images produced through fusion might simulate multispectral images acquired in higher spatial resolutions (Nikolakopoulos, 2008).

In general, fusion techniques can be divided into two classes: (1) colour-related techniques; and (2) statistical/numerical techniques. The first is based on the composition of three image channels in the RGB colour interval and more sophisticated colour creation, such as the Intensity Hue Saturation (IHS) and Brovey techniques. The approaches developed on a statistical channel basis include correlation and other techniques such as Principal Components Analysis (PCA) and regression.

The Principal Components Analysis technique consists of an orthogonal linear model of transformation of the original values into a set of non-correlated values, known as Principal Components (PC). When applied to images, the procedure attributes new values to each pixel, producing a new set of bands. The PC bands don't store correlation with one another and stay ordered according to the variance value of each image about the total set's variance. PC1 is the image with the highest conflict, PC2 has the second-highest variance, and so on. The variance expresses the amount of redundancy observed between the original images, which can aggregate a higher likelihood of identifying specific targets in the context of remote sensing.

PCA is useful in image codification tasks, dataset reduction, image enhancement, change detection and image fusion. When originated from SAR data, the principal component images show potential for topographical mapping, mainly for three-dimensional prints (Pohl and Van Genderen, 1998). It's a statistical technique transforming a set of multivariate data of intercorrelated variables into a dataset of new non-correlated linear combinations (Pohl and Van Genderen, 1998).

In an applied manner, there are two approaches for application to remote sensing images, selective and standard. The standard approach uses PCA for all the spectral bands available as input data and aims to fuse images sensed by different instruments. The selective approach is employed to enhance the spatial resolution of panchromatic images. In this case, the PCA fusion is first applied to a lower resolution multispectral image (which we wish to enhance). Then, the first principal component, which resembles a panchromatic image (98-99% of the variance is contained in the first three main components), is substituted by a panchromatic image of higher spatial resolution and then the reverse PCA is applied. This step generates a new idea of a higher spatial resolution (Pohl and Van Genderen, 1998; Pal, Majumdar and Bhattacharya, 2007; Nikolakopoulos, 2008).

The PCA approach is sensitive to the choice of the region to be analyzed. The correlation coefficient reflects the proximity relationship of homogenous samples. However, changes in band values caused by different land coverages influence the correlations and, consequently, the resulting image.

An assessment for the employment of this technique is related to the criteria's definition and the evaluation capacity of one's product. Both fields, employment, and assessment, are challenging to be defined and depend considerably on empirical results (Pohl and Van Genderen, 1998).

CLASSIFICATION: RANDOM FOREST

The use of machine learning algorithms for supervised classification is widely publicized for thematic mapping of multispectral images, both in environmental and privately funded research. Furthermore, scientific literature shows attempts to aggregate data for the enrichment of different soil usage classifications, especially for vegetation coverage. In this regard, research consisted of comparing the results of pure multispectral data classification and the classification results of images obtained via the analysis of the principal components of multispectral and radar data.

Developed as an extension of the CART (Classification and Regression Trees) software, RF is a technique that seeks to improve this model's prediction accuracy. It consists of creating a forest of predictive trees, where each tree is generated from a random vector, sampled independently and with the same distribution for all trees. The divisions inside each tree are determined based on a subset of co-variables chosen randomly from the existing set (Junior et al., 2016). RF classifications are generally considered stabler when compared to CART and other parametrized techniques, such as Maximum Likelihood, because of the use of *bootstrapping* (a method of hypotheses construction based on parameters of interest) and then the use of a random subset of data for the construction of the random forest model.

Several studies compare the results of executed classifications using RF with different parameters on their model. However, little has been written on the RF's sensitivity to the different training data selection strategies used. Various aspects of sample collection strategies can play essential roles in the classification result.

Supervised image classification requires the collection of training and validation samples for the production of both thematic and characteristics of interest maps. Independently of the classifier choice, accuracy assessment is used to determine the classification's quality. Many factors may affect results in an accuracy assessment, including the training sample's size, the number of defined classes, the quality of the training data in the characterization of classes to be mapped and the dimensionality of the data. When an image classification and accuracy assessment is made, training and validation data must be statistically separated (in a way that does not represent a spatial grouping). In addition, samples must be representative of the respective classes and abundant (Millard and Richardson, 2015).

Ideally, a randomly distributed sample should be used for obtaining training and validation data, which may consume much time and be challenging to implement in the event of requiring field validation for each sampled region. However, when training and validation data are not randomly distributed, such as in the polygon application in homogenous areas, these samples violate the independence hypothesis, leading to an optimistic bias classification. In this case, the assessed accuracy will be inflated. Furthermore, in this study, the heterogeneity principle may suffer in research due to the region of interest being mainly occupied by native forests and silviculture.

This paper presents a methodology that seeks the automated mapping of vegetation areas to detect the need for trimming, given the vegetation's height. Therefore, through remote sensing (fusion of the Radar Sentinel-1 and Sentinel-2 images), mapping of the vegetation located around the study area (TL Novo Horizonte) is generated, indicating the ideal moment of intervention so that the branches do not disturb the transmission system, damaging or causing interruption to the energy supply of a determined region. In a reality that requires travel to the location for inspection of thousands of kilometres of transmission lines to assess the need for vegetation removal – native or planted – the monitoring by remote sensors presents itself as a useful resource for increasing efficiency. In this context, a preliminary investigation on the possibility of aggregating details to the mappings obtained by supervised classification, achieved by the fusion of multispectral and radar data, is presented in this paper.

