ANOMALY DETECTION BASED METHOD FOR LANDSCAPE DYNAMICS MONITORING: THE AMAZON AND PANTANAL CASE

Vinícius Luiz Silva Gino[1], *Rogério Galante Negri*[2], *Felipe Nascimento Souza*[3] *e José Victor Orlandi Simões*[4]

[1]UNESP, Instituto de Ciência e Tecnologia, vinicius.gino@unesp.br; [2]UNESP, Instituto de Ciência e Tecnologia, rogerio.negri@unesp.br; [3]UNESP, Instituto de Ciência e Tecnologia, fn.souza@unesp.br; e [4]UNESP, Instituto de Ciência e Tecnologia, jose.orlandi@unesp.br

RESUMO

As tecnologias de Sensoriamento Remoto e a Inteligência Artificial emergem com potencial para definir novas ferramentas de monitoramento ambiental a partir de sua combinação. O presente trabalho propõe uma abordagem que explora métodos de Detecção de Anomalias aplicados a imagens de Sensoriamento Remoto para identificar mudanças espaço-temporais na superfície da Terra. O potencial da abordagem introduzida é explorado em um estudo de caso sobre a análise das mudanças da paisagem com aplicação dos métodos One-Class SVM e Isolation Forest em imagens MODIS e Landsat para regiões de Cáceres-MT e Altamira-PA, sujeitas às mudanças climáticas intensificadas pelos impactos de atividades antrópicas.

Palavras-chave – Sensoriamento Remoto, Detecção de Anomalias, Floresta Amazônica, Pantanal.

ABSTRACT

Remote Sensing technologies and Artificial Intelligence based environmental monitoring tools emerge with potential in climate changes context. The present work proposes an approach that explores Anomaly Detection methods applied to Remote Sensing images to identify spatial and temporal changes on the Earth's surface. The potential of this approach was introduced, explored and assessed in a case study on the analysis of landscape changes with the application of the One-Class SVM and Isolation Forest methods in MODIS and Landsat for Cáceres-MT and Altamira-PA areas, subject to climate change intensified by the impacts of human activities.

Key words – *Remote Sensing, Anomaly Detection, Amazon Forest, Pantanal.*

1. INTRODUCTION

The Environmental Monitoring is involved in main goals of UN 2030 Agenda for sustainable development [1]. The accelerated consumption of natural resources in past decades results in significantly impacts in landscape and ecosystem services [2]. Unfortunately, Brazil lies at the center of debates regarding environmental questions, since it has the largest tropical forest in the world and unique feature biomes, Amazon and Pantanal, respectively.

The intensification of illegal mining, logging and agricultural activities in Legal Amazon resulted resulted in higher deforestation rates since 2014 [3]. As impacts of these events, the Pantanal extrreme droughts between 2019 and 2020 can be highlighted [4].

In this sense, remote sensors rises as an important tool to monitoring spectral and temporal dynamics of Earth surface [5]. Anomaly Detection comprises a kind of unsupervised Machine Learning technique that may be applied in Remote Sensing spectral indices data to automatically identify the temporal changes over landscape and water bodies [6].

This project proposes an approach the use of Anomaly Detection methods over Remote Sensing data to identify surface changes. Furthermore, this research implements a prototype of an "anomaly monitoring and warning system" fed by images acquired by the MODIS and Landsat programs/satellites. As a case study, assess the municipality of Altamira-PA and Cáceres-MT in Amazon and Pantanal biomes, respectively.

2. MATERIALS AND METHODS

2.1. Anomaly Detection Methods and Spectral Indices

Among the different techniques that permeate Machine Learning, Anomaly Detection identifies elements with significantly distinct behavior compared to other observations. In this sense, two methods have been successfully employed in remote sensing studies: Isolation Forest (IF) and One-Class Support Vector Machine (OC-SVM) [7].

Support Vector Machines (SVM) based methods are very popular in Remote Sensing applications. Its solid mathematical formulation, simple algorithmic structure and high generalizability are some of the outstanding features of this approach [8]. From the original conception of the SVM method several variants have been proposed, among which OC-SVM can be highlighted for its motivation by the problem of quantile estimation and anomaly detection [9]. It is worth noting that the OC-SVM is parameterized by $\nu \in [0, 1]$ and other parameters related to the adopted kernel function. Further details on kernel functions are discussed in [10].

