Time-series metrics applied to land use and land cover mapping with focus on landslide detection

Tatiana Dias Tardelli Uehara,[a,][*] Thales Sehn Körting,[a] Anderson dos Reis Soares,[b] and Renata Pacheco Quevedo[a]

[a]National Institute for Space Research, Earth Observation and Geoinformatics Division, Sao Jose dos Campos, Brazil

[b]Cognizant Technology Solutions, São Paulo, Brazil

Abstract. Landslides are a recurring phenomenon in Brazil and have caused many socio-economic losses and casualties. To monitor them, land use and land cover (LULC) and landslide inventory maps are essential to identifying high susceptibility areas. In this sense, the main aim of this study is to produce LULC classification focused on landslide detection via semi-automatic methods, using data mining techniques with remote sensing time-series imagery. For that, different indices, such as the normalized difference vegetation index, the normalized difference built-up index (NDBI), and the soil adjusted vegetation index were extracted from Sentinel-2 imagery. Basic, polar, and fractal metrics were extracted from the time series. From the Shuttle Radar Topography Mission digital elevation model, six geomorphometric features were extracted. Then, classification was performed with random forest with four different approaches: mono-temporal, bi-temporal, metrical, and all. In every approach, the NDBI index or metric derived from it presented the highest importance, and the slope was ranked among the six first predictors. The all approach showed the highest overall accuracy (OA) (88.96%), followed by metrical (87.90%), bi-temporal (82.59%), and mono-temporal (74.95%). Briefly, the metrical approach presented the most beneficial result, presenting high OA and low levels of commission and omission errors. © *The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JRS.16.034518]

1 Introduction

Landslides are a natural gravity-driven phenomenon, which consists of the movement of a mass of soil or rock from the top of the hill toward the bottom. The occurrence of landslides is registered all over the world, frequently in mountainous areas, where there has been extreme precipitation events, earthquakes, or snow melting.[1] Landslides happen with recurrence in Brazil and have been the reason for many socioeconomic losses and casualties, for instance, the consequences of the so-called Mega disaster in 2011 at Nova Friburgo in Rio de Janeiro state, Brazil.[2]

Another example was seen in Rolante River catchment, in Rio Grande do Sul, Brazil, in which an extreme precipitation event triggered a massive occurrence of landslides in January 2017. Previous works have detected more than 300 landslide scars deriving from this episode.[3,4]

To reduce the landslide risk, land use and land cover (LULC) and inventory maps are relevant information. The LULC maps assist in the detection of areas that have experienced anthropogenic interventions that may induce landslides. According to Refs. 5–7, the land use change is recognized as one of the most important factors influencing the occurrence of rainfall-induced landslides. LULC may act as a predisposing factor for landslide occurrence since LULC characteristics such as vegetation suppression and construction building on high slope areas can

---

*Address all correspondence to Tatiana Dias Tardelli Uehara, tatiana.tardelli@gmail.com

increase the susceptibility.[5,8–10] In addition, LULC may also assist in estimating landslide consequences, according to potential loss and damage.[11]

The landslide inventory map consists of identifying and registering information about mass movement location, spatial distribution, types, and patterns, which can be useful in landslide susceptibility modeling.[12,13] Moreover, high-quality landslide inventories are also of utmost importance to calibrate and validate statistical landslide susceptibility and hazard models as well as to evaluate the performance of physically based slope stability models.[14–17] Thus, the inventory is crucial to support urban planning and disaster risk reduction.[18]

The inventory maps can be done by either conventional methods (field mapping and manual vectorization by visual interpretation) or state-of-art techniques. The manual work and refinement done using conventional methods usually guarantees a very accurate product, however the process can be time and resource consuming.[13,19] Semi-automatic methods, on the other hand, can provide a rapid mapping via change detection and pattern recognition algorithms,[13] especially through remote sensing images.

Earth observation sensors can provide data in several observation frequencies, such as daily or weekly. Most of the methodologies used for LULC classification, however, use a dataset only composed of cloud-free and mono-temporal observations.[20] Nowadays, a large amount of research has identified the benefits of incorporating a multi-temporal approach, recognizing its many potentials, to deal with change detection and the monitoring of targets, which present seasonal behavior (i.e., croplands). Hence, time series have been progressively used for LULC mapping and to identify the nature of land cover changes.[21] Remote sensing medium spatial resolution optical time-series data have demonstrated a high capacity for characterizing environmental phenomena, describing trends as well as discrete change events. The inclusion of time series change in the land cover mapping process provides information on class stability and informs on logical class transitions, both temporally and categorically.[22]

Previous studies show that different types of data have been explored to detect landslide scars. Cloud coverage is a great challenge to deal with when optical remote sensing imagery is used for landslide detection. This occurs because many of the events are triggered by rainfall, so it is common that the images from dates close to the events appear with most of the area covered by clouds. Many researches used synthetic aperture radar (SAR) and interferometric SAR (InSAR), in which the sensors can overcome this issue.[23,24] Moreover, in addition to the contribution of SAR data, some difficulties are still present as for landslide detection in densely forested areas. Considering that, the use of light detection and ranging technologies has been gaining more space, presenting significant results.[25,26] In addition to the aforementioned issues, the most commonly used type of data to detect landslide scars is still the optical remote sensing images. Several studies have focused on spaceborne or airborne (very) high-resolution optical images, once it can provide information on a more detailed scale, usually resulting in higher accuracy values.[27,28]

Out of remote sensing images, much spectral information can be extracted, and from an earth observation time series, a diversity of properties can be calculated. A variety of predictors has been successfully used on landslide classification. First, regarding spectral indices, the normalized difference vegetation index (NDVI) and the soil adjusted vegetation index (SAVI), proved to be effective tools for landslide detection and LULC mapping.[12,29] Also, the normalized difference build-up index (NDBI) has been applied to many LULC mapping researches, in addition to some other recent papers on landslide detection that considered this index as a potential variable.[30] Moreover, the digital elevation model (DEM) and its derived geomorphometric attributes have significantly contributed to the landslide scars detection, as presented by Li et al.[31] and Pradhan and Mezaal.[32] Metric variables can also be extracted from time series and incorporated in the classification inputs, among them, there are the basic and polar metrics,[33] in addition to the fractal[34] and phenological metrics.[35]

To tackle the challenge of landslide detection through semi-automatic methods, machine learning and deep-learning classifiers are commonly used, such as the support vector machine (SVM), maximum likelihood (ML), artificial neural network, random forest (RF), and the convolutional neural networks. SVM and ML have been used to identify landslides in São Paulo state coast (SP, Brazil), in which the SVM presented better performance than ML, especially when associated with the NDVI.[36] Moreover, studies show that decision tree algorithms have

been used in landslide detection[37] and LULC mapping,[38] and presented suitable results. The RF algorithm can be used both for classification and regression modeling. As a classifier, it has been widely used for the identification of landslide scars.[27,39] A research of data mining-aided automatic landslide detection compared the performance of SVM and RF, having the latter presented a higher classification accuracy.[40]

It is important to elucidate that there are many approaches for landslide scars detection to produce inventory maps. Most methods for producing landslide scars identification are based on change detection, producing binary results (scar versus non-scar). Nonetheless, based on the fact that both products (LULC and inventory map) are essential materials for landslide susceptibility modeling (the most effective method in supporting territorial planning and disasters risk reduction), the methodology here applied unites both final products in one single workflow. One should notice that, in addition to the inventory is derived from the LULC map, it does not mean that the same combination of predictors will present the best performance for both purposes. In other words, accuracy is analyzed separately, so an LULC classification, which considers all classes, could have the best performance in one approach; however, for the landslide class, specifically, some other approach could present a better accuracy.

Even though time series have already contributed for remote sensing imagery classification, the unsolved question is: until what point the addition of variables is beneficial to the classification, taking into account both accuracy analysis and the procedure's complexity/processing time? In this light, which classes present a relevant improvement in the accuracy with the increase of predictors? Which of them presents a high performance in class separability with low quantity of variables?

Furthermore, to choose the most suitable spectral index, many factors may need to be considered. When dealing with time series, the use of vegetation indices, such as the NDVI, is very common, once it helps to reduce the number of variables in the time series by joining information from the red and near infrared bands in one single predictor. The NDVI is indeed a widely used index for general classifications, as the LULC maps; however, in the scenario of vegetation analysis when a landslide occurs, would that be the best index to be used? According to the literature, NDVI, SAVI, and NDBI have been extensively used for LULC mapping, thus selected for this work.[12,29,30]

Moreover, remote sensing time-series metrics have been gaining more relevance in the scientific community as the availability of historical data from free Earth observation imagery grows as Landsat, Sentinel, and China-Brazil Earth-Resources Satellite 4A (CBERS-4A) data. Time-series metrics have shown great potential for classification especially for agricultural purposes, as exposed by bendini_2019 with crop identification in the Brazilian Cerrado. This has to do with the considerable variability of this type of class during the time. As the landslides have a fundamental change factor over the period, time-series metrics might be helpful to detect them. Even though many studies have used time series in landslide research, most of them focus on monitoring the landslide evolution[41] and especially of slow mass movements (creep) using SAR data.[42] However, research in optical time series and time-series metrics for landslide detection is still recent, so more investigation is required over whether these metrics could be helpful to identify landslide scars. Moreover, the polar metrics are still not widely used in the remote sensing community, which can be seen as a gap with considerable potential to be developed.

In this context, the main aim of this research is to develop a methodology to generate the landslide inventory map as a product derived from the LULC map for the Rolante River catchment area. This means that the landslide scars consist of a class of land cover that is assigned during the classification process. In a bi-temporal example, the landslide class corresponds, specifically, to vegetation land cover in time 1 and bare soil in time 2. The LULC classification and the landslide inventory map were generated through semi-automatic methods, based on remote sensing time-series imagery, geomorphometric attributes, and data mining techniques. For that, the specific objectives were to (1) analyze the performance of long, dense and irregular time series for LULC classification of landslide prone areas; (2) evaluate which classes present a relevant improvement in accuracy with the increase of predictors and which of them present high accuracy results even with low quantity of predictors; (3) investigate the predictors and indexes with the greatest importance to be used in these types of areas; (4) explore the potential of time-series metrics for classification and landslide detection; (5) compare different

classification approaches that range from the most simple and reduced dataset to more complex and dense ones; and (6) analyze the best generated landslide inventory and LULC map, according to this research methods, for the Rolante River catchment area.

2 Materials and Methods

The methodology developed for this research is shown in Fig. 1. The study was structured in five main steps: (i) data selection and pre-processing; (ii) feature extraction; (iii) sampling method; (iv) classification; and (v) validation and analysis. These steps are explained in the following sections.

2.1 *Study Area*

The study area covers the whole extension of the Rolante River hydrographic basin, a sub-basin of the Sinos River catchment, located in the state of Rio Grande do Sul, Brazil (Fig. 2). Its drainage area comprehends 828 km^2, with elevation values varying between 20 and 1040 m.
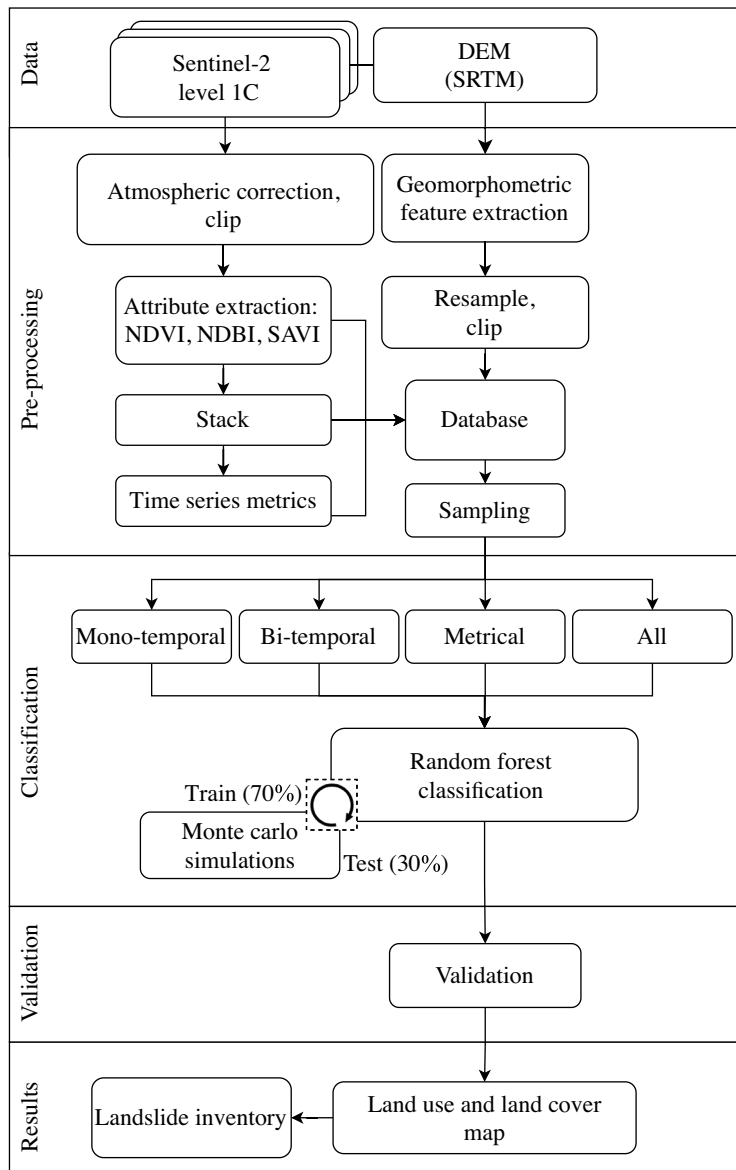
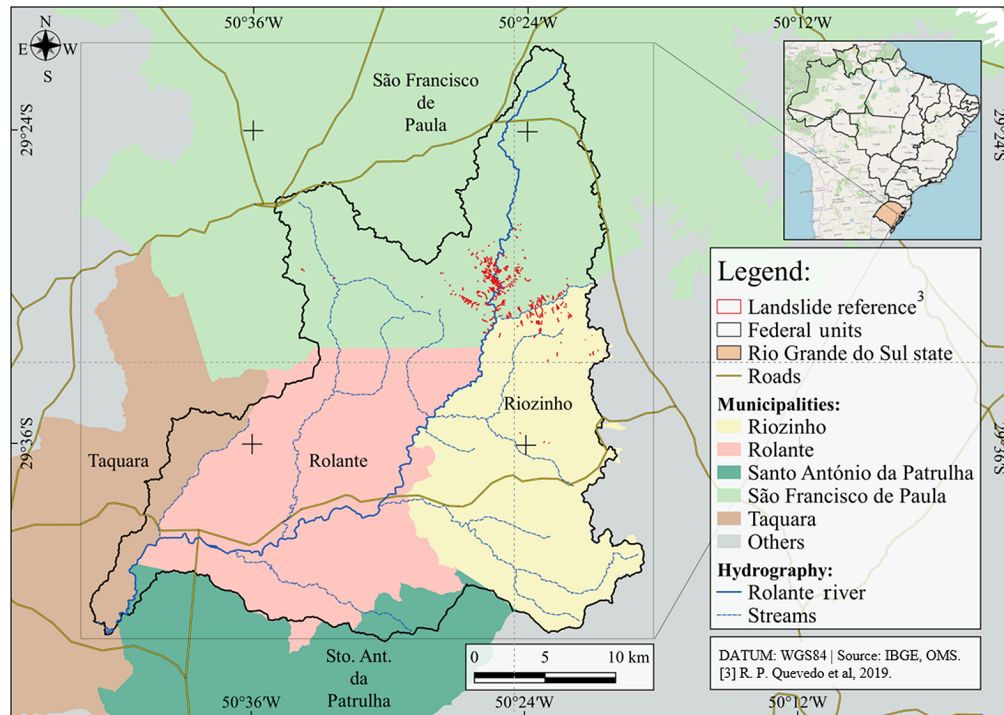**Fig. 1** Methodology flowchart.

