



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-mxx/aaaa/00.00.00.00-XXX

**AGRICULTURAL LAND CLASSIFICATION BASED
ON PHENOLOGICAL INFORMATION FROM DENSE TIME-
SERIES LANDSAT-LIKE IMAGES IN THE BRAZILIAN
CERRADO**

Hugo do Nascimento Bendini

PhD Thesis submitted to the Post
Graduation in Remote Sensing,
supervised by Drs. Leila Maria
Garcia Fonseca and Thales Sehn
Körting, approved in July 13, 2018.

URL do documento original:
<<http://urlib.net/xx/yy>>

INPE
São José dos Campos
2018

Cataloging in Publication Data

Bendini, Hugo do Nascimento.
Cutter Agricultural Land Classification based on Phenological
Information from Dense Time-Series Landsat-like Images in the Brazilian Cerrado
/ Hugo do Nascimento Bendini. - São José dos Campos: INPE, 2018.

Thesis (Doctorate in Remote Sensing) – Instituto Nacional de
Pesquisas Espaciais, São José dos Campos, 2018.
Guiding: Drs. Leila Maria Garcia Fonseca and Thales Sehn Körting

1. big-data. 2. time-series analysis. 3. agricultural land use
classification. 4. multi-sensor. 5. remote sensing.

I. Title

CDU



Esta obra foi licenciada sob uma Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada.

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

FOLHA DE APROVAÇÃO
CONFECCIONADA PELO SPG E INCLUÍDA PELO SID

"A ciência pode classificar e nomear os órgãos de um sabiá

mas não pode medir seus encantos.

A ciência não pode calcular quantos cavalos de força existem

nos encantos de um sabiá.

Quem acumula muita informação perde o condão de adivinhar: divinare.

Os sabiás divinam."

(Manoel de Barros)

To the Brazilian society, and to all those are concerned with sustainable development in the Brazilian Cerrado biome.

ACKNOWLEDGEMENTS

I thank the institutions that provided infrastructure and financial support for the development of my thesis, especially, the Remote Sensing and Digital Image Processing Divisions at the National Institute for Space Research and the Geomatics Lab at the Humboldt University (HU) in Berlin. I also thank CAPES for the financial support.

I would like to express my gratitude to my supervisors, Prof. Dr. Leila Maria Garcia Fonseca and Prof. Dr. Thales Sehn Körting, who were essential for the pragmatic direction of my research, supporting my initiatives and creating opportunities. Their interest in international collaboration gave me the possibility for research abroad. I am also glad to Dra. Ieda Del'Arco Sanches for the support and motivation during these years.

Many thanks to Prof. Dr. Patrick Hostert for support me with such hospitality during my stay at the HU Geomatics Lab in Berlin. His valuable advices were crucial for my thesis.

Many thanks to my professors, Elisabete Caria Moraes, Flávio J. Ponzoni, Lênio S. Galvão, Antonio Roberto Formaggio, Laércio M. Namikawa, Karine R. Ferreira, Lúbia Vinhas, Antônio Miguel V. Monteiro, Eymar S. S. Lopes, Emiliano F. Castejon, Camilo D. Rennó, Gilberto R. Queiroz and also to Dalton de M. Valeriano, Luis Eduardo P. Maurano, Jeferson de S. Arcanjo and Alfredo José B. Luiz. I feel grateful for how much I learned from you, your insights, motivation and stimulating discussions, which were important to support my decisions. Thanks to Prof. Dr. Bernardo F. T. Rudorff and MSc. Joel Risso, for the opportunity to join the Rally da Safra team 2016, which was crucial for collecting data for the study cases.

Thanks to all my colleagues that I had a pleasure to meet during my studies. And thank all my family. My parents, João Marcos and Teresa Cristina, deserve special thanks for their support, encouragement and affection. Without them, I would not be on this place today.

ABSTRACT

Brazil has an important role in the world in terms of food production and the largest native forest, providing essential environmental services for the planet and humanity. However, this highlights the challenge of creating an economic development model that takes into account the environmental conservation. Brazil has already demonstrated successful experiences in Amazon deforestation reduction, but other biomes of great environmental importance, such as the Cerrado, has been under great pressure of agricultural expansion. Satellite image time series can be used to derive phenological information of vegetation, and considering the high heterogeneity of crop types and their respective planting calendars in Brazil, is essential for crop classification and monitoring. Our hypothesis in this thesis is that phenological information can be extracted from Landsat-like dense image time series, allowing the development of a method for agriculture mapping with more detail. We tested the integration of different satellite, such as Landsat-8, Landsat-7 and CBERS-4, combined with different smoothing techniques, to generate EVI (Enhanced Vegetation Index) image time series at high frequency in order to extract the phenological metrics. A hierarchical classification approach using the Random Forest algorithm was developed to produce detailed agricultural maps. The classification results are promising (higher than 80% of overall accuracy) and showed the feasibility of applying the method on a large scale and over a longer period of time for the Cerrado biome. In addition, the phenological information obtained by the method showed a potential to be used in the understanding of different agricultural practices adopted by farmers in property level.

Keywords: big-data, time-series analysis, agricultural land use classification, multi-sensor, remote sensing.

CLASSIFICAÇÃO DE ÁREAS AGRÍCOLAS BASEADA EM INFORMAÇÕES FENOLÓGICAS DE SÉRIES TEMPORAIS DE IMAGENS *LANDSAT-LIKE* NO CERRADO BRASILEIRO

RESUMO

O Brasil tem um papel importante no mundo em termos de produção de alimentos e a maior floresta nativa, fornecendo serviços ambientais essenciais para o planeta e para a humanidade. No entanto, isso destaca o desafio de criar um modelo de desenvolvimento econômico que leve em consideração a conservação ambiental. O Brasil já demonstrou experiências bem-sucedidas na redução do desmatamento da Amazônia, mas outros biomas de grande importância ambiental, como o Cerrado, estão sob grande pressão de expansão agrícola. Séries temporais de imagens de satélite podem ser usadas para derivar informações fenológicas da vegetação. Considerando a diversidade de culturas agrícolas e seus respectivos calendários de plantio no Brasil, essas informações são essenciais para a classificação e monitoramento agrícola. Nossa hipótese é que informações fenológicas podem ser extraídas de séries temporais de imagens de resolução espacial *Landsat-like*, permitindo o desenvolvimento de método para mapeamento detalhado da agricultura. Testamos a integração de diferentes satélites, como Landsat-8, Landsat-7 e CBERS-4, combinados com diferentes técnicas de suavização para gerar séries temporais de imagem EVI (*Enhanced Vegetation Index*) em alta frequência e extrair as métricas fenológicas. Uma abordagem de classificação hierárquica usando o algoritmo *Random Forest* foi aplicada para produzir os mapas. Os resultados da classificação são promissores (acima de 80% da acurácia) e mostraram a viabilidade de aplicar o método em larga escala e por um longo período para o Bioma Cerrado. Além disso, as informações fenológicas mostraram potencial para serem utilizadas na compreensão de diferentes práticas agrícolas adotadas pelos agricultores no Cerrado, em escala de propriedade.

Palavras-chave: big data, análise de séries temporais, classificação de uso agrícola, multi-sensor, sensoriamento remoto.

LIST OF FIGURES

	<u>Page</u>
Figure 1.1 – Overview of the general methodology, specific objectives of the thesis and the respective linked chapters.....	7
Figure 2.1 – False color (bands 5, 6 and 4 in red, green and blue respectively) Landsat imagery of the study area (note the overlapping area of two adjacent scenes 219/75 and 220/75).	11
Figure 2.2 – Some of the seasonality parameters generated by TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value.	14
Figure 2.3 – EVI time series spectral profile for an annual crop sample located into a triple cropping system of potato, bean and corn. Dotted line represents the time series outliers and null values removed and the thick line the double logistic filtered time series.	16
Figure 2.4 – EVI time series spectral profile for a semi-perennial crop sample of cassava.	17
Figure 2.5 - EVI time series spectral profile for a semi-perennial crop sample of sugarcane.....	17
Figure 2.6 - EVI time series spectral profile for a perennial crop sample of rubber trees.....	18
Figure 2.7 - EVI time series spectral profile for perennial crop sample of one of the citrus farm (similar to mango).	19
Figure 2.8 – EVI time series spectral profile for a planted forest sample of eucalyptus.....	20
Figure 2.9 – 3D-scatterplot of length of season, base value and left derivative features of the first seasons. The zero values was removed for best visualization.....	21
Figure 3.1 – Location of the study area in São Paulo state, Brazil.....	26

Figure 3.2 – Cropped images used on the correlation analysis. (a) LANDSAT-8 OLI EVI (06 August 2015) and (b) CBERS-4 MUX EVI (August 4th, 2015).	29
Figure 3.3 – General scheme of the methodology used to build the integrated time series. On the left, a time series of EVI (the red line is the predicted time series using the equation to predict OLI reflectance from MUX and the blue line is the original Landsat time series); the integrated time series is on the right, which points out the positions where the replacement has occurred.	30
Figure 3.4 – Example of how local outliers were removed from the NDVI time series. The cyan lines are the positions where cloud and cloud shadow were detected by Fmask. The black line is the integrated time series and the green line is the filtered integrated time series using Equation 3.1 followed by Equation 3.2.	32
Figure 3.5 – Scatterplot of the pair of cropped images used to predict OLI reflectance from MUX reflectance. (a) EVI and (b) NDVI.	34
Figure 3.6 – Results of different approaches for noise removal in an EVI time series. In (a) the black line is the original Landsat-8 time series, the blue line is the integrated time series and the black thin line is filtered integrated time series. In (b) the blue line is the filtered integrated and the black line is the Landsat-8 filtered time series.	36
Figure 4.1 – Location of Itobi municipality, in São Paulo state, Brazil.	42
Figure 4.2 – Smoothed EVI time series with the different algorithms using the raw data. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons detected by the TIMESAT’s algorithm to extract the phenological attributes.	45
Figure 4.3 – Smoothed EVI time series with the different smoothing algorithms using the filtered EVI time series. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons	

detected by the TIMESAT’s algorithm to extract the phenological attributes.	46
Figure 4.4 – Within-class mean standard deviation for the different classes using the different smoothing algorithms with the raw EVI time series.	47
Figure 4.5 – Within-class mean standard deviation for the different classes using the different smoothing algorithms with the filtered EVI time series.....	48
Figure 5.1 – Study sites on the Cerrado biome.....	54
Figure 5.2 – Averaged EVI phenological profiles for each Land Cover class in the season 2015–16 (black lines), with their respective standard deviations (blue margins).	63
Figure 5.3 – Averaged EVI phenological profiles for each Crop Group (Level 3) class in the season 2015–2016 (black lines), with their respective standard deviations (blue margins).	65
Figure 5.4 – Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 3 classification (First crop / Second crop, First crop / Winter crop, Single crop, and Single / Non-commercial crop) based on the training pixels.) based on the training pixels.	68
Figure. 5.5 – Pixel-wise EVI time series plotted on polar coordinates using the representation proposed by Körting et al. (2013) for the classes of level 3 classification (Crop Group).....	69
Figure 5.8 – Map of region C, obtained at level 2 (Land Cover: Perennial crop and Non-crop classes) and level 3 (Crop Group) of the hierarchical classification approach. For the production of the maps, we grouped the different classes of natural vegetation into one.	75
Figure 5.9 – Classification results based on seasonal phenological parameters for the municipalities of Luís Eduardo Magalhães-BA (region A1), Campo Verde-MT (region B) and Itobi-SP (region C), obtained at level 2 (Land Cover: “Annual crop and Semi-Perennial crop” only), level 3 (Crop Group) (a, c, e) and	

level 4 (Crop Rotation) (b, d, f) of the hierarchical classification approach.	76
Figure A1. Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 1 classification (Annual crop and Semi-perennial crop; Perennial crop and Non crop) based on the training pixels.	92
Figure A2. Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 2 classification (Annual crop and Semi-perennial crop) based on the training pixels.	93

LIST OF TABLES

	<u>Page</u>
Table 2.1 – Confusion matrix: (a) Perennial crop, (b) Annual crop, (c) Planted Forest and (d) Semi-perennial crop. EC = errors of commission; EO = errors of omission.....	20
Table 3.1 – Availability of Landsat-8 (Path/Row 219/75) and CBERS-4 (Path/Row 155/124) imagery from August 2015 to August 2016.	27
Table 3.2 – Spectral Band Specifications for Landsat-8 OLI and CBERS-4 MUX.	28
Table 3.3 – Characteristics of the pair of MUX and OLI images used for correlation analysis.	28
Table 3.4 – Accuracy of classification for the different data set classifications.	37
Table 4.1 – Accuracy assessment statistics of the classifications considering the different smoothing algorithms using the EVI raw time series including global accuracy (GA), kappa statistic, and the producer's accuracy (PA) of each class.	44
Table 4.2 – Accuracy assessment statistics including global accuracy and (kappa statistic) for classifications, considering the different smoothing algorithms using the EVI filtered time series, and the producer's accuracy (PA) of each class.	47
Table 5.1 – Definition of the four levels of detail as defined in the hierarchical classification scheme. The number of available reference polygons for training and validation is shown in parenthesis.....	56
Table 5.2. Overall accuracy and class wise f1-scores obtained for the different hierarchical classification levels.	71
Table B1. Mean Confusion Matrix of classification Level 1 (Annual crop and Semi-perennial crop) resulting from the Monte Carlo process, with 1000 simulations.	94

Table B2. Mean Confusion Matrix of classification Level 2 (Annual crop and Semi-perennial crop) resulting from the Monte Carlo process, with 1000 simulations.	94
Table B3. Mean Confusion Matrix of classification Level 3 (First crop / Second crop, First crop / Winter crop, Single crop, and Single / Non-commercial crop) resulting from the Monte Carlo process, with 1000 simulations.	95

LIST OF ACRONYMS AND ABBREVIATIONS

ACCA	Automated Cloud Cover Assessment
AG	Asymmetric Gaussian
CBERS	China Brazil Earth Resources Satellite
CEPAGRI	Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura / Center for Meteorological and Climatic Research Applied to Agriculture
CEPEA	Centro de Estudos Avançados em Economia Aplicada / Center for Advanced Studies on Applied Economics
CONAB	Companhia Nacional de Abastecimento / National Supply Company
DEM	Digital Elevation Model
DL	Double Logistic
DOY	Day of Year
EC	Error of Comission
EO	Error of Omission
EO	Earth Observation
Eos	End of Season
EROS	Earth Resources Observation and Science
ESPA	Center Science Processing Architecture
ETM	Enhanced Thematic Mapper
EVI	Enhanced Vegetation Index
GA	Global Accuracy
GDP	Gross Domestic Product
GPS	Global Positioning System
IBGE	Instituto Brasileiro de Geografia e Estatística / Brazilian Institute of Geography and Statistics
INPE	Instituto Nacional de Pesquisas Espaciais / National Institute for Space Research
L1T	Level-1 Terrain Corrected
Lder	Left Derivative
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing
Linteg	Left Integer
LoS	Length of Season

LSP	Land Surface Phenology
MAPA	Ministério da Agricultura, Pecuária e Abastecimento / Ministry of Agriculture, Livestock and Supply
Mid	Middle of Season
MMA	Ministério do Meio Ambiente / Ministry of Environment
MSI	Multispectral Instrument
MUX	Multispectral Camera
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
OLI	Operational Land Imager
PA	Producer's Accuracy
PAM	Produção Agrícola Municipal / Municipal Agricultural Production
RBF	Radial Basis Function
Rder	Right Derivative
RF	Random Forest
SG	Savitzky-Golay
Sinteg	Small Integral
SLC	Scan Line Corrector
Sos	Start of Season
SRF	Spectral Response Function
SRTM	Shuttle Radar Topography Mission
UNICAMP	Universidade de Campinas / University of Campinas
USGS	United States Geological Survey
USP	Universidade de São Paulo / São Paulo University
UTC	Universal Time Coordinated
WFI	Wide-Field Imager
WRS	Worldwide Reference System

CONTENTS

	<u>Page</u>
1 INTRODUCTION	1
1.1. Document Organization	6
2 USING LANDSAT 8 IMAGE TIME SERIES FOR CROP MAPPING IN CASA BRANCA MUNICIPALITY, BRAZIL	9
2.1. Introduction	10
2.2. Materials and Methods	11
2.2.1. Study area.....	11
2.2.2. Remotely sensed data	12
2.2.3. Removal of outliers and null values.....	13
2.2.4. Seasonal data extraction.....	13
2.2.5. Training and validation data	15
2.2.6. Random Forest classification	15
2.3. Results and Discussion	15
2.3.1. Crop phenological curves and TIMESAT features	15
2.3.2. Classification performance	20
2.4. Final Considerations	23
3 ASSESMENT OF A MULTI-SENSOR APPROACH FOR NOISE REMOVAL ON LANDSAT-8 OLI TIME SERIES USING CBERS-4 MUX DATA TO IMPROVE CROP CLASSIFICATION BASED ON PHENOLOGICAL PARAMETERS	23
3.1. Introduction	24
3.2. Methodology	25
3.2.1. Study area.....	25
3.2.2. Remotely sensed data	26
3.2.3. Correlation analysis between Landsat-8 OLI and CBERS-4 MUX.....	28
3.2.4. Building the multi-sensor time series.....	30
3.2.5. Filtering the time series	31
3.2.6. Extracting phenological features	32
3.2.7. Classification	33

3.3.	Results and Discussion	34
3.4.	Final Considerations.....	38
4	EVALUATION OF SMOOTHING METHODS ON LANDSAT-8 EVI TIME SERIES FOR CROP CLASSIFICATION BASED ON PHENOLOGICAL PARAMETERS	40
4.1.	Introduction.....	40
4.2.	Work Methodology.....	41
4.3.	Results and Discussions.....	43
4.4.	Final Considerations.....	49
5	ROBUST CLASSIFICATION OF CROPLANDS IN THE BRAZILIAN CERRADO BASED ON PHENOLOGICAL INFORMATION FROM DENSE LANDSAT TIME SERIES	50
5.1.	Introduction.....	50
5.2.	Materials and Methods	53
5.2.1.	Study areas.....	53
5.2.2.	Definitions for the hierarchical classification scheme	55
5.2.3.	Remotely sensed data	58
5.2.4.	Landsat dense time series	58
5.2.5.	Phenometrics	59
5.2.6.	Random Forest classification	62
5.3.	Results and Discussion	63
5.3.1.	Phenological profiles.....	63
5.3.2.	Phenometrics	66
5.3.3.	Land cover and agricultural mapping	70
5.3.4.	Outlook.....	77
6	SYNTHESIS	78
6.1.	Can we use CBERS-4 MUX data to remove noise in a Landsat-8 OLI NDVI and EVI image time series to improve a crop classification method based on phenological information? ...	79
6.2.	Which smoothing method is better for phenological metrics-based crop classification with Landsat-8 OLI EVI image time series in the Cerrado biome?	79

6.3. Can we integrate Landsat-8 OLI and Landsat-7 ETM+ data to generate a dense high spatial resolution image time series, for classifying agricultural land use in the Cerrado?	80
6.4. Are Landsat based phenological metrics suitable to analyze crop management practices of the major crops in the Cerrado?...	80
6.5. Future work.....	81
REFERENCES.....	81
Appendix A	92
Appendix B	94

1 INTRODUCTION

The Brazilian agribusiness represents around 22.5% of the total Gross Domestic Product (GDP) of the country (CEPEA-USP/CNA, 2015). Even though Brazil's GDP retracted 3.8% in 2015 due to both recession and political unrest, the agricultural sector's GDP increased by 1.8% and was the only sector to report any job growth (SPERA, 2017; LEWIS, 2016; MAPA, 2016). At the same time, Brazil draws global attention as a top emitter of carbon dioxide (CO²) from land-use change and deforestation, mostly due to agriculture expansion (GALFORD et al., 2013), which causes an impact on ecosystem services, biodiversity, water cycle and the carbon cycle (FEARNSIDE, 2006).

However, the agricultural scenario in Brazil has changed considerably in the last decades, especially with respect to crop-management practices. We can understand the context of these changes by some studies in the literature that point out that the adoption of management practices can increase productivity through agricultural intensification (CONWAY; TOENNIESSEN, 1999; TILMAN et al. 2002; GODFRAY et al. 2010). Agriculture can be intensified by the expansion to new planting areas or increasing productivity in consolidated areas, improving some crop-management practices such as irrigation, no-tillage and the adoption of double-cropping system (OLIVEIRA et al., 2014). One important case that illustrates this scenario in Brazil is that deforestation rates in Mato Grosso state decreased 90% between 2004 and 2014 (INPE, 2017) whereas crop production continued to increase thanks to the adoption of intensive agricultural practices (IBGE, 2016) (ARVOR et al., 2017). On the other hand, in the region of Matopiba¹, known as Brazil's "Last Agricultural Frontier", where occurred the most clearing of natural vegetation on the

¹ In May of 2015, 337 municipalities in the north-eastern Cerrado states of Maranhão, Tocantins, Piauí, and Bahia were officially designated "Matopiba" by the Brazilian government, and a bill was ratified committing the government to investing in infrastructure, agricultural technology, and the expansion of the rural middle class. Matopiba is the newest and potentially last agricultural frontier region within the Cerrado (SPERA et al., 2017).

Cerrado biome during the past decade, 85% of the production areas are still based on soy monoculture (SPERA et al., 2017).

Therefore, agriculture mapping and monitoring is important to understand the intensification process and its evolution over time, considering the different crop-management practices. Additionally, it enables the generation of fast and accurate information, which is strategic and crucial for policies and for decision making in the investment market. The availability of this information is also important to support food security programs, preventing food shortages and enabling transparency in the agricultural market.

Given that, the adoption of new crop-management practices currently represents the most promising alternative to increase crop production with limited impacts on the natural environment. Monitoring procedures to evaluate the implementation and communication to policy makers and society are necessary to subsidise policies, such as the Low Carbon Agriculture Plan (ABC Plan) (MAPA, 2012), which is a policy to reduce agricultural emissions.

Yet, mapping agriculture in a regional-scale allows the estimation of productivity on a more accurate way. Once taking into account the capacity of adaptations that the farmers can have in response to variations on climatic conditions and incentive on crop production, this information is important to predict the impact of agricultural policies and long-term economic effects in agriculture (ASSUNÇÃO; CHEIN, 2009).

The potential use of remotely sensed data from Earth Observation (EO) satellites to map and monitor agriculture is mainly due to the most cost-effective means for gathering spatially explicit, timely, detailed and reliable information about vegetation over large land areas with high revisit frequency (ATZBERGER et al., 2013). Exploring the temporal frequency of satellite data allows finding relations between the spectral information and the phenological aspects of different crop types (REED et al., 1994). This is essential for crop classification, since there is a high similarity between different crops and it is a complex task to discriminate them (PEÑA-

BARRAGÁN et al., 2011). Besides mapping, phenological information of crops is also important for many other applications, such as estimation of net primary production (KIMBALLA et al., 2004; TRAMONTANA et al., 2015), decision-making about water and fertilizer supply (DINGKUHN; LEGAL, 1996; SAKAMOTO et al., 2005; ZHENG et al., 2015). Furthermore, a resolution of Brazil's Central Bank (Resolution 4427 of June 25, 2015²) was recently implemented authorizing the use of remote sensing for rural credit operations inspection purposes. Among the requirements for compliance with this resolution is that the method must allow for crop type identification and its development in each phenological phase, what highlights the importance of deriving phenological information from agricultural crops through remote sensing techniques.

With over half of Brazil's agricultural land falling within the Cerrado biome limits, the consequent land clearing in this region has been central to the development and strength of the country's agricultural sector (SPERA, 2017). The Cerrado biome, also known as the Brazilian Savannah, is the second larger biome in Brazil, after Amazon, with an extent of more than 2 million km² (MMA, 2016). Considered as a global biodiversity hotspot, the Cerrado also provides environmental services of global importance, such as carbon storage and climate regulation (ALCAMO, J.; HASSAN, R, 2003). A weak land conservation status has led to large-scale conversions from natural to agricultural land that already affected more than 40% of the Cerrado, which is likely to aggravate in the future (FERREIRA et al., 2012; SANO et al., 2010). The intense scenario of agricultural expansion falls under the need for the development of accurate methods for mapping the distribution of agricultural areas and its evolution over time, so that we can ensure a sustainable development of agriculture and the preservation of the biome.

² http://www.bcb.gov.br/pre/normativos/res/2015/pdf/res_4427_v1_O.pdf

Recently, TerraClass project was implemented in order to map the land use in deforested areas and understand these spatial patterns of land use and land cover, firstly in the Amazon biome (ALMEIDA et al., 2016), and recently in the Cerrado (INPE, 2017). Different classes were considered, and specifically agriculture was mapped at intermediate thematic detail, including annual crops, perennial crops (which includes coffee, citrus, and sugarcane) and planted forest. However, although the generated maps have wide applicability in understanding the key phenomena and of landscape transformation processes constraints of biomes (FONSECA-MORELLO, 2015; ADAMI et al, 2015; RISSO et al, 2012), the methodology still relies on manual procedures for visual interpretation, which involves subjectivity, as well as a lot of effort and time for execution. The methodology used on TerraClass for mapping annual agriculture (COUTINHO et al., 2013) is limited to 250 meters of spatial resolution and depends on the definition of thresholds that need to be manually defined for each region. Although there is a crop calendar recommended by some institutions and even required by some financial agencies, farmers do not follow it strictly. So, approaches based on thresholds definition with fixed time intervals do not work well, once they can vary through different regions. Mapping croplands using phenological information derived from coarse resolution satellite image time series has been extensively studied (SAKAMOTO et al., 2005; WARDLOW et al, 2007; RUDDORF et al, 2010; ESQUERDO et al, 2011; ARVOR et al, 2011; COUTINHO et al, 2013; BORGES; SANO, 2014). The Advanced Very High Resolution Radiometer (AVHRR), and the Moderate Resolution Imaging Spectroradiometer (MODIS), provide high temporal resolution images with wide coverage, and have been essential in the development of these studies. However, their moderate spatial resolution (250 – 1000 meters) limits the development of spatially accurate agricultural maps due to the spectral mixture of different targets, and also do not allow the detection of small fields, being thus restricted to regions where agriculture is practiced in large scale. This is a major problem, especially for regions where crop fields are typically small, diverse in shapes, and management approaches vary, which deem this

approaches based on coarse resolution data ineffective (ZHANG et al., 2016).

The free data police for Landsat (WULDER et al., 2012), CBERS-4, and Sentinel images as well as the advances on methodologies of higher computational performance have provide a propitious environment to work with image time series analysis and intensive data technologies. It's also possible to combine data acquired from sensors onboard Landsat series, such as the Enhanced Thematic Mapper (ETM+) and Operational Land Imager (OLI), which has provided improvements on the newly collection management of the Landsat archive available on the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (USGS, 2016; GERACE et al., 2017; MICIJEVIC et al., 2016). These improvements produced a tiered collection of consistently geolocated and radiometrically calibrated products that are suitable for time series analyses.

The scientific community has been widely working on land cover and land use mapping using Landsat-based approaches (MULLER et al., 2015; RUFIN et al., 2015; ZHENG et al, 2015; PEÑA et al, 2015; PAN et al, 2015). Even though, there are still many challenges in mapping agricultural lands with higher spatial resolution images. Higher spatial resolution sensors, such those onboard Landsat satellite, do not have sufficient temporal resolution to detect all typical short variations in agricultural areas.

