Brazilian Cerrado native vegetation mapping with Google Earth Engine

Mapeamento da vegetação nativa do Cerrado com Google Earth Engine

Nayara Estrabis[1]*, Hemerson Pistori[1,2], Wesley Gonçalves[1], José Marcato Junior[1]

ABSTRACT

Cerrado is the second largest biome of South America, with the major extension located in Brazil. This biome is considered a world biodiversity hotspot due to the rich and important biodiversity contrasting with the high threat of destruction. Monitoring using remote sensing approaches is a crucial tool for maintaining and preserving this large-scale biome. Through this context, this study compared and assessed different scenarios with Landsat 8 OLI multispectral bands and Vegetation Indices (EVI, NDVI, and SAVI) for the Cerrado mapping in Mato Grosso do Sul state, Brazil. An amount of 512 sample polygons were distributed into 2 classes: native vegetation and non-native vegetation. The Google Earth Engine platform was applied to the training and classification processes using the Random Forest method. The results showed that using Landsat 8 OLI bands obtained better results than the vegetation indices, with an overall accuracy and kappa index of 97.08% and 0.94, respectively. The mapping verified the existence of 26,80% of Cerrado original native vegetation in Mato Grosso do Sul state in 2019.

Keywords: Brazilian Savanah; Vegetation indices; Landsat 8 OLI; Vegetation mapping; Machine learning

RESUMO

O Cerrado é o Segundo maior bioma da América do Sul, com maior extensão localizado no Brasil.Este bioma é considerado um hotspot mundial em biodiversidade devido sua rica e importante biodiversidade contrastando com a alta ameaça de destruição. O monitoramento usando abordagens de sensoriamento remoto são cruciais para sua manutenção e preservação. Neste context, este estudo comparou e avaliou diferente cenários com bandas multiespectrais Landsat 8 OLI e Índices de Vegetação (EVI, NDVI e SAVI) para o mapeamento do Cerrado no estado de Mato Grosso do Sul, Brasil. Uma quantidade de 512 polígonos amostrais foram distribuídos em 2 classes: vegetação nativa e vegetação não nativa. A plataforma Google Earth Engine foi aplicada aos processos de treinamento e classificação utilizando o método Random Forest. Os resultados mostraram que o uso das bandas Landsat 8 OLI obteve melhores resultados do que os índices de vegetação, com acurácia geral e índice kappa de 97,08% e 0,94, respectivamente. O mapeamento verificou a existência de 26,80% de vegetação nativa original do Cerrado no estado de Mato Grosso do Sul em 2019.

Palavras-chave: Cerrado; Índices de vegetação; Landsat 8 OLI; Mapeamento de vegetação; Machine learning

[1] Federal University of Mato Grosso do Sul

*E-mail: nayara.estrabis@ufms.br

[2] Dom Bosco Catholic University

INTRODUÇÃO

A biome is a large area with a microclimate defined by different characteristics such as distinct plants, animals, soil, clime, topography, phytophysiognomies, and ecology (COUTINHO, 2006). These regions are habitats for endemic species and ecosystems. However, the intensive exploration of natural resources, human activities, and non-planned urbanization has caused environmental deterioration. In addition, the consequences of these and other environmental impacts can be irreversible.

The Cerrado is a Brazilian biome located in most of the country's central region. Considered the second largest in South America and Brazil, its extension covers 23,9% (IBGE, 2022a; MMA, 2019) of the Brazilian territory, approximately 2 million km² (RATTER et al., 1997). It is distributed in 12 states from north to south of the country (Maranhão, Piauí, Tocantins, Goiás, Rondônia, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Bahia, Distrito Federal, São Paulo e Paraná) and there are incidences in 3 more states (Amapá, Roraima, and Amazonas).

This biome is also called and known as Brazilian Savanah. According to Coutinho (2006), the Cerrado is composed of Savanah areas and a complex mosaic from different types of other biomes, from campestral (cerrado limpo) to the forest (cerradão). These characteristics make the Cerrado a local with high biodiversity, varieties of soils, geomorphologies, vegetation, and climate.

One of these characteristics is biodiversity, composed of more than 4000 plants and vertebrate endemic species, 11627 native vegetation cataloged species (MARRIS, 2005; DURIGAN et al., 2007; STRASSBURG et al., 2017, FERREIRA et al., 2017; MMA, 2019). The Cerrado is essential for the hydric resources, being a water fountain for the three largest South American watersheds: Amazon Basin, Plata Basin, and San Francisco River Basin (MMA, 2019). These watersheds are also located in other biomes areas, showing their importance for other biomes.

The native vegetation is crucial for this biome and, consequently, for maintaining Brazil's water resources. The native vegetation in Cerrado has some deep roots in tree species. This characteristic acts as a sponge or water transport channel to recharge the aquifers, contrasting with short roots from crops or exposed soil, where water evaporates before reaching the aquifers (SCHOLZ et al., 2002; FELLET, 2017).

Although its great importance and rich biodiversity, Cerrado is considered a world biodiversity hotspot due to the existing threats (STRASSBURG et al., 2017; DURINGAN

et al., 2007; MARRIS, 2005). Expansions in the pasture, agriculture, burnings, and deforestation have been some factors of Cerrado native vegetation reduction, generating impacts on its ecosystem and biodiversity (BEUCHLE et al., 2015; PEREIRA; GAMA, 2010; DURINGAN et al., 2007; WARD et al., 1992). In the last 40 years, this biome has been the center of intensive agriculture and expansion pasture. These changes affected Cerrado's hydrologic balance (SPERA, 2016) and reduced native vegetation, generating negative consequences for the biome.

Zalles et al. (2018) identified an increase in agricultural areas and the reduction of tree cover in Cerrado after 2006. Beuchle et al. (2015) identified changes in vegetal cover in Cerrado and Caatinga between 1990 and 2010. The authors verified a continuous pattern of Cerrado native vegetation reduction, with an average annual variation of 0.6%. Furthermore, they detected that the quantity of native vegetation remaining is less than the number of other types of soil cover. Between 2001 and 2014, Cerrado presented a higher expansion in croplands, 52% of the total area of Brazil expansion (ZALLES et al., 2018).

The vegetal cover changing in Cerrado has been neglected compared to the attention given to the Amazonia (BEUCHLE et al., 2015; MARRIS, 2005). The Cerrado has the lowest rate of protected area than the other biomes. This situation becomes even more worrying with the possibility of the Cerrado not existing anymore in 2050, according to Strassburg et al. (2017).

Monitoring and controlling the activities in Cerrado areas are crucial for its preservation and conservation. The mapping of these areas is necessary to evaluate the conditions of the native vegetation, land use, and land cover. In this context, the use of machine learning techniques applied in imageries increased in the remote sensing domain (CASTRO et al., 2017; CRESSON, 2019).

