



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/06.23.21.49-TDI

## SEMI-AUTOMATIC METHODS FOR LAND USE AND LAND COVER MAPPING USING TIME SERIES WITH FOCUS ON LANDSLIDE SCARS DETECTION

Tatiana Dias Tardelli Uehara

Master's Dissertation of the  
Graduate Course in Remote  
Sensing, guided by Dr. Thales  
Sehn Körting, approved in June  
11, 2021.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34T/44TPHRS>>

INPE  
São José dos Campos  
2021

**PUBLISHED BY:**

Instituto Nacional de Pesquisas Espaciais - INPE  
Coordenação de Ensino, Pesquisa e Extensão (COEPE)  
Divisão de Biblioteca (DIBIB)  
CEP 12.227-010  
São José dos Campos - SP - Brasil  
Tel.:(012) 3208-6923/7348  
E-mail: pubtc@inpe.br

**BOARD OF PUBLISHING AND PRESERVATION OF INPE  
INTELLECTUAL PRODUCTION - CEPPII (PORTARIA N°  
176/2018/SEI-INPE):****Chairperson:**

Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra  
(CGCT)

**Members:**

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)  
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e  
Ciência Espaciais (CGCE)  
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e  
Pesquisas Aplicadas (CGIP)  
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

**DIGITAL LIBRARY:**

Dr. Gerald Jean Francis Banon  
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

**DOCUMENT REVIEW:**

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)  
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

**ELECTRONIC EDITING:**

Ivone Martins - Divisão de Biblioteca (DIBIB)  
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/06.23.21.49-TDI

**SEMI-AUTOMATIC METHODS FOR LAND USE AND  
LAND COVER MAPPING USING TIME SERIES WITH  
FOCUS ON LANDSLIDE SCARS DETECTION**

Tatiana Dias Tardelli Uehara

Master's Dissertation of the  
Graduate Course in Remote  
Sensing, guided by Dr. Thales  
Sehn Körting, approved in June  
11, 2021.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34T/44TPHRS>>

INPE  
São José dos Campos  
2021

Cataloging in Publication Data

---

Uehara, Tatiana Dias Tardelli.

Ue3s      Semi-automatic methods for land use and land cover mapping  
using time series with focus on landslide scars detection / Tatiana  
Dias Tardelli Uehara. – São José dos Campos : INPE, 2021.  
xxii + 92 p. ; (sid.inpe.br/mtc-m21d/2021/06.23.21.49-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional  
de Pesquisas Espaciais, São José dos Campos, 2021.

Guiding : Dr. Thales Sehn Körting.

1. Mass movements. 2. Image time series. 3. Time series  
metrics. 4. Random Forest. 5. Machine Learning. I.Title.

CDU 528.8:519.246.8

---



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

MINISTÉRIO DA  
CIÊNCIA, TECNOLOGIA  
E INOVAÇÕES**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**  
Serviço de Pós-Graduação - SEPGR**DEFESA FINAL DE DISSERTAÇÃO DE TATIANA DIAS TARDELLI UEHARA**  
**BANCA Nº 174/2021, REG 795870/2019**

No dia 11 de junho de 2021, as 09 horas, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestra em Sensoriamento Remoto. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

**Título: "SEMI-AUTOMATIC METHODS FOR LAND USE AND LAND COVER MAPPING USING TIME SERIES WITH FOCUS ON LANDSLIDE SCARS DETECTION"**

Eu, Camilo Daleles Rennó, Presidente da Banca Examinadora, assino esta ATA, em nome de todos os membros, com o consentimento dos mesmos.

Dr. Camilo Daleles Rennó - Presidente - INPE

Dr. Thales Sehn Körting - Orientador - INPE

Dr. Laércio Massaru Namikawa - Membro Interno - INPE

Dr. Fernando Shinji Kawakubo - Membro Externo - USP

Dr. Anderson Reis Soares - Membro Externo - Cognizant Technology Solutions



Documento assinado eletronicamente por **Camilo Daleles Rennó, Tecnologista**, em 15/06/2021, às 13:13 (horário oficial de Brasília), com fundamento no art. 6º do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site <http://sei.mctic.gov.br/verifica.html>, informando o código verificador **7592169** e o código CRC **FDC77ADB**.



*“It often requires more courage to dare to do right than to fear to do wrong”.*

ABRAHAM LINCOLN



*To all people that believe and defend the **public**  
**education** of quality, and all who lost their loved ones  
due to the **Covid-19** pandemic.*



## ACKNOWLEDGEMENTS

I am always grateful to my family, Rosana, Sílvia, Víctor, Beatriz, Dulce, and Billy for supporting my choices, cheering and believing in my potential. They are the solid base from where I can develop.

I express my enormous gratitude to my advisor Dr. Thales Sehn Körting, who has always been present during this research, contributing with new ideas and demonstrating support on emotional, technical and intellectual fields. As an advisor, his outstanding characteristics were the patience to teach what I didn't know, the compliments as a recognition of what I could do well, the constructive critics to help in my personal development, the constant motivation to keep going and believe in my work, and the support as a friend when I struggled. All these characteristics were fundamental to conclude this master dissertation. I thank Dr. Anderson Soares, Dr. Camilo Rennó, Dr. Fernando Kawakubo, and Dr. Laércio Namikawa, members of the Dissertation Defense Committee, for all the constructive suggestions that helped improving this research. I would like to thank Dr. Fernando Nadal, member of the Geography Department of USP, for awakening my interest in the Geomorphology field, specially the landslides.

I am extremely thankful to all friends of SER-2019! Each of them showed I different life perspective, and together we could build a big and colourful mosaic of love, comfort, parties, complicity and care. By seeing some of you passionate about science, I started loving it too. I have learned so much from you guys. We shared so many nice experiences and hope to keep you in my life. A special thank to Fernanda Theodoro, who inspires me by her determination, and has always encouraged me to be my best version, celebrating my achievements and giving comfort during the hard moments. I also thank all colleagues of our laboratory, with whom I have shared the same study room, tea/coffee breaks, and great ideas.

I would like to thank the Brazilian National Institute for Space Research (INPE), and specially the Earth Observation and Geoinformatics Division, for all the infrastructure and intellectual support. I thank all professors from the Remote Sensing Department, which have effectively contributed to my learning process. This research was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for which I am grateful.



## ABSTRACT

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” at Nova Friburgo in 2011. To prevent them, Land Use and Land Cover (LULC) and landslide inventory maps are essential to identify areas of high susceptibility. Most inventories are made using visual interpretation methods. However this approach is time and resource consuming. In this context, the aim of this study is to use data mining techniques with remote sensing time series imagery to produce a LULC classification and detect landslide scars via semi-automatic methods. The study area covers the whole extension of the Rolante River hydrographic basin, located at Rio Grande do Sul state, Brazil. In January 5<sup>th</sup>, 2017, more than 300 landslides occurred due to an extreme precipitation event. Sentinel-2 from 2015 to 2020 was used, which resulted in 122 dates, with a temporal resolution of approximately 13 days. A variety of attributes were generated, being them spectral indices, as the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Built-up Index (NDBI), and the Soil Adjusted Vegetation Index (SAVI). Basic, polar and fractal metrics were extracted from the time series. From the Digital Elevation Model (DEM) provided by SRTM, six geomorphometric features were extracted (Slope, Aspect, Plan curvature, Profile curvature, General Curvature and the Topographic Wetness Index). After that, classification was performed by the Random Forest (RF) algorithm. Four different input approaches were analysed: Mono-temporal, Bi-temporal, Metrical and All. Each approach consists of a different dataset input, in which the first approach takes into consideration only 10 attributes and the last one, 436 attributes. Considering the attributes importance ranking, the NDBI index or metric derived from that index presented the highest position for every approach, over NDVI and SAVI. Among the geomorphometric attributes, Slope was ranked among the 6 first attributes for all of them. Comparing the approaches performances through the overall accuracy analysis, All approach showed the highest value (88.96%), followed by Metrical (87.90%), Bi-temporal (82.59%), and Mono-temporal (74.95%). The approaches that presented the highest error rates for landslide class were Mono-temporal (7.69%) for omission and Bi-temporal (14.67%) for commission error. 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, presenting high overall accuracy values and low levels of commission and omission errors.

Keywords: Mass movements. Image time series. Time series metrics. Random Forest. Machine Learning.



# MÉTODOS DE MAPEAMENTO SEMI-AUTOMÁTICO DO USO E COBERTURA DO SOLO POR SÉRIES TEMPORAIS COM FOCO NA DETECÇÃO DE CICATRIZES DE DESLIZAMENTOS

## RESUMO

Os deslizamentos de terra acontecem com frequência no Brasil e têm sido a causa de muitas perdas socioeconômicas e fatalidades, a exemplo das consequências do chamado “Mega-desastre” em Nova Friburgo em 2011. Para evitá-los, os mapas de Uso e Cobertura do Solo (LULC) e os inventários de deslizamentos são essenciais para identificar áreas com alta suscetibilidade. A maioria dos inventários é feita usando métodos de interpretação visual, no entanto, essa abordagem consome demasiado tempo e recursos. Neste contexto, o objetivo deste estudo é utilizar técnicas de mineração de dados com imagens de séries temporais de sensoriamento remoto para produzir uma classificação LULC e detectar cicatrizes de deslizamentos por meio de métodos semiautomáticos. A área de estudo cobre toda a extensão da bacia hidrográfica do Rio Rolante, localizada no estado do Rio Grande do Sul, Brasil. Em 5 de janeiro de 2017, mais de 300 deslizamentos de terra ocorreram devido a um evento de precipitação extrema. Foram utilizadas imagens do Sentinel-2 de 2015 a 2020, que resultaram em 122 datas, com resolução temporal de aproximadamente 13 dias. Diversos atributos foram gerados, sendo eles índices espectrais, como o Índice de Vegetação por Diferença Normalizada (NDVI), o Índice de Construção da Diferença Normalizada (NDBI) e o Índice de Vegetação Ajustado do Solo (SAVI). Métricas básicas, polares e fractais foram extraídas das séries temporais. Do Modelo Digital de Elevação (DEM) fornecido pela SRTM, seis atributos geomorfométricos foram extraídos (declividade, aspecto, curvatura do horizontal, curvatura vertical, curvatura geral e índice de umidade topográfico). Em seguida, a classificação foi realizada pelo algoritmo Random Forest (RF). Quatro abordagens diferentes de dados de entrada foram analisadas: *Mono-temporal*, *Bi-temporal*, *Metrical* e *All*. Cada abordagem consiste em uma entrada de conjunto de dados diferente, em que a primeira abordagem leva em consideração apenas 10 atributos e a última, 436 atributos. Considerando o ranking de importância dos atributos, o índice NDBI ou métrica derivada desse índice apresentou o primeiro lugar para todas as abordagens, acima do NDVI e do SAVI. Dentre os atributos geomorfométricos, a declividade foi classificada entre os 6 primeiros atributos para todos os casos. Comparando o desempenho das abordagens por meio da análise da acurácia global, a abordagem *All* apresentou o maior valor (88,96%), seguidas por *Metrical* (87,90%), *Bi-temporal* (82,59%) e *Mono-temporal* (74,95%). As abordagens que apresentaram as maiores taxas de erro para a classe de deslizamento foram *Mono-temporal* (7,69%) para omissão e *Bi-temporal* (14,67%) para erro de comissão. Além disso, a partir da interpretação de todos os resultados de acurácia, e compreendendo-se que o objetivo era fornecer tanto o mapa de LULC quanto o produto de inventário de deslizamento, a abordagem *Metrical* apresentou o resultado mais benéfico, mostrando altos valores de acurácia global e níveis baixos de erros de comissão e omissão.

Palavras-chave: Movimentos de Massa. Séries Temporais de Imagens. Métricas de Séries Temporais. Random Forest. Machine Learning.



## LIST OF FIGURES

	<u>Page</u>
2.1	Types of mass movements. . . . . 8
2.2	Landslides at irregular occupation, Salvador (BA - Brazil). . . . . 9
2.3	“Mega disaster” landslides at Nova Friburgo (RJ - Brazil). . . . . 10
2.4	Representation of a time series. . . . . 11
2.5	Examples of basic time series metrics. . . . . 12
2.6	Linear and polar plots. . . . . 15
2.7	Steps that compose the KDD process. . . . . 16
2.8	Hughes phenomenon. . . . . 17
2.9	Random Forest structure. . . . . 19
2.10	Sentinel-2 satellite. . . . . 20
2.11	Sentinel-2 spectral channels. . . . . 21
3.1	Study area, the Rolante River Catchment. . . . . 25
3.2	Geographical characterization of the study area. . . . . 26
4.1	Methodology flowchart. . . . . 29
4.2	Feature extraction. . . . . 32
4.3	Geomorphometric attributes at Rolante River’s Basin. . . . . 34
4.4	Classification approaches. . . . . 37
4.5	Dates distribution. . . . . 37
4.6	Bi-temporal approach. . . . . 38
4.7	Cloud coverage issue. . . . . 39
4.8	Monte Carlo Simulations. . . . . 40
5.1	NDVI metrics. . . . . 46
5.2	NDBI metrics. . . . . 47
5.3	SAVI metrics. . . . . 48
5.4	NDBI metrics in details for Landslides scars. . . . . 49
5.5	NDBI metrics in details for Agriculture class. . . . . 50
5.6	NDBI metrics in details for Silviculture class. . . . . 51
5.7	Landslides linear and polar graphs. . . . . 53
5.8	Forest linear and polar graphs. . . . . 53
5.9	Silviculture linear and polar graphs. . . . . 54
5.10	Agriculture linear and polar graphs. . . . . 54
5.11	Landslides boxplot analysis. . . . . 55
5.12	Forest and Silviculture boxplot analysis. . . . . 56

5.13	Agriculture boxplot analysis. . . . .	57
5.14	Urban boxplot analysis. . . . .	58
5.15	Water boxplot analysis. . . . .	58
5.16	Classification maps. . . . .	60
5.17	Confusion matrices. . . . .	61
5.18	Landslides in details. . . . .	64
5.19	Landslide difference map. . . . .	65
5.20	Landslide vs Silviculture. . . . .	66
5.21	Classification maps in details. . . . .	68
A.1	LULC map from Mono-temporal approach. . . . .	88
A.2	LULC map from Bi-temporal approach. . . . .	89
A.3	LULC map from Metrical approach. . . . .	90
A.4	LULC map from All approach. . . . .	91
A.5	Landslide Inventory Maps. . . . .	92

## LIST OF TABLES

	<u>Page</u>
4.1 Example of confusion matrix. . . . .	41
5.1 Attributes' Importance Ranking (IR). . . . .	43
5.2 Kappa index and Overall Accuracy. . . . .	62
5.3 Omission errors. . . . .	62
5.4 Commission errors. . . . .	63



## LIST OF ABBREVIATIONS

ANN	–	Artificial Neural Network
ASF	–	Alaska Satellite Facility
BOA	–	Bottom-of-Atmosphere
CFS	–	Correlation-based Feature Selection
CFS	–	Convolutional Neural Networks
DEM	–	Digital Elevation Model
DFA	–	Detrended Fluctuation Analysis
DM	–	Data Mining
DT	–	Decision Tree
DTM	–	Digital Terrain Model
FOV	–	Field-of-view
HE	–	Hurst Exponent
InSAR	–	Interferometric SAR
KDD	–	Knowledge Discovery in Databases
LiDAR	–	Light Detection And Ranging
LULC	–	Land Use and Land Cover
OSAVI	–	Optimized Soil Adjusted Vegetation Index
ML	–	Maximum Likelihood
MODIS	–	Moderate-Resolution Imaging Spectroradiometer
MSI	–	MultiSpectral Instrument
NDBI	–	Normalized Difference Built-up Index
NDVI	–	Normalized Difference Vegetation Index
RF	–	Random Forest
ROI	–	Region Of Interest
SAR	–	Synthetic Aperture Radar
SAVI	–	Soil Adjusted Vegetation Index
SITS	–	Satellite Image Time Series
SRTM	–	Shuttle Radar Topography Mission
STD	–	Standard Deviation
SVM	–	Support Vector Machine
SWIR	–	Short Wave Infra-Red
TOA	–	Top-of-Atmosphere
UTM	–	Universal Transverse Mercator
VNIR	–	Visible and Near-Infra-Red



# CONTENTS

	<u>Page</u>
<b>1 INTRODUCTION</b> . . . . .	<b>1</b>
<b>2 THEORETICAL REFERENCE</b> . . . . .	<b>7</b>
2.1 Landslides . . . . .	7
2.2 Time series . . . . .	10
2.2.1 Time series metrics . . . . .	13
2.3 Data mining . . . . .	15
2.3.1 The Random Forest classification algorithm . . . . .	18
2.4 Sampling methods . . . . .	19
2.5 Sentinel-2/MSI . . . . .	19
2.6 SRTM digital elevation model . . . . .	22
<b>3 STUDY AREA</b> . . . . .	<b>25</b>
<b>4 METHODOLOGY</b> . . . . .	<b>29</b>
4.1 Data and pre-processing . . . . .	30
4.2 Feature extraction . . . . .	30
4.3 Sampling . . . . .	35
4.4 Classification . . . . .	35
4.5 Validation and analysis . . . . .	40
<b>5 RESULTS AND DISCUSSION</b> . . . . .	<b>43</b>
5.1 Importance ranking . . . . .	43
5.2 Time Series metrics . . . . .	44
5.3 Land Use and Land Cover maps . . . . .	59
<b>6 CONCLUSION</b> . . . . .	<b>69</b>
<b>REFERENCES</b> . . . . .	<b>73</b>
<b>APPENDIX A - MAPS</b> . . . . .	<b>87</b>



## 1 INTRODUCTION

Landslides are a natural gravity driven phenomenon, which consists in the movement of a mass of soil or rock from the top of the hill towards the bottom. The occurrence of landslides is registered all over the world, frequently in mountainous areas where there have been extreme precipitation events, earthquakes or snow melting. 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 (CARDOZO, 2018).

On January 5<sup>th</sup>, 2017, an extreme precipitation event triggered a massive occurrence of landslides in the Rolante River catchment in Rio Grande do Sul state, Brazil. Previous works have detected more than 300 landslide scars deriving from this episode (QUEVEDO et al., 2019; GAMEIRO et al., 2019).

In order to reduce the landslide risk, inventory and Land Use and Land Cover (LULC) maps are relevant information. The landslide inventory map consists of identifying mass movement scars, which can provide much information about past events, as location, types and patterns, assisting to build landslide susceptibility models (RAMOS-BERNAL et al., 2018). 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 (GUZZETTI et al., 2008; BLAHUT et al., 2010; PETSCHKO et al., 2013; SCHMALTZ et al., 2017). Thus, the inventory is crucial to support urban planning and disaster risk reduction (LUPIANO et al., 2019).

The completeness of a landslide inventory is frequently related to the land cover distribution of an area (SCHMALTZ et al., 2017), once characteristics as vegetation removal and construction building on high slope areas can increase the susceptibility. The LULC maps assist the detection of areas that have experienced anthropogenic interventions that may induce landslides. According to Glade (2003), the land use change is recognized as one of the most important factors influencing the occurrence of rainfall-induced landslides. LULC may act as predisposing factors of landslide occurrence, besides, may also conduct the spatial distribution of landslide consequences (GARIANO et al., 2017; GLADE, 2003; BEGUERÍA, 2006; PROMPER et al., 2014). The LULC maps are a fundamental aspect to the vulnerability analysis of an area. Not only triggering changes can be identified (i.e, deforestation on hillslope), but also can assist to the social vulnerability measurement. According to Birkmann (2006),

a vulnerability indicator is defined as a variable which is an operational representation of a system characteristic or quality able to provide information regarding the susceptibility, coping capacity and resilience of a system to an impact resulting from a natural hazard. Considering that, the social vulnerability can be characterized by the same attributes on a scale more closely focused on the social, including attributes related to issues of livelihood, housing, income, education, security and gender among many others (TAPSELL et al., 2010).

The inventory maps can be done by either conventional methods (field mapping and manual vectorization by visual interpretation) or state-of-art techniques. Conventional methods are time and resource (price of man-hour) consuming (QIN et al., 2018); however, are still very commonly used, once they usually provide more accurate results. Semi-automatic methods, on the other hand, can provide a rapid mapping via change detection and pattern recognition algorithms (GUZZETTI et al., 2012), specially through remote sensing images.

Earth observation sensors can provide data in several observation frequencies, such as daily or weekly. Most LULC classification approaches, however, are designed for cloud-free and mono-temporal observations (RUSSWURM; KÖRNER, 2018). Nowadays, many researches have identified the benefits of incorporating a multi-temporal approach, recognizing its many potentials, as to deal with change detection and the monitoring of targets which present seasonal behavior (i.e, croplands). Hence, time series have been progressively used to LULC mapping and to identify the nature of land cover changes (PARIS et al., 2019). Remote sensing medium spatial resolution optical time series data have demonstrated 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 (GÓMEZ et al., 2016).

Out of an earth observation time series, a diversity of properties can be calculated, which are generally named as attributes. From remote sensing images, many spectral information can be extracted. The spectral indices are created in order to emphasize an specific target, which can be generated by making operations with the most relevant spectral bands, finally resulting in one single product. A variety of attributes have been successfully used on classification related to landslide. Firstly, regarding spectral indices, the Normalized Difference Vegetation Index (NDVI), and the Soil Adjusted Vegetation Index (SAVI), proved to be effective tools for landslide de-

tection and LULC mapping (SALLEH et al., 2018; ISHAK, 2018; YANG et al., 2012). Also, the Normalized Difference Build-up Index (NDBI) has been applied to many LULC mapping researches, besides some other recent papers on landslide susceptibility analysis that considered this index as a potential variable (PHAM et al., 2020; CHANG et al., 2020). 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. (2017) and Pradhan and Mezaal (2017). Metric attributes can also be extracted from time series and incorporated in the classification inputs, among them there are the basic and polar metrics (KÖRTING et al., 2013), besides the fractal, and phenological metrics (EKLUNDH; JÖNSSON, 2012).

Studies related to the landslide theme are, in general, associated to landslide susceptibility modelling. Research on that field is ancient, however is constantly being updated as new technologies and methodologies are developed. This happens because, besides all the effort on monitoring, in some cases this phenomenon can be still extremely unpredictable and devastating. This is specially applied to unplanned urban areas, usually in developing countries of tropical weather, where the irregular auto-construction on hill slopes is a common practice, and heavy rainfall is frequent.

To tackle the challenge of landslide detection through semi-automatic methods, machine learning and deep learning classifiers are commonly used, as the Support Vector Machine (SVM), Artificial Neural Network (ANN), Maximum Likelihood (ML), Random Forest (RF), and the Convolutional Neural Networks (CNN). 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 to the NDVI (MANFRÉ et al., 2014). Studies have compared ANN and SVM to mapping landslides and the results have shown no significant differences between both methods (MOOSAVI et al., 2014). The RF algorithm can be used both for classification and regression modelling. As a classifier, it has been widely used for the identification of landslide scars (GHORBANZADEH et al., 2019; CHEN et al., 2018). A research of data mining-aided automatic landslide detection compared the performance of SVM and RF, having the latter presented a higher classification accuracy (MEZAAL; PRADHAN, 2018). Moreover, studies show that Decision Tree (DT) algorithms have been used in landslide detection (SIYAHGHALATI et al., 2016) and LULC mapping (FRIEDL; BRODLEY, 1997), outperforming other classifiers.

Previous studies show that different types of data have been explored to detect landslides 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 (MONDINI et al., 2019; ZHAO et al., 2018). Moreover, besides the contribution of SAR data, some difficulties are still present as for landslide detection in dense forested areas. Considering that, the use of Light Detection And Ranging (LiDAR) technologies have been gaining more space, presenting significant results (PAWŁUSZEK et al., 2019; GORSEVSKI et al., 2016). Besides the aforementioned issues, the most commonly used type of data to detect landslide scars are 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 (GHORBANZADEH et al., 2019; YI; ZHANG, 2020).

It is important to elucidate that there are many approaches for landslide scars detection to produce inventory maps. However, in this study we assume the inventory map as a secondary product derived from the LULC map. 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. Most methods for producing landslide scars identification are based on change detection, producing binary results (scar Vs non scar). Nonetheless, based on the fact that both products (LULC and inventory map) are an essential material for landslide susceptibility modeling (the most effective method for preventing disasters), the methodology here applied unites both final products in one single workflow. One should notice that, besides the inventory is derived from the LULC map, it does not mean that the same combination of attributes will present the best performance for both purposes. In other words, accuracy is analysed separately, so a 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.

Although the contribution of time series for remote sensing imagery classification has already been confirmed by many researches, the question unsolved is: Until what point the addition of attributes is beneficial to the classification, taking into account both accuracy analysis and the procedure's complexity/processing time? Under this light, which classes present a relevant improvement in the accuracy with the increase of attributes? Which of them present high accuracy values with low

quantity of attributes?

Furthermore, in order to choose the most suitable spectral index, many factors may need to be considered. When dealing with time series, the use of vegetation indices, as the NDVI, is very common, once it helps to reduce the quantity of attributes in the time series by joining information from the red and near infrared bands in one single attribute. The NDVI is indeed a widely used index for general classifications, as the LULC maps; however, in the scenario of landslide susceptibility, 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.

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 Moderate-Resolution Imaging Spectroradiometer (MODIS) data. Time series metrics have shown great potential for classification specially for agricultural purposes, as exposed by [Bendini et al. \(2019\)](#) 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. However, research in this field is still recent, so more investigation is required over whether the time series 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 aim of this research is to generate the LULC classification and the landslide inventory map for the Rolante River catchment area through semi-automatic methods, based on remote sensing time series imagery, geomorphometric attributes and data mining techniques. The proposal consists of comparing different classification approaches and searching for the optimal combination of attributes for heterogeneous and landslide prone areas. Attributes are here interpreted as any input given to the classification algorithm, which means it can be a time series metric, one date of the NDVI image or the DEM, for example. Besides, metrics are here interpreted only as the time series metrics (basic, polar, fractal), which are a sub-group inside the wider category “attributes”. So, more specifically, the objectives are:

- Analyse the performance of long, dense and irregular time series for LULC classification of landslide prone areas.

- Evaluate which classes present a relevant improvement in accuracy with the increase of attributes, and which of them present high accuracy results even with low quantity of attributes.
- Investigate the most suitable index, among NDVI, SAVI and NDBI to be used in this type of areas.
- Explore the potential of time series metrics for classification and landslide detection.
- Inspect the most relevant set of attributes to be used in this case.
- Compare different classification approaches that range from the most simple and reduced dataset to more complex and dense ones.
- Generate the best landslide inventory and LULC map, according to this research methods, for the Rolante River catchment area.

## 2 THEORETICAL REFERENCE

This chapter introduces relevant concepts and methods that are considered in this research. It starts explaining the definition of landslides and contextualizing with the Brazilian scenario. Secondly, explanation is done about time-series and its metrics. Third, the concept of data mining is exposed, followed by some comments on the Random Forest classification algorithm. Finally, the description of the Sentinel-2/MSI and the DEM by SRTM.

