Remote Sensing and Machine Learning on Anomaly Detection at high spectral and temporal dynamics regions in Brazil

Vinícius L. S. Gino, Rogério G. Negri, Felipe N. Souza

[1]Instituto de Ciência e Tecnologia (ICT)
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)
12247-004 – São José dos Campos – SP – Brazil

{vinicius.gino,rogerio.negri,fn.souza}@unesp.br

Abstract. *In climate changes context Remote Sensing tools are widely used and widespread in research. In this sense, Artificial Intelligence rises offering possible improves for environmental monitoring applications using techniques such as Machine Learning for Anomaly Detection applied to Remote Sensing imagery to identify the spatio-temporal changes over the Earth's surface. This approach is explored in three high dynamic regions in Brazil assessing deforestation, fires and technological disaster areas using One-Class SVM and Isolation Forest methods over MODIS, Landsat and Sentinel images.*

1. Introduction

One of the biggest global challenges is breaking issues like greenhouse gases emission, deforestation, and other disasters impulsed by unstoppable consumption of natural resources [Steffen et al. 2015]. In this sense, the "United Nations 2030 Agenda" provides a multidimensional and holistic vision of this subject, where sustainable development goals rule how to combine human well-being with economic prosperity and environmental protection to guide public policies to mitigate impacts on the environment [Pradhan et al. 2017]. Unfortunately, Brazil lies at the center of debates regarding environmental questions, since it has the largest tropical forest in the world.

The impacts of climate changes are a painful question in the country, once Brazil had the largest area of deforestation between 2010 and 2015 [MacDicken et al. 2016]. The Amazon Forest previously mentioned is influenced by atmospheric and meteorological factors, such El Nino and La Nina whose intensification, combined with human interactions, caused extreme events in past years [Jimenez et al. 2019]. Thus, the biodiversity found in unique biomes such Pantanal is at the mercy of threats caused by land use changes and climate dynamics related with atmospherical air moves from Amazon [Marengo et al. 2021].

In addition, technological disasters are a significant cause of surface dynamics in Brazil. The reason of this approach comes from the recent technological disasters caused by the failures on mining dams in Mariana [do Carmo et al. 2017] and Brumadinho [Rotta et al. 2020], which resulted in the death of hundreds of people in addition to significant environmental impacts. Face to these events, the development of strategies and tools to analyze and monitor mining dams has demanded attention.

In this scenario, Remote Sensing technology rises as a convenient tool for observing and analyzing the Earth's surface. Beyond allowing register the information in different spectral wavelengths, the remote sensors also allow wide spatial and temporal analysis [Jensen 2009]. Additionally to Remote Sensing data, the Machine Learning techniques encompass the construction of algorithms able to identify and extract information from large bases of data, which includes diverse studies and applications with Remote Sensing data [Lary et al. 2016]. Anomaly Detection comprises a kind of unsupervised Machine Learning technique that may be applied in Remote Sensing data to automatically identify the temporal changes and dynamics over the Earth's surface [Guo et al. 2016].

In the light of the presented discussions, this study addresses the use of Anomaly Detection and Remote Sensing data to identify regions with high spectral-temporal dynamics. Furthermore, this research proposes and implements a prototype of an "anomaly monitoring and warning system" fed by images acquired by the MODIS, Sentinel and Landsat programs/satellites. Functionalities of the Google Earth Engine platform support such implementation. A study case focuses on analyzing the regions at the center of environmental debate such the city of Altamira, Para, in Amazon Forest and areas affected by the dams collapse in Mariana and Brumadinho.

2. Theory background

2.1. Anomaly Detection

Among the different techniques that permeate Machine Learning, Anomaly Detection identifies events/elements with significantly distinct behavior compared to other observations. Usually, such techniques have been used in the identification of bank fraud, checking for intruders in security systems, and in supporting medical analysis [Gu et al. 2019]. In addition to these applications, anomaly detection techniques are highlighted as a potential tool for the environmental monitoring [Dereszynski and Dietterich 2011].

The Breaks For Additive Season and Trend (BFAST) [Lambert et al. 2013], Local Outlier Factor (LOF) [Ma et al. 2013], Elliptic Envelope [Hoyle et al. 2015] and One-Class Support Vector Machine (OC-SVM) [Chen et al. 2001] and Isolation Forest (IF) [Liu et al. 2008] are example of Anomaly Detection methods found in the literature. In special, the two latter mentioned methods have been successfully employed in remote sensing studies [Rembold et al. 2013, Holloway and Mengersen 2018].