Google Earth Engine (GEE) is one of the tools that enable us to perform large areas mapping and monitoring using machine learning methods, including Random Forest (RF), Support Vector Machine (SVM), CART, and Fast Naïve Bayes. GEE is a free and cloud platform (TSAI et al., 2018), enabling spatial data analysis that requires high computer performance in a fast processing and interactive way (GORELICK et al., 2017), besides allowing to edit and create algorithms (KUMAR; MUTANGA, 2018). This platform is considered one of the most significant advances in the Earth observation monitoring process (KUMAR; MUTANGA, 2018). It has been explored for land use and

land cover mapping large extensions due to its fast processing (DUONG et al., 2018; KUMAR; MUTANGA, 2018; NYLAND et al., 2018; TSAI et al., 2018).

Despite the advantages and possibilities provided by GEE, although it is increasing, there are few studies applying it in the mapping of Brazilian regions. There are studies in Semi-Arid (GOLDBLATT et al., 2017), mining areas in the Brazilian Amazon (LOBO et al., 2018), temperature analysis in a region of São Paulo (Bourscheidt, 2017), pasture areas in Brazil (Parente; Ferreira, 2018), studies in Pantanal region (PEREIRA et al, 2018) and Matopiba, a region in Cerrado (Guerra et al., 2017).

An online and free mapping platform was created using GEE, called MapBiomas v.7.0 (Projeto MapBiomas, 2022). This platform has produced annual maps of land cover in Brazil since 1985, with the last version and mapping referring to 2021. The classification is conducted using Landsat imageries and Random Forest (RF) algorithm.

The main aim of this study is to compare the use of Vegetation Indices and Landsat 8 OLI bands for Cerrado native vegetation mapping in Mato Grosso do Sul state. The study evaluation is important to find the best performance to map the vegetation characteristics and provide more accurate mappings. Also, the methodology can provide better information about Cerrado native vegetation, a threatened biome and so essential for the environment ecosystem and society.

MATERIALS AND METHODS

Study Area

The study area corresponds to the Cerrado in Mato Grosso do Sul state. Mato Grosso do Sul is a Brazilian state located in the middle west region. It has frontiers with important Brazilian economic regions. This state is composed of Pantanal, Atlantic Forest, and Cerrado biomes. The Cerrado biome area corresponds to 61% of its territory (see Figure 1) (IBGE, 2022a). In this region, the hydrography is composed of the Paraná watershed in the east and the Paraguay watershed in the west. These watersheds are divided by Serra de Maracajú, popularly known as the great water divisor of the state. The state has 5,112,932,61 hectares of environmental protection area (APA), and other conservation unities areas such as 148,347.30 ha of Private Natural Heritage Reserve and 327,027.06 ha of full protection conservation units (IMASUL, 2022).

The land use of Mato Grosso do Sul has more than 3 million hectares for crops (permanent, temporary, and flower crop), where the temporary crops represented the

major area and more than 16 million hectares for pasture, according to preliminary results from IBGE (2022b) for 2017. Also, it has approximately more than 6 million hectares of forest or natural vegetation, where natural vegetation is about 285 thousand hectares. A little more than 5 million hectares are designated for permanent preservation or legal reserve, and approximately 950 thousand hectares are planted forest. Approximately 385 thousand hectares has agricultural and forest system.

Among the municipalities of this state located in the Mato Grosso do Sul Cerrado biome, there are cities with intense production of soybean, sugar cane, corn, eucalyptus (also, international reference), and cattle. These characteristics allowed Mato Grosso do Sul to become one of the most agricultural-producing states in Brazil. According to the state government website Portal Governo do Estado de Mato Grosso do Sul (2019), the state has twelve municipalities in the ranking of the 100 most agricultural producers' cities in Brazil.

Figure 1: Cerrado biome location in Mato Grosso do Sul.
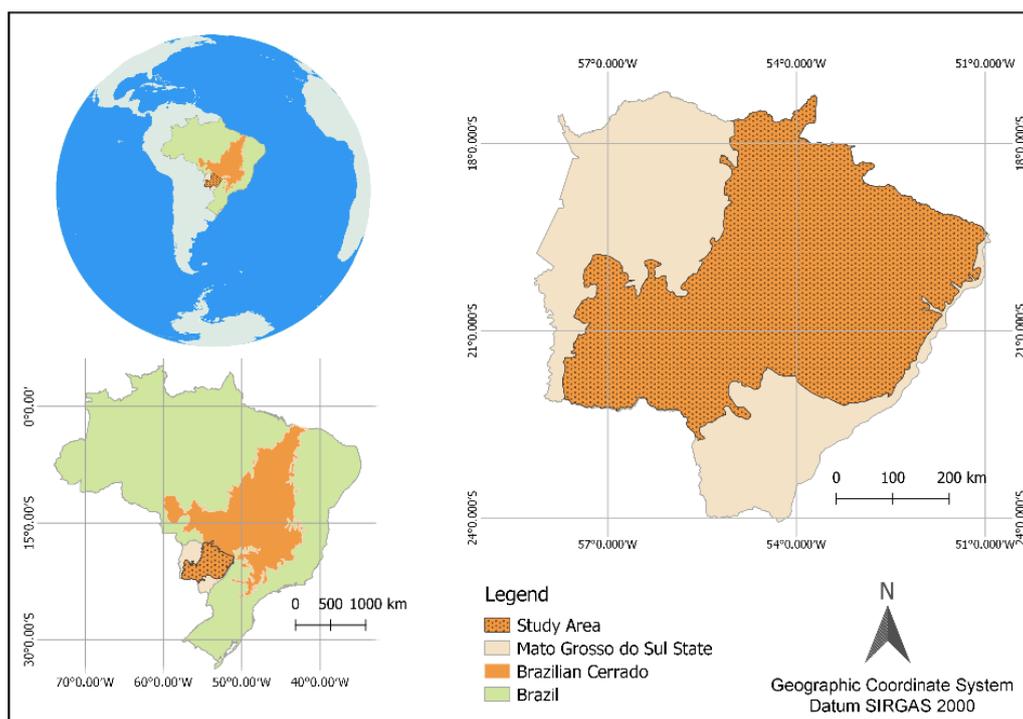
Source: The authors, 2022.

Image Dataset

Landsat satellite collection has the goal of observing natural Earth resources. Landsat 8 is the last version of this satellite collection active. Its characteristics are high accuracy positional (IRONS et al., 2012; ROY et al., 2014 and STOREY et al., 2014);

time of revisit of 16 days; 30 meters spatial resolution; multispectral bands (Table 1); near infra-red (NIR) band, two on short wave infra-red (SWIR) band and; orthorectified image (USGS, 2018). The Landsat 8 OLI bands selected for this study were bands 2 to 7, shown in Table 1.