### 2.1 Landslides

Landslides are widespread natural geomorphological processes and represent a gravity-driven component of erosion (DAVIES, 2015). They are downward movements of soil and rocks from the slope triggered by earthquakes, snow melting or heavy rain, which can also be caused or intensified by anthropic activities (GUZZETTI et al., 2012). There are a variety of definitions for mass movements, however they are generally based on the following criteria (AUGUSTO FILHO, 1995):

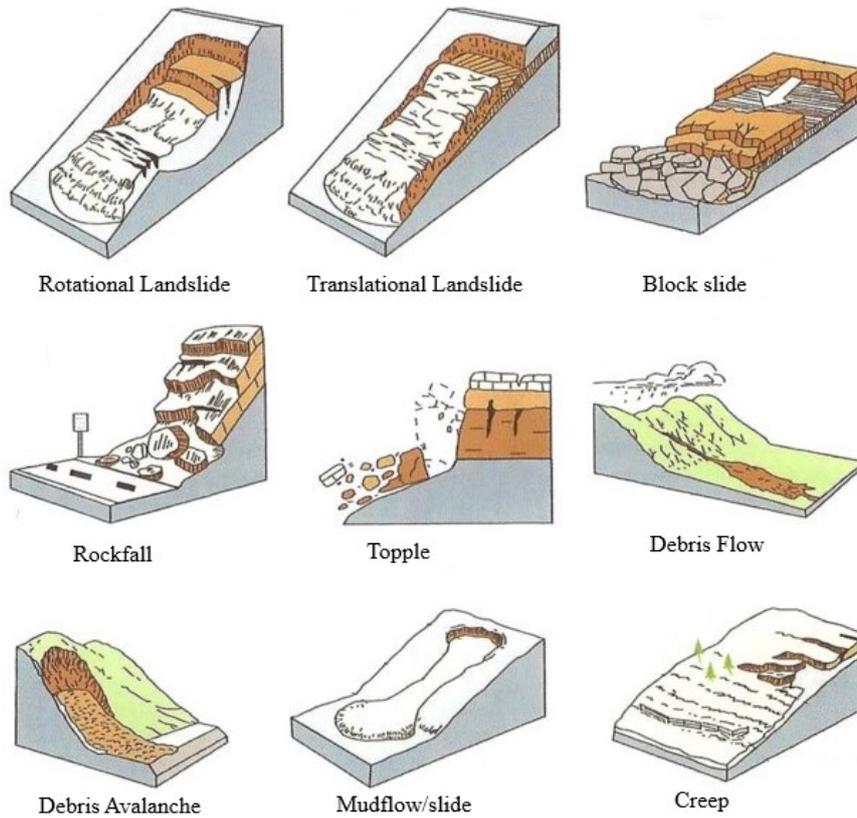
- a) **Motion kinematics:** speed, direction and sequence of displacements of matter in relation to the stable terrain;
- b) **Type of material:** earth, rocks, debris;
- c) **Geometry:** size and shape of displaced material.

Mass movements associated to slopes are assembled in many classes as shown in Figure 2.1. This study focuses specifically on landslides (or slides); however, one should know that all of the types typically occur simultaneously.

These phenomena can cause significant economic and human losses; hence, the importance of the risk management, mitigation and monitoring. One can define risk as a probability of undergoing damages by one or several hazards, and a disaster as a realized risk (NATHAN, 2005). In order to convert a risk into a disaster there must be specific conditions of vulnerability. Traditionally, vulnerability has been used only as a natural factor, independently from the context. In fact, it is both context-dependant and subject-dependant, once one is vulnerable to something, in a given place and at a given time (NATHAN, 2005). Tapsell et al. (2010) emphasizes the concept of social vulnerability when dealing with natural hazards. Landslide hazard is a function of susceptibility (spatial propensity to landslide activity) and

temporal frequency of landslide triggers, and its assessment may be done on different scales (UNISDR, 2017).

Figure 2.1 - Types of mass movements.



SOURCE: Adapted from Schwab et al. (2005).

Different from extreme natural events, like hurricanes or earthquakes, which have their causes mainly based on natural conditions, landslides present a deep association with LULC. Increased landslide hazard, for instance, has been related to the improper cut-and-fill construction of self-built housing on steep slopes, after the removal of vegetation (MENDES et al., 2018). According to Monteiro et al. (2015), considering it as a process, not as an event, could be helpful to comprehend how disasters are historically constructed on a spatial-temporal scale. In this research, besides considering it as a process, sometimes landslides will be referred as “event” in order to standardize and simplify the writing.

In Brazil, the landslides events were responsible for the second highest number of deaths between 1991 and 2012 (UNIVERSIDADE FEDERAL DE SANTA CATARINA - UFSC, 2013). A data survey from 1997 to 2017 showed that more than 202 deaths were registered only in the Metropolitan area of São Paulo city during this 20 years period (IPT, 2017). In the Brazilian case, the hazardous scenario with high number of human losses arises from extreme precipitation events associated with a weak infrastructure.

There is a serious issue related to the housing deficit due to the fast expansion of the urban population without an efficient urban planning. In 2012, São Paulo presented the largest housing deficit in Brazil; and while the state had about 4 million houses, the state capital (the city of São Paulo) had about 11 million inhabitants in an area of a little more than 1500 km<sup>2</sup> (LISTO; VIEIRA, 2012). The housing deficit led to the dense occupation of inappropriate areas, as high slopes, creating slums in risk areas (Figure 2.2). Naturally landslide prone due to the declivity, these areas become riskier when the vegetation is removed, and a dense amount of irregular constructions are built instead. Since they are usually illegally occupied areas, no assistance is given from the Government; hence, no appropriate sewage system is created, which leads to the soaking of the soil, increasing the probability of landslides. In these cases, one single event can generate many casualties.

Figure 2.2 - Landslides at irregular occupation, Salvador (BA - Brazil).



SOURCE: Dias (2015).

Many episodes have marked the Brazilian history with human and economic losses as a result of that. On January 11<sup>th</sup> and 12<sup>th</sup>, 2011, a so-called “Mega disaster” occurred in the mountainous region of Rio de Janeiro state (Fig. 2.3). From the 23 municipalities affected, seven were stated as public calamity situation (VASSOLER, 2013): Areal, Bom Jardim, Nova Friburgo, São José do Vale do Rio Preto, Sumidouro, Petrópolis and Teresópolis. This episode resulted in 947 registered deaths, 300 missing and millions of homeless people, besides the severe economic losses and infrastructure destruction (CEMADEN, 2016). Nova Friburgo, Teresópolis and Petrópolis municipalities recorded the greatest number of casualties (CARDOZO, 2018). In Nova Friburgo, 3.000 landslides were mapped, having the majority in within the urban area; while for the other municipalities, it occurred in the rural areas (BUSCH; AMORIM, 2011).

Figure 2.3 - “Mega disaster” landslides at Nova Friburgo (RJ - Brazil).



SOURCE: Vallin (2011).

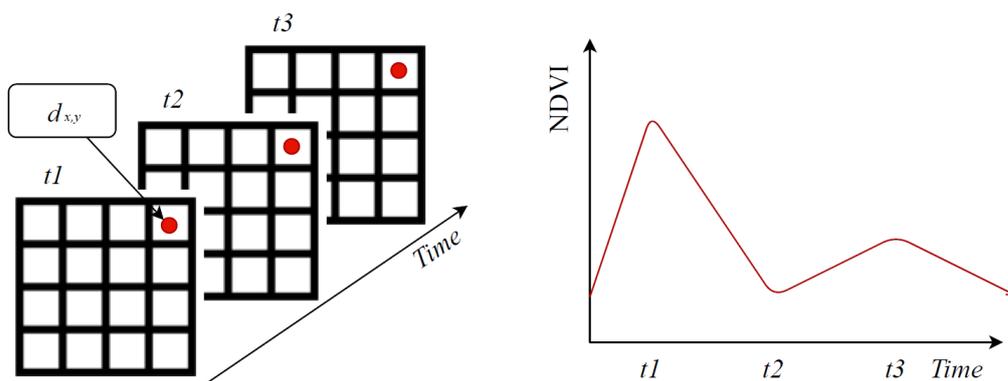
## 2.2 Time series

Orbital satellites generate images from the same part of the globe several times a year. The time series of remotely sensed imagery basically consists in the images

stacking, taking into consideration their ordered acquisition date.

Time series are composed of a set of observations  $d_{x,y}(t)$ , where each of them has a specific coordinate in space  $(x, y)$  and was captured in a specific moment in time  $(t)$ ; as illustrated in Figure 2.4.

Figure 2.4 - Representation of a time series.



$t_n$  corresponds to the date of each observation.  $d_{x,y}$  is the coordinate in space. The smoothed curve represents the NDVI values for each date.

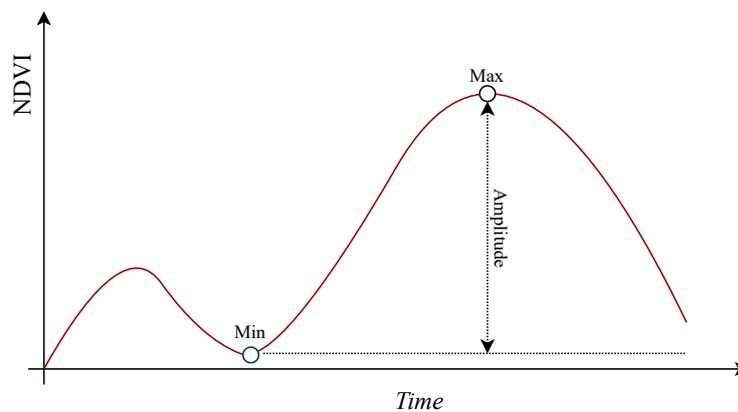
SOURCE: Adapted from Eklundh and Jönsson (2012).

There are several ways to extract information from time series. Udelhoven (2010) emphasized the problem of storing millions of temporal records in remote sensing databases; however, reducing the detection of hidden patterns at various levels of abstraction only by visual inspection. Considering that, many strategies have been developed involving image analysis techniques for data processing and information extraction. Time series can be used to make inferences, being necessary to set up a hypothetical probability model to represent the data (BROCKWELL; DAVIS, 2016). After that, it is possible to estimate parameters and check how well the model fits to the data, for example.

Time series have demonstrated a significant capacity for characterization of environmental phenomena, describing trends, seasonal behavior and abrupt change events. For that reason, some metrics can be extracted from the series, helping to enhance

the classification models. Some of those can be basic or polar metrics (KÖRTING et al., 2013), as well as fractal and seasonality metrics (EKLUNDH; JÖNSSON, 2012). The basic metrics consist of statistical measures, for instance, mean, standard deviation, maximum and minimum values of the curve. The polar ones are extracted by a cyclical representation of the time series, as eccentricity, angle and area (KÖRTING et al., 2013). More details about this process are explained further in Section 2.2.1. Some basic metrics are illustrated in Figure 2.5.

Figure 2.5 - Examples of basic time series metrics.



SOURCE: Authors's production.

The availability of free access to medium spatial resolution (Landsat-like) times series raised many researches, improving results when compared to mono-temporal approaches. Many authors have used time series for LULC mapping, as well as for change detection analysis (AYELE et al., 2018; ZHU, 2017). Time series have been used for the mapping of agricultural land use systems (BELLÓN et al., 2017), to analyse the impact of drought on native vegetation (OKIN et al., 2018), to monitor deforestation and trends of aeolian desertified lands (SCHULTZ et al., 2016; WANG et al., 2017).

Concerning landslides, many studies have been developed with SAR, for instance the ones presented by Hu et al. (2016), Dai et al. (2016), Jiang et al. (2016). Optical Satellite Image Time Series (SITS) have been used to monitor and detect creep (slow-moving landslides) (STUMPF et al., 2017; LACROIX et al., 2019). To detect regular landslides, bi-temporal optical imagery have been widely used (LV et al., 2018; QIN et al., 2018; GERENTE et al., 2017a; LI et al., 2016). In these cases, change detection

methods are applied to one image before and one after the event, identifying the scars. Besides bi-temporal, few studies reveal the usage of more observations to increment the temporal component, as presented by [Huang et al. \(2019\)](#), where images from five dates were used.

### 2.2.1 Time series metrics

The *stmetrics*<sup>1</sup>, package available in Python, allows the extraction of features from the time series. The package currently includes three modules to perform feature extraction. The first module is composed of basic metrics, that are derived from the time series using common statistical approaches, in which 10 metrics are available:

- Max - Maximum value of the time series.
- Min - Minimum value of the time series.
- Mean - Average value of the time series.
- Std - Standard deviation of the time series.
- Sum - Sum of values over a cycle. Usually is an indicator of the annual production of vegetation.
- Amplitude - The difference between the time series's maximum and minimum values.
- First slope - Maximum value of the first slope of the cycle.
- Skew - Measures the asymmetry of the time series.
- AMD - Absolute Mean Derivative.
- AbsSum - Absolute Sum of values over of the time series.

The second module implements the polar metrics proposed by [Körting et al. \(2013\)](#). Currently, 7 metrics are available:

- Area - Area of the closed shape.
- Angle - The main angle of the closed shape created after transformation.

---

<sup>1</sup><https://github.com/brazil-data-cube/stmetrics>

- Area s1 - Partial area of the shape, proportional to quadrant 1 of the polar representation.
- Area s2 - Partial area of the shape, proportional to quadrant 2 of the polar representation.
- Area s3 - Partial area of the shape, proportional to quadrant 3 of the polar representation.
- Area s4 - Partial area of the shape, proportional to quadrant 4 of the polar representation.
- Polar balance - The standard deviation of the areas per season, considering the 4 seasons.
- Gyration radius - Equals the average distance between each point inside the shape and the shape's centroid.

These metrics are derived from a time wheel legend proposed by [Edsall et al. \(1997\)](#). As exposed by [Soares et al. \(2020\)](#), 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, and  $T$  is a time interval  $t_1, \dots, t_N$ , 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:

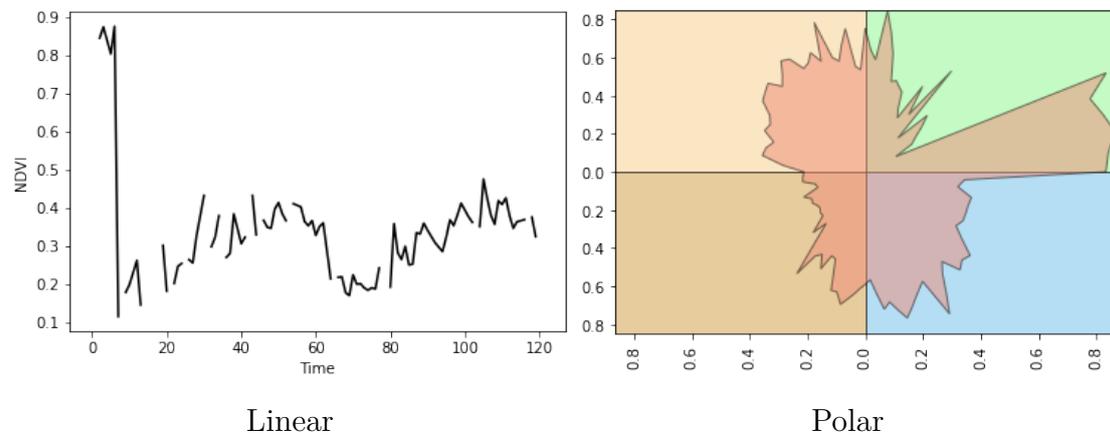
$$a_i = v_i \cos\left(\frac{2\pi i}{N}\right) \in A, i = 1, \dots, N. \quad (2.1)$$

and

$$o_i = v_i \sin\left(\frac{2\pi i}{N}\right) \in O, i = 1, \dots, N. \quad (2.2)$$

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. An example of polar plot is represented in [Figure 2.6](#). The same time series is represented in both linear and polar plots. In the latter, the start point is the bottom of the top right quadrant, advancing in anti-clockwise direction. From this plot, the metrics area calculated, as the area of each feature inside the quadrants, for example.

Figure 2.6 - Linear and polar plots.



SOURCE: Authors's production.

The third module implements the fractal metrics. Currently, 3 metrics are available:

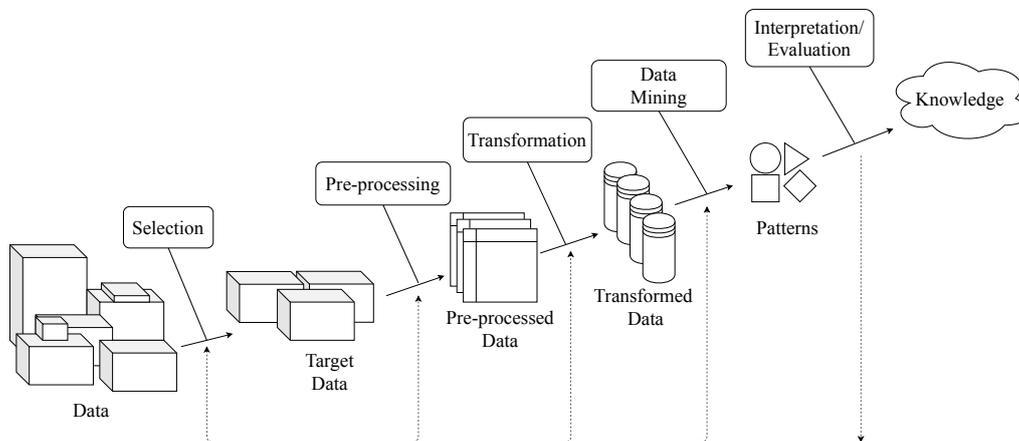
- Hurst - Computes the Hurst Exponent (HE) by a standard re-scaled range (R/S) approach. HE is a self-similarity measure that assesses long-range dependence in a time series. It can be used to determine whether the time series is more, less, or equally likely to increase if it has increased in previous steps.
- DFA - Detrended Fluctuation Analysis measures the Hurst parameter H, which is very similar to the Hurst exponent. The main difference is that DFA can be used for non-stationary time series.
- Katz - Computes the fractal dimension using Katz algorithm.

### 2.3 Data mining

Data Mining (DM) is a process which consists of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns (or models) over the data (FAYYAD et al., 1996). It is a core process inside the wider concept of Knowledge Discovery in Databases (KDD) coined by Piatetsky-Shapiro (1990). KDD refers to the overall process of discovering knowledge from data, while DM refers to a specific step. According to Fayyad et al. (1996), the distinction between KDD process and DM step (within the

process) is important, once the additional steps in KDD, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that knowledge is derived from the data. DM component of KDD relies specially on known techniques of machine learning, pattern recognition, and statistics to find patterns from data. Figure 2.7 illustrates the steps needed to convert raw data into useful knowledge. It is important to highlight that KDD is not a linear process; which means that steps can be repeated in order to refine the relation between data and patterns.

Figure 2.7 - Steps that compose the KDD process.



SOURCE: Adapted from Fayyad et al. (1996).

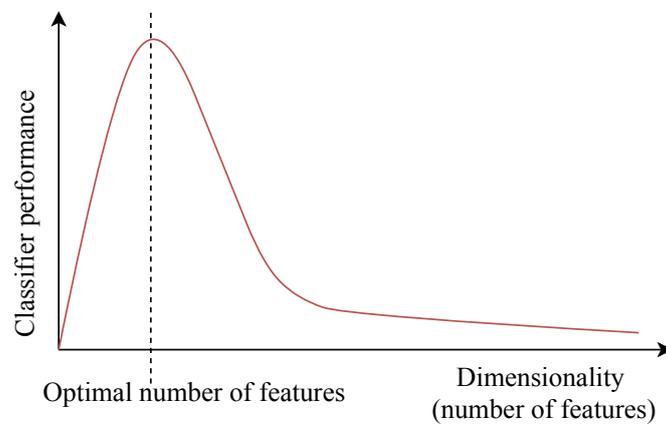
According to Larose and Larose (2014), these are the main DM tasks: description, estimation, prediction, classification, clustering and association. Regarding remote sensing, the classification consists in a technique of extracting information from digital images, and is responsible for giving a meaning or labeling a pixel (or a set of them) according to its proprieties, being them spectral, spatial, etc (NOVO, 2010).

Many attributes can be extracted from digital images. The reflectance values from the images derive an attribute itself, and by combining values from different spectral bands new attributes can be generated (i.e, vegetation indices). A variety of combinations and transformations have been developed in order to extract more information from the original material.

The advances in feature extraction techniques have resulted in a great capability of generating and storing big amounts of data. However, it also includes irrelevant or redundant data, which do not aid discrimination and can be discarded. Removing the least effective features is referred to as feature selection. [Hall \(1999\)](#) defines feature selection as a learning step that focuses on the most useful data aspects for analysis and feature prediction. In that sense, the exclusion of unnecessary attributes allows the time reduction for generating the model and increases the classification overall accuracy ([DASH; LIU, 1997](#)).

Hughes phenomenon ([HUGHES, 1968](#)), also known as “the curse of dimensionality”, demonstrates that the accuracy of a classifier depends on the number of training samples. It shows that, as the number of features increases, the classifier performance increases as well until we reach the optimal number of features (Figure 2.8). Adding more features based on the same size as the training set will then degrade the classifier performance.

Figure 2.8 - Hughes phenomenon.



SOURCE: Adapted from [Hsu \(2007\)](#).

In order to deal with this effect, some classifiers have already included in their algorithms the feature selection process, as the DT and RF. Classification algorithms can be divided in two main groups: unsupervised and supervised. Unsupervised image classification is the process by which each image in a dataset is identified to be a member of one of the inherent categories present in the image collection without

the use of labelled training samples (OLAODE et al., 2014). It searches for patterns and structures capable to join similar data. Examples of unsupervised algorithms are the K-means and ISODATA.

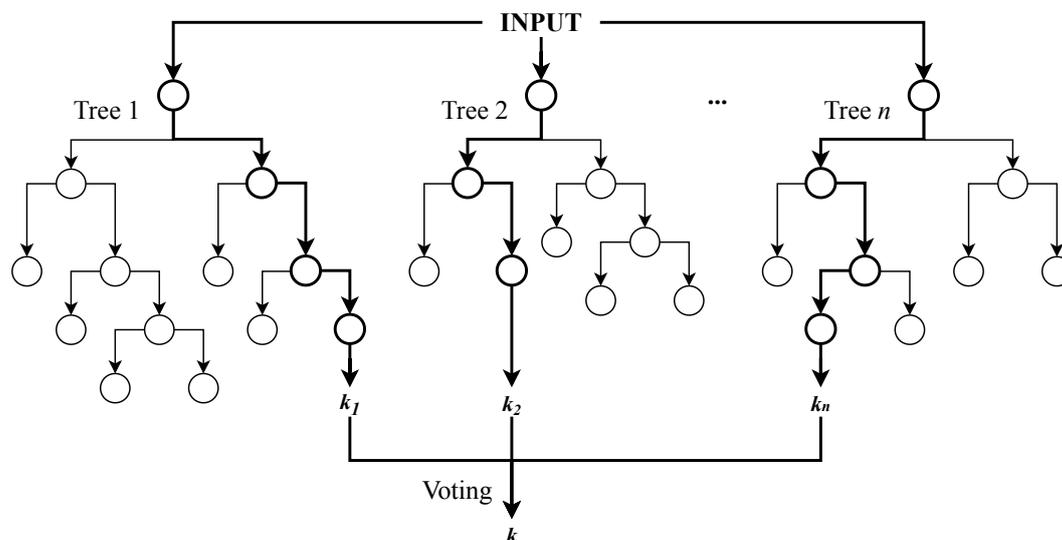
On the other hand, supervised methods work with pre-established classes, having their algorithms been trained beforehand. A parametric supervised method adopts the assumption that the classes can be modelled by probability distributions and, as a consequence, are described by the parameters of those distributions (RICHARDS, 2013). An example is the ML classification algorithm. Other techniques, in which neither distribution models nor parameters are relevant, are referred to as non-parametric methods. Examples are the SVM, DT and RF.

### 2.3.1 The Random Forest classification algorithm

The RF consists of a decision tree classifiers combination where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unique vote for the most popular class to classify an input vector (BREIMAN, 2001). The trees are created by drawing a subset of training samples through replacement (a bagging approach); which means that the same sample can be selected several times, while others may not be selected at all (BELGIU; DRĂGUT, 2016). Some proportion of the samples are used to train the trees, while the remaining part is used for an internal cross-validation technique, estimating how well the resulting model is performing (BREIMAN, 2001). Those samples are called in-bag and out-of-bag samples, respectively.

The algorithm creates trees that have high variance and low bias, based on two user-defined parameters: the number of trees ( $Ntrees$ ) and the number of features ( $Mtry$ ). The final classification decision is taken by the majority value of the class assignment calculated by all trees. Researches have pointed out that classification accuracy is less sensitive to  $Ntree$  than to the  $Mtry$  parameter (KULKARNI; SINHA, 2012; GHOSH et al., 2014). Another important observation by Guan et al. (2013) is that, once the RF classifier is computationally efficient and does not overfit,  $Ntree$  can be as large as possible. Many studies have proved the effectiveness of this classifier for LULC mapping and landslide detection (PELLETIER et al., 2016; WANG; LU, 2019; CHEN et al., 2017).

Figure 2.9 - Random Forest structure.



Each  $k$  represents a final class labeled by the trees.

SOURCE: Adapted from Verikas et al. (2016).

## 2.4 Sampling methods

There are many sampling design options, and each of them fulfills different objectives. According to Stehman (2009), three key decisions that strongly influence the choice of sampling design are whether to use strata, clusters, or to implement a systematic/simple random selection protocol. Stratification corresponds to the partitioning of the Region Of Interest (ROI) so each assessment unit is assigned to a single stratum. The stratum in classification, is usually related to the class in the map. Olofsson et al. (2014) points out that stratification is implemented for two primary purposes. Firstly, when the strata are of interest for reporting results (e.g., accuracy and area are reported by land cover class or by geographic sub-region). Secondly, to improve the precision of the accuracy and area estimates; for example, when strata are created for reporting accuracy by strata, the stratified design allows specifying a sample size for each stratum guaranteeing that a precise estimate is obtained for each stratum.

## 2.5 Sentinel-2/MSI

The Copernicus Sentinel-2 mission (Fig. 2.10) comprises a constellation of two identical polar-orbiting satellites placed in the same sun-synchronous orbit, phased at

180° to each other (ESA, 2015). Their orbit stays at a mean altitude of 786km, which allows a wide swath reaching a 290km field-of-view (FOV). The coverage limits are between latitudes 56° south and 84° north. The revisit time is 10 days at the equator with one satellite, and five days with the combination of both satellites under cloud-free conditions.

Figure 2.10 - Sentinel-2 satellite.



SOURCE: ESA (2015).

The MultiSpectral Instrument (MSI) works passively, by collecting sunlight reflected from the Earth. The incoming light beam is split at a filter and focused onto two separate focal plane assemblies within the instrument; one for Visible and Near-Infra-Red (VNIR) bands and one for Short Wave Infra-Red (SWIR) bands (ESA, 2015). The MSI imager contains 13 spectral bands (443 nm – 2190 nm), which are represented in Figure 2.11. The spatial resolutions are 10 m, for four bands; 20 m, for six bands; and 60 m, for three bands (REGAN et al., 2010).

According to Jensen (2009), radiometric resolution is defined as the sensitivity of a remote sensing detector to differences in signal strength as it records the radiant flux reflected, emitted, or back-scattered from the terrain. It defines the number of just discriminable signal levels, usually represented in bits ranging from 8 to 16. The greater the radiometric resolution, the more accurate the sensed image could be. The MSI instrument provides 12 bits imagery, enabling the image to be acquired over a range of 0 to 4095 ( $2^{12} - 1$ ) potential light intensity values.



- e) Imagery compression of the resultant Level-1C imagery via the JPEG2000 algorithm and a GML geographic imagery-encoded header.

In resume, Level-1C product provides orthorectified TOA reflectance, with sub-pixel multispectral registration. Cloud and land/water masks are included in the product. Level-2A product provides orthorectified Bottom-Of-Atmosphere (BOA) reflectance, with sub-pixel multispectral registration. A Scene Classification map (cloud, cloud shadow, vegetation, soils/deserts, water, snow, etc.) is incorporated in the product.