Fig. 2 Study area, the Rolante River Catchment.

The main course of this basin is the Rolante River, which received this name due to the great impact of the water during the flood period.[43] The main cities in this region are: Riozinho, Rolante, and São Francisco de Paula. The major activities are agriculture and livestock, with the presence of native forest, silviculture, and anthropogenic rural occupation.[44]

Concerning pedological aspects, the basin contains four major types of soil: dystrophic red-yellow argisol, eutrophic red nitosol, eutrophic litholic neosol, and dystrophic humic cambisol. About the geomorphological characteristics, the area can be divided in five different units: alluvial-colluvial plains, JacuíRiver depression, Serra Geral baselines, Serra Geral unit, and Campos Gerais plateau. Moreover, regarding the geological aspects, it presents four units: holocene alluvial deposits, Botucatu unit, Serra Geral unit, and Serra Geral—Caxias facies. The climate is very humid subtropical, characterized by abundant annual precipitation varying between 1700 and 2000 mm and temperatures of 14°C to 17°C.[45]

Previous works have detected around 300 landslide scars that occurred after an extreme precipitation event on January 5, 2017.[3,4] The rain registered 90 to 272 mm and lasted for about 4 h.[46] Rolante city was affected after a natural dam disruption on the Mascarada river, a tributary of the Rolante river. This dam was generated by the accumulation of debris driven from the hillside.

The criteria to select this area for the research consists of two main factors. First, the significant presence of already mapped landslide scars. Second, the magnitude of the event, which allows the identification of the scars via orbital imagery of 10-m spatial resolution, freely available by the European Space Agency (ESA).

2.2 Data and Pre-processing

The data required to identify landslides scars depend on the characteristics of each study case. For this research, the majority of the landslides present in the Rolante River basin had, in general, 20- to 60-m width, which could be seen from 10-m resolution imagery. Considering that, Sentinel-2 (A and B) imagery was chosen, especially because it provides free orthorectified reflectance products with 10 m of spatial resolution and a five-day temporal resolution. Level 2A [Bottom Of Atmosphere (BOA)] product was not available for the date of the event

(January 2017), so Level 1C was used. Moreover, a 30-m spatial resolution DEM from the Shuttle Radar Topography Mission (SRTM) was used.

To get the surface reflectance product from Sentinel-2 Level 1C imagery, the atmospheric correction was performed using the Sen2Cor algorithm, which is used by ESA to provide the BOA Sentinel images. This system was implemented in R by Ranghetti et al.,[47] and the "sen2r" package was used to perform this correction. In this process, cloud and cloud shadow cover were masked. Images with cloud coverage above 80% were removed, so instead of the regular five days of temporal resolution provided by Sentinel-2A and Sentinel-2B satellites, a general temporal resolution of 13 days was achieved. Further, an outlier removal filter was applied to smooth each time series separately (NDVI, NDBI, and SAVI). This filter removes spikes and consists in identifying values that present more than 10% of difference from the same pixel in the previous and the following dates, if positive, it is replaced by the mean value of these two. The time interval available for the study area ranges from November 11, 2015, to June 3, 2020; which, in total, resulted in 122 satellite observations dates. The general temporal resolution of this time series is around 13 days. This can be explained by two main factors: (a) some images were discarded because of cloud coverage; (b) Sentinel-2B was launched in 2017, almost two years after Sentinel-2A. Both satellites together make a five-day resolution product; however, each of them has a revisiting time of 10 days.

2.3 Feature Extraction

The feature extraction process, shown in Fig. 3, is based on landslide detection literature, as presented by Gerente et al.[48,49] The NDVI, developed by Rouse et al.,[50] was used because it presents a drastic reduction in its values when the landslide occurs. In addition to that, once it is a composition of bands, it allows the use of the red and near-infrared (NIR) bands in one single feature. This reduction of predictors is specifically beneficial when dealing with dense time series. It is used to detect varying densities of vegetation coverage, which can be applied for mass movement events.[51] The literature shows a lot of successful research using the NDVI for land cover mapping.[52,53] The normalized difference built-up index (NDBI), proposed by Zha et al.,[54] was used in the classification of urban areas. It is composed by the red and short wave infrared (SWIR) bands. Once built-up areas present low NDVI values, these usually show high spectral similarity with bare soil and landslide scars. Furthermore, the SAVI index, developed by Huete,[55] was also used. It incorporates an adjustment factor (L), based on the amount of vegetation, from 0 (for high vegetation) to 1 (for low vegetation), to adjust the soil background
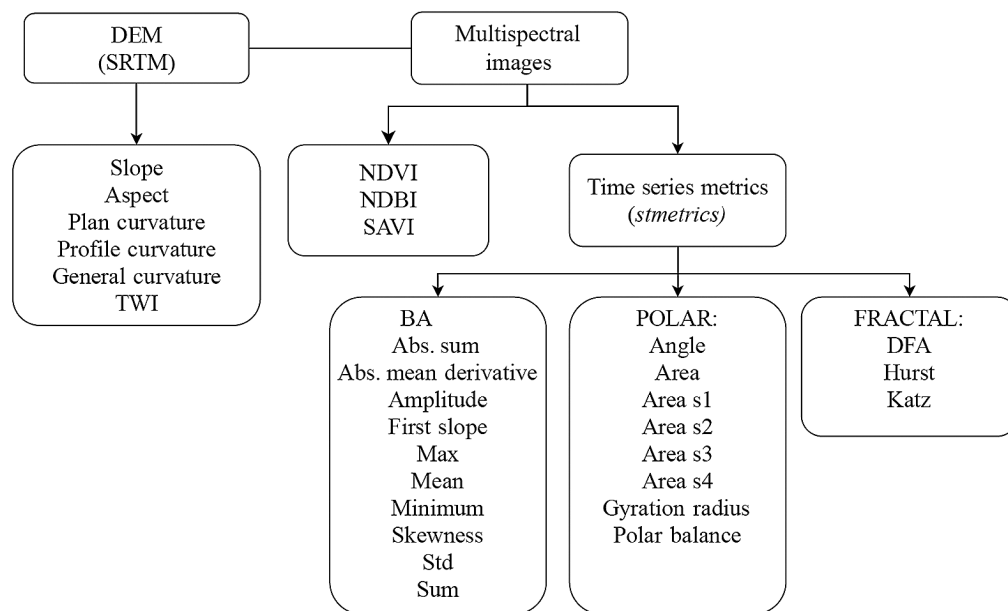

Fig. 3 Feature extraction.

effect. In the absence of extrinsic knowledge, an intermediate adjustment factor of 0.5 has been suggested and generally applied.[56] A variety of papers mention the SAVI index as a useful tool to detect landslides, as the one by Mohd Salleh et al.,[57] which analyzed vegetation anomalies in the NDVI and SAVI index as bio-indicators for landslide activity mapping. The formulas of the aforementioned indices are represented as follows:

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}}, \tag{1}$$

$$\text{NDBI} = \frac{\rho_{\text{swir}} - \rho_{\text{red}}}{\rho_{\text{swir}} + \rho_{\text{red}}}, \tag{2}$$

$$\text{SAVI} = (1 + \text{L}) \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}} + \text{L}}, \tag{3}$$

where ρ_{nir} is the reflectance value for the NIR band (B8); ρ_{red}: reflectance value for the red band (B4); ρ_{swir} is the reflectance value for the SWIR band (B11); and L is constant inversely correlated with the leaf area index.

From each dataset (NDVI, NDBI, and SAVI), time-series metrics were extracted using the stmetrics (https://github.com/brazil-data-cube/stmetrics) package available in Python. The package currently includes three modules to perform feature extraction: basic, polar, and fractal. The polar metrics are derived from a time wheel legend proposed by Edsall et al.[58] As exposed by Soares et al.,[59] to compute the polar features, each time series has its values projected to angles in the interval $[0,2\pi]$. A time series is a function $f(x, y, T)$ where (x, y) is the spatial position of a point, T is a time interval $t_1, \ldots t_N$, and N is the number of observations. The time series can be visualized by a set of values $v_i \in V$ in time t_i. Therefore, its polar representation is defined by a function $g(V)\{A, O\}$ (A corresponds to the abscissa axis in the Cartesian coordinates and O to the ordinate axis) where

$$a_i = v_i \cos\left(\frac{2\pi i}{N}\right) \in A, \quad i = 1, \ldots N, \tag{4}$$

and

$$o_i = v_i \sin\left(\frac{2\pi i}{N}\right) \in O, \quad i = 1, \ldots N. \tag{5}$$

In both equations, $\frac{2\pi i}{N}$ is an arbitrary angle that depends on the acquisition date and v_i is the corresponding time series value. Considering that $a_{n+1} = a_1$ and $o_{n+1} = o_1$, a closed shape is obtained.

The system for automated geoscientific analyses (SAGA) tool, available in the QGIS software, allows the extraction of the geomorphometric attributes from the DEM. The use of these attributes is based on previous studies, which have been associated with the landslide occurrence.[30,60,61] The extracted attributes are slope, aspect, plan curvature, profile curvature, general curvature, and the topographic wetness index (TWI). Once the Sentinel-2 data have 10 m of spatial resolution, the geomorphometric attributes were resampled from 30 to 10 m using the nearest neighbor parameter in QGIS.

The elevation is represented in meters and corresponds to the pixel value at the DEM, it is an altimetric data. The slope is defined as the zenital declivity angle, having its values varying from 0 deg to 90 deg, although it is commonly expressed in percentage.[62] The profile curvature (deg/m) is related to the profile shape of the hillside, referring to the convex/concave characteristic of the hill. The plan curvature (deg/m) corresponds to the hill shape when observed in its horizontal projection and is defined as a second-order derivative from the contour lines. The aspect is defined as the azimuthal angle corresponding to the highest declivity of the terrain, in the descendent direction.[62] To use it for classification purposes, it is recommended that the aspect should be converted to directional components. It is measured in degrees, from 0 deg to 360 deg, and varies according to the orientation (N-S and E-W). The TWI was developed by within[63] the runoff model TOPMODEL. The TWI is defined as

$$\mathrm{TWI} = \ln(a/\tan\beta), \qquad (6)$$

where a is the upslope area per unit contour length; and $\tan\beta$ is the local slope of the ground surface.

Locations with a large upslope area receive a high index value and are expected to have relatively higher water availability than locations with a small upslope area that are assumed to have relatively lower water availability and therefore receive a small index value.[64] The TWI is designed to quantify the effect of local topography on hydrological processes and for modeling the spatial distribution of soil moisture and surface saturation.[65]

2.4 Sampling

A stratified sampling was applied, in which each class was sampled proportionally to their area based on previous classification works of the region.[3,44,66] The proportion was not exact, but served as a base to guide the number of samples. The samples are pixel-based and were separated in the proportion 70/30 for training and validation. The sampling was manually realized based on high resolution imagery from Google Earth. To minimize errors regarding spatial auto-correlation, the collection of samples took into consideration a standardized geographic block, so no area is under or over sampled. A regular rectangular grid was used to collect the samples, in which every polygon ($25\ \mathrm{km}^2$) should contain a minimum amount of samples for each class, guaranteeing a satisfactory distribution of the samples on the whole catchment area. In some cases, as the landslide class, which are more concentrated in a specific region, some polygons of the grid did not contain their samples.

Regarding the sample size,[67] suggests a minimum of 50 samples for each class, increasing to 75 to 100 for large areas (more than a million acres) or if the classification has a large number categories (i.e., more than 12 classes). Even though no specific number of samples was established for the construction of the sample set, every class had at least more than 50 points. Some classes as landslides and water, had notably less presence in the area compared to forest and silviculture, which had considerably fewer sample points. As a matter of fact, a first attempt was made to divide the landslide class into two sub-classes "upper" and "deposit," to differentiate the area where the scar starts, in the higher part of the hill, and the area of material's deposit. However, in the 10-m spatial resolution data, this could not be separated, so only one class was created.

Previous to the choice of the classes that would compose the classification, a detailed study of the area was done. A solid knowledge of the elements that are contained in that landscape was built and all possible disturbances that would present similar fingerprints on spectral, temporal, and metrical dimensions of the landslide class were identified. After this analysis, the main classification labels were chosen, being: landslide, forest, silviculture, agriculture, pasture, bare soil, water, and urban area. Once the number of sample points for each class is substantially different, which could create a negative consequence for under sampled classes, the Monte Carlo simulations method was used in the validation process guaranteeing a stable result and will be explained further in this chapter.