On the GEE Platform, we used images from different days and months of the first semester of 2019. The data range selected between February 1, 2019, and July 30, 2019, and cloud cover of less than 2% provided a better mosaic composition. The amount of 58 images that represented 348 bands composed the Landsat 8 OLI mosaic.

Table 1: Landsat 8 OLI spectral bands characteristics.

Bands	Name	Spectral (µm)
2	Blue	0.452 – 0.512
3	Green	0.533 – 0.590
4	Red	0.636 – 0.673
5	NIR	0.851 – 0.879
6	SWIR-1	1.566 – 1.651
7	SWIR-2	2.07 – 2.294

Source: adapted from USGS (2018).

Classification Process

Classes Identification and Samples Delimitation

Two classes were considered in the classification process: 1. native vegetation and 2. non-native vegetation. For the native-vegetation classes, we considered areas composed of the canopy of trees. The non-native-vegetation area comprises pasture, agriculture, watercourses, silviculture, and urban areas.

The characterization of classes allowed the sample delimitation to compose the dataset. A total of 512 sample polygons were split into two groups: 340 (66%) samples for training, 170 per class, and 172 (34%) for validation, 86 per class. The location of the samples is shown in Figure 2: training (yellow) and validation (red) samples. The training samples are in a different place than the validation.

Figure 2: Samples Location in Study Area
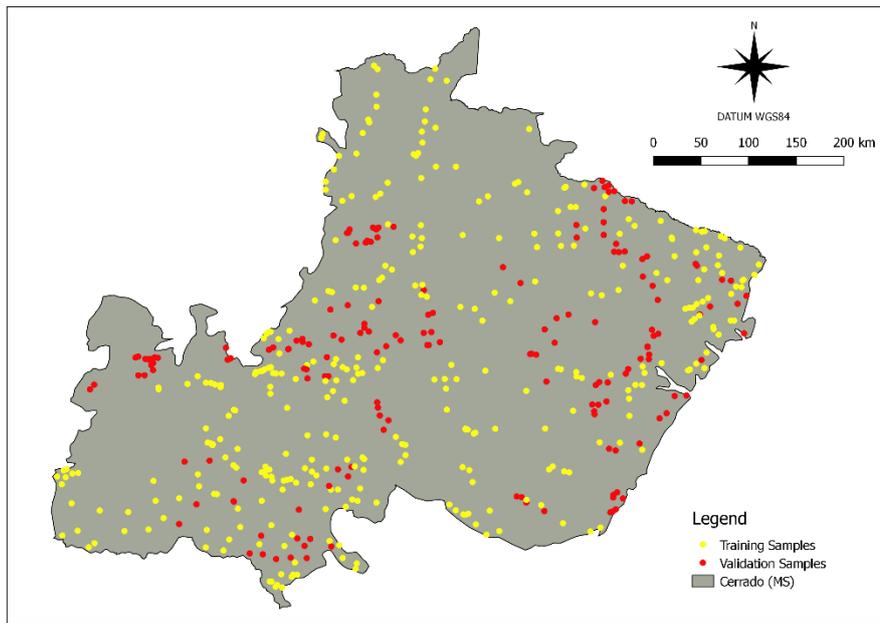
Source: The authors, 2022.

The polygons were drawn by hand on GEE platform, with different sizes and shapes. The training polygons samples cover an average of 0.8 km² per sample, and the validation polygons cover an average of 1.35 km² per sample. In Table 2, it is possible to compare the covered area of the polygons and also the number of pixels.

Table 2: Total areas and pixel quantity of the training and validation samples

Class	Training		Validation	
	m²	pixel	m²	pixel
Native – Vegetation	136,045,410	151,161.60	73,759,243.03	81,954.71
Non–native vegetation	229,102,270.02	254,558.08	70,847,010.10	78,718.90

Scenarios

Several scenarios were considered combining the original bands and vegetation indices. The following vegetation indices were considered in the classification process: Normalized Difference Vegetation Index (NDVI) proposed by Rouse Jr (1974), Enhanced Vegetation Index (EVI) proposed by Huete et al. (2002), Soil Adjusted Vegetation Index (SAVI), proposed by Huete (1988).

The NDVI (Equation 1) is a numerical index widely used for vegetation detection obtained through the NIR and Red bands (Table). The index value varies from -1 to 1, where the higher value obtained means the reflection of the area is more detectable in the NIR band as dense vegetation.

$$NDVI = \frac{NIR - \text{Red}}{NIR + \text{Red}} \tag{1}$$

The EVI vegetation index, although similar to NDVI, presents the calculation of corrections coefficients, adjustments variables, and the use of the Blue band that represents atmospheric corrections and canopy background noise corrections (VERMONT et al., 2016). Also, the values for the coefficients and variables are $G = 2.5$, $C_1 = 6$, $C_2 = 7.5$ and $L = 1$. The corresponding equation is presented below (Equation 2).

$$EVI = G \frac{NIR - \text{Red}}{NIR + C1 * \text{Red} - C2 * \text{Blue} + L} \tag{2}$$

where: G is the gain factor, C1 and C2 are resistance atmospheric correction coefficients, and L is the adjustment for the canopy background.

The SAVI vegetation index (Equation 3) reduced the soil brightness influence in low-cover vegetation areas. This correction is represented by the coefficients: $L = 0.5$, as suggested by USGS (2019), to adhere to more land cover types.

$$SAVI = \frac{NIR - \text{Red}}{NIR + \text{Red} + L} * (1 + L) \tag{3}$$

where: L is the adjustment for the soil brightness correction factor.

The scenarios combining the indices and Landsat OLI spectral bands are described in Table 3.

Table **3**: Experimental Scenarios

Scenarios	Bands
I	Spectral bands (2 to 7)
II	NDVI
III	EVI
IV	SAVI
V	Scenario I and II
VI	Scenario I and III

VII	Scenario I and IV
VIII	Scenario I to IV

Image classification based on Machine Learning

The dataset explained in the section "Image Dataset", is composed of training and validation samples, where the polygons samples for training are in different locations than the validation samples. This differentiation is necessary to avoid an inducted result. The samples for the machine learning method were selected in the GEE platform using the satellite image.

In GEE, we used Random Forest (RF) algorithm to perform the training and classification. The RF algorithm is a predictor tree set, where each tree has an independent random vector, and the information selection occurs by choosing the most popular class for determining data (BREIMAN, 2001). The platform provides a classifier pack, where RF was considered with the following configurations: number of trees of 100 and 10 for the minimum size of a terminal node. According to the considerations of Breiman (2001) about the number of trees and some tests performed by Estrabis et al. (2019), we decided to adopt 100 trees for the RF classification.