The Isolation Forest (IF) [11] comprises a low-computational cost method able to overcome the difficulties when dealing with large databases. This method has been used in Remote Sensing studies [12] and other analyses involving digital image processing [13]. In summary, the IF embodies an ensemble of decision trees, in this case, called "isolated tree" (IT). According to the conceptual idea behind this method, when the data/objects are submitted to classification in a decision tree scheme, the anomalies tend to present a short path to the root node. The expected length of this path is strictly dependent on the number of decision trees in the ensemble and the size of the dataset [14].

The Anomaly Detection Methods (ADM) was applied in spectral indices obtained from Remote Sensing different wavelenghts data combinations. This research used Normalized Difference Vegetation Index (NDVI) computed by $\frac{x_{NIR} - x_{Red}}{x_{NIR} + x_{Red}}$. NDVI has various application purposes, for example, monitoring and mapping crops, droughts, pest damage, agricultural productivity, hydrological modeling, and others [15].

2.2. Conceptual formalization

Figure 1 depicts a general overview of the proposed method for multitemporal anomaly detection.

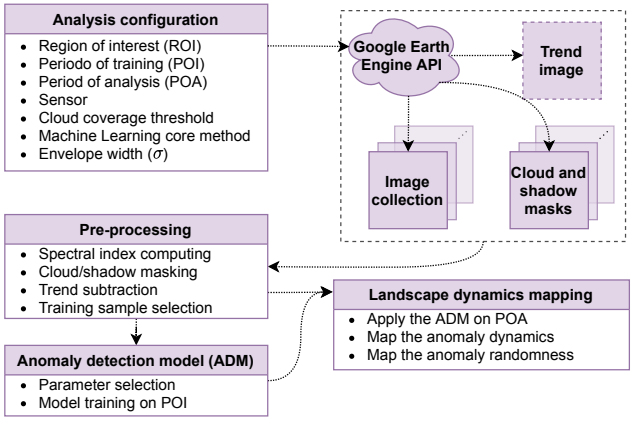


Figura 1: Overview of the proposed method.

The Python 3.8 was the programming language adopted to implement the proposed method, as the monitoring prototype. Additionally, the Scikit-Learn library was used to apply the Anomaly Detection methods. Moreover, the Pandas library was employed to organize the information.

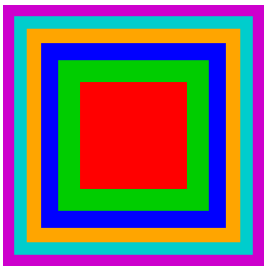
The Anomaly detection models are trained with basis on observed values of a previously defined spectral index (i.e., NDVI or NDWI) in $[-\alpha\sigma, +\alpha\sigma]$, where σ is thee standard deviation of considered spectral index and $\alpha = 0.5$ is a constant adopted to control the training set regularity.

Lastly, the Google Earth Engine (GEE) Application Programming Interface (API) is used to access the Remote Sensing image catalogs and obtain the multitemporal image series according to the defined period, region, and sensor, based in Python. Landsat and Sentinel data are considered in this study. The cloud occurrence threshold of 5% inside the region of analysis is admitted to disregarding useless scenes.

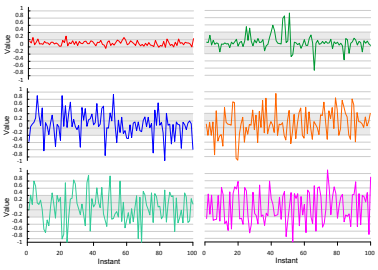
3. DATA TYPES

3.1. Synthetic data

To validate the purpose and optimize previously mentioned Anomaly Detection methods to Remote Sensing applications, this project is assessed using 100 synthetic images segmented into six degrees of anomaly probability for equal areas. Figure 2 depicts values of x observed over the time series for randomly selected positions s into each region.



(a) The conceptual image structure. The regions $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4, \mathcal{R}_5, \mathcal{R}_6$ are represented in red, green, blue, orange, cyan and magenta, respectively.



(b) Values x observed for positions s selected into each conceptual image's region. The color of each line assign the selected position to one region.

Figura 2: Synthetic data composition.

3.2. Geospatial data

Figure 3 presents the two study area locations considered in this experiment. The first study area (Area 1) comprises the downtown region of Cáceres, Mato Grosso state, Brazil. The Pantanal's landscape was deeply changed after extreme drougths as results of climate dynamics registered beetween 2019 and 2020. At last, the second area (Area 2) contains a portion of Altamira, Pará state, Brazil, a region located in Amazon biome which is under strong deforestation over the last decades.