As a variant of the well-known and attractive Support Vector Machine (SVM) method, the OC-SVM [Chen et al. 2001] deals with quantile estimation and anomaly detection problems. Conceptually, starting from a set of observations $\mathcal{Z}$, the OC-SVM method provides a model capable of classifying the objects as part of a set of non-anomalous elements according to a probability ν of false-positive occurrence.

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 [Shawe-Taylor et al. 2004].

The Isolation Forest (IF) [Liu et al. 2008] 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 [Li et al. 2019] and other analyses involving digital image processing [Alonso-Sarria et al. 2019].

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 [Lesouple et al. 2021].

The definition of an IT starts from a sample set $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$, where $\mathbf{x}_i = [x_{i1}, \ldots, x_{id}]^T \in \mathbb{R}^d$ with components express a specific attribute in m observations. This dataset may also be represented as a matrix $\mathbf{X}$ whose columns are the vectors $\mathbf{x}_i$, for $i = 1, \ldots, m$. The nodes of a IT may be either internal or external. While the earlier have two descendants, the external node has no descendent and are called "leaf". Based in this structure, the IT sequentially randomly select a value p in the q-th attribute to split $\mathbf{X}$ into two descendants. After recursively perform this process, the IT is defined. As stop criterion for the IT expansion, is assumed: (i) the IT reaches its length limit; (ii) $|\mathbf{X}| =1$; or (iii) all the columns of $\mathbf{X}$ are equal.

Regarding the IT structure, the Anomaly Detection process is performed by scores assigned to each $\mathbf{x}_i$ according to the root-to-leaf path length that such vector pass-through the IT, represented by $h(\mathbf{x}_i)$. The average estimate of $h(\mathbf{x}_i)$ for the external nodes is the same as an unsuccessful search in a Binary Search Tree, expressed as:

$$c(m) = 2H(m-1) - \frac{2(m-1)}{m} \tag{1}$$

where $H(i) = ln(i) + 0.5772156649$ is a harmonic number [Havil 2003] and $c(m)$ is the average estimate of $h(\cdot)$ considering the m observations. In turn, the anomaly score is:

$$s(\mathbf{x}_i, m) = 2^{-\left(\frac{E(h(\mathbf{x}_i))}{c(m)}\right)} \tag{2}$$

where $E(h(\mathbf{x}_i)) = \frac{1}{q}\sum_{i=1}^{q} h(\mathbf{x}_i)$ is the mean of $h(\mathbf{x}_i)$ from a collection of ITs.

Therefore, it can be inferred that if $E(h(\mathbf{x}_i))$ tends to zero, the score tends to 1, representing then an anomaly. On the other hand, when $h(\mathbf{x}_i)$ tends to $m-1$, s tends to 0, showing very likely regular data. Furthermore, when $E(h(\mathbf{x}_i))$ tends to $c(m)$, $s(\mathbf{x}_i, m)$ tends to 0.5 and then there is no anomaly distinction.

2.2. Spectral Indices

A spectral index comprises a combination of two or more spectral bands to provide a particular representation of the Earth's surface. Among a plethora of spectral indices proposed in the literature, the vegetation indices take into account the spectral response of chlorophyll targets concerning electromagnetic radiation from the Sun [Moreira 2000].

One of the most used vegetation indices for canopy characterization is the Normalized Difference Vegetation Index (NDVI) [Rouse et al. 1974], which uses the red and infrared bands as input data. This index has various application purposes, for example, monitoring and mapping crops, droughts, pest damage, agricultural productivity, hydrological modeling, and others [Xue and Su 2017].

The Normalized Difference Water Index (NDWI) [Gao 1996] comprises a spectral index based on the region of electromagnetic spectrum sensitive to water presence. Its use allows detecting particulate matter and suspended sediments in water columns.

Let consider $\mathcal{I}(s) = \mathbf{x}$ where the components x_{Green}, x_{Red} and x_{NIR} stands for the radiometric response at the green, red and near-infrared wavelengths. The NDVI and NDWI values at the position s is computed by $\dfrac{x_{NIR} - x_{Red}}{x_{NIR} + x_{Red}}$ and $\dfrac{x_{Green} - x_{NIR}}{x_{Green} + x_{NIR}}$, respectively.