Validation

In the validation process, we compared the classified data with the reference data, where it was possible to obtain values of agreements and disagreements, called the confusion matrix. This process was performed using QGIS software (QGIS Development Team, 2019), analyzing the confusion matrix and estimating the kappa indexes and accuracies. The kappa index proposed by Cohen (1960) is an evaluation of the degree of agreement of the classes, as shown in Equation (4). The interpretation of the agreement's strength of the Kappa index was established as proposed by Landis and Koch (1977). Another validation metric is accuracy. The probability of a sample being correctly classified is called general accuracy (GA), estimated using Equation (5). The producer accuracy (PA), in Equation (6), shows how well an area can be mapped on the Earth and indirectly indicate the omission error. The user accuracy (UA) in Equation 7 is associated with the reliability of the map and shows how well the map represents what something is on the land; indirectly, it indicates the commission errors (STORY; CONGALTON, 1986).

$$K = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

Where: Po is the agreement proportions, and Pe is the agreement proportions expected by chance.

$$GA = \frac{\sum_{i=1}^{a} cs}{\sum_{i=1}^{a} rs} \tag{5}$$

$$PA = \frac{cs}{rs} \tag{6}$$

$$UA = \frac{cs}{as} \tag{7}$$

Where: "cs" is the number of correct samples, "rs" is the total number of reference samples, "a" is the total number of classes, and "as" is the total amount of samples classified in the class.

The producer and user accuracy values indicate indirect commission and omission errors. Producer accuracy is associated with commission errors, where, according to Souza and Sampaio (2005), classes are mistakenly included in other classes during the classification. The user accuracy is related to omission errors, where classes are omitted or forgotten during the classification.

RESULTS AND DISCUSSIONS

All the classifications based on the scenarios were performed on GEE, and the maps of the classifications for each scenario are shown in Figure 3. The green color represents the native-vegetation areas, and the yellow color represents the non-native-vegetation areas.

Figure 3: Comparison of the scenario maps. Scenario I: Landsat OLI bands, scenario II: NDVI, scenario III: EVI, scenario IV: SAVI, scenario V: NDVI + Landsat OLI bands, scenario VI: EVI + Landsat OLI bands, scenario VII: SAVI + Landsat OLI bands, scenario VIII: Lansat OLI bands + NDVI + EVI + SAVI.
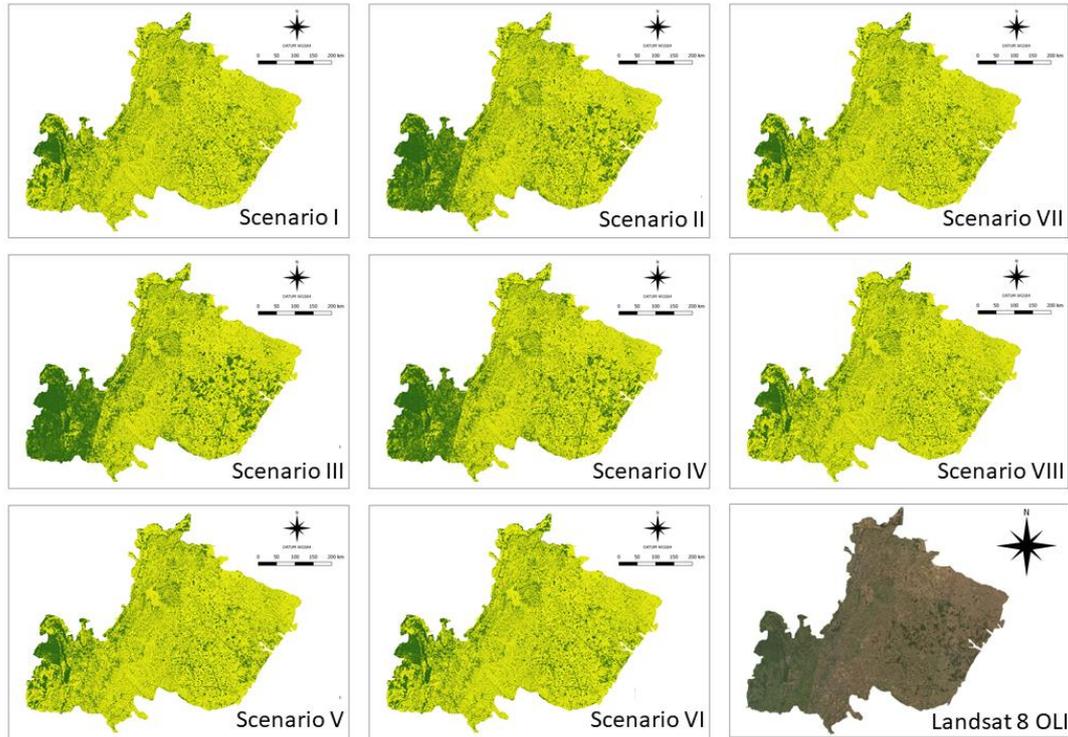
The confusion matrix was estimated in percentage, and the results are shown in Table 4. The diagonal represents the target or the agreements between classification and the validation (in bold), and the other values represent the confusion.

Table 4: Scenarios Confusion Matrix

			Reference	
		Class	1	2
Classified	Scenario I	1	**98.84**	4.76
		2	1.16	**95.24**
	Scenario II	1	**97.47**	12.11
		2	2.53	**87.89**
	Scenario III	1	**99.54**	14.89
		2	0.46	**85.12**
	Scenario IV	1	**97.46**	14.88
		2	2.54	**85.12**
	Scenario V	1	**98.85**	4.78
		2	1.15	**95.22**
	Scenario VI	1	**98.81**	4.76

		2	1.18	**95.24**
	Scenario VII	1	**98.82**	4.74
		2	1.18	**95.26**
	Scenario VIII	1	**97.46**	14.88
		2	2.54	**85.12**

The classifications provided correctly at least 97% of native-vegetation classified as native-vegetation, and less than 3% as non-native-vegetation. The scenario III (EVI) presented the best result for class 1 – native vegetation. Although the values were around 97%, the scenarios II, III, IV, and VIII presented the lowest agreement for class 2, non-native vegetation. Regarding class 2, differences of around 10% occurred in non-native-vegetation classification. The scenarios II, III, IV, and VIII presented classifications with more confusion among other scenarios, with almost 15% of confusion with the native-vegetation class. In general, comparing the agreement results (Table 4) for all scenarios, scenario VIII generated the lowest results. The classification image from EVI obtained the highest agreements for the native-vegetation class and one of the lowest agreements for the non-native-vegetation class, comparing the results for the confusion matrix.

In Table 5 is shown the values for the producer's accuracy, user's accuracy, and kappa index for each class in each scenario, and overall accuracy for each classification per scenario.