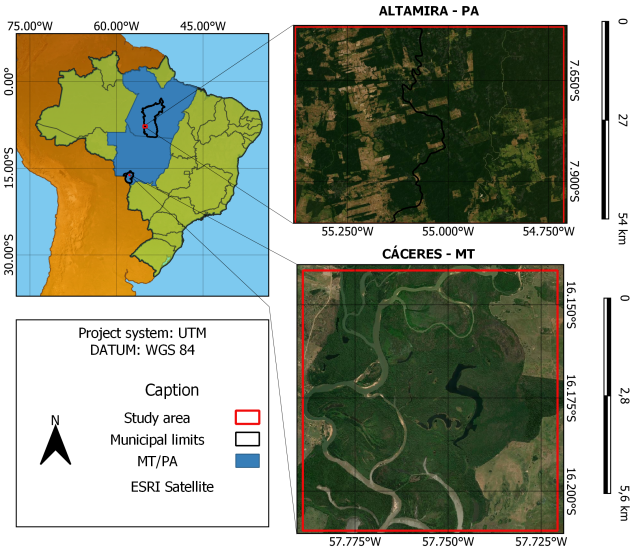


Figura 3: Spatial location of study areas.

Parameter	Altamira-PA	Cáceres-MT
Sensor	MODIS	Landsat-8 OLI
Resolution [m]	250	30
Period	2010-2021	2013-2021
Number of images	276	30
std [σ]	0.085	0.112
mean [μ]	-0.015	-0.016

Tabela 1: Image time series informations.

In order to assess the results obtained by applying the proposed framework on each study area and respective image series, since the study areas have been modified by

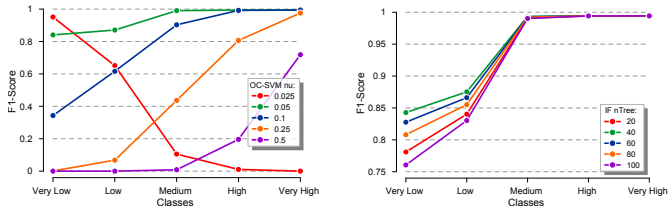
a constant deforestation process, we employed as ground-truth samples the polygons from the annual deforestation inventory provided by the Real-Time System for Detection of Deforestation (DETER) project and Monitoring Program for the Amazon and Other Biomes (PRODES) [3].

4. RESULTS AND DISCUSSION

4.1. Synthetic data

Once defined the synthetic image series, different parameter configurations were tested by a grid-search procedure with 10-fold cross-validation and the respective results were assessed in terms of F1-Score [16] measure by considering six "classes of dynamic". Such classes stands form "no anomalies", "very low", "low", "medium", "high" and "very high" when the anomaly frequencies are 0%, 1–20%, 21–40%, 41–60%, 61–80% and 81–100%, respectively.

Figure 4 shows the OC-SVM and IF performances on each "classes of dynamic" when considering distinct parameters (ν and n_t). Independently of method or parameter, the "medium", "high" and "very high" classes are perfectly identified (i.e., F1-Score = 1). Reversely, the method and parameter choice are relevant when mapping the classes "very low" and "low", where $\nu = 0.05$ and $n_t = 40$ comprises good parameter choices for OC-SVM and IF, respectively.



(a) OC-SVM. (b) IF.

Figura 4: Performances of OC-SVM and IF method on identify the distinct anomaly frequency classes according to distinct parameter configuration.

4.2. Study areas

The same procedure occurred in Remote Sensing data for study areas, different parameter configurations were tested considering the six "classes of dynamic". However, the validation of this process are made through change comparison maps, which reference samples are collected from image time series for Cáceres-MT and from PRODES database for Altamira-PA region.

4.2.1. Cáceres-MT

Figure 5 depicts the results of multitemporal analysis, which provides reference samples of changed and unchanged areas, additionally this figure shows the "anomaly count" maps considering classes of dynamic and multiple parameters for Isolation Forest and One-Class SVM.

