BRAZILIAN MIDWEST NATIVE VEGETATION MAPPING BASED ON GOOGLE EARTH ENGINE

N. V. Estrabis [1, *], L. Osco [1], A. P. Ramos [2], W. N. Gonçalves [1], V. Liesenberg [3], H. Pistori [4], J. Marcato Junior[1]

[1] Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Mato Grosso do Sul, Brazil - (nayara.estrabis, wesley.goncalves, jose.marcato@ufms.br)@ufms.br; pradoosco@gmail.com
[2] Environmental and Regional Development, University of Western São Paulo, Presidente Prudente, Brazil - anaramos@unoeste.br
[3] Department of Forest Engineering, Santa Catarina State University, Santa Catarina, Brazil - veraldo.liesenberg@udesc.br
[4] Department of Computer Engineering, Dom Bosco Catholic University, Mato Grosso do Sul, Brazil - pistori@ucdb.br

KEY WORDS: Native Vegetation, Landsat 8 OLI, Google Earth Engine, Vegetation Indices, Atlantic Forest, Random Forest

ABSTRACT:

Google Earth Engine (GEE) platform is an online tool, which generates fast solutions in terms of image classification and does not require high performance computers locally. We investigate several data input scenarios for mapping native-vegetation and non-native-vegetation in the Atlantic Forest region encompassed in a Landsat scene (224/076) acquired on November 28, 2019. The data input scenarios were: I- spectral bands (blue to shortwave infrared); II- NDVI (Normalized Difference Vegetation Index); III- mNDWI (modified Normalized Difference Water Index); IV- scenarios I and II; and V- scenarios I to III. Our results showed that the use of spectral bands added NDVI and mNDWI (scenario V) provided the best performance for the native-vegetation mapping, with accuracy of 96.64% and kappa index of 0.91.

1. INTRODUCTION

The native vegetation has been converted by human activities into agriculture, pasture areas and urbanization. These activities contribute for reducing the native vegetation area and generate impacts on the environmental, which sometimes can be irreversible.

In Brazil, the native vegetation is characterized by six biomes: Amazon, Cerrado (Brazilian Savanah), Caatinga, Atlantic Forest, Pampas and Pantanal. The Brazilian Midwest is composed of Amazon, Cerrado, Atlantic Forest and Pantanal biomes. In this region, two world hotspots in biodiversity are located: Atlantic Forest and Cerrado biomes, considered as such due to their high biodiversity, in contrasting with their high threat (Myers et al., 2000).

The Atlantic Forest was considered one of the largest Americas's rainforest (Ribeiro et al., 2009). This biome has around 20,000 plant species (MMA, 2019), more than 2.9% of global vertebrates' species (Myers et al, 2000) and endemic species. However, after more than 5 centuries of destruction and explorations (Bogoni et al., 2018), only 29% of native vegetation exists, considering all stages of regenerations and existence in the biome, according to the Brazilian Ministry of Environment (MMA, 2019). The fragments and habitat loss are one of the causes for species reductions, an example of which is observed in the impacts on beetle diversity in this biome (Araujo et al., 2015). Also, there are other changes that directly and indirectly impact on the ecosystem in this biome and the society.

The human activities control, and monitoring are essentials to Atlantic Forest conservation and preservation. The Google Earth Engine (GEE) is a remote sensing tool capable to process large amount of data, e.g. satellites images and vectors, with fast processing and no need of a high-performance computers (Kumar and Mutanga, 2018). The cloud-computing platform is easy to use, and it is also supporting the society (Gorelick et al., 2017) and environmental challenges.

The GEE has been explored in different studies, such as the Glacial Lake mapping (Chen et al., 2017), lake dynamics' monitoring (Zhou et al., 2019), flood preventing (Liu et al, 2018), land cover changes (Sidhu et al, 2018), forest plantation's mapping (Koskinen et al., 2019), and vegetation and land use types' mapping (Tsai et al., 2018). Although there are some studies in Brazil, e.g., in pasture areas (Parente, Ferreira 2018), mangrove status (Diniz et al., 2019), Cerrado native vegetation (Estrabis et al., 2019), in Amazon mining areas (Lobo et al., 2018), in Matopiba (Pereira, Gama, 2018), the GEE is little explored in Brazilian Midwest regions.