Table 5: Classifications Accuracies and Kappa Index

Scenario	Producer's Accuracy (%)		User's Accuracy (%)		Kappa Hat	
	Class		Class		Class	
	1	2	1	2	1	2
I	98.84	95.24	95.58	98.75	**0.91**	0.98
II	97.47	87.89	89.34	97.09	0.78	0.94
III	**99.54**	85.12	87.44	**99.44**	0.74	**0.99**
IV	97.46	87.90	89.34	97.08	0.78	0.94
V	98.85	95.22	95.56	98.75	**0.91**	0.98
VI	98.81	95.24	95.58	98.72	**0.91**	0.97
VII	98.82	**95.26**	**95.59**	98.73	**0.91**	0.98
VIII	98.84	95.15	95.50	98.74	**0.91**	0.98
	Overall Accuracy				Kappa Index	
I	**97.077**				**0.9415**	
II	92.775				0.8552	
III	92.472				0.8489	
IV	92.774				0.8552	
V	97.069				0.9413	
VI	97.063				0.9412	

| VII | 97.074 | 0.9414 |
| VIII | 97.032 | 0.9406 |

The classification for the scenario I presented accuracies of producer and user above 95%. These results showed a good mapping of the classes (producer's accuracy) and reliability, achieving almost 100% for all classes. Also, these values indicated commission and omission errors of less than 5%. The kappa hat achieved agreements above 0.90, reaching almost 1 for class 2, considered almost perfect agreement.

The scenario II represents a classification using the NDVI. The results for this scenario showed agreements with a slight reduction compared to scenario I in the producer's accuracy for class 1 and the user's accuracy for class 2, a reduction of 7.35% for class 2 in the producer's accuracy, and 11.34% for class 1 for user's accuracy. It is possible to observe an increase in confusion in non-native vegetation classification, classified as native vegetation, which indirectly indicates an increase in omission errors for class 2. Also, the reduction of class 1 in user's accuracy increased the commission errors and the reduction in mapping reliability indirectly. The kappa hat for class 1 was reduced compared with scenario I, considered as substantial agreements and almost perfect agreements for class 2.

The EVI classification, scenario III, presented the highest result for the producer's accuracy for class 1 and class 2 in the user's accuracy. However, this classification obtained the lowest result for the class 2 producer's accuracy and class 1 in the user's accuracy. These observations indicate that the classification provided well detection of the native-vegetation class, however, with lower reliability. For class 2, although the classification provides higher reliability, it did not represent so well what it is on Earth. The kappa hat resulted in the lowest value for native-vegetation among, considered as substantial agreements and almost perfect agreements for non-native-vegetation.

The SAVI classification, scenario IV, performed similarly to scenario II. The producer's accuracies showed better classification of the native-vegetation class than the non-native vegetation class, and the user's accuracy demonstrated better reliability for class 2 compared to class 1. The values indicated more omission errors for class 2, around 12.1%, and less than 3% for class 1. Also, the commission errors were higher for class 1, reaching 10.66%, and less than 3% for class 2. The kappa hat indicated more disagreement for class 1 and more agreement for class 2, considered "substantial agreements" and "almost perfect agreements", respectively.

The combinations between the vegetation indices and Landsat bands improve the classifications by only using the vegetation indexes. The scenario V obtained a considerable increase, around 5%, in the producer's accuracy for class 2 and an increase in the user's accuracy for class 1, representing greater class identification and mapping reliability. The kappa hat showed more agreements for class 1 than only the NDVI classification. Also, these improvements can be observed in scenario VII, with results similar to scenario V and the same kappa hat values. However, in scenario VI, occurred a slight reduction in the producer's accuracy for class 1 and the user's accuracy for class 2. Also, in scenario VI were observed considerable improvements for other accuracies, around 10%, generating better classification reliability. The kappa hat performed with a similar result, all values above 0.97 for class 2 and 0.91 for class 1, were considered "almost perfect agreements".

The scenario VIII, performed with the combination of all scenarios, generated results similar to scenario I (without the vegetation indices). These values indicated that using all vegetation indices associated with Landsat bands did not significantly improve the classification performed in the scenario I (only Landsat bands). Also, the results were similar to the classifications where the vegetation indices were included with the Landsat bands for each vegetation index (scenarios V, VI, and VII), resulting in similar values. The producer's and user's accuracies maintained a good classification of what is and mapping reliability above 95%. The kappa hat indicated "almost perfect agreements" for both classes, with confusion under 1%.

The vegetation indices images are shown in Figure 4, where it is possible to observe how the different elements are identified in the vegetation indices evaluated in this study. Class 2 was specified in different samples that constitute the classes water, pasture areas, agriculture, urban, and silviculture areas.

Different from the other vegetation indices (NDVI and SAVI), the water in EVI classification presented as a class that has clearly vegetation in its composition. It can be occurred because of the presence of eutrophication in the sample collected from water bodies, as we can see in 4.c.EVI.