Figure 6 shows the *Kappa scores* for change comparison samples, whose expected results were the compatibility between "Changed" areas and high anomaly counts, the same were prospected to "Unchanged" regions and low anomaly

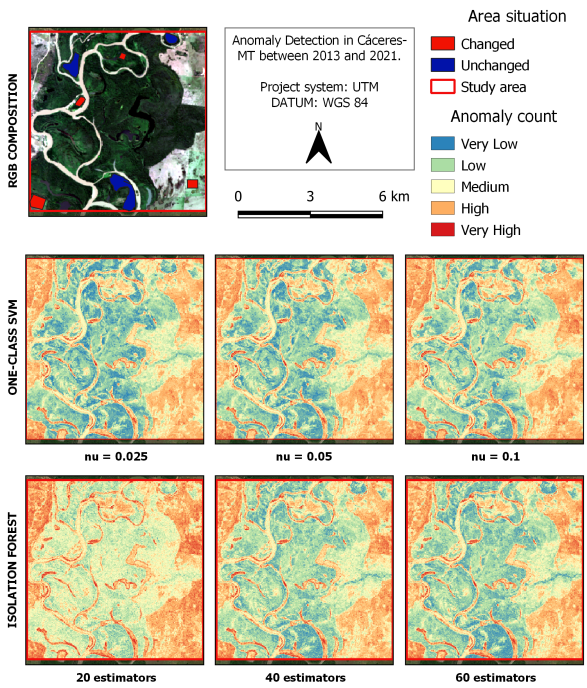
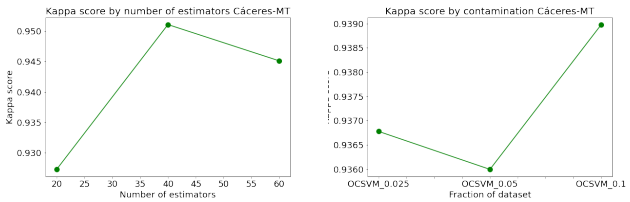


Figura 5: Anomaly Detection results for Cáceres-MT.

counts. This metric was calculated considering: (i) "Very Low" and "Low" labels as "No change"; (ii) "Medium" label as transition areas; (iii) "High" and "Very High" labels as "Changes".



(a) IF. (b) OCSVM.

Figura 6: *Kappa scores* for change comparison maps in Cáceres-MT.

Regarding the evaluation, all *Kappa scores* obtained corresponded to the expectations, being superior to 0.9 in all parameter configuring. The study area comprises regions such as wetlands, agricultural crops and the Paraguai River, whose registered the lowest level in 50 years during the period of study. In addition, Pantanal was under an intensive droughts and fires over last 5 years, which be considered to analyze the results.

4.2.2. Altamira-PA

The municipality of Altamira in Pará, with the largest territorial extension in Brazil, is subject to intense exploitation natural resources, evidenced by the deforestation of an area approximately equal to 500.000 football fields between 2017 and 2020, according to PRODES data.

Figure 7 depicts the results of multitemporal analysis considering PRODES reference samples of deforestation beetween 2010 and 2021 to validate the purposed methods. The evaluation of prototype from this study area were

constructed considering triannual intervals of anomalies and change samples over 2010 through 2021.

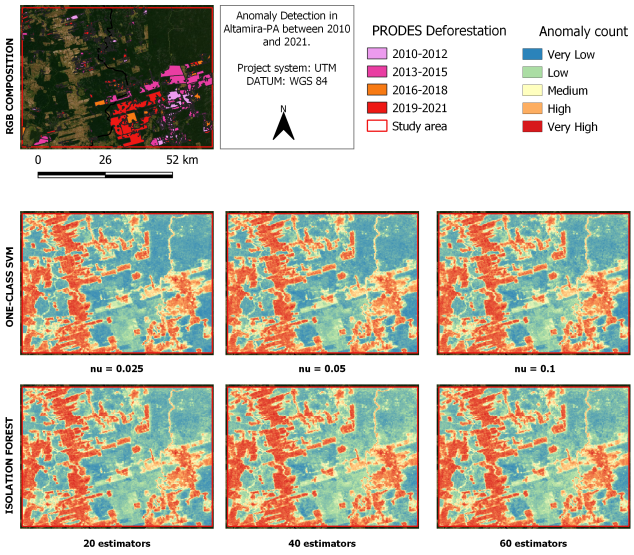
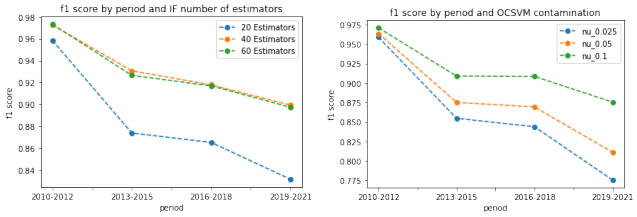


Figura 7: Anomaly Detection results for Altamira-PA.