2.5 Classification

The RF classification algorithm was used for this procedure. Among a variety of classification algorithms, the RF is frequently used in remote sensing applications. Furthermore, this classifier can be successfully used to select and rank the variables with the greatest ability to discriminate between the target classes.[68] Considering that one of the aims of this study is to search for the best combinations of indexes and predictors, this classifier was chosen. This process was conducted using the scikit-learn package in Python programming language. The classification was realized under four distinct approaches, in which each of them considered a different set of input predictors. As a result, four classification products were compared (Fig. 4). The input approaches are:

1. Mono-temporal: single date image (after the landslides occurrence) with NDVI, NDBI, and SAVI + geomorphometric attributes. Total of 10 predictors.
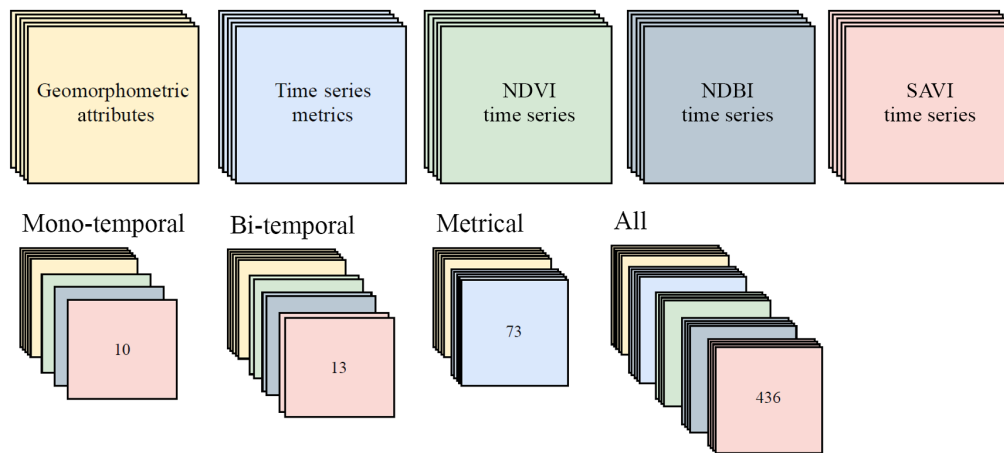
Fig. 4 Classification approaches. The types of inputs are separated by colors, and represented respectively: geomorphometric attributes (yellow), time series metrics (blue), NDVI time series (green), NDBI time series (gray), and SAVI time series (red). Below, the dataset of each approach is represented and the number inside the box reveals the quantity of predictors. Note that for mono and bi-temporal approaches, only one or two layers of NDVI, NDBI, and SAVI are illustrated.

2. Bi-temporal: two dates images (before and after the landslides occurrence) with NDVI, NDBI, and SAVI + geomorphometric attributes. Total of 13 predictors.
3. Metrical: time-series metrics for NDVI, NDBI, and SAVI + geomorphometric attributes. Total of 73 predictors.
4. All: 122 dates time series with NDVI, NDBI, and SAVI + time-series metrics for NDVI, NDBI and SAVI + geomorphometric attributes. Total of 436 predictors.

The time-series datasets were composed of indexes extracted from images within the period of November 2015 to Mai 2020. From each date, the NDVI, NDBI, and SAVI index were extracted, which resulted in three time-series datasets. From each time series, separately, the times series metrics were extracted. To form the dataset called "All," e.g., the three time series datasets (NDVI, NDBI, and SAVI), the three time series metrics datasets extracted from the latter and the geomorphometric attributes were piled together into one single stacked input. This means that when the classification is performed, all data are considered, and each layer is interpreted as a variable.

Once most of the landslide scars occurred during an extreme precipitation event in January 5, 2017, the aim was to find the closest images from that date to build the mono and bi-temporal datasets. However, this period presents high levels of cloud coverage, interfering with the images quality. For that reason, the chosen closest images to the date of the event are from June 9, 2016 (before), and March 11, 2018 (after). This reveals the difficulty of choosing a few images to classify landslide-prone areas. Figure 5 shows the two dates chosen for the bi-temporal approach.

One of the RF outputs is the predictors' importance ranking (IR), which is an ordered list related to the relevance of each variable to the classification. In a range of 0 to 1, all predictors are assigned to a value according to their importance, and the total sum of the predictors results in 1.

The proposed methodology focus on finding the most accurate and simple approach to tackle the classification problem. The idea of starting from a single-image dataset, progressively increasing until the time series, is applied to guarantee the statement "the simpler, the better." No unnecessary procedure should be performed before testing less complicated steps before. In this study, once one of the objectives is to understand the contribution of the time series for the LULC classification and landslide detection, the evaluation of these four approaches is fundamental.

2.6 *Validation and Analysis*

The reference material for the validation of the products was based on Google Earth's high spatial resolution imagery, as suggested by Olofsson et al.,[69] and an inventory map provided by Quevedo et al.[3] and updated for this research. This inventory was realized by visual
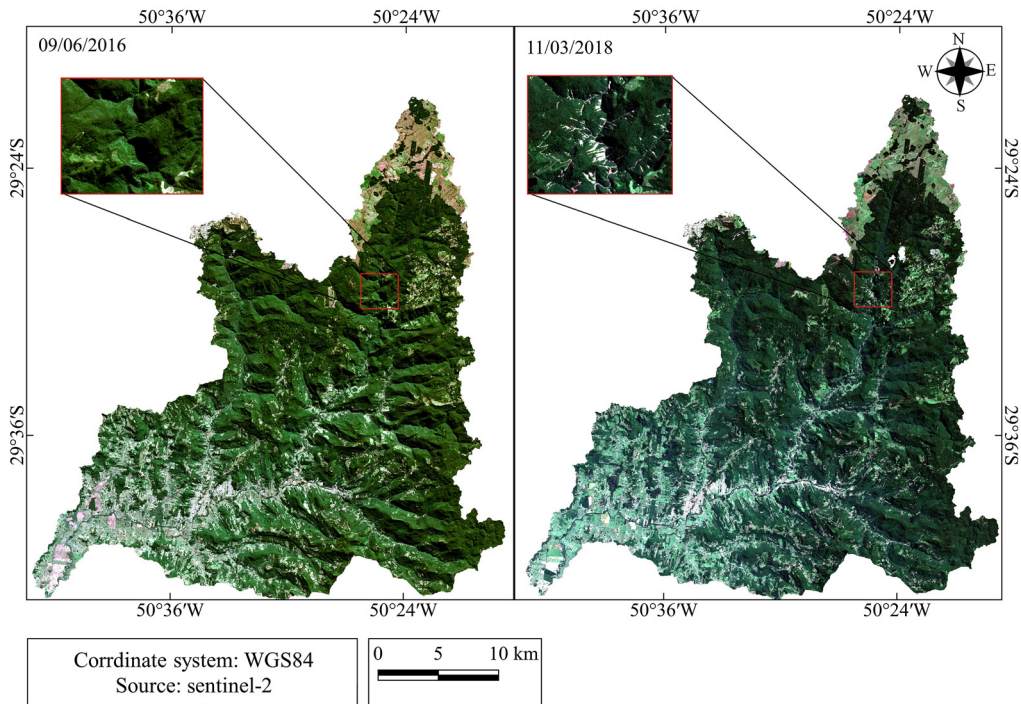
Fig. 5 Bi-temporal approach. Chosen images from before and after the landslides occurrence. It is worth noting that the closest cloud-free image before the event is from June 09, 2016 and the closest after is from March 11, 2018. There is 1 year and 9 months difference from one image to another.

interpretation of high spatial resolution imagery. The reference was used for the construction of a sampling set for the classification.

To compare different accuracy results, it is important to establish a credibility interval. In case, this interval is not considered, a comparison based on one single value can lead to mistaken conclusions. This happens because different training and validation set of samples can produce different results. Figure 6 shows how this procedure was conducted via Monte Carlo simulations.

The classifications evaluation and performance analysis was held under the statistical measures derived from the confusion matrix. The confusion matrix reveals every validation sample point, comparing their true label and the predicted one. From this matrix, the most relevant confusion between classes can be detected and interpreted. The overall accuracy (OA) is extracted from the matrix according to the formula expressed below, which ranges from 0 to 1 (100% of accuracy). The Cohen's kappa index is[70] another method used to evaluate the model. The kappa index is a measure of true agreement, which indicates the proportion of agreement beyond that expected by chance.[71] In other words, the achieved beyond-chance agreement as a proportion of the possible beyond-chance agreement.

3 Results

3.1 *Predictors' Importance Ranking*

The IR is extracted directly from the RF workflow. For each dataset (mono-temporal, bi-temporal, metrics, and All), the classification was performed, and the predictors' ranking comes as an output of the RF classifier using scikit-learn package. As mentioned before, each layer of the dataset is interpreted as a variable, so it could be a specific image from the whole NDBI time series, e.g., NDBI (June 09, 2016).

Table 1 gives the top 10 predictors, by each classification approach, and their corresponding importance values. More focus should be given to the top five predictors, once the mono-temporal approach considers a total of 10 predictors only. To clarify, all selected metrics that
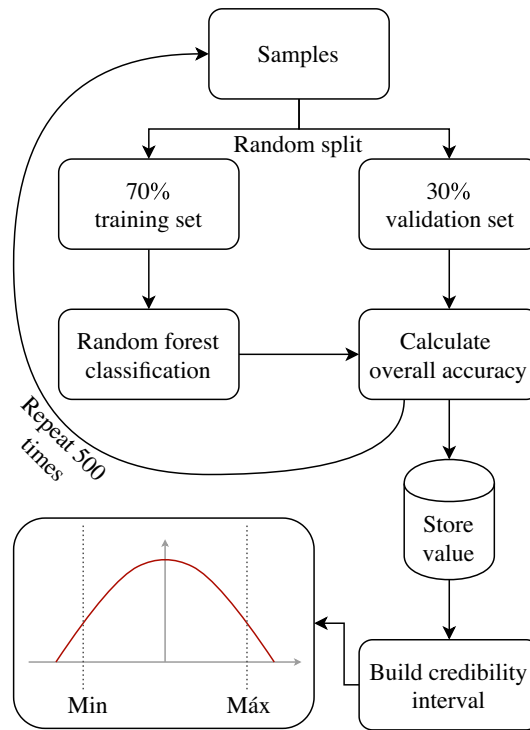
Fig. 6 Monte Carlo simulations.

Table 1 Predictors' IR.

	Mono-temporal (10 at.)		Bi-temporal (13 at.)		Metrical (73 at.)		All (436 at.)	
	IR	Predictor	IR	Predictor	IR	Predictor	IR	Predictor
1 deg	0.247	NDBI (March 11, 2018)	0.162	NDBI (June 09, 2016)	0.056	Area s2	0.0201	NDBI (June 09, 2016)
2 deg	0.195	NDVI (March 11, 2018)	0.148	NDBI (March 11, 2018)	0.036	Slope	0.0177	Slope
3 deg	0.144	SAVI (March 11, 2018)	0.146	NDVI (March 11, 2018)	0.035	Min	0.0163	NDVI (June 09, 2016)
4 deg	0.130	Slope	0.107	SAVI 2018	0.030	First slope	0.0154	NDVI (December 26, 2018)
5 deg	0.087	DEM	0.106	NDVI 2016	0.029	Area s4*	0.0137	Min
6 deg	0.046	TWI	0.079	Slope	0.028	Polar balance	0.0135	Area s1
7 deg	0.040	Aspect	0.071	SAVI 2016	0.027	Area s4	0.0121	Area s2
8 deg	0.039	Prof. curvature	0.053	DEM	0.027	Sum	0.0120	NDBI (November 22, 2015)
9 deg	0.038	Plan curvature	0.027	TWI	0.026	Mean	0.0107	NDBI (September 17, 2016)
10 deg	0.031	Gen. curvature	0.024	Prof. curvature	0.026	Abs. Sum	0.0101	NDBI (March 21, 2019)

appear at metrical and All top 10 are related to the NDBI. Only area s4, from metrical, was derived from the NDVI time series. For that reason, to reduce the amount of information in the table, the difference between NDBI and NDVI metrics is represented by a *. Metrics related to the SAVI index were not selected among the top 10 for neither approaches.

All selected metrics that appear at metrical and All approaches are related to the NDBI. Only area s4*, from metrical, was derived from the NDVI time series. The most important predictors for all four approaches are either an NDBI image itself or a metric extracted from its time series. In order, from mono-temporal to All approaches, the first predictors selected are the NDBI image from 2018, the NDBI image from 2016, the area s2 metric from the NDBI, and the NDBI image from 2016. As explained above, the NDBI metrics represented the majority of metrics predictors in the top 10 for metrical and All. The NDVI also presented high IR values in all approaches, being among the top 5. The NDVI image from 2018 was selected by the two first approaches, the area s4 from the NDVI for metrical, and the NDVI images from 2016 and 2018 for All. For the mono and bi-temporal approach, the date from 2018 selected to compose the dataset is March 11, 2018, once it was the first date after the event that presented a cloud free image for the whole catchment area. However, analyzing All's top 10 selection, one can see that the March 11, 2018 date was not chosen. Instead, the first image selected after the event occurrence was the NDVI from December 26, 2018, which is nine months after. This may reveal that not necessarily the closest image to the event's date is the most appropriate to be chosen as an predictor, even both being cloud free images.

Regarding the geomorphometric predictors, the slope stood out, being the first of them, and sometimes unique, to be chosen in every approach. It was selected among the top five for all of them, in addition to the bi-temporal approach, which got the sixth place. With respect to the metrics, it is evident that the polar metrics showed significant importance. For metrical approach, the first predictor chosen was the area s2, and in addition, NDVIs Area s4 is also among the top five. Moreover, polar balance and NDBIs area s4 were selected. About All's approach, from the three metrics selected among the top 10, two of them are polar, NDBIs area s1 and area s2.

3.2 Time-Series Metrics

There are many possibilities for analyzing the performance of a classification map. In this research, the discussion will be conducted by the analysis of the confusion matrix and its statistical products, such as the OA, kappa index, and commission and omission errors. Moreover, a deeper visual analysis of the maps is described, showing in detail some differences among the approaches regarding each class. A similar discussion is developed, showing the contribution of the metrics for the classes prediction. In addition, the metrics values for each class are analyzed via boxplots, linear, and polar graphics.

From each time series (NDVI, NDBI, and SAVI), 22 metrics were extracted. All metrics had their values normalized, being their range between 0 and 1. For analyzing the potential of the metrics for identifying each class, some NDBI metrics were chosen to be exposed. The criterion to select only NDBI metrics was the fact that this index presented the most relevant results at the predictor's IR (Table 1). Moreover, this visual analysis is just realized for the classes that presented the most revealing characteristics when interpreting the metrics images, being landslides, agriculture, and silviculture classes. It is important to emphasize that the final formats of the time series metrics are raster files where in each pixel the value will correspond to the output computed from a mathematical equation taking into account all the observations of the time series for this pixel.

The metrics in which Landslide scars stood out are represented in Fig. 7. The red points indicate the location of some of the landslides. The first three metrics, katz, polar balance, and first slope, showed a quite similar result, presenting values close to 0 (blue) for this class.