The native vegetation preservation and conservation are essentials, which makes the production of reliable vegetation mapping important. In this context, with the potential of GEE, the main aim of this study is to evaluate the use of GEE for native vegetation mapping in Brazilian Midwest. For this analysis, we considered different data input scenarios. In order to achieve this goal, the paper is organized as follow. Section 2. describes the methodology adopted in this study. Section 3 presents and discuss the results. Section 4 summarizes the main conclusion.

2. METHODOLOGY

2.1 Study Area

The study area (Figure 1) comprises the South region of Mato Grosso do Sul state. It is located in the Midwest of Brazil in Atlantic Forest biome. The area covers a part of a Conservation Unit named "Parque Estadual das Várzeas do Rio Ivinhema", which all extension covers 733.45 km² according to Mato Grosso do Sul Environmental Institute (Imasul, 2015). This

* Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-303-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165607

region is composed by small cities, where the agriculture and livestock represent the most expressive economy activities.

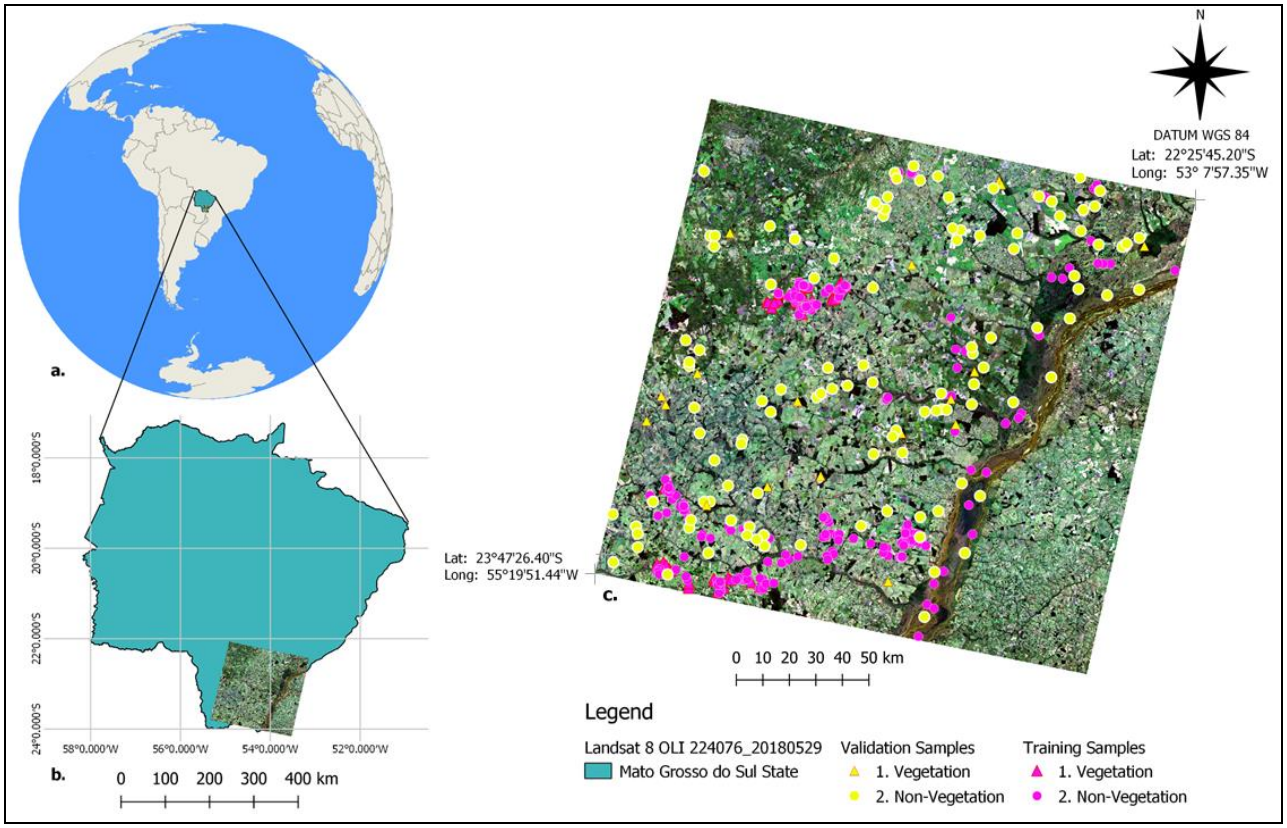

Figure 1: a. study area located in Brazil, b. study area located in Mato Grosso do Sul State, and c. Training (in magenta) and validation (in yellow) polygons samples centers location.