Figure 8 shows the *F1-scores* for PRODES samples, whose expected results were the compatibility between deforestation areas and high anomaly counts.



(a) IF. (b) OCSVM.

Figura 8: *F1-scores* for change comparison maps in Altamira-PA.

5. CONCLUSIONS

Based on the presented results, it is possible to verify that the proposed approach be able to identify anomalies that correspond to targets with high spectral-temporal dynamics. It is also noted that both Machine Learning methods were able to identify key regions in some of the main brazilian biomes for validate the applied methods, as deforestation areas in Altamira indicated by PRODES and wetland areas in Mato Grosso Pantanal.

In theory, according to the validation by synthetic images, it is more difficult for IF and OCSVM to recognize non-anomaly patterns. The IF method was able to distinguish anomalies allied to lower computational costs using $n_t \in [20, 60]$. The OC-SVM, in turn, was more sensitive for detecting changes, often classifying unchanged regions as anomalies, but still with strong agreement between changes and anomalies using $\nu \in [0.025, 0.1]$.

In future works, other numerical validation techniques may be addressed to determine the correlation between the detection of anomalies and the intensity of changes, in order to improve the proposed prototype applications.

6. REFERENCES

[1] Katherine Anderson, Barbara Ryan, William Sonntag, Argyro Kavvada, and Lawrence Friedl. Earth observation in service of the 2030 agenda for sustainable development. *Geo-spatial Information Science*, 20(2):77–96, 2017.

[2] Will Steffen, Wendy Broadgate, Lisa Deutsch, Owen Gaffney, and Cornelia Ludwig. The trajectory of the anthropocene: the great acceleration. *The Anthropocene Review*, 2(1):81–98, 2015.

[3] Luiz Fernando FG Assis, Karine Reis Ferreira, Lubia Vinhas, Luis Maurano, Claudio Almeida, Andre Carvalho, Jether Rodrigues, Adeline Maciel, and Claudinei Camargo. Terrabrasilis: a spatial data analytics infrastructure for large-scale thematic mapping. *ISPRS International Journal of Geo-Information*, 8(11):513, 2019.

[4] Jose A Marengo, Ana P Cunha, Luz Adriana Cuartas, Karinne R Deusdará Leal, Elisangela Broedel, Marcelo E Seluchi, Camila Miranda Michelin, Cheila Flávia De Praga Baião, Eleazar Chuchón Ângulo, Elton K Almeida, et al. Extreme drought in the brazilian pantanal in 2019–2020: characterization, causes, and impacts. *Frontiers in Water*, 3:639204, 2021.

[5] John R Jensen. *Remote sensing of the environment: An earth resource perspective*. Pearson Education India, 2 edition, 2009.

[6] Qiandong Guo, Ruiliang Pu, and Jun Cheng. Anomaly detection from hyperspectral remote sensing imagery. *Geosciences*, 6(4):56, 2016.

[7] Jacinta Holloway and Kerrie Mengersen. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10(9):1365, 2018.

[8] L. Bruzzone and C. Persello. A novel context-sensitive semisupervised SVM classifier robusttomislabeled training samples. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7):2142–2154, 2009.

[9] Yunqiang Chen, Xiang Sean Zhou, and Thomas S Huang. One-class svm for learning in image retrieval. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 1, pages 34–37. IEEE, 2001.

[10] John Shawe-Taylor, Nello Cristianini, et al. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[11] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008.

[12] Shutao Li, Kunzhong Zhang, Puhong Duan, and Xudong Kang. Hyperspectral anomaly detection with kernel isolation forest. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):319–329, 2019.

[13] Francisco Alonso-Sarria, Carmen Valdivieso-Ros, and Francisco Gomariz-Castillo. Isolation forests to evaluate class separability and the representativeness of training and validation areas in land cover classification. *Remote Sensing*, 11(24):3000, 2019.

[14] Julien Lesouple, Cédric Baudoin, Marc Spigai, and Jean-Yves Tourneret. Generalized isolation forest for anomaly detection. *Pattern Recognition Letters*, 149:109–119, 2021.

[15] Jinru Xue and Baofeng Su. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017, 2017.

[16] C. J. Van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, USA, 2nd edition, 1979.