Regarding the metrics' types, first slope, mean, and absolute mean derivative are basic; Katz is fractal; and polar balance, area s1, and area s2 are polar. Area s1 and area s2 were selected because of their contrast for that class, in which landslides do not appear in the first and are revealed in the latter. Mean shows a similar result to area s2, where landslides present a higher value (yellow), compared to the rest of the image, mostly with low values (blue).
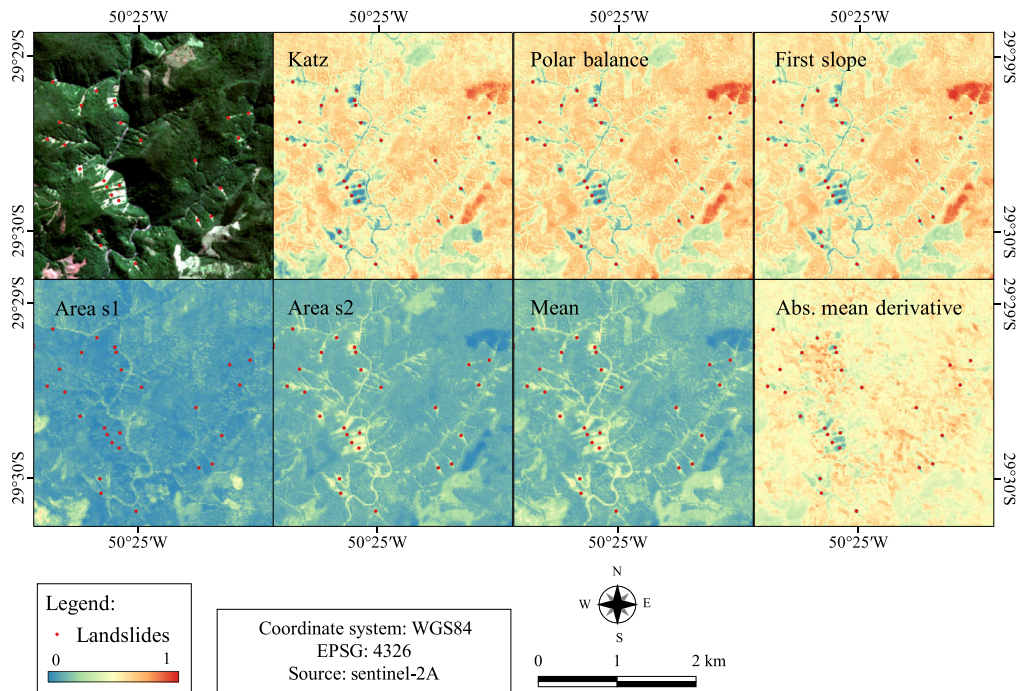
Fig. 7 NDBI metrics in details for landslide scars. Red points represent landslide areas of reference.
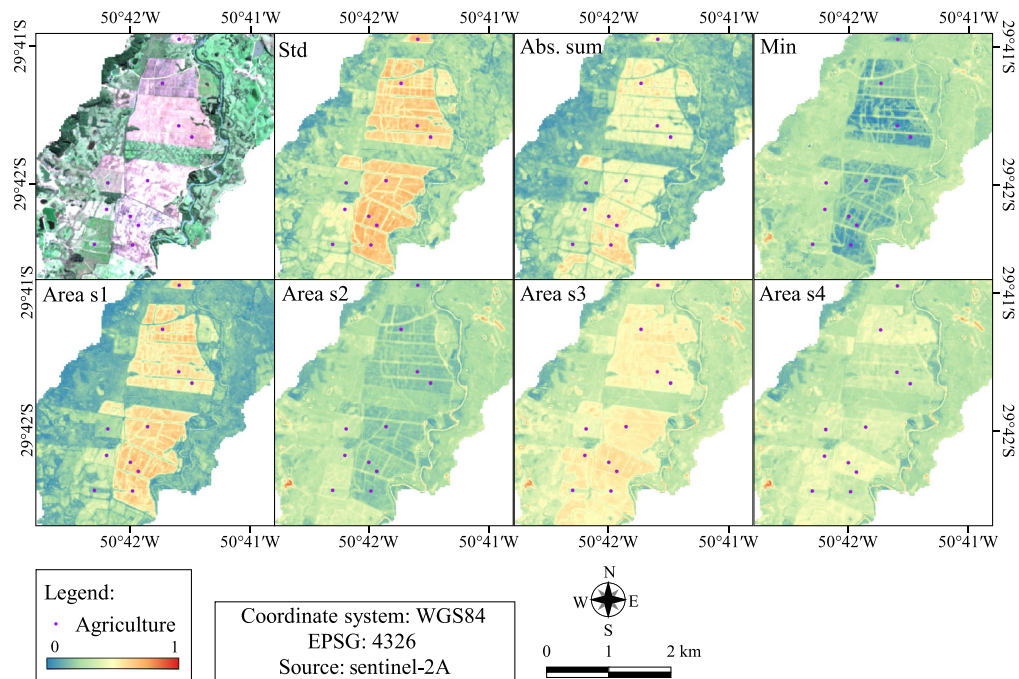

Fig. 8 NDBI metrics in details for agriculture class. Purple points represent agriculture areas of reference.

The metrics that presented the most relevant results for agriculture are exposed in Fig. 8. The std metric revealed very high values for most of the agricultural areas (orange/red). This is related to the great variability of the crops dynamics along the agricultural cycles. For absolute sum, agriculture also shows higher values compared to the rest of the image. An opposite result is the Min metric, in which this class presented the lowest values, probably related to the harvesting
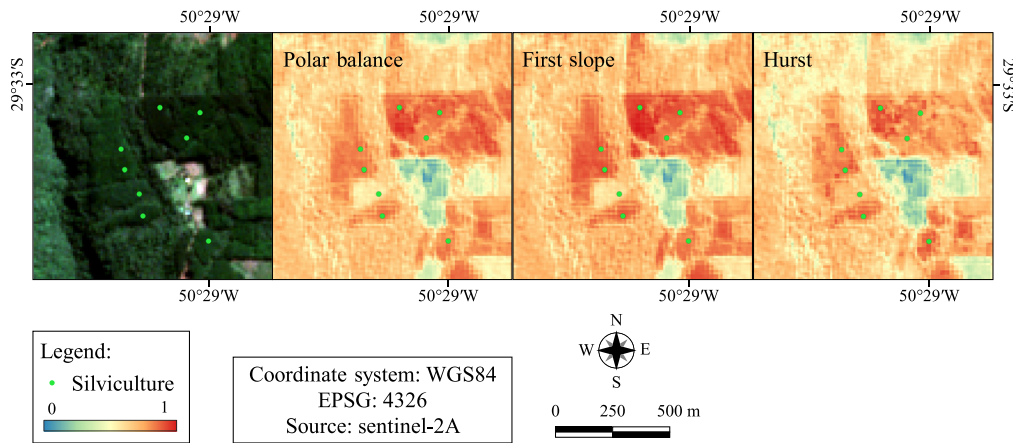
Fig. 9 NDBI metrics in details for silviculture class. Green points represent silviculture areas of reference.

periods. Finally, areas s1 to areas s4, confirm the result presented by the Std metric, in which the variability seems to be a key characteristic of agriculture.

The metrics for silviculture class are shown below (Fig. 9). Polar balance (polar), first slope (basic), and hurst (fractal) were the most revealing ones. The silviculture is represented by the light green dots. In all cases, this class presented the highest values compared to its' surroundings. A contrast can be identified compared to the forest, which presents a lower response but still in the orange/red range of values. This emphasizes the contribution of polar, basic and fractal metrics for discriminating silviculture from forest.

Basic and polar metrics are extracted from linear and polar graphs, which represent the time series. These graphs provide a relevant illustration, under two different perspectives, regarding the classes' spectral behavior along the time. As an example, classes landslide (Fig. 10), forest (Fig. 11), silviculture (Fig. 12), and agriculture (Fig. 13) are exposed below. The linear graphs are composed by the X axis, representing the time (each date of the time series from 2015 to 2020), and the Y axis, represents the NDVI values. The polar graph is divided into four rectangles, each of them representing a quarter period of the time series. Values from the first quarter period of the time series are plotted in the top-right green rectangle, in an anticlockwise direction, starting from the horizontal line that divides top-right (green) and bottom-right (blue) rectangles. In other words, the green rectangle contains values from the first season of the time series; the orange rectangle, from the second season; the brown rectangle, from the third season; and the blue rectangle, from the last season. In the polar graphs, values are plotted from the center point (intersection point among the four rectangles) to the borders. This means that the closer to the center, the lower the NDVI value; and the further from the center, the higher the NDVI value. Analyzing the landslide graphs (Fig. 10), one can notice that there are high values of NDVI at the
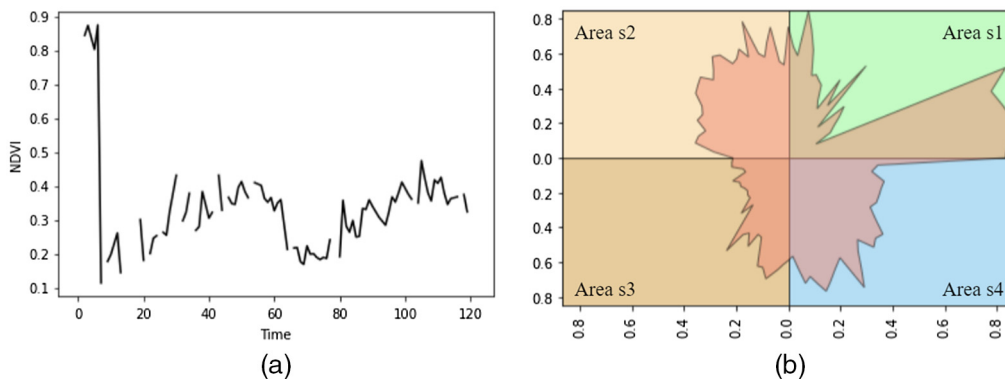

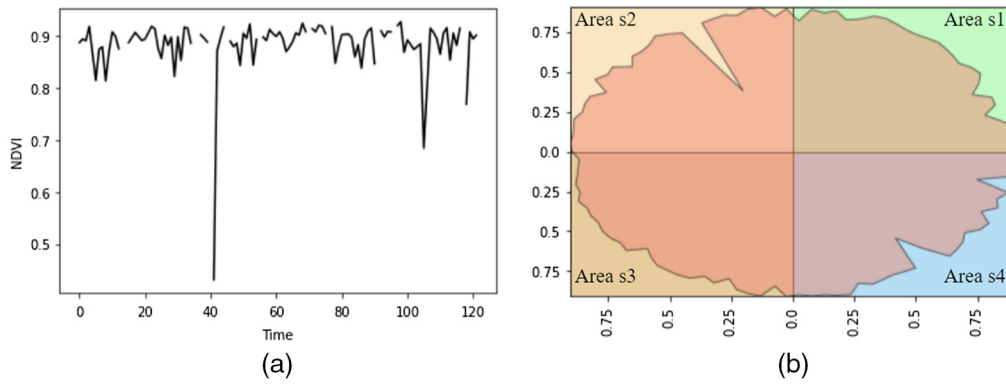**Fig. 10** Landslides linear and polar graphs: (a) linear and (b) polar.

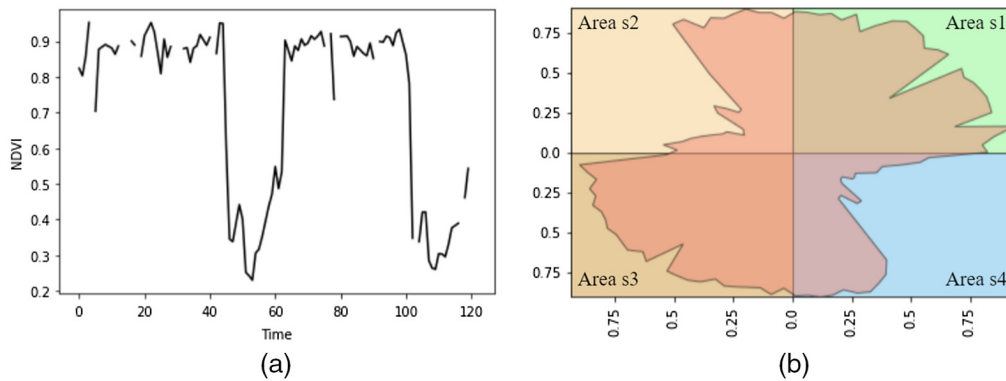**Fig. 11** Forest linear and polar graphs: (a) linear and (b) polar.

Fig. 12 Silviculture linear and polar graphs: (a) linear and (b) polar.
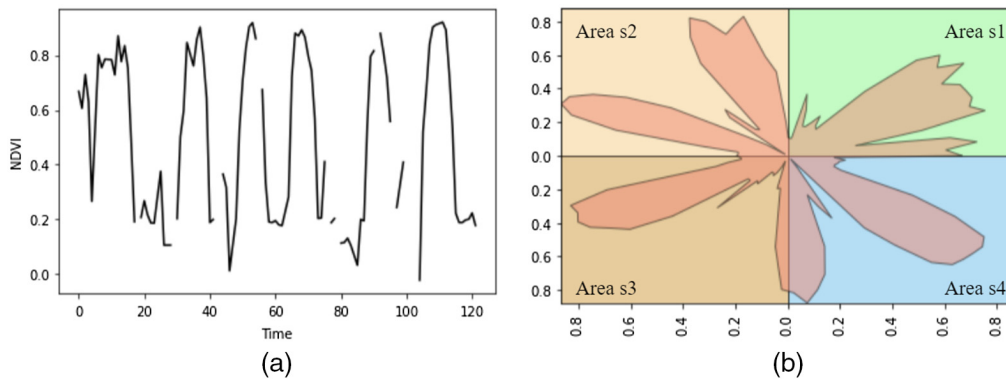

Fig. 13 Agriculture linear and polar graphs: (a) linear and (b) polar.

beginning of the time series, followed by an abrupt reduction. After that, two small cycles of growth and decrease are observed.

Forest graphs (Fig. 11) indicate a stable profile of high NDVI values. Some outliers can be observed by abrupt changes to low values, followed by high values again. In the polar graph, these outliers are represented by the teeth (triangles shapes), observed in the second and fourth rectangles.

The silviculture profile (Fig. 12) is represented in the graphs by high and two moments of very low NDVI values. This reveals the growing and harvesting periods. In the polar graph, the growing seasons are shown in the first and third rectangles and the harvest in the second and fourth. Agriculture class (Fig. 13) is also represented by growing and harvesting periods. However, for this class, the crops present shorter development cycles. In the linear graph, besides no similar value is repeatedly present, as in the stable forest profile, there is a regular pattern
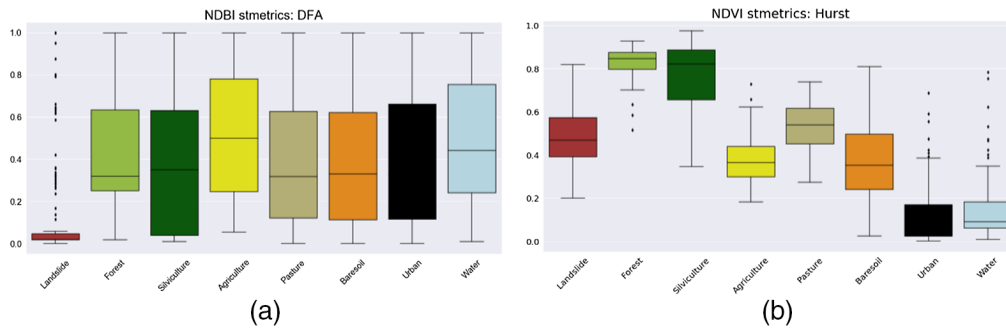
Fig. 14 Landslides boxplot analysis: (a) NDBI DFA and (b) NDBI hurst. Note that for the first metric, the landslide class box is concentrated on the bottom part with the lowest values.

frequency. This pattern is composed of short periods of high, followed by short periods of low NDVI values. In the polar graph, one can notice that, in general, there is more than one cycle present in each season, revealing the expressive variability of this class' behavior.