Since the Landsat-8 orbit is 8 days apart from the Landsat 7 one, these two sensors together provide 8-day repeatable observations. However, due to the fact that ETM+ data had been degraded by the Scan Line Corrector (SLC)-off gaps, there is less wall-to-wall data available between 2012 (the end of Landsat-5 TM lifetime) and 2013 (Landsat-8 launching). Some methods have been developed to reconstruct the missing information generating smoothed and gap-filled time series of higher spatial resolution data (SHEN et al., 2015; SCHIEWDER et al., 2016; MARUJO et al., 2017). But there are still problems for mapping agricultural lands with high detailed information about cropping systems and crop types. There is also a possibility of combining data from other satellite sensors, such as MUX

(Multispectral Camera) onboard CBERS-4 (China Brazil Earth Resources Satellite) to generate higher temporal resolution (i.e., 8 days) satellite image time series, what we are calling dense satellite image time series.

Within this context, the objective of this thesis is to develop, describe and assess a method based on phenological metrics derived from dense Landsat-like image time series to classify agricultural land use in the Cerrado biome. To develop this work, we assume the hypothesis that phenological metrics from high-spatial resolution image dense time series makes it possible to have progress toward more refined agricultural land classification and with accuracies higher than the average from the recent literature (> 80%).

To reach this objective, the following set of tasks will be developed:

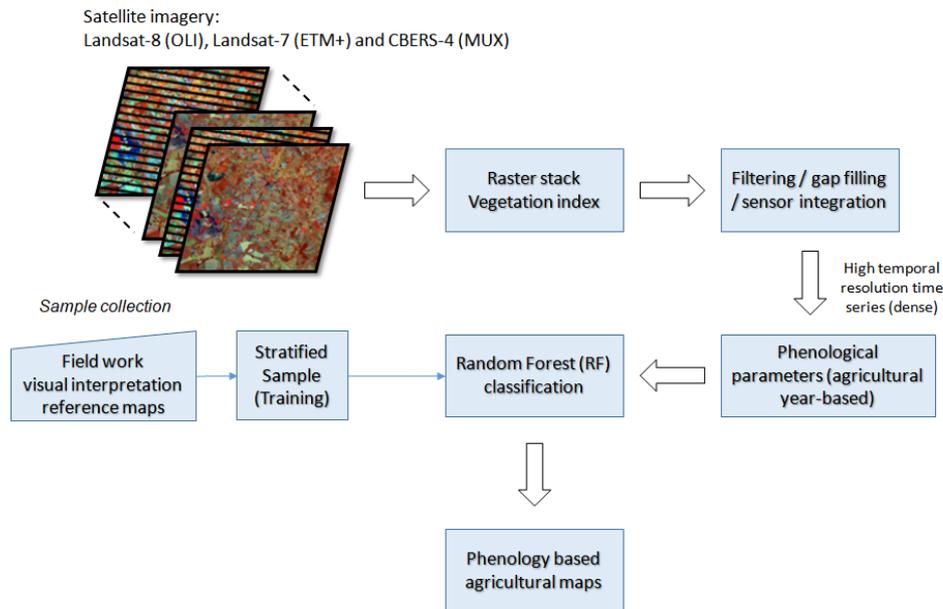
- Investigate noise removal in Landsat-8 OLI time series using CBERS-4 MUX data to improve a crop classification method based on phenological features;
- Evaluate different time series smoothing algorithms for agricultural use discrimination;
- Integrate Landsat-8 OLI and Landsat-7 ETM+ data to generate a dense Landsat-like image time series;
- Explore Landsat based phenological metrics to analyze the major crop management practices in the Cerrado biome.

1.1. Document Organization

The methodology proposed in this thesis to classify the agricultural land is complex and involves many steps of processing to choose the adequate image processing procedures. Therefore, we organized the thesis by describing the experiments that describe each step of the methodology in chapters. These experiments are based on case studies. Figure 1.1 shows

an overview of the methodology and how we organized the chapters to answer specific questions about the methodology. The chapters are based on articles, which were published in International Journals and Conferences, except Chapter 5, which is being prepared for submission.

Figure 1.1 – Overview of the general methodology, specific objectives of the thesis and the respective linked chapters.



Issues addressed in each chapter

Chapter 3

Can we use CBERS-4 MUX data to remove noise in a Landsat-8 OLI NDVI and EVI image time series to improve a crop classification method based on phenological information?

Chapter 4

Which smoothing method is better for phenological metrics-based crop classification with Landsat-8 OLI EVI image time series in the Cerrado biome?

Chapter 5 – Question 1

Can we integrate Landsat-8 OLI and Landsat-7 ETM+ data to generate a dense high spatial resolution image time series, for classifying agricultural land use in the Cerrado?

Chapter 5 – Question 2

Is it suitable to explore Landsat based phenological metrics to analyse crop management practices of the major crops in the Cerrado?

In Chapter 2, we present the experiment that investigates the construction of dense Landsat time series, taking advantage of the overlapping area between two adjacent Landsat scenes (LUIZ et al., 2015). Afterwards, phenological metrics (phenometrics) derived from Enhanced Vegetation Index (EVI) time series are used to classify agricultural areas in a region of the Brazilian Cerrado. EVI was chosen because it highlights the green vegetation canopy, taking into account the background and atmospheric effects. We used a simple linear interpolation for removing outliers and noise in the time series. The phenometrics were obtained by the TIMESAT software (JÖNSSON; EKLUNDH, 2002; 2004), since it has been reported as a promising strategy for classifying agricultural areas (BORGES; SANO, 2014; TOMÁS et al., 2015). The following classes were considered in this study: Annual Crop, Semi-Perennial Crop, Perennial Crop and Planted

Forestry, the same classes described in the survey Municipal Agricultural Production (PAM) produced by the Brazilian Institute of Geography and Statistics (IBGE, 2016). The high accuracies achieved on this experiment motivated us to keep working with the phenological metrics derived from dense image time series. But as this methodology had the dependence of overlapping scenes, we still needed to find a way to expand it for larger areas, taking advantage of other sensors such as CBERS-4 MUX or even sensors from the Landsat family.

In this way, in Chapter 3 we investigated a Landsat-8 OLI time series noise removal method, using CBERS-4 MUX data to improve the image time series temporal resolution and consequently the crop classification with the phenological metrics. An algorithm was built to look for the nearest MUX image for each Landsat image, based on user defined time span. The algorithm checks for cloud contaminated pixels on the Landsat time series using Fmask (ZHU; WOODCOCK, 2012) and replaces the contaminated pixels to build the integrated time series (Landsat-8 OLI + CBERS-4 MUX). In this experiment, we evaluate the use of EVI and also the Normalized Difference Vegetation Index (NDVI) to derive phenometrics. The phenometrics were extracted from the time series samples by different methods (original time series and multi sensor time series, with and without filtering) and subjected to data mining process based on Random Forest classification (BREIMAN, 2001). We used TIMESAT software for the phenometrics extraction, and the Double Logistic function was selected for the time series parameterization, based on the recommendation of Borges and Sano (2014). However, this recommendation was done based on an experiment using MODIS image time series.

Thus, in Chapter 4 we tested the three different time series smoothing methods implemented in TIMESAT software – Savitzky-Golay (SG), asymmetric Gaussian function (AG) and double-logistic function (DL) (JÖNSSON; EKLUNDH, 2002, 2004). We also evaluate their impact on the agricultural land use classification, in order to recommend a smoothing method for the phenometrics extraction from Landsat-like image time series, using TIMESAT.

Finally, in Chapter 5, we described a complete methodology developed in this thesis based on phenometrics derived from dense satellite image time series to classify agricultural land use in three different regions in the Cerrado biome: northeast of São Paulo State, west of Bahia State and southeast of Mato Grosso State. As the results using CBERS-4 data were not satisfactory, we used Landsat Enhanced Thematic Mapper (L7/ETM+) and Operational Land Imager (L8/OLI) sensors, applying a weighted ensemble of Radial Basis Function (RBF) convolution filters as a kernel smoother to fill data gaps such as cloud cover and Scan Line Corrector (SLC)-off data (SCHWIEDER et al., 2016). Through this approach, we created a dense Enhanced Vegetation Index (EVI) data cube with an 8-day temporal resolution. We used only EVI also because the results of Chapter 4 showed that with this vegetation index we could reach better results than using NDVI. We used a classification approach based on a hierarchical scheme with four levels, from land covers (pasture, planted forest, annual crop, semi-perennial crop and natural vegetation) to crop type level.

In Chapter 6, we present conclusions and future work.

2 USING LANDSAT 8 IMAGE TIME SERIES FOR CROP MAPPING IN CASA BRANCA MUNICIPALITY, BRAZIL³

³ This chapter is based on the paper: BENDINI, H. N.; SANCHES, I. D. A.; KÖRTING, T. S.; FONSECA, L. M. G.; LUIZ, A. J. B.; FORMAGGIO, A. R. Using Landsat 8 Image Time Series for

2.1. Introduction

Agriculture has significant participation in the Brazilian economy; it is the main responsible for the positive trade balance of the country. Given the high availability of arable land, and taking into account the growing demand for food in the world, Brazil has been consolidating as a big player on the world agricultural scenario. On the other hand, important biomes in the country have essential ecological functions for humanity, and therefore must be preserved. In this way, agricultural mapping is strategic for offering subsidies to generate information as estimates of acreage and production, as well as allowing a better understanding of the distribution of croplands, and its impact on the environment. Remote sensing is an important tool for its ability to generate information on large scale in a cost-effective way. With advances in data processing and storage technologies, and the availability of long-term image series, remote sensing is undergoing a paradigm shift, in which time series techniques stand out for allowing taking into account seasonal variations of the analysed target. This approach is useful for vegetation studies, especially in agricultural areas, since vegetation cover is quite dynamic in time, and the ability to capture these variations is essential to discriminate different types of crops, through its phenological characteristics. Until now, only time series of vegetation indices like NDVI or EVI using MODIS data were well explored for this purpose (SAKAMOTO et al., 2005; WARDLOW et al., 2007; ESQUERDO et al., 2011; ARVOR et al., 2011; RISSO et al., 2012; KÖRTING, 2013; COUTINHO et al., 2013; BORGES; SANO, 2014; TOMÁS et al., 2015; NEVES et al., 2016). However, there is still a demand for more detailed maps, which are possible from time series with finer spatial resolutions, such as Landsat-like images (ZHENG et al., 2015; PEÑA et al., 2015; PAN et al., 2015). In this context, the objective of this study is to employ phenological metrics obtained by

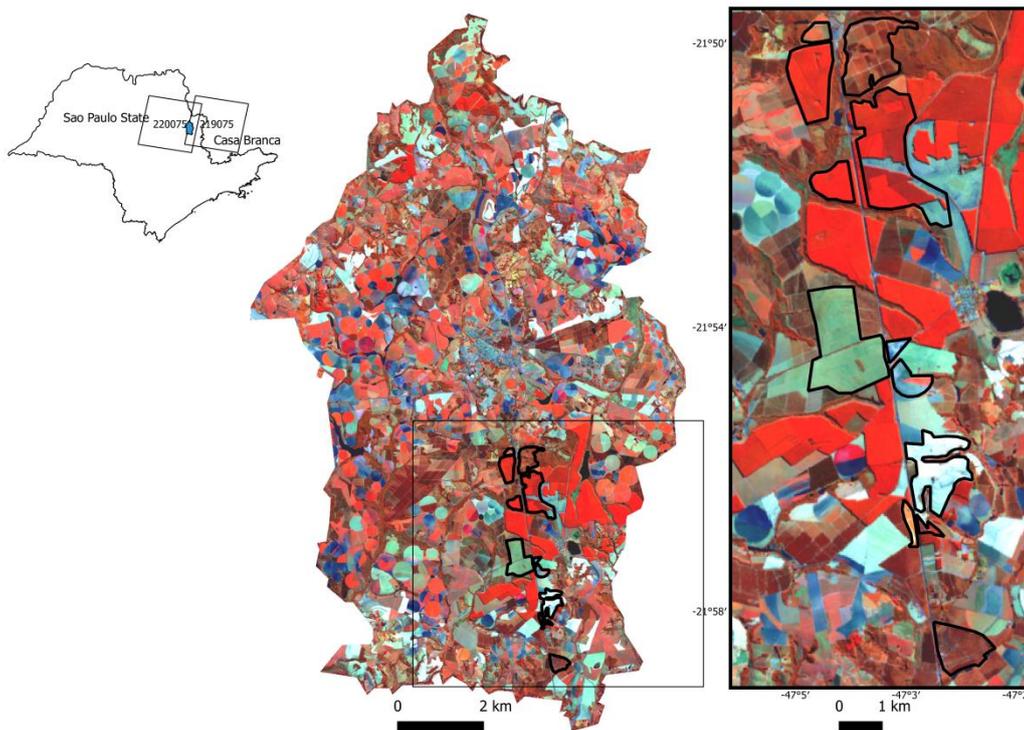
time series of vegetation indices from Landsat-8 OLI for the classification of agricultural land use in the municipality of Casa Branca, located in the Cerrado biome in the state of São Paulo, Brazil.

2.2. Materials and Methods

2.2.1. Study area

We conducted our study in the south of Casa Branca municipality, in state of São Paulo, Brazil (Figure 2.1). Such region is located in an overlapping area of two adjacent Landsat-8 scenes (219/75 and 220/75), providing a temporal resolution of 8 days as suggested by Luiz et al. (LUIZ et al., 2015).

Figure 2.1 – False color (bands 5, 6 and 4 in red, green and blue respectively) Landsat imagery of the study area (note the overlapping area of two adjacent scenes 219/75 and 220/75).



Casa Branca city has a tropical wet and dry climate (Aw, according to the Köppen Climate Classification System) with average annual temperature of 21.5° C and a seasonal rainfall pattern with most rainfall occurring from

October to March. The average annual precipitation is 109.18 mm (CEPAGRI/UNICAMP, 2016).

In this region, farmers grow a variety of crops along all the year. Major field crops in this area are sugarcane, corn, bean, potato, soybean, peanuts, sorghum and cassava. There is also a significant production of citrus, mango, coffee, rubber and eucalyptus (IBGE, 2016). Farmers grow crops in double cropping systems and even in triple cropping systems, mainly on the irrigated areas. The usual planting crop dates is from October to December and harvesting from February to April, but planting crops in late fall (May – July) and harvest in the next spring is also observed, especially in the irrigated areas.

2.2.2. Remotely sensed data

A total of 46 scenes of Landsat-8 OLI (Path 219 and 220/ Row 75) between July 2014 (Day of year – DOY 206) and July 2015 (DOY 209) were obtained at the US Geological Survey platform (USGS) Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA), which are provided with geometric correction level 1 (L1T) and converted to surface reflectance by the algorithm of atmospheric correction LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing) (MASEK et al., 2006). We then generate EVI (Equation 2.1) (HUETE et al., 2002) layers for each image.

$$EVI = \frac{NIR-R}{NIR+(6*R-7.5*B)+1} \quad (2.1)$$

where NIR = near-infrared surface reflectance band (band 5)

R = red surface reflectance band (band 4)

B = blue surface reflectance band (band 2)

EVI was chosen because it's capacity for highlighting the spectral response related to green vegetation canopy, taking into account the background and atmospheric effects.

2.2.3. Removal of outliers and null values

No cloud free images were identified in November, February and March. We applied a simple linear interpolation for removal of outliers and null values (Equation 2.2).

$$x_t = \frac{x_{t-1} + x_{t+1}}{2} \quad (2.2)$$

where x_t = an EVI observation of the time series at time t

x_{t-1} = an observation at time $t - 1$

x_{t+1} = at time $t + 1$

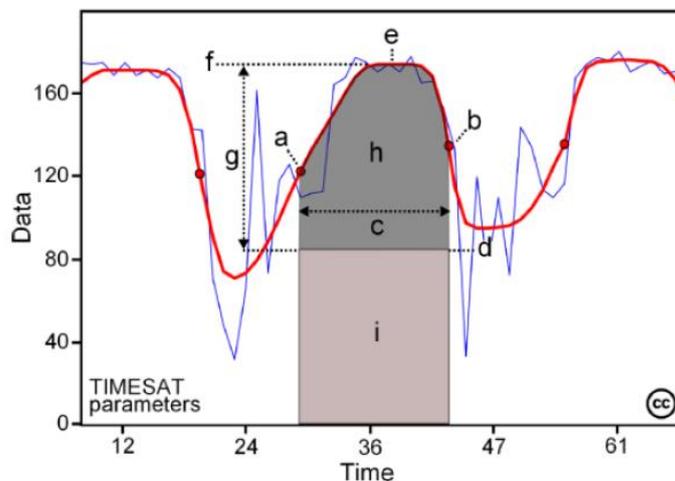
x_t observation is replaced by the average of x_{t-1} and x_{t+1} if x_t is less than 0,01. This method, however, is not capable of removing consecutive outliers. After this, we also smoothed the time series considering the double logistic filter (ZHANG et al., 2003; JÖNSSON; EKLUNDH, 2004). This function is recommended for smoothing image time series on cropland areas in the Brazilian Cerrado (BORGES; SANO, 2014).

2.2.4. Seasonal data extraction

Phenological metrics in EVI time series were obtained by the TIMESAT software (3.2 version), where seasonal data are extracted for each of the growing seasons of the central year (Figure 2.2). During a period of n years there may be $n - 1$ full seasons together with two fractions of a season in the beginning and end of the time series. So, to extract seasonality parameters from one year of data the time series has been duplicated to span three years, as recommended by Jönsson and Eklundh (2015). Figure 2.2 shows the schema of the seasonality parameters generated by TIMESAT. In this work we assume that the seasonality parameters are the same of phenological metrics. The time for the beginning of season (a), or start of the season (sos), and the end of season (eos) (b) is the time for which the left and right edge, respectively, has increased to a defined level (often a certain fraction of the seasonal amplitude) measured from the minimum level on the corresponding side. The length of the season (c) is

the time from the start to the end of the season. Base value (d) is given as the average of the left and right minimum values. The middle of season (e) is computed as the mean value of the times for which, respectively, the left edge has increased to the 80 % level and the right edge has decreased to the 80 % level.

Figure 2.2 – Some of the seasonality parameters generated by TIMESAT: (a) beginning of season, (b) end of season, (c) length of season, (d) base value, (e) time of middle of season, (f) maximum value, (g) amplitude, (h) small integrated value, (h+i) large integrated value.



Source: Jönsson and Eklundh (2015).

The maximum value (f), or the peak of phenological cycle, is the largest data value for the fitted function during the season. The seasonal amplitude (g) is difference between the maximum value and the base level. The left derivative is calculated as the ratio of the difference between the left 20% and 80% levels and the corresponding time difference, and the right derivative (i.e. the rate of decrease at the end of the season) is the absolute value of the ratio of the difference between the right 20% and 80% levels and the corresponding time difference. The rate of decrease is thus given as a positive quantity. Large seasonal integral (h+i) is integral of the function describing the season from the season start to the season end. And the small seasonal integral (h) is the integral of the difference between the function describing the season and the base level from season start to

season end (JÖNSSON; EKLUNDH, 2015). For more details see Jönsson and Eklundh (2002, 2004).

2.2.5. Training and validation data

Field campaigns were carried out for collecting training samples, totalling 16,243 pixels. The classes considered were Annual Crop (potato and corn, on a double crop system), Semi-Perennial Crop (sugarcane and cassava), Perennial Crop (citrus, mango and rubber) and Planted Forestry (eucalyptus). The polygons that represent each class were directly digitized over the Landsat-8 images.

2.2.6. Random Forest classification

The variables obtained by TIMESAT were subjected to data mining using the Random Forest algorithm (BREIMAN, 2001). This algorithm is a classification technique in which the data set is randomly divided in several subsets of smaller size by means of applying bootstrap, and from each subset is developed a decision tree. All trees contribute to the classification of the object under study, by voting on which class the target attribute must belong. Random Forest algorithm has been widely used in remote sensing (CLARK et al, 2010; MÜLLER et al, 2015; PEÑA et al, 2015) because of its advantages in efficiently handle large databases, providing estimates on most relevant variables, and allowing the identification of outliers (RODRIGUEZ-GALIANO et al., 2012).

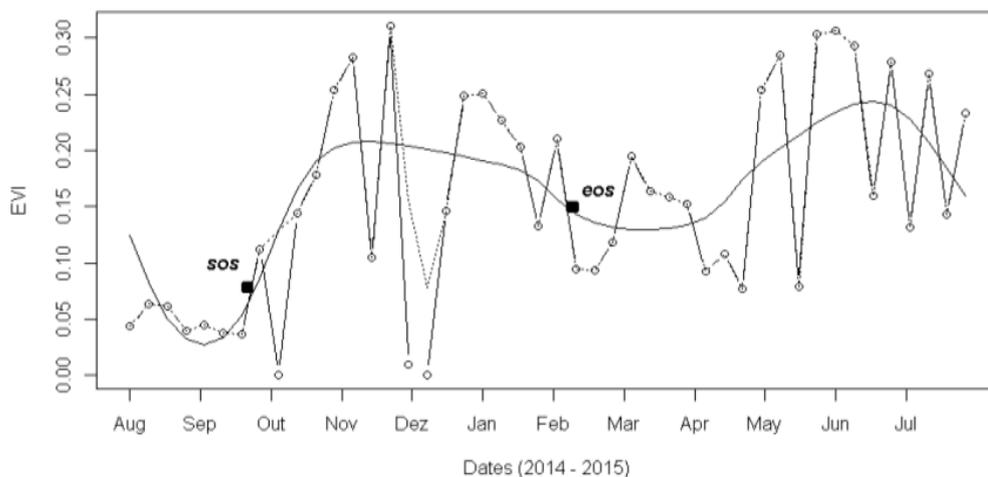
2.3. Results and Discussion

2.3.1. Crop phenological curves and TIMESAT features

Initially, we have done an evaluation of the time series spectral profiles of different types of targets. The phenological features were characterized by median values observed in all the samples for each class. Figures 2.3 to 2.8 shows the time series spectral profiles of the different targets and the results of the two approaches for noise removal, where the dotted line represents the time series outliers and null values removed (Equation 2.2) and the thick line the double logistic filtered time series. The samples of annual crop were

located into two different farms. However, both were on pivot center irrigation systems. We observed in the field work that in one of them it was a triple crop system (potato, bean and corn), where potatoes were harvested between July and the mid-August. This was followed by corn planting on September, and harvested between January and February. We also observed a short-cycle bean between February and March, followed by corn planting from May. On the most time series samples of this farm only one season was modelled by TIMESAT, corresponding to corn planting in late September (Day of the Year – DOY 265) and harvested in early February (DOY 41). The middle of season was observed in mid-November (DOY 325) (Figure 2.3).

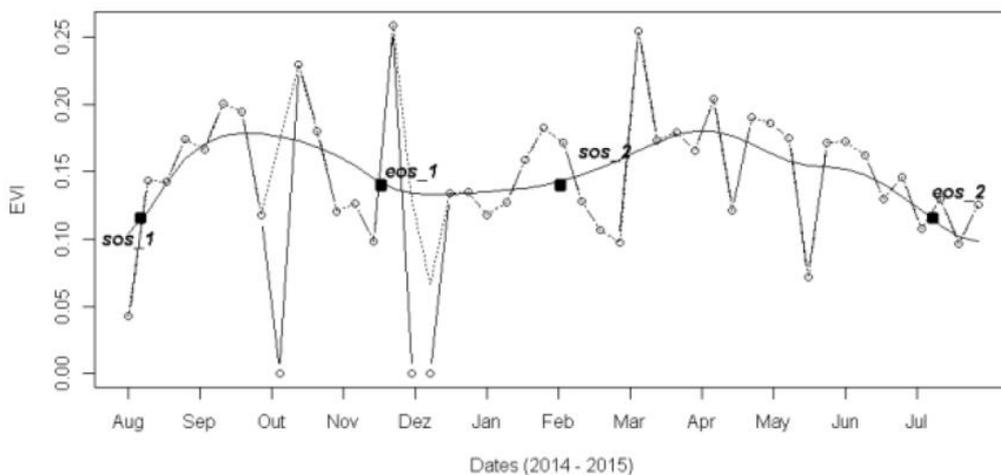
Figure 2.3 – EVI time series spectral profile for an annual crop sample located into a triple cropping system of potato, bean and corn. Dotted line represents the time series outliers and null values removed and the thick line the double logistic filtered time series.



It was not possible to identify potatoes seasons because it was not completely inside the time series span. It was also not possible to capture the bean season, as well the late fall corn season. We hypothesize that that is due to the short cycle of bean and mainly for the high intensity of cloud contaminated pixels on this period. Although potatoes were harvested later in the other farm, it was not possible to detect their season. Only the corn season was detected, starting between late November and mid-December (DOY 347) and the end of season it was on mid-April (DOY 102). The middle of season was observed in late January (DOY 26). On the semi-perennial

crop samples corresponding to cassava farms it was observed that in 77% of pixels was detected only one season. The start of season was in late September (DOY 270) and the middle of season was detected in mid-February (DOY 45). Based on field observations, we hypothesize that this season can be associated to the end of cassava senescence, characterized by the start of a new greenness period. The end of season coincided with the end of harvest, when the area becomes complete cover by weeds. In the other 23% of pixels it was possible to detected two seasons, where the second season began between early December (DOY 347) and early April (DOY 60) and ending on late June (DOY 179). This indicates a predominant phenological behavior of weed, which have a great biomass gain after a rainfall period (Figure 2.4).

Figure 2.4 – EVI time series spectral profile for a semi-perennial crop sample of cassava.



We observed that sugarcane (semi-perennial crop class) was planted in June. However, the complete ground cover occurs only in November, when there were almost no cloud-free images. Therefore, the start of season was identified in early December (DOY 338) and the end of season in mid-May (DOY 135), corresponding to the harvest (Figure 2.5).

Figure 2.5 - EVI time series spectral profile for a semi-perennial crop sample of sugarcane.

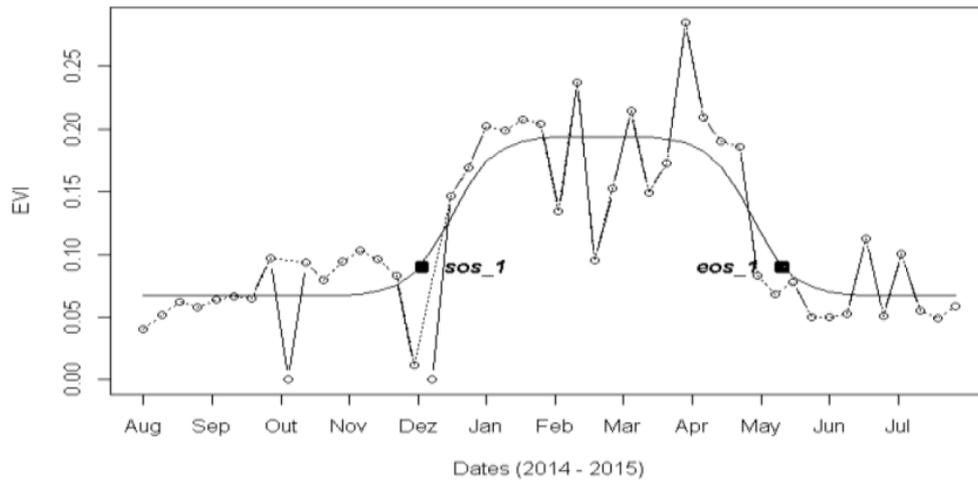
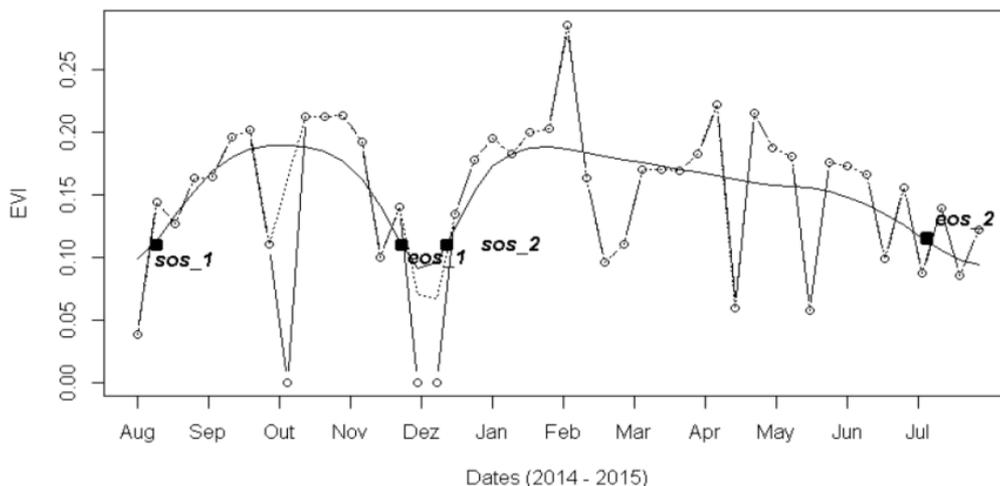


Figure 2.6 shows the time series spectral profile of a perennial crop sample, corresponding to a rubber tree farm. We can see two seasons' detection, where on the first the start of season was in the middle of August (DOY 229) and the end in December (DOY 352).