We used the GEE platform for the image acquisition and classification. The GEE has a large repository with different image collections from different satellites. All the training and validation samples used for the classification process were performed on GEE.

2.2 Image Selection

We used a satellite image from Landsat 8 Collections, sensor Operational Land Imagers (OLI), available on repository of GEE. The Landsat image has a spatial resolution of 30 meters. It is orthorectified and has some spectral bands, in the visible, near infrared band (NIR), and shortwave infrared (SWIR). The revisit interval is 16 days. We use the bands specified in Table 1.

Band Number	Spectral region	Wavelength (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	2.107 – 2.294

Table 1. Landsat image bands used for the image classification, and their respective wavelength intervals

2.3 Classification Process

After the image selection, two classes were defined: class 1 – native-vegetation; and class 2 - non vegetation. The non-native-vegetation class characterizes most part of land cover and activities practiced in the study area, such as silviculture, agriculture areas, pasture, water and exposed soil. An amount of 45 samples were considered for native-vegetation class and 242 samples were considered for non-native-vegetation class. A total of 287 training polygon samples were hand-draw, with different sizes and distribution in the image (see Figure 1.c). The number of samples for the native-vegetation class was smaller than the non-native-vegetation class due to the small amount of vegetation areas found than the non-vegetation areas, as to express the local reality we chose to use unbalanced samples quantities.

We used Random Forest (RF) algorithm for the training and classification. This algorithm was proposed by Breiman (2001), and it is an ensemble learning which uses prediction trees independently and the same distribution for all trees. In this study the RF's parameter 'trees' was processed with 100 trees. It was performed tests with other quantities of trees (10, 50 and 250 trees), providing similar results. Taking into considerations some tests performed by Breiman (2001), it was decided to use the algorithm with 100 trees.

Vegetation and water spectral indices were considered in the classification process: Normalized Difference Vegetation Index (NDVI) proposed by Rouse Jr (1974) (Equation 1) and modified

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-303-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165607

Normalized Difference Water Index (mNDWI) (Equation 2). The NDVI is an index large used in vegetation investigations due to the capacity to obtain the chlorophyll content obtained through the differences between infrared and red reflectances. The mNDWI is an index proposed by Xu (2006), in which the SWIR1 band replaces the NIR band. According to the author, the replacement of this band improves the water detection when compared to the NDWI. We considered the mNDWI due to presence of wetlands in the study area.

$$NDVI = \frac{(NIR - \mathrm{Re}\,d)}{(NIR + \mathrm{Re}\,d)} \quad (1)$$

$$mNDWI = \frac{(Green - SWIR1)}{(Green + SWIR1)} \quad (2)$$

where Green = Green Band (B3)
Red = Red Band (B4)
NIR = NIR Band (B5)
SWIR1 = SWIR 1 Band (B6)

We considered four scenarios (Table 2) to evaluate the better performance for the native vegetation classification.

Scenarios	Data Input
I	All spectral Bands
II	NDVI
III	mNDWI
IV	Scenarios I and II
V	Scenarios I to III

Table 2. Experimental scenarios

2.4 Validation

The validation process was performed in QGIS Software (QGIS Development Team, 2016) using 20 validation samples for the native-vegetation and 118 for the non-native-vegetation class. As in the training, the validation samples were polygons hand-drawn with different sizes and located in different areas than the training samples.

The classifications with different scenarios were analysed and compared through the kappa index, proposed by Cohen (1960) and for each class (kappa hat) to measure the agreements proportions and accuracies as explained by Story and Congalton (1986). We adopted overall accuracy, producer's and user's accuracy. The overall accuracy is the ratio between all the agreements and all the samples.