The metrics' contribution to classification was also interpreted via boxplot graphs. It is worth noting that the boxplots were not realized based on the classification results but on the extraction of information from all sample points for each class by each time-series metric output. These graphs provide a variety of information, as the lower and upper quartiles (represented by the limits of the bounding box), the median (line inside the box), and outliers (points). To understand if a specific metric can contribute to the identification of a class, it is recommended to find the boxplot with the least variability (small boxes), with the least amount of classes sharing the same range of values (boxes should be placed in different vertical positions). Boxplots can supply with information for setting threshold values to differ one class from another. Regarding landslides (Fig. 14), the detrended fluctuation analysis (DFA) metric extracted from the NDBI shows the most differentiating potential of this class from the others, in which its' upper quartile reaches up to 0.1, whereas for the other classes, it is above 0.6.

Forest and silviculture (Fig. 15) presented very similar results in most of the metrics. Both of them showed very high values for NDVIs polar balance, with the lowest quartile above 0.65,
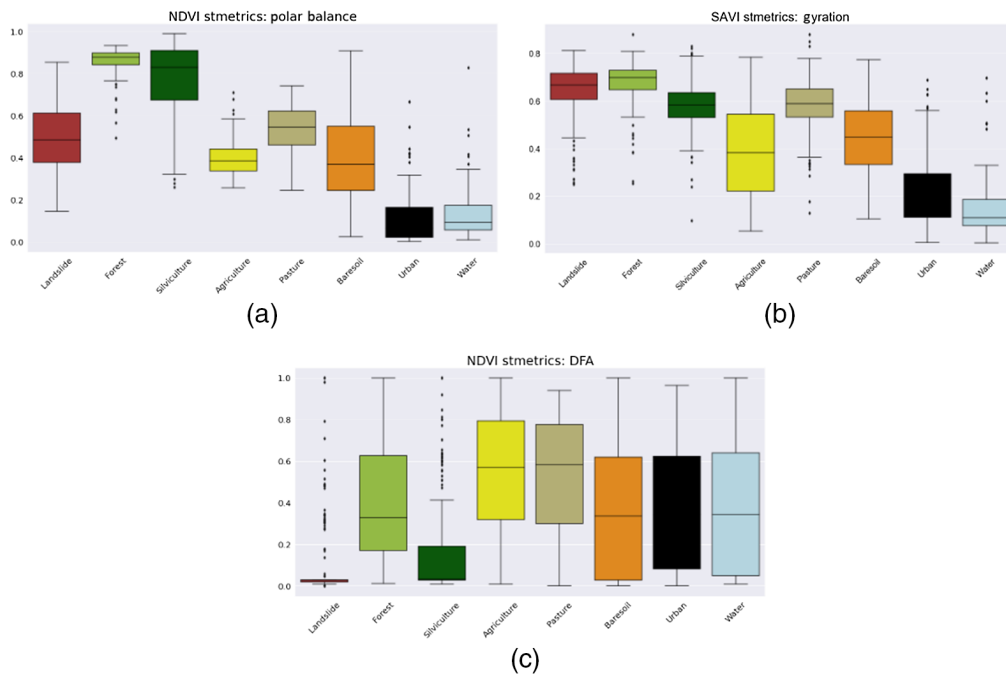

Fig. 15 Forest and silviculture boxplot analysis: (a) NDVI polar balance; (b) SAVI gyration; and (c) NDVI DFA. Note that forest and silviculture classes present the highest values for the Polar Balance extracted from the NDVI time series.
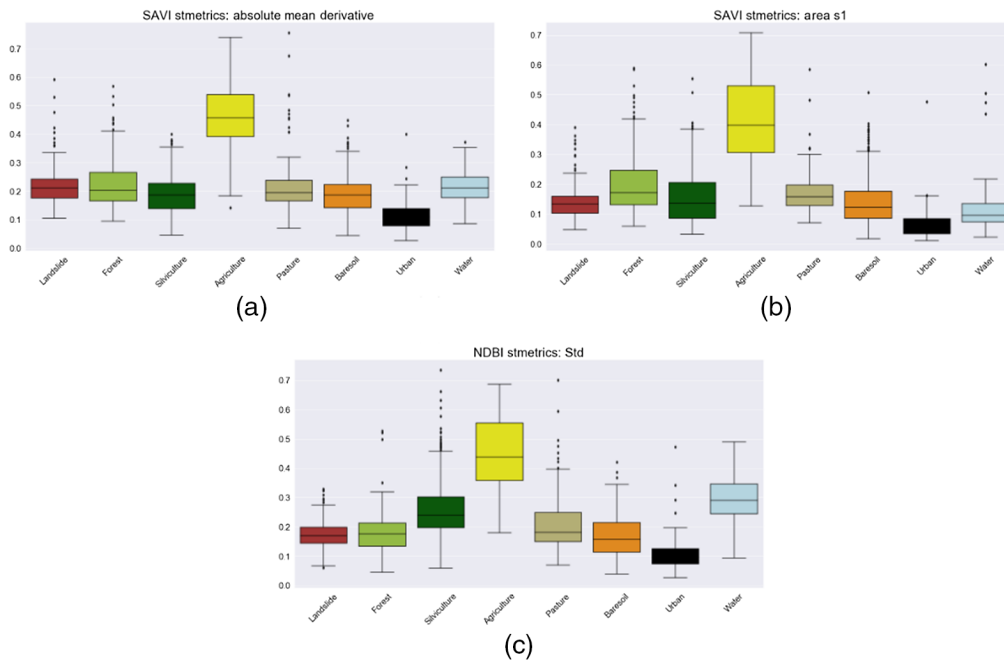
Fig. 16 Agriculture boxplot analysis: (a) SAVI Abs. mean derivative; (b) SAVI area s1; and (c) NDVI std. It is worth noting that agriculture class presents the highest values showing a well defined difference from other classes represented in the boxes.

whereas the other classes reached upper quartile values around 0.60. After separating the two of them from the rest, their differentiation can be done through the SAVIs gyration, in which forest's median is around 0.68 and silviculture's, 0.58. Another method for differing forest from silviculture is using the NDVIs DFA metric, in which the medians are around 0.32 and 0.05, respectively. The similarity between forest and silviculture might be due to their high and dense vegetation characteristics. Both polar balance and gyration are polar metrics, and DFA, fractal.

Agriculture class presented the most outstanding results for the boxplots analysis (Fig. 16). The metrics that better revealed this differentiation were SAVIs absolute mean derivative, SAVIs area s1, and NDBIs Std. In these metrics, the agriculture class shows medians values of at least 0.1 higher than the other classes. Moreover, in all cases its' lower quartile is higher than the other's upper quartile. Finally, in SAVIs absolute mean derivative metric its' lower quartile is above most of the other classes maximum range value. In addition area s1, which is polar, all other metrics are the type basic.

3.3 Land Use and Land Cover Maps

The LULC classification products derived from each approach is shown in Fig. 17. In this visual analysis, the salt-and-pepper effect appears more intensively on the mono-temporal map and gets reduced as the number of predictors increases. This can be seen more effectively on the southwest part of the basin, where a significant part of agriculture is placed. At the mono-temporal approach, the agricultural areas seem to spread to the whole area, whereas All approach shows more limited and well designed polygons of agriculture.

The confusion matrix is an effective tool to analyze the misclassifications between classes. The confusion matrix for each approach is shown in Fig. 18. Considering class by class, the landslide presented a significant proportion of confusion with forest, silviculture, bare soil, and water. The confusion with Water is just present in the mono-temporal approach, whereas in the others it no longer appears. Another class that requires attention is the agriculture, which showed high values of confusion, especially in the mono-temporal and bi-temporal approaches. For the mono-temporal, it presented more pixels classified as bare soil than the agriculture itself, followed by a notable confusion with pasture, urban, and silviculture. The confusion maintains in
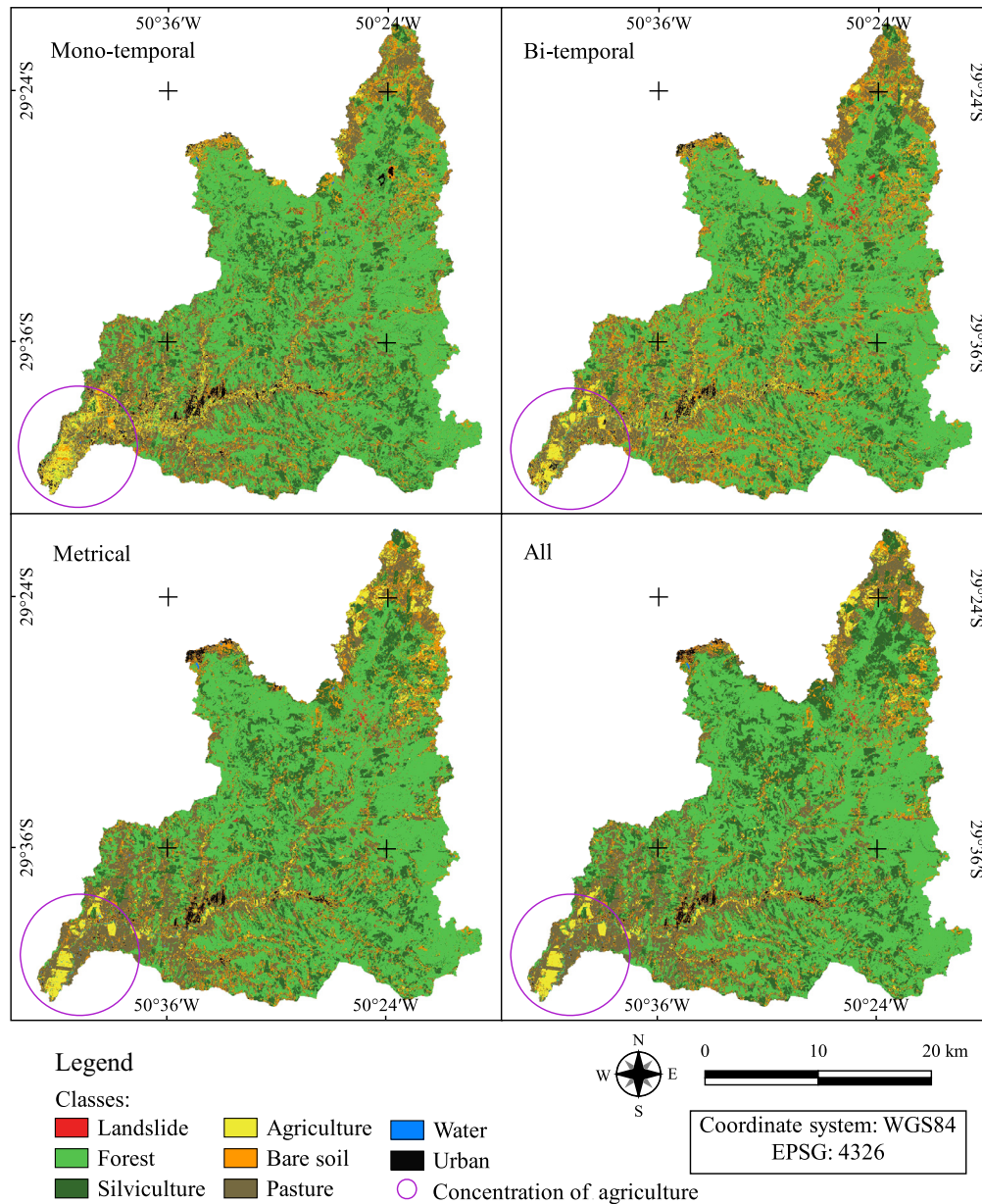
Fig. 17 Classification maps.

the bi-temporal approach, even though in a reduced level, where agriculture is misclassified with bare soil, followed by urban and pasture. On the other hand, this situation changes drastically when compared to the scenario for metrical and All, in which the confusion between the classes is low, presenting more errors for pasture and bare soil, respectively. Table 2 shows kappa indices, OA values, and their associated confidence interval.

First of all, the kappa values show an expressive improvement comparing the approaches. Starting from the mono-temporal, with 0.70, to the bi-temporal, with 0.79, around 0.09 is increased in the value, which represents more than 10% of improvement. From bi-temporal to metrical, a relevant increase is also noticed, however less intense, going from around 0.79 to 0.86. The highest value is presented by the All approach, with around 0.87. Comparing all kappa values, it seems that it increases in an exponential behavior, showing large improvement steps in the first two approaches and stabilizing after metrical. The same interpretation is visible in the OA values, where 7.64% is increased from mono to bi-temporal, 5.31% from bi-temporal to metrical, and only 1.06% from metrical to All.
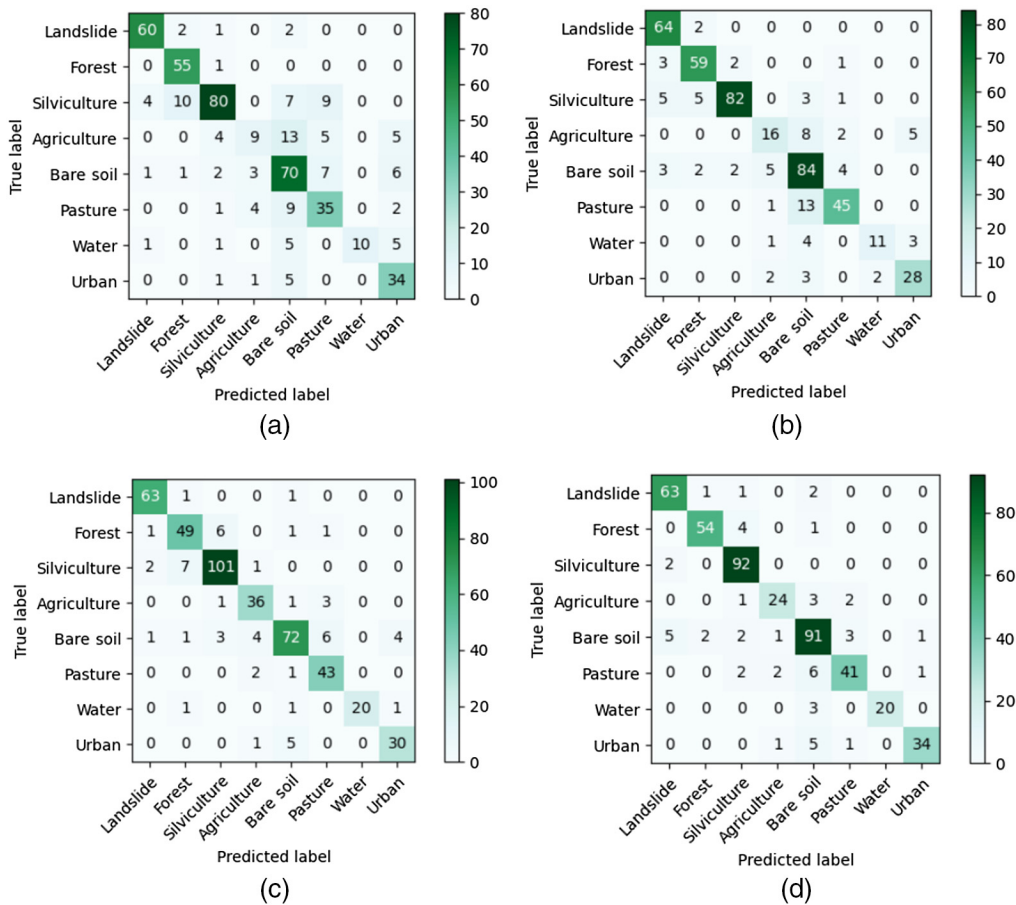
Fig. 18 Confusion matrices. (a) Mono-temporal; (b) bi-temporal; (c) metrical; and (d) all. The diagonal shows the correctly classified pixels, the stronger the green tone color, the higher the number of correctly classified pixels.