3. Proposal of multitemporal anomaly detection

3.1. Conceptual formalization

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

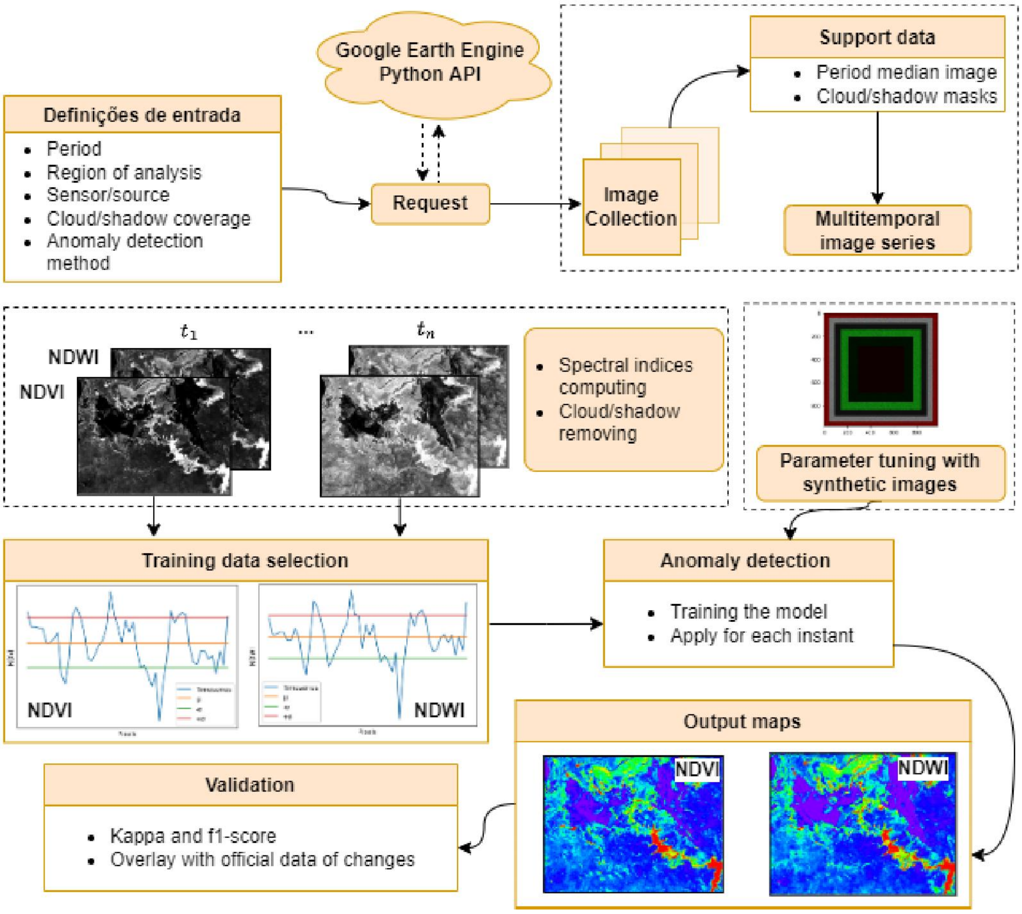


Figure 1. Overview of the proposed method.

Accordingly to this structure, as an initial step, it is defined the period of analysis, the region of interest, a cloud cover threshold, and a remote sensor as a data source. The anomaly detection method is also defined in the initial step. Such configuration (except the anomaly detection model) is submitted as a request to the Google Earth Engine (GEE), which consequently returns a collection of images that gives place to a multitemporal image series. A median image and cloud/shadow cover masks are determined from such image series as support data for posterior use. In a second stage, the NDVI and NDWI are computed at each instant and then subtracted from the median image of period for that study area to translate all the data around a common central tendency (i.e., the zero). Moreover, information from areas affected by cloud and shadow occurrences are disregarded after applying the previously defined masks. After, the NDVI and NDWI translated values in $[-\alpha\sigma, +\alpha\sigma]$ are used to train an anomaly detection model F and classify the complete dataset. The σ is the dataset standard deviation and $\alpha \in \mathbb{R}$ is an adopted scale factor. Lastly, a map about the multitemporal dynamics is produced according to an anomaly counting over the analyzed period. Also, a map of p-value based on the "run test of randomness" [Siegel and Castellan 1988] allows identifying regions with high confidence regarding the occurrence of the changes.