Figure 2.6 - EVI time series spectral profile for a perennial crop sample of rubber trees.

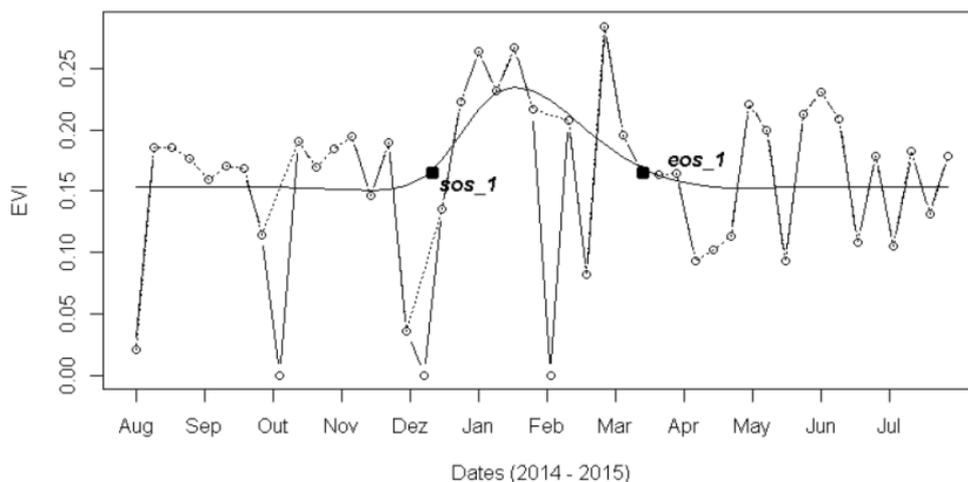


The second season began between December and February (DOY 29) and the end of season occurred in late July (DOY 190). This occurred in 61% of the pixels inside this farm. This behavior was not expected for rubber trees, considered a perennial crop. However, this can be explained by the high frequency of cloud contaminated images between November and February. This noise could not be corrected by the smoothing and noise removal employed in this study. Additionally, the rubber trees phenology on this

region has a senescence period between July and August, followed by a greenness period after September. The 49% remaining pixels located in this farm presents time series with a single season between August (sos = DOY 231) and late July (eos = DOY 173).

Time series of pixels located on one of the two citrus farms that represents the perennial crops class presents a subtle difference on phenological features. It was detected one short season starting on early December (DOY 343) and ending on mid-March (DOY 73). The other citrus farm, as well on a mango farm, it was observed that the model detected a longer season, between December (DOY 338) and June (DOY 171). Among the other differences it was also observed bigger amplitudes on the first citrus farm (median = 0.073 EVI) than the other perennial crops farms (median = 0.059 and 0.055 for citrus and mango, respectively). Furthermore, the middle of season on the first citrus farm occurred in January (DOY 18) and the EVI peak found was 0.22 EVI, while in the other areas the middle occurred in March (DOY 61 and 85, for citrus and mango), with a 0.18 EVI peak.

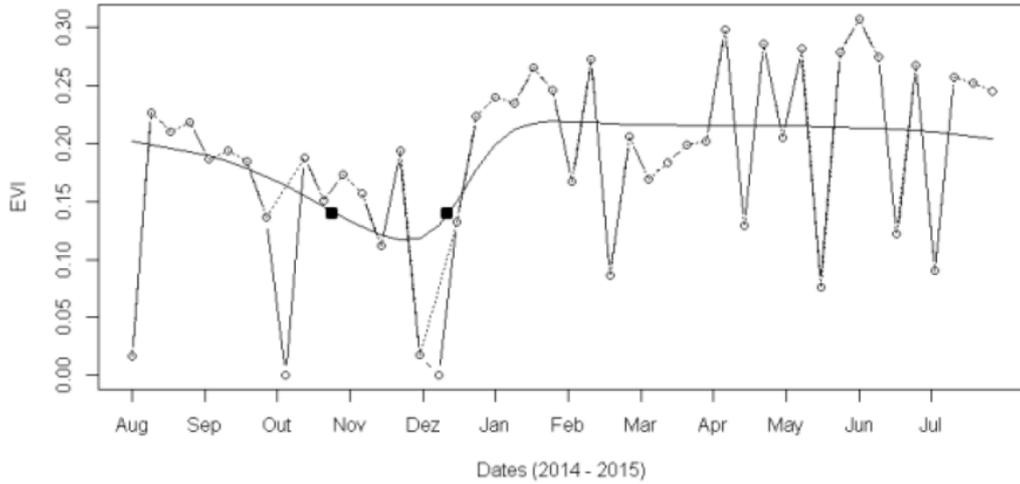
Figure 2.7 - EVI time series spectral profile for perennial crop sample of one of the citrus farm (similar to mango).



Finally, on the eucalyptus planted forests, 60% of pixels showed time series where it was not possible to detect any season, so the values of

phenological features were zero. In the other 40% pixels' time series it was detected one season that began on early February (DOY 41) and finished on July (DOY 185), as showed in the Figure 2.8.

Figure 2.8 – EVI time series spectral profile for a planted forest sample of eucalyptus.



2.3.2. Classification performance

A 10 fold cross-validation technique was applied in the data set and 15,499 of the 16,243 pixels were classified correctly, resulting in an accuracy of 95.42% and Kappa 0.9284. The confusion matrix with the errors of commission and of omission is presented on Table 1.1.

Table 2.1 – Confusion matrix: (a) Perennial crop, (b) Annual crop, (c) Planted Forest and (d) Semi-perennial crop. EC = errors of commission; EO = errors of omission.

Classification	Reference					Total	EC%
	a	b	c	d			
a	6563	40	184	127	6914	5.1	
b	3	717	7	10	737	2.7	
c	265	17	6303	40	6625	4.9	
d	26	21	4	1916	1967	2.6	
Total	6857	795	6498	2093	16243		
EO%	4.3	9.8	4	2			

Most misclassification happened between perennial crop and planted forest and between perennial crop and semi-perennial crop, due to the rubber trees spectral similarity with the two other classes. The major omission

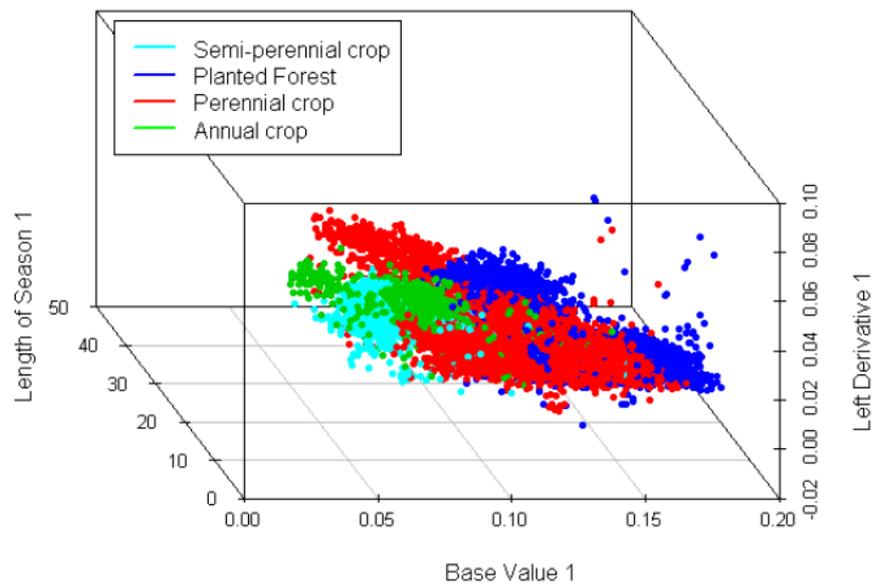
errors were observed on annual crop and perennial crop classes. We observed that 5.03% of pixels in the annual crop class were omitted, being wrongly classified as perennial crop.

As we discussed before, a significant number of samples of annual crop class presented time-series where it was detected only one season (between December and March) with a 0.22 EVI peak. This is similar to some annual crop class pixels, what should be contributed for the omission of these samples. With respect to the semi-perennial crop class, major confusion was observed on the perennial crop, where around 6.1% of pixels were wrongly classified.

The major commission errors were founded on the perennial crop (5.1%) and planted forest (4.9%) classes. An inclusion of 2.6% of pixels of the planted forest and 1.84% of semi-perennial classes was also observed in the perennial crop class. Around 4% of pixels of perennial agriculture were included on planted forest class.

However, we considered that the explored features configured a good separability between classes. Figure 2.9 shows the 3D-scatterplot with length of season, base value and left derivative features of the first seasons.

Figure 2.9 – 3D-scatterplot of length of season, base value and left derivative features of the first seasons. The zero values were removed for best visualization.



The left derivative is the rate of increase at the beginning of the season. Higher values of left derivative indicate that there is a fast rise of the greenness.

We can see in Figure 2.9 that lower values of length of season and base value were found for annual crop samples. High values of left derivative were observed for this class (> 0.035 EVI). The low values of base value (< 0.08) indicate that the EVI values before and right after the season were also low. This is expected for cropland areas. Besides the omitted zero values of planted forest class samples, there are medium and high base values (0.13 – 0.17 EVI), medium values of length of season and left derivative (0.013 – 0.029 EVI). Pixels of the semi-perennial class presented low values of base value and left derivative, and medium values of length of season. In the perennial crop class, it is possible to visualize two clusters on the 3D-scatterplot, determined by length of season values. Both left derivative and base values were medium. By investigating the errors spatial distribution, it also seems that it may be associated with heterogeneity of the monitored areas, resulting in spectral mixing. In addition, for the extraction of the phenological parameters using TIMESAT, a generic model was considered for all classes, and the small number of samples may also have contributed to errors. The definition of time series span it is also very important criteria to take into account.

2.4. Final Considerations

These preliminary results are encouraging and demonstrate the potential of phenological parameters obtained from time series of OLI vegetation indices for agricultural land use classification. We showed the potential use of higher temporal resolution Landsat-like images for crop mapping, what will soon become reality, once Landsat, Sentinel-2A and CBERS-4 data can be combined to generate consistent time series to produce land use maps. Further analyses are needed to apply this approach in large areas and to test different time span and different vegetation indices.

3 ASSESMENT OF A MULTI-SENSOR APPROACH FOR NOISE REMOVAL ON LANDSAT-8 OLI TIME SERIES USING CBERS-4 MUX

DATA TO IMPROVE CROP CLASSIFICATION BASED ON PHENOLOGICAL PARAMETERS⁴

3.1. Introduction

Given the large availability of arable land, and the growing demand for food in the world, Brazil has consolidated as a big player on the global agricultural scene. Remote sensing is an important tool used within agriculture, regarding its ability to generate information on large scale in a cost-effective manner. Therefore, agricultural mapping has become strategic since it provides better understanding of cropland distribution, and its impact on the environment. With advances in data processing and storage technologies as well as the availability of consistent and continuous long-term image series, remote sensing is undergoing a paradigm shift. Time series techniques stand out for allowing seasonal variation accounts of the analysed target.

Although the use of time series for cropland classification has been well explored using MODIS sensor (SAKAMOTO et al., 2005; ARVOR et al., 2011; KÖRTING, 2013; RISSO et al., 2012; BORGES; SANO, 2014; NEVES et al., 2016), there is still a demand for more detailed maps, which are made possible from time series with finer spatial resolutions, such as Landsat-like images (ZHENG et al., 2015; PEÑA et al., 2015; PAN et al., 2015; BENDINI et al., 2016). As the temporal resolution of Landsat-like satellites is still low (e.g. 16 days), an open question in the scientific literature is about how to deal with the noise in the time series. The noise is characterized by negative outliers, which possibly result from either cloud cover or cloud shadow contamination or atmospheric scattering. To deal

⁴ This chapter is based on the paper: BENDINI, H. N.; FONSECA, L. M. G.; KÖRTING, T. S.; MARUJO, R. F. B.; SANCHES, I. D. A.; ARCANJO, J. S. Assesment Of A Multi-Sensor Approach for Noise Removal on Landsat-8 OLI Time Series Using CBERS-4 Mux Data to Improve Crop Classification Based on Phenological Parameters. *Brazilian Journal of Cartography* (2017), N° 69/5, Special Issue GEOINFO 2017: 947-957 Brazilian Society of Cartography, Geodesy, Photogrammetry and Remote Sense ISSN: 1808-0936.

with it, some approaches were developed, which include cloud and cloud shadow flags generated from the Automated Cloud Cover Assessment (ACCA) algorithm (IRISH et al., 2006) and Fmask algorithm (ZHU; WOODCOCK, 2012). However, both ACCA and Fmask sometimes fail to detect thin clouds i.e. cirrus and the edges of cumulus clouds (LYMBURNER et al., 2016). In this case, methods based on thresholds (HAMUNYELA et al., 2013; BENDINI et al., 2016; LYMBURNER et al., 2016) or smoothers (PAN et al., 2015) can be used.

There is also the possibility to take advantage of multi-sensor data, considering the large amount of available remote sensing images. In a previous investigation, we showed the potential use of higher temporal resolution Landsat-like images for crop mapping (BENDINI et al., 2017b). Recently the China Brazil Earth Resources Satellite (CBERS) program launched CBERS-4 that carries in the payload module, among others, the Multispectral Camera (MUX).

In this work, we investigated an algorithm for noise removal in Landsat-8 OLI time series using CBERS-4 MUX data to improve a crop classification method based on phenological features.

3.2. Methodology

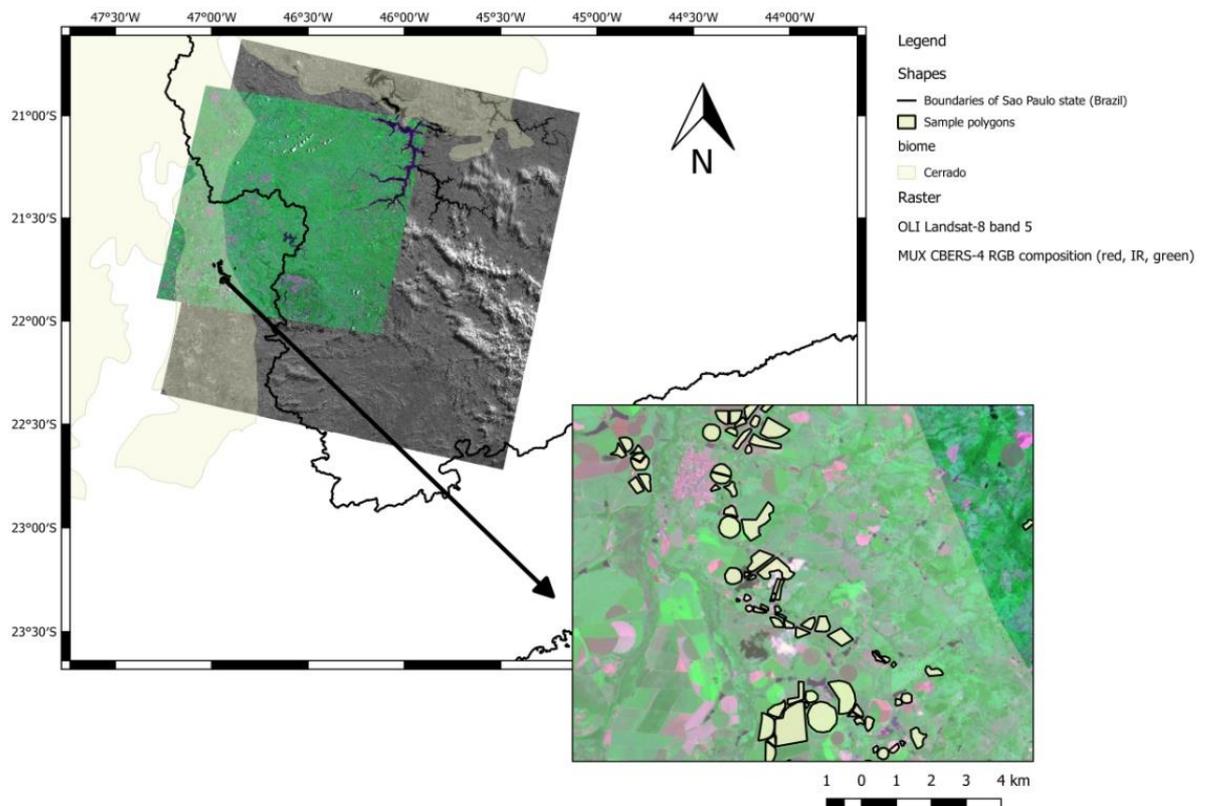
This section describes the methodology and is divided in a description of the study area, characteristics of data used on this work, the correlation analysis between both sensors to deal with the spectral differences, how the integrated time series were constructed, the attribute extraction and the classification.

3.2.1. Study area

The study area is situated in Itobi municipality in São Paulo state (southeast of Brazil), in a Cerrado biome region (Figure 3.1). As the focus is on croplands, we selected a region of interest where the main land cover is agriculture, silviculture, and pasture. In this region, farmers grow a variety of crops throughout the year. Major field crops in this area are sugarcane,

corn, bean, potato, soybean, sugar beet and onions. There is also production of mango, avocado and eucalyptus. Farmers grow crops in double cropping systems and even in triple cropping systems, mainly in irrigated areas. The usual planting for summer crops occurs from October to December and harvesting from February to April. We also observed crops growing in late fall (May – July) and harvesting in the next spring especially in irrigated areas.

Figure 3.1 – Location of the study area in São Paulo state, Brazil.



3.2.2. Remotely sensed data

A total of 24 scenes of Landsat-8 OLI (WRS 2 – Worldwide Reference System 2, Path/Row 219/75) between August 2015 and August 2016 were processed to Level 1 Terrain Corrected (L1T) by the USGS EROS Science Processing Architecture (ESPA) (DEVRIES et al. 2015; DEVRIES et al. 2015a).

Landsat-8 data were corrected using L8SR, a newly developed algorithm that takes advantage of Landsat-8 new sensor characteristics (U.S. GEOLOGICAL SURVEY, 2015; VERMOTE, 2016). Cloud (pixel value 4), cloud shadow (pixel value 2), snow (pixel value 3), water (pixel value 1) and clear (pixel value 0) masks were provided for Landsat-8 data using Cfmask, a C implementation of the Fmask algorithm (ZHU; WOODCOCK, 2012; ZHU et al., 2015).

CBERS-4 MUX imagery has been provided by INPE. A total of 9 scenes of CBERS-4 MUX (CBERS WRS Path/Row 155/124) were acquired in the same period. Table 3.1 shows OLI and MUX images available from August 2015 to August 2016.

The images were radiometrically corrected and geometrically adjusted and refined by using control points and the SRTM 30m v. 2.1 digital elevation model (DEM) (Level 4).

Table 3.1 – Availability of Landsat-8 (Path/Row 219/75) and CBERS-4 (Path/Row 155/124) imagery from August 2015 to August 2016.

Month/Year	Sensor	Acquisition dates (day of year)	Number of scenes
Aug – Dec 2015	OLI	218, 234, 250, 266, 282, 298, 314, 330, 346, 362	10
	MUX	215, 241, 267, 345	4
Jan – Aug 2016	OLI	13, 29, 45, 61, 77, 93, 109, 125, 141, 157, 173, 189, 205, 237	14
	MUX	32, 110, 162, 188, 240	5

The atmospheric correction was proceeded using the 6S model (Second Simulation of a Satellite Signal in the Solar Spectrum) (VERMOTE et al. 1997). For the MUX imagery, the cloud cover for the region of interest was visually assessed. The specification of the Landsat-8 OLI and CBERS-4 MUX spectral bands used in this can be seen on Table 3.2.

Table 3.2 – Spectral Band Specifications for Landsat-8 OLI and CBERS-4 MUX.

Band	Landsat-8 OLI (μm)	CBERS-4 MUX (μm)
Blue	B2: 0.45 – 0.51	B5: 0.45 – 0.52
Green	B3: 0.53 – 0.59	B6 0.52 – 0.59
Red	B4: 0.64 – 0.67	B7: 0.63 – 0.69
Near Infrared (NIR)	B5: 0.85 – 0.88	B8: 0.77 – 0.89

The greatest difference in spectral bandwidths between the two sensors are on the NIR band, but there are also significant differences in spectral response function (SRF) profiles between corresponding CBERS-4 MUX and Landsat-8 OLI spectral bands (PINTO et al., 2016).

3.2.3. Correlation analysis between Landsat-8 OLI and CBERS-4 MUX

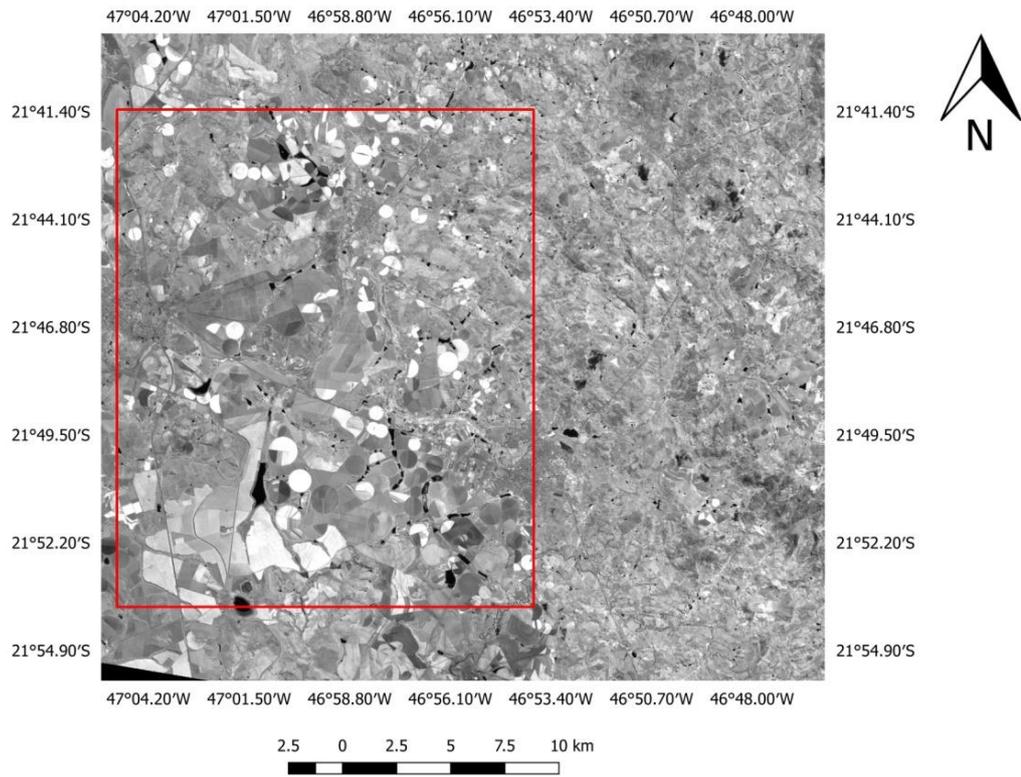
First we selected a pair of MUX and OLI images, considering the temporal proximity between them. The characteristics of the two images are shown in Table 3.3.

Table 3.3 – Characteristics of the pair of MUX and OLI images used for correlation analysis.

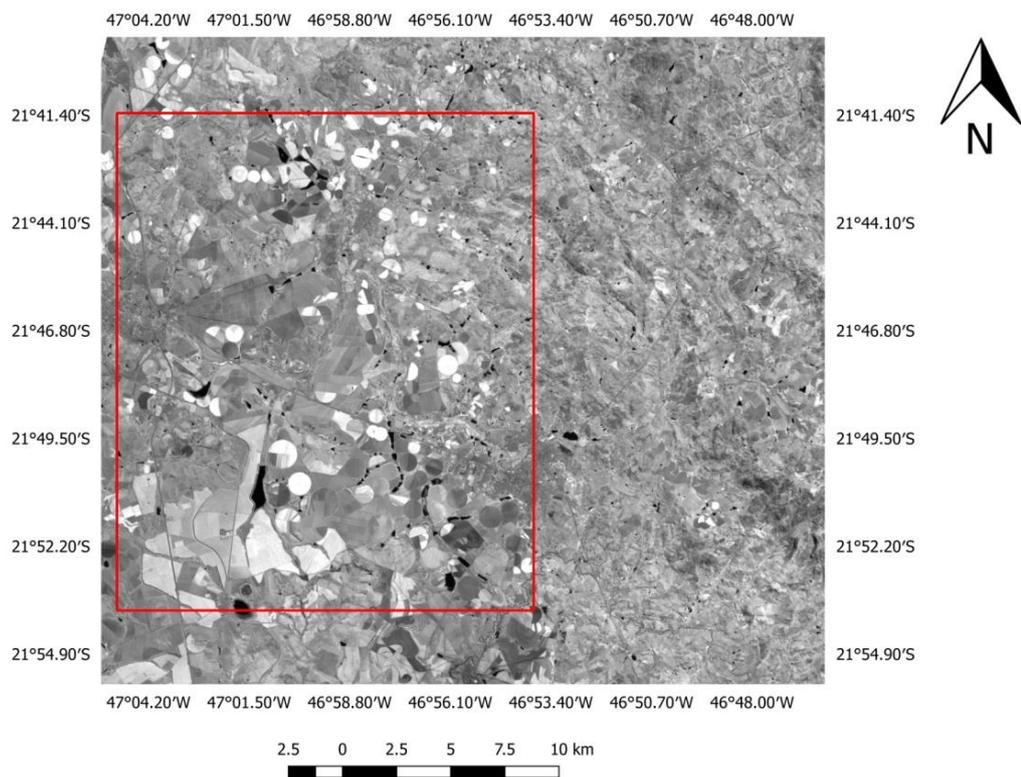
Satellite/Sensor	Landsat-8 OLI (μm)	CBERS-4/MUX
Date	06-Aug-15	04-Aug-15
Acquisition Time (UTC)	13:03:18	13:26:11
Path/Row	219/75	155/124
Sun elevation	40.61°	43.37°
Sun azimuth	41.58°	36.05°
Look Angle	NADIR	NADIR

Considering the spatial resolution difference between the images (30 meters for OLI and 20 meters for MUX), we resampled MUX images to 30 meters, using the nearest neighbour interpolation. To deal with cloud contamination, we used the Fmask image and visual assessment to crop a cloud free region on OLI and MUX surface reflectance images, respectively (Figure 3.2).