MATERIALS AND METHODS

TL - Novo Horizonte is situated in the State of Paraná and crosses the towns of Bocaiúva do Sul, Campina Grande do Sul and Tunas do Paraná. It operates on the tension of 138kV, having a safety distance determined by NBR 5422 of 15m on either side of its main axis. The environment in which this TL is located belongs to the Mata Atlântica biome, precisely in areas of contact between the Dense Montane Ombrophilous Forest physiognomy, characterized by uniform canopies around 20m high, and the Mixed Montane Ombrophilous Forest physiognomy, more heterogeneous, but with species of relevant size, such as Araucaria Angustifolia and ocotea porosa. These vegetation physiognomies can put the system at risk because of the contact with cables and towers.

Table 1: Classes mapped by MapBiomas in 2019 in TL Novo Horizonte's safety clearance

Soil Usage	Area (m²)	Area (%)
Forest Formation	1.262.447,23	81,15
Foresting	133.758,87	8,60
Agriculture and Pasture Mosaic	131.849,26	8,48
Pasture	22.283,38	1,43
Other non-Vegetated Areas	4.354,72	0,28
River, Lake and Ocean	806,46	0,05
Urban Infrastructure	167,66	0,01
Total	1.555.667,58	100,00

The Forest Formation, Foresting and Agriculture and Pasture Mosaic classes reach more than 98% of TL Novo Horizonte's safety clearance coverage.

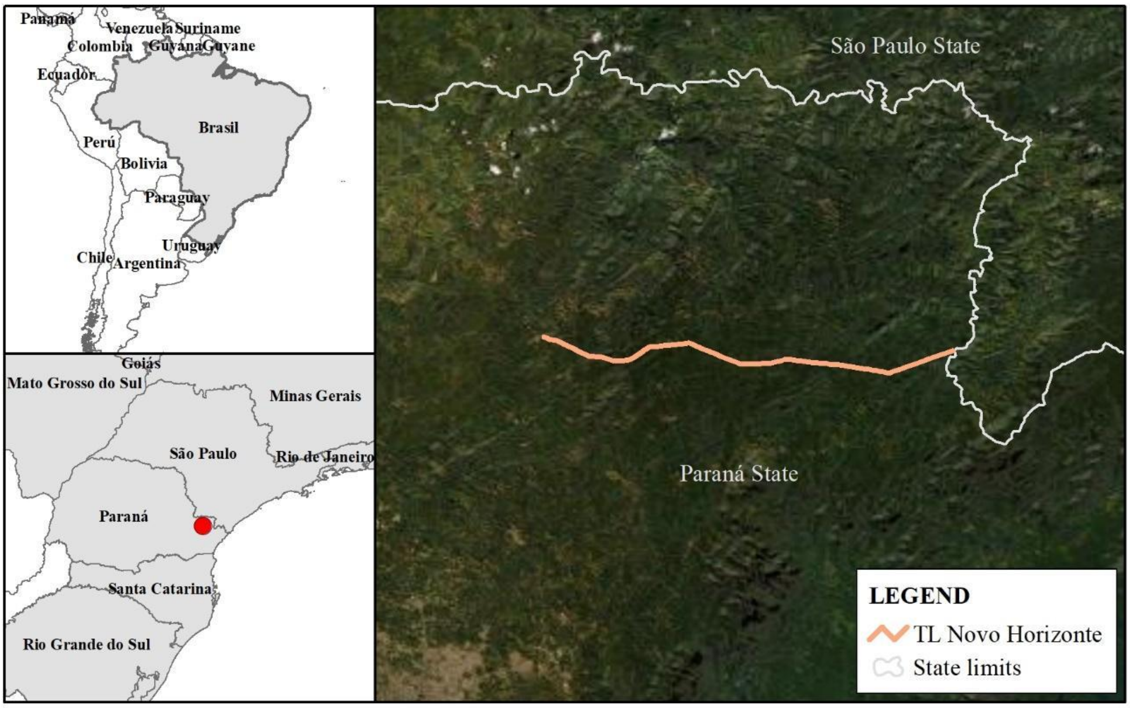

Figure 3: Location of TL Novo Horizonte.

UTILIZED TECHNIQUES

Taking into account the nature of the backscatter coefficient data, which has a central role in the analysis made in this paper, fusion methods based on pixel colours (or on their hue) were discarded. This exclusion is justified because these techniques aim to generate a new image, which may present visual characteristics that allow the visualization of imperceptible targets in ordinary compositions. Another point that justifies the limitation is that the hue attributed to the backscatter coefficient has no role other than mere visualization. In this study, the mapping was made by supervised classification and not by visual analysis of final data.

In this sense, PCA fusion was selected because taking the backscatter coefficient values into account offers solid computational performance for a statistical technique, on top of the possibility of reducing the dataset.

The PCA was done using the raster PCA tool of the RStoolbox package of the R programming language based on the covariances matrix (unpatterned PCA). The patterned PCA utilizes the correlation matrix, which requires pixel samples, for executing the princomp function native to the R environment, from which the prediction will extend to the entire raster. In the case of the unpatterned PCA, there isn't a need of samples and the covariance matrix will be calculated be used in the calculation of the princomp function and extended to the remainder of the raster.

The fusion of multispectral bands (Sentinel-2) and radar channels (Sentinel-1) was executed in the R environment on version 4.1.0, applying the Principal Component Analysis (PCA) method offered by the RStoolbox package.

For the classification, the RF algorithm was chosen because of the computational performance, ease of application and popularity, which can make the technique's replication simpler. The classification was also executed in the R environment, using the randomForest library.