3.2. Implementation details

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 the 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 20% inside the region of analysis is admitted to disregarding useless scenes.

4. Experiments

4.1. Synthetic data

To validate the purpose and optimize previously mentioned Anomaly Detection methods from Remote Sensing applications, this project is assessed using 100 synthetic images segmented into six degrees of anomaly probability for equal areas.

Once defined the 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 [Rijsbergen 1979] 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.

4.2. Study area and Remote Sensing data

In order to assess the method proposed and discussed at Section 3, it is carried a practical application regarding the analysis of temporal dynamics in regions divided into 2 groups: Mariana-MG and Brumadinho-MG affected after the respective dam collapses; Altamira-PA one of largests cities of Brazil in territorial area whose vegetation are threat by human exploration and climate changes. Figure 2 shows the area locations.

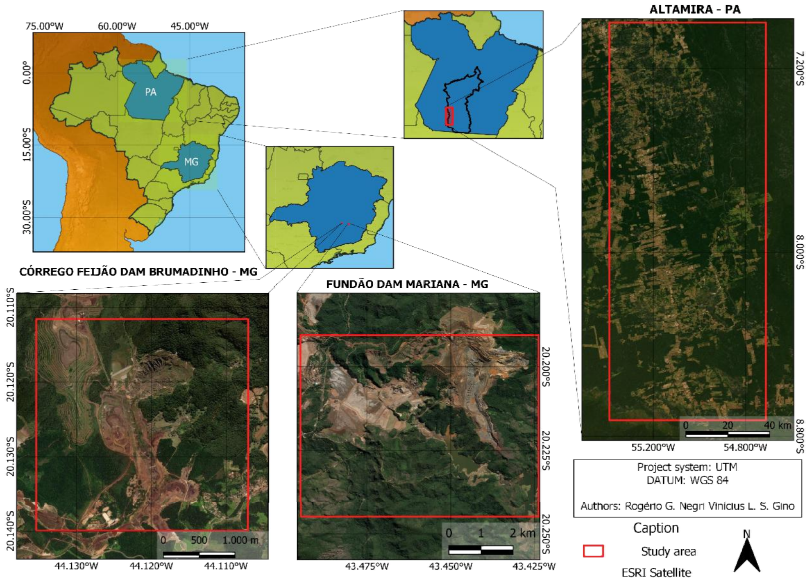


Figure 2. Spatial location of study areas.

It is worth highlighting that the Mariana (Fundão) and Brumadinho (Córrego do Feijão Mine I) dams are located in Minas Gerais (MG). These areas are considered strategic for the development of mining activity in Brazil, a sector responsible for 4% of the national GDP and the generation of more than 2 million indirect jobs [IBRAM 2020]. The disruption of these structures impacted the surrounding landscape, initially surrounded by vegetation characteristic of the Atlantic Forest biome. Moreover, these dams were built following the upstream heightening, which is less costly but with the greater risk of disruptions [Thomé and Passini 2018].

In turn, Altamira is located in the state of Para, north region of Brazil, inside Amazon Forest. Based in Brazilian Amazon Deforestation Monitoring Program (PRODES) data, since 2015 the deforestation rate increases in Amazon biome caused by many factors that reach that region, such agricultural frontier dynamics, land grabbing activity and illegal mining, being a key point to climate changes intensification [Silva Junior et al. 2021].