Figure 4: Samples and vegetation indices visualization
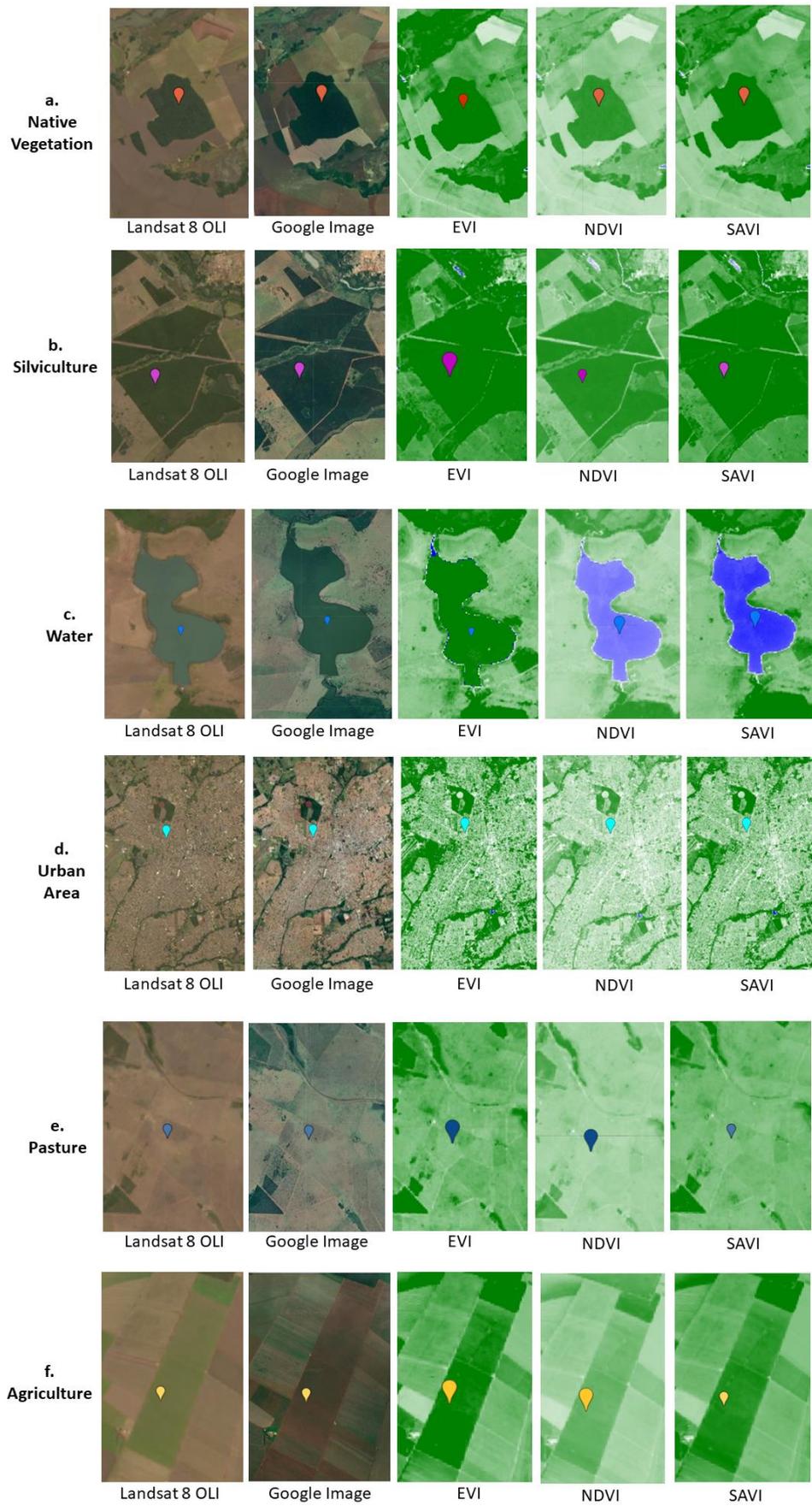
This interpretation occurred with EVI, presented in 4.c, of eutrophic waters, which changed the perception of the water to something with vegetation. However, although for the other indices, this water sample was identified as a different element of vegetation. The atmospheric corrections performed by the calculation of EVI could turn more real the identification of the presence of vegetation in diverse areas, in other words, more sensible to vegetation presence, however, can increase the probability of any confusion of the algorithm classifier. The urban and pasture areas samples are presented in vegetation indices as light green or near-white color, visually different from native vegetation, silviculture, or agriculture samples.

Figure 5: Scenarios comparison
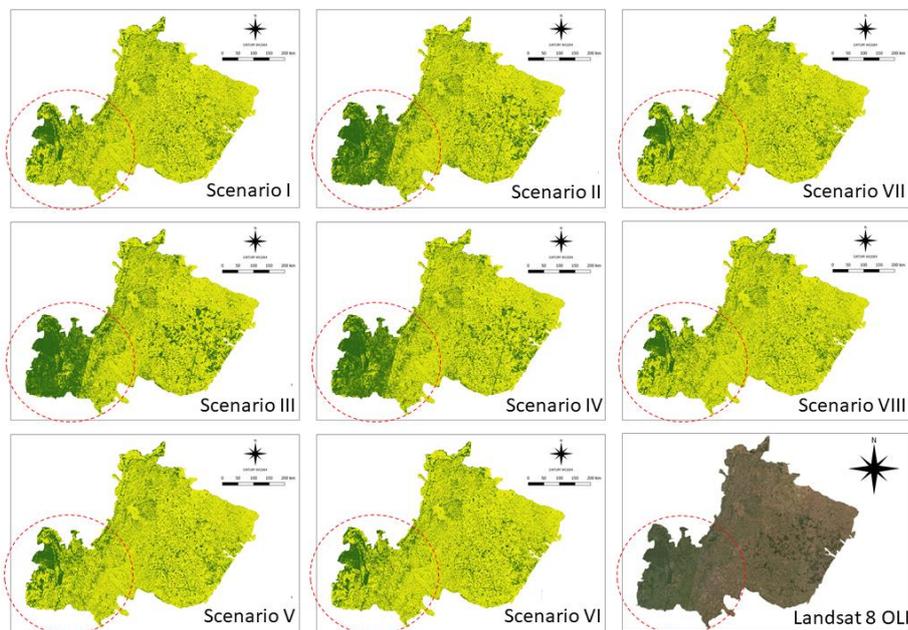
All scenarios results are presented in Figure 5. The classification only considering vegetation indices generated results slightly different among them. An area (in red circle) was identified where different results were obtained. Visually on Landsat images, this specific area has coloration greener than other areas, e.g., pasture areas, and this characteristic could influence the vegetation indices generation, which has the aim to highlight the vegetation and improve its detection. Then, when it was highlighted by the vegetation indices, the RF algorithm classified it as native vegetation and not as non-native vegetation (e.g., pasture areas).

The balance among the classification was observed when these vegetation indices were associated with Landsat bands, reducing the confusion. One of the reasons for this

improvement could be that the RF algorithm can decide better when all sets of other spectral bands are used than only the vegetation indices. Consequently, a better classification was achieved. Also, the presence of the other elements with vegetation in their composition could contribute to the confusion of the classifier.

Finally, the validation results indicated the scenario I was the best scenario for the native vegetation mapping using Landsat 8 OLI bands 2 to 7. The map of the Cerrado native vegetation in the first semester of 2019 in Mato Grosso do Sul state is presented in Figure 6.