UTILIZED DATASETS

Preliminarily all the waves provided by the Sentinel-2 platform, with the exception of bands B1, B9 and B10 – aerosol, water vapour (steam) and cirrus, were adopted as data of interest for the classification of multispectral images. In addition, to the Sentinel-1 data, the channels of backscatter coefficient used in the PCA fusion technique's application were VV and VH, available for the region of interest, and the ratio between VH and VV (VH/VV).

Table 2 presents details of the bands used from both platforms. The adoption of all multispectral bands which bring information on the Earth's surface is justified by the intention of aggregating the maximum amount of information throughout the classification process.

Table 2: Characteristics of the utilized data

Sensor	Spatial resolution (m)	Band Number	Band Name	Wavelength (nm)
Sentinel-2	10	B2	Blue	490
		B3	Green	560
		B4	Red	665
		B8	Near infrared	842
	20	B5	Red Edge 1	705
		B6	Red Edge 2	740
		B7	Red Edge 3	783
		B8A	Red Edge 4	865
		B11	SWIR 1	1610
		B12	SWIR 2	2190
	60	B1	Aerosol	443
		B9	Water Vapour	940
		B10	Cirrus	1375
Sentinel-1	5 x 20	VV	Co-polarized	3,8 to 7,5e+7
	5 x 20	VH	Cross-polarized	3,8 to 7,5e+7

The multispectral image was obtained utilizing its portal Copernicus and used in this paper. From the sensing date of the multispectral image, detailed in Table 3, an adequate period was defined for obtaining the radar image of the synthetic opening provided by the Sentinel-1 platform (Table 3).

Table 3: Characteristics of the utilized data

	Sentinel-2	Sentinel-1
Acquisition	09/02/2019	26/01/2019 to 23/02/2019
Satellite	A	A e B
Sensor	MSI	SAR-C
Processing Level	2A	GRD

That image was obtained using the Google Earth Engine service, formed by the averages of the pixel values from the sensed images between the dates of 28/09/2019 and 26/10/2019, a period which includes two weeks before and two weeks after the date of the multispectral image sensing.

Since this work compares classifications from different datasets, the process flow remains almost identical in both cases, as shown in Figure 4. However, the differences in processing are present when the analysis-ready data is obtained, therefore, in the raw data pre-processing, in the multispectral band and/or backscatter coefficient channel selection, and in the stacking or application of the PCA fusion technique.

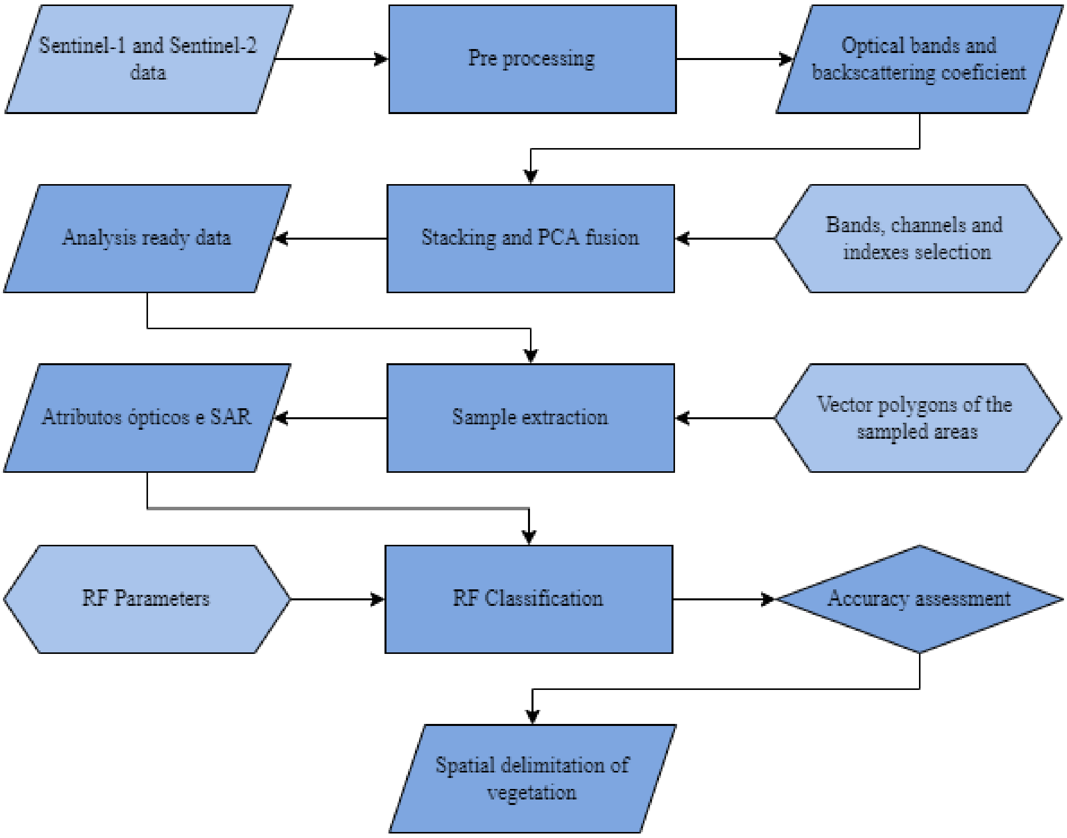

Figure 4: Flowchart of the technique

As soon as the data is ready for analysis, samples are extracted, which will be randomly partitioned for training and classification testing. For example, this investigation used a partitioning of 70% of samples going towards training and another partition (30% of samples) for testing.