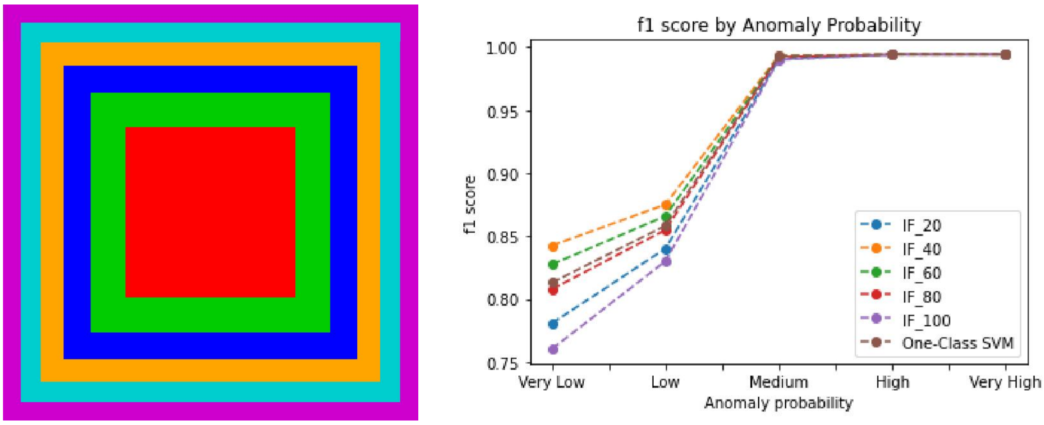
In this sense, Remote Sensing data used in this framework considered the area of dams structure and detailing capability given by each sensor: MODIS, Landsat-8 OLI and Sentinel MSI. Table 4.2 show the sensor used for each study area, the period of analysis and the number of images collected in GEE API considering a threshold in cloud occurrence of 10% for each instant.

Area	Sensor	Resolution [m]	Period	Images
Altamira	MODIS	250	2010-2021	278
Brumadinho	Sentinel MSI	10	2016-2021	67
Mariana	Landsat-8 OLI	30	2013-2021	79

Table 1. Data information from each study area.

4.3. Results and discussion

Using synthetic data, Figure 3(b) displays the f1-scores comparison assessed to define the Anomaly Detection Methods parameters and table 2 shows them.



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

(b) F1-score values observed for OC-SVM $\nu = 0.05$ and Isolation Forest estimators. The color of each line assign the selected parameters to Anomaly Detection methods.

Figure 3. Synthetic data composition and parameter optimization.

OC-SVM		Isolation Forest	
Parameter	Value	Parameter	Value
kernel	rbf	number of estimators	40
kernel coefficient (γ)	*auto*	max samples, contamination, bootstrap, verbose, warm start	default
upper bound on the fraction of training errors (ν)	0.05		
tolerance, shrinking, cache size, verbose, max iterations	default		

Table 2. Anomaly Detection Methods parameters.

Figures 4, 5 and 7 depicts a bi-temporal comparison using color compositions and the respective multitemporal dynamic maps in terms of “anomaly detection counting” and “p-value”. The first one was obtained by percentage discretization of anomaly detection data, where the lowest 20% represents “Very low” label, and so on. The NDVI and NDWI values are considered to obtain the results for Brumadinho, Altamira and Mariana areas, respectively.

To validate and comparate the results generated by the proposed method, reference samples collected from change maps of moments before and after dam failures were divided into (i) No change areas; (ii) Change areas. These samples were applied at Anomaly Detection maps, which can be observed in the histograms highlighted by Figure 6, whose expected results were the decrease of “No changes” bars as they increase “Changes” bars along anomaly count axis.

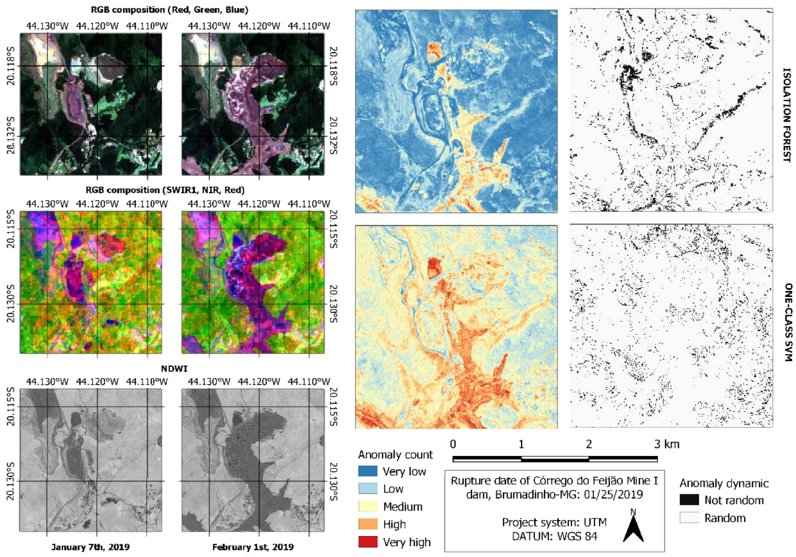


Figure 4. Results using NDWI for Brumadinho dam area.