The producer's accuracy is the ratio between the agreements and the reference samples, showing how well an area located on Earth was mapped and indirectly it expresses the omission errors. The user's accuracy is the ratio between the agreements and the classified data samples. It measures the reliability in which a mapping represents something that exists on the ground.

3. RESULTS AND DISCUSSION

For the validation process, the confusion matrix (Table 3) was estimated. For each scenario, it is possible to observe the agreements (in bold) and the confusions.

		Reference		
		Class	1	2
Classified	Scenario I	1	**89.76**	0.86
		2	10.24	**99.14**
	Scenario II	1	**89.17**	16.25
		2	10.83	**83.75**
	Scenario III	1	**58.80**	7.78
		2	41.20	**92.22**
	Scenario IV	1	**90.07**	0.81
		2	9.93	**99.19**
	Scenario V	1	**88.87**	0.92
		2	11.13	**99.08**

Table 3. Confusion matrix for the conducted experiments

The results for scenarios I, II, IV and V were similar, providing more agreements between the classes. Scenario III (mNDWI), had the lowest agreements for the native-vegetation class, however the non-native-vegetation class agreements were better than scenario II (NDVI). The scenario II obtained one of the highest value of agreements for native-vegetation class (more than 89%) and the lowest for non-native-vegetation (more than 15%).

The scenario II (NDVI), a vegetation index, demonstrated a good performance for class 1, however, analysing the class 2 it obtained more than 15% of confusions between class two and class one. One of reasons for this confusion is the presence of other types of vegetation existing on wetlands, silviculture, agriculture and pasture considered as class 2, non-native-vegetation, mainly silviculture, which also has spectral similarities with native vegetation.

The scenario III, which considers the water index mNDWI, resulted in the lowest agreements for class 1 and more than 90% of agreements for class 2. This scenario presented more than 40% of vegetation samples confused with non-native-vegetation. One of the reasons is the presence of wetlands regions and other elements in the image as silviculture, agriculture and pasture, which were not different enough for this classification.

The scenario IV provided results similar to scenario 1 (Table 3), however represented the best performance. The agreements for native-vegetation class reached 90.07% and 99.19% for non-native-vegetation class, indicating less than 10% of errors. The combination of Landsat bands and NDVI slightly improved the results.

Scenario V (Landsat 8 bands and the indices) presented similar results for native-vegetation class compared to scenarios I, II and IV, approximately 89% of agreements for native-vegetation class and 99% for non- native-vegetation class.

The scenario I, comprising only the Landsat bands, reached almost 90% of agreements for the native-vegetation class and 99% for the non- native-vegetation class. These results showed errors less than 11% for native-vegetation class and less than 1% for non- native-vegetation class.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-303-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165607