Figure 6: Cerrado Native Vegetation Mapping in Mato Grosso do Sul
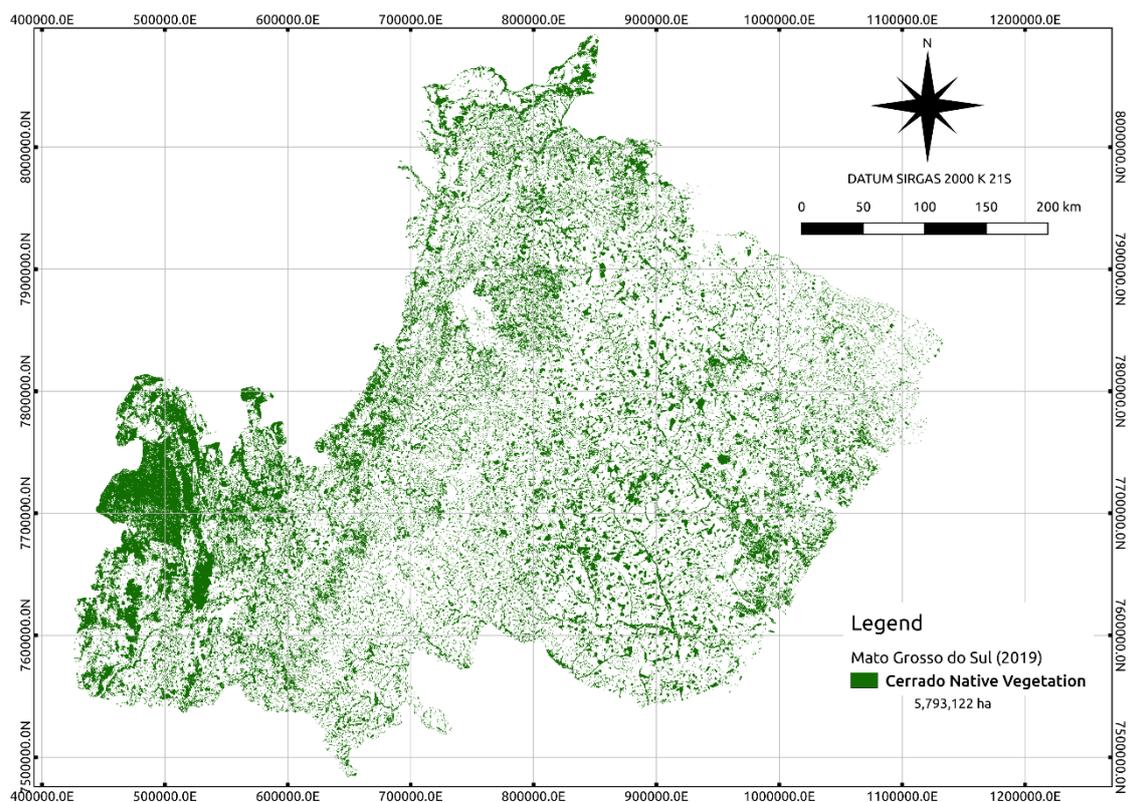
The mapping results indicated that the Cerrado native vegetation in Mato Grosso do Sul covered 5,793,122 ha, and the non-native vegetation corresponded to 15,847,258 ha. The total area mapped was 21,640,380 ha, demonstrating that the presence of the Cerrado biome in this state for 2019 was 26,80%.

CONCLUSIONS

The GEE can generate native vegetation mapping with high accuracies using only Landsat spectral bands or NDVI, EVI, and SAVI, and in addition, these bands combined

with each other. The applications of vegetation indices and spectral bands from Landsat 8 in different scenarios indicate that the use of bands 2 to 7 from Landsat 8 OLI provides the best classification of Cerrado Native Vegetation mapping. Classifications using only vegetation indices present more confusion, reducing the reliability of the native vegetation mapping. The EVI classification is sensitive for eutrophic water sources inducing a wrong classification as dense vegetation, and pasture areas can classify as dense vegetation.

In Mato Grosso do Sul, the presence of the Cerrado Native Vegetation mapping for the first semester of 2019 was 5,793,122 ha, representing 26,80% of this biome area in this state.

ACKNOWLEDGMENT

REFERENCES

BEUCHLE, R.; GRECCHI, R. C.; SHIMABUKURO, Y. E.; SELIGER, R.; EVA, H. D.; SANO, E.; ACHARD, F. Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach. **Applied Geography**, v. 58, 2015, pp. 116–127.

BOURSCHEIDT, V. Análise de tendência da temperatura de superfície a partir de imagens Landsat 5: contribuições da plataforma Google Earth Engine. **Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto SBSR**. Santos, 2017. pp.7401-7407.

BREIMAN, L. Random Forests. **Machine Learning**, vol. 45, 2001. pp. 5-32.

CASTRO, J. D. B.; FEITOZA, R. Q, ROSA, L. C. L., DIAZ, P. M. A.; SANCHES I. D. A., A Comparative Analysis of Deep Learning Techniques for Sub-Tropical Crop Types Recognition from Multitemporal Optical/SAR Image Sequences. *30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI).* Niteroi, 2017, pp. 382-389.

COHEN, J. A Coefficient of Agreement for Nominal Scales. **Educational and Psychological Measurement**, v. 20, 1960, n. 1, pp. 37–46.

COUTINHO, Leopoldo Magno. O conceito de bioma. *Acta Bot. Bras*. 2006, vol.20, n.1, pp.13-23.

CRESSON, R. A Framework for Remote Sensing Images Processing Using Deep Learning Techniques. **IEEE Geoscience and Remote Sensing Letters**. vol. 16, no. 1, 2019. pp. 25-29.

DUONG, P.; TRUNG, T; NASAHARA, K.; TADONO, T. JAXA High-Resolution Land Use/Land Cover Map for Central Vietnam in 2007 and 2017**. Remote Sensing**., v. 10, 2018, n. 9, pp. 1406.

DURIGAN, G.; SIQUEIRA, M.F de; FRANCO, G.A.D.C. Threats to the Cerrado remnants of the state of São Paulo, Brazil. **Scientia Agricola**, v. 64, 2007, n. 4, pp. 355–363.

FELLET, J. Como as raízes do Cerrado levam água a torneiras de todas as regiões do Brasil. BBC Brasil. 2017. Site <https://www.bbc.com/portuguese/brasil-39391161>, acessado em janeiro de 2019.

FERREIRA, F.G.; MACHADO, E. L. M.; SILVA NETO, C. DE M. E ; SILVA JÚNIOR, M. C.; MEDEIROS, M. M.; GONZAGA, A. P. D.; SOLÓRZANO, A.; VENTUROLI, F.; FAGG. J. M. F. Diversity and indicator species in the cerrado biome, Brazil. **Australian Journal of Crop Science AJCS**, v. 11, 2017, n. 08, pp. 1042-1050.

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.

GUERRA, J. B.; SCHULTZ, B.; SANCHES, I. D. Mapeamento automático da expansão da agricultura anual no Matopiba entre 2002 e 2015 utilizando a plataforma Google Earth Engine. **Anais do XVIII Simpósio Brasileiro de Sensoriamento Remoto SBSR.** Santos, 2017. pp.6850-6857.

HUETE, A.R . A Soil Adjusted Vegetation Index (SAVI). **Remote Sensing of Environment**, vol. 25,1988, no. 3, pp. 295-309.

IBGE, Instituto Brasileiro de Geografia e Estatística, Brasil em síntese: território. 2022a, Available from < https://brasilemsintese.ibge.gov.br/ territorio.html>. Accessed in 2022.

IBGE, Instituto Brasileiro de Geografia e Estatística, Mato Grosso do Sul. 2022b. Available from <https://cidades.ibge.gov.br/brasil/ms/pesquisa/24/76693>. Accessed in 2022.

IMASUL, Instituto do Meio Ambiente de Mato Grosso do Sul. Unidades de Conservação Municipais. 2022. Available from: <https://www.imasul.ms.gov.br/ gestao-de-unidades-de-conservacao/ >. Accessed in 2022.