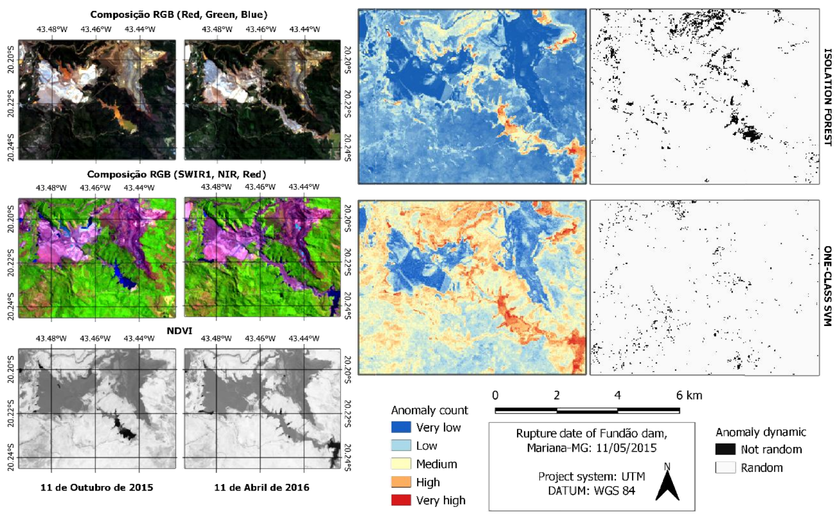


Figure 5. Results using NDVI for Mariana dam area.

Concerning that results, the validation is closed with a build of confusion matrix where x axis is represented by changed and unchanged regions in time series, in turn in the y axis represents the samples of changes and no changes. The anomaly count data, exposed in five labels, was clustered in a binary classification, were 'Medium' count of anomalies was interpreted as transition areas. 'Very Low' and 'Low' labels were grouped as 'Unchanged' while 'High' and 'Very High' labels as 'Changed'. Table 3 shows the validation metrics *Kappa Score* and *Accuracy*.

For assessing Altamira results, PRODES deforestation data between 2010 and 2021 was considered and overlayed under anomaly detection counting maps by three

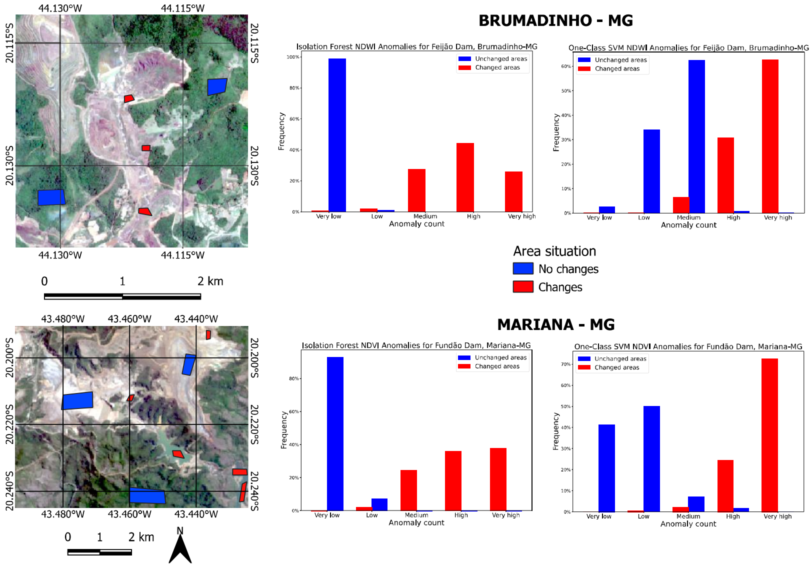


Figure 6. Anomaly count for reference samples and change map comparison.