Sentinel-2 provides data feeding services for applications in the Copernicus priority areas of land monitoring, emergency management and security (ESA, 2015). In behalf of Copernicus programs, data from Sentinel-2 supports services such as land management, agriculture, forestry as well as disaster management and humanitarian relief operations.

## 2.6 SRTM digital elevation model

A Digital Elevation Model (DEM) can be defined as a numeric representation of a topographic surface arranged as a set of regularly spaced points, normally in a square grid or hexagonal pattern, expressed as three-dimensional coordinates (KENNIE; PETRIE, 1990; OKSANEN et al., 2006). In other words, it is a digital representation of the land surface elevation with respect to a given reference datum.

The Shuttle Radar Topography Mission (SRTM) is a joint project between the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space Administration (NASA), which aimed to produce digital topographic data for 80% of the Earth's land surface. The SRTM was carried on-board the space shuttle *Endeavour* from February 11 to 22, 2000, which orbited Earth 16 times each day during the 11-day mission, completing 176 orbits (USGS, 2021a). It considers all land areas between 60° north and 56° south latitude, with data points located every 1-arc-second (approximately 30 meters) on a latitude/longitude grid. The absolute vertical accuracy of the elevation data is around 16 meters (at 90% confidence).

These data can be used for military, civil, and scientific purposes. It provides an accurate product, which can benefit different types of needs, as flood control, soil conservation, reforestation, volcano monitoring, earthquake research, and glacier movement monitoring.

In order to produce the DEM, SRTM made use of a technique called radar interfer-

ometry. In this method, two radar images are taken from slightly different locations. The calculation of surface elevation is realized based on the differences between these images. To get two radar images, of the same spot, taken from different locations the SRTM hardware consisted of one radar antenna in the shuttle payload bay and a second radar antenna attached to the end of a mast extended 60 meters out from the shuttle (USGS, 2021b).

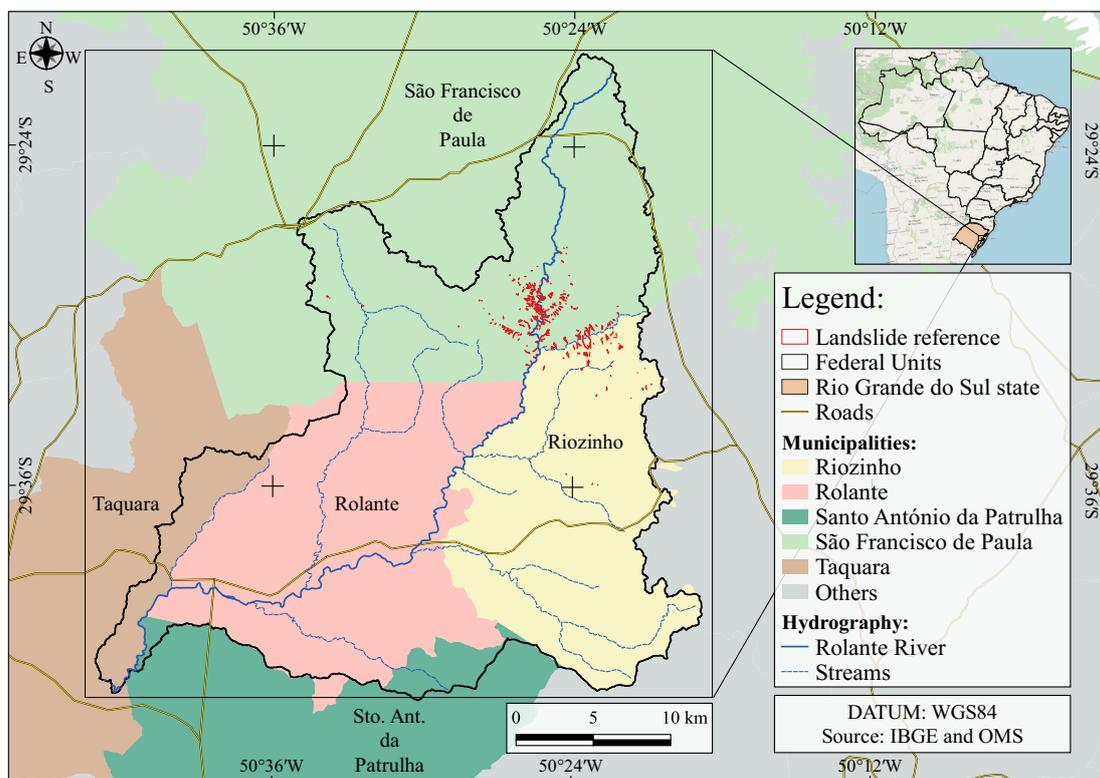
SRTM was launched in an orbit with an inclination of 57 degrees, which allowed all of the Earth's land surface that lies between 60 degrees north and 56 degrees south latitude to be covered by it. This represents about 80 percent of the Earth's land surface.



### 3 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. Its drainage area comprehends 828km<sup>2</sup>, with elevation values varying between 20m to 1040m. 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 (PETRY, 2003). 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 (LUERCE, 2012).

Figure 3.1 - Study area, the Rolante River Catchment.

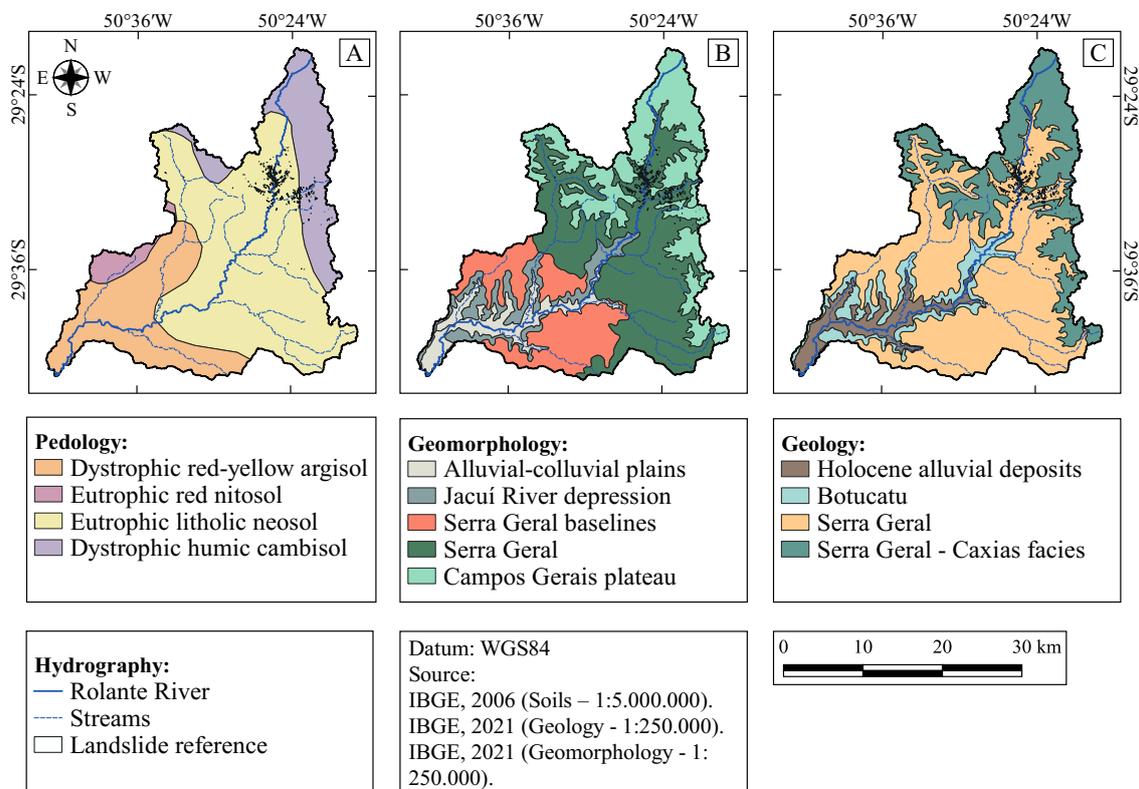


SOURCE: Author's production.

Figure 3.2 shows more details on the geographical characterization of the study area. 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, Jacuí River depression, Serra Geral baselines, Serra Geral unit, 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. According to [Rossato \(2011\)](#), the climate is very humid subtropical, characterized by abundant annual precipitation varying between 1700mm to 2000mm and temperatures of 14°C to 17°C.

Figure 3.2 - Geographical characterization of the study area.



Pedology map is represented by figure “A”, while Geomorphology and Geology maps are represented by “B” and “C”, respectively. Note that the landslide scars are in black color.

SOURCE: Author’s production.

Previous works have detected around 300 landslide scars at the catchment ([GAMEIRO et al., 2019](#); [QUEVEDO et al., 2019](#)). The scars occurred after an extreme precipitation event on January 5<sup>th</sup>, 2017. The rain registered 90mm to 272mm and lasted for about four hours ([SEMA, 2017](#)). 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.

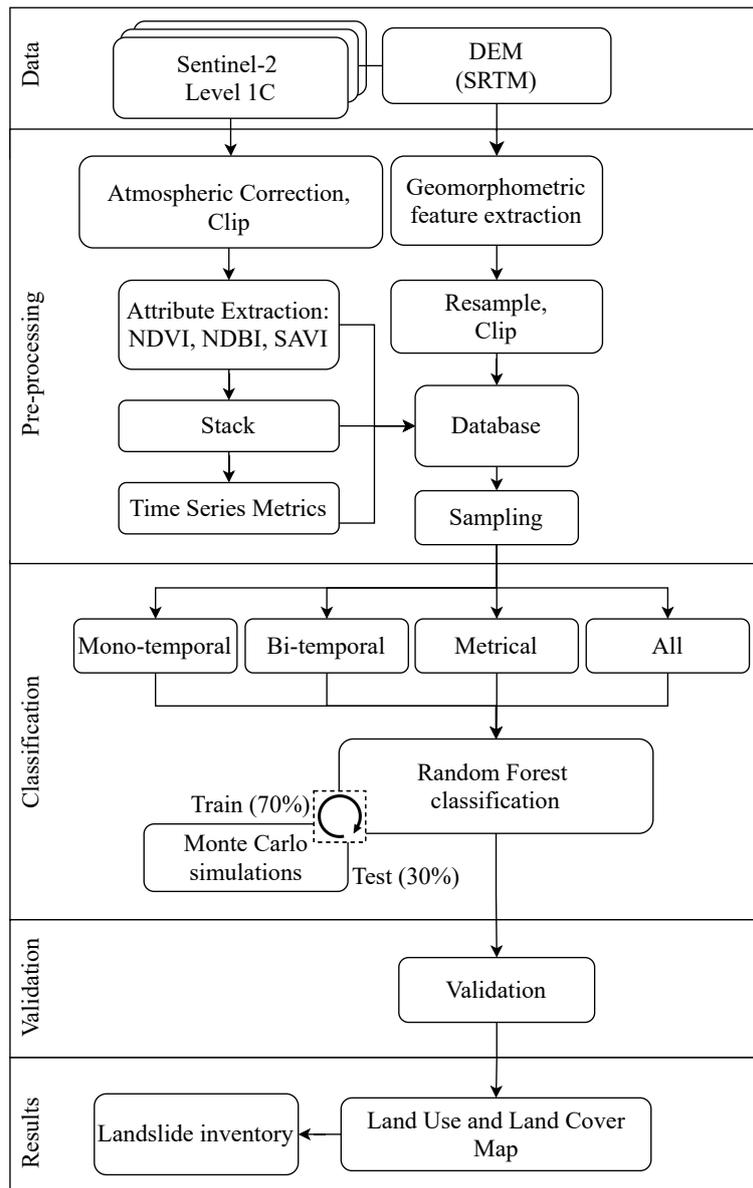
The criteria to select this area for the research consists in two main factors. First, the significant presence of already mapped landslide scars. Second, the magnitude of the event, which allows the identification of the scars via orbital imagery of 10m spatial resolution, freely available by the European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



## 4 METHODOLOGY

The methodology developed for this research is represented in Figure 4.1. This chapter will clarify each part of the process in details.

Figure 4.1 - Methodology flowchart.



SOURCE: Author's production.

## 4.1 Data and pre-processing

The identification of landslide scars usually present better results when high spatial resolution images are used (JOYCE et al., 2009). Considering that, Sentinel-2 (A and B) imagery was chosen, especially because it provides free orthorectified reflectance products with 10m of spatial resolution and a 5-day temporal resolution. Level 2A (BOA) product was not available for the date of the event (Jan, 2017), so Level 1C was used. Moreover, a 30m spatial resolution DEM from SRTM was used.

In order 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. (2020), 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, which resulted in a less refined temporal resolution than 5 days. Further, an outlier removal filter was applied to smooth the time series. This filter consists in identifying values that present more than 10% of difference from the previous and the following values, if positive, it is replaced by the mean value of these two neighbors. The time interval available for the study area ranges from November 11<sup>th</sup>, 2015, to June 3<sup>rd</sup>, 2020; which, in total, resulted in 122 dates. The general temporal resolution of this time series is around 13 days. This can be explained by two main factors: a) some images was discarded because of cloud coverage; b) Sentinel-2B was launched in 2017, almost two years after Sentinel-2A. Both satellites together make a 5-day resolution product; however, each of them has a revisiting time of 10 days.

## 4.2 Feature extraction

The feature extraction process, illustrated in Figure 4.2, is based on landslide detection literature, as presented by Gerente et al. (2017a), Gerente et al. (2017b), Joyce et al. (2009). The NDVI, developed by Rouse et al. (1974), was used because it presents a drastic reduction in its values when the landslide occurs. Besides 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 attributes is specifically beneficial when dealing with dense time series. It is used to detect varying densities of vegetation coverage which can be applied for natural disasters (BHANDARI et al., 2012). The literature shows many successful researches using the NDVI for land cover mapping (JEEVALAKSHMI et al., 2016; KONG et al., 2016). The Normalized Difference Built-up Index (NDBI), proposed by Zha et al. (2003), was used to 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 (1988), 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), in order to adjust the soil back-ground effect. In the absence of extrinsic knowledge, an intermediate adjustment factor of 0.5 has been suggested and generally applied (LAWRENCE; RIPPLE, 1998). A variety of papers mention the SAVI index as an useful tool to detect landslides, as the one by Salleh et al. (2019), 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:

$$\text{NDVI} = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}} \quad (4.1)$$

$$\text{NDBI} = \frac{\rho_{\text{swir}} - \rho_{\text{red}}}{\rho_{\text{swir}} + \rho_{\text{red}}} \quad (4.2)$$

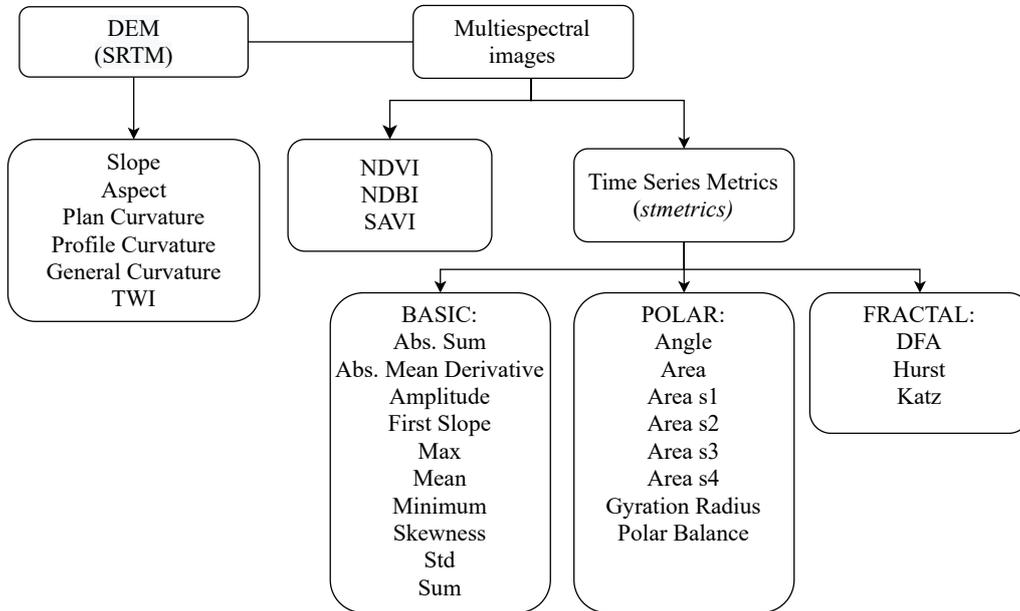
$$\text{SAVI} = (1 + L) \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}} + L} \quad (4.3)$$

Where:

- $\rho_{\text{nir}}$ : reflectance value for the NIR band;
- $\rho_{\text{red}}$ : reflectance value for the Red band;
- $\rho_{\text{swir}}$ : reflectance value for the SWIR band;
- $L$ : constant inversely correlated with the Leaf Area Index (LAI).

The SAGA tool, available in the QGIS software, allows the extraction of the geomorphometric features from the DEM. The use of these attributes is based on previous studies, which have been associated to the landslide occurrence (SOTHE et al., 2017; PETSCHKO et al., 2012; SARO et al., 2016). The extracted attributes are Slope, Aspect, Plan Curvature, Profile Curvature, General Curvature, and the Topographic Wetness Index (TWI). Figure 4.3 illustrates these geomorphometric attributes. Once the Sentinel-2 data has 10m of spatial resolution, the geomorphometric attributes were resampled from 30m to 10m using the Nearest Neighbor parameter in QGIS.

Figure 4.2 - Feature extraction.



SOURCE: Author's production.

The elevation attribute 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  $0^\circ$  to  $90^\circ$ , although is commonly expressed in percentage (VALERIANO; ROSSETTI, 2008). The vertical curvature ( $^\circ/\text{m}$ ) is related to the profile shape of the hillside, referring to the convex/concave characteristic of the hill. The horizontal curvature ( $^\circ/\text{m}$ ) corresponds to the hill shape when observed in its horizontal projection. Is defined as a second-order derivative from the contour lines. It can be described as a variation of the aspect within a certain distance. The aspect is defined as the azimuthal angle corresponding to the highest declivity of the terrain, in the descendent direction (VALERIANO; ROSSETTI, 2008). It is measured in degrees, from  $0^\circ$  to  $360^\circ$ . The topographic wetness index (TWI) was developed by Beven and Kirkby (1979) within the runoff model TOPMODEL. The TWI is defined by:

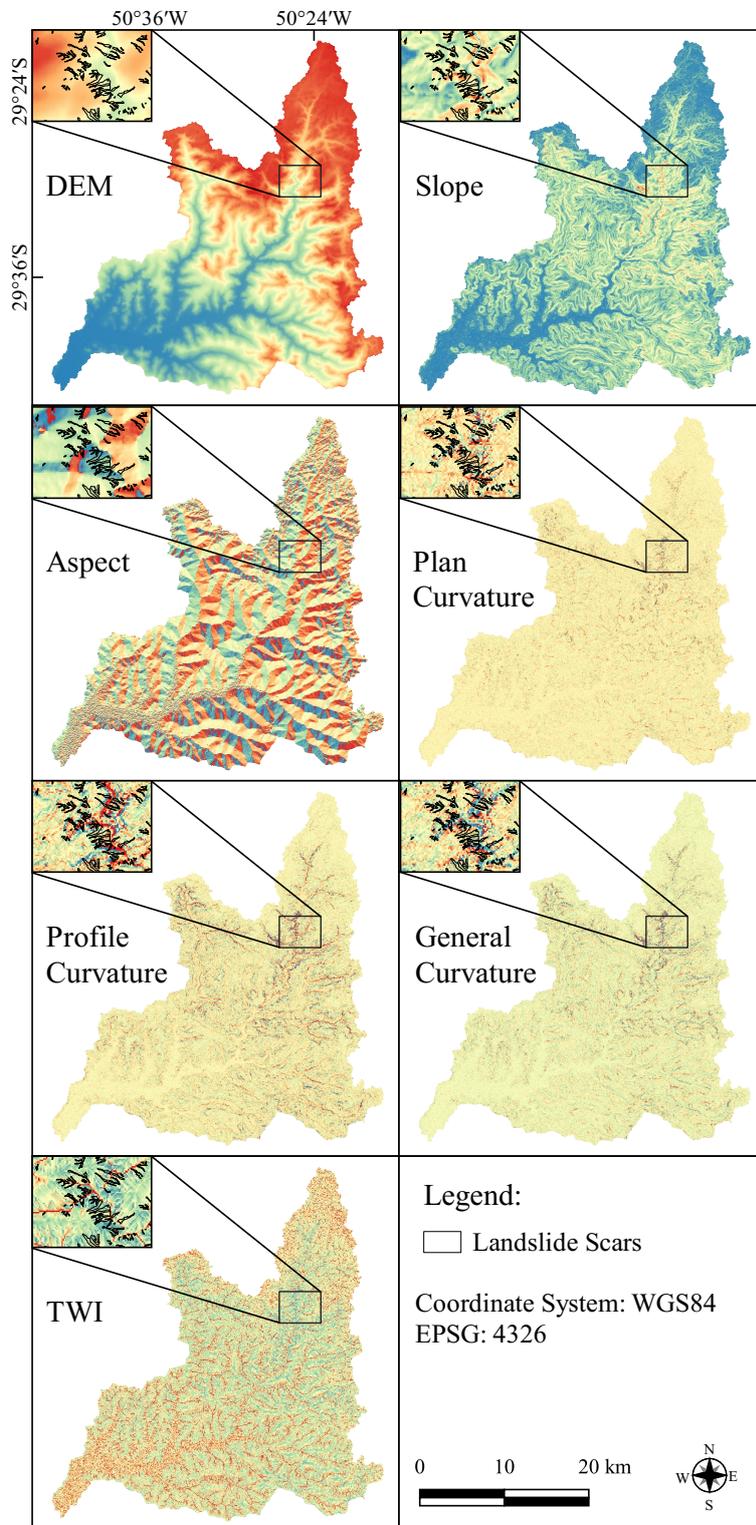
$$TWI = \ln(a/\tan\beta) \quad (4.4)$$

where:

- $a$ : upslope area per unit contour length;
- $\tan\beta$ : 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 (SØRENSEN; SEIBERT, 2007). 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 (QIN et al., 2011).

Figure 4.3 - Geomorphometric attributes at Rolante River's Basin.



SOURCE: Author's production.

### 4.3 Sampling

Considering the explanations on Sec. 2.4, a stratified sampling was applied. In that sense, each class was sampled proportionally to their area based on previous classification works of the region (UEHARA et al., 2019; LUERCE, 2012). 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 30/70 for training and validation. The sampling was manually realized based on high resolution imagery from Google Earth. In order 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 ( $25\text{km}^2$ ) 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 an specific region, some polygons of the grid did not contain their samples.

Regarding the sample size, Congalton et al. (1991) suggests a minimum of 50 samples for each class, increasing to 75 - 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, having considerably less sample points. As a matter of fact, a first attempt was made to divide the Landslide class into two sub-classes “upper” and “deposit”, in order 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 10m spatial resolution data this could not be separated, so only one class was created.

The classification labels are: 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.

### 4.4 Classification

The Random Forest classification algorithm was used for this procedure (for classifier’s description refer to Sec. 2.3.1). This process was conducted using the *scikit-*

*learn*<sup>1</sup> package in Python programming language. The classification was realized under four distinct approaches, in which each of them considered a different set of input attributes. As a result, four classification products were compared (Fig. 4.4). The input approaches are:

- Mono-temporal: single date image (after the landslides occurrence) with NDVI, NDBI and SAVI + geomorphometric attributes. Total of 10 attributes.
- Bi-temporal: two dates images (before and after the landslides occurrence) with NDVI, NDBI and SAVI + geomorphometric attributes. Total of 13 attributes.
- Metrical: time series metrics for NDVI, NDBI and SAVI + geomorphometric attributes. Total of 73 attributes.
- All: 122 dates time series with NDVI, NDBI and SAVI + time series metrics for NDVI, NDBI and SAVI + geomorphometric attributes. Total of 436 attributes.

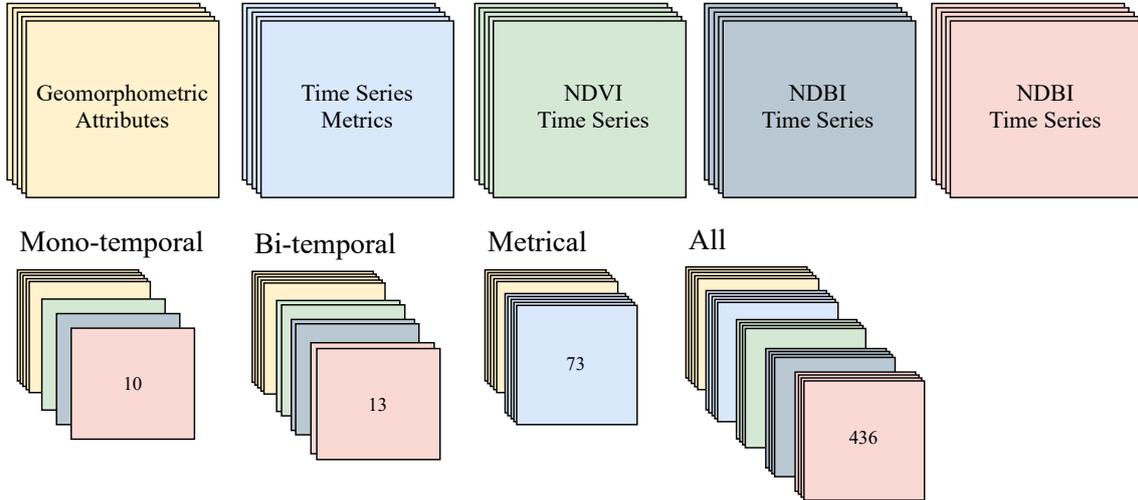
Once most of the landslide scars occurred during an extreme precipitation event in 2017, Jan, 5<sup>th</sup>, the aim was to find the closest images from that date do build the Mono and Bi-temporal datasets. However, this period presents high levels of cloud coverage, interfering in the images quality. For that reason, the chosen closest images to the date of the event are from June 9<sup>th</sup>, 2016 (before) and March 11<sup>th</sup>, 2018 (after). This reveals the difficulty of choosing few images to classify landslide prone areas. Figure 4.5 shows the distribution of the dates used in the approaches. Figure 4.6 shows the two dates chosen for the Bi-temporal approach. Figure 4.7 shows all images available between these two dates that were rejected because of cloud coverage issues.

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 in order 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.

---

<sup>1</sup><https://scikit-learn.org/stable/>

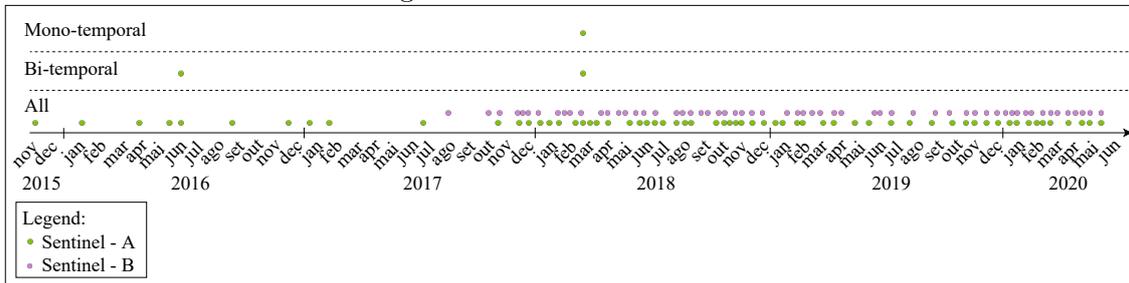
Figure 4.4 - Classification approaches.