CLASSES AND SAMPLES

The region of interest for this investigation, the surroundings of TL Novo Horizonte, is characterized by a predominance of native forest formations, silviculture and land usages which follow the dynamics of this kind of activity – soil and grass. The Native Forest, Foresting, Grass, Soil and Water classes were adopted to represent this coverage. This classification is representative enough of the phenomenon which is being mapped.

The training samples were obtained without the proportionality criterion concerning the presence of each class in the classified scene. A more representative number of samples was taken from the Native Forest and Foresting classes, while the other classes, which occupy smaller areas in the region of interest, were less represented, mainly grass and water. Due to this paper's objective being to compare classifications of multispectral and PCA fusion data, the samples used in both classifications are the same.

RESULTS AND DISCUSSIONS

Throughout TL Novo Horizonte, a considerable alternation was verified between the forest remnants and the silviculture areas, representing an intense dynamic in the change of soil usage. For example, in Figure 5 two towers located above an area of exposed soil caused by the removal of planted forestry can be observed.

Figure 5: Towers and safety clearance of TL Novo Horizonte over the RGB composition of the multispectral image (Sentinel-2).

This phenomenon brings with it two possibilities (inferences) which demand attention throughout their occurrences: the first is the maximum height that the planted forest must reach below the transmission line before its extraction, and the second is the possibility of soil erosion, depending on the exposure of the soil and the intensity of rainfall, as well as rifts and holes caused by tree stump removal, characterized by the mechanized removal of the stem of trees after their suppression.

In the top portion of Figure 6, which illustrates an RGB composition of multispectral data and a vertex of TL Novo Horizonte's layout, it is possible to note the presence of native forest fragments (lighter green), silviculture (darker green), as well as of soil (in the roads used for transporting wood). The multispectral data classification is shown in the same Figure's bottom portion.

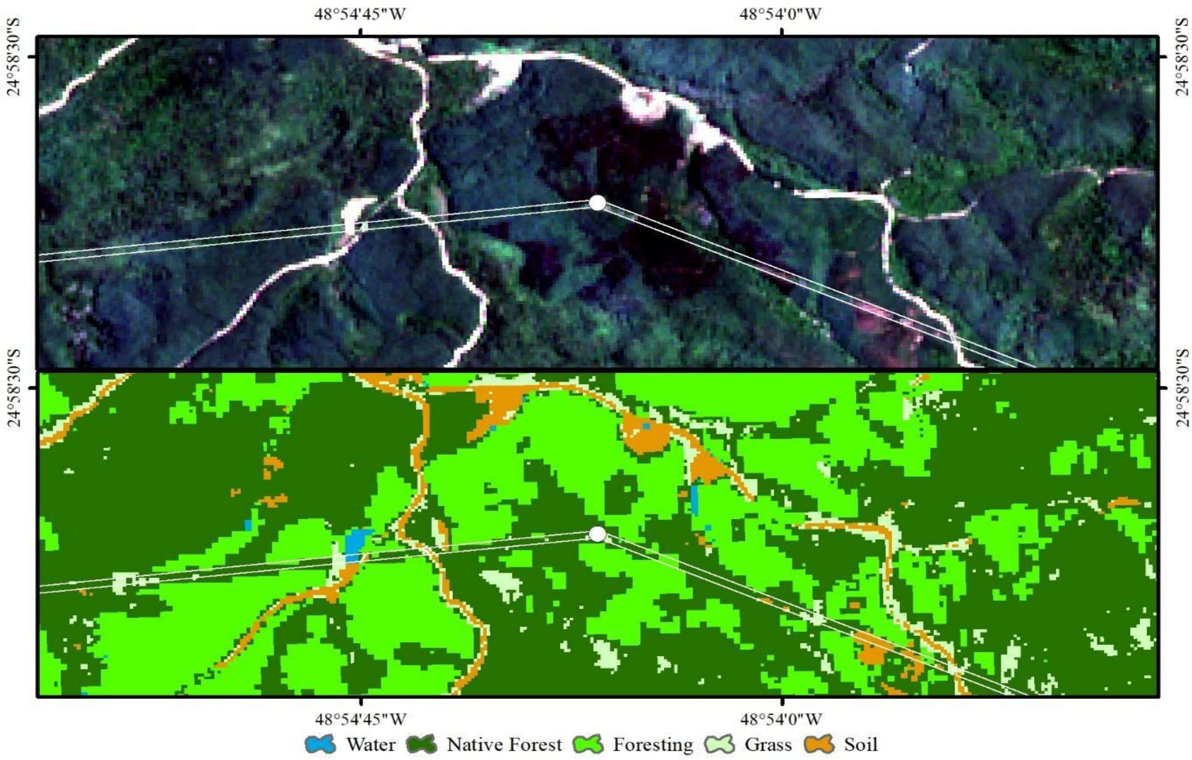

Figure 6: A Sentinel-2 (TCI composition) image and Random Forest Classification of the multispectral data of TL Novo Horizonte's surroundings, in 1:20.000 scale**.**

The multispectral data classification's confusion matrix expressed an overestimated accuracy, as shown in Table 4. This result is due to samples of Native Forest and Foresting classes having been considerably more numerous than the other classes. Another factor that justifies the development is that these classes comprise most of the region of interest.

Table 4: The confusion matrix of the multispectral data classification generated by the Random Forest algorithm.