Study area	Index	Method	Accuracy	*Kappa*
Brumadinho	NDWI	IF	0.849	0.747
		OC-SVM	0.651	0.316
Mariana	NDVI	IF	0.867	0.775
		OC-SVM	0.943	0.888

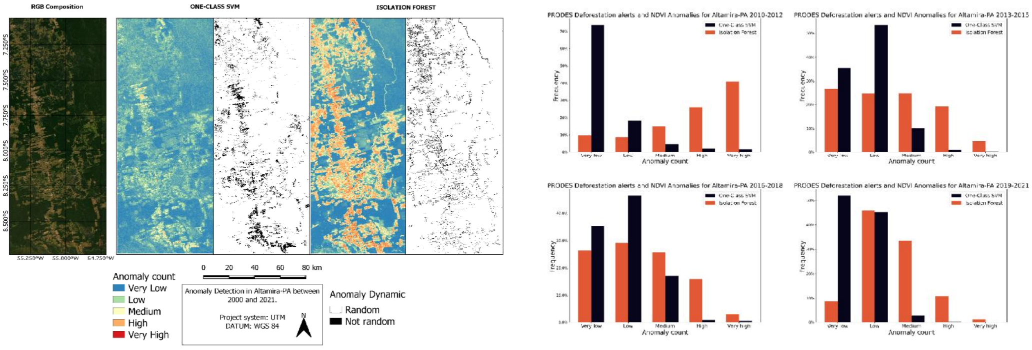
Table 3. Validation metrics for each method and study area

year periods through previously mentioned dates. Figure 7 shows the Anomaly Detection maps and validation with PRODES data.

Focusing on the “anomaly detection counting” maps, it is possible to observe that while the IF method delivers more consistent results, OC-SVM tends to overestimate the frequency of anomaly occurrence, evidenced in validation of dam areas and in agricultural fields in Altamira. Low-dynamic regions, like vegetation and exposed soil, are also highlighted when the proposed method is equipped with the IF model.

Regarding the p-value maps, under a 5% significance, the pixels in black represents regions with not random behavior in terms of anomaly/regular occurrence over time. Consequently, such regions demand attention when analyzing the obtained maps. Among the possible causes are seasonal changes showed by targets like water bodies and vegetation. In general, the p-value mapping results achieved with the IF model are more consistent than those using the OC-SVM.

The whole process involved considerable computational costs. The reference machine was a desktop with 16 GB RAM and 500 GB of SSD memory. Generally, the time of image processing and application of Anomaly Detection Methods can vary between 1.5 and 3 hours depending on spatial resolution of Remote Sensing data and study area dimensions.



(a) Results using NDVI for Altamira area.

(b) Anomaly probability from PRODES deforestation areas in Altamira.

Figure 7. Anomaly Detection results and PRODES overlay from Altamira region.

5. Conclusions

Based on the presented results, it is possible to verify that the proposed method, viewed as an environmental monitoring system prototype, could identify anomalies that correspond to targets with high spectral-temporal dynamics.

It is noteworthy that the assessed Anomaly Detection models have different precision. The IF method was able to distinguish with better contrast the regions of anomalies and regular and provide more consistent p-value maps (useful to identify seasonal changes). OC-SVM method, in turn, was more sensible to change detection often classifying unchanged regions as anomalies.

In future works could be addressed the suppression of seasonal trends and the combination of multiple spectral indices to improve the proposed prototype. Furthermore, the expansion of study areas and availability of this approach and availability of results for the general public in easy visualization and manipulation tools could be explored.

Acknowledgments

The authors thank FAPESP (grant 2018/01033-3, 2020/14664-1 and 2021/01305-6) and CNPq for their financial support of this research.

References

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

Chen, Y., Zhou, X. S., and Huang, T. S. (2001). 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.

Dereszynski, E. W. and Dietterich, T. G. (2011). Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. *ACM Transactions on Sensor Networks (TOSN)*, 8(1):1–36.

do Carmo, F. F., Kamino, L. H. Y., Junior, R. T., de Campos, I. C., do Carmo, F. F., Silvino, G., Mauro, M. L., Rodrigues, N. U. A., de Souza Miranda, M. P., Pinto, C. E. F., et al. (2017). Fundão tailings dam failures: the environment tragedy of the largest technological disaster of brazilian mining in global context. *Perspectives in ecology and conservation*, 15(3):145–151.

Gao, B.-C. (1996). NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266.

Gu, J., Wang, L., Wang, H., and Wang, S. (2019). A novel approach to intrusion detection using svm ensemble with feature augmentation. *Computers & Security*, 86:53–62.

Guo, Q., Pu, R., and Cheng, J. (2016). Anomaly detection from hyperspectral remote sensing imagery. *Geosciences*, 6(4):56.

Havil, J. (2003). Gamma: exploring euler's constant. *The Australian Mathematical Society*, page 250.

Holloway, J. and Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, 10(9):1365.