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 (grey), SAVI Time Series (red). Below, the dataset of each approach is represented and the number inside the box reveals the quantity of attributes. Note that for Mono and Bi-temporal approach only one or two layers of NDVI, NDBI and SAVI are illustrated.

SOURCE: Authors' production.

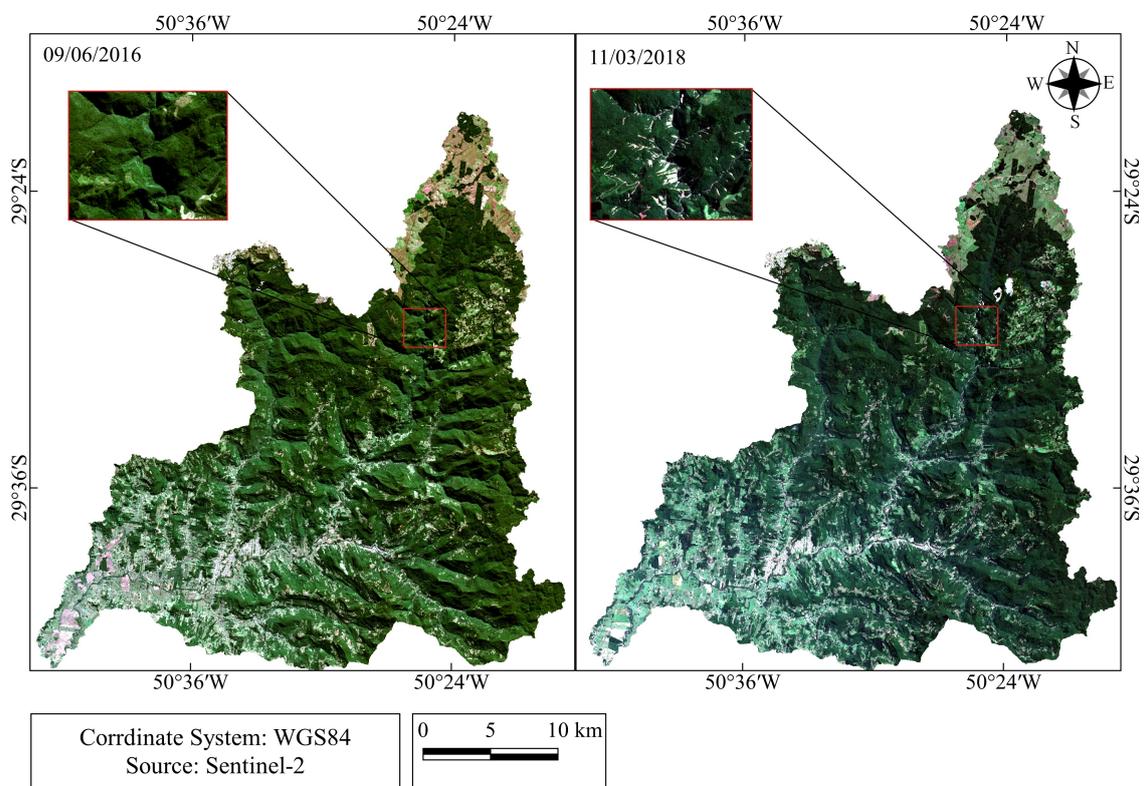
Figure 4.5 - Dates distribution.



Each point represents the date of the image used for the corresponding approach. Green points refer to Sentinel-2A images, while pink points to Sentinel-2B.

SOURCE: Authors' production.

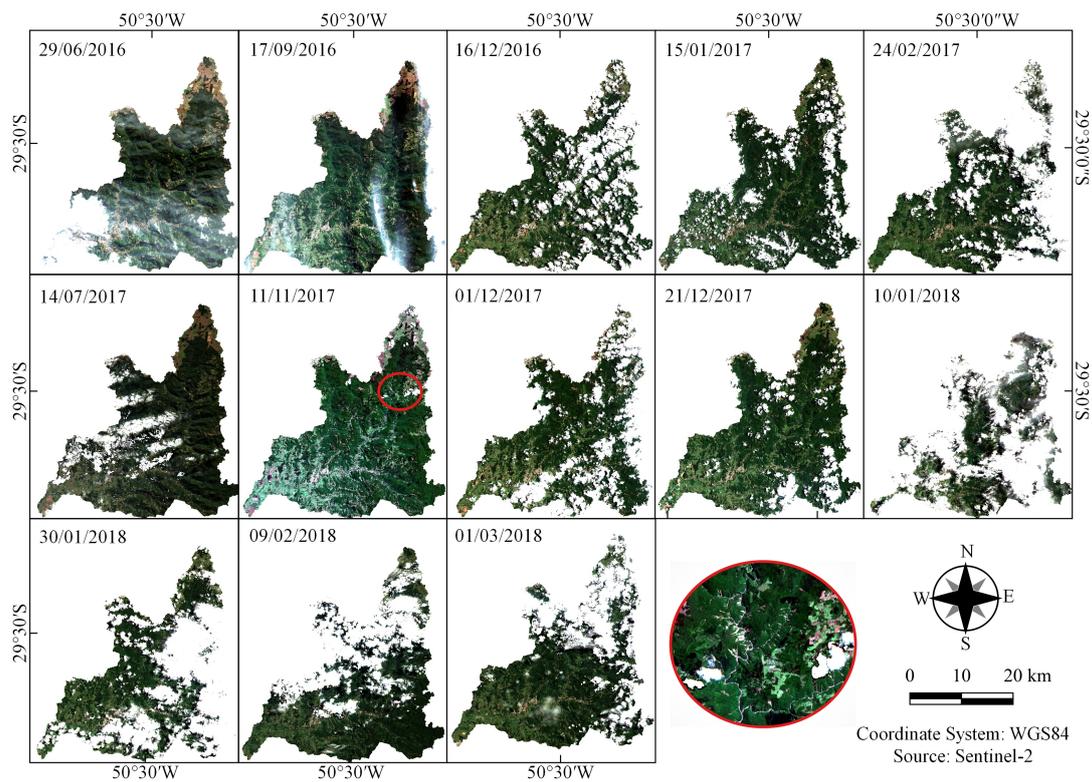
Figure 4.6 - Bi-temporal approach.



Chosen images from before and after the landslides occurrence.

SOURCE: Authors' production.

Figure 4.7 - Cloud coverage issue.



The landslides event occurred on Jan 5<sup>th</sup>, 2017. However, the closest images from that date have a significant cloud coverage. Special attention is given to Nov 11<sup>th</sup>, 2017, in which the image appears clear, but presents clouds exactly above the landslide scars (note the red circle zoom).

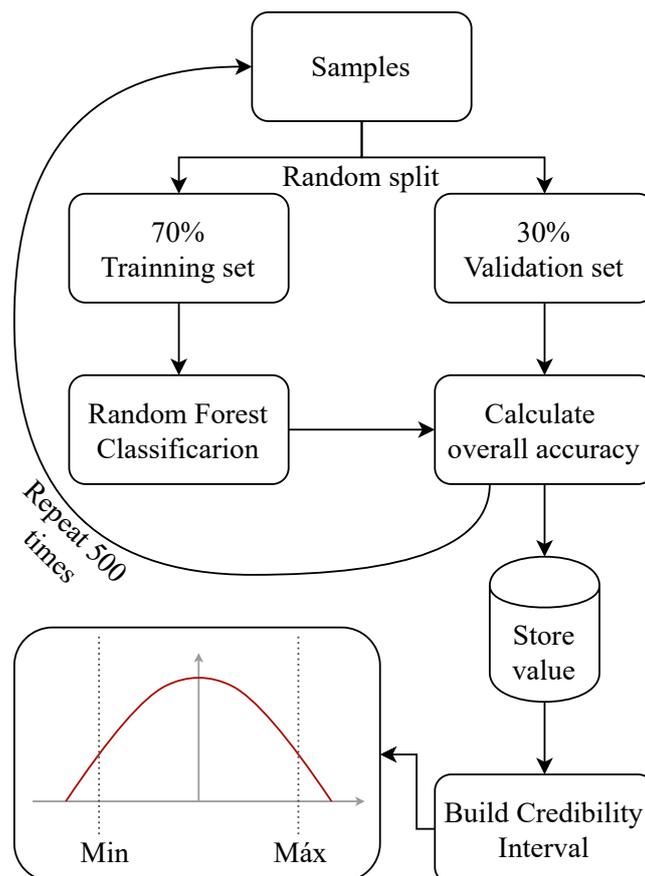
SOURCE: Authors' production.

## 4.5 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. \(2014\)](#), and an inventory map provided by [Quevedo et al. \(2019\)](#) and updated for this research. This inventory was realized by visual interpretation of high spatial resolution imagery. The reference was used for the construction of a sampling set for the classification.

In order 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 4.8 shows how this procedure was conducted via Monte Carlo simulations.

Figure 4.8 - Monte Carlo Simulations.



SOURCE: Authors' production.

Monte Carlo simulation is a type of simulation that relies on repeated random sampling and statistical analysis to compute the results (RAYCHAUDHURI, 2008). In this case, from all samples available, the set is randomly split in the proportion 70% for training and 30% for validation. The training set is used to perform the RF classification, while the validation samples are used exclusively to validate and calculate the overall accuracy value. This value is stored and the process is repeated 500 times. The credibility interval is built for 5% of significance using all values stored. A classification accuracy value inside this interval was used for comparison among the approaches.

The classifications evaluation and performance analysis was held under the statistical measures derived from the confusion matrix. An example of confusion matrix is showed in Table 4.1, so the statistical measures can be explained. 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 is extracted from the matrix according to the formula expressed below, which ranges from 0.0 to 1.0 (100% of accuracy).

Table 4.1 - Example of confusion matrix.

Predicted \ True Label	C <sub>1</sub>	C <sub>2</sub>	...	C <sub>c</sub>	Total
C <sub>1</sub>	x <sub>11</sub>	x <sub>12</sub>	...	x <sub>1c</sub>	x <sub>1+</sub>
C <sub>2</sub>	x <sub>21</sub>	x <sub>22</sub>	...	x <sub>2c</sub>	x <sub>2+</sub>
...	...	...	...	...	...
C <sub>c</sub>	x <sub>c1</sub>	x <sub>c2</sub>	...	x <sub>cc</sub>	x <sub>c+</sub>
Total	x <sub>+1</sub>	x <sub>+2</sub>	...	x <sub>+c</sub>	n

Where:

- C<sub>c</sub>: class *c*.
- x<sub>ij</sub>: number of points from *j* (true label), predicted as class *i* (predicted).
- x<sub>kk</sub>: total number of points correctly classified in class *k*.
- x<sub>+j</sub>: total number of points evaluated from class *j* in the true label.
- x<sub>i+</sub>: total number of points evaluated from class *i* in the prediction.

- $n$ : total number of correctly predicted points.

$$OA = \frac{\sum_{k=1}^c x_{kk}}{n} \quad (4.5)$$

Furthermore, the commission and omission errors are extracted from the matrix. This errors are analysed by class, the commission error for the class “ $C_1$ ” refers to the prediction of a pixel as “ $C_1$ ”, when in the reference it truly belongs to the class “ $C_2$ ”. The omission error, is the opposite, when a pixel truly belongs to the class “ $C_1$ ”, but is predicted as “ $C_2$ ”.

The Cohen’s kappa index (COHEN, 1960) is another method used to evaluate the model. According to Sim and Wright (2005), kappa is a measure of “true” agreement, which indicates the proportion of agreement beyond that expected by chance. In other words, the achieved beyond-chance agreement as a proportion of the possible beyond-chance agreement, and can be calculated by the following formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \quad (4.6)$$

where  $P_o$  is the proportion of observed agreements and  $P_e$  is the proportion of agreements expected by chance.

## 5 RESULTS AND DISCUSSION

### 5.1 Importance ranking

The Importance Ranking (IR) is an ordered list related to the relevance of each attribute to the classification. In a range of 0.0 to 1.0, all attributes are assigned to a value according to their importance, and the total sum of the attributes results in 1.0. This information is provided by the RF algorithm during the classification process. Table 5.1 shows the top 10 attributes, by each classification approach, and their corresponding importance values. More focus should be given to the top 5 attributes, once the Mono-temporal approach considers a total of 10 attributes only. Clarifying, all selected metrics which 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, in order 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.

Table 5.1 - Attributes’ Importance Ranking (IR).

<i>Mono-temporal (10 at.)</i>			<i>Bi-temporal (13 at.)</i>		<i>Metrical (73 at.)</i>		<i>All (436 at.)</i>	
	IR	Attribute	IR	Attribute	IR	Attribute	IR	Attribute
1 <sup>o</sup>	0.247	NDBI (11/03/2018)	0.162	NDBI (09/06/2016)	0.056	Area s2	0.0201	NDBI (09/06/2016)
2 <sup>o</sup>	0.195	NDVI (11/03/2018)	0.148	NDBI (11/03/2018)	0.036	Slope	0.0177	Slope
3 <sup>o</sup>	0.144	SAVI (11/03/2018)	0.146	NDVI (11/03/2018)	0.035	Min	0.0163	NDVI (09/06/2016)
4 <sup>o</sup>	0.130	Slope	0.107	SAVI 2018	0.030	First Slope	0.0154	NDVI (26/12/2018)
5 <sup>o</sup>	0.087	DEM	0.106	NDVI 2016	0.029	Area s4 *	0.0137	Min
6 <sup>o</sup>	0.046	TWI	0.079	Slope	0.028	Polar Balance	0.0135	Area s1
7 <sup>o</sup>	0.040	Aspect	0.071	SAVI 2016	0.027	Area s4	0.0121	Area s2
8 <sup>o</sup>	0.039	Prof. Curvature	0.053	DEM	0.027	Sum	0.0120	NDBI (22/11/2015)
9 <sup>o</sup>	0.038	Plan Curvature	0.027	TWI	0.026	Mean	0.0107	NDBI (17/09/2016)
10 <sup>o</sup>	0.031	Gen. Curvature	0.024	Prof. Curvature	0.026	Abs. Sum	0.0101	NDBI (21/03/2019)

All selected metrics which 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 attribute for all four approaches are either a NDBI image itself or a metric extracted from its time series. In order, from Mono-temporal to All approach, the first attributes 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 attributes 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 11/03/2018, once it was the first date after the event which presented a cloud free image for the whole catchment area. However, analysing All's top 10 selection, one can see that the 11/03/2018 date was not chosen. Instead, the first image selected from after the event occurrence was the NDVI from 26/12/2018, 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 attribute, even both being cloud free images.

Regarding the geomorphometric attributes the Slope stood out, being the first of them, and sometimes unique, to be chosen in every approach. It was selected among the top 5 for all of them, besides the Bi-temporal approach, which got the sixth place. In respect to the metrics, it is evident that the polar metrics showed significant importance. For Metrical approach, the first attribute chosen was the Area s2, and, besides, NDVI's Area s4 is also among the top 5. Moreover, Polar Balance and NDBI's Area s4 were selected. About All's approach, from the three metrics selected among the top 10, two of them are polar, NDBI's Area s1 and s2.

## 5.2 Time Series metrics

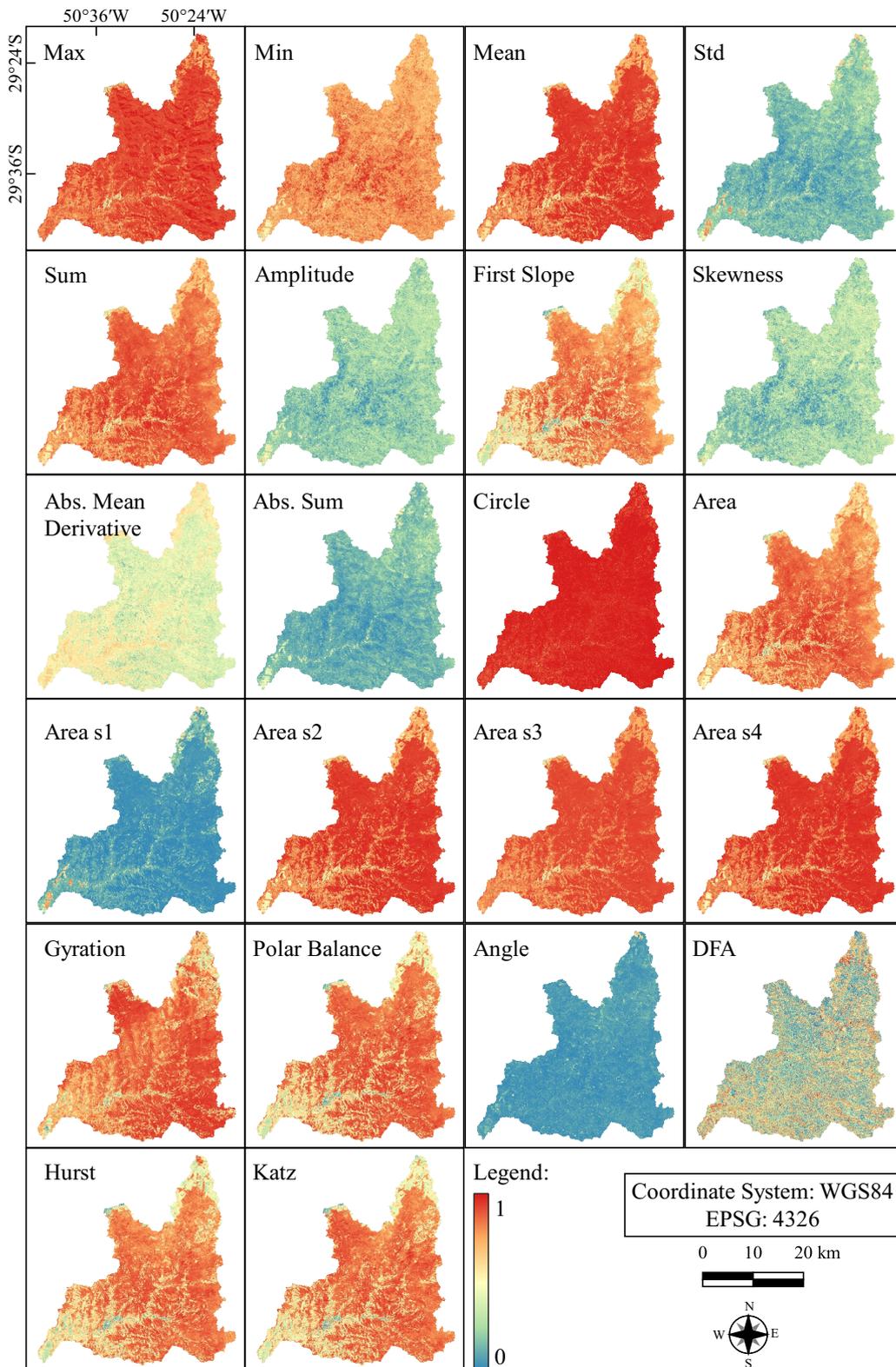
There are many possibilities of analysing the performance of a classification map. In this research, the discussion will be conducted by the analysis of the confusion matrix and its statistical products, such as the overall accuracy, kappa index, and commission and omission errors. Moreover, a deeper visual analysis of the maps is described, showing in details 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 analysed via boxplots, linear and polar graphics.

From each time series (NDVI, NDBI and SAVI), 22 metrics were extracted. The following figures reveal the results for the NDVI (Figure 5.1), the NDBI (Figure 5.2) and the SAVI (Figure 5.3). All metrics had their values normalized, being their range between 0.0 and 1.0. In the maps, blueish colors represent lower values than the reddish ones. Further in this chapter more detailed explanations will be given regarding the metrics contribution to each class prediction. However, a quick visual analysis may provide some relevant interpretations of the metrics.

Analysing the three indices results, it is notable that for all of them Area s1 has presented a considerably different output compared to Area s2, s3 and s4. This might be due to the fact that the landslide event happened chronologically on the

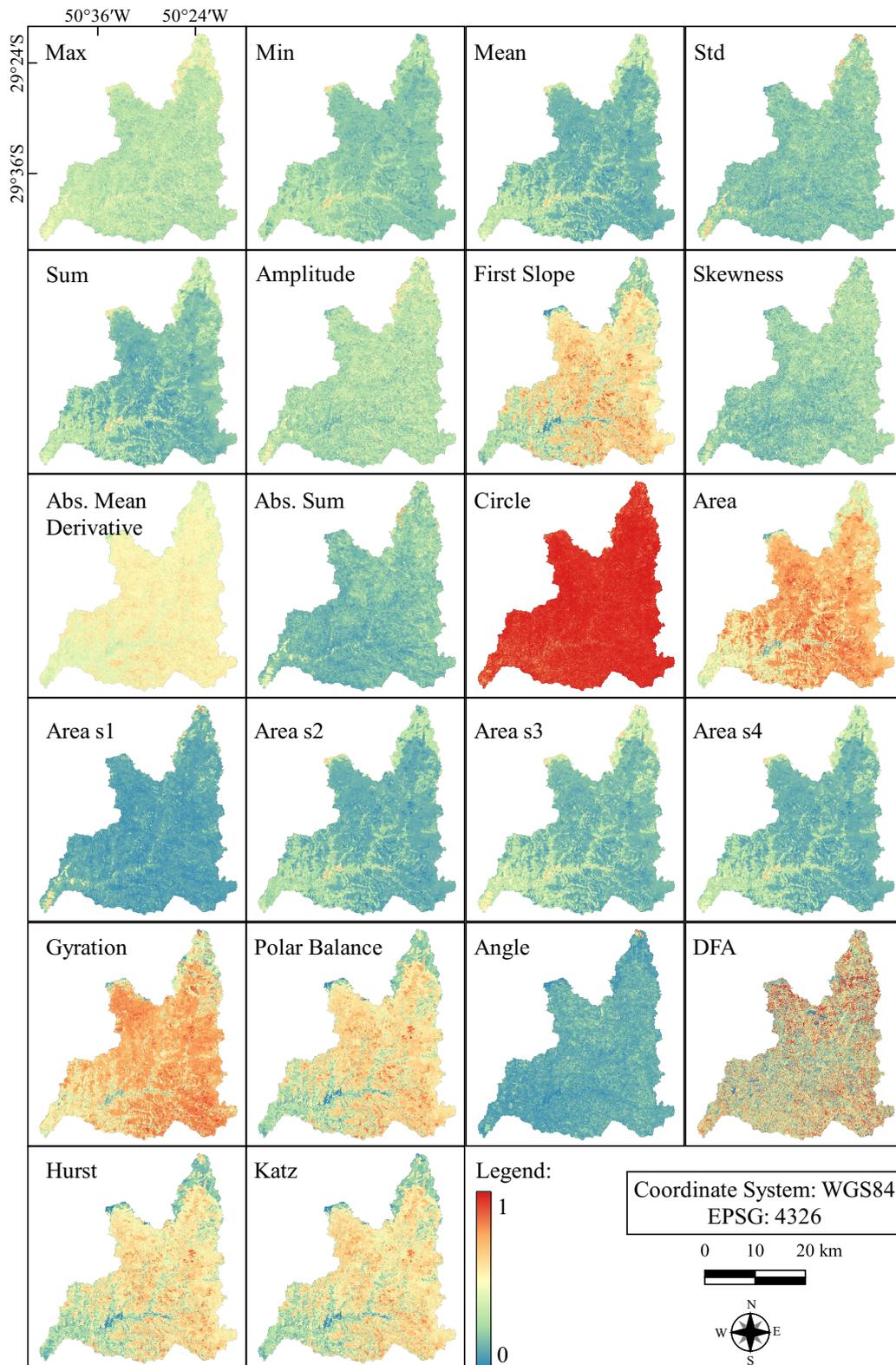
first quarter block of the time series. This means that the date of the event is between the first 30 observations. Thus, once landslides can cause a significant change in the landscape, the event itself is the hypothesis for explaining the difference among Area s1 and the others. In general, the NDVI index presented higher values for all metrics, followed by SAVI, and NDBI, respectively.

Figure 5.1 - NDVI metrics.



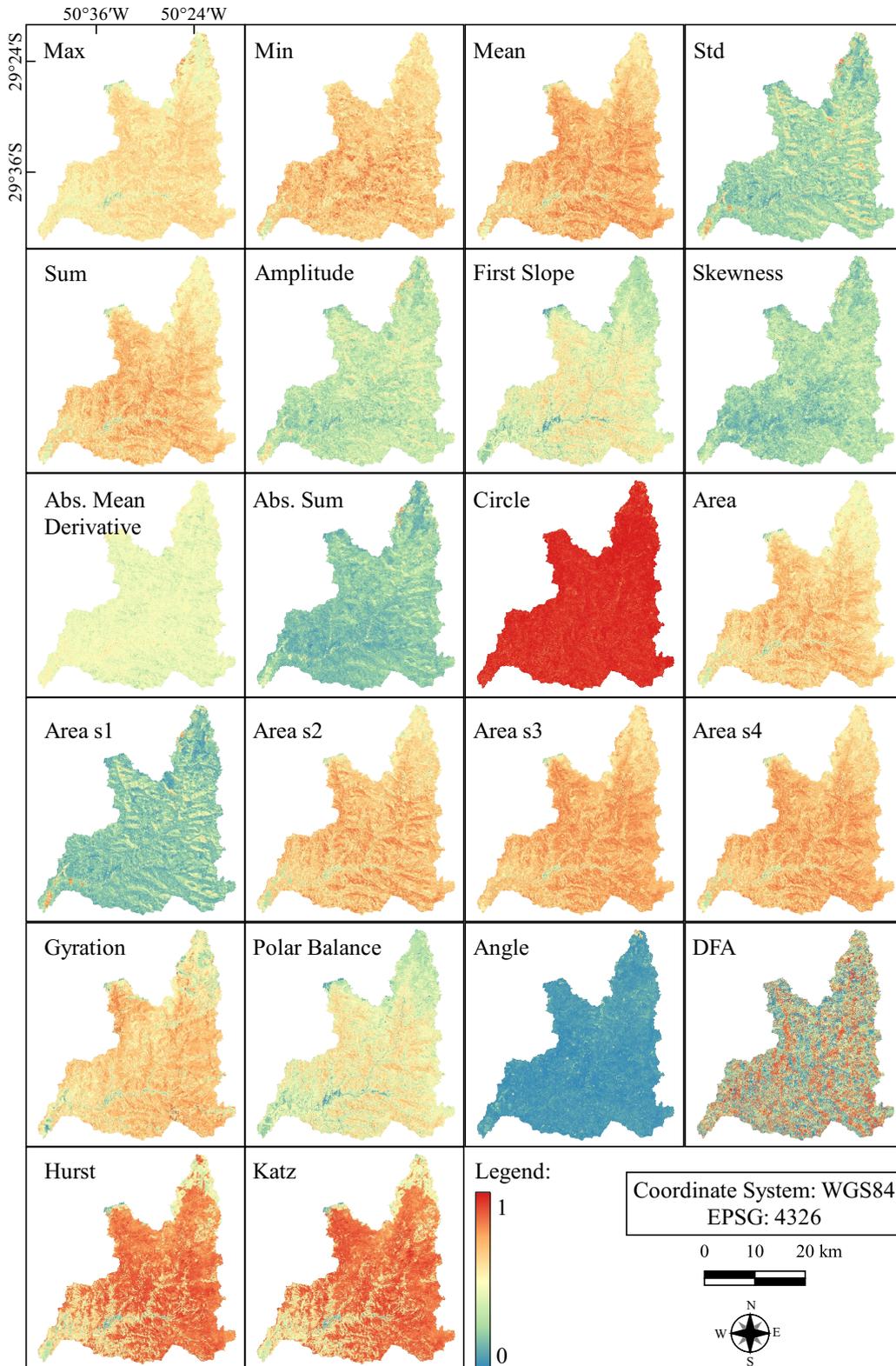
SOURCE: Author's production.