	Water	**Native Forest**	**Foresting**	**Grass**	**Soil**	**User Accuracy**
Water	378	0	0	0	0	1,0000
Native Forest	0	1899	1	0	0	0,9995
Foresting	0	0	3362	0	0	1,0000
Grass	0	1	0	259	1	0,9923
Soil	0	1	0	0	210	0,9953
Producer Accuracy	1,0000	0,99894792	0,999702646	1,0000	0,9953	

Figure 7 shows the same multispectral image as Figure 6 in the upper portion – information that was used in the PCA fusion's processing – and the lower portion is the resulting data from the PCA fusion's processing by the Random Forest algorithm.

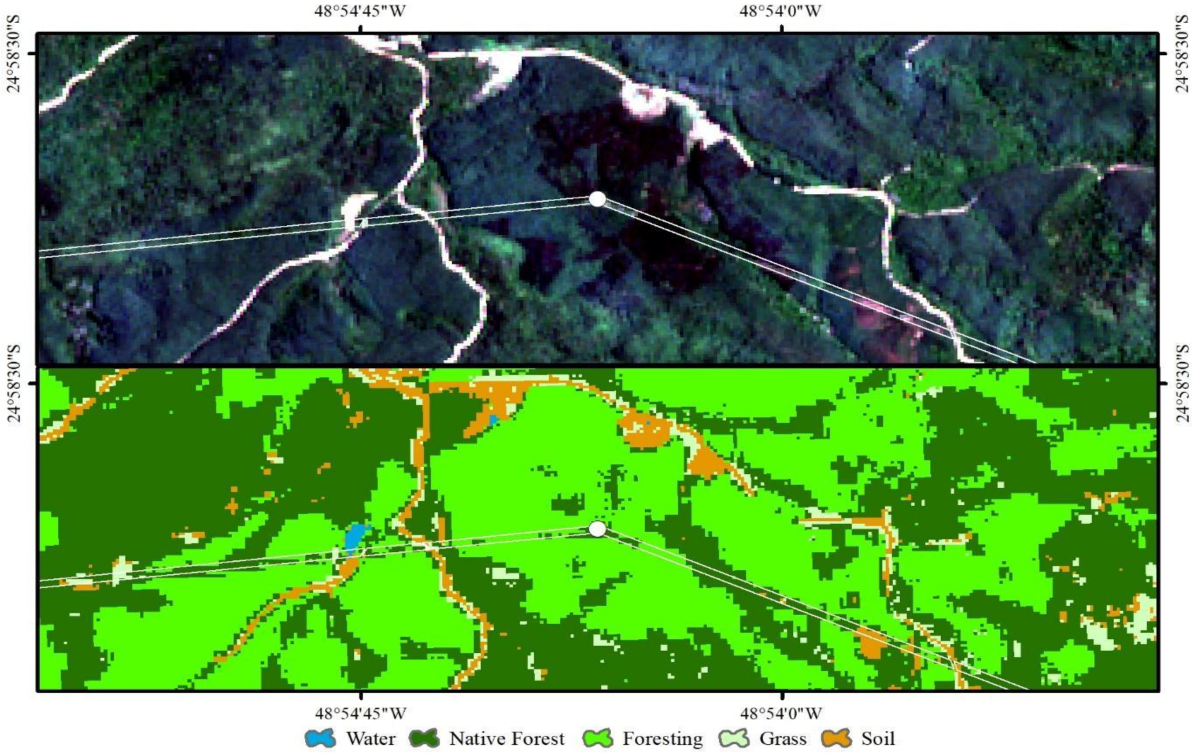

Figure 7: Sentinel-2 image (TCI composition) and Random Forest Classification of the fusion data (S1+S2) of TL Novo Horizonte's surroundings, scaled 1:20.000.

The confusion matrix of the PCA fusion data's classification (Table 5) expressed a behaviour similar to the confusion matrix shown in Table 4 due to the same vectorial polygons being used for the extraction of training samples for the multispectral data's classification.

Table 5: The confusion matrix of the multispectral data's classification generated by the Random Forest algorithm

	Water	**Native Forest**	**Foresting**	**Grass**	**Soil**	**User Accuracy**
Water	378	0	0	0	0	1,0000
Native Forest	0	1898	1	1	0	0,9989
Foresting	0	1	3361	0	0	0,9997
Grass	0	5	0	256	0	0,9808
Soil	0	2	0	0	209	0,9905
Producer Accuracy	1,0000	0,9958	0,9997	0,9961	1,0000	

Although there are weaknesses in the accuracy's assessment of the classifications made due to the samples, it was possible to infer substantial differences between them when it came to identifying classes and their outlines. Generally, the PCA fusion's data classification identified a more precise delimitation of the native forest class throughout the region of interest. Figure 8 shows three slices with "a" and "b" analysis points. The left slice represents the composition adapted for visual analysis. Point "a" marks the height of the remainder of the native forest preserved from West to East. In the middle slice, representing the multispectral data's classification, the area is more rudimentary in its delimitation and wider in the multispectral data's variety. In the proper slice representing the PCA fusion's data classification, the remainder classified is more faithful to the composition.