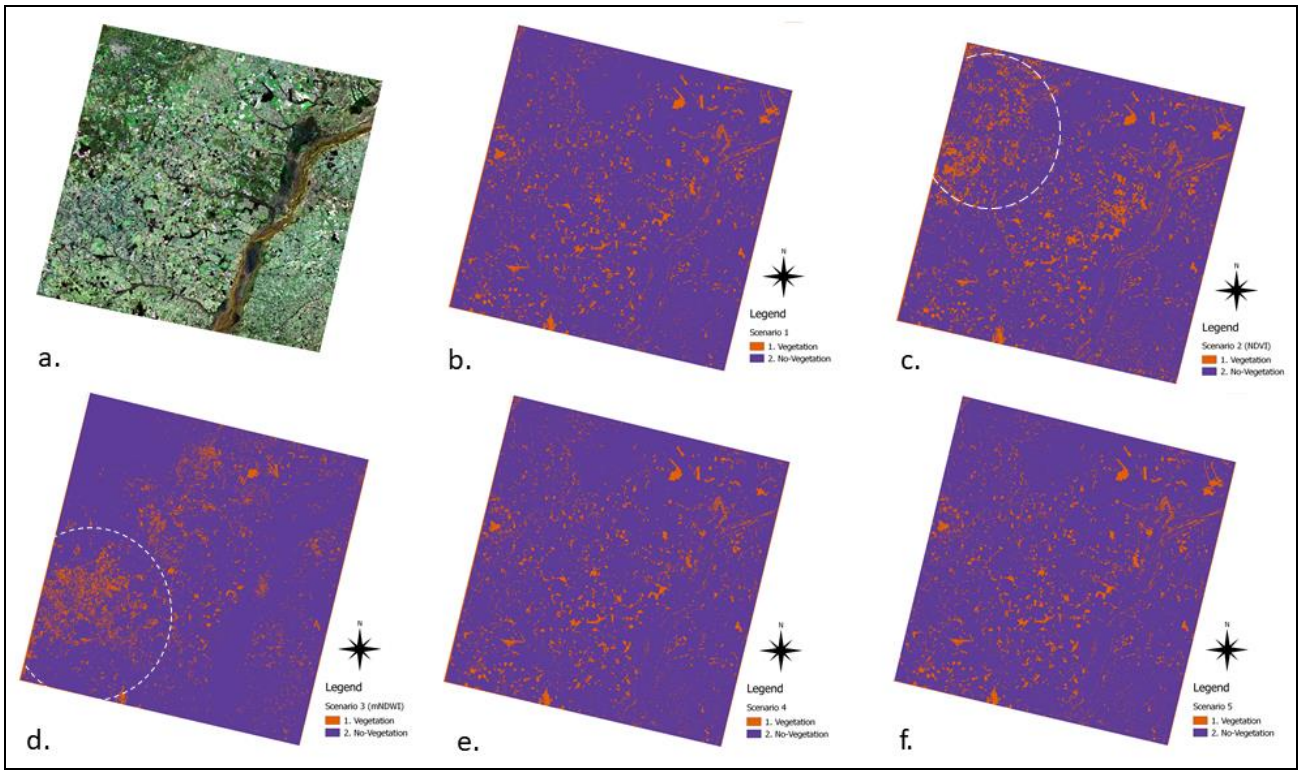

Figure 3: a. Image Landsat investigated, b. classification - scenario I, c. classification - Scenario II (NDVI), d. classification - scenario III (mNDWI), e. classification - scenario IV (I + II) and f. classification - scenario V (I + II+III).

Scenario	Producer's Accuracy (%)		User's Accuracy (%)		Kappa Hat	
	class		class		class	
	1	2	1	2	1	2
I	89.86	99.14	97.59	96.18	**0.97**	0.86
II	89.17	83.75	68.04	95.22	0.56	0.83
III	58.80	92.22	74.57	85.23	0.65	0.47
IV	88.87	99.04	97.41	95.82	0.96	0.85
V	**90.07**	**99.19**	**97.74**	**96.26**	**0.97**	**0.87**
	Overall Accuracy				Kappa Index	
I	96.55				0.91	
II	85.27				0.67	
III	82.88				0.55	
IV	96.23				0.90	
V	**96.64**				**0.91**	

Table 4. Accuracy and Kappa Index for validation results

According to Table 4, the scenario V presented the best results for each class and overall evaluations. The producer's accuracy reached more than 90% for class 1 and 99% for class 2, representing the lowest omission errors. Regarding the user's accuracy, we achieved results of approximately 98% for class 1 and 96% for class 2, indirectly representing commission errors less than 5%. The kappa index reached 0,97 for class 1 and 0.87 for class 2. Both values are considered as "almost perfect" agreements (Landis, Koch, 1977).

The scenario I presented almost the same results than scenario V, with slight differences. The producer's and user's accuracy reached respectively almost 90% and 96% for class 1 and 99% and 96% for class 2. These values represent omissions and commission errors less than 5%. The kappa hat reached 0,99 for class 1 and 0.86 for class 2, representing "almost perfect" agreements.

The scenario II presented producer's accuracy of 89% for class 1, similar to scenario I, and approximately 80% for class 2. These values indicated that omissions errors were lower than 17%. The user's accuracy for vegetation class was around 68%, and, consequently the commission errors were higher than 30%. This value indicated for this classification that only 68% of the vegetation is correctly classified as it really is, on the ground. The user's accuracy for class 2 were approximately 95%, representing commissions errors lower than 5%. The kappa index was 0.56 for the native-vegetation class, the lowest value of kappa index for all the scenarios.