Figure 5.2 - NDBI metrics.



SOURCE: Author's production.

Figure 5.3 - SAVI metrics.

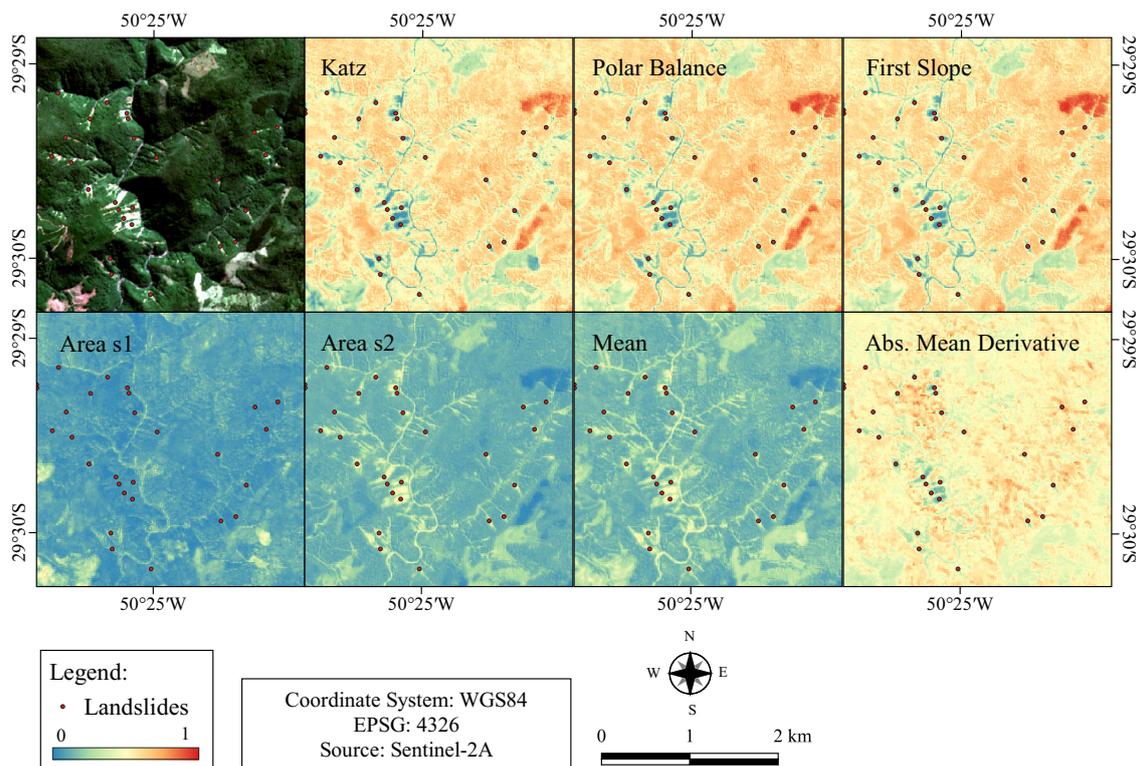


SOURCE: Author's production.

For analysing 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 attribute's importance ranking (see Sec. 5.1). 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.

The metrics in which Landslide scars stood out are represented in Figure 5.4. 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 0.0 (blue) for this class.

Figure 5.4 - NDBI metrics in details for Landslides scars.



Red points represent Landslide areas of reference.

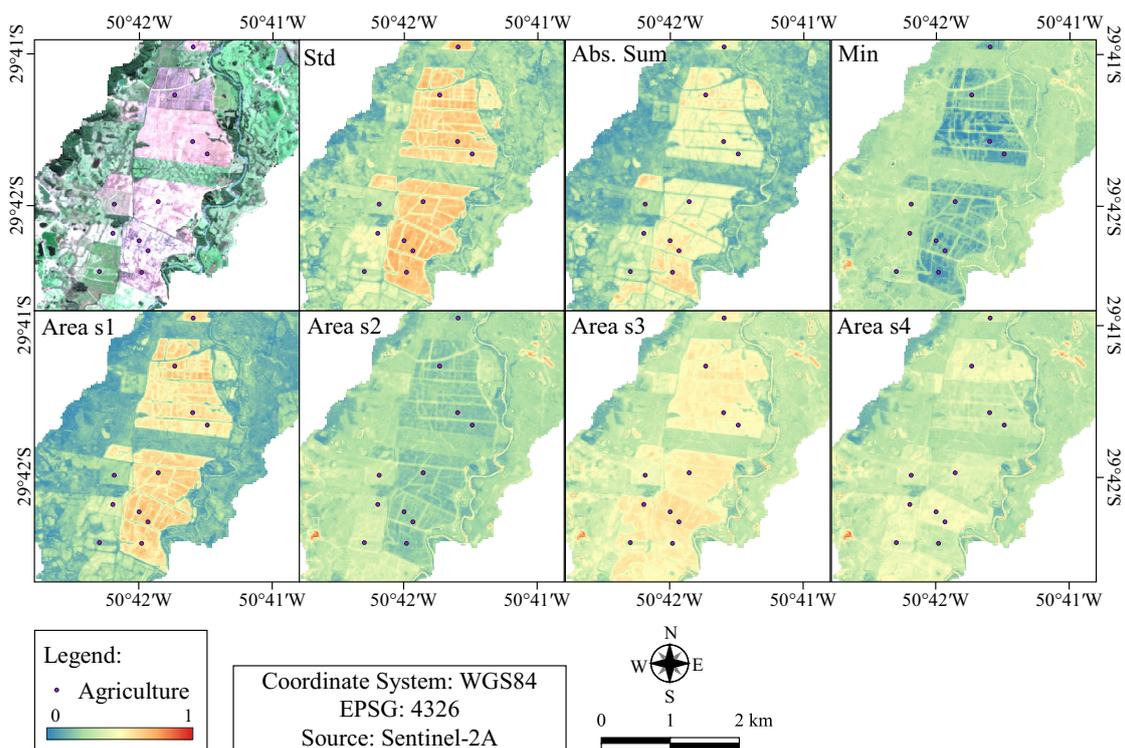
SOURCE: Author's production.

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

quite similar result to Area s2, where Landslides present a higher value (yellow), comparing to the rest of the image, mostly with low values (blue). 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.

The metrics that presented the most relevant results for Agriculture are exposed in Figure 5.5. 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 comparing 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 periods.

Figure 5.5 - NDBI metrics in details for Agriculture class.



Purple points represent Agriculture areas of reference.

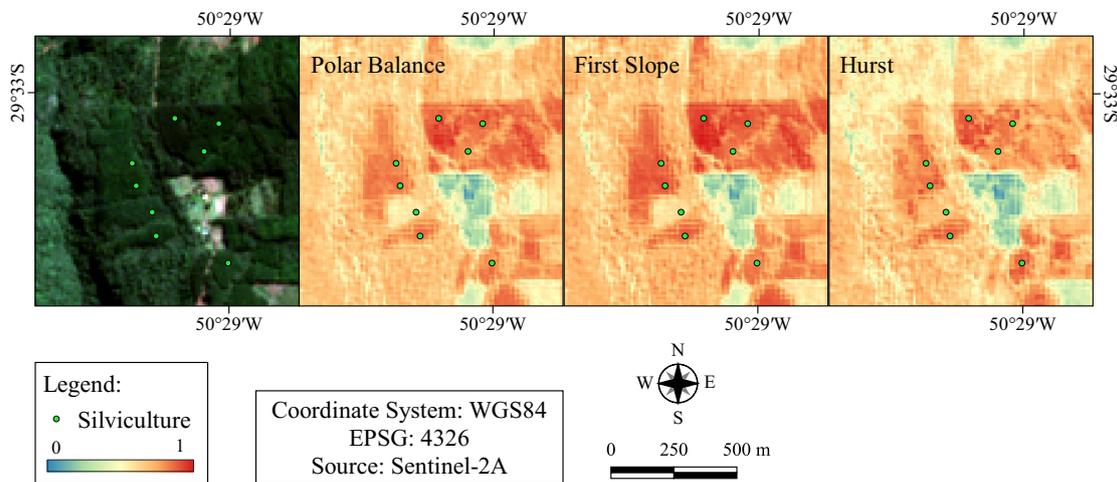
SOURCE: Author's production.

Finally, Areas s1 to s4, confirm the result presented by the Std metric, in which the variability seems to be a key characteristic for Agriculture. 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 metrics for Silviculture class are showed below (Figure 5.6). 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 comparing to its' surroundings. A contrast can be identified comparing to the Forest, which presents a lower response, but still in the orange/red range of values. This emphasises the contribution of polar, basic and fractal metrics for discriminating Silviculture from Forest.

Figure 5.6 - NDBI metrics in details for Silviculture class.



Green points represent Silviculture areas of reference.

SOURCE: Author's production.

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 exam-

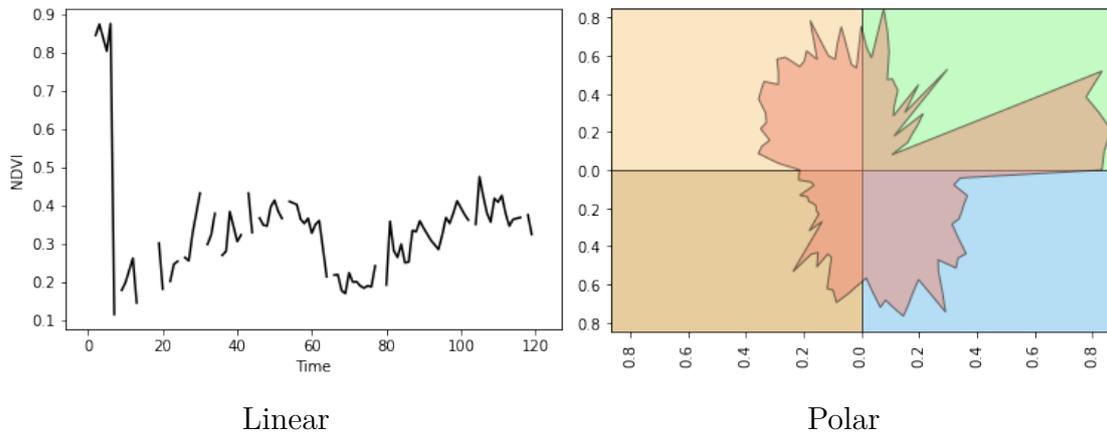
ple, classes Landslide (Figure 5.7), Forest (Figure 5.8), Silviculture (Figure 5.9), and Agriculture (Figure 5.10) are exposed below. The linear graphs are composed by the X axis, representing the time (each date of the time series from 2015 to 2020), and the Y axis, representing the NDVI values. The polar graph is divided in 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 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.

Analysing the Landslide graphs (Figure 5.7), one can notice that there are high values of NDVI in the beginning of the time series, followed by an abrupt reduction. After that, two small cycles of growth and decrease are observed. This profile represents 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.

Forest graphs (Figure 5.8) 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. Very different from Landslide, the Forest polar graph illustrates a wide and continuous circle, emphasizing the stability of this class behavior along the time.

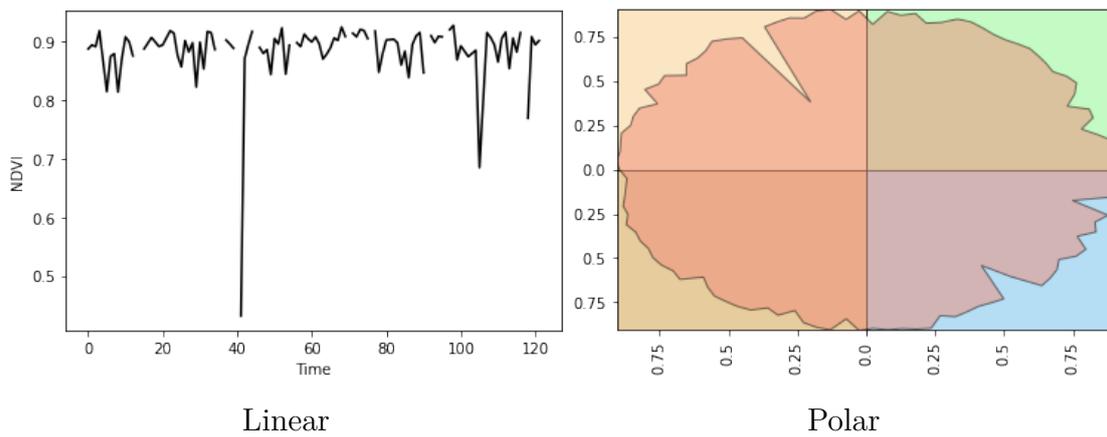
The Silviculture profile (Figure 5.9) is represented in the graphs by high, and two moments of very low NDVI values. This reveals the growing and harvesting periods. In the polar graph, the growing seasons are illustrated in the first and third rectangles, while the harvest in the second and fourth. Agriculture class (Figure 5.10)

Figure 5.7 - Landslides linear and polar graphs.



SOURCE: Author's production.

Figure 5.8 - Forest linear and polar graphs.

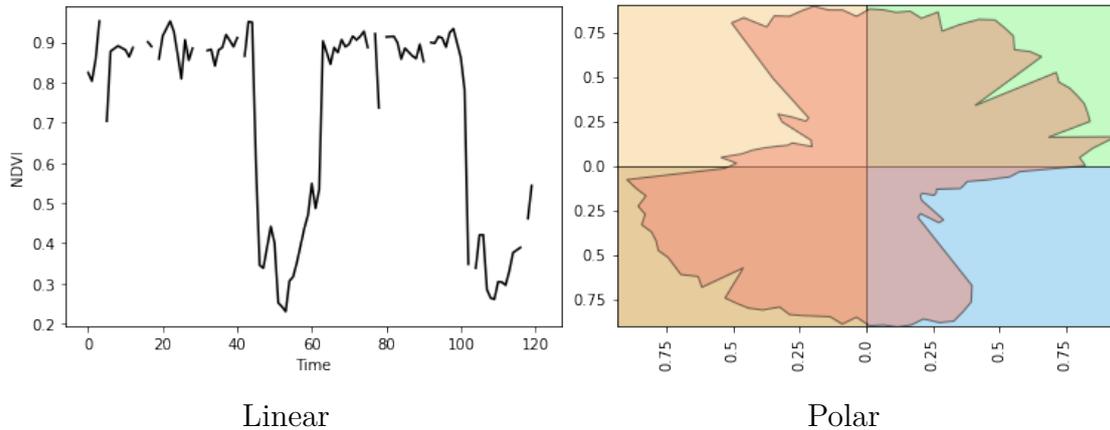


SOURCE: Author's production.

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

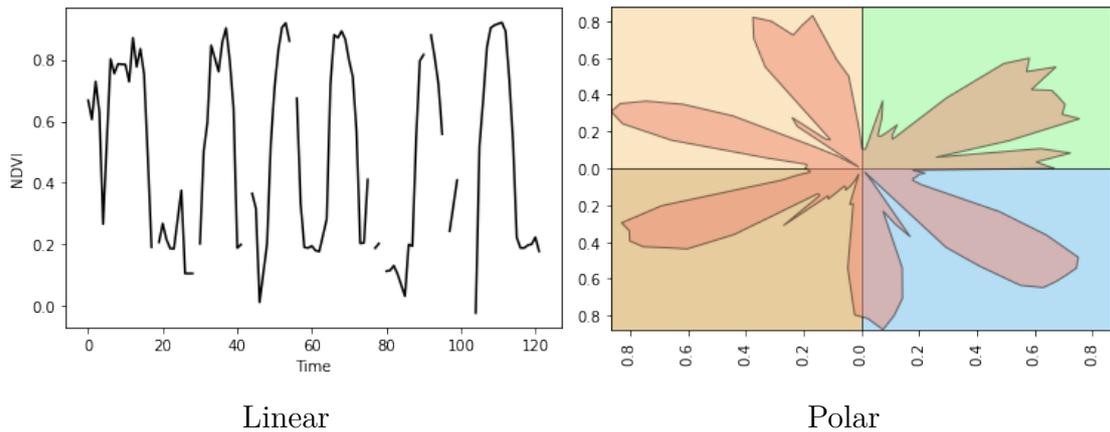
The metrics' contribution for classification was also interpreted via boxplot graphs. This graphs provide a variety of information, as the lower and upper quartiles (represented by the limits of the bounding box), the median (line inside the box), and

Figure 5.9 - Silviculture linear and polar graphs.



SOURCE: Author's production.

Figure 5.10 - Agriculture linear and polar graphs.



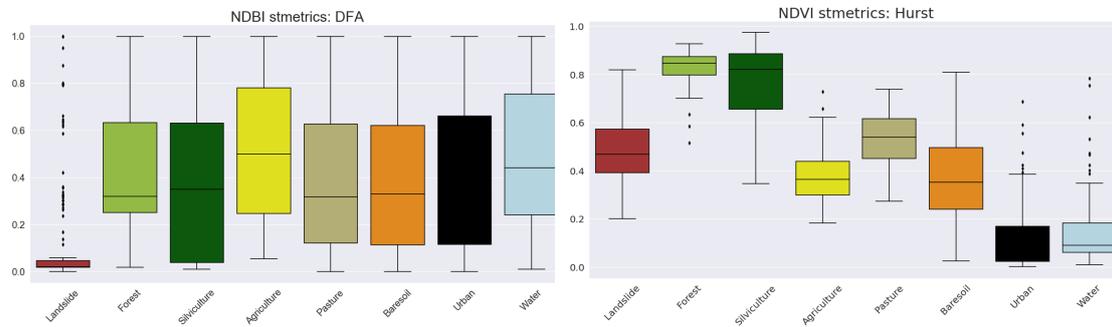
SOURCE: Author's production.

outliers (points). In order to understand if an 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 thresholds values to differ one class from another.

Regarding Landslides (Figure 5.11), the 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, while for the other classes, it is above 0.6. 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 NDVI's 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 were both of the type fractal.

Figure 5.11 - Landslides boxplot analysis.

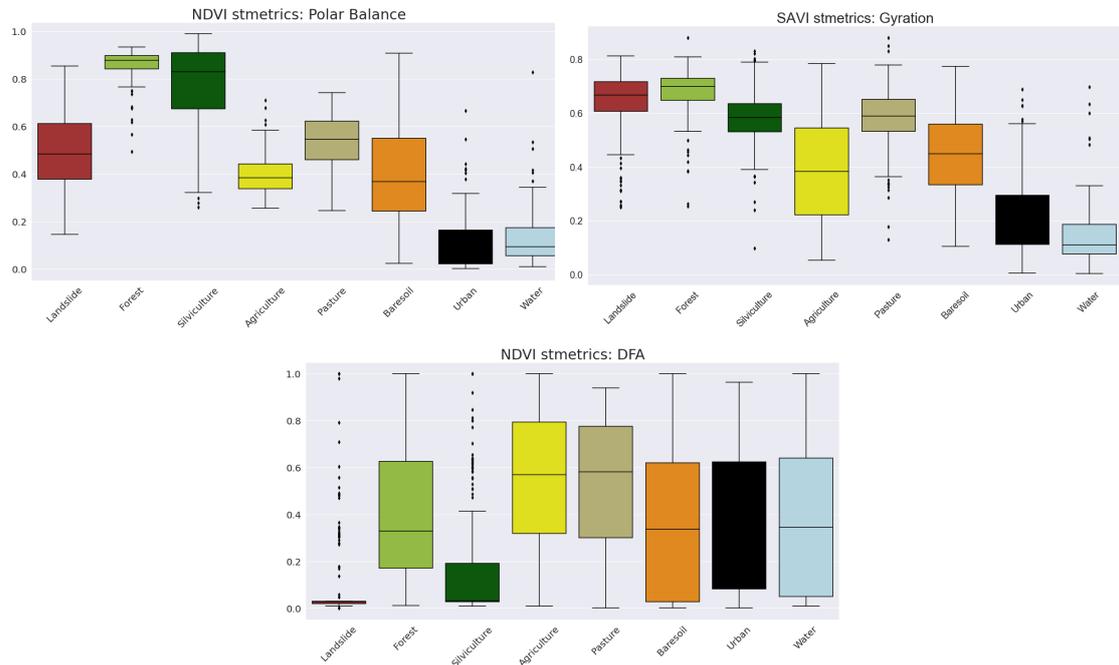


SOURCE: Author's production.

Forest and Silviculture (Figure 5.12) presented very similar results in most of the metrics. Both of them showed very high values for NDVI's Polar Balance, with the lowest quartile above 0.65, while 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 SAVI's Gyration, in which Forest's median is around 0.68 and Silviculture's, 0.58. Another method for differing Forest from Silviculture is by using the NDVI's 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 (Figure 5.13). The metrics that better revealed this differentiation were NDBI's and SAVI's Area s1, NDBI's and SAVI's Std, and SAVI's Absolute Mean Derivative. In all five metrics demonstrated below, the Agriculture class shows median values at least 0.1 higher than the other classes. Moreover, in all cases its' lower quartile is

Figure 5.12 - Forest and Silviculture boxplot analysis.



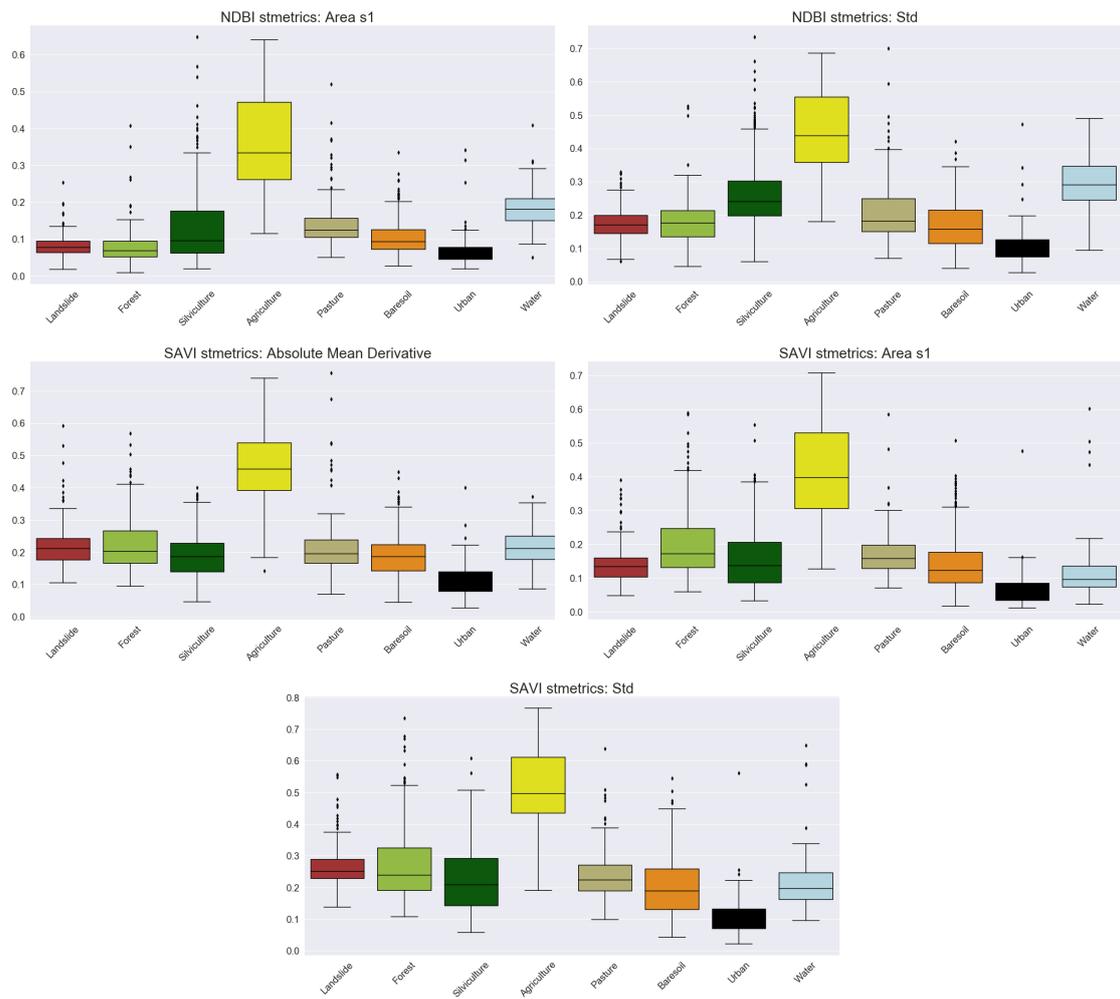
SOURCE: Author's production.

higher than the other's upper quartile. Finally, in SAVI's Absolute Mean Derivative metric its' lower quartile is above most of the other classes maximum range value. Beside Area s1, which is polar, all other metrics are the type basic.

Regarding the Urban class (Figure 5.14), the only metric that out-stands it is the NDBI's Area s2, with lower quartile value around 0.52, while the others' upper quartile are all below. However, another strategy could be to separate Urban and Water from the rest, by setting a threshold for values below 0.62 on NDVI's Max, or higher than 0.5 for NDBI's Area s2. After that, the difference between both classes could be solved by the NDVI's Std, where Urban and Water median values present around 0.3 of difference.

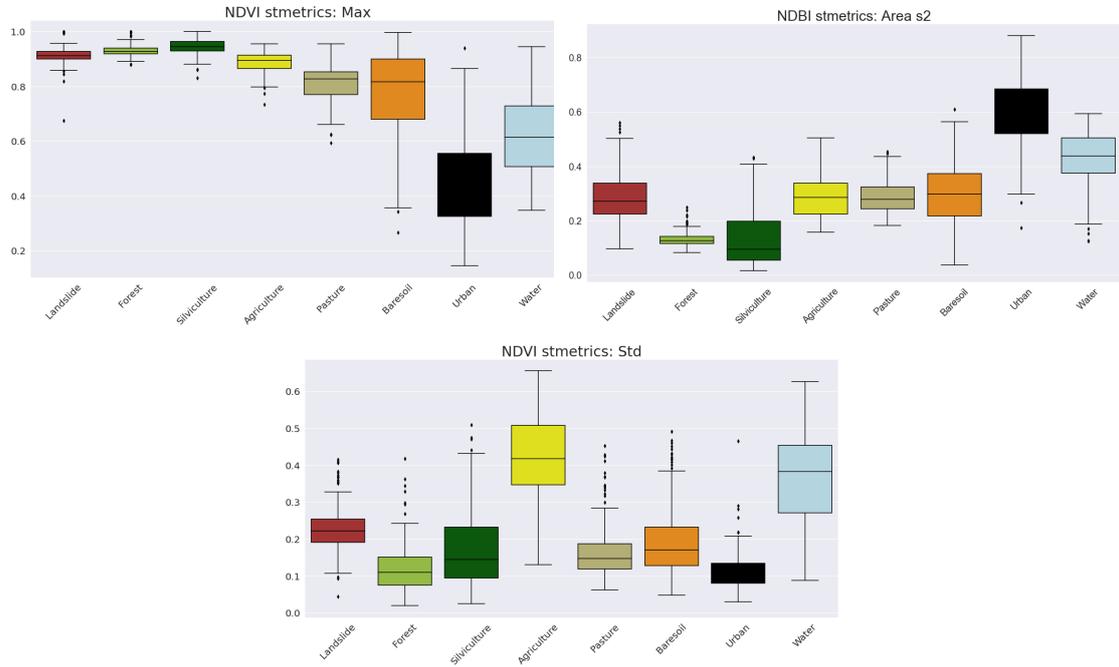
For Water class (Figure 5.15), no metric could provide an outstanding result for this class uniquely. However, the NDVI's Absolute Mean Derivative could be used to set a threshold separating Agriculture and Water from the other classes, by selecting only values above 0.3. Then, NDVI's Katz or Polar Balance could be used to differ between the two of them. Besides Urban and Water present very similar results for this two last metrics, the Urban class would not be included in the first threshold defined by the Absolute Mean Derivative, with upper quartile reaching up to 0.22.