Table 2 Kappa index and OA.

	Kappa (κ)	OA (%)	Credibility interval for OA
Mono-temporal	0.7043	74.95	74.78% to 75.11%
Bi-temporal	0.7938	82.59	82.50% to 82.79%
Metrical	0.8574	87.90	87.84% to 88.09%
All	0.8693	88.96	88.81% to 89.05%

Another important factor to analyze a classification accuracy is the evaluation of the commission and omission errors. In that sense, Tables 3 and 4 show the statistical products depicted for each separate class and approach.

Concerning the omission errors, the mono-temporal approach presented the best result for only one of the classes, forest, with 1.79%. The bi-temporal approach showed the lowest error rate only for landslide class, with 3.03%, followed by metrical, with 3.08%. Metrical also presented the best results for agriculture (12.20%), pasture (6.52%), water (13.04%), and urban (16.67%). All's approach performed best with the classes silviculture (2.13%), bare soil (13.33%), and water (13.04%), with the same value as metrical for this class. Thus, metrical had the best results concerning omission errors, with the lowest error rates in five of the eight classes. Furthermore, even though mono and bi-temporal presented very satisfactory results for

Table 3 Omission errors.

	Mono-temporal (%)	Bi-temporal (%)	Metrical (%)	All (%)
Landslide	7.69	3.03	3.08	5.97
Forest	1.79	9.23	15.52	8.47
Silviculture	27.27	14.58	9.01	2.13
Agriculture	75.00	48.39	12.20	20.00
Bare Soil	22.22	16.00	20.88	13.33
Pasture	31.37	23.73	6.52	21.15
Water	54.55	42.11	13.04	13.04
Urban	17.07	20.00	16.67	17.07

Table 4 Commission errors.

	Mono-temporal (%)	Bi-temporal (%)	Metrical (%)	All (%)
Landslide	9.09	14.67	5.97	10.00
Forest	19.12	13.24	16.95	5.26
Silviculture	12.09	4.65	9.01	9.80
Agriculture	47.06	36.00	18.18	14.29
Bare Soil	36.94	26.96	12.20	18.02
Pasture	37.50	15.09	18.87	12.77
Water	0.00	15.38	0.00	0.00
Urban	34.62	22.22	14.29	5.56

one of the classes, they also showed the highest omission error percentage for other classes; for instance, 75% and 48.39% for agriculture, and 54.55% and 42.11% for water, respectively.

Regarding the commission errors, the mono-temporal approach presented the best result only for water, with 0%, similar to metrical and All. The bi-temporal approach showed the lowest error rate only for silviculture, with 4.65%. Metrical presented the best results for landslide (5.97%), bare soil (12.20%), and water (0%). All's approach performed best with the classes forest (5.97%), agriculture (14.29%), pasture (12.77%), water (0%), and urban (5.56%). Thus, All had the best results concerning commission errors, with the lowest error percentages in five of the eight classes. Moreover, as in omission error, despite mono and bi-temporal presenting very satisfactory results for one of the classes, they also showed the highest omission error percentage for other classes; for instance, 47.06% and 36% for agriculture, 36.94% and 26.96% for bare soil, and 37.50% and 15.09% for pasture, respectively.

In addition, a relevant information regarding the landslide prediction is that, even though the bi-temporal presented the best results for omission error, the difference from the metrical was very narrow, about 0.05%. However, in regard to commission error, metrical presented the best result (5.97%) with about 8.7% of difference from bi-temporal (14.67%), which showed the highest values among all approaches.

An example of this commission error comparison concerning the landslide class is shown in Fig. 19. The landslide inventory reference is represented by the red polygons on top of the Sentinel image. The four maps below represent each classification product derived from the approaches. Comparing the bi-temporal approach to the others, it is visible that it could indeed detect all or most of the landslide scars. However, it is also noticeable that it confused most of the
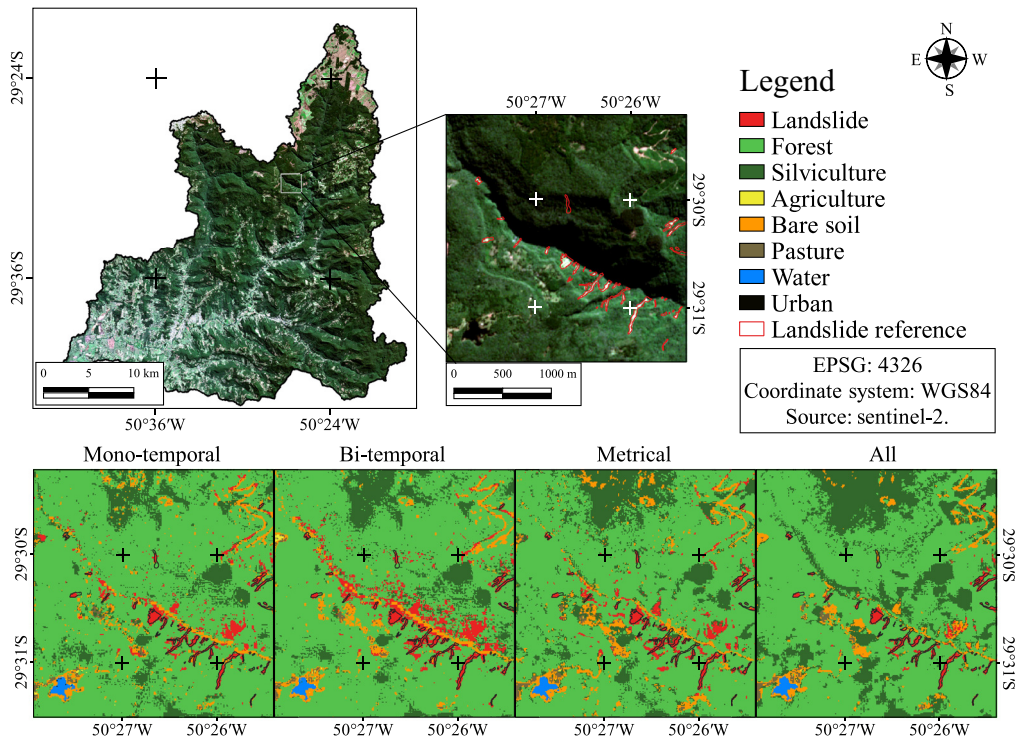
Fig. 19 Landslides in details. The reference landslide scars are represented by the polygons with a black line on the four maps on the bottom.

forest area from the hillside with landslides. In this specific area, visually analyzing, All approach performed better, presenting defined polygons (low salt-and-pepper effect) well and very satisfactory detection of the landslide scars, compared to the inventory reference.

The difference map is another method used to analyze the classification performance regarding the landslide scars. In Fig. 20, the inventory reference is represented by the dark blue polygons on top of the Sentinel image. From that vector data, a raster reference was created, in which every pixel inside the polygons is represented in red as landslide class, and everything else is labeled as non-landslide and represented in gray. The first four maps below show the classification result for each approach. To clarify, in this case, the classification map from Fig. 17 was used, and the landslide was separated from all the other classes, which were gathered in a unique class called non-landslide. The last four maps below represent the output of a simple difference between the reference and the prediction, produced with band arithmetic. The blue color stands for correctly predicted areas, whereas brown and orange represent omission and commission classification errors, respectively.

Observing the difference maps, it is clear that the amount of commission errors is significantly superior to omission errors. This information is also confirmed by the commission and omission error table, where the first presents considerably higher values than the latter. Furthermore, comparing the four approaches, All presented the best result, with significantly more areas of correctly prediction.

The confusion matrices (Fig. 18) reveal that landslides showed confusion with silviculture for the four approaches, especially for mono and bi-temporal. This analysis can be developed in Fig. 21. The two first Sentinel images represent an area only composed of forest, containing silviculture in the central part, and no landslides scars. Silviculture class is clearly apparent in the 2018s image, where it has been harvested. In addition, its texture in 2016s image is also notably different from the surrounding forest. The classification results are displayed on the four maps below.

Mono-temporal approach presented the least satisfactory output, with most of the silviculture area classified as pasture, bare soil, and landslides scars. Also, the bi-temporal approach wrongly classified a considerable amount of silviculture pixels as landslide. Metrical and All presented
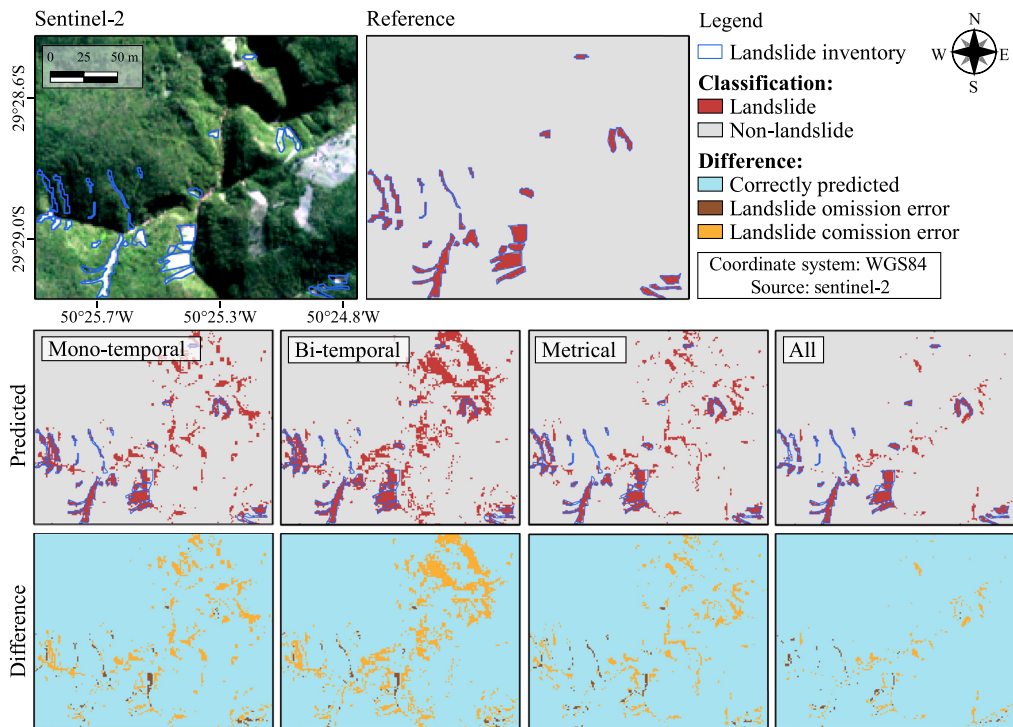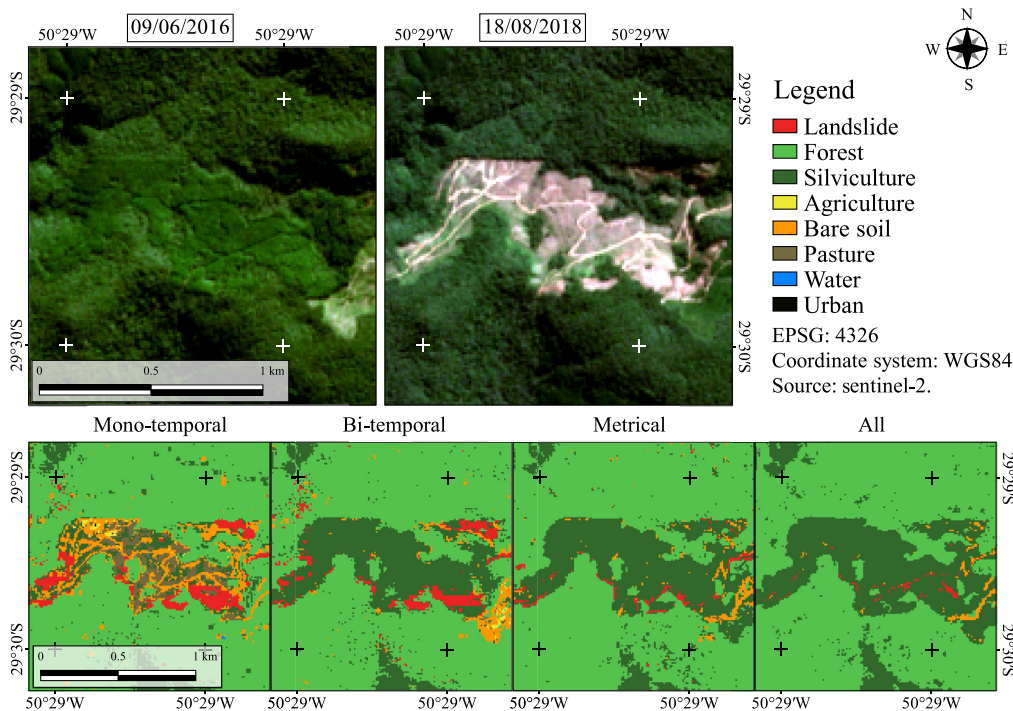
Fig. 20 Landslide difference map.