The producer's accuracy for scenario III were approximately 58% for class 1 and 92% for class 2. These results showed that in the class 1 more omission errors occurred in the class than in the others scenarios, and in the class 2 less than 10% of omission errors occurred. The user's accuracy was approximately 75% for class 1 and 85% for class 2, representing commission errors of 25% and 15%, respectively. The kappa index was 0.65 for class 1 and 0.47 for class 2, one of the lowest values compared to the scenarios I, II, IV and V.

The scenario IV provided producer's accuracy of 88.87% for class 1 and 99.04% for class 2. The user's accuracy was 97.41% for class 1 and 95.82% for class 2. These results indicated omissions and commissions errors lower than 12%. The kappa index for each class reached 0.96 for native-vegetation and 0.85 for class 2. These values are slightly lower than those obtained in scenario I.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W12-2020-303-2020 | © Authors 2020. CC BY 4.0 License.
Primary publication at IEEE Xplore: https://doi.org/10.1109/LAGIRS48042.2020.9165607

In general, the scenarios I and V provided the highest results and performance compared to the others scenarios, with an overall accuracy and kappa index higher than 96% and 0.91, respectively. According to Cohen (1960), this value for kappa index represents "almost perfect agreements". However, the scenario V provided slightly better performance. This scenario represented the highest agreements performed by classification with RF.

The classification results for all the scenarios are presented in Figure 3. The images classified presented some areas with differences on classification. In the scenario II, it is possible to observe an area on the left size (inside of white circle) that was poorly detected by the other scenarios. Parts of these area correspond agricultural areas, classified as native vegetation by the NDVI band. The scenario III with mNDWI detected more concentration of native vegetation (inside the white circle) in region above the area highlighted by the circle in scenario II. This area corresponds to some wetlands areas existent in pastures lands and near vegetation of forest type, however most of the areas in these concentration do not represent the forest type native vegetation. The scenario I, IV and V balanced these characteristics, and the classification map generated, the values of accuracy and kappa index were similar. However, the classification of scenario V presented less omission and commission errors, producing a more accurate mapping.

4. CONCLUSION

The native vegetation can be mapped with high accuracy and kappa index using RF algorithms on Google Earth Engine. The processing speed is efficient, depending on the internet connection. This advantage is very helpful and contributes for the classifier evaluation.

Our different scenarios demonstrated that the use of the bands 2 to 7 from Landsat 8 OLI images with the NDVI and mNDWI is the composition that provides the best performance for native vegetation mapping with high accuracy and kappa index.

We suggest for future work the extension of these study to the entire Brazilian Midwest and other regions, also the use of different classifiers and parameters.

ACKNOWLEDGEMENTS

Acknowledgements CAPES, CNPQ and FUNDECT for supporting the project and a Master Scholarship. This research was funded by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil) through a Master Scholarship (Finance Code 001), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and FUNDECT (Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul), through research grants (p. 456149/2014-7, p. 433783/2018-4 and p. 59/300.066/2015). The work was also partially funded by CAPES through the program Print-CAPES

REFERENCES

Araujo, L. S., Komonen, A., Lopes-Andrade, C., 2015. Influences of landscape structure on diversity of beetles associated with bracket fungi in Brazilian Atlantic Forest. *Biological Conservation*, 191, 659–666. doi.org/10.1016/j.biocon.2015.08.026.

Bogoni, J. A., Pires, J. S. R., Graipel, M. E., Peroni, N., Peres, C. A., 2018. Wish you were here: How defaunated is the Atlantic Forest biome of its medium- to large-bodied mammal fauna? *PLOS ONE*, 13(9). doi.org/10.1371/journal.pone.0204515.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5–32. doi.org/:10.1023/a:1010933404324.

Chen, F., Zhang, M., Tian, B., Li, Z., 2017. Extraction of Glacial Lake Outlines in Tibet Plateau Using Landsat 8 Imagery and Google Earth Engine. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9), 4002–4009. doi.org/10.1109/jstars.2017.2705718.