Figure 5.13 - Agriculture boxplot analysis.



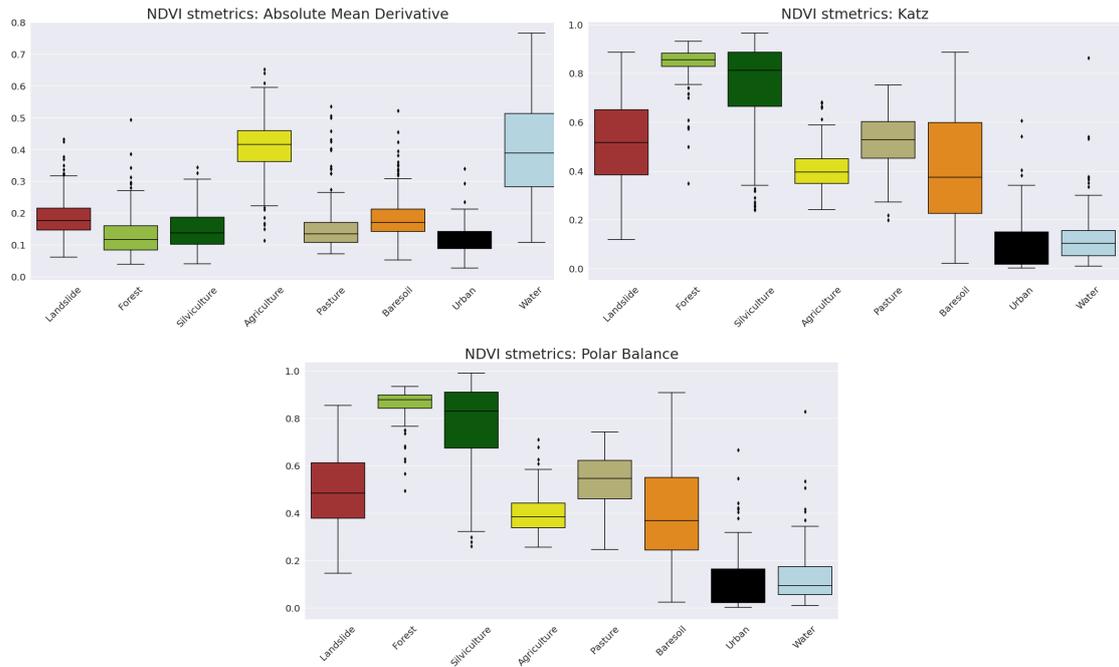
SOURCE: Author's production.

Figure 5.14 - Urban boxplot analysis.



SOURCE: Author's production.

Figure 5.15 - Water boxplot analysis.



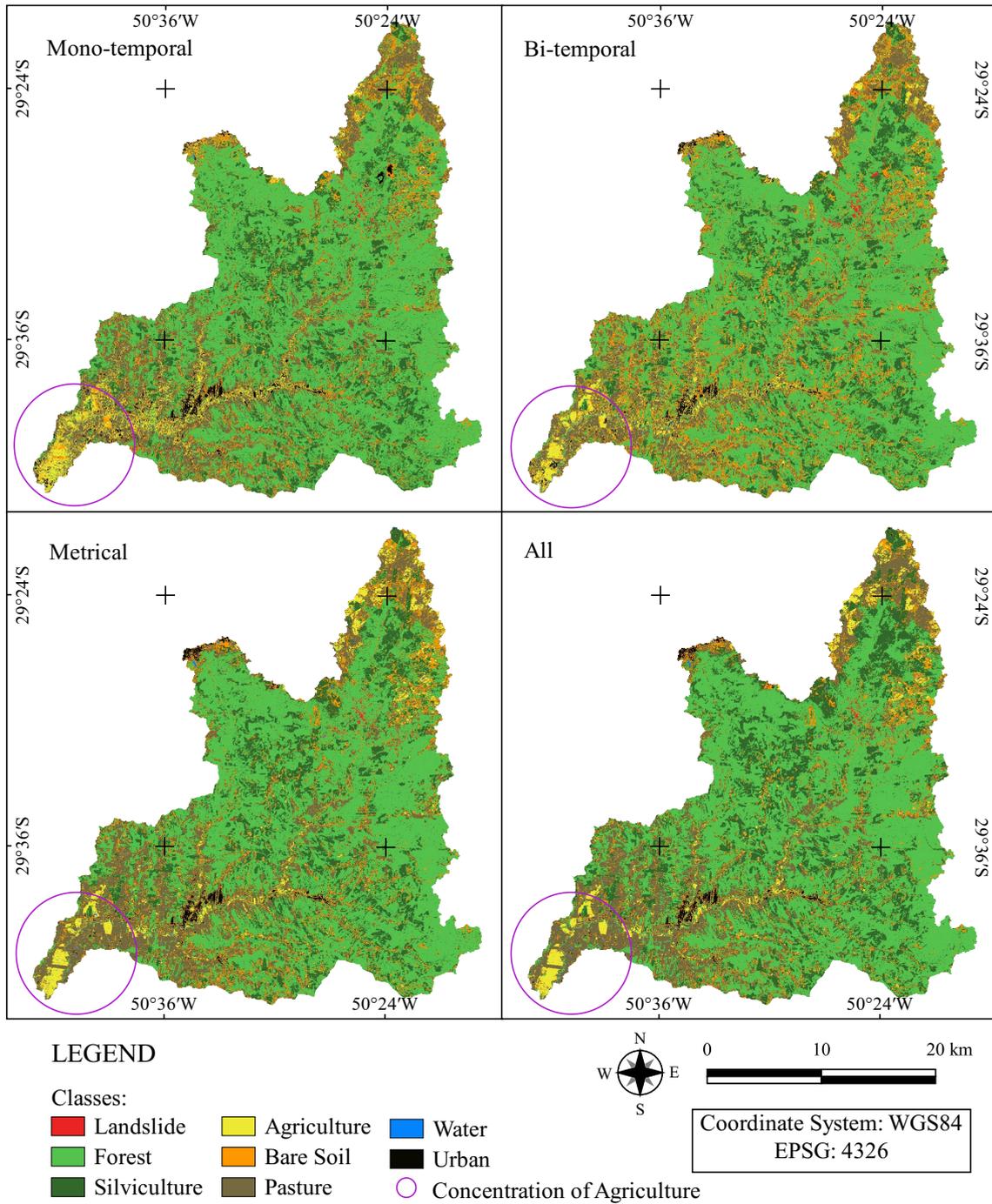
SOURCE: Author's production.

### 5.3 Land Use and Land Cover maps

Figure 5.16 shows the LULC classification products derived from each approach. Once the area is significantly large, this map can only provide an overview of the classification. However, more detailed visual comparison is available further in this chapter. Moreover, each LULC product as a separate map is provided in section Appendix ( Figure A.1, A.2, A.3 and A.4). Even though, no deep consideration can be done due to the scale of the maps, a simple comparison regarding the approaches is possible through this figure. In this visual analysis, the salt-and-pepper effect appears more intensively at the Mono-temporal map, and get reduced as the number of attributes increases. This can be seen more effectively at the south-west 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, whilst All approach shows more limited and well designed polygons of agriculture.

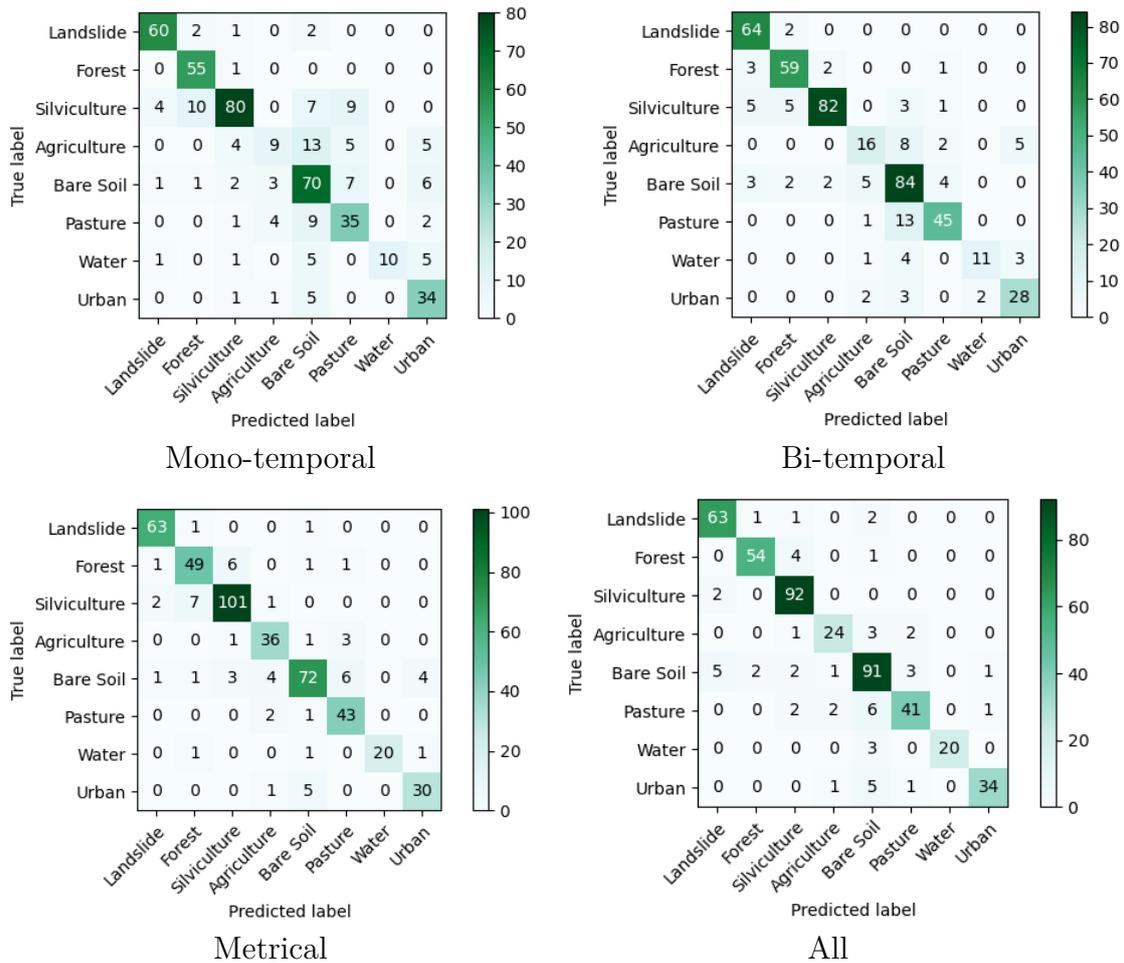
The confusion matrix is an effective tool to analyse the misclassifications between classes. It allows the calculation of a variety of statistical indices, which contribute to the interpretation of the classification algorithm performance. The confusion matrix for each approach is presented in Figure 5.17. 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, while in the others it no longer appears. Another class that requires attention is the Agriculture, which showed high values of confusion, specially 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 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.

Figure 5.16 - Classification maps.



SOURCE: Author's production.

Figure 5.17 - Confusion matrices.



The diagonal shows the correctly classified pixels, the stronger the green tone color, the higher is the number of correctly classified pixels.

SOURCE: Author's production.

As stated in Section 4.5, each classification approach was conducted under 500 Monte Carlo simulations, in order to guarantee an accuracy value between a confidence interval. This confidence intervals assure that the accuracy values present significant differences (or not), when comparing different approaches. Table 5.2 shows kappa indices, OA values, and their associated confidence interval.

Table 5.2 - Kappa index and Overall Accuracy.

	<i>Kappa (<math>\kappa</math>)</i>	<i>OA</i>	<i>Confidence Interval for OA</i>
Mono-temporal	0.7043	74.95%	74.78% - 75.11%
Bi-temporal	0.7938	82.59%	82.50% - 82.79%
Metrical	0.8574	87.90%	87.84% - 88.09%
All	0.8693	88.96%	88.81% - 89.05%

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 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.

Another important factor to analyse a classification accuracy, is by the evaluation of the commission and omission errors. In that sense, Tables 5.3 and 5.4 shows this statistical products depicted for each separate class and approach.

Table 5.3 - Omission errors.

	<i>Mono-temporal</i>	<i>Bi-temporal</i>	<i>Metrical</i>	<i>All</i>
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%

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 than Metrical for this class. Thus, Metrical had the best results concerning omission errors, with lowest error rates in five from the eight classes. Furthermore, even though Mono and Bi-temporal presented very satisfactory results for 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.

Table 5.4 - Commission errors.

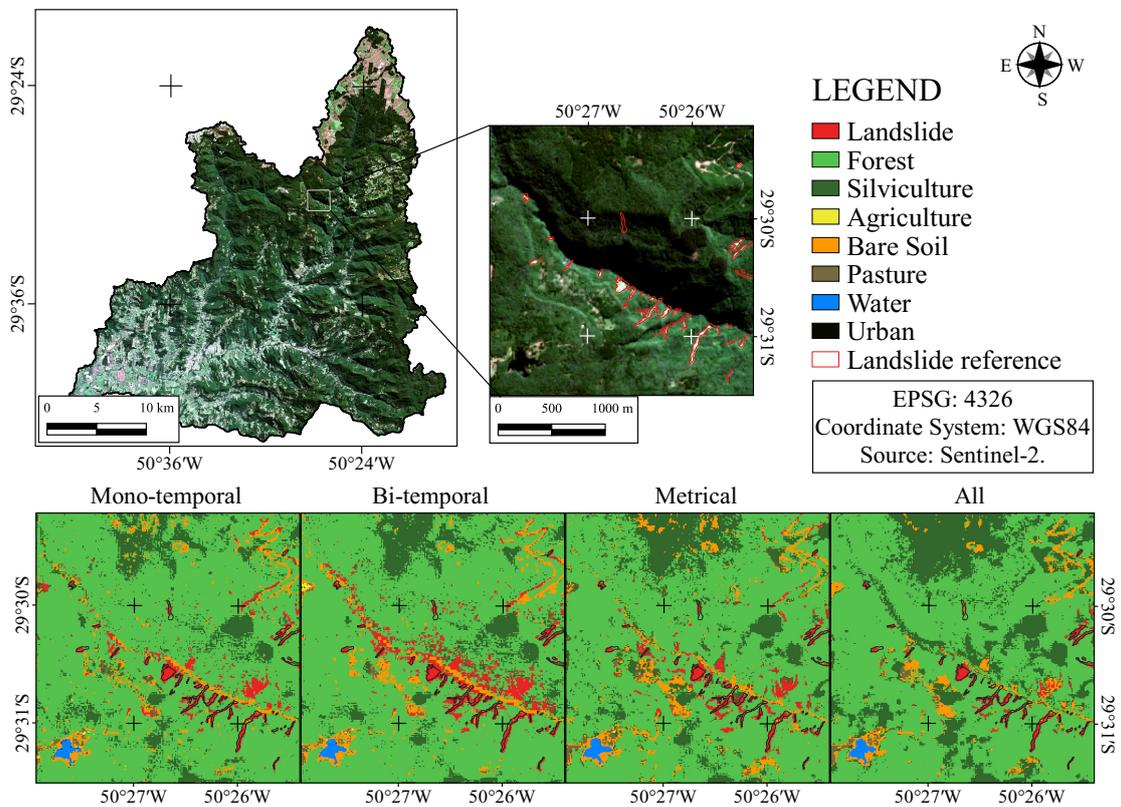
	<i>Mono-temporal</i>	<i>Bi-temporal</i>	<i>Metrical</i>	<i>All</i>
Landslide	9.09%	14.67%	5.97%	10.00%
Forest	19.12%	13.24%	16.95%	5.26%
Silviculture	12.09%	4.65%	9.01%	9.80%
Agriculture	47.06%	36.00%	18.18%	14.29%
Bare Soil	36.94%	26.96%	12.20%	18.02%
Pasture	37.50%	15.09%	18.87%	12.77%
Water	0.00%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

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 lowest error percentages in five from the eight classes. Moreover, as in omission error, despite Mono and Bi-temporal presented 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.

Figure 5.18 depicts an example of this commission error comparison concerning the Landslide class. 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 forest area from the hillside with Landslides. In this specific area, visually analysing, All approach performed better, presenting well defined polygons (low salt-and-pepper effect), and very satisfactory detection of the landslide scars, compared to the inventory reference.

Figure 5.18 - Landslides in details.

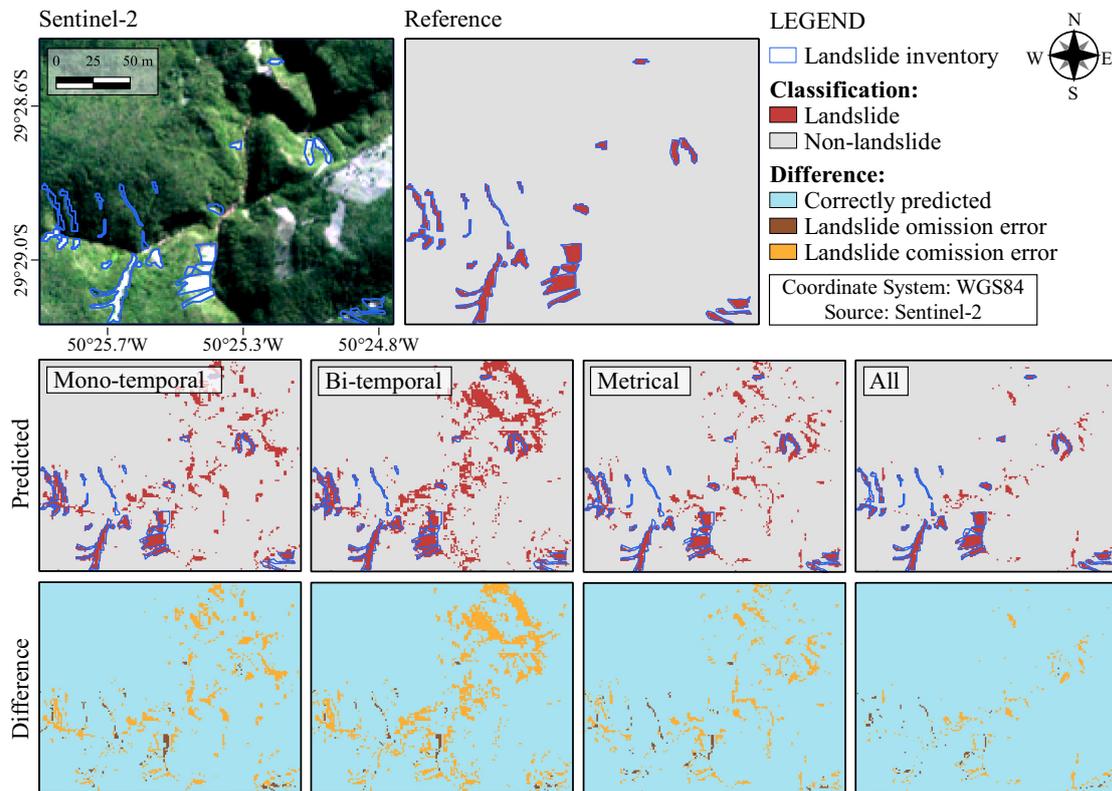


SOURCE: Author's production.

The difference map is another method used to analyse the classification performance regarding the Landslide scars. In Figure 5.19, 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 (Reference), in which, every pixel inside the polygons

are represented in red as Landslide class, and everything else is labeled as Non-Landslide and represented in grey. The first four maps below show the classification result for each approach. To clarify, in this case, the classification map from Figure 5.16 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, while brown and orange represent omission and commission classification errors, respectively. The inventory maps extracted from each approach are in the section Appendix, Figure A.5.

Figure 5.19 - Landslide difference map.



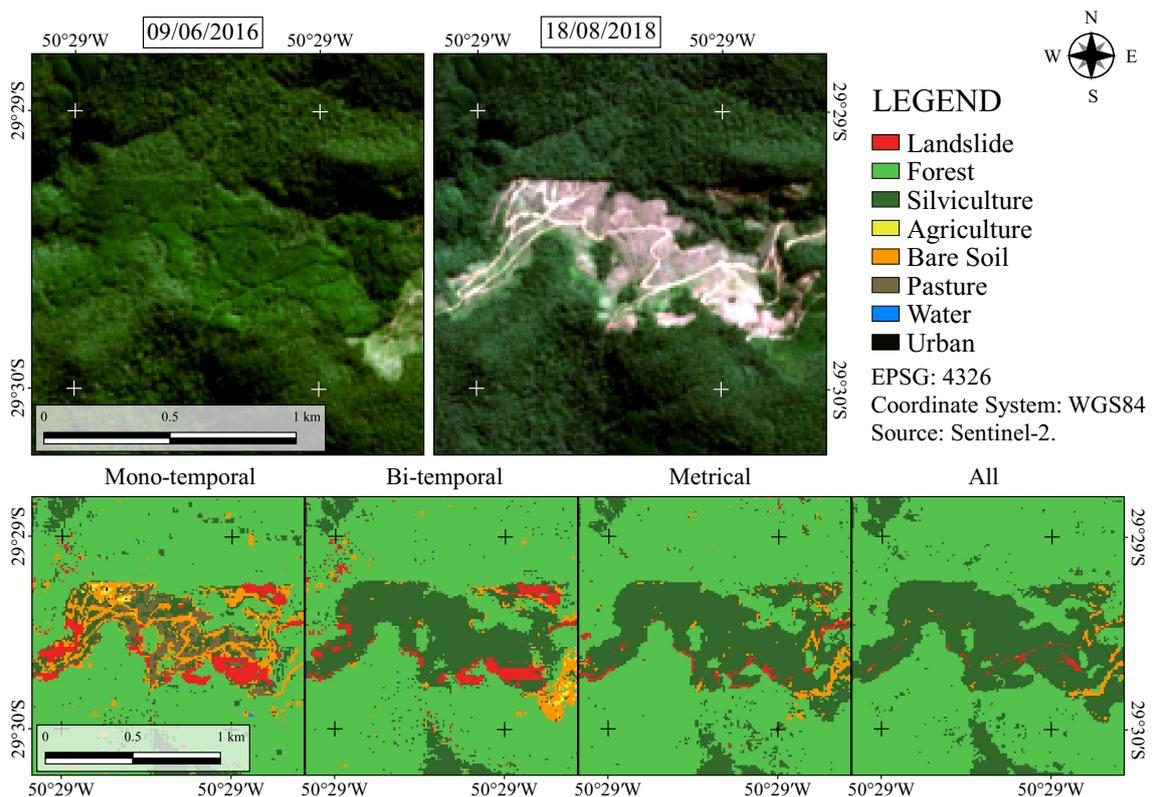
SOURCE: Author's production.

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 (Figure 5.17) reveal that Landslides showed confusion with Silviculture for the four approaches, especially for Mono and Bi-temporal. This analysis can be developed in Figure 5.20. The two first Sentinel images represent an area only composed by Forest, containing Silviculture in the central part, and no Landslides scars. Silviculture class is clearly apparent in the 2018's image, where it has been harvested. Besides, its' texture in 2016's image is also notably different from the surrounding Forest. The classification results are displayed on the four maps below.

Figure 5.20 - Landslide vs Silviculture.



SOURCE: Author's production.

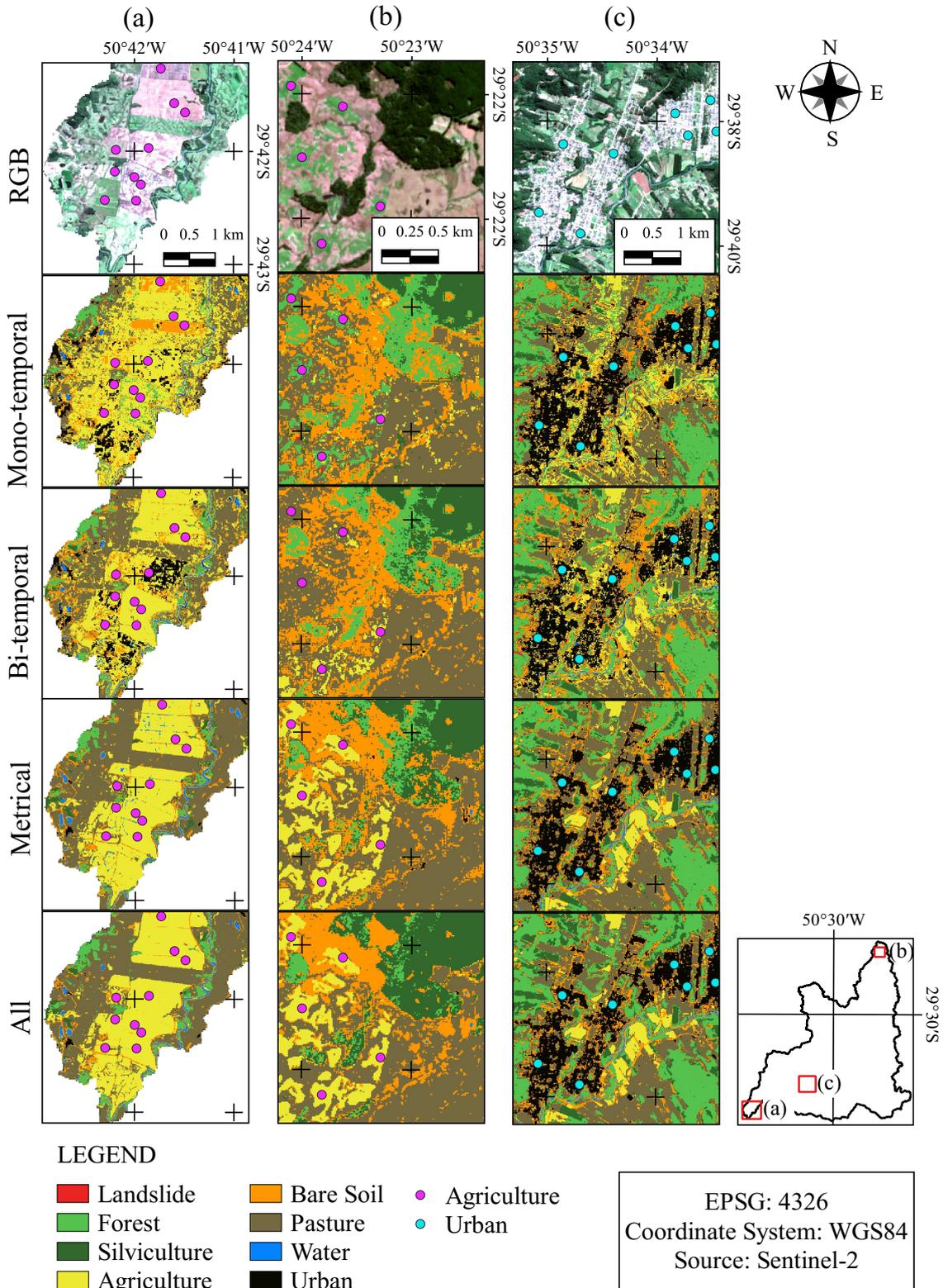
Mono-temporal approach presented the least satisfactory output, with most of the Silviculture area classified as Pasture, Bare Soil and Landslides scars. Also, the Bi-

temporal approach wrongly classified a considerable amount of Silviculture pixels as Landslide. Metrical and All presented the best results, with the most part correctly predicted as Silviculture. This results reveal, again, that one of the most challenging obstacles to landslide detection is to reduce the commission errors. Once the Mono-temporal approach did 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 presented in Figure 5.21. 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, while “c”, Urban. Each row shows the result for an specific approach. The pink points represent Agriculture areas of reference, while 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. Besides, both of these approaches also misclassified Agriculture pixels with Urban, which do 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 present 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.