Fig. 21 Landslide versus silviculture. It is worth noting that in the mono-temporal approach the whole silviculture area was classified as pasture or bare soil, whereas from the bi-temporal to all approach the error is substantially reduced.

the best results, with the most part correctly predicted as silviculture. These results reveal, again, that one of the most challenging obstacles to landslide detection is to reduce the commission errors. Once the mono-temporal approach does not have the previous information of silviculture before the harvest, it is comprehensible that the prediction classifies the area as bare soil and

pasture, for example. Bi-temporal approach presented a more accurate result (visually), however the only reason for this improvement is the fact that this specific silviculture area has been harvested exactly on the same period of time where the two dates were used for bi-temporal input. However, if the harvesting had been done before the first date, maybe it would have presented a result close to mono-temporal.

Another visual analysis comparison among the approaches is shown in Fig. 22. The three columns refer to "a," "b," and "c" areas spatially represented by the red rectangles. Both "a" and "b" focus on the analysis of the agriculture class, and "c" is urban. Each row shows the result
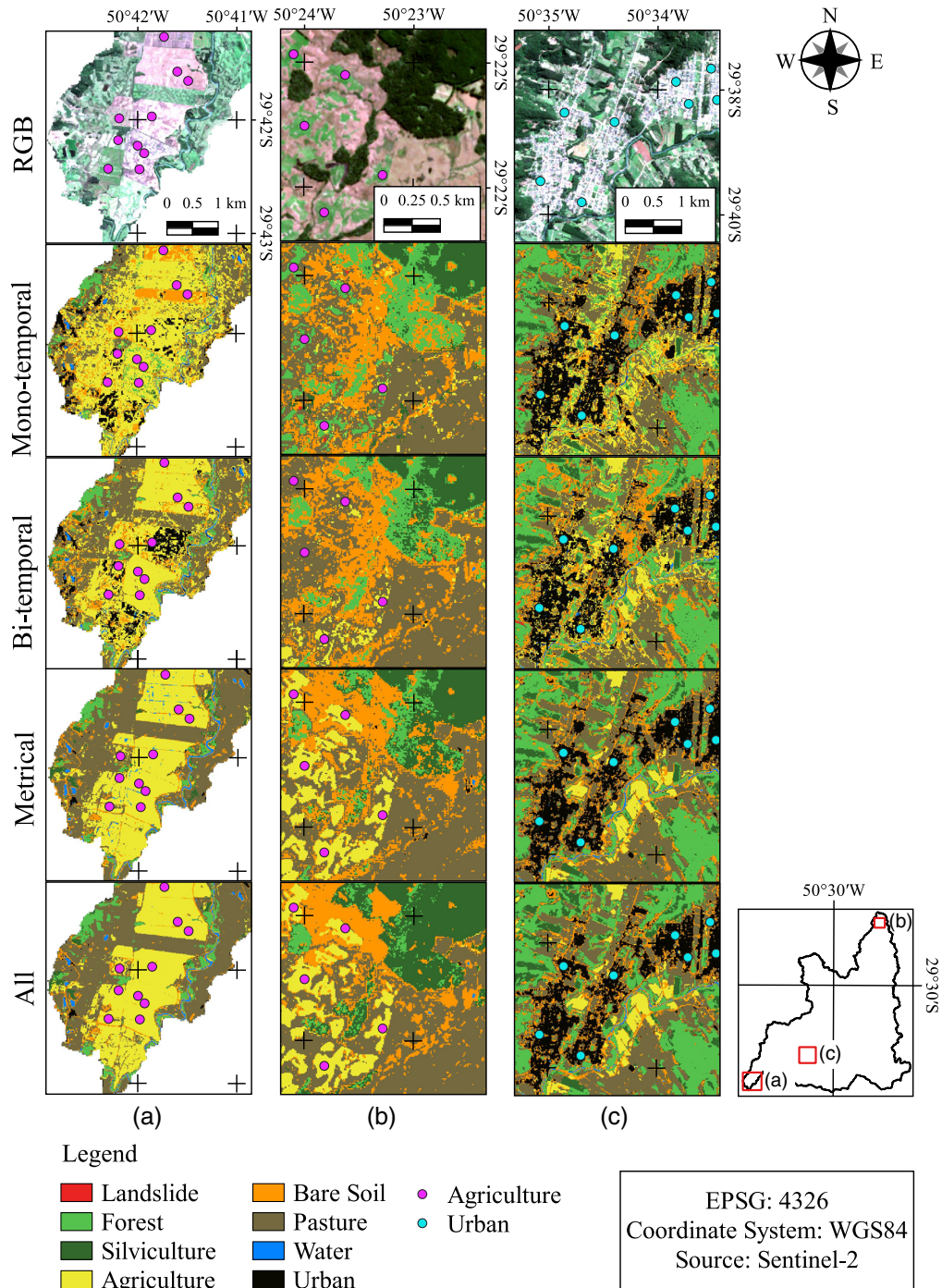

Fig. 22 Classification maps in details. (a) and (b) Examples of agricultural areas. (c) Example for urban areas. The pink and blue points are references for each class. The RGB image is a composition of the following Sentinel-2 bands: R (B4), G (B3), and B (B2).

for a specific approach. The pink points represent Agriculture areas of reference, whereas the blue points, Urban. In "a" area, it is observable that from mono-temporal to All, the agriculture polygons progressively gain a more solid and well designed shape. The salt-and-pepper effect is more expressive at mono and bi-temporal results, which practically classified the whole area as agriculture. In addition, both of these approaches also misclassified agriculture pixels with urban, which does not exist in this area. In "b" area, the agriculture areas are only well detected by metrical and All approaches, being confused with forest, pasture, and bare soil by the other methods. An hypothesis for that might be the fact that for agricultural purposes, the phenological cycle, provided by the time series, is very relevant. If the agricultural crop is still low or very high, and if that is the only information available for mono or bi-temporal approaches, it is comprehensible that confusion might happen with these classes.

Finally, in "c" area, the Urban class can be identified as the pixels concentrated in the quarter blocks divided by linear streets at the Sentinel image. Once again, the metrical and All approaches presented more well defined polygons, with less spared pixels (salt-and-pepper). The urban cluster at these approaches presents consolidated black polygons, with some bare soil, which could be explained by unpaved streets. However, on mono and bi-temporal, many agricultural areas are represented in the middle of the urban center, which does not matches the reference analyzed in Google Earth.

4 Discussion

Considering the time-series metrics, polar metrics showed high importance positions in the rank (Table 1). For metrical approach, the first predictor chosen was the NDBIs area s2 and NDVIs Area s4 is among the top five. Moreover, NDBIs polar balance and NDVIs area s4 were selected on the top 10. Considering All's approach, from the three metrics selected among the top 10, two of them are polar, NDBIs area s1 and s2. The slope was among the top 10 predictors for every approach. This can be explained by the fact that landslide scars tend to occur in areas where slope values are high, which can be decisive for separating this class from bare soil. In addition, the NDBI was also very present in the first positions of the ranking. Once this index is focused on built-up areas, its contribution can be understood by the fact that three types of classes in this study might present very similar spectral responses, being bare soil, landslide, and urban area. Brazilian roofs are usually made of ceramic, which is mostly clay, explaining the reason for the confusion among these classes. In this sense, the NDBI presented useful values to improve the classification.

Visual analysis of the time-series metrics revealed that, especially for landslides, agriculture, and silviculture, some metrics could support their detection, as, e.g., Katz, Std, and First Slope, respectively. Moreover, variability seems to be a key characteristic of agriculture, as shown in areas s1 to s4 and Std (Fig. 8). As each area represents a period in the time series, the comparison among the four of them could reveal a grown crop in the first period (high values for area s1), harvest season in the second (low values for area s2), followed by another cycle of growing crop (values increase in area s3), and harvest (values decrease in area s4). Once the time series considered in this study is very large, it is possible that more than one agricultural season is present in each period represented by the areas (i.e., seeding and harvesting in the same area period). Among the selected metrics, std, absolute sum, and min are basic; areas s1 to s4 are polar.

The profiles in Fig. 10 shows an area of high vegetation, probably forest, that suffered the removal of the vegetation due to the landslide event. The attempts of reconstructing the vegetation in the area are indicated by this two growing cycles, however, the high susceptibility condition for landslides, as high slope values, might hinder the vegetation settlement. The decreasing curves, might point to the repetition of the landslide in this area, maintaining the landslide scar. This reveals the difficulty of vegetation recovery in landslide scars, and the possibility of multiple occurrence of the event in the same spot. Very different from landslide, the forest polar graph illustrates a wide and continuous circle, emphasizing the stability of this class behavior along the time (Fig. 11).

In addition, the boxplots interpretation provided more insights on how the time series metrics could contribute to the classification. As an example, some metrics that stood out were the

absolute mean derivative for agriculture, area s2 for urban, and DFA for landslides. In the latter (Fig. 14), even though it presents a considerable quantity of outliers, the result for landslide is notably different from the others. Once silviculture also shows low values for this metric, one strategy for landslides detection could be the use of the DFA for separating landslides and silviculture from the rest, and then use the NDVIs hurst metric to distinguish one from another. In the latter, landslide values, in general, range around 0.40 to 0.58, when silviculture ranges from 0.65 to 0.88. Maybe, the low values for landslides and silviculture might occur because of their abrupt change response in the time series when the vegetation is removed, purposely in the case of the latter. One should notice that for landslide detection, the metrics that revealed the best results where both of the type fractal.

Comparing the approaches' performances through the OA analysis, All approach showed the highest value (88.96%), followed by metrical (87.90%), bi-temporal (82.59%), and mono-temporal (74.95%). Focusing on landslide detection, from which the inventory is made, the bi-temporal approach presented the lowest omission error rate (3.03%), followed by metrical (3.08%), All (5.97%), and mono-temporal (7.69%). And concerning the commission error metrical showed the lowest values (5.97%), followed by mono-temporal (9.09%), All (10%), and bi-temporal (14.67%). Out of these results, it is possible to conclude that the main challenge in this methodology for landslide detection is the reduction of the commission error. It is remarkable that the bi-temporal approach, presented the lowest level for omission error, however the highest for commission. Furthermore, from the interpretation of all of the accuracy results, with the understanding that the goal was to provide both LULC and landslide inventory products, the metrical approach presented the most beneficial result, once it showed the second best result for OA and omission error; and the best result for commission error. It should be noted that other metrics can be used to perform the accuracy assessment, for instance, standard error should be included to weight confusion matrix cells by the classified surface for each class, as recommended by Stehman and Foody.[72]

To compare the proposed method with other landslide detection approaches, a recent study also based on Sentinel-2 multi-temporal imagery and the same accuracy assessment metrics was analyzed. Yang et al.[73] used a dataset from 2015 to 2018 applied to *k*-means and achieved user's accuracy of 90.8% (commission error of 9.2%) and producer's accuracy of 83.6% (omission error of 16.4%) for landslide detection. In our research the best results were 94.1% for user's accuracy (commission error of 5.9%) and 97% for producer's accuracy (omission error of 3%). Different from our results,[73] presented more omission error rates than commission. The lack of studies using time series, not only bi-temporal datasets, to detect landslide scars reveals the scientific gap in this subject. However, deep-learning methods using bi-temporal approaches have achieved outstanding results, as shown by Lei et al.[74] with OA of 91.9% against our best result of 88.9%.

Regardless of all the contribution and results obtained by this methodology, field work, and visual interpretation from specialists are indeed very important and should not be completely replaced by the semi-automatic approaches. These techniques can provide accurate information that might not be achievable only by semi-automatic methods. However, the main suggestion could be to combine these techniques. For future studies, it is recommended to apply this methodology with different predictors and replicate it to other study areas, with different characteristics, such as regions with more urban areas. It is highly recommended to use a time series with regular time spacing so that interpolation and smoothing methods can be applied to the time series to avoid irregularities due to clouds, for example. Furthermore, landslide scars can present very similar response as other disturbances, as fire scars, for example. To deal with this type of issue, it is recommended to do a very detailed previous study of the area, as well as designing the classification model to account for identification and characterization of disturbances producing behavior, and trying to figure out which variables represent the best discriminators.

Even though the RF algorithm realizes a selection of variables based on their IR, it is noted that the SAVI index was not selected among the top 10 for neither approaches. This might be due to the high-cross correlation between SAVI and NDVI. So, for future studies, it is recommended to perform a cross-correlation analysis to select only the most significant predictors. Moreover, for a more complete analysis of the classification results, calculating uncertainty maps are highly recommended with, among others, Shannon entropy, RF maximum probability estimates, and

confusion index. In addition, the proposed methodology should be tested with other types of orbital data besides Sentinel-2, e.g., CBERS-4 and 4A, Sentinel-1 (SAR), and Landsat-8.

5 Conclusions

The main aim of this research was to develop a methodology that generates the LULC and landslide inventory maps through a semi-automatic process, using a time-series assessment. Even though free access to medium spatial resolution orbital imagery is constantly updated and available for use, landslides inventories are still largely realized by visual interpretation of single or bi-temporal datasets of high spatial resolution images. Not only is this process time-consuming and manually realized via polygon vectorization from a visual interpretation method but usually requires updated images from (very-) high spatial resolution, which, in general, needs to be purchased by the final user, not available for free, which could present high costs. The idea that the "best" inventory map can only be achieved by human visual interpretation has been questioned with the emergence of new technologies and the advance into the artificial intelligence era, reducing the distance among conventional and state-of-art methods. Nowadays, the world shows that data availability is no longer a limitation because many types of data are massively provided from countless sources every day. The main challenge is how to manage and transform data into information. In this sense, the 21st century has been evolving toward the machine learning, big data, and cloud processing techniques pursuing the final goal, which is to build knowledge.

Taking all above into account, this work contributes to the use of state-of-art methods and data to generate landslide susceptibility assessment supply's materials. The methodology proposed uses geomorphometric attributes and dense and irregular time series of different spectral indices from remote sensing imagery and three modules of time-series metrics (basic, polar, and fractal)e applied to RF classification algorithm, all developed in a free and open source way. Analyzing the predictors' IR, among the four input approaches compared (mono-temporal, bi-temporal, metrical, and All), each of them ranked an NDBI image or metric extracted from this index as the most important predictor of the classification. The NDVI or predictors related to that index also presented high relevance. The geomorphometric attributes, especially the slope, was present in all of the approaches among the top 6 in the importance list.

This research answered some questions on how time series and time series metrics could contribute to LULC mapping and landslide detection. Moreover, the IR may support the decision of what predictors should be used for classification on that scientific field. The effort of this work was to incorporate the abundant availability of free orbital optical data of high spatial resolution in a semi-automatic procedure to present an alternative to the conventional mono/bi-temporal visual interpretation method. The idea was to explore the potential of this type of data and classification algorithm looking forward to assist in the advance of automatized methods that make use of free and open source material, in favor of accessible, faster, cheaper, and accurate products for landslide susceptibility assessment.