Figure 3.2 – Cropped images used on the correlation analysis. (a) LANDSAT-8 OLI EVI (06 August 2015) and (b) CBERS-4 MUX EVI (August 4th, 2015).



(a)



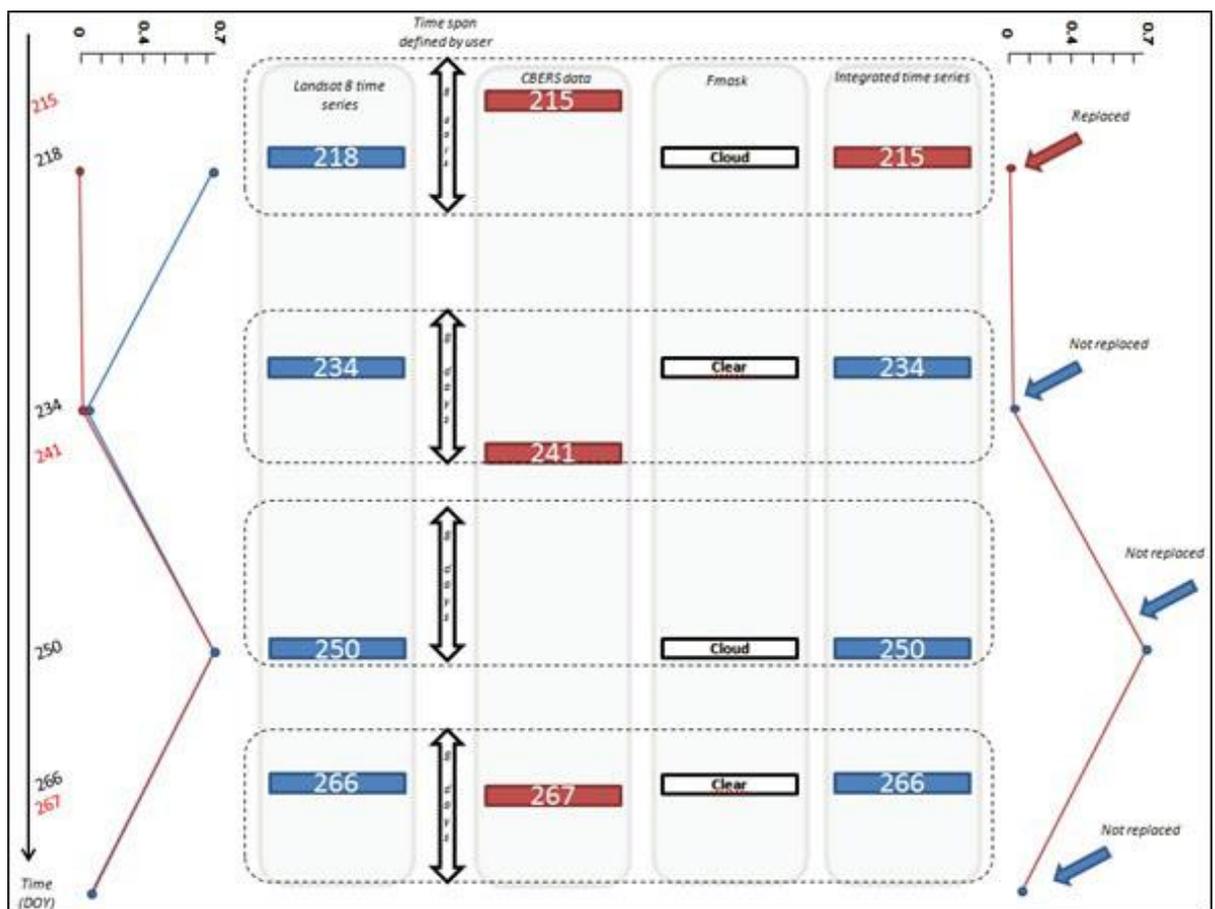
(b)

We analysed the correlations between cropped MUX and OLI images, for each selected vegetation index (EVI and NDVI). In order to determine an equation to predict OLI reflectance from MUX reflectance, linear regressions were built.

3.2.4. Building the multi-sensor time series

An algorithm was built to look for the nearest MUX image to each Landsat image, based on a user defined time span. In this case, we used time span of 8 days. Figure 3.3 shows a general scheme of the proposed method.

Figure 3.3 – General scheme of the methodology used to build the integrated time series. On the left, a time series of EVI (the red line is the predicted time series using the equation to predict OLI reflectance from MUX and the blue line is the original Landsat time series); the integrated time series is on the right, which points out the positions where the replacement has occurred.



After detecting the nearest MUX images for each Landsat image, the algorithm checks the pixels contaminated by cloud and cloud shadow in the Landsat time series using Fmask images. When a contaminated pixel is detected in the time series, it is replaced by a predicted OLI reflectance, if it is within the defined time window.

3.2.5. Filtering the time series

We also applied a combined filtering approach for noise removal to the Landsat time series in order to assess the classification improvement compared to that of integrated time series. The approach was put forth by interpolating noise values with the average between the nearest neighbours in time, considering the Fmask quality data (Equation 3.1) and negative outliers based on a threshold as recommended by Hamunyela et al. (2013) (Equation 3.2).

$$x_t = \frac{x_{t-1} + x_{t+1}}{2} \quad \{if \text{ fmask} = 2 \text{ OR } \text{ fmask} = 4\} \quad (3.1)$$

$$x_t = \frac{x_{t-1} + x_{t+1}}{2} \quad \{if \ x_t - x_{t-1} < -0.01x_{t-1} \ \& \ x_t - x_{t+1} < -0.01x_{t+1}\} \quad (3.2)$$

where x_t = an observation of the time series at time t

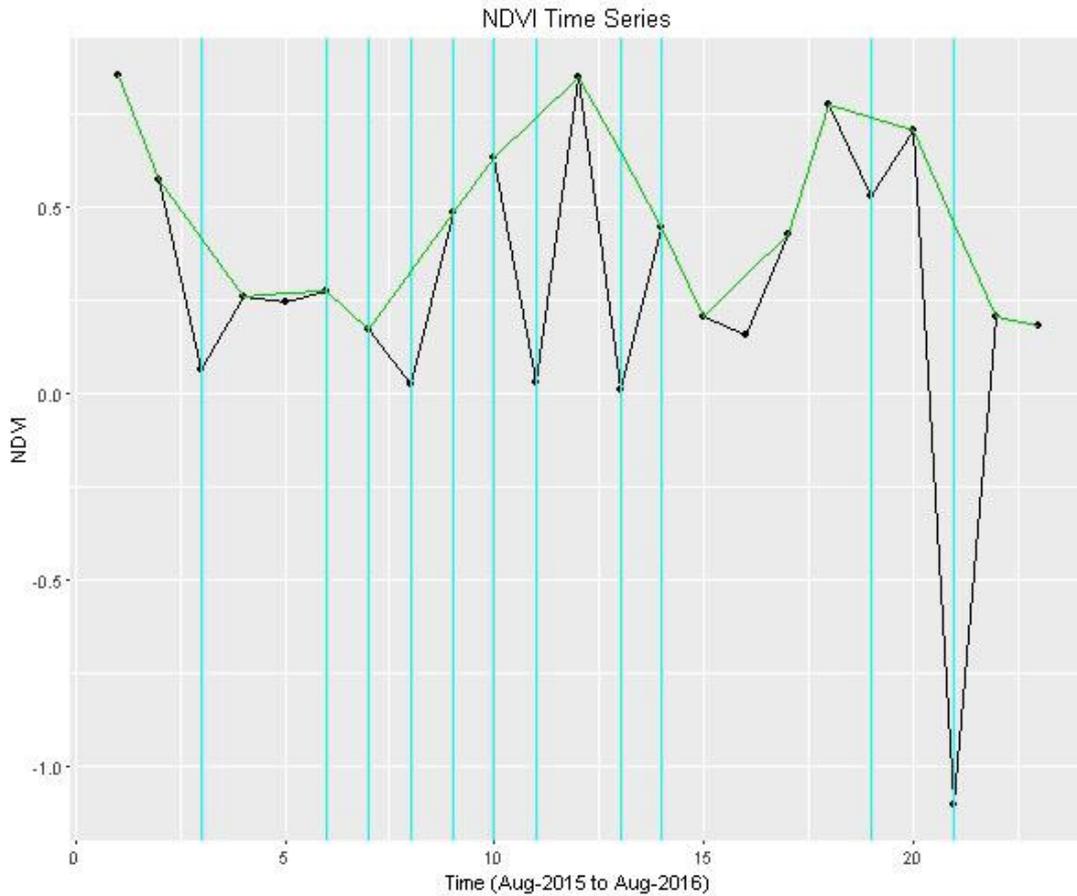
x_{t-1} = an observation at time $t - 1$

x_{t+1} = at time $t + 1$

observation x_t is replaced as an outlier with the average of x_{t-1} and x_{t+1} if the difference between x_t and x_{t-1} is less than 1% of x_{t-1} , and the difference between x_t and x_{t+1} if is less than -1% of x_{t+1} .

This method, however, is not capable of removing consecutive outliers. Figure 3.4 shows an example of how local outliers were removed from the NDVI and EVI time series.

Figure 3.4 – Example of how local outliers were removed from the NDVI time series. The cyan lines are the positions where cloud and cloud shadow were detected by Fmask. The black line is the integrated time series and the green line is the filtered integrated time series using Equation 3.1 followed by Equation 3.2.



3.2.6. Extracting phenological features

We selected 100 well-known polygon samples in the study area, considering the classes of annual agriculture (potato, corn, sugar beet, onion, bean and soybean), perennial agriculture (avocado and mango), semi-perennial agriculture (sugarcane), grassland and native forest.

We extracted NDVI and EVI time series of pixels from each sample polygon in the study area. Phenological metrics in time series were obtained by the TIMESAT v3.2 software (JÖNSSON; EKLUNDH, 2004), where seasonal data are extracted for each of the growing seasons of the central year (Figure 2.2). During a period of n years there may be $n - 1$ full seasons together with two fractions of a season in the beginning and end of the time

series. So, to extract seasonality parameters from one year of data, the time series has been duplicated to span three years, as recommended by Jönsson and Eklundh (2015). For the phenological metrics extraction, we smoothed (out) the time series considering the double logistic filter (ZHANG et al., 2003; JÖNSSON; EKLUNDH, 2004). This function is recommended for smoothing image time series on cropland areas in the Brazilian Cerrado (BORGES; SANO, 2014).

Figure 2.2 illustrates the schema of the seasonality parameters generated by TIMESAT. In this study, we assume that the seasonality parameters are the same as the phenological metrics.

3.2.7. Classification

We subjected the phenological metrics obtained on TIMESAT to data mining using the Random Forest (RF) algorithm (BREIMAN, 2001) considering each input: 1) Original Landsat EVI time series; 2) Filtered Landsat EVI time series; 3) Integrated EVI time series; 4) Filtered Integrated EVI time series; 5) Original Landsat NDVI time series, 6) Filtered Landsat NDVI time series, 7) Integrated NDVI time series and 8) Filtered Integrated NDVI time series.

The RF algorithm is a classification technique in which the data set is randomly divided into several subsets of smaller size, and from each subset a decision tree is built.

Random Forest algorithm has been widely used in remote sensing applications since it efficiently handles large databases (MÜLLER et al, 2015; PEÑA et al, 2015). Besides, it provides estimates on the most relevant variables, allowing the identification of outliers (RODRIGUEZ-GALIANO et al., 2012).

There was a total of 31 training pixels for annual agriculture, 15 pixels for perennial agriculture, 26 pixels for semi-perennial agriculture, 14 pixels for grassland and 14 pixels for native forest. The results were evaluated using confusion matrix index and global accuracy (WITTEN; FRANK; HALL,

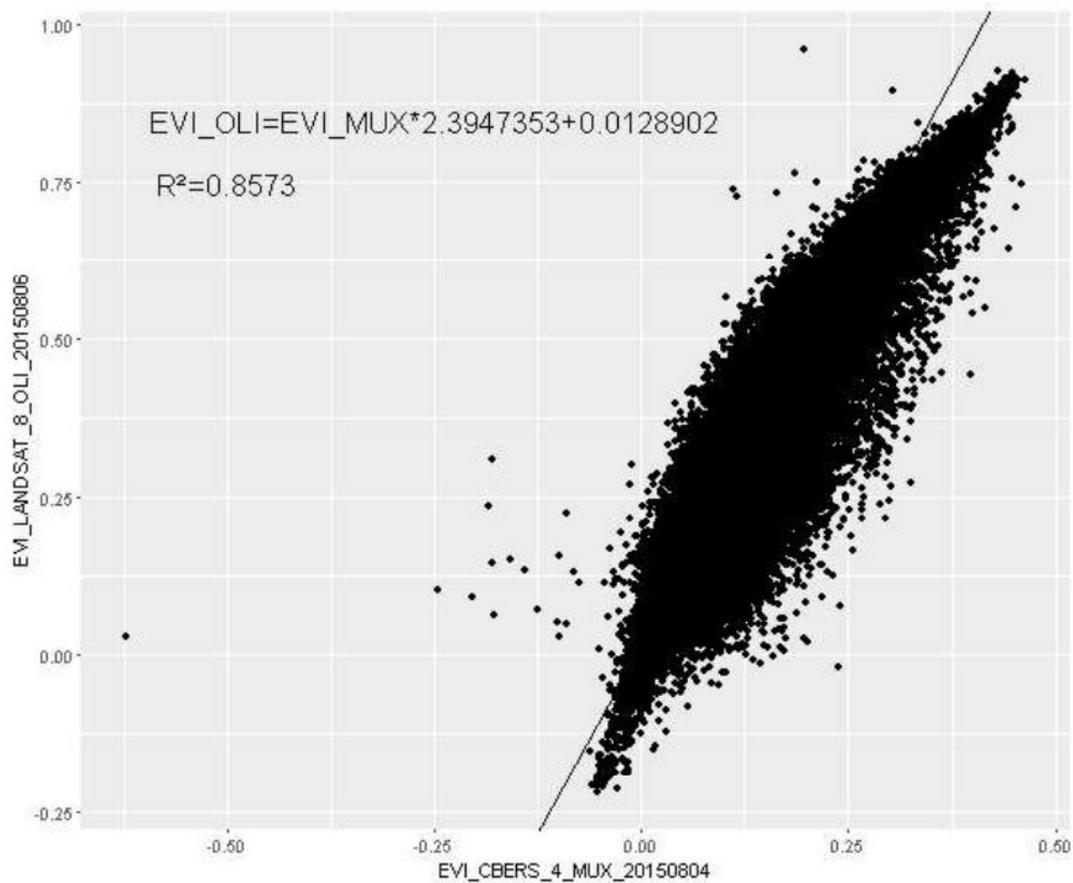
2011). The models were executed considering a 10-fold cross validation method. The classification results were obtained using the software package WEKA (HALL et al., 2009).

3.3. Results and Discussion

The results of the correlation analysis between the cropped images are shown in Figure 3.5, for each selected vegetation index: a) EVI and b) NDVI.

Figure 3.6 shows the results of different approaches for noise removal in an EVI time series.

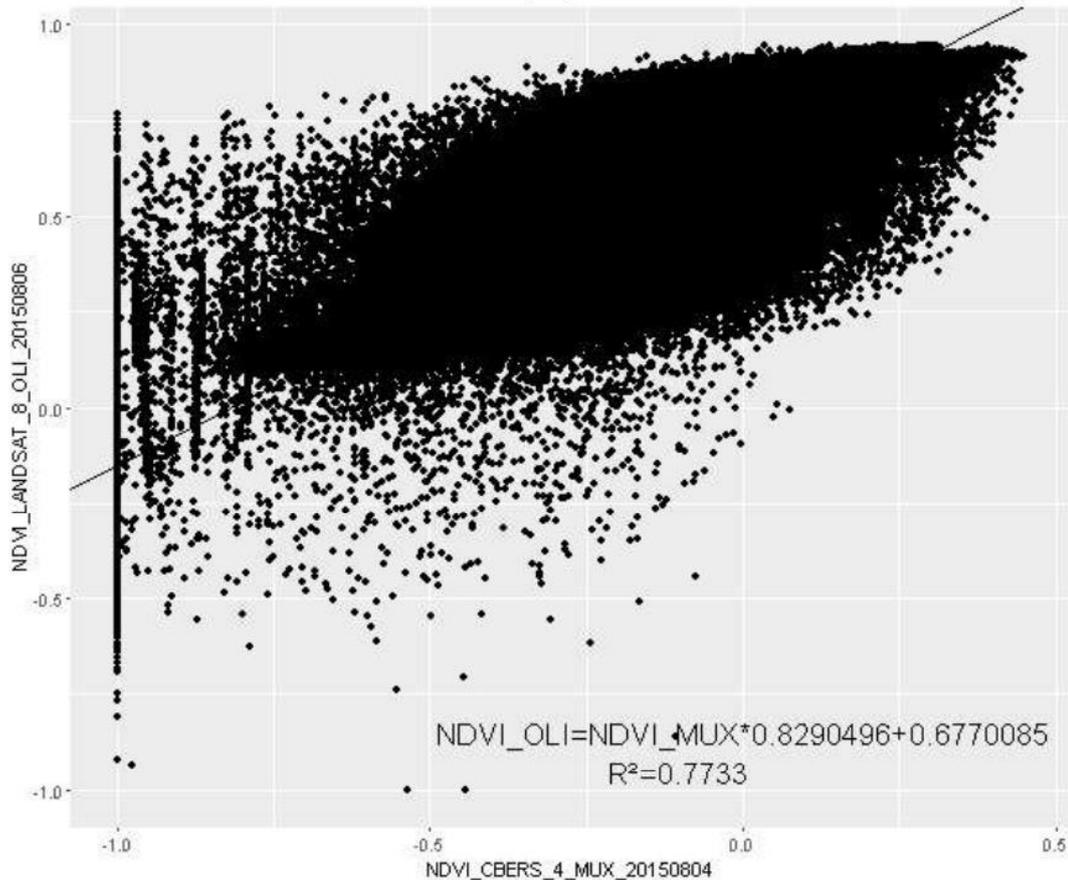
Figure 3.5 – Scatterplot of the pair of cropped images used to predict OLI reflectance from MUX reflectance. (a) EVI and (b) NDVI.



(a)

continue

Figure 3.5 – Conclusion.

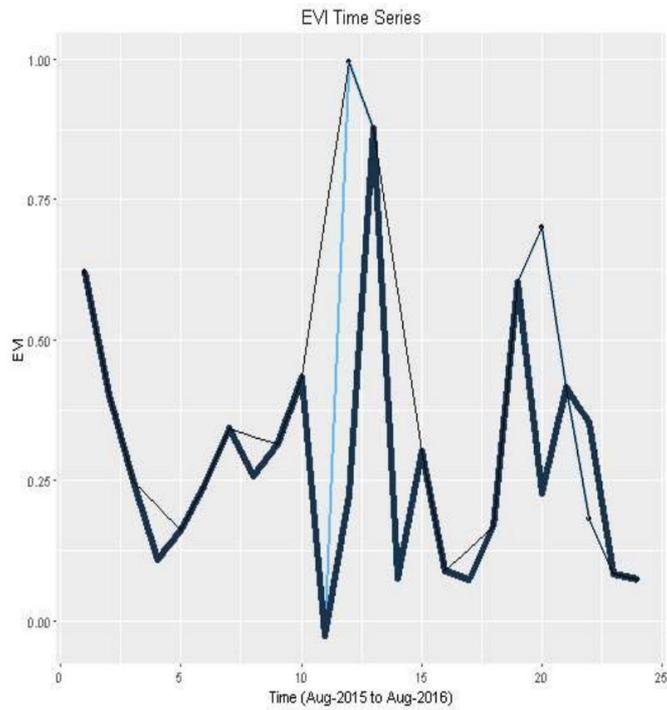


(b)

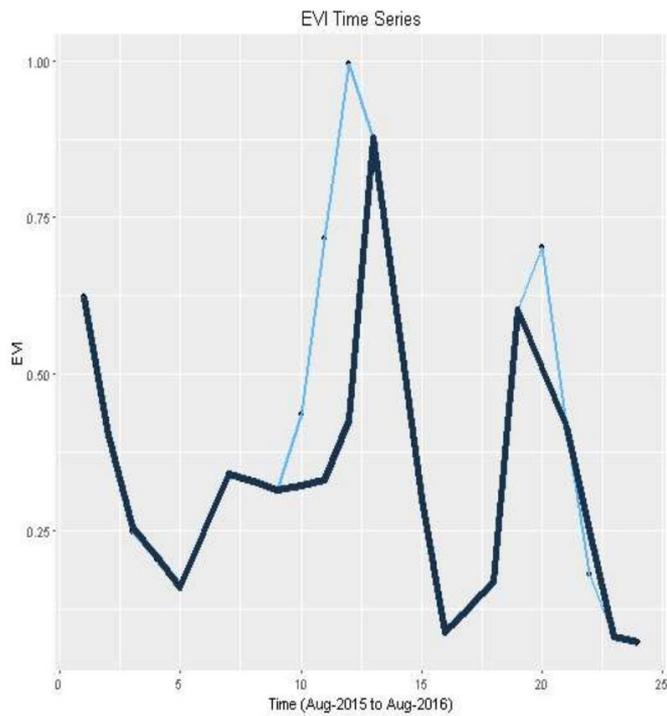
The linear regressions equations to predict OLI reflectance from MUX reflectance are also shown. The goodness of fit for EVI and NDVI are respectively 0.8573 and 0.7733. We can see that both EVI and NDVI values of Landsat-8 are higher than CBERS-4.

As we can see in Figure 3.6, the integrated time series can deal with noise, replacing cloud and cloud shadow contaminated pixels with clear pixels of MUX images, and allowing improvement of the time series according to the phenological behavior of the vegetation, which is significant regarding the capability of TIMESAT of extracting features.

Figure 3.6 – Results of different approaches for noise removal in an EVI time series. In (a) the black line is the original Landsat-8 time series, the blue line is the integrated time series and the black thin line is filtered integrated time series. In (b) the blue line is the filtered integrated and the black line is the Landsat-8 filtered time series.



(a)



(b)

Concerning the 100 analysed pixels, 11.96% of all cloud and cloud shadow contaminated observations were replaced using CBERS-4 MUX images.

A 10 fold cross-validation technique was applied using different training sets (Original Landsat EVI time series; Filtered Landsat EVI time series; Integrated EVI time series; Filtered Integrated EVI time series; Original Landsat NDVI time series, Filtered Landsat NDVI time series, Integrated NDVI time series and Filtered Integrated NDVI time series). The different data set classification accuracy is presented in Table 3.4. Concerning the NDVI time series, the multi-sensor approach accuracy was 64% using the filtering approach (Equation 3.1 and 3.2), as opposed to 68% without the filtering step.

Table 3.4 – Accuracy of classification for the different data set classifications.

Time series Data sets	NDVI	EVI
Integrated	68%	73%
Filtered Integrated	64%	78%
Filtered Landsat	70%	76%
Original Landsat	60%	70%
Integrated	68%	73%

However, when using only Landsat-8 data, the accuracy was 60%. But when combining the filtering approaches of Equation 3 and 4, the classification accuracy with Landsat-8 time series reached 70%.

In relation to EVI time series, multi-sensor approach produced an accuracy higher than those when using the original Landsat-8-time series (respectively, 73% and 70%), as well as when combined with filtering approaches. The classification accuracy using the filtered integrated time series (78%) was slightly better than that using Landsat-8 time series (76%).

Holden and Curtis (2016) observed the effect of combining data from the two sensors (L7 ETM+ and L8 OLI). Once L7 ETM+ has the same spectral bandwidths of CBERS-4 MUX, we can use some of their conclusions. For example, NDVI relies on the contrasting relationship between the near infrared band and the red band. They observed that there is a strong and

consistent positive bias in NDVI, with Landsat-8 having higher NDVI. Here we observed that there is also a strong bias, but not consistent, as this relation is not observed for the smaller values. We observed that the EVI values are also higher for Landsat-8 than CBERS-4, but the correlation between them is higher than NDVI. The EVI differs from NDVI by utilizing the blue band as an additional normalizing factor that corrects the red band for atmospheric influences. The bias in the blue band between Landsat-8 and CBERS-4 nullifies the bias in the red and near infrared band, resulting in a more correlated EVI across sensors (HOLDEN; CURTIS, 2016). This is probably the reason explaining the higher correlation for EVI, and consequently the best classification results when using the EVI integrated time series. Arvor et al. (2011) reported that EVI is more sensitive than NDVI for high biomass regions (HUETE et al., 1999). According to Chen et al. (2004), this is due to the fact that the NDVI saturates faster than the EVI, resulting in the NDVI being less effective in separating crops. Moreover, the EVI is less affected by atmosphere and soil disturbances (HUETE et al., 2002). We can see that small differences on the time series values lead to changes in the results of the smoothers improved by TIMESAT. Furthermore, finding differences on the extracted parameters can modify the results of classification. As the correlation between the MUX NDVI and OLI NDVI tend to be smaller, it can modify the amplitude of the signal, resulting in significant changes on the smooth time series. We can also see in Figure 3.5 that the goodness of fit between the Landsat NDVI and MUX NDVI are significantly lower than in respect EVI. As observed by Pinto et al. (2016), the greatest spectral bandwidths difference between the sensors is on the NIR band. But there is also significant spectral response function (SRF) profiles differences between corresponding CBERS-4 MUX and Landsat-8 OLI spectral bands.

3.4. Final Considerations

This experiment had the objective of investigating a method for noise removal on Landsat-8 OLI time series using CBERS-4 MUX data to improve a crop classification method based on phenological features. We observed a slight increase in the classification accuracy when using the proposed method. The results for EVI were consistently more accurate compared to NDVI. The best result was observed with the EVI integrated filtered time series (78%), followed by the filtered Landsat EVI time series (76%). This work did not compare thoroughly the two sensors, but we can see that there are significant differences. We suggest that image normalization procedures are strongly recommended to equate the surface reflectance from CBERS-4 to Landsat-8. A per-scene relative correction should also be performed to incorporate the spatial variability of the sensor differences and the seasonal variation. We can also infer that the different methods of atmospheric correction and ancillary datasets may be affecting the results; as well problems of misregistration between the images, resampling and the use of just one pair of images for determine the equation to predict OLI reflectance from MUX reflectance can also be sources of errors. More studies using other footprints and for longer time series are needed to better comprehend the relation between the OLI and MUX images and the effects of the different filtering approaches, as well to understand these effects on the results of smoothing proceeded by TIMESAT with double logistic functions. It is also suggested to test the other smoothing approaches implemented by TIMESAT as the Asymmetric Gaussian functions and Savitzky-Golay.

4 EVALUATION OF SMOOTHING METHODS ON LANDSAT-8 EVI TIME SERIES FOR CROP CLASSIFICATION BASED ON PHENOLOGICAL PARAMETERS⁵

4.1. Introduction

With the availability of free and continuous satellite imagery, allowing the construction of consistent time series of vegetation index images, remote sensing undergoes a paradigm shift with regard to monitoring changes in land use and land cover, mainly vegetation and agriculture applications, where it is necessary to take into account phenological variations.