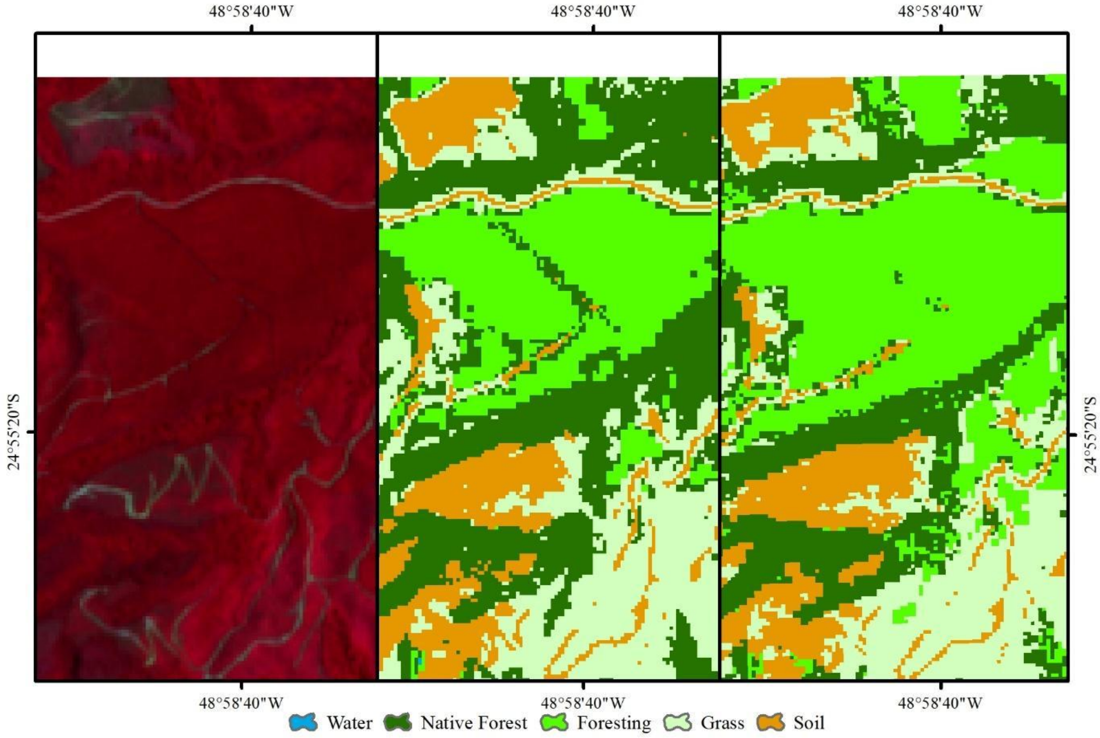

Figure 8: Sentinel-2 image (R composition: B8, G: B4, B: B3, left), Random Forest Classification of multispectral data (middle) and Random Forest Classification of fusion data (S1+S2) of TL Novo Horizonte's surroundings scaled 1:20.000.

On point "b" represented in Figure 8, on the left slice, it is possible to observe the presence of a road in a silviculture area. In the middle slice, which represents the multispectral data's classification, that same road appears with most pixels classified as native forest. The road is generalized and almost not outlined on the right slice – classification on PCA fusion's data. One more validation was done on a point denominated "c" for grass and soil classes' analyses, as illustrated in Figure 9. In the left slice (RGB composition of analyses) in the portion to the North and West of point "c", reddish colour is noticeable, which represents vegetation in the first phases of ecological succession, and a darker colour to the South and East of point "c", which refers to the soil. The right slice illustrates the classification obtained via the PCA fusion's data, which better represents the dynamics of occupation in the region of interest related to the suppression of the Foresting and Native Forest classes. Another point observed is that in the multispectral data's classification, grass is absent. In the PCA fusion's data classification, this class appears with greater fidelity.

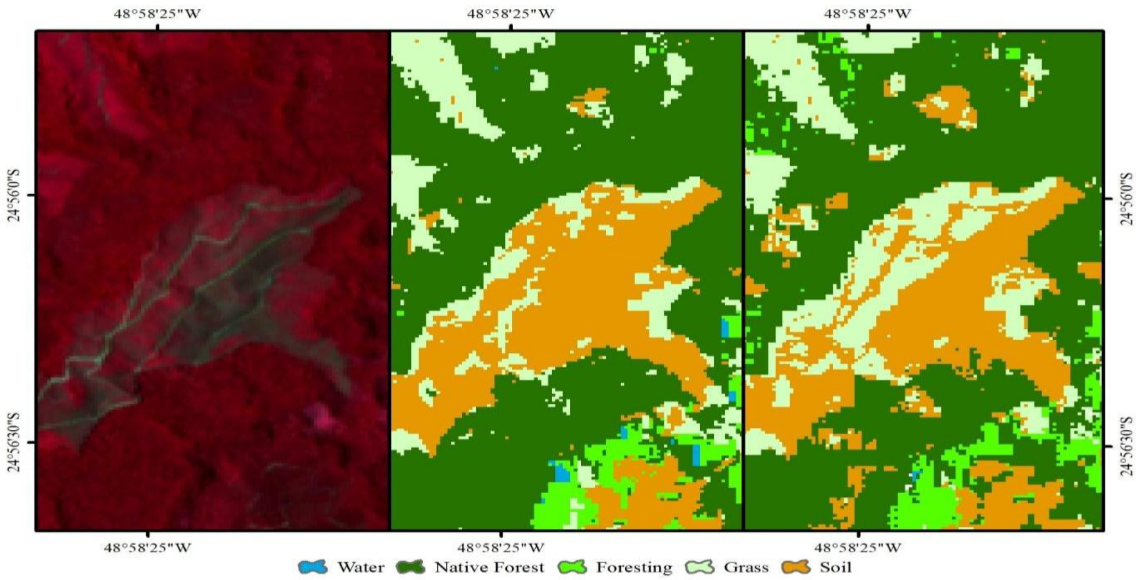

Figure 9: Sentinel-2 image (R composition: B8, G: B4, B: B3, left), Random Forest Classification of multispectral data (middle) and Random Forest Classification of fusion data (S1+S2) of TL Novo Horizonte's surroundings, scaled 1:20.000.