Hoyle, B., Rau, M. M., Paech, K., Bonnett, C., Seitz, S., and Weller, J. (2015). Anomaly detection for machine learning redshifts applied to sdss galaxies. *Monthly Notices of the Royal Astronomical Society*, 452(4):4183–4194.

IBRAM (2020). Informações sobre a economia mineral brasileira 2020. Technical report, Instituto Brasileiro de Mineração.

Jensen, J. R. (2009). *Remote sensing of the environment: An earth resource perspective*. Pearson Education India, 2 edition.

Jimenez, J. C., Marengo, J. A., Alves, L. M., Sulca, J. C., Takahashi, K., Ferrett, S., and Collins, M. (2019). The role of enso flavours and tna on recent droughts over amazon forests and the northeast brazil region. *International Journal of Climatology*.

Lambert, J., Drenou, C., Denux, J.-P., Balent, G., and Cheret, V. (2013). Monitoring forest decline through remote sensing time series analysis. *GIScience & Remote Sensing*, 50(4):437–457.

Lary, D. J., Alavi, A. H., Gandomi, A. H., and Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1):3–10.

Lesouple, J., Baudoin, C., Spigai, M., and Tourneret, J.-Y. (2021). Generalized isolation forest for anomaly detection. *Pattern Recognition Letters*, 149:109–119.

Li, S., Zhang, K., Duan, P., and Kang, X. (2019). Hyperspectral anomaly detection with kernel isolation forest. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):319–329.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE.

Ma, H., Hu, Y., and Shi, H. (2013). Fault detection and identification based on the neighborhood standardized local outlier factor method. *Industrial & Engineering Chemistry Research*, 52(6):2389–2402.

MacDicken, K., Jonsson, Ö., Piña, L., Maulo, S., Contessa, V., Adikari, Y., Garzuglia, M., Lindquist, E., Reams, G., and DâAnnunzio, R. (2016). Global forest resources assessment 2015: how are the world's forests changing?

Marengo, J. A., Cunha, A. P., Cuartas, L. A., Deusdará Leal, K. R., Broedel, E., Seluchi, M. E., Michelin, C. M., De Praga Baião, C. F., Chuchón Ângulo, E., Almeida, E. K., et al. (2021). Extreme drought in the brazilian pantanal in 2019–2020: Characterization, causes, and impacts. *Frontiers in Water*, 3:13.

Moreira, R. d. C. (2000). Influência do posicionamento e da largura de bandas de sensores remotos e dos efeitos atmosféricos na determinação de índices de vegetação. *São José dos Campos. 181p. Dissertação (Mestrado em Sensoriamento Remoto)-INPE*.

Pradhan, P., Costa, L., Rybski, D., Lucht, W., and Kropp, J. P. (2017). A systematic study of sustainable development goal (sdg) interactions. *Earth's Future*, 5(11):1169–1179.

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, 5(4):1704–1733.

Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann, USA, 2nd edition.

Rotta, L. H. S., Alcantara, E., Park, E., Negri, R. G., Lin, Y. N., Bernardo, N., Mendes, T. S. G., and Souza Filho, C. R. (2020). The 2019 brumadinho tailings dam collapse: Possible cause and impacts of the worst human and environmental disaster in brazil. *International Journal of Applied Earth Observation and Geoinformation*, 90:102119.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., et al. (1974). Monitoring vegetation systems in the great plains with erts. *NASA special publication*, 351(1974):309.

Shawe-Taylor, J., Cristianini, N., et al. (2004). *Kernel methods for pattern analysis*. Cambridge university press.

Siegel, S. and Castellan, N. (1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill international editions statistics series. McGraw-Hill.

Silva Junior, C. H., Pessoa, A., Carvalho, N. S., Reis, J. B., Anderson, L. O., and Aragão, L. E. (2021). The brazilian amazon deforestation rate in 2020 is the greatest of the decade. *Nature Ecology & Evolution*, 5(2):144–145.

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., and Ludwig, C. (2015). The trajectory of the anthropocene: the great acceleration. *The Anthropocene Review*, 2(1):81–98.

Thomé, R. and Passini, M. L. (2018). Barragens de rejeitos de mineração: características do método de alteamento para montante que fundamentaram a suspensão de sua utilização em minas gerais. *Ciências Sociais Aplicadas em Revista*, 18(34):49–65.

Xue, J. and Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *J. Sensors*, 2017:1353691:1–1353691:17.