Figure 5.21 - Classification maps in details.



Figures (a) and (b) show examples of agricultural areas. Figure (c) illustrates the 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).

SOURCE: Author's production.

## 6 CONCLUSION

The major aim of this research was to develop and describe a methodology that provides, with the highest accuracy possible, two main products to landslide susceptibility assessment through a semi-automatic process; LULC and landslide inventory maps. Even though free access medium resolution (Landsat-like) 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 this process is time-consuming and manually realized, but usually requires updated images from (very-)high spatial resolution, which, in general, have 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 (AI) era, reducing the distance among conventional and state-of-art methods. Nowadays, the world shows that data availability is no longer a limitation, once many types of data are massively provided from countless sources everyday. The main challenge is how to manage and transform data into information. In this sense, the 21<sup>st</sup> century has been evolving towards 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 on the use of state-of-art methods and data to generate landslide susceptibility assessment supply’s materials. The methodology here proposed uses geomorphometric attributes, dense and irregular time series of different spectral indices from remote sensing imagery and three modules of time series metrics (basic, polar and fractal), applied to RF classification algorithm, all developed in a free and open source way. Analysing the attributes importance ranking, among the four input approaches compared (Mono-temporal, Bi-temporal, Metrical and All), each of them ranked a NDBI image or metric extracted from this index as the most important attribute to the classification. The NDVI or attributes related to that index also presented high relevance. The geomorphometric attributes, specially the Slope, was present in all of the approaches among the top 6 in the importance list.

Regarding the time series metrics, polar metrics showed high importance positions in the rank. For Metrical approach, the first attribute chosen was the NDBI’s Area s2, and NDVI’s Area s4 is among the top 5. Moreover, NDBI’s Polar Balance and NDVI’s 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, NDBI’s Area s1 and

s2. Moreover, the visual analysis revealed that specially for Landslides, Agriculture and Silviculture, some metrics could support their detection, as, for example, Katz, Std, and First Slope, respectively. Besides, 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.

Comparing the approaches performances through the overall accuracy 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 (10.00%), 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 overall accuracy and omission error; and the best result for commission error.

This research answered some questions on how time series and time series metrics could contribute to LULC mapping and landslide detection. Moreover, the importance ranking may support on the decision of what attributes 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 medium spatial resolution in a semi-automatic procedure, in order 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 on 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.

Nonetheless, field work and visual interpretation from specialists is 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 attributes and other regions with more urban areas. Moreover, for a more complete analysis of the classification results, uncertainty maps are highly recommended. Besides, the aggregation of other type of orbital data despite the Sentinel-2 should be tested, for example CBERS-4 and 4A, Sentinel-1 (SAR) and Landsat-8.



## REFERENCES

AUGUSTO FILHO, O. **Escorregamentos em encostas naturais e ocupadas: análise e controle**. São Paulo: ABGE/IPT, 1995. 7

AYELE, G. T.; TEBEJE, A. K.; DEMISSIE, S. S.; BELETE, M. A.; JEMBERRIE, M. A.; TESHOME, W. M.; MENGISTU, D. T.; TESHALE, E. Z. Time series land cover mapping and change detection analysis using geographic information system and remote sensing, northern ethiopia. **Air, Soil and Water Research**, v. 11, p. 1178622117751603, 2018. 12

BEGUERÍA, S. Changes in land cover and shallow landslide activity: a case study in the spanish pyrenees. **Geomorphology**, v. 74, n. 1-4, p. 196–206, 2006. 1

BELGIU, M.; DRĂGUȚ, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016. 18

BELLÓN, B.; BÉGUÉ, A.; SEEN, D. L.; ALMEIDA, C. A. D.; SIMÕES, M. A remote sensing approach for regional-scale mapping of agricultural land-use systems based on ndvi time series. **Remote Sensing**, v. 9, n. 6, p. 600, 2017. 12

BENDINI, H. do N.; FONSECA, L. M. G.; SCHWIEDER, M.; KÖRTING, T. S.; RUFIN, P.; SANCHES, I. D. A.; LEITAO, P. J.; HOSTERT, P. Detailed agricultural land classification in the brazilian cerrado based on phenological information from dense satellite image time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 82, p. 101872, 2019. 5

BEVEN, K. J.; KIRKBY, M. J. A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. **Hydrological Sciences Journal**, v. 24, n. 1, p. 43–69, 1979. 32

BHANDARI, A.; KUMAR, A.; SINGH, G. Feature extraction using normalized difference vegetation index (ndvi): a case study of jabalpur city. **Procedia Technology**, v. 6, p. 612–621, 2012. 30

BIRKMANN, J. Indicators and criteria for measuring vulnerability: Theoretical bases and requirements. In: BIRKMANN, J. (Ed.). **Measuring vulnerability to natural hazards: towards disaster resilient societies**. Tokyo: United Nations University Press, 2006. p. 55–77. 1

- BLAHUT, J.; WESTEN, C. J. V.; STERLACCHINI, S. Analysis of landslide inventories for accurate prediction of debris-flow source areas. **Geomorphology**, v. 119, n. 1-2, p. 36–51, 2010. 1
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. 18
- BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. [S.l.]: Springer, 2016. 11
- BUSCH, A.; AMORIM, S. N. D. **A tragédia da região serrana do Rio de Janeiro em 2011: procurando respostas**. Brasília: Escola Nacional de Administração Pública (ENAP), 2011. 10
- CARDOZO, C. P. **Spatially integrated modelling approach to landslide risk assessment: a case study of the Nova Friburgo disaster - RJ, Brazil**. Tese (Doutorado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2018. 1, 10
- CENTRO NACIONAL DE MONITORAMENTO E ALERTAS DE DESASTRES NATURAIS - CEMADEN. **Histórico da criação do CEMADEN**. 2016. Available from: <http://www.cemaden.gov.br/historico-da-criacao-do-cemaden/>. 10
- CHANG, Z.; DU, Z.; ZHANG, F.; HUANG, F.; CHEN, J.; LI, W.; GUO, Z. Landslide susceptibility prediction based on remote sensing images and gis: comparisons of supervised and unsupervised machine learning models. **Remote Sensing**, v. 12, n. 3, p. 502, 2020. 3
- CHEN, F.; YU, B.; LI, B. 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**, v. 15, n. 3, p. 453–464, 2018. 3
- CHEN, T.; TRINDER, J. C.; NIU, R. Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China. **Remote Sensing**, v. 9, n. 4, p. 333, 2017. 18
- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960. 42
- CONGALTON, R. G. et al. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, New York, v. 37, n. 1, p. 35–46, 1991. 35

DAI, K.; LI, Z.; TOMÁS, R.; LIU, G.; YU, B.; WANG, X.; CHENG, H.; CHEN, J.; STOCKAMP, J. Monitoring activity at the Daguangbao mega-landslide (China) using Sentinel-1 TOPS time series interferometry. **Remote Sensing of Environment**, v. 186, p. 501–513, 2016. 12

DASH, M.; LIU, H. Feature selection for classification. **Intelligent Data Analysis**, v. 1, n. 3, p. 131–156, 1997. 17

DAVIES, T. Landslide Hazards, Risks, and Disasters: Introduction. In: SHRODER J. F.; DAVIES, T. (Ed.). **Landslide hazards, risks and disasters**. [S.l.: s.n.], 2015. p. 1–16. 7

DIAS, M. **Salvador**. PORTALNOAR, 2015. Available from: <<https://portalnoar.com.br/com-chuva-forte-salvador-registra-novos-deslizamentos/>>. Access in: 11 feb. 2020. 9

EDSALL, R.; MACEACHREN, A.; PEUQUET, D. Assessing the effectiveness of temporal legends in environmental visualization. In: GIS/LIS, 01., 1997. **Proceedings...** [S.l.], 1997. p. 28–30. 14

EKLUNDH, L.; JÖNSSON, P. Timesat 3.2 with parallel processing-software manual. **Lund University**, v. 52, 2012. 3, 11, 12

EUROPEAN SPATIAL AGENCY (ESA). **ESA sentinel online**: Multispectral instrument (msi) overview. 2015. Available from: <<https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>>. Access in: fev 21 2020. 20, 22

FAYYAD, U. M. et al. Knowledge discovery and data mining: towards a unifying framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. **Proceedings...** [S.l.], 1996. v. 96, p. 82–88. 15, 16

FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. **Remote Sensing of Environment**, v. 61, n. 3, p. 399–409, 1997. 3

GAMEIRO, S.; QUEVEDO, R. P.; OLIVEIRA, G.; RUIZ, L.; GUASSELLI, L. Análise e correlação de atributos morfológicos e sua influência nos movimentos de massa ocorridos na Bacia do Rio Rolante, RS. In: SIMPÓSIO BRASILEIRO

DE SENSORIAMENTO REMOTO, XIX, 19., 2019, Santos. **Anais...** São José dos Campos: INPE, 2019. p. 2880–2883. 1, 26

GARIANO, S. L.; PETRUCCI, O.; GUZZETTI, F. The role of rainfall and land use/cover changes in landslide occurrence in Calabria, Southern Italy, in the 20th Century. In: WORKSHOP ON WORLD LANDSLIDE FORUM. **Proceedings...** [S.l.]: Springer, 2017. p. 339–345. 1

GERENTE, J.; PLETSCHE, M. A. J. S.; SÖTHER, C.; FRANCISCO, C. N. Classificação semiautomática de cicatrizes de movimentos de massa por meio de técnicas de detecção de mudanças. **Revista Brasileira de Geomorfologia**, v. 18, n. 4, 2017. 12, 30

GERENTE, J.; SÖTHER, C.; NEGRÃO, P.; KÖRTING, T. S. Mass movements' scars classification using data mining techniques. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos, Brazil. **Anais...** São José dos Campos: INPE, 2017. p. 3553–3560. 30

GHORBANZADEH, O.; BLASCHKE, T.; GHOLAMNIA, K.; MEENA, S. R.; TIEDE, D.; ARYAL, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. **Remote Sensing**, v. 11, n. 2, p. 196, 2019. 3, 4

GHOSH, A.; FASSNACHT, F. E.; JOSHI, P. K.; KOCH, B. A framework for mapping tree species combining hyperspectral and lidar data: role of selected classifiers and sensor across three spatial scales. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, p. 49–63, 2014. 18

GLADE, T. Landslide occurrence as a response to land use change: a review of evidence from New Zealand. **Catena**, v. 51, n. 3-4, p. 297–314, 2003. 1

GÓMEZ, C.; WHITE, J. C.; WULDER, M. A. Optical remotely sensed time series data for land cover classification: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 116, p. 55–72, 2016. 2

GORSEVSKI, P. V.; BROWN, M. K.; PANTER, K.; ONASCH, C. M.; SIMIC, A.; SNYDER, J. Landslide detection and susceptibility mapping using lidar and an artificial neural network approach: a case study in the cuyahoga valley national park, ohio. **Landslides**, v. 13, n. 3, p. 467–484, 2016. 4

GUAN, H.; LI, J.; CHAPMAN, M.; DENG, F.; JI, Z.; YANG, X. Integration of orthoimagery and lidar data for object-based urban thematic mapping using

random forests. **International Journal of Remote Sensing**, v. 34, n. 14, p. 5166–5186, 2013. 18

GUZZETTI, F.; ARDIZZONE, F.; CARDINALI, M.; GALLI, M.; REICHENBACH, P.; ROSSI, M. Distribution of landslides in the Upper Tiber River basin, central Italy. **Geomorphology**, v. 96, n. 1-2, p. 105–122, 2008. 1

GUZZETTI, F.; MONDINI, A. C.; CARDINALI, M.; FIORUCCI, F.; SANTANGELO, M.; CHANG, K.-T. Landslide inventory maps: new tools for an old problem. **Earth-Science Reviews**, v. 112, n. 1-2, p. 42–66, 2012. ISSN 0012-8252. 2, 7

HALL, M. A. **Correlation-based feature selection for machine learning**. Thesis (PhD) — University of Waikato, Hamilton, 1999. 17

HSU, P.-H. Feature extraction of hyperspectral images using wavelet and matching pursuit. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 62, n. 2, p. 78 – 92, 2007. ISSN 0924-2716. Available from: <<http://www.sciencedirect.com/science/article/pii/S0924271607000020>>. 17

HU, X.; WANG, T.; PIERSON, T. C.; LU, Z.; KIM, J.; CECERE, T. H. Detecting seasonal landslide movement within the Cascade landslide complex (Washington) using time-series SAR imagery. **Remote Sensing of Environment**, v. 187, p. 49–61, 2016. 12

HUANG, Q.; WANG, C.; MENG, Y.; CHEN, J.; YUE, A. Landslide monitoring using change detection in multitemporal optical imagery. **IEEE Geoscience and Remote Sensing Letters**, v. 17, n. 2, p. 1–5, 2019. ISSN 1545-598X. 13

HUETE, A. Huete, ar a soil-adjusted vegetation index (savi). remote sensing of environment. **Remote Sensing of Environment**, v. 25, p. 295–309, 1988. 31

HUGHES, G. On the mean accuracy of statistical pattern recognizers. **IEEE Transactions on Information Theory**, v. 14, n. 1, p. 55–63, 1968. 17

INSTITUTO DE PESQUISAS TECNOLÓGICAS. **Levantamento do IPT de mortes causadas por deslizamentos é tema de reportagem no UOL e Jornal Nacional**. 2017. Available from:

<[http://www.ipt.br/noticias\\_interna.php?id\\_noticia=1190](http://www.ipt.br/noticias_interna.php?id_noticia=1190)>. Access in: may 16 2021. 9

ISHAK, N. I. B. **Vegetation anomaly index from remote sensing for landslide activities mapping**. Thesis (PhD) — Universiti Teknologi Malaysia, 2018. 3

JEEVALAKSHMI, D.; REDDY, S. N.; MANIKIAM, B. Land cover classification based on NDVI using Landsat8 time series: a case study Tirupati region. In: INTERNATIONAL CONFERENCE ON COMMUNICATION AND SIGNAL PROCESSING, 2016. **Proceedings...** [S.l.]: IEEE, 2016. p. 1332–1335. 30

JENSEN, J. R. **Remote sensing of the environment: an earth resource perspective 2/e**. [S.l.]: Pearson Education India, 2009. 20

JIANG, Y.; LIAO, M.; ZHOU, Z.; SHI, X.; ZHANG, L.; BALZ, T. Landslide deformation analysis by coupling deformation time series from sar data with hydrological factors through data assimilation. **Remote Sensing**, v. 8, n. 3, p. 179, 2016. 12

JOYCE, K. E.; BELLISS, S. E.; SAMSONOV, S. V.; MCNEILL, S. J.; GLASSEY, P. J. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. **Progress in Physical Geography**, v. 33, n. 2, p. 183–207, 2009. 30

KENNIE, T.; PETRIE, G. Terrain modelling-application fields and terminology. In: PETRIE G.; KENNIE, T. (Ed.). **Terrain Modelling in Surveying and Civil Engineering**. [S.l.]: McGraw-Hill, 1990. p. 1–3. 22

KONG, F.; LI, X.; WANG, H.; XIE, D.; LI, X.; BAI, Y. Land cover classification based on fused data from gf-1 and modis ndvi time series. **Remote Sensing**, v. 8, n. 9, p. 741, 2016. 30

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. Geodma—geographic data mining analyst. **Computers & Geosciences**, v. 57, p. 133 – 145, 2013. ISSN 0098-3004. Available from: <<http://www.sciencedirect.com/science/article/pii/S0098300413000538>>. 3, 12, 13

KULKARNI, V. Y.; SINHA, P. K. Pruning of random forest classifiers: a survey and future directions. In: INTERNATIONAL CONFERENCE ON DATA SCIENCE & ENGINEERING, 2012. **Proceedings...** [S.l.]: IEEE, 2012. p. 64–68. 18

- LACROIX, P.; ARAUJO, G.; HOLLINGSWORTH, J.; TAIPE, E. Self-entrainment motion of a slow-moving landslide inferred from Landsat-8 time series. **Journal of Geophysical Research: Earth Surface**, v. 124, n. 5, p. 1201–1216, 2019. 12
- LAROSE, D. T.; LAROSE, C. D. **Discovering knowledge in data: an introduction to data mining**. [S.l.]: John Wiley & Sons, 2014. 16
- LAWRENCE, R. L.; RIPPLE, W. J. Comparisons among vegetation indices and bandwise regression in a highly disturbed, heterogeneous landscape: Mount St. Helens, Washington. **Remote Sensing of Environment**, v. 64, n. 1, p. 91–102, 1998. 31
- LI, Y.; HUANG, J.; PU, H.; HAN, Z.; LI, W.; YAN, B. Semiautomatic landslide detection using remote sensing and slope units. **Transportation Research Record**, v. 2604, n. 1, p. 104–110, 2017. 3
- LI, Z.; SHI, W.; MYINT, S. W.; LU, P.; WANG, Q. Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. **Remote Sensing of Environment**, v. 175, p. 215–230, 2016. 12
- LISTO, F. d. L. R.; VIEIRA, B. C. Mapping of risk and susceptibility of shallow-landslide in the city of São Paulo, Brazil. **Geomorphology**, v. 169, p. 30–44, 2012. 9
- LUERCE, T. D. **Geoturismo na bacia hidrográfica do Rio Rolante/RS: um estudo acerca das Quedas d'água**. 181 p. Dissertation (Master in Geography) — Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2012. 25, 35
- LUPIANO, V.; RAGO, V.; TERRANOVA, O. G.; IOVINE, G. Landslide inventory and main geomorphological features affecting slope stability in the Pícentino river basin (Campania, southern Italy). **Journal of Maps**, v. 15, n. 2, p. 131–141, jul 2019. ISSN 17445647. 1
- LV, Z. Y.; SHI, W.; ZHANG, X.; BENEDIKTSSON, J. A. Landslide inventory mapping from bitemporal high-resolution remote sensing images using change detection and multiscale segmentation. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 5, p. 1520–1532, 2018. 12

MANFRÉ, L. A.; SHINOHARA, E. J.; SILVA, J. B.; SIQUEIRA, R. N. D. P.; GIANNOTTI, M. A.; QUINTANILHA, J. Method for landslides identification at the São Paulo state coast, Brazil. **Geociências**, v. 33, n. 1, p. 172–180, 2014. 3

MENDES, R. M.; ANDRADE, M. R. M. de; TOMASELLA, J.; MORAES, M. A. E. de; SCOFIELD, G. B. Understanding shallow landslides in Campos do Jordão municipality - Brazil: disentangling the anthropic effects from natural causes in the disaster of 2000. **Natural Hazards & Earth System Sciences**, v. 18, n. 1, 2018. 8

MEZAAL, M. R.; PRADHAN, B. Data mining-aided automatic landslide detection using airborne laser scanning data in densely forested tropical areas. **Korean Journal of Remote Sensing**, , v. 34, n. 1, p. 45–74, 2018. 3

MONDINI, A. C.; SANTANGELO, M.; ROCCHETTI, M.; ROSSETTO, E.; MANCONI, A.; MONSERRAT, O. Sentinel-1 sar amplitude imagery for rapid landslide detection. **Remote Sensing**, v. 11, n. 7, p. 760, 2019. 4

MONTEIRO, A.; CARDOZO, C.; LOPES, E. Sentidos territoriais: a paisagem como mediação em novas abordagens 15 metodológicas para os estudos integrados em riscos de desastres. In: SIQUEIRA, A.; VALENCIO, N.; SIENA, M.; MALAGOLI, M. A. (Ed.). **Riscos de desastres relacionados à água**. São Paulo: Rima, 2015. 8

MOOSAVI, V.; TALEBI, A.; GEOMORPHOLOGY, B. S. Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. **Geomorphology**, v. 204, p. 646–656, 2014. Available from: <<https://www.sciencedirect.com/science/article/pii/S0169555X13004704>>. 3

NATHAN, F. Vulnerabilities to natural hazard: case study on landslide risks in la paz. In: WORLD INTERNATIONAL STUDIES CONFERENCE, 2005. **Proceedings...** [S.l.], 2005. 7

NOVO, E. Sensoriamento remoto: princípios e aplicações. 2.ed. **São Paulo: Edgard Blücher**, v. 308, 2010. 16

OKIN, G. S.; DONG, C.; WILLIS, K. S.; GILLESPIE, T. W.; MACDONALD, G. M. The impact of drought on native southern california vegetation: remote sensing analysis using modis-derived time series. **Journal of Geophysical Research: Biogeosciences**, v. 123, n. 6, p. 1927–1939, 2018. 12

OKSANEN, J. et al. **Digital elevation model error in terrain analysis**. [S.l.]: Finnish Geodetic Institute, 2006. 22

OLAODE, A.; NAGHDY, G.; TODD, C. Unsupervised classification of images: a review. **International Journal of Image Processing**, v. 8, p. 2014–325, 09 2014. 18

OLOFSSON, P.; FOODY, G. M.; HEROLD, M.; STEHMAN, S. V.; WOODCOCK, C. E.; WULDER, M. A. Good practices for estimating area and assessing accuracy of land change. **Remote Sensing of Environment**, v. 148, p. 42–57, 2014. 19, 40

PARIS, C.; BRUZZONE, L.; FERNÁNDEZ-PRIETO, D. A novel approach to the unsupervised update of land-cover maps by classification of time series of multispectral images. **IEEE Transactions on Geoscience and Remote Sensing**, 2019. 2

PAWŁUSZEK, K.; MARCZAK, S.; BORKOWSKI, A.; TAROLLI, P. Multi-aspect analysis of object-oriented landslide detection based on an extended set of lidar-derived terrain features. **ISPRS International Journal of Geo-Information**, v. 8, n. 8, p. 321, 2019. 4

PELLETIER, C.; VALERO, S.; INGLADA, J.; CHAMPION, N.; DEDIEU, G. Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas. **Remote Sensing of Environment**, v. 187, p. 156–168, 2016. 18

PETRY, O. G. **Cartografia das áreas de risco no Rio Rolante-RS e estratégias de recomposição da mata ciliar**. 132 p. Dissertation (Master in Geography) — Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, 2003. 25

PETSCHKO, H.; BELL, R.; BRENNING, A.; GLADE, T. Landslide susceptibility modeling with generalized additive models—facing the heterogeneity of large regions. In: EBERHARDT E.; FROESE, C. T. K. L. S. (Ed.). **Landslides and engineering slopes: protecting society through improved understanding**. Boca Raton: CRC, 2012. p. 769–777. 31

PETSCHKO, H.; BELL, R.; LEOPOLD, P.; HEISS, G.; GLADE, T. Landslide inventories for reliable susceptibility maps in lower Austria. In: MARGOTTINI C.; CANUTI, P. S. K. (Ed.). **Landslide science and practice**. [S.l.]: Springer, 2013. p. 281–286. 1

- PHAM, V. D. et al. Convolutional neural network—optimized moth flame algorithm for shallow landslide susceptible analysis. **IEEE Access**, v. 8, p. 32727–32736, 2020. 3
- PIATETSKY-SHAPIRO, G. Knowledge discovery in real databases: a report on the ijcai-89 workshop. **AI Magazine**, v. 11, n. 4, p. 68–68, 1990. 15
- PRADHAN, B.; MEZAAL, M. R. Optimized rule sets for automatic landslide characteristic detection in a highly vegetated forests. In: PRADHAM, B. (Ed.). **Laser scanning applications in landslide assessment**. [S.l.]: Springer, 2017. p. 51–68. 3
- PROMPER, C.; PUISSANT, A.; MALET, J.-P.; GLADE, T. Analysis of land cover changes in the past and the future as contribution to landslide risk scenarios. **Applied Geography**, v. 53, p. 11–19, 2014. 1
- QIN, C.-Z.; ZHU, A.-X.; PEI, T.; LI, B.-L.; SCHOLTEN, T.; BEHRENS, T.; ZHOU, C.-H. An approach to computing topographic wetness index based on maximum downslope gradient. **Precision Agriculture**, v. 12, n. 1, p. 32–43, 2011. 33
- QIN, Y.; LU, P.; LI, Z. Landslide inventory mapping from bitemporal 10 m sentinel-2 images using change detection based markov random field. **International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 42, n. 3, 2018. 2, 12
- QUEVEDO, R. P.; GUASSELLI, L. A.; OLIVEIRA, G. G.; RUIZ, L. F. C. Modelagem de áreas suscetíveis a movimentos de massa: avaliação comparativa de técnicas de amostragem, aprendizado de máquina e modelos digitais de elevação. **Geosciences**, v. 38, n. 3, p. 781–795, 2019. 1, 26, 40
- RAMOS-BERNAL, R. N.; VÁZQUEZ-JIMÉNEZ, R.; ROMERO-CALCERRADA, R.; ARROGANTE-FUNES, P.; NOVILLO, C. J. Evaluation of unsupervised change detection methods applied to landslide inventory mapping using aster imagery. **Remote Sensing**, v. 10, n. 12, p. 1987, 2018. 1
- RANGHETTI, L.; BOSCHETTI, M.; NUTINI, F.; BUSETTO, L. sen2r: an r toolbox for automatically downloading and preprocessing sentinel-2 satellite data. **Computers Geosciences**, v. 139, p. 104473, 2020. Available from: <<http://sen2r.ranghetti.info>>. 30

- RAYCHAUDHURI, S. Introduction to Monte Carlo simulation. In: WINTER SIMULATION CONFERENCE, 2008. **Proceedings...** [S.l.], 2008. p. 91–100. 41
- REGAN, A.; SILVESTRIN, P.; FERNANDEZ, D.; GABRIELE, A.; LEVEQUE, N. Sentinel convoy: synergetic observations with satellites flying in formation with european operational missions. In: SMALL SATELLITE AND SERVICES SYMPOSIUM, 2010, Madeira, Portugal. **Proceedings...** [S.l.], 2010. 20, 21
- RICHARDS, J. A. **Remote sensing digital image analysis: an introduction.** [S.l.: s.n.], 2013. 1–494 p. ISBN 9783642300622. 18
- RIO GRANDE DO SUL. SECRETARIA ESTADUAL DO MEIO AMBIENTE - SEMA. **Diagnóstico preliminar: descritivo dos eventos ocorridos no dia 5 de janeiro de 2017 entre as regiões dos municípios de São Francisco de Paula e Rolante/RS.** Porto Alegre, 2017. 26 p. 26
- ROSSATO, M. S. **Os climas do Rio Grande do Sul: variabilidade, tendências e tipologia.** Tese (Doutorado em Geografia) — Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. 26
- ROUSE, J.; HAAS, R.; SCHELL, J.; DEERING, D. **Monitoring vegetation systems in the great plains with ERTS.** [S.l.]: NASA, 1974. 309 p. 30
- RUSSWURM, M.; KÖRNER, M. Multi-temporal land cover classification with sequential recurrent encoders. **ISPRS International Journal of Geo-Information**, v. 7, n. 4, p. 129, 2018. 2
- SALLEH, M. M.; ISMAIL, Z.; ARIFF, S. M.; RAHMAN, M. A.; KHANAN, M. A.; ASMADI, M. A.; RAZAK, K. A. Spatial prediction models for landslide activity mapping using vegetation anomalies. **International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences**, 2019. 31
- SALLEH, M. R. M.; ISHAK, N. I.; RAZAK, K. A.; RAHMAN, M. Z. A.; ASMADI, M. A.; ISMAIL, Z.; KHANAN, M. F. A. Geospatial approach for landslide activity assessment and mapping based on vegetation anomalies. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 42, n. 4/W9, 2018. 3
- SARO, L.; WOO, J. S.; KWAN-YOUNG, O.; MOUNG-JIN, L. The spatial prediction of landslide susceptibility applying artificial neural network and logistic

regression models: a case study of inje, korea. **Open Geosciences**, v. 8, n. 1, p. 117–132, 2016. 31

SCHMALTZ, E. M.; STEGER, S.; GLADE, T. The influence of forest cover on landslide occurrence explored with spatio-temporal information. **Geomorphology**, v. 290, p. 250–264, 2017. 1

SCHULTZ, M.; CLEVERS, J. G.; CARTER, S.; VERBESSELT, J.; AVITABILE, V.; QUANG, H. V.; HEROLD, M. Performance of vegetation indices from landsat time series in deforestation monitoring. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p. 318–327, 2016. 12

SCHWAB, J.; GORI, P.; JEER, S. **Landslide hazards and planning**. [S.l.]: American Planning Association, 2005. ISBN 1932364129. 8

SIM, J.; WRIGHT, C. C. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. **Physical Therapy**, v. 85, n. 3, p. 257–268, 2005. 42

SIYAHGHALATI, S.; SARAF, A. K.; PRADHAN, B.; JEBUR, M. N.; TEHRANY, M. S. Rule-based semi-automated approach for the detection of landslides induced by 18 september 2011 Sikkim, Himalaya, earthquake using IRS LISS3 satellite images. **Geomatics, Natural Hazards and Risk**, v. 7, n. 1, p. 326–344, 2016. 3

SOARES, A. R.; BENDINI, H. N.; VAZ, D. V.; UEHARA, T. D.; NEVES, A. K.; LECHLER, S.; KÖRTING, T. S.; FONSECA, L. M. Stmetrics: a python package for satellite image time-series feature extraction. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2020. **Proceedings...** [S.l.]: IEEE, 2020. p. 2061–2064. 14

SØRENSEN, R.; SEIBERT, J. Effects of DEM resolution on the calculation of topographical indices: TWI and its components. **Journal of Hydrology**, v. 347, n. 1-2, p. 79–89, 2007. 33

SOTHE, C.; CAMARGO, E. C. G.; GERENTE, J.; RENNÓ, C. D.; MONTEIRO, A. M. V. Uso de modelo aditivo generalizado para análise espacial da suscetibilidade a movimentos de massa. **Revista do Departamento de Geografia**, v. 34, p. 68–81, 2017. 31

STEHMAN, S. V. Sampling designs for accuracy assessment of land cover. **International Journal of Remote Sensing**, v. 30, n. 20, p. 5243–5272, 2009. 19

STUMPF, A.; MALET, J.-P.; DELACOURT, C. Correlation of satellite image time-series for the detection and monitoring of slow-moving landslides. **Remote Sensing of Environment**, v. 189, p. 40–55, 2017. 12

TAPSELL, S. et al. **Social vulnerability to natural hazards**. London: Citeseer, 2010. 2, 7

UDELHOVEN, T. Timestats: a software tool for the retrieval of temporal patterns from global satellite archives. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 4, n. 2, p. 310–317, 2010. 11

UEHARA, T. D. T.; CORREA, S. P. L. P.; QUEVEDO, R. P.; KÖRTING, T. S.; DUTRA, L. V.; RENNÓ, C. D. Classification algorithms comparison for landslide scars. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS (GEOINFO), 20. **Proceedings...** São José dos Campos: INPE, 2019. p. 158. 35

UNITED NATION OFFICE FOR DISASTER RISK REDUCTION. **Words into action guidelines: national disaster risk assessment**. UNISDR, 2017.