Cohen, J. A., 1960. Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, v. 20, 1960, n. 1, pp. 37 - 46. doi.org/10.1177/001316446002000104

Colombo, A.F., Joly, C.A., 2010. Brazilian Atlantic Forest lato sensu: The most ancient Brazilian forest, and a biodiversity hotspot, is highly threatened by climate. *Brazilian Journal of Biology*, 70(3), pp. 697-708.

Diniz, C., Cortinhas, L., Nerino, G., Rodrigues, J., Sadeck, L., Adami, M., Souza-Filho, P., 2019. Brazilian Mangrove Status: Three Decades of Satellite Data Analysis. *Remote Sens*, 11(7), 808. doi.org/10.3390/rs11070808.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. doi.org/:10.1016/j.rse.2017.06.031.

Instituto do Meio Ambiete de Mato Grosso do Sul – Imasul, 2015. Unidades de Conservações Estaduais. https://www.imasul.ms.gov.br/gestao-de-unidades-de-conservacao/unidades-de-conservacao-estaduais/ (24 October 2019).

Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., Pekkarinen, A., Käyhkö, N., 2019. Participatory mapping of forest plantations with Open Foris and Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 63–74. doi.org/10.1016/j.isprsjprs.2018.12.011.

Kumar, L., Mutanga, O. 2018. Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sens*, 10(10), 1509. doi.org/10.3390/rs10101509.

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

Liu, C.-C., Shieh, M.-C., Ke, M.-S., Wang, K.-H., 2018. Flood Prevention and Emergency Response System Powered by Google Earth Engine. *Remote Sens*, 10(8), 1283. doi.org/10.3390/rs10081283.

MMA - Ministério do Meio Ambiente, 2019. Mata Atlântica. https://www.mma.gov.br/biomas/mata-atl%C3%A2ntica_emdesenvolvimento (24 October 2019)

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Kent, J., 2000. Biodiversity hotspots for conservation

priorities. *Nature*, 403(6772), 853–858. doi.org/10.1038/35002501.

Parente, L., Ferreira, L., 2018. Assessing the Spatial and Occupation Dynamics of the Brazilian Pasturelands Based on the Automated Classification of MODIS Images from 2000 to 2016. *Remote Sens*, 10(4), 606. doi.org/10.3390/rs10040606.

Pereira, A., Gama, V., 2010. Anthropization on the Cerrado biome in the Brazilian Uruçuí-Una Ecological Station estimated from orbital images. *Brazilian Journal of Biology*, 70(4), 969–976. doi.org/10.1590/s1519-69842010000500008.

QGIS Development Team, 2016. QGIS Geographic Information System (QGIS) Software, Version 2.14.22. Open Source Geospatial Foundation Project. http://qgis.osgeo.org (21 October 2019).

Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J., Hirota, M. M., 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, 142(6), 1141–1153. doi.org/10.1016/j.biocon.2009.02.021.

Rouse Jr J.W., Hass, R.H., Schell, J.A., Deering, D.W. 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, v. 351, p. 309, 1974. In: Proceedings Earth Resources Technology Satellite-1 Symposium, 3. Washington, D.C.: NASA. Goddart Space Flight Center; . v.1. p.309-17.

Sidhu, N., Pebesma, E., and Câmara, G., 2018. Using Google Earth Engine to detect land cover change: Singapore as a use case. *European Journal of Remote Sensing*, 51(1), 486-500. doi.org/10.1080/22797254.2018.1451782.

Story, M. and Congalton, R.G., 1986. Accuracy Assessment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing*, 52, 397-399.

Tsai, Y., Stow, D., Chen, H., Lewison, R., An, L., Shi, L., 2018. Mapping Vegetation and Land Use Types in Fanjingshan National Nature Reserve Using Google Earth Engine. *Remote Sens*, 10(6), 927. doi.org/10.3390/rs10060927.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033. doi.org/10.1080/01431160600589179.

Zhou, Y., Dong, J., Xiao, X., Liu, R., Zou, Z., Zhao, G., & Ge, Q., 2019. Continuous monitoring of lake dynamics on the Mongolian Plateau using all available Landsat imagery and Google Earth Engine. *Science of The Total Environment*, 689, 366-380. doi.org/10.1016/j.scitotenv.2019.06.341.