Due to the possibility of working with consistent time series, each pixel of image time series can be treated as a signal, so that signal processing techniques and econometrics can be applied, such as the decomposition of time series in trend and seasonality components, and extraction of parameters of these components to land cover and land use classification (ARVOR et al., 2011; ZHENG et al., 2015), change detection and trajectories analysis (VERBESSELT et al., 2010). One of the important steps for this type of study is the preprocessing of the time series for noise removal, usually caused by the presence of clouds. Several algorithms for noise removal in time series have been used, among them, the Savitzky-Golay smoother (CHEN et al., 2004), asymmetric Gaussian functions (JÖNSSON; EKLUNDH, 2004) and Double-logistic (ZHANG et al., 2003; JÖNSSON; EKLUNDH, 2004). In addition to the application of these smoothing algorithms, cloud masks combined with outlier interpolation techniques can also be used together (HAMUNYELA et al., 2013; BENDINI et al., 2016), as well as multi-sensor approaches for replacing contaminated pixels by cloud-free pixels (BENDINI et al., 2017b). Some studies in the

⁵ This chapter is based on the paper: BENDINI, H. N.; FONSECA, L. M. G.; KÖRTING, T. S.; SANCHES, I. D. A.; MARUJO, R. F. B. Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters. In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 18., 2017, Santos, Brazil. Proceedings... São José dos Campos: INPE, 2017. p. 4267-4274. ISBN: 978-85-17-00088-1.

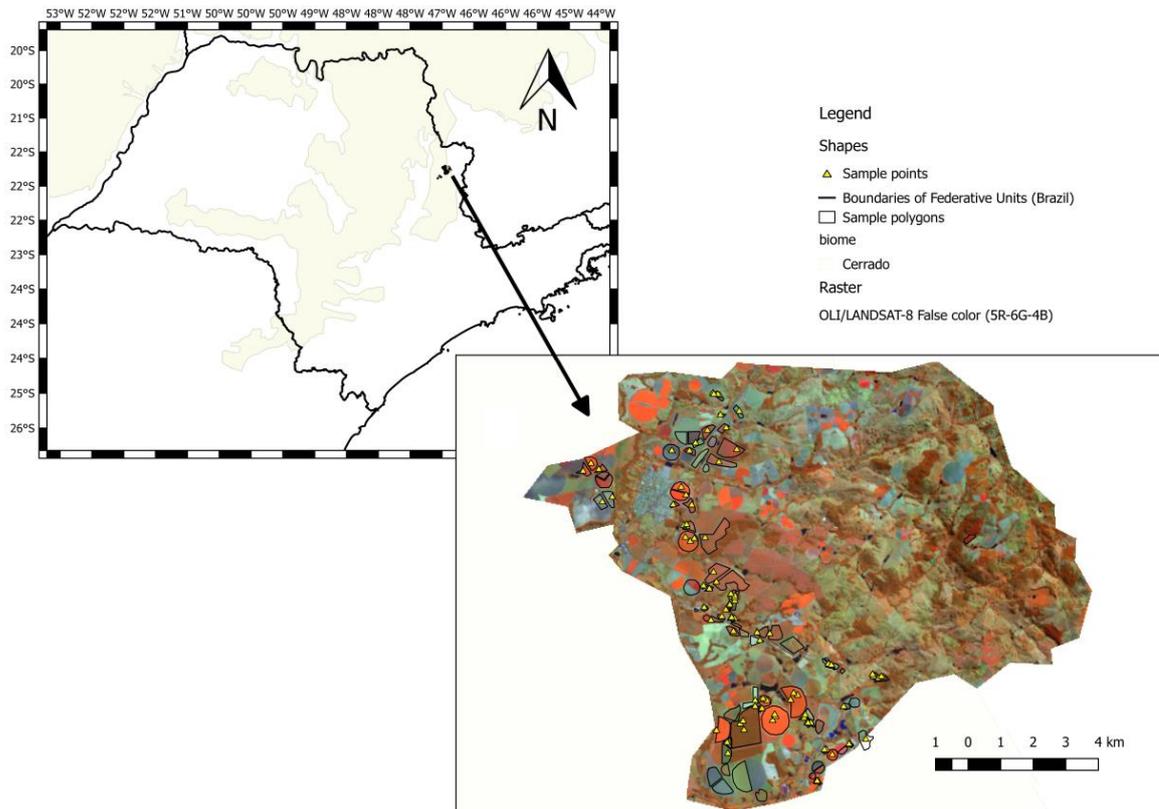
recent literature seek to compare these different smoothing approaches, but most of them focused on coarse spatial resolution satellite image time series, such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (ATZBERGER; EILLERS, 2011; BORGES; SANO, 2014) and, furthermore, they do not take into account the effect that these different smoothers have on the classification performance. Bendini et al. (2017b) evaluated the application of Landsat-8 OLI and MUX/CBERS-4 sensor integration method to obtain vegetation indices time series with lower cloud contamination to improve a crop classification in the Cerrado biome. They used phenological attributes extracted from time series smoothed by the double-logistic algorithm, since this algorithm was referenced by Borges and Sano (2014) as being the most suitable for smoothing series in agricultural areas in the Cerrado, considering MODIS images. However, there is a lack of information to confirm if this comparison is valid for Landsat-like image time series. Thus, the objective of this study was to evaluate the different time series smoothing methods, Savitzky-Golay, asymmetric Gaussian and double-logistic functions, combined or not with filtering techniques, and the impact of its use on the classification of agricultural use in a region of the Cerrado, using phenological parameters extracted from EVI Landsat-8 image time series.

4.2. Work Methodology

We conducted our study in the Itobi municipality, in São Paulo state, Brazil (Figure 4.1). Field campaigns were carried out for collecting training samples, totaling 100 pixels. The classes considered were Annual Crop (potato, soybean, corn, onions and sugar beet on a double crop system), Semi-Perennial Crop (sugarcane), Perennial Crop (avocado, mango and Brazilian grape tree), Natural Forest and Grassland. A total of 24 scenes of Landsat-8 OLI (WRS 2 – Worldwide Reference System 2, Path/Row 219/75) between August 2015 and August 2016 were processed to Level 1 Terrain Corrected (L1T). These were corrected for atmospheric conditions to identify and mask cloud and cloud shadows by the USGS EROS Science Processing Architecture (ESPA) (DEVRIES et al. 2015; DEVRIES et al. 2015a). Landsat-8 data were corrected using L8SR, a newly developed

algorithm that takes advantage of some of Landsat-8's new sensor characteristics (USGS, 2016; VERMOTE, 2016). Cloud (pixel value 4), cloud shadow (pixel value 2), snow (pixel value 3), water (pixel value 1) and clear (pixel value 0) masks were provided for Landsat-8 data using Cfmask, a C implementation of the Fmask algorithm (ZHU; WOODCOCK, 2012; ZHU et al., 2015).

Figure 4.1 – Location of Itobi municipality, in São Paulo state, Brazil.



We first applied a combined filtering approach for noise removal on the Landsat time series. The approach was put forth by interpolating the noise values with the average between the nearest neighbors in time, considering the Fmask quality data and negative outliers based on a threshold as recommended by Hamunyela et al. (2013). This method, however, is not capable of removing consecutive outliers.

For both cases (filtered and raw data) we implemented three smoothing algorithms through the TIMESAT software package including the Savitzky-Golay, asymmetric Gaussian and double-logistic functions (JÖNSSON; EKLUNDH, 2004), followed by the extraction of the phenological attributes

using TIMESAT as well. More details about its implementation and the phenological attributes extraction can be seen at Jönsson and Eklundh (2015).

We subject the phenological metrics obtained on TIMESAT to data mining using the Random Forest (RF) algorithm (BREIMAN, 2001) considering each method: raw Landsat EVI time series smoothed by 1) Savitzky-Golay function, 2) by Asymmetric Gaussian and 3) Double logistic function; and the filtered Landsat EVI time series smoothed by 4) Savitzky-Golay function, 5) Asymmetric Gaussian and 6) Double- logistic function.

The RF algorithm is a classification technique in which the data set is randomly divided into several subsets of smaller size by means of applying bootstrap, and from each subset a decision tree is developed. All trees contribute to the classification of the object under study, by voting on which class the target attribute must belong. Random Forest algorithm has been widely used in remote sensing (MÜLLER et al, 2015; PEÑA et al, 2015) because of its advantages in efficiently handling large databases, providing estimates on the most relevant variables, and allowing the identification of outliers (RODRIGUEZ-GALIANO et al., 2012).

There were a total of 31 training pixels for the annual agriculture class, 15 pixels for perennial agriculture, 26 pixels for semi-perennial agriculture, 14 pixels for grassland and 14 pixels for native forest.

The results were evaluated by the confusion matrix index, as global accuracy (GA), Kappa and producer's accuracy (PA) (WITTEN et al, 2011). The models were executed considering a 10-fold cross validation method. The classification results were obtained using the software package WEKA (HALL et al., 2009).

4.3. Results and Discussions

The classification results considering the different smoothing methods over the original raw EVI time series are showed in Table 4.1.

Table 4.1 – Accuracy assessment statistics of the classifications considering the different smoothing algorithms using the EVI raw time series including global accuracy (GA), kappa statistic, and the producer's accuracy (PA) of each class.

<i>Smoother/PA (%)</i>	<i>Annual Agriculture</i>	<i>Natural Forest</i>	<i>Perennial Agriculture</i>	<i>Semi-perennial Agriculture</i>	<i>Grassland</i>	<i>GA% (kappa)</i>
<i>DL</i>	84.00%	75.00%	53.57%	90.00%	61.90%	70.00% (0.62)
<i>AG</i>	89.66%	78.57%	76.92%	73.68%	66.67%	79.00% (0.73)
<i>SG</i>	92.86%	85.71%	78.57%	82.35%	92.31%	86.00% (0.82)

The best classification result was found using the Savitzky-Golay smoothed data, with an overall accuracy of 86% and kappa of 0.82. This result is followed by 79% (kappa of 0.73) with Asymmetric Gaussian method, and 70% (kappa of 0.62) using the Double-logistic function.

In respect to per-class classification results, we can see that the only class that had lower producer's accuracy using the Savitzky-Golay smoothed time series it was the Semi-Perennial Agriculture, with 82.35%. The higher producer's accuracy for this class was found with the Double-logistic method, with 90%.

Figure 4.2 shows the smoothed EVI time series with the different smoothing algorithms using the filtered time series and the points of start and end of seasons detected by the TIMESAT's algorithm to extract the phenological attributes.

The classification results considering the different smoothing methods over the filtered EVI time series can be observed in Table 4.2.

Figure 4.2 – Smoothed EVI time series with the different algorithms using the raw data. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons detected by the TIMESAT’s algorithm to extract the phenological attributes.

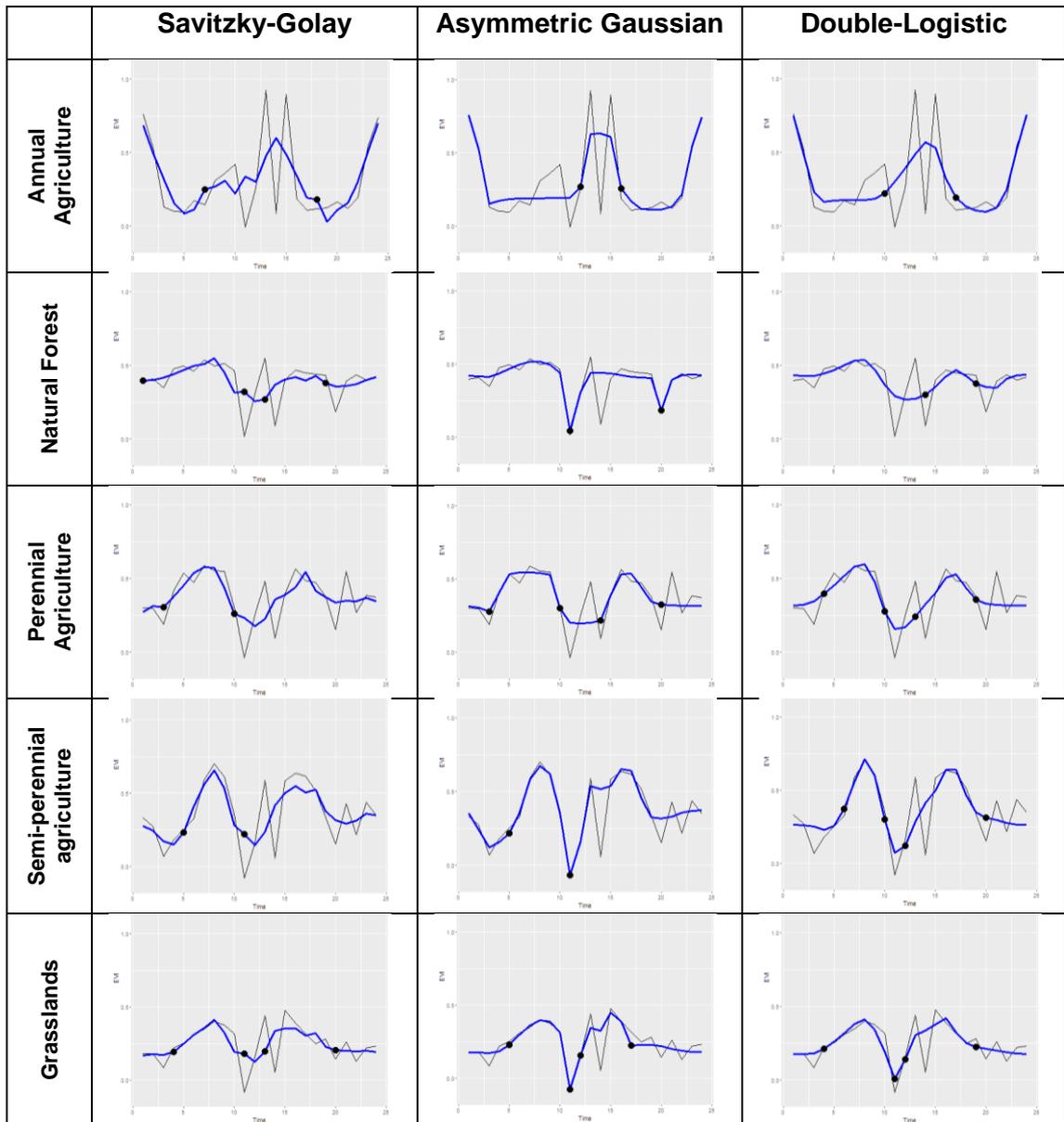
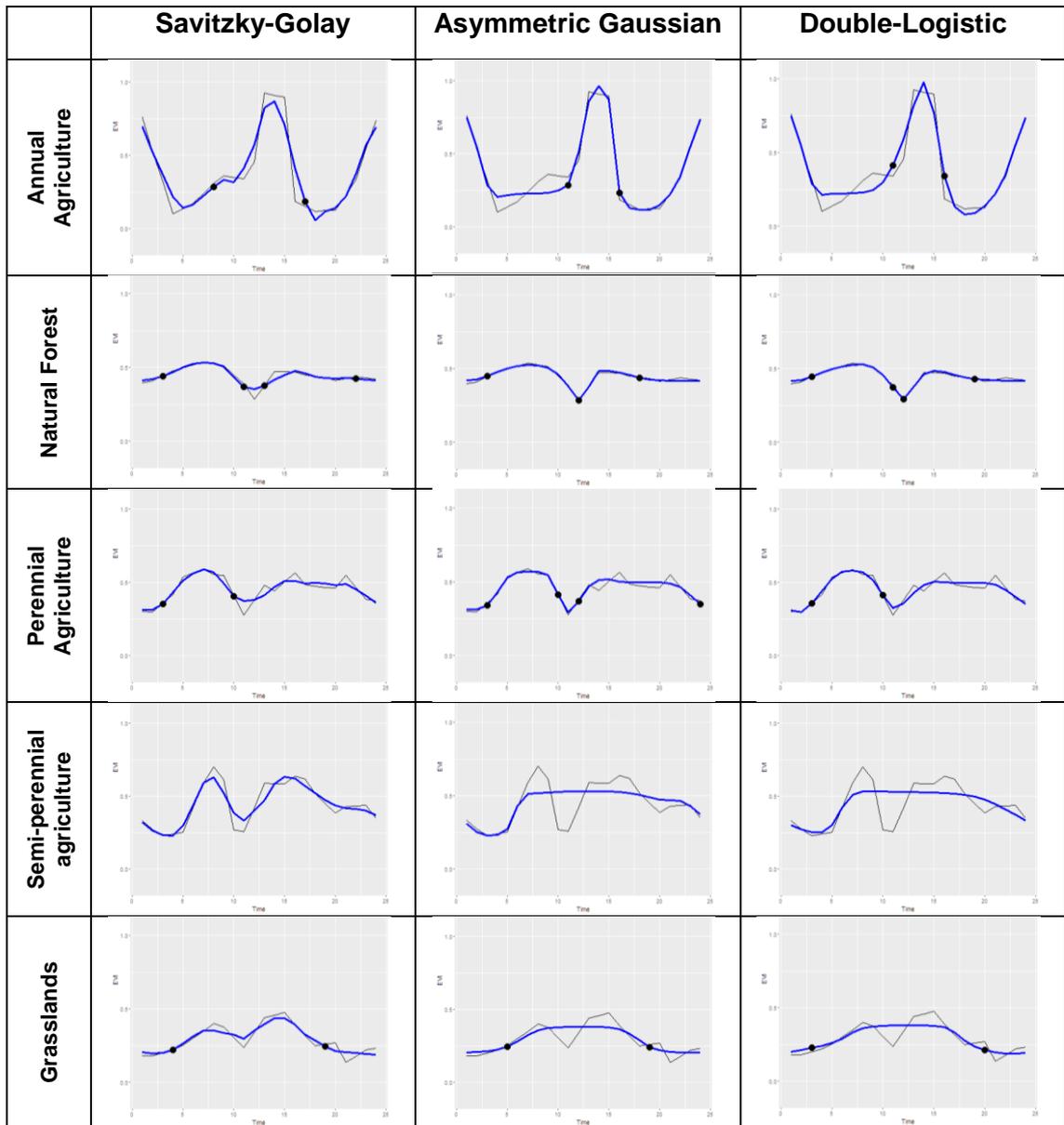


Figure 4.3 shows the Smoothed EVI time series with the different smoothing algorithms using the filtered time series and the points of start and end of seasons detected by the TIMESAT’s algorithm to extract the phenological attributes.

Figure 4.3 – Smoothed EVI time series with the different smoothing algorithms using the filtered EVI time series. The blue line is the smoothed time series, the black thin line is the raw time series, and the points are the start and end of seasons detected by the TIMESAT’s algorithm to extract the phenological attributes.



The best classification result was found using the Asymmetric Gaussian smoothed data, with an overall accuracy of 81% and kappa of 0.75 (Table 4.2). This results are followed by 76% (kappa of 0.69) with Double-logistic method, and 70% (kappa of 0.61) using the Savitzky-Golay function. In respect to per-class classification results, we can see that the only class that had lower producer’s accuracy using the Asymmetric Gaussian smoothed time series it was the Semi-perennial agriculture class, with 61.54%.

Table 4.2 – Accuracy assessment statistics including global accuracy and (kappa statistic) for classifications, considering the different smoothing algorithms using the EVI filtered time series, and the producer's accuracy (PA) of each class.

<i>Smoother/PA (%)</i>	<i>Annual Agriculture</i>	<i>Natural Forest</i>	<i>Perennial Agriculture</i>	<i>Semi-perennial Agriculture</i>	<i>Grassland</i>	<i>GA% (kappa)</i>
<i>DL</i>	86.67%	68.75%	76.00%	69.23%	68.75%	76.00% (0.69)
<i>AG</i>	90.32%	71.43%	78.57%	61.54%	92.86%	81.00% (0.75)
<i>SG</i>	78.79%	57.14%	64.52%	72.73%	72.73%	70.00% (0.61)

We can see clearly looking to Figure 4.2 and 4.3 that when the filtering approach was applied before the smoothing, the model fits better the time series, but this are not related to the improvement of the classification results. We calculated the mean standard deviation considering all the pixel time-series within each class in order to measure the intra-class variability. Figure 4.4 shows the within-class mean standard deviation for the different classes, considering data from different smoothing algorithms using the raw and the filtered EVI time series.

Figure 4.4 – Within-class mean standard deviation for the different classes using the different smoothing algorithms with the raw EVI time series.

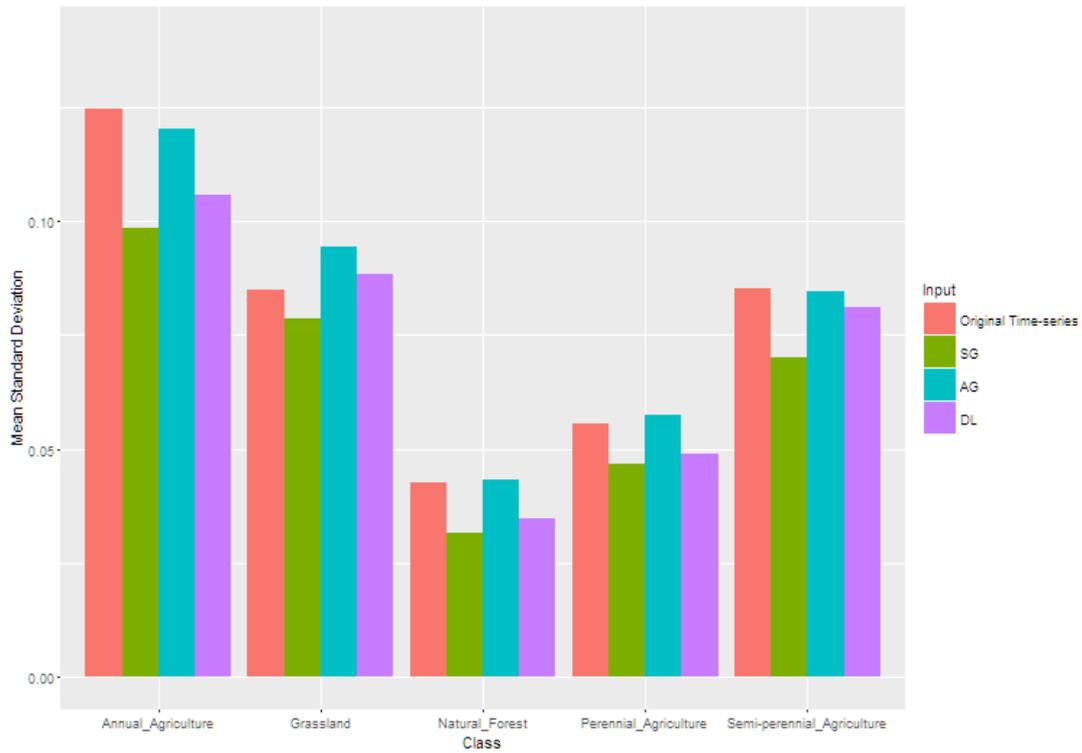
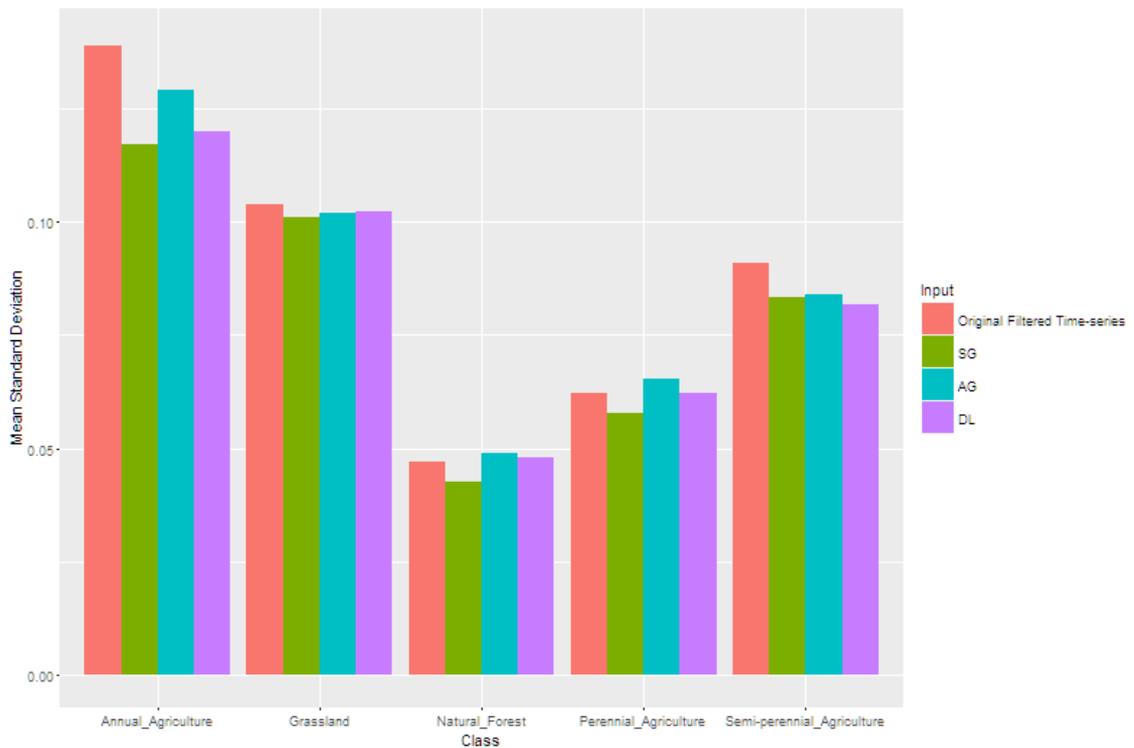


Figure 4.5 – Within-class mean standard deviation for the different classes using the different smoothing algorithms with the filtered EVI time series.



We can observe in Figure 4.4 and 4.5 that the EVI time-series previously filtered presents bigger mean standard deviations within-class for all the classes for both smoothing methods, for example, for the Annual Agriculture

class, the mean standard deviation measures increases from 0.124 to 0.139, 0.098 to 0.117, 0.120 to 0.129 and 0.106 to 0.120 for the original time-series, SG, AG and DL, respectively. Among the three smoothing algorithms, SG algorithm generated the smallest mean standard deviation for most classes. This analysis allows us to evaluate how homogeneous are the results of each fitting procedure. We expect that lower mean standard deviation values indicates that the method does not generate much new EVI values, except for the outliers and pixels with quality issues, and by consequence provides more homogeneous results. Similar results were observed by Shao et al. (2016) using MODIS data. Considering the raw EVI time-series, the smaller the mean standard deviation within-class the higher the producer's accuracy, except for the Semi-Perennial Agriculture class. But this relation cannot be observed in the previously filtered time-series.

4.4. Final Considerations

In this experiment we evaluated three different time series smoothing methods, Savitzky-Golay, asymmetric Gaussian function and Double-logistic function, combined or not with filtering techniques. Besides, we evaluated the impact of their use in the agricultural use classification, in a region of the Brazilian Cerrado, using phenological parameters extracted from one year Enhanced Vegetation Index (EVI) Landsat-8 image time series. The smoothing method that provided the highest classification accuracy was the Savitzky-Golay applied to the raw time series (86% and kappa=0.82), followed by the asymmetric Gaussian applied to the filtered time series (81% and kappa=0.754).

5 ROBUST CLASSIFICATION OF CROPLANDS IN THE BRAZILIAN CERRADO BASED ON PHENOLOGICAL INFORMATION FROM DENSE LANDSAT TIME SERIES⁶

5.1. Introduction

Agricultural mapping and monitoring enable rapid generation of accurate information about agricultural supplies, which is strategic and crucial for policies and for decision-making in the investment market. Availability of this information is important for supporting food security programs, preventing food shortages and enabling transparency in the agricultural market. Brazilian agribusiness represents around 22.5% of the country's total GDP (CEPEA-USP/CNA, 2015). Even though Brazil's GDP retracted by 3.8% in 2015 due to a recession and political unrest, the agricultural sector's GDP increased by 1.8% and was the only sector to report any job growth (SPERA et al., 2017; MAPA, 2016).