Available from: <[https://www.unisdr.org/files/52828\\_nationaldisasterriskassessmenthazar%5B1%5D.pdf](https://www.unisdr.org/files/52828_nationaldisasterriskassessmenthazar%5B1%5D.pdf)>. Access in: 22 may. 2021. 8

UNITED STATES GEOLOGICAL SURVEY (USGS). **USGS EROS archive - digital elevation - Shuttle Radar Topography Mission (SRTM) 1**

**Arc-Second Global**. USGS, 2021. Available from:

<[https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-srtm-mission-summary?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-srtm-mission-summary?qt-science_center_objects=0#qt-science_center_objects)>. Access in: apr 03 2021. 22

\_\_\_\_\_. **USGS EROS archive - digital elevation - SRTM mission summary**.

USGS, 2021. Available from: <[https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-srtm-mission-summary?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-srtm-mission-summary?qt-science_center_objects=0#qt-science_center_objects)>. Access in: apr 03 2021. 23

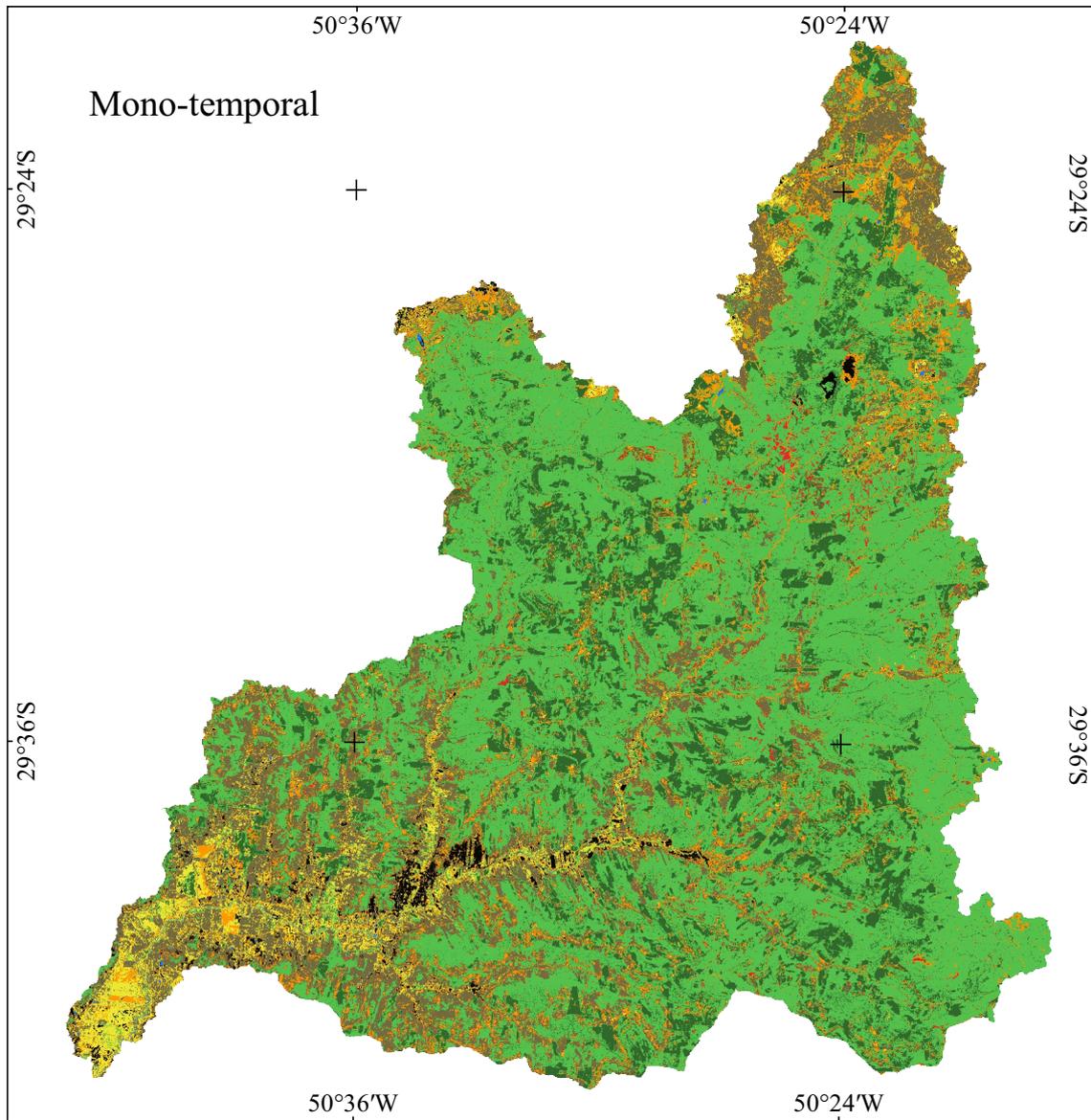
UNIVERSIDADE FEDERAL DE SANTA CATARINA - UFSC. **Atlas brasileiro de desastres naturais 1991 a 2012: volume Brasil**. Florianópolis: UFSC, 2013. 9

VALERIANO, M. D. M.; ROSSETTI, D. D. F. **Topodata: guia para utilização de dados geomorfológicos locais**. São José dos Campos: INPE, 2008. 32

- VALLIN, M. Das 170 áreas com alto risco de deslizamento na serra, só oito passaram por obras de recuperação de encosta. **Veja**, 2011. Available from: <<https://bit.ly/3mjuRcx>>. Access in: 16 may. 2021. 10
- VASSOLER, R. **Ações da vigilância epidemiológica nos desastres naturais. Experiência na Região Serrana em 2011**. 2013. Available from: <<http://www.riocomsaude.rj.gov.br/Publico/MostrarArquivo.aspx>>. 10
- VERIKAS, A.; VAICIUKYNAS, E.; GELZINIS, A.; PARKER, J.; OLSSON, M. C. Electromyographic patterns during golf swing: activation sequence profiling and prediction of shot effectiveness. **Sensors**, v. 16, p. 592, 04 2016. 19
- WANG, C.; LU, L. Object-based random forest classification for detecting plastic-mulched landcover from gaofen-2 and landsat-8 oli fused data. In: INTERNATIONAL CONFERENCE ON AGRO-GEOINFORMATICS, 8., 2019. **Proceedings...** [S.l.], 2019. p. 1–5. 18
- WANG, Y.; ZHANG, J.; TONG, S.; GUO, E. Monitoring the trends of aeolian desertified lands based on time-series remote sensing data in the Horqin Sandy Land, China. **Catena**, v. 157, p. 286–298, 2017. 12
- YANG, W.; WANG, M.; SHI, P. Using modis ndvi time series to identify geographic patterns of landslides in vegetated regions. **IEEE Geoscience and Remote Sensing Letters**, v. 10, n. 4, p. 707–710, 2012. 3
- YI, Y.; ZHANG, W. A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal rapideye satellite imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 13, p. 6166–6176, 2020. 4
- ZHA, Y.; GAO, J.; NI, S. Use of normalized difference built-up index in automatically mapping urban areas from tm imagery. **International Journal of Remote Sensing**, v. 24, n. 3, p. 583–594, 2003. 30
- ZHAO, C.; KANG, Y.; ZHANG, Q.; LU, Z.; LI, B. Landslide identification and monitoring along the Jinsha River catchment (Wudongde reservoir area), China, using the InSAR method. **Remote Sensing**, v. 10, n. 7, p. 993, 2018. 4
- ZHU, Z. Change detection using landsat time series: a review of frequencies, preprocessing, algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 370–384, 2017. 12

## APPENDIX A - MAPS

Figure A.1 - LULC map from Mono-temporal approach.



LEGEND

Classes:

- |              |             |       |
|--------------|-------------|-------|
| Landslide    | Agriculture | Water |
| Forest       | Bare Soil   | Urban |
| Silviculture | Pasture     |       |

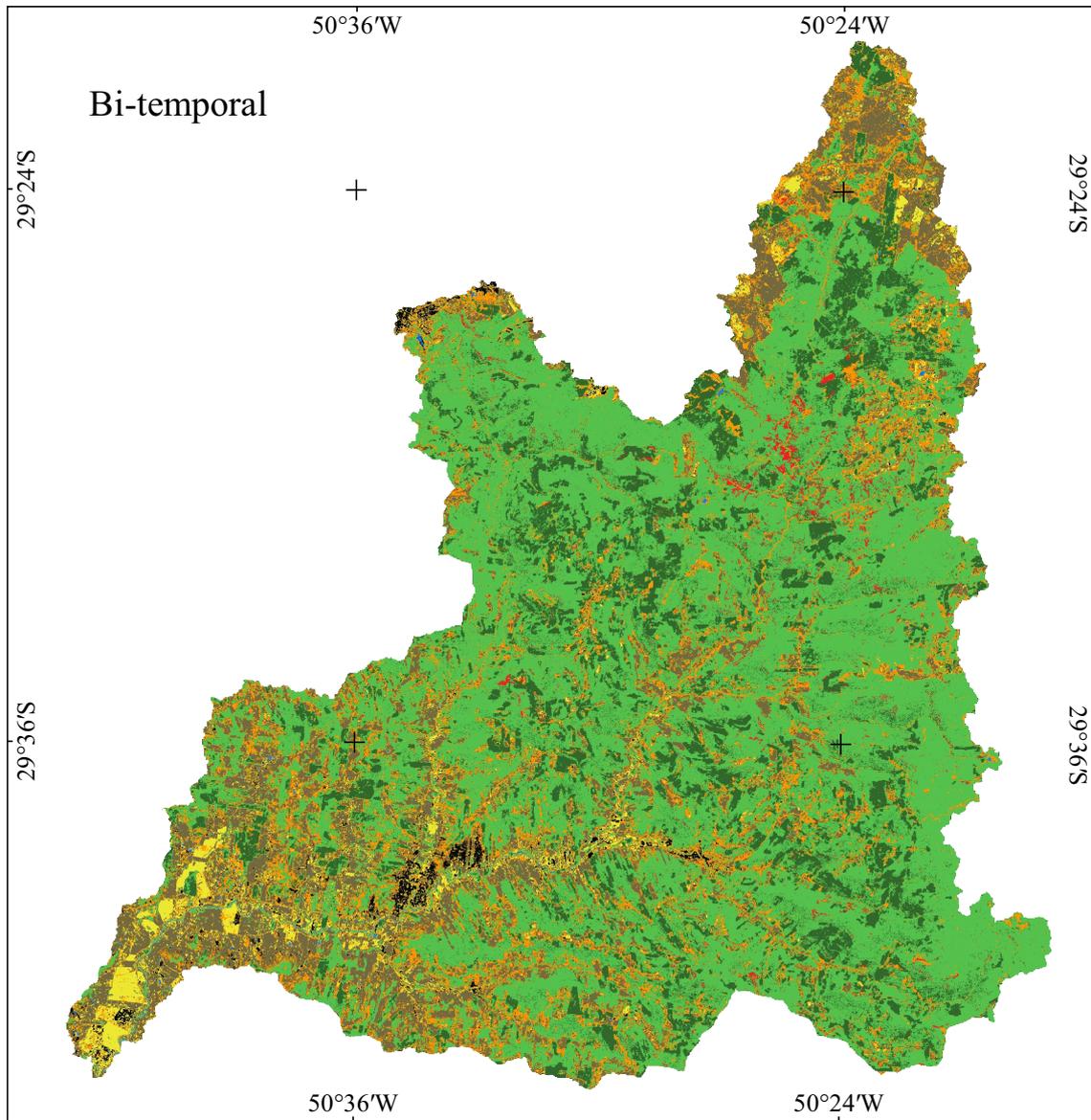


0 10 20 km

Coordinate System: WGS84  
EPSG: 4326

SOURCE: Author's production.

Figure A.2 - LULC map from Bi-temporal approach.



LEGEND

Classes:

- |              |             |       |
|--------------|-------------|-------|
| Landslide    | Agriculture | Water |
| Forest       | Bare Soil   | Urban |
| Silviculture | Pasture     |       |

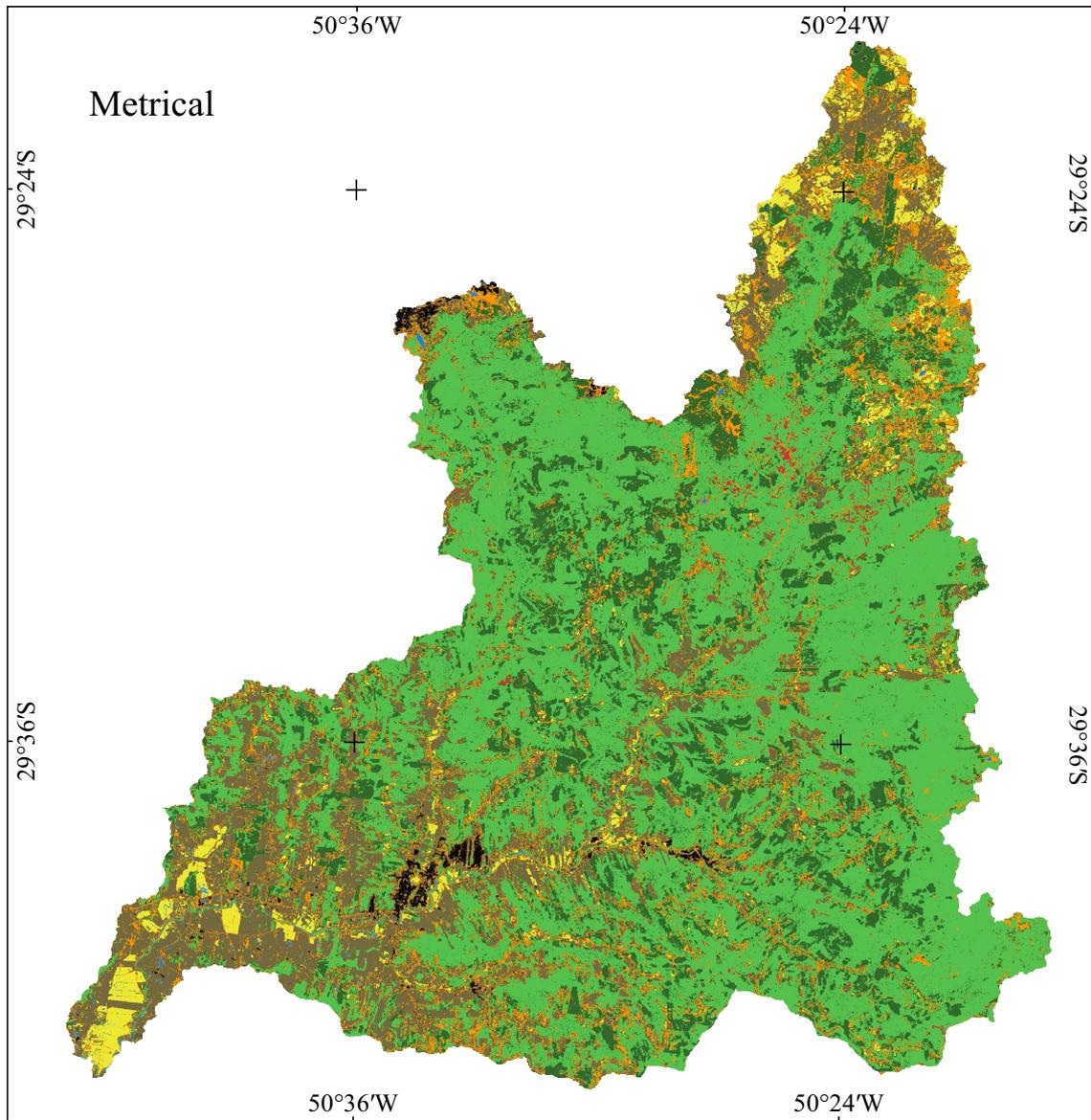


0 10 20 km

Coordinate System: WGS84  
EPSG: 4326

SOURCE: Author's production.

Figure A.3 - LULC map from Metrical approach.



LEGEND

Classes:

- |              |             |       |
|--------------|-------------|-------|
| Landslide    | Agriculture | Water |
| Forest       | Bare Soil   | Urban |
| Silviculture | Pasture     |       |

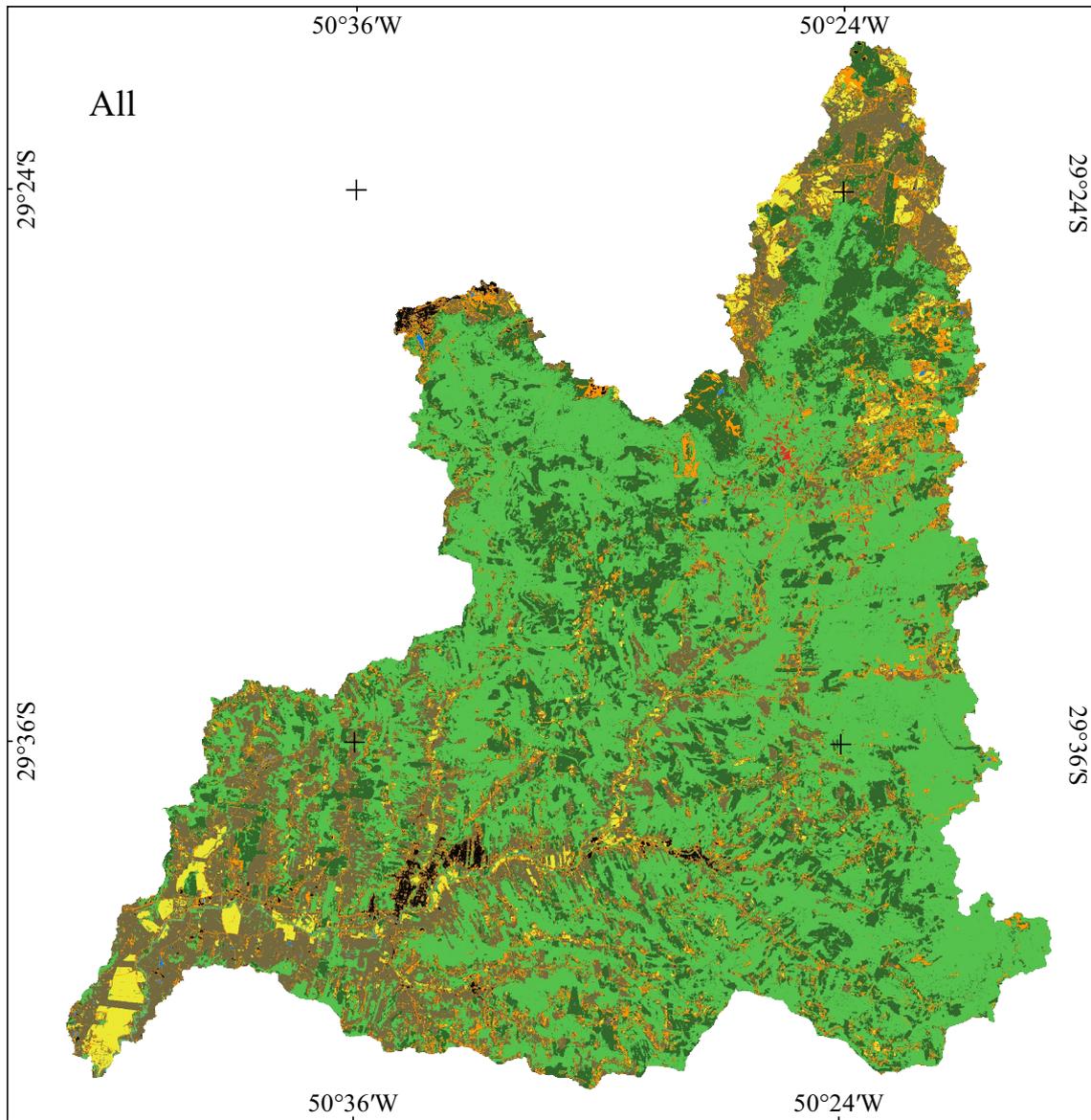


0 10 20 km

Coordinate System: WGS84  
EPSG: 4326

SOURCE: Author's production.

Figure A.4 - LULC map from All approach.



LEGEND

Classes:

- |              |             |       |
|--------------|-------------|-------|
| Landslide    | Agriculture | Water |
| Forest       | Bare Soil   | Urban |
| Silviculture | Pasture     |       |

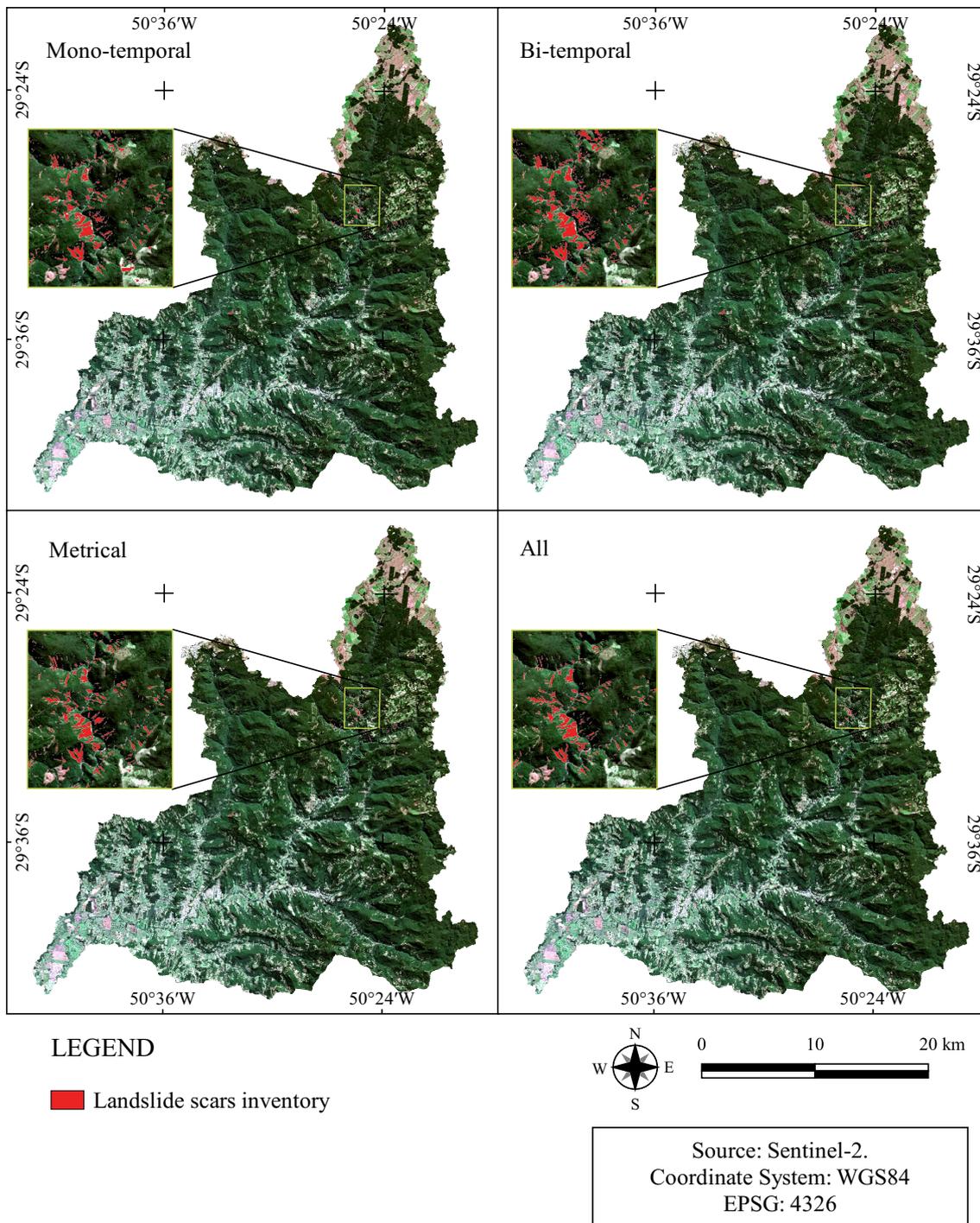


0 10 20 km

Coordinate System: WGS84  
EPSG: 4326

SOURCE: Author's production.

Figure A.5 - Landslide Inventory Maps.



SOURCE: Author's production.