Throughout the multispectral data classification, there are areas classified as water, especially in places where the image has worse lighting. Generally, border regions of native forest or silviculture formations, or even regions where there is a sharp difference of vegetation height, as well as accentuated terrain unevenness.

Regarding the multispectral data classification, the PCA fusion classification's data (right, Fig. 10) shows much greater accuracy, significantly reducing this specific error.

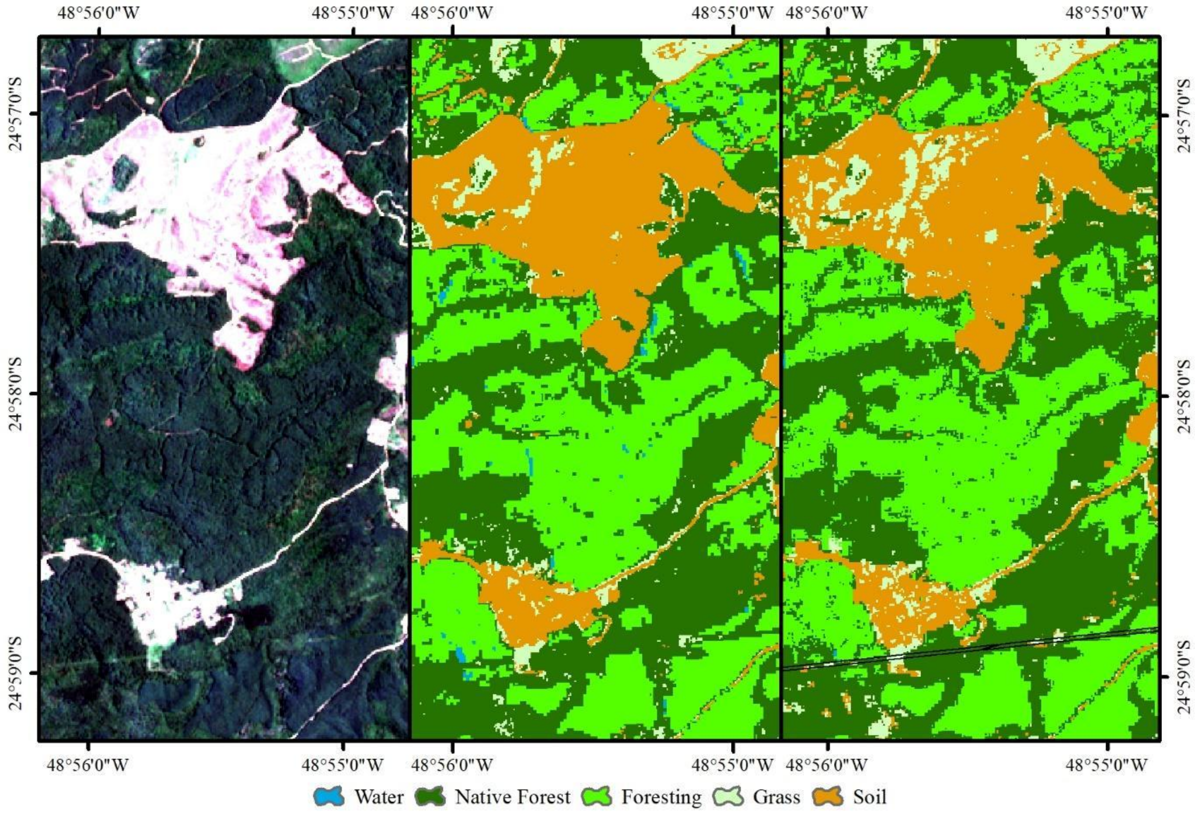

Figure 10: Sentinel-2 image (True Colour Image Composition, left), Random Forest Classification of multispectral (middle) and Random Forest Classification of fusion data (S1+S2) of TL Novo Horizonte's surroundings, scaled 1:50.000.

4. CONCLUSION

As a result of applying the algorithm to the two images (multispectral bands and principal components), satisfactory results were obtained concerning the definition of the mapped classes. When analyzing the results, it is noted that they were almost the same after applying the Majority Filter algorithm, intending to generalize the "salt and pepper" noise and the crude delimitations between the classes. However, a visual analysis indicates a higher detailing in the classification of the principal components' image, especially in the regions of borders between the forest remnants and the regions of planted forest. This fact may indicate an addition of texture characteristics to the sensed vegetation, suggested by the data from the Sentinel-1 platform, which eventually can bring an aggregation of quality in vegetation mapping activities if enhanced. Finally, applying the methodological procedure aimed at vegetation mapping by Principal Components of data provided by the Sentinel-1 and Sentinel-2 platforms shows itself to be promising at preserving the integrity of the TLs and recognizing the soil coverage with a higher degree of certainty. Therefore, for the preliminary stage of which this research, the main result is identified as the knowledge of the aggregation of detail by the PCA fusion's data, the supervised classification, executed using the Random Forest algorithm when compared to the multispectral data's category created by the same algorithm and using the same parameters. This brings a positive perspective to the research's continuation. Using all the multispectral bands and backscatter channels was adopted as the starting point for analyzing the classifier's answers. In the second stage, new combinations of spectral bands and backscatter channels, as well as indexes composed of this data, should be tested and evaluated empirically to obtain better classification results and to outline the remaining native and planted forests.