With over half of Brazil's agricultural land falling within the Cerrado biome boundary, land clearing in this region has been central to the development and strength of the country's agricultural sector (SPERA et al., 2017). The Cerrado biome, also known as the Brazilian savanna, is the second-largest biome in Brazil after Amazon, with an extent of more than 2 million km² (MMA, 2016). Considered as a global biodiversity hotspot, the Cerrado biome also provides environmental services of global importance, such as carbon storage and climate regulation. Despite that, the Cerrado has lost 88 Mha (46%) of its native vegetation cover and with a projection that 31-34% of the remaining biome is likely to be cleared by 2050 (STRASSBURG et al., 2017). This scenario of intense agricultural expansion underlines the need for methods that accurately map the distribution of agricultural areas

⁶ This chapter is based on the paper: BENDINI, H. N.; FONSECA, L. M. G.; SCHWIEDER, M.; KÖRTING, T. S.; RUFIN, P.; SANCHES, I. D. A.; LEITÃO, 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, 101872, 2019: <https://doi.org/10.1016/j.jag.2019.05.005>

and its evolution over time, so that we can ensure sustainable agricultural development and the preservation of the biome.

Approaches like the TerraClass project were implemented to map natural areas, as well as land use in managed areas, first in the Amazon biome (ALMEIDA et al., 2016) and recently in the Cerrado (INPE, 2017). Agriculture in particular was mapped at intermediate thematic detail, such as annual crops, perennial crops (including coffee, citrus, and sugarcane) and planted forest. However, although the generated maps have wide applicability for understanding the main phenomena and processes involved in landscape transformations within those biomes (FONSECA-MORELLO, 2015; ADAMI et al., 2015; RISSO et al., 2012) it still relies on manual procedures of visual interpretation, which entails not only subjectivity but also considerable investment of time and effort.

Remote sensing data have been shown to be a valuable tool for land cover and land use mapping, as they capture spatially explicit land cover changes in a synoptic manner. In particular sensors with a high temporal resolution reflect the seasonal behavior of vegetation, which is essential for mapping the highly dynamic and spectrally similar cover types in the agricultural domain. Sensors like the Advanced Very High Resolution Radiometer (AVHRR) and the Moderate Resolution Imaging Spectroradiometer (MODIS) with up to daily revisit times have been widely used for agricultural mapping (SAKAMOTO et al., 2005; WARDLOW et al., 2007) and several studies highlighted the benefits of these time series for agricultural mapping in Brazil (ESQUERDO et al., 2011; RUDORFF et al., 2010; ARVOR et al., 2011; BORGES; SANO, 2014).

However, even though their moderate spatial resolution (250–1000 meters) has been shown to be sufficient for monitoring large scale agricultural practices, it does not allow for the detection of smaller fields, due to the spectral mixture of the different targets. This limits their use for the development of accurate agricultural maps with a high level of thematic and spatial detail. To overcome this limitation, several studies have focused on mapping approaches that use Landsat-like images that have a spatial

resolution which is sufficient to capture even small scale agricultural practices (ZHENG et al., 2015; PEÑA; BRENNING, 2015; PAN et al., 2015). These developments were catalyzed since space agencies began to provide satellite imagery in an operational way (i.e., consistent methods for atmospheric and geometric correction), and the extensive data holdings of the Landsat archive were made freely accessible (WULDER et al., 2012).

Along with advances in computational processing performance, and the possibility to combine data from different sensor systems the development of advanced methods of land cover and land use mapping that make use of the Landsat approaches (MÜLLER et al., 2015; RUFIN et al., 2015), specifically those involving dense time series (SCHWIEDER et al., 2016; BENDINI et al., 2017b; BENDINI et al., 2016). The Landsat sensor family has been observing Earth since 1974 and can thus be used to map land use and land cover changes over more than 40 years. However, due to SLC issues with the Landsat-7 ETM+ sensor, there is less wall-to-wall data available between 2012 (the end of the lifetime of Landsat-5 TM) and 2013 (launch of Landsat-8).

The use of crop phenological parameters extracted from image time series can be an important strategy for developing agriculture mapping methods. However, the 16-day temporal resolution of Landsat imagery is not sufficient to derive continuous phenological information, nor is it conducive to the acquisition of cloud-free images in high frequency.

Bendini et al. (2016) showed that phenological parameters can be derived at a 30-meter spatial resolution, and a higher temporal resolution enables mapping of agricultural classes with good accuracies (> 80%) (BENDINI et al., 2016). They used the overlap region of WRS2 footprints of Landsat-8 to generate 8-day-interval image time series and derived phenological parameters for classifying agriculture within a small study area in the northeast of São Paulo State, Brazil. However, with this approach the methodology was limited to these overlapped areas and could not be applied in large scale.

Schwieder et al. (2016) explored the potential of dense Landsat time series with 8-day temporal resolution to derive phenological information for mapping vegetation gradients in the Cerrado. Based on radial basis convolution filters they filled data gaps in combined Landsat ETM+/OLI time series to derive continuous pixel-wise temporal profiles that capture the seasonal behavior of vegetation physiognomies in the Cerrado (SCHWIEDER et al., 2016). Their results highlighted the benefits and limitations of this approach to map the main spatial patterns of the observed physiognomies based on their phenological differences. However, the transferability of the approach to managed agricultural lands was not assessed.

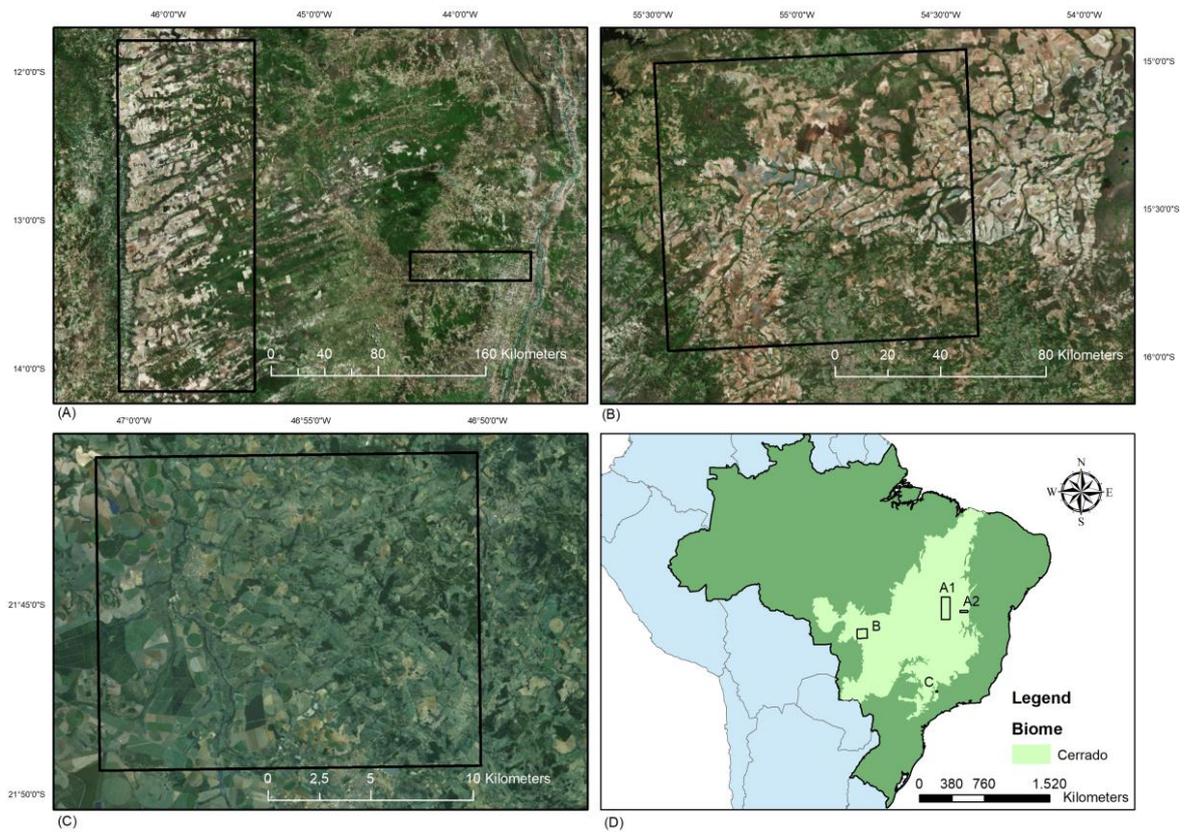
The main objective of this work is to describe and assess a method, based on phenometrics derived from dense satellite image time series, for classifying agricultural land use in the Cerrado biome. We sought to answer the following research questions: Are Landsat-based phenometrics suitable for classifying agricultural land use in the Cerrado biome? Can we separate different semantic levels of a hierarchical classification scheme with this method? Are Landsat based phenometrics beneficial to derive important agricultural phenological information of the major crops in the Cerrado biome?

5.2. Materials and Methods

5.2.1. Study areas

We tested the method in three different study areas within the Cerrado (Figure. 5.1) to analyze the method's robustness. The study areas were in the west of Bahia State (regions A1 and A2), southeast of Mato Grosso State (region B) and northeast of São Paulo State (region C). Despite all of these areas being in the Cerrado biome, they differ with respect to climate conditions, vegetation and agricultural practices.

Figure 5.1 – Study sites on the Cerrado biome.



The areas in southeastern Mato Grosso and western Bahia (Figure 5.1; regions A1, A2, and B) consist mostly of large-scale and market-oriented agriculture. The most common crops are soy, maize and cotton, but there is also significant production of coffee and banana, as well as small areas producing beans, rice and cassava. The study area in southeastern Mato Grosso is characterized by intensive soy/maize and soy/cotton double-cropping rotations, while western Bahia features mostly single-cropping regimes of soy and cotton. The study area in northeastern São Paulo State (Figure 5.1; region C) is considered a smallholder agricultural zone, characterized by high intra and inter-field spatial variability. Therefore, despite its small size, this study area is characterized by heterogeneous landscapes. Farmers in this region grow a variety of crops throughout the year.

The majority of crops in this area are sugarcane, potato, maize, soy, beans, carrot and onions. There are also areas of perennial crops such as avocado, Brazilian grapefruit, lemon, mango, coffee and planted forests (eucalyptus)

(IBGE, 2016). Farmers grow crops mainly in double-cropping systems, mostly in irrigated areas. The usual planting period is from October to December, with harvesting from February to April, but farmers also plant winter crops in late fall (May–July), especially in the irrigated areas, and harvest them in the following spring.

Field surveys characterizing the major cropping systems were conducted during the 2015–2016 cropping season. We registered GPS waypoints in the study area, choosing the points based on their proximity to roads, on the size of the field, and on the representativeness of existing cropping systems. For the study area in northwestern Bahia, we conducted the survey together with the “Rally da Safra 2016” team (AGROCONSULT, 2018) during the growing peak (end of February and March). For the other study sites, we visited the area throughout the year, recording information about crop rotation systems. The field work in Campo Verde is described in Sanches et al. (2018) (SANCHES et al., 2018).

5.2.2. Definitions for the hierarchical classification scheme

We collected data on land uses and land cover at different levels of thematic detail (see Table 5.1) to test the proposed approach using a hierarchical classification scheme. Lebourgeois et al. (2017) showed that this approach could improve classification results (LEBOURGEOIS et al., 2017), relative to the classical approach (where a complete training dataset is used at each level).

GPS waypoints were also registered for different non-crop classes (pasture, planted forest, natural vegetation, etc.). As our methodology uses phenological information for classification, we considered the Perennial crop class and non-crop classes together due to the evergreen phenological behavior of perennial crops. We divided the natural vegetation class into three main Cerrado physiognomies, based on the definition of TerraClass Cerrado (INPE, 2017): forest, savanna and natural grasslands (RIBEIRO; WALTER, 2008).

We also used the TerraClass maps and photo interpretation of Google Earth imagery to collect additional samples for the non-crop classes. Field boundaries were then digitized over high resolution Google Earth imagery to obtain a ground polygon database. We generated 841 polygons (681 of Annual and Semi-perennial crop classes and 160 of Perennial crop and non-crop classes).

Table 5.1 – Definition of the four levels of detail as defined in the hierarchical classification scheme. The number of available reference polygons for training and validation is shown in parenthesis.

Level 1 Cropland	Level 2 Land Cover	Level 3 Crop Group	Level 4 Crop Rotation
<i>Annual crop and Semi-perennial crop (681)</i>	<i>Annual crop (637)</i>	<i>First crop / Second crop (348)</i>	<i>Soy / Bean (2)</i>
			<i>Soy / Maize (184)</i>
			<i>Soy / Cotton (141)</i>
			<i>Soy / Soy (1)</i>
			<i>Maize / Soy (6)</i>
		<i>First crop / Winter crop (30)</i>	<i>Maize / Carrot (2)</i>
			<i>Maize / Onion (4)</i>
			<i>Maize / Potato (3)</i>
			<i>Maize / Beans (5)</i>
			<i>Soy / Potato (14)</i>
		<i>Single crop (177)</i>	<i>Maize (28)</i>
			<i>Soy (119)</i>
			<i>Cotton (27)</i>
			<i>Millet (1)</i>
			<i>Sorghum (1)</i>
		<i>Single crop / Non-commercial crop (82)</i>	<i>Cotton / Millet (1)</i>
			<i>Soy / Millet (7)</i>
<i>Soy / Brachiaria (6)</i>			
<i>Soy / Sorghum (4)</i>			
<i>Soy / Weed (8)</i>			
<i>Maize / Weed (1)</i>			
<i>Maize / Regrowing (4)</i>			
<i>Semi-perennial crop (44)</i>			
<i>Perennial crop and Non-crop (160)</i>	<i>Perennial crop (31)</i>		
	<i>Planted forest (15)</i>		
	<i>Pasture (63)</i>		
	<i>Forest (26)</i>		
	<i>Natural grasslands (10)</i>		
	<i>Savanna (15)</i>		

The average polygon areas fluctuated from 8.1 hectares for region C to 130 hectares for regions A1, A2 and B. We selected approximately 50 randomly sampled pixels from each polygon, avoiding the edges (using a 30-meter buffer), resulting in a set of 54,151 samples.

The land cover classes (Level 2) are based on the nomenclature of the Systematic Survey of Agricultural Production of the Brazilian Institute of Geography and Statistics, or PAM-IBGE (IBGE, 2016). Annual crops are defined as short- or medium-duration crops, completing their vegetative cycle within one year and requiring replanting after harvest to produce. Perennial crops have a longer growing season and produce successive harvests without the need for replanting. Semi-perennial crops, though not listed as one of the classes in the PAM-IBGE, are defined as crops of average duration—their vegetative cycle lasts between 1 and 2 years, and they require replanting after harvest to produce (e.g., sugarcane). Planted forest is a forest formation that is planted and regularly harvested by humans (*Eucalyptus* sp., *Pinus* sp., etc.). The Crop Group (Level 3) is defined by the main agricultural practices in the Cerrado region. The “First Crop / Second crop” class is the most common. It consists of a main crop during the summer season (October–March) and a second crop, usually maize, between March and June.

The “First Crop / Winter crop” is most common in regions characterized by smallholders and the presence of irrigated agriculture; the farms also rotate a main crop, such as soy or maize, with winter crops like onions, potatoes, beans and carrots. “Single crop” is the traditional agricultural system that consists in only one crop season during the year and commonly still uses soil tillage. Finally, “Single crop / Non-commercial crop” consists on planting one main crop per agricultural year, but with the use of a non-commercial crop (i.e., sorghum, brachiaria or millet) to make feasible the no-till farming. The Crop Rotation level (Level 4) consists in the most detailed level, with crop type definitions.

5.2.3. Remotely sensed data

We used all available ETM+ and OLI data for the study areas (Path/Row 226/070, 225/070, 226/071, 225/071, 220/068, 220/069, 220/070, 219/069 and 219/075), acquired between April 2013 and April 2017. Assuming an 8-day temporal resolution, this 4-year time frame contains 186 potential observations. The images were obtained from the US Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). These data are provided with level 1 geometric correction (L1TP). Landsat-7 imagery was converted to surface reflectance by the atmospheric correction algorithm LEDAPS (Landsat Ecosystem Disturbance Adaptive Processing) (MASEK et al., 2016), and Landsat 8 data were corrected using L8SR, a newly developed algorithm that takes advantage of Landsat-8's new sensor characteristics (USGS, 2016; VERMOTE et al., 2016). Cloud and cloud shadow pixels were masked out from using the Fmask algorithm (ZHU; WOODCOK, 2012; ZHU et al., 2015). From the resulting surface reflectance products, we derived the Enhanced Vegetation Index (EVI) (LIU et al., 1995), which is known to decouple the canopy background from the signal (HEUTE et al., 2002).

5.2.4. Landsat dense time series

Limiting factors of a dense time series are sensor errors, such as the scan line corrector failure in the case of the Landsat-7 ETM+ (MAXWELL et al., 2007), and cloud cover. To overcome these constraints, Schwieder et al. (2016) used a weighted ensemble of Radial (Gaussian) Basis Function (RBF) convolution filters to approximate the missing data in a Landsat time series. To approximate the given EVI observations into dense 8-day time series without data gaps, we used the RBF approach (SCHWIEDER et al., 2016) with some adaptations. Let $f(t)$ be a time series, where $t \in \{1, \dots, N\}$. The approximated values $y(t)$ are calculated by Equation 5.1.

$$y(t) = \frac{\sum_{i=1}^T [f(t) \otimes K_i] W_i}{\sum_{i=1}^T W_i} \quad (5.1)$$

where \otimes is an operator for convolution, the kernels for the convolution (K_i) are given by the Gaussian function (Equation 5.2), and T is the total number of standard deviations (σ) that can be used for the kernel calculation.

$$K_i = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma_i} \right)^2} \quad x \in \{1, \dots, L\} \quad (5.2)$$

The size of the kernel window (L) is given by σ , which directly translates into the number of observations (at 8-day intervals). The total kernel width is limited to the points in time, which delineate 90% of the area under the Gaussian kernel. Increasing σ expands the kernel window and lowers the kernel values. We used three different kernels with $\sigma_1 = 0.5, \sigma_2 = 1, \sigma_3 = 3$. The final approximation is the weighted average of the results of the three temporal convolution filters, where the weights (W_i) are also calculated by the kernel convolution (Equation 5.3), but applied in $d(t)$, which is a vector that expresses data availability (Equation 5.4). The more data available in each kernel window relative to the total kernel window, the higher its weights in the final aggregation of the different kernels.

$$W_i = d(t) \otimes K_i \quad (5.3)$$

where,

$$d(t) \begin{cases} 0, & \text{if } f(t) = NA \\ 1, & \text{if } f(t) \neq NA \end{cases} \quad (5.4)$$

Unlike Schwieder et al. (2016), we did not use a priori outlier detection. After an expert-driven visual inspection, we observed that using this outlier detection masked some expected intrinsic variations in the phenological profiles of agricultural targets (abrupt greening, induced senescence or harvesting). We also considered that, as the outliers not relate to harvesting are mostly related to clouds or cloud shadows, the cloud masking should be sufficient to deal with those outliers. Finally, we used a spline function to interpolate possible remaining missing data, with the package “zoo” in R (R DEVELOPMENT CORE TEAM, 2017; ZEILES; GROTHENDIECK, 2005).

5.2.5. Phenometrics

Phenological parameters were obtained by the TIMESAT v3.2 software (JÖNSSON, P.; EKLUNDH, L., 2004), where seasonal data are extracted from the time series for each growing season of the focal year. During a period of n years, there may be $n - 1$ full seasons, as well as two fractions of a season at the beginning and end of the time series. In Brazil, the agricultural year of most crops is defined as between August of a given year and October of the following year (CONAB, 2010); therefore, when extracting seasonality parameters, we used the period of August 4, 2015, to October 1, 2016, as the focal year. We fitted the time series using the Savitzky-Golay filter (JÖNSSON; EKLUNDH, 2004; ZHANG et al., 2003) with a window size of 4. Among the three available approaches in TIMESAT, we found in Chapter 4 that this function is recommended for cropland areas in the Brazilian Cerrado (BENDINI et al., 2017a). During the fitting procedure, a primary maximum is always given, and a secondary maximum may be found. If the amplitude ratio between the secondary maximum and the primary maximum exceeds the user-defined threshold (seasonality parameter), it will detect two annual seasons.

Since our focus was detection of agricultural areas, the seasonality parameter was set to 0, forcing the software to treat the data as if there were two annual seasons. A set of 13 phenometrics were derived for each season (S1 and S2).

The start of season (SoS) and the end of season (EoS) are the dates when the values at the left and right edges, respectively, have increased to a defined percentage of the minimum value on the corresponding side. We selected a value of 10% of the vegetation growth amplitude to detect the SoS and EoS (SHANG et al., 2017). The start value (StartVal) and end value (EndVal) are the EVI values for SoS and EoS, respectively (JÖNSSON, P.; EKLUNDH, L., 2015). The length of the season (LoS) is the time between the start and the end of the season. The base value (Base) is the average of the left and right minimum values. The middle of the season (Mid) is computed as the mean value of observation dates, for which the left edge has increased to the 80% level and the right edge has decreased to the 80%

level. The maximum value (Peak) is the largest data value for the fitted function during the season. The seasonal amplitude (Amp) is the difference between the maximum value and the base level. The left derivative (Lder) and right derivative (Rder) are the ratios of the difference between the left or right 20% and 80% levels and the corresponding time difference. The large seasonal integral (Linteg) is the integral of the function describing the season from start to end. The small seasonal integral (Sinteg) is the integral of the difference between the Linteg and the base level from start to end of the season.

Besides the TIMESAT phenometrics, we also used the phenometrics proposed by Körting et al. (2013) (KÖRTING, 2013), which are also called polar features, since the purpose is to represent the time series by projecting the values onto angles in the interval $[0, 2\pi]$.

Let a cycle be the function $f(x) = (x, y, T)$, 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 in such a cycle. The cycle can be visualized as a set of values $v_i \in V$, where v_i is a possible value of $f(x, y)$ in time t_i . Let its polar representation be defined by the function $g(V) \rightarrow \{A, O\}$ (A corresponds to the abscissa axis in the Cartesian coordinates, and O to the ordinate axis) where:

$$a_1 = v_i \cos\left(\frac{2\pi}{N}\right) \in A, i = 1, \dots, N \quad (5.5)$$

and

$$o_1 = v_i \sin\left(\frac{2\pi}{N}\right) \in O, i = 1, \dots, N \quad (5.6)$$

Considering $a_N + 1 = a_1$ and $o_N + 1 = o_1$, we can obtain the coordinates of a closed shape. We then calculate the area of the resulting shape for each of the quadrants ($[\pi, 3\pi/2]$, $[\pi/2, \pi]$, $[0, \pi/2]$ and $[3\pi/2, 2\pi]$), which we refer to here as Q1, Q2, Q3 and Q4, respectively, and are supposed to represent the seasons.

5.2.6. Random Forest classification

After the feature extraction, we used our field database to train Random Forest (RF) (BREIMAN, 2001) and obtained a classifier for each nomenclature level. We used a hierarchical classification approach by which “Annual crop and Semi-perennial crop” / “Perennial crop and Non-crop” domains are isolated at level 1, and land cover is classified by correspondence at level 2. Then we isolated the “Annual crop” class and used it to classify land use by correspondence for each domain for the subsequent nomenclature levels.

RF is a classification technique in which the dataset is randomly divided into several smaller subsets, and a decision tree is built from each subset. The training dataset is randomly divided into several smaller subsets through bootstrapping — i.e., a bagging-based approach (random sampling with replacement) (HAN et al., 2011) — and a decision tree is built for each subset. All the trees contribute to classification by voting on the class to which a target attribute belongs. Random forests need two parameters to be tuned including the number of trees (ntree), and the number of variables (mtry). The RF algorithm has been widely used in remote sensing applications (CLARK et al., 2010; PEÑA; BRENNING, 2015; MÜLLER et al., 2015; BENDINI et al., 2017b) due to advantages such as efficiently handle large databases, providing estimates on most relevant variables, and allowing the identification of outliers (RODRIGUEZ-GALIANO et al., 2012). Furthermore, RF classifier is considered stable and relatively efficient, involving few user-defined parameters and yielding overall accuracy levels that are either comparable to or better than other classifiers (LAWRENCE et al., 2006; CHAN et al., 2008; PAL, 2005). The “randomForest” package in R was used for our classification tasks (LIAW; WIENER, 2002). To assess accuracy, we used an exhaustive method based on the Monte Carlo simulation, which performs random experiments to solve mathematical models and complex problems. The goal is to simulate a real system based on the large samples theory (RUBINSTEIN; KROESE, 2008). To execute the Monte Carlo simulation, 1000 simulations were carried out by randomly selecting 70% of the samples from the total

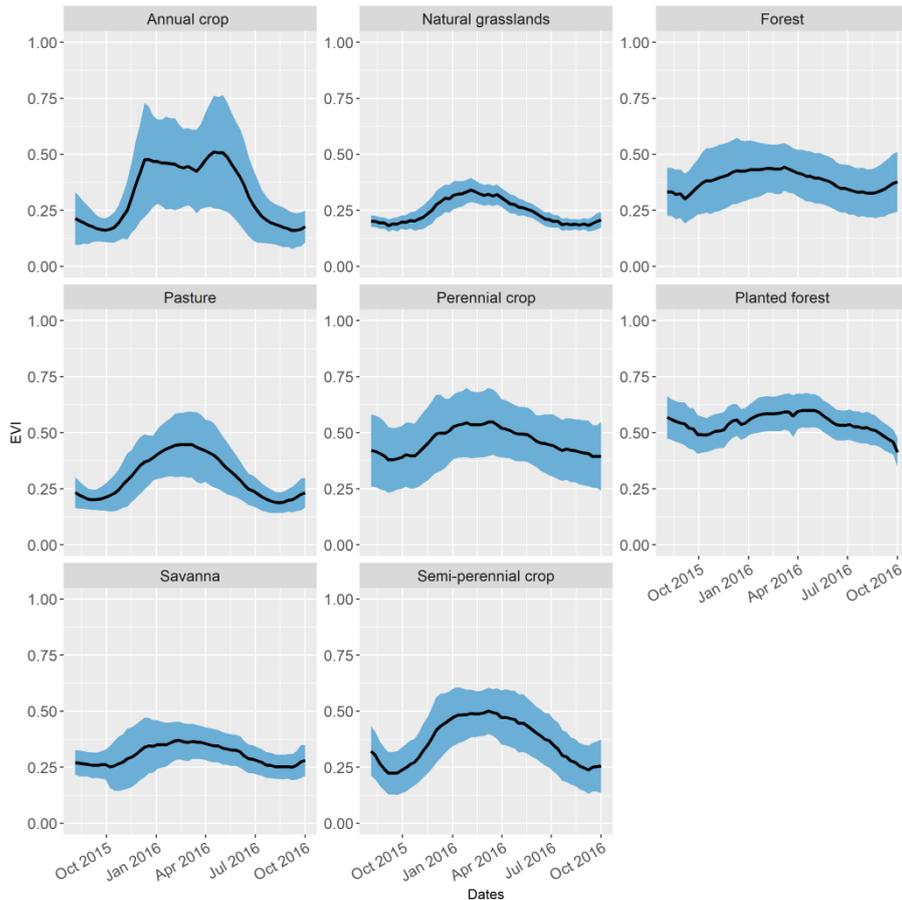
and using these to train a RF classification model for each hierarchical level (Table 1); the remaining 30% were used for validation. For each subdivision, a confusion matrix was calculated, and the average confusion matrix was used to derive the overall accuracy and the class f1-scores.