Acknowledgments

References

1. D. M. Cruden, "Cruden, DM, Varnes, DJ, 1996, landslide types and processes, transportation research board, us national academy of sciences, special report, 247: 36–75," (1993).

2. C. P. Cardozo, "Spatially integrated modelling approach to landslide risk assessment: a case study of the Nova Friburgo disaster—RJ, Brazil," (2018).

3. R. P. Quevedo et al., "Modelagem de áreas suscetveis a movimentos de massa: avaliação raziliane de técnicas de amostragem, aprendizado de máquina e modelos digitais de elevação," *Geosciences* **38**(3), 781–795 (2019).

4. S. Gameiro et al., "Análise e correlação de atributos morfométricos e sua influência nos movimentos de massa ocorridos na Bacia do Rio Rolante, RS," in *Anais SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, XIX*, INPE, São José dos Campos (2019).

5. T. Glade, "Landslide occurrence as a response to land use change: a review of evidence from New Zealand," *Catena* **51**(3–4), 297–314 (2003).

6. F. F. Ávila et al., "The influence of land use/land cover variability and rainfall intensity in triggering landslides: a back-analysis study via physically based models," *Nat. Hazards* **105**(1), 1139–1161 (2021).

7. V. Canavesi and R. C. d. S. Alvalá, "Changes in vegetation cover in reforested areas in the state of São Paulo, Brazil and the implication for landslide processes," *ISPRS Int. J. Geo-Inf.* **1**(2), 209–227 (2012).

8. S. L. Gariano, O. Petrucci, and F. Guzzetti, "The role of rainfall and land use/cover changes in landslide occurrence in Calabria, Southern Italy, in the 20th Century," in *Proc. Workshop on World Landslide Forum*, Springer, pp. 339–345 (2017).

9. S. Beguera, "Changes in land cover and shallow landslide activity: a case study in the Spanish Pyrenees," *Geomorphology* **74**(1–4), 196–206 (2006).

10. C. Promper et al., "Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios," *Appl. Geogr.* **53**, 11–19 (2014).

11. T. Mrozek et al., "Landslide susceptibility and risk assessment in a non-mountainous region: a case study of Koronowo, Northern Poland," *Geol. Q.* **60**, 758–769 (2016).

12. R. N. Ramos-Bernal et al., "Evaluation of unsupervised change detection methods applied to landslide inventory mapping using aster imagery," *Remote Sens.* **10**(12), 1987 (2018).

13. F. Guzzetti et al., "Landslide inventory maps: new tools for an old problem," *Earth-Sci. Rev.* **112**(1–2), 42–66 (2012).

14. F. Guzzetti et al., "Distribution of landslides in the Upper Tiber River basin, central Italy," *Geomorphology* **96**(1–2), 105–122 (2008).

15. J. Blahut, C. J. Van Westen, and S. Sterlacchini, "Analysis of landslide inventories for accurate prediction of debris-flow source areas," *Geomorphology* **119**(1–2), 36–51 (2010).

16. H. Petschko et al., "Landslide inventories for reliable susceptibility maps in lower Austria," in *Landslide Science and Practiceid*, P. S. K. Margottini and C. Canuti, Eds., pp. 281–286, Springer (2013).

17. E. M. Schmaltz, S. Steger, and T. Glade, "The influence of forest cover on landslide occurrence explored with spatio-temporal information," *Geomorphology* **290**, 250–264 (2017).

18. V. Lupiano et al., "Landslide inventory and main geomorphological features affecting slope stability in the Picentino river basin (Campania, southern Italy)," *J. Maps* **15**, 131–141 (2019).

19. Y. Qin, P. Lu, and Z. Li, "Landslide inventory mapping from bitemporal 10 m Sentinel-2 images using change detection based Markov random field," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **42**(3), 1447–1452 (2018).

20. M. Rußwurm and M. Körner, "Multi-temporal land cover classification with sequential recurrent encoders," *ISPRS Int. J. Geo-Inf.* **7**(4), 129 (2018).

21. C. Paris, L. Bruzzone, and D. Fernández-Prieto, "A novel approach to the unsupervised update of land-cover maps by classification of time series of multispectral images," *IEEE Trans. Geosci. And Remote Sens.* **57**(7), 4259–4277 (2019).

22. C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification: a review," *ISPRS J. Photogramm. Remote Sens.* **116**, 55–72 (2016).

23. A. C. Mondini et al., "Sentinel-1 SAR amplitude imagery for rapid landslide detection," *Remote Sens.* **11**(7), 760 (2019).

24. C. Zhao et al., "Landslide identification and monitoring along the Jinsha River catchment (Wudongde reservoir area), China, using the InSAR method," *Remote Sens.* **10**(7), 993 (2018).

25. K. Pawłuszek et al., "Multi-aspect analysis of object-oriented landslide detection based on an extended set of lidar-derived terrain features," *ISPRS Int. J. Geo-Inf.* **8**(8), 321 (2019).

26. P. V. Gorsevski et al., "Landslide detection and susceptibility mapping using lidar and an artificial neural network approach: a case study in the Cuyahoga valley National Park, Ohio," *Landslides* **13**(3), 467–484 (2016).

27. O. Ghorbanzadeh et al., "Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection," *Remote Sens.* **11**(2), 196 (2019).

28. Y. Yi and W. Zhang, "A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal rapideye satellite imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **13**, 6166–6176 (2020).

29. V. S. da Silva et al., "Methodological evaluation of vegetation indexes in land use and land cover (lulc) classification," *Geol. Ecol. Landscapes* **4**(2), 159–169 (2020).

30. K. K. Singh and A. Singh, "Detection of 2011 Sikkim earthquake-induced landslides using neuro-fuzzy classifier and digital elevation model," *Nat. Hazards* **83**(2), 1027–1044 (2016).

31. Y. Li et al., "Semiautomatic landslide detection using remote sensing and slope units," *Transport. Res. Rec.* **2604**(1), 104–110 (2017).

32. B. Pradhan and M. R. Mezaal, "Optimized rule sets for automatic landslide characteristic detection in a highly vegetated forests," in *Laser Scanning Applications in Landslide Assessment*, B. Pradham, ed., pp. 51–68, Springer (2017).

33. T. S. Körting, L. M. G. Fonseca, and G. Cmara, "GeoDMA—geographic data mining analyst," *Comput. Geosci.* **57**, 133–145 (2013).

34. S. Dong, H. Li, and D. Sun, "Fractal feature analysis and information extraction of woodlands based on MODIS NDVI time series," *Sustainability* **9**, 1215 (2017).

35. L. Eklundh and P. Jönsson, *Timesat 3.2 with Parallel Processing-Software Manual*, Vol. 52, Lund University (2012).

36. L. A. Manfré et al., "Method for landslides identification at the São Paulo state coast, Brazil," *Geociências* **33**(1), 172–180 (2014).

37. S. Siyahghalati et al., "Rule-based semi-automated approach for the detection of landslides induced by 18 september 2011 Sikkim, Himalaya, earthquake using IRS LISS3 satellite images," *Geomat. Nat. Hazards Risk* **7**(1), 326–344 (2016).

38. M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.* **61**(3), 399–409 (1997).

39. F. Chen, B. Yu, and B. Li, "A practical trial of landslide detection from single-temporal Landsat8 images using contour-based proposals and random forest: a case study of national Nepal," *Landslides* **15**(3), 453–464 (2018).

40. M. R. Mezaal and B. Pradhan, "Data mining-aided automatic landslide detection using airborne laser scanning data in densely forested tropical areas," *Korean J. Remote Sens.* **34**(1), 45–74 (2018).

41. D. Hölbling et al., "Mapping and analyzing the evolution of the butangbunasi landslide using landsat time series with respect to heavy rainfall events during typhoons," *Appl. Sci.* **10**(2), 630 (2020).

42. Y. Kang et al., "Insar monitoring of creeping landslides in mountainous regions: a case study in Eldorado National Forest, California," *Remote Sens. Environ.* **258**, 112400 (2021).

43. O. G. Petry, "Cartografia das áreas de risco no rio rolante-rs e estratégias de recomposição da mata ciliar," 132 (2003).

44. T. D. Luerce, "Geoturismo na bacia hidrográfica do rio rolante/rs: um estudo acerca das quedas d'água," 181 (2012).

45. M. S. Rossato, "Os climas do rio grande do sul: variabilidade, tendências e tipologia," (2011).

46. "Diagnóstico preliminar: descritivo dos eventos ocorridos no dia 5 de janeiro de 2017 entre as regiões dos municípios de são razilian de paula e rolante/rs," (2017).

47. L. Ranghetti et al., "sen2r: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data," *Comput. Geosci.* **139**, 104473 (2020).

48. J. Gerente et al., "Classificação semiautomática de cicatrizes de movimentos de massa por meio de técnicas de detecção de mudanças," *Revista Brasileira de Geomorfologia* **18**(4), 801–812 (2017).

49. J. Gerente et al., "Mass movements' scars classification using data mining techniques," in *Anais Simpósio Brasileiro de Sensoriamento Remoto*, INPE, São José dos Campos, pp. 3553–3560 (2017).

50. J. Rouse et al., *Monitoring Vegetation Systems in the Great Plains with ERTS*, NASA (1974).

51. A. Bhandari, A. Kumar, and G. Singh, "Feature extraction using normalized difference vegetation index (NDVI): a case study of Jabalpur city," *Proc. Technol.* **6**, 612–621 (2012).

52. D. Jeevalakshmi, S. N. Reddy, and B. Manikiam, "Land cover classification based on NDVI using Landsat8 time series: a case study Tirupati region," in *Proc. Int. Conf. Commun. And Signal Process.*, IEEE, pp. 1332–1335 (2016).

53. F. Kong et al., "Land cover classification based on fused data from GF-1 and MODIS NDVI time series," *Remote Sens.* **8**(9), 741 (2016).

54. Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from tm imagery," *Int. J. Remote Sens.* **24**(3), 583–594 (2003).

55. A. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.* **25**, 295–309 (1988).

56. R. L. Lawrence and W. J. Ripple, "Comparisons among vegetation indices and bandwise regression in a highly disturbed, heterogeneous landscape: Mount St. Helens, Washington," *Remote Sens. Environ.* **64**(1), 91–102 (1998).

57. M. Mohd Salleh et al., "Spatial prediction models for landslide activity mapping using vegetation anomalies," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLII-4/W16**, 441–449 (2019).

58. R. Edsall, A. MacEachren, and D. Peuquet, "Assessing the effectiveness of temporal legends in environmental visualization," in *Proc. GIS/LIS*, pp. 28–30 (1997).

59. A. R. Soares et al., "Stmetrics: a python package for satellite image time-series feature extraction," in *Proc. Int. Geosci. And Remote Sens. Symp.*, IEEE, pp. 2061–2064 (2020).

60. T. R. Martha et al., "Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods," *Geomorphology* **116**(1–2), 24–36 (2010).

61. S. Tavakkoli Piralilou et al., "Landslide detection using multi-scale image segmentation and different machine learning models in the higher Himalayas," *Remote Sens.* **11**(21), 2575 (2019).

62. M. D. M. Valeriano and D. D. F. Rossetti, *Topodata: guia para utilização de dados geomorfológicos locais*, INPE, São José dos Campos (2008).

63. K. J. Beven and M. J. Kirkby, "A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du razi versant," *Hydrol. Sci. J.* **24**(1), 43–69 (1979).

64. R. Sørensen and J. Seibert, "Effects of DEM resolution on the calculation of topographical indices: TWI and its components," *J. Hydrol.* **347**(1–2), 79–89 (2007).

65. C.-Z. Qin et al., "An approach to computing topographic wetness index based on maximum downslope gradient," *Precis. Agric.* **12**(1), 32–43 (2011).

66. T. D. T. Uehara et al., "Classification algorithms comparison for landslide scars," in *Proc. Brazilian Symp. Geoinf.*, INPE, São José dos Campos, p. 58 (2019).

67. R. G. Congalton et al., "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.* **37**(1), 35–46 (1991).

68. M. Belgiu and L. Drăgut, "Random forest in remote sensing: a review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).

69. P. Olofsson et al., "Good practices for estimating area and assessing accuracy of land change," *Remote Sens. Environ.* **148**, 42–57 (2014).

70. J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.* **20**(1), 37–46 (1960).

71. J. Sim and C. C. Wright, "The kappa statistic in reliability studies: use, interpretation, and sample size requirements," *Phys. Ther.* **85**(3), 257–268 (2005).

72. S. V. Stehman and G. M. Foody, "Key issues in rigorous accuracy assessment of land cover products," *Remote Sens. Environ.* **231**, 111199 (2019).

73. W. Yang et al., "Using Sentinel-2 time series to detect slope movement before the jinsha river landslide," *Landslides* **16**(7), 1313–1324 (2019).
74. T. Lei et al., "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.* **16**(6), 982–986 (2019).
75. H. do Nascimento Bendini et al., "Detailed agricultural land classification in the razilian cerrado based on phenological information from dense satellite image time series," *Int. J. Appl. Earth Observ. Geoinf.* **82**, 101872 (2019).

Tatiana Dias Tardelli Uehara received her BS degree in geography in 2019 from the University of São Paulo (USP), and her MS degree in remote sensing in 2021 from Brazil's National Institute for Space Research (INPE). She was granted a scholarship for an exchange program at Kyung Hee University, Seoul, South Korea. Her research areas include image multi-temporal analysis, time-series metrics, landslide detection, machine learning, and data mining. She works currently as a geospatial data scientist at EarthDaily Agro.

Thales Sehn Körting is a PhD in remote sensing, with an MS degree in applied computing (both titles obtained at Brazil's National Institute for Space Research—INPE). He is also a computer engineer (at FURG). Nowadays, he is a researcher at INPE, and his research areas include remote sensing image segmentation, multi-temporal analysis, image classification, data mining algorithms, and artificial intelligence.

Anderson dos Reis Soares is a PhD in remote sensing Brazil's National Institute for Space Research – INPE, with a MS degree in geodetic science obtained at the Federal University of Pernambuco. Moreover, he is also a technologist in geoprocessing by the Federal Institute of Education Science and Technology of Paraiba. Nowadays, he is a specialist consultant at Cognizant working and researching remote sensing image segmentation, spatiotemporal analysis, and image classification.

Renata Pacheco Quevedo received her BS degree in geography in 2016 and an MSc degree in remote sensing in 2019, both from Federal University of Rio Grande do Sul, Porto Alegre, Brazil, and carried out academic exchanges at Nacional University of Cuyo (2014), Argentina, and Autonomous University of Madrid (2018), Spain. She is mainly devoted to research in the field of remote sensing, geological hazards, machine learning, and spatial analysis. Currently, she is a PhD candidate in remote sensing at National Institute for Space Research, São José dos Campos, Brazil. Her research interests include landslide susceptibility analysis, machine learning, and image processing techniques.