ACKNOWLEDGMENT

The authors would like to thank the CPFL group for technical and financial support through the Research and Development project PD-00063-3075/2020 (ANEEL/Brazil) via the Research and Development Program from the National Agency of Electric Energy (ANEEL). The authors also would like to thank Santa Catarina State University (UDESC) for supporting this project, especially GeoLab – Laboratory of Geoprocessing.

AUTHOR'S CONTRIBUTIONS:

The authors confirm contribution to the paper as follows: study conception and design: Hypolito Suarez Fernandez, Jessica Gerente, Francisco Henrique de Oliveira; data collection: Hypolito Suarez Fernandez, Jessica Gerente; analysis and interpretation of results: Hypolito Suarez Fernandez, Lucas Antonio Providelo, Guilherme Marcchiori; draft manuscript preparation: Hypolito Suarez Fernandez, Francisco Caruso Gomes Junior, Yang Liu. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest

SOFTWARE/DATA AVAILABILITY

Software: ESA SNAP version 8.0; R version 4.1.2 (libraries: caret, CaTools, dplyr, e1071, gdalUtils, randomForest, raster, reshape2, rgeos, rgdal, RStoolbox)

REFERENCES

Brazilian Association of Technical Norms (ABNT). (1985). Technical Norm 5422 – Transmission Lines.

Chen, L. *et al.* (2018) 'Estimation of forest above-ground biomass by geographically weighted regression and machine learning with sentinel imagery', *Forests*, 9(10). https://dx.doi.org/10.3390/f9100582.

Debastiani, A. B. *et al.* (2019) 'Evaluating SAR-optical sensor fusion for above-ground biomass estimation in a Brazilian tropical forest', *Annals of Forest Research*, 62(1), pp. 109–122. https://dx.doi.org/10.15287/afr.2018.1267.

Dymond, J. *et al.* (2019) 'Mapping Physiognomic Types of Indigenous Forest using Space-Borne SAR, Optical Imagery and Air-borne LiDAR', *Remote Sensing*, 11(16). https://dx.doi.org/10.3390/rs11161911.

Dyring, E. (1973). Principles of Remote Sensing. Ambio, 11(3), 57–69. https://doi.org/10.4324/9780203714522-9

Erinjery, J. J., Singh, M. and Kent, R. (2018) 'Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery', *Remote Sensing of Environment*, 216, pp. 345–354. https://dx.doi.org/10.1016/j.rse.2018.07.006.

Hirschmugl, M. *et al.* (2020) 'Use of SAR and optical time series for tropical forest disturbance mapping', *Remote Sensing*, 12(4). https://dx.doi.org/10.3390/rs12040727.

Institute for Applied Economic Research (IPEA). (2020). Contextualization of the Brazilian electrical sector and the infrastructure's planning in the long run, 69, pages 7-14.

Junior, W. de C. *et al.* (2016) 'Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas', *Pesquisa Agropecuaria Brasileira*, 51(9), pp. 1428–1437. https://dx.doi.org/10.1590/S0100-204X2016000900041.

Law #9.427, of 26 December 1996. Institutes the National Electrical Energy Agency – ANEEL, orders the regime for electrical energy service concessions and institutes other measures. Brasilia. 1996. Recovered on 23 January, 2022, from www.planalto.gov.br/ccivil_03/leis/l9427cons.htm

Law #9.648, of 27 May 1998. Alters rules of Laws #3.890-A, of 25 April 1961, #8.666, of 21 June 1993, #8.987, of 13 February 1995, #9.074, of 7 July 1995, #9.427, of 26 December 1996, and authorizes the Executive Power to promote the restructuring of the Brazilian Electrical Centres – ELETROBRÁS and its subsidiaries and institutes other measures. Brasília. 1998. Recovered on 20 January 2022, from https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=445096

Liu, Y. *et al.* (2019) 'Estimation of the forest stand mean height and above-ground biomass in Northeast China using SAR Sentinel-1B, multispectral Sentinel-2A, and DEM imagery', *ISPRS Journal of Photogrammetry and Remote Sensing*, 151, pp. 277–289. https://dx.doi.org/10.1016/j.isprsjprs.2019.03.016.

Millard, K. and Richardson, M. (2015) 'On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping', *Remote Sensing*, 7(7), pp. 8489–8515. https://dx.doi.org/10.3390/rs70708489.

National Electrical Energy Agency (ANEEL). (2019). Normative Resolution 846/2019. Approves procedures, parameters and criteria for the imposition of penalties on the electrical energy sector's agents and regulates on general directives of the Agency's monitoring. Brasilia. 2019. Recovered on 19 January 2022, from https://www.in.gov.br/web/dou/-/resolucao-normativa-n-846-de-11-de-junho-de-2019-164060539

Nikolakopoulos, K. G. (2008) 'Comparison of nine fusion techniques for very high resolution data', *Photogrammetric Engineering and Remote Sensing*, 74(5), pp. 647–659. https://dx.doi.org/10.14358/PERS.74.5.647.

Pal, S. K., Majumdar, T. J. and Bhattacharya, A. K. (2007) 'ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation', *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(5), pp. 281–297. https://dx.doi.org/10.1016/j.isprsjprs.2006.10.001.

Pohl, C. and Van Genderen, J. L. (1998) *Review article Multisensor image fusion in remote sensing: Concepts, methods and applications*, *International Journal of Remote Sensing*. https://dx.doi.org/10.1080/014311698215748.

Spracklen, B. and Spracklen, D. V. (2021) 'Synergistic use of sentinel-1 and sentinel-2 to map natural forest and acacia plantation and stand ages in north- central vietnam', Remote Sensing, 13(2), pp. 1–19. https://dx.doi.org/10.3390/rs13020185