5.3. Results and Discussion

5.3.1. Phenological profiles

We derived pixel-wise EVI profiles from the 8-day temporal resolution Landsat time series for the period 2013–2017. Figure 5.2 shows averaged profiles for the target period of 2015–2016, with the respective standard deviations for each class of level 2 (Land Cover) based on the training pixels of all the study sites.

Figure 5.2 – Averaged EVI phenological profiles for each Land Cover class in the season 2015–16 (black lines), with their respective standard deviations (blue margins).



Arvor et al. (2011) observed similar patterns for the Annual crop class while using EVI data with coarser spatial resolution and analyzing the temporal profiles of this land cover class in the Cerrado (ARVOR et al., 2011).

The natural vegetation classes (Natural grasslands, Savanna and Forest) showed a general seasonal trend in the EVI profiles: a dry season from July to September followed by a green-up in mid-October, then a green season until June.

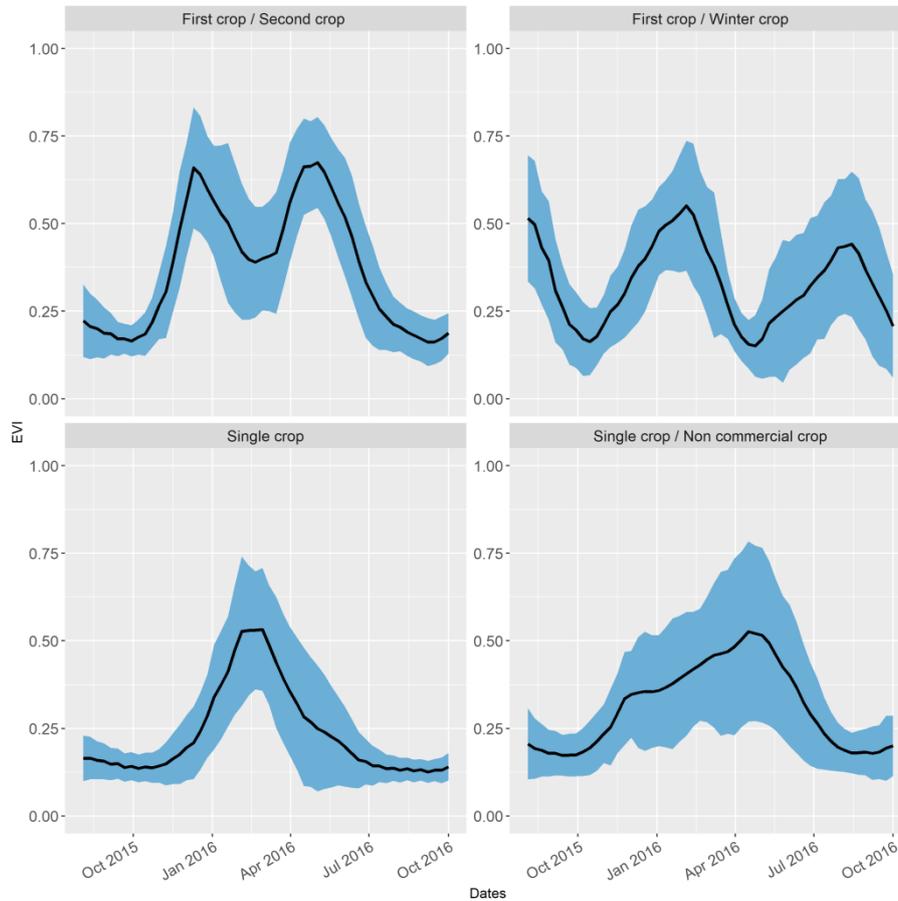
Ratana et al. (2005) found similar results for different physiognomies in the Cerrado (RATANA et al., 2005), using coarse spatial resolution data. However, their results with respect to the forest classes do not agree with those found by Arvor et al. (2011), who found higher EVI values (ARVOR et al., 2011). Schwieder et al. (2016) also found different results (SCHWIEDER et al., 2016), albeit using tasseled cap transformation, but we can verify a shift between the green seasons for the natural vegetation classes. These differences may be associated with the differences between these classes along different latitudes, due primarily to the precipitation regime and terrain.

Furthermore, as mentioned before, we used the three main Cerrado physiognomic classes, which aggregate the different forest physiognomies into one class, and all differences between them are incorporated. However, our goal was not to map natural vegetation in detail.

Even so, our fitted temporal profiles also show the potential to derive land surface phenology (LSP) in savanna ecosystems at the Landsat spatial resolution, thus ratifying the methodology proposed by Schwieder et al. (2016) (SCHWIEDER et al., 2016).

Figure 5.3 displays the averaged pixel-based EVI temporal profiles and standard deviations for all level 3 classes (Crop Group), based on the training dataset for all study sites.

Figure 5.3 – Averaged EVI phenological profiles for each Crop Group (Level 3) class in the season 2015–2016 (black lines), with their respective standard deviations (blue margins).



The temporal profiles for the double-cropping system of the “First crop / Second crop” class is mostly represented by a first season of soy, or occasionally maize, generally planted from late September to mid-November or even in early December (particularly in western Bahia). The harvest period occurs from late January to mid-March. The second season commonly consists of maize or cotton. Planting occurs between the end of February and beginning of March, depending on when the first-season crops are harvested. The harvest period of the second-season crops occurs from June to early July. These results agree with the findings of other authors (ARVOR et al., 2011; OLIVEIRA et a., 2014). The first season of the “First crop / Winter crop” class can begin approximately one month later than that of the “First crop / Second crop” class. The slope of the green-up curve is subtly less pronounced, because the first crop in this class is usually maize, which has a longer green-up period than soy (NGUY-ROBERTSON

et al., 2012). In the southeastern region of São Paulo, the winter crops are represented mostly by potatoes or minor crops like onions and carrots; these are planted from mid-March to June and harvested by late September or early October.

The single-crop class is represented mostly by cotton, soy and maize. When cultivated in a single-cropping system, cotton crops are planted between October and December and harvested from April to July. This agricultural practice is common in the region, where the agriculture intensification process is not well consolidated, as we mentioned before (SPERA et al., 2016). Finally, the crops of the “Single crop / Non-commercial crop” class are planted from March to May, after the late soy, and harvested between June and July. These classes were not well described by the EVI temporal profiles, because we expected the first-season crops to present a higher peak between November and March instead of between April and June. This fact emphasizes the need to consider that, in some regions, farmers grow non-commercial crops between October and November and plant the main crops between February and March. The non-commercial crops are used to maintain a constant vegetation cover (millet, sorghum and brachiaria) on the field during the dry season, reducing soil erosion and improving soy productivity when a no-tillage agricultural system is used (ARVOR et al., 2011).

5.3.2. Phenometrics

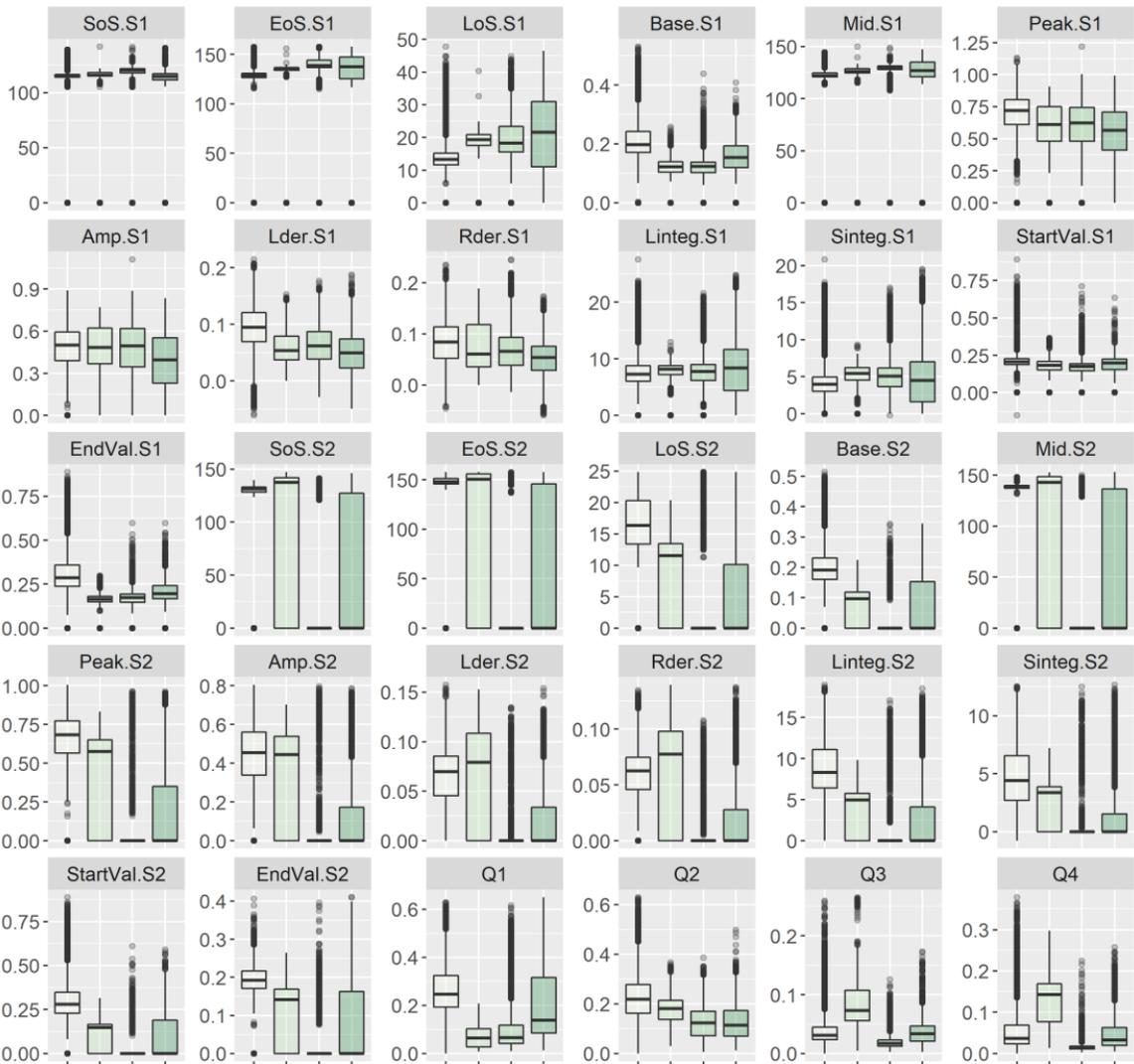
We used the derived EVI temporal profiles using TIMESAT to calculate pixel-wise phenological parameters for the season 2015 – 2016 and describe the course of the phenological profiles. The distribution of these parameters in each class highlighted their phenological differences (see Figure A1 in Appendix A for EVI phenological parameters for the level 1 classification based on the set of training pixels). The phenometrics of the first season (EoS.S1, LoS.S1, Amp.S1, Lder.S1, Rder.S1 and Q3) in particular show a clear separability between the two classes. Length of season revealed what we expected for agriculture: the complete phenological cycle is defined by the planting and harvesting activities and,

for most crops, generally lasts around 90 to 180 days. The main season of the non-crop class is strongly related to the natural vegetation greenness period, which extends from the rainy season to the post-summer period (200 to 280 days, depending on the region). The left and right derivatives are related to greening-up and browning, respectively. As expected, due to the abrupt changes in the EVI time series during planting and harvesting, the crop class presented much higher values for these variables. Except for the outliers, the “Perennial crop and Non-crop” class did not show a second season.

The variables amplitude, right derivative, left derivative and Q3 for the first season can be used to separate annual crops and semi-perennial crops for level 2 classification (see Figure A2 in Appendix A for EVI phenological parameters for level 2 classification based on the set of training pixels). Sugarcane crops constitute the Semi-perennial class; thus, higher values were expected for length of season, since a normal sugarcane cycle is about 9–12 months (for one-year sugarcane and winter sugarcane) or 18 months (one-and-a-half-year sugarcane). By integrating different varieties of sugarcane in our samples, the phenometrics describing the start, end, and length of season for this class varied widely. In general, the parameters used for season detection in TIMESAT were not adjusted for semi-perennial crops, since our work focused on annual crops. Despite this, the hierarchical classification efficiently separated semi-perennial crops from annual crops. We highlight the importance of the variables generated with polar representation (Q3 and Q4), which does not depend on season detection—these variables showed a high potential for increasing the separability of these classes.

Figure 5.4 displays the distribution of the EVI phenological parameters for the level 3 classification (First Crop / Second Crop, First crop / Winter crop, Single crop, and Single / Non-commercial crop) based on the set of trainings pixels.

Figure 5.4 – Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 3 classification (First crop / Second crop, First crop / Winter crop, Single crop, and Single / Non-commercial crop) based on the training pixels.) based on the training pixels.

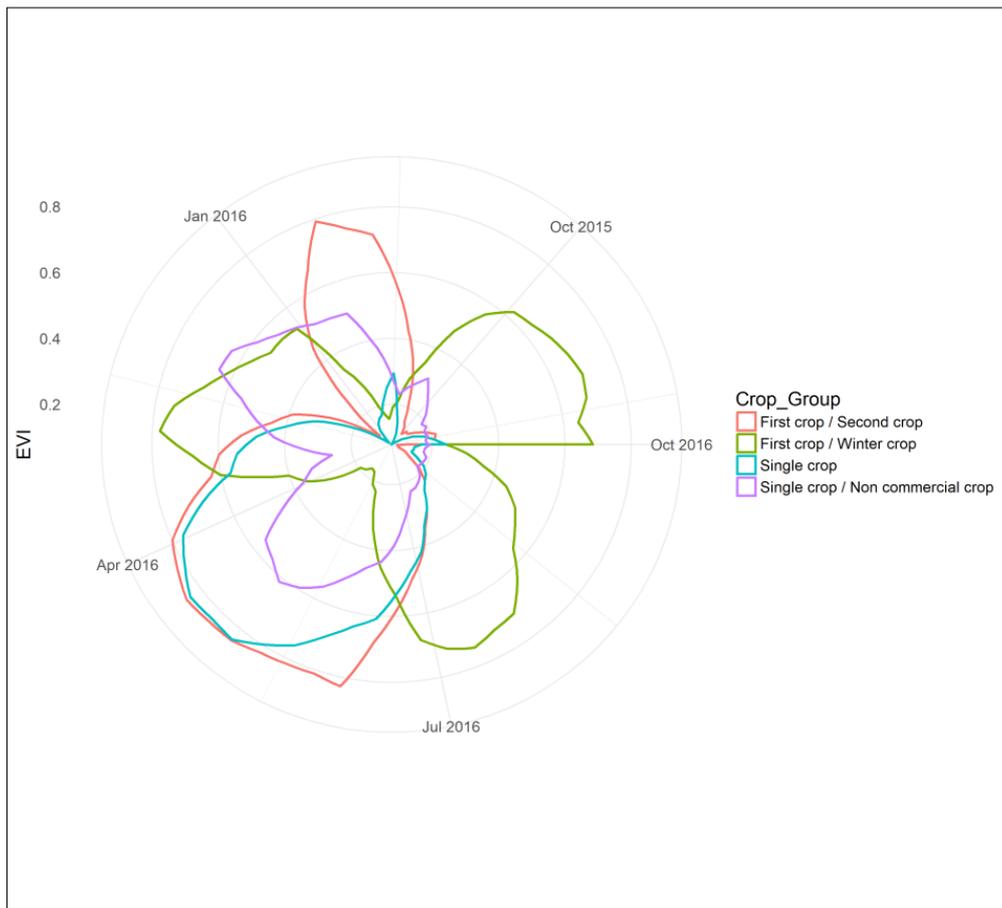


Class First crop / Second crop First crop / Winter crop Single crop Single crop / Non commercial

Figure 5.4 shows a large variation in most of the parameters for the second season of the “First crop / Winter crop” and the “Single crop / Non-commercial crop” classes. This may be explained by the higher diversity of crops cultivated in these classes compared to the “First crop / Second crop” class, in which the major crops planted in the second season are maize and cotton. Base, start and end values for “First crop / Winter crop” are low, which is possibly related to the fact that potato crops require intensive tillage operations for planting, while maize, cotton and soy in double-cropping systems are planted mostly without tillage. We verified that base, start and

end values are generally lower for the “Single crop” class, which is represented largely by soy, and the majority of the samples are in region A, in western Bahia. This may be related to the fact that, in single-cropping systems, soy is usually planted through conventional systems, i.e. soil tillage, which agrees with the results of Spera et al. (2016) showing that, in 2015, 85% of the large-scale agriculture in the MATOPIBA region was based on single-cropping systems (SPERA et al., 2016). “First crop / Winter crop” class had high Q3 and Q4 values. As explained before, Q3 and Q4 correspond to the area of curve in the 1st and 4th quadrants of the EVI time series plotted on polar coordinates (Figure. 5.5).

Figure. 5.5 – Pixel-wise EVI time series plotted on polar coordinates using the representation proposed by Körting et al. (2013) for the classes of level 3 classification (Crop Group).



In Figure 5.5, the curves for the “First crop / Winter crop” are considerably pronounced in the 1st and 4th quadrants, which represent the periods from early August to mid-November and June to October. This is explained by the fact that the average season of the winter crops extends from April to October, when the other Crop Group classes are either still being planted, greening-up or in senescence and being harvested.

In general, the phenometrics explain what was shown in the EVI temporal profiles, and they conform to our expectation based on the crop calendars, field observations, and reports by other authors (ARVOR et al., 2011; OLIVEIRA et al., 2014). They also demonstrate their potential for the description of phenological parameters, which is important for many applications in agriculture, such as risk assessment for rural credit, yield estimation, and understanding the effects of policy, trade, and global and technological changes on food security.

5.3.3. Land cover and agricultural mapping

The overall accuracy and class f1-scores for each hierarchical level are presented in Table 5.2 (see Tables B1, B2, B3 and B4 in Appendix B for the complete confusion matrices of each classification level). The ntree parameter values of each RF classification model (Level 2, 3 and 4) were respectively 50, 70 and 90.

Figure 5.6 presents the map of regions A1 and A2 (western Bahia) obtained at level 2 (Land cover) and level 3 (Crop Group) of the hierarchical classification. The cropping patterns provide information about the spatial distribution of the croplands.

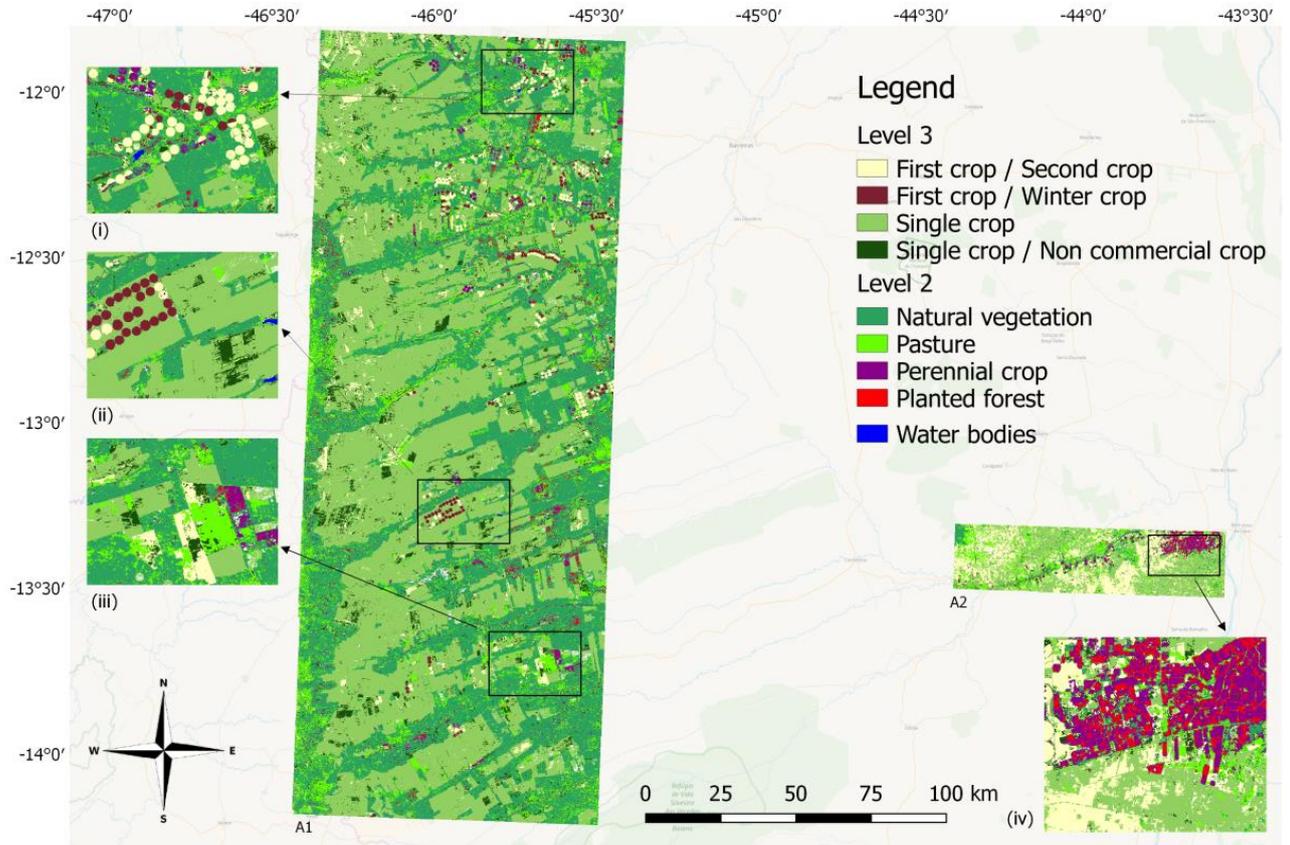
Table 5.2. Overall accuracy and class wise f1-scores obtained for the different hierarchical classification levels.

L1 (0.979)*	L2 (0.993 and 0.912)**	L3 (0.964)*	L4 (0.951)*
<i>Annual crop and Semi-perennial crop (0.988)</i>	<i>Annual crop (0.996)</i>	<i>First crop / Second crop (0.981)</i>	<i>Soy / Bean (0.737)</i>
			<i>Soy / Maize (0.956)</i>
			<i>Soy / Cotton (0.958)</i>
			<i>Soy / Soy (0.999)</i>
			<i>Maize / Soy (0.996)</i>
		<i>First crop / Winter crop (0.993)</i>	<i>Maize / Carrot (0.994)</i>
			<i>Maize / Onion (0.979)</i>
			<i>Maize / Potato (0.996)</i>
			<i>Maize / Beans (0.993)</i>
			<i>Soy / Potato (0.995)</i>
		<i>Single crop (0.966)</i>	<i>Maize (0.917)</i>
			<i>Soy (0.962)</i>
			<i>Cotton (0.950)</i>
			<i>Millet (0.654)</i>
<i>Sorghum (0.524)</i>			
<i>Single crop / Non-commercial crop (0.875)</i>	<i>Cotton / Millet (0.762)</i>		
	<i>Soy / Millet (0.761)</i>		
	<i>Soy / Brachiaria (0.854)</i>		
	<i>Soy / Sorghum (0.566)</i>		
	<i>Soy / Weed (0.901)</i>		
	<i>Maize / Weed (0.964)</i>		
	<i>Maize / Regrowing (0.976)</i>		
<i>Perennial crop and Non-crop (0.923)</i>	<i>Semi-perennial crop (0.926)</i>		
	<i>Perennial crop (0.915)</i>		
	<i>Planted forest (0.901)</i>		
	<i>Pasture (0.942)</i>		
	<i>Forest (0.909)</i>		
	<i>Natural grasslands (0.848)</i>		
	<i>Savanna (0.872)</i>		

* The overall accuracy of the model for the respective classification level.

** The overall accuracies of the model for the classification level of the respective hierarchical classes.

Figure 5.6 – Map of regions A1 and A2, obtained at level 2 (Land Cover: Perennial crops and Non-crop classes) and level 3 (Crop Group) of the hierarchical classification approach. The insets (i), (ii), (iii) and (iv) show zoomed-in areas of the maps. For the production of the maps, we grouped the different classes of natural vegetation into one.



This map also shows that although the single-cropping system is not recommended, because it increases soil and water pollution and reduces soil fertility (CARVALHO et al., 2009), it is still being widely used in this region. The soils in this region are sandier and less physically suitable than in Mato Grosso, and the region receives less rainfall; these aspects pose a challenge to adopting double-cropping regimes (SPERA et al., 2017). Double-cropping systems occur mostly in irrigated areas or in the western side of the region A1, where precipitation is higher (DOURADO et al., 2013). Perennial crops were mapped in the northern study area, especially in irrigated areas (usually coffee). An area in the eastern part of region A2 also has a large concentration of perennial crops, which, according to our field knowledge, is mostly banana and papaya.

The presence of perennial crops in this area may be related to this region's proximity to the São Francisco River, making it better for irrigation. Confusion between perennial crops and planted forest is possibly due to the evergreen phenological behavior of these classes (Figure 5.2). The Natural grasslands, Savanna and Pasture classes contain much salt-and-pepper noise (Figure 5.7). A large area of natural vegetation was included in the Annual crop class in the southern area of the region.

We observed high amplitude in the EVI time series for these areas, which suggests a great seasonal variation. We believe that more studies should investigate whether this result is related to this area's proximity to transition zones between the Cerrado and Caatinga biomes. Moreover, some patterns of deforestation are visible—the clearing of forest vegetation, leaving a predominance of grasses and shrubs, may increase seasonality (SANO et al., 2010). We used the Modified Normalized Difference Water Index (MNDWI) (XU, 2005) to derive a mask of the water bodies; this revealed some small dams, most of which are close to center-pivot irrigated areas (see insets in Figure 5.6).

Figure 5.7 presents the map of region B (southeast of Mato Grosso) obtained at levels 2 and 3 of the hierarchical classification. The Crop Group map shows that most of the agricultural areas are based on double-cropping systems (First crop / Second crop class), but some small areas of single cropping and of single non-commercial crops remain, mostly in the northeast.

Figure 5.7 – Map of region B, obtained at level 2 (Land Cover: Perennial crops and Non-crop classes) and level 3 (Crop Group) of the hierarchical classification approach. For the production of the maps, we grouped the different classes of natural vegetation into one.