IRONS, J. R., DWYER, J. L., BARSI, J. A. The next Landsat satellite: The Landsat Data Continuity Mission. **Remote Sensing of Environment**, v. 122, p. 11–21. 201

KUMAR, L.; MUTANGA, O. Google Earth Engine Applications Since Inception: Usage, Trends, and Potential, **Remote Sensing**, vol, 10, 2018. pp. 1509.

LANDIS, J. R.; KOCH, G. G. The Measurement of Observer Agreement for Categorical Data. **Biometrics**, vol. 33, 1977. pp. 159-174.

LOBO, F. de L.; SOUZA-FILHO, P. W. M.; NOVO, E. M. L. de M.; CARLOS, F. M.; BARBOSA, C. C. F. B. Mapping Mining Areas in the Brazilian Amazon Using MSI/Sentinel-2 Imagery (2017). **Remote Sensing**, vol.10, 2018, p. 1178.

MARRIS, E. The forgotten ecosystem. **Nature**, vol. 437, 2005, n 7061. pp. 944–945.

MING, D.; ZHOU, T.; WANG, M.; TAN, T. Land cover classification using random forest with genetic algorithm-based parameter optimization. **Journal of Applied Remote Sensing**.vol. 10, n.3, 2016.

MMA, Ministério do Meio Ambiente. O Bioma Cerrado. 2019. Available from < http://www.mma.gov.br/biomas/cerrado >. Accessed in 2019.

NYLAND, K.E.; GUNN, G. E.; SHIKLOMANOV, N. I.; ENGSTROM, R.N.; STRELETSKIY, A. Land Cover Change in the Lower Yenisei River Using Dense Stacking of Landsat Imagery in Google Earth Engine. **Remote Sens.**, vol. 10, 2018, n. 8. p. 1226.

PARENTE, L.; FERREIRA, L. Assessing the Spatial and Occupation Dynamics of the Brazilian Pasturelands Based on the Automated Classification of MODIS Images from 2000 to 2016. **Remote Sensing**, vol. 10, 2018, n.4. p. 606.

PEREIRA, A.C.; GAMA, G V.F. Anthropization on the Cerrado biome in the Brazilian Uruçuí-Una Ecological Station estimated from orbital images. **Brazilian Journal of Biology**. vol. 70, 2010, n. 4. pp. 969-976.

PEREIRA, G. H. de A. et al. Uso da plataforma Google Earth Engine para análise multitemporal de imagens SAR para detecção de variações de áreas inundadas no Pantanal. **Anais 7º Simpósio de Geotecnologias no Pantanal**, Embrapa Informática Agropecuária/INPE, Jardim, 2018. p. 274-282.

Projeto MapBiomas – Coleção 7.0 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil, acessado em 2022 através do link: https://mapbiomas.org/

QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project., 2019.

RATTER, J. The Brazilian Cerrado Vegetation and Threats to its Biodiversity. **Annals of Botany**. vol. 80, 1997, n. 3, pp. 223–230.

ROY, D. P.; WULDER, M. A.; LOVELAND, M. C.; WOODCOCK, C. E.; LLEN, R. G.; ANDERSON, M. C.; HELDER, D.; IRONS, J. R.; JOHNSON, D. M.; KENNEDY, R.; SCAMBOS, T. A.; SCHAAF, C. B.; SCHOTT, J. R.; SHENG, Y.; VERMOTE, E. F.; BELWARD, A. S.; BINDSCHADLER, R.; COHEN, W. B.; GAO, F.; HIPPLE, J. D.; HOSTERT, P.; HUNTINGTON, J.; JUSTICE, C. O.; KILIC, A.; KOVALSKYY, V.; LEE, Z. P.; LYMBURNER, L.; MASEK, J. G.; MCCORKEL, J.; SHUAI, Y.; TREZZA, R.; VOGELMANN, J.; WYNNE, R. H.; ZHU, Z. Landsat-8: Science and Product Vision for Terrestrial Global Change Research. **Remote Sensing of Environment**, v. 145, p.154-172. 2014.

SCHOLZ, F. G. Hydraulic redistribution of soil water by Neotropical savanna trees. **Tree Physiol**. vol.22, 2002. pp.603-612.

SPERA, S. A.; GALFORD, G. L.; COE, M. T., MACEDO, M. N.; MUSTARD, J. F. Land-use change affects water recycling in Brazil's last agricultural frontier. **Global Change Biology**. vol.22, 2016, 3405–3413.

STOREY, J.; CHOATE, M.; LEE, K. Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance. **Remote Sensing**, v. 6, p. 11127-11152. 2014.

STORY, M.; CONGALTON, R. G. Accuracy assessment: A user's perspective. **Photogrammetric Engineering and Remote Sensing**. vol.52, 1986, p.397−399.

STRASSBURG, B. B. N.; BROOKS, T.; FELTRAN-BARBIERI, R.; IRIBARREM, A.; CROUZEILLES, R.; LOYOLA, R.; LATAWIEC, A.E.; OLIVEIRA FILHO, F.J.B.; SCARAMUZZA, C. A. DE M.; SCARANO, F. R.; SOARES-FILHO, B.; BALMFORD, A. Moment of truth for the Cerrado hotspot. **Nature Ecology & Evolution**. vol. 1, 2017. pp. 0099.

TSAI, Y.H.; STOW, D.; CHEN, H.L.; LEWISON, R.; AN, L.; SHI, L. Mapping Vegetation and Land Use Types in Fanjingshan National Nature Reserve Using Google Earth Engine. **Remote Sensing**, vol. 10, 2018, pp. 927.

USGS, United State Geological Survey. Section 2 – Observatory Overview. 2018. Site <https://landsat.usgs.gov/landsat-8-l8-data-users-handbook-section-2>. acessado em dezembro 2018.

VERMOTE, E., JUSTICE, C.; CLAVERIE, M.; FRANCH, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. **Remote Sensing of Environment**, 2016, 185, 46-56.

WARD, D. E.; SUSOTT, R. A; KAUFFMAN, J. B ; BABBITT, R. E.; CUMMINGS, D. L.; DIAS, B.; HOEBEN, N.; KAUFMAN, Y. J.; RASMUSSEN, R. A.; SETZER, A. W. Smoke and fire characteristics for cerrado and deforestation burns in Brazil: BASE-B Experiment. **Journal of Geophysical Research**. vol. 97, 1992. pp. 14601.

ZALLES, V.; HANSEN, M. C.; POTAPOV, P. V.; STEHMAN, S. V.; TYUKAVINA, A.; PICKENS, A.; SONG, X.; ADUSEI, B.; OKPA, C; AGUILAR R., JOHN, N.;, CHAVEZ, S. Near doubling of Brazil's intensive row crop area since 2000. **Proceedings of the National Academy of Sciences**. 2018.