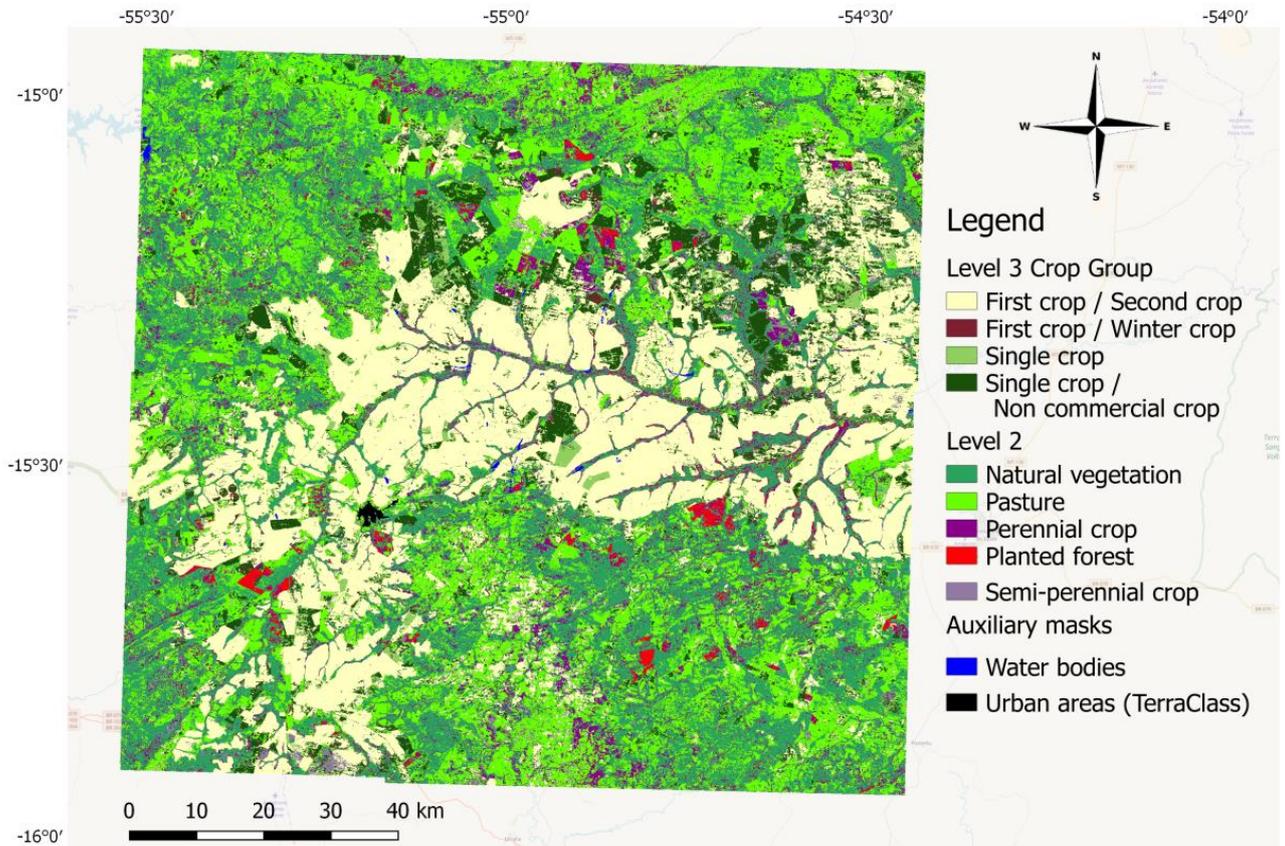
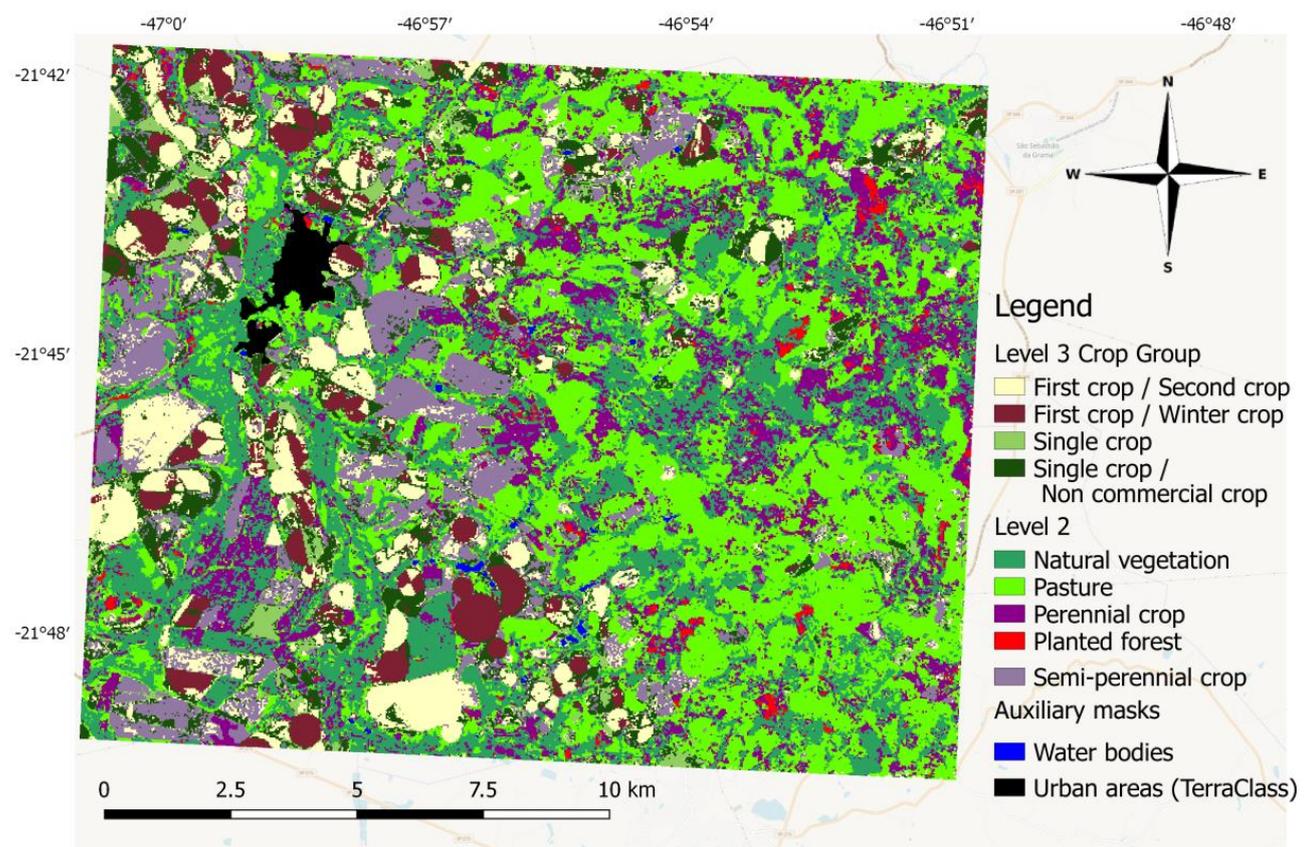


Figure 5.7 illustrates the predominance of pasture in the peripheral areas, though there is some confusion between natural vegetation types. We can see some patterns of noise on the west side of this area, the shape of which can be compared to SLC stripes. We found that pixels with high cloud intensity, combined with SLC noise, result in less clean observations; consequently, the RBF algorithm fails to correctly approximate the time series. Unexpectedly some perennial crops are visible in the south. Inspection of the Google Earth images verified some confusion between perennial crops and forest in the south. This can be confirmed by analyzing the confusion matrix for the respective classes. We can also observe semi-perennial crops in the south, which includes the municipality of Jaciara, one of the biggest sugarcane producers of Mato Grosso State (IBGE, 2016). This demonstrates the capacity of the proposed method to map semi-perennial crops. Figure 5.8 presents the map of region C (northeast of São

Paulo), obtained at levels 2 and 3. The map shows a great heterogeneity of classes, because this is a smallholder agricultural zone, where farmers grow different crops throughout the year. Most of the agricultural areas are concentrated on the western side of the study area, which belongs to the Cerrado biome and is characterized by a flat topography. The scenario in this area is one of intensive agriculture, with double-cropping systems (“First crop / Second crop” and “First crop / Winter crop” classes) using center-pivot irrigation, and different crop systems are adopted within the same area. Semi-perennial crops, as well as some perennial crops, are also observed. The perennial crops here are mostly citrus and coffee (on the eastern side).

Figure 5.8 – Map of region C, obtained at level 2 (Land Cover: Perennial crop and Non-crop classes) and level 3 (Crop Group) of the hierarchical classification approach. For the production of the maps, we grouped the different classes of natural vegetation into one.



In the eastern region, we observed some confusion between forest and perennial crops, which may be related to the transition into the Atlantic

Forest biome as well as to the irregular terrain, the shadow effects of which may affect the EVI time series. Pasture areas were consistently mapped in the east of the study region.

We selected three municipalities within the study regions to visualize the level 4 classification results. The municipalities are Luís Eduardo Magalhães-BA (region A1), Campo Verde-MT (region B) and Itobi-SP (region C). Figure 5.9 shows the map of the respective municipalities with the level 2, 3 and 4 classifications.

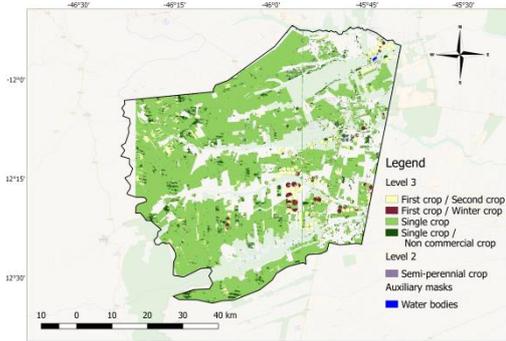
Campo Verde presented a consolidated agricultural scenario, with a large area of double-cropping systems of soy/maize or soy/cotton. This finding is in agreement with other studies (ARVOR et al., 2011; OLIVEIRA et al., 2014). The northeast features single cropping and single crop / non-commercial crop systems. We also mapped rotation systems of soy and potatoes in some areas. Although we did not collect any samples of potato crops in this region, other authors found increased potato production in irrigated systems in the area (DOURADO et al., 2012).

Soy planted in a single-cropping system was the main crop mapped in western Bahia, which is in line with our field observations. The central-east region, however, has large areas of cotton crops and double-cropping systems of soy with maize in center-pivot irrigated areas. In Itobi municipality in São Paulo, we find mainly double-cropping of soy or maize with potatoes in center-pivot irrigated areas, but there are also smaller areas of double-cropping systems with winter crops such as beans and onions.

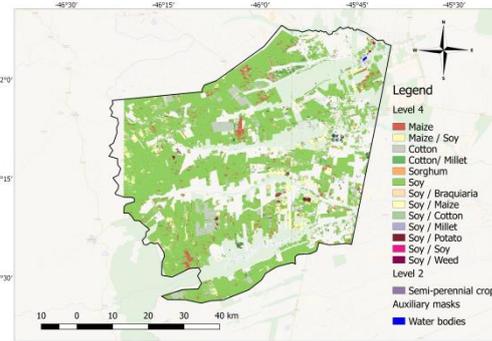
The map also shows areas with rotation of soy and cotton, but since we did not observe any cotton crops during our field surveys, we consider this to be a classification error. We can confirm this by the confusion matrix, which shows confusion between double-cropping systems of soy/maize and soy/cotton.

Figure 5.9 – Classification results based on seasonal phenological parameters for the municipalities of Luís Eduardo Magalhães-BA (region A1), Campo Verde-MT

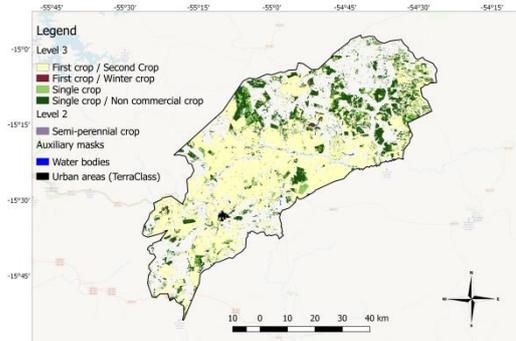
(region B) and Itobi-SP (region C), obtained at level 2 (Land Cover: “Annual crop and Semi-Perennial crop” only), level 3 (Crop Group) (a, c, e) and level 4 (Crop Rotation) (b, d, f) of the hierarchical classification approach.



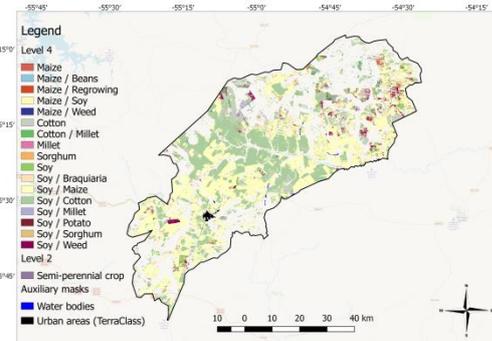
(a)



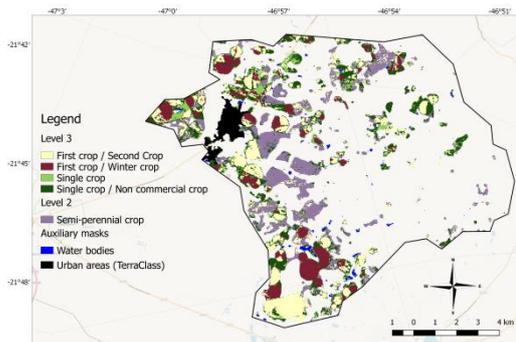
(b)



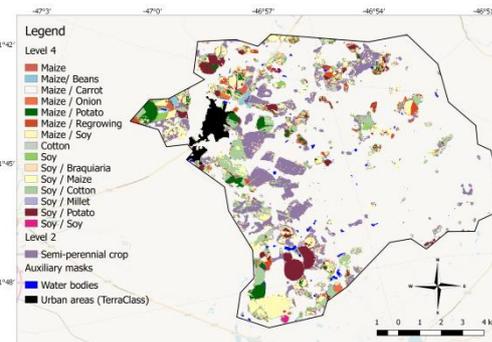
(c)



(d)



(e)



(f)

5.3.4. Outlook

Our results show that the proposed Landsat-based method of hierarchical classification can map agricultural land use in the Cerrado biome. Furthermore, the results reinforce the method's potential for mapping natural vegetation. Despite the limitation of using TIMESAT to extract phenometrics for semi-perennial crops, the hierarchical classification enabled accurate separation of semi-perennial and annual crops. The use of variables generated with polar representation (Q3 and Q4), which does not depend on season detection, revealed a high potential for increasing the separability for these classes, especially the crop group classification.

This work follows the assumption that our samples covered the most representative crop group classes in the Brazilian Cerrado biome. However, the results showed that a more detailed analysis should be done for the "Single crop / Non-commercial crop" class, with a focus on the different planting periods of the main crop season. Depending on the planting period, the non-commercial crop cycle might be detected in the previous crop year, revealing a potential "Non-commercial crop / Single crop" class. There was also some confusion in regions near areas of transition to different biomes (Caatinga and Atlantic Forest). The EVI time series for these classes showed that the vegetation in these regions has high amplitude due to its large seasonal variation, which may explain this confusion. This fact points to the need for different approaches to mapping agriculture in these areas.

We are aware of the method's dependence on data availability. However, the upcoming availability of Sentinel-2 and CBERS-4 data will increase the density of clear observations, though there will also be challenges in the harmonization of these datasets. Even though these satellites will be essential for overcoming these issues in the future, the presented methodology based on Landsat data still provides options for mapping historical agricultural land use, allowing land-use change mapping and analysis.

6 SYNTHESIS

The main objective of this work was to develop, to describe, and to assess a method based on phenological metrics derived from dense high spatial resolution image time series for classifying, with significantly detail, agricultural land use in the Cerrado biome. The following sections will revise the main contributions of this thesis by answering some scientific questions and then discuss further research topics.

6.1. Can we use CBERS-4 MUX data to remove noise in a Landsat-8 OLI NDVI and EVI image time series to improve a crop classification method based on phenological information?

We observed a slight increase in the classification accuracy when using the method described in Chapter 3, in which we used CBERS-4 MUX data to remove noise in Landsat-8 OLI time series. The results for EVI were consistently more accurate compared to NDVI, reinforcing what was expected following the main findings on the literature, and this result supported our decision on keep working with EVI on the next steps. There is a great potential for using CBERS-4 imagery, but image normalization procedures are needed to make it compatible with Landsat-8, as well it's important to consider that different atmospheric correction methods and ancillary datasets may affected the results. In this thesis we contributed to show the feasibility about integrating CBERS-4 MUX data and Landsat-8 to produce image time series for agricultural land classification. However, this integration process still demands more improvements, and also it is not enough to generate agricultural land classification with higher accuracy. In the same way, the proposed method can be applied for integrating other sensors such as CBERS-4 WFI, Landsat 7 ETM+, and Sentinel MSI.

6.2. Which smoothing method is better for phenological metrics-based crop classification with Landsat-8 OLI EVI image time series in the Cerrado biome?

In Chapter 4 we evaluated three different time series smoothing methods, Savitzky-Golay, asymmetric Gaussian function and Double-logistic function, combined or not with filtering techniques. The evaluation was performed through their use in the agricultural land classification based on the phenological parameters extracted from one-year Enhanced Vegetation Index (EVI) Landsat-8 image time series. The smoothing method that provided the highest classification accuracy was the Savitzky-Golay applied to raw time series (86% and kappa=0.82), followed by the asymmetric Gaussian applied to filtered time series (81% and kappa=0.754).

6.3. Can we integrate Landsat-8 OLI and Landsat-7 ETM+ data to generate a dense high spatial resolution image time series, for classifying agricultural land use in the Cerrado?

In Chapter 5, we used a weighted ensemble of RBF convolution filters as a kernel smoother to approximate missing data, thereby generating an 8-day temporal resolution data cube using data from the Landsat 7/ETM+ and Landsat 8/OLI sensors, as well as deriving the phenometrics to run a RF classification in a hierarchical scheme. We showed that phenometric attributes are feasible to classify agricultural land in the Cerrado biome with high classification accuracy. To our knowledge, this work is the first one to produce detailed, consistent and reproducible agricultural crop maps in the Brazilian Cerrado using Landsat data. The results can also be used to assess the interplay between production and protection in the Brazilian Cerrado biome.

6.4. Are Landsat based phenological metrics suitable to analyze crop management practices of the major crops in the Cerrado?

In chapter 5, we demonstrated that the phenological metrics are feasible to describe phenological information about the major crops that Brazilian farmers grow in the Cerrado biome. This is one important contribution as

this information can be useful for understanding agricultural practices, supporting risk assessment analyses for rural credit, yield estimation, and understanding the effects of policy, trade, and global and technological change on food security.

6.5. Future work

Based on the results of this thesis, we described some specific topics for future work:

- Test multi-sensor integration methods with different sensors, such as CBERS-4 WFI and Sentinel-2A and 2B MSI;
- Develop infrastructure and apply the methodology for the entire Cerrado and for a longer period of time, comparing with the official statistics for the period;
- Relate phenological metrics with productivity estimates;
- Compare with other classification approaches, making use of Deep Learning techniques;
- Compare the maps obtained with official databases.

REFERENCES

ADAMI, M.; GOMES, A. R.; COUTINHO, A. C; ESQUERDO, J. C. D. M.; VENTURIERI, A. Dinâmica do uso e cobertura da terra no estado do Pará entre os anos de 2008 a 2012. In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 17., 2015, João Pessoa, Brazil. **Proceedings...**, São José dos Campos: INPE, 2015. p.7028-7035.

AGROCONSULT. **Rally da safra**. 2016. Available from: <http://www.rallydasafra.com.br/>. Access in: 01 Dec. 2016.

ALCAMO, J.; HASSAN, R. **Ecosystems and human well-being millennium ecosystem assessment**. Washington, US: Island Press, 2003. 212p. (Millennium Ecosystem Assessment).

ALMEIDA, C.; COUTINHO, A.; ESQUERDO, J.; ADAMI, M.; VENTURIERI, A.; DINIZ, C.; DESSAY, N.; DURIEUX, L.; GOMES, A. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazonica**, v. 46, p. 291-302, 2016.

ARVOR, D.; JONATHAN, M.; MEIRELLES, M. S. O. P.; DUBREUIL, V.; DURIEUX, L. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. **International Journal of Remote Sensing**, v. 32, n. 22, p. 7847-7871, 2011.

ARVOR, D.; TRITSCH, I.; BARCELLOS, C.; JÉGOU, N.; DUBREULIA, V. Land sustainability on the South-Eastern Amazon agricultural frontier: recent progress and the challenges ahead. **Applied Geography**, v. 80, p. 86-97, 2017.

ASSUNÇÃO, J.; CHEIN, F. Climate change, agricultural productivity and poverty. **Environment and Development Economics**, v. 21, n. 5, p.581-602, 2016.

ATZBERGER, C.; EILERS, P. H. C. Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements. **International Journal of Remote Sensing**, v. 32, n. 13, p.3689-3709, 2011.

ATZBERGER, C. Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. **Remote Sensing**, v. 5, n. 2, p.949-981, 2013.

BENDINI, H. N.; SANCHES, I. D. A.; KÖRTING, T. S.; FONSECA, L. M. G.; LUIZ, A. J. B.; FORMAGGIO, A. R. Using Landsat 8 image time series for crop mapping in a region of Cerrado, Brazil. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v.41, n. B8, p.845-850, 2016.

BENDINI, H. N.; FONSECA, L. M. G.; KÖRTING, T. S.; SANCHES, I. D. A.; MARUJO, R. Evaluation of smoothing methods on Landsat-8 EVI time series for crop classification based on phenological parameters. In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 18., 2017, Santos, Brazil. **Proceedings...** São José dos Campos: INPE, 2017a. p.4267-4274.

BENDINI, H. N.; FONSECA, L. M. G.; KÖRTING, T. S.; MARUJO, R.; SANCHES, I. D. A.; ARCANJO, J. Assessment of a multi-sensor approach for noise removal on Landsat-8 OLI time series using CBERS-4 MUX data

to improve crop classification based on phenological features. **Brazilian Journal of Cartography**, v. 69, n. 5, p.947-957, 2017b.

BENDINI, H. N.; FONSECA, L. M. G.; SCHWIEDER, M.; KÖRTING, T. S.; RUFIN, P.; SANCHES, I. D. A.; LEITÃO, 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, 101872, 2019.

BORGES, E. F.; SANO, E. E. Séries temporais de EVI do MODIS para o mapeamento de uso e cobertura vegetal do oeste da Bahia. **Bulletin of Geodetic Sciences**, v. 20, n. 3, p.526-547, 2014.

BRASIL. MINISTRY OF AGRICULTURE, LIVESTOCK AND SUPPLY - MAPA. **GDP for agriculture highs 1.8% in 2015. 2016**. Available from: <http://www.agricultura.gov.br/noticias/pib-da-agropecuaria-tem-alta-de-1-8-em-2015>. Access in: Dec. 2017.

BRASIL. MINISTRY OF AGRICULTURE, LIVESTOCK AND SUPPLY - MAPA. **Sector plan for mitigation and adaptation to climate change for the consolidation of a low carbon economy in agriculture: ABC (Low Carbon Agriculture)**. 2012. Available from: <http://www.agricultura.gov.br/assuntos/sustentabilidade/plano-abc/arquivo-publicacoes-plano-abc/download.pdf>. Access in: July 2017.

BRASIL. MINISTRY OF ENVIRONMENT - MMA. **O bioma Cerrado**. Available from: <http://www.mma.gov.br/biomas/cerrado.html/>. Access in: Oct. 2016.

BRAZIL NATIONAL SUPPLY COMPANY - CONAB. **Periodicity of the brazilian crops**. Message received by leda D. A. Sanches on May 2017.

BRAZILIAN INSTITUTE OF GEOGRAPHY AND STATISTICS - IBGE.: **Municipal agricultural production**. 2016. Available from: <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>. Access in: Oct 2016.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n.1, p. 5-32, 2001.

CARVALHO, J. L. N.; CERRI, C. E. P.; FEIGL, B. J.; PÍCCOLO, M. C.; GODINHO, V. P.; UWE, H.; CERRI, C. C. Conversão do cerrado em agricultura no sudoeste da Amazônia: estoques de carbono e fertilidade do solo. **Scientia Agricola**, v. 66, n. 2, p.233-241, 2009.

CENTER FOR ADVANCED STUDIES IN APPLIED ECONOMICS OF SÃO PAULO UNIVERSITY; CONFEDERATION OF AGRICULTURE AND LIVESTOCK OF BRAZIL - CEPEA-USP/CNA. **Agribusiness GDP - data from 1994 to 2011**. 2013. Available on: <https://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>. Access in: 10 Nov. 2015.

- CHAN, J. C. W.; PAELINCKX, D. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. **Remote Sensing of Environment**, v. 112, n. 6, p. 2999-3011, 2008.
- CHEN, J.; JÖNSSON, P.; TAMURA, M.; GUA, Z.; MATSUSHITA, B.; EKLUNDH, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. **Remote Sensing of Environment**, v. 91, n. 3-4, p.332-344, 2004.
- CLARK, M. L.; AIDE, T. M.; GRAU, H. R.; RINER, G. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: a case study in the Dry Chaco ecoregion of South America. **Remote Sensing of Environment**, v. 114, n. 1, p.2816-2832, 2010.
- CONWAY, G.; TOENNIESSEN, G. Feeding the world in the twenty-first century. **Nature**, v. 402, n. 55/58, 1999.
- COUTINHO, A. C.; ESQUERDO, J. C. D. M.; OLIVEIRA, L. S.; LANZA, D. A. Methodology for systematical mapping of annual crops in Mato Grosso Do Sul State (Brazil). **Geografia**, Rio Claro, v. 38, n.1, p.45-54, 2013.
- DINGKUHN, M.; LE GAL, P., Effect of drainage date on yield and dry matter partitioning in irrigated rice. **Field Crops Research**, v. 46, n. 1/3, p.117-126, 1996.
- DOURADO, C. S.; OLIVEIRA, S. R. M.; AVILA, A. M. H. Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia. **Bragantia**, v. 72, n. 2, p.192-198, 2013.
- DOURADO, C. S.; OLIVEIRA, S. R. M.; AVILA, A. M. H. Classificação de anos secos e chuvosos em zonas pluviometricamente homogêneas no Estado da Bahia. In: INTERNSHIP COLLOQUIUM OF EMBRAPA AGRICULTURAL INFORMATICS, 8., 2012, Campinas, Brazil. **Proceedings...** Campinas: EMBRAPA, 2012. p. 51-54.
- ESQUERDO, J. C. D. M.; ZULLO JÚNIOR, J.; ANTUNES, J. F. G. Use of NDVI/AVHRR time-series profiles for soybean crop monitoring in Brazil. **International Journal of Remote Sensing**, v. 32, n. 13, p.3711 – 3727, 2013.
- FEARNSIDE P.M. Desmatamento na Amazônia: dinâmica, impactos e controle. **Acta Amazônica**, v. 36, n. 3, 2006.
- FERREIRA, M. E.; FERREIRA, J. L. G.; MIZIARA, F.; SOARES-FILHO, B. Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservatio. **Journal of Land Use Science**, v. 8, n.4, p.403-421, 2012.
- FONSECA-MORELLO, T. Agricultural use of fire in the Brazilian Amazon: assessing the role of farms' boundaries. In: BRAZILIAN SYMPOSIUM ON

REMOTE SENSING, 17., 2015, João Pessoa, Brazil. **Proceedings...** São José dos Campos: INPE, 2015. p. 5058-5064.

GALFORD, G. L.; SOARES-FILHO, B.; CERRI, C. E. P. Prospects for land-use sustainability on the agricultural frontier of the Brazilian Amazon. **Philosophical Transactions of the Royal Society**, v. 368, n. 1619, p.1-5, 2013.

GERACE, A.; MONTANARO, M.; CONNALL, R. Leveraging intercalibration techniques to support stray-light removal from Landsat 8 Thermal Infrared Sensor data. **Journal of Applied Remote Sensing**, v. 12, n. 1, p.1-13, 2017.

GIBBS, H. K.; RAUSCH, L.; MUNGER, J.; SCHELLY, I.; MORTON, D. C.; NOOJIPADY, P.; SOARES-FILHO, B.; BARRETO, P.; MICOL, L.; WALKER, N. F. Brazil's Soy Moratorium. **Science**, v. 347, n. 6220, p.377-378, 2015.

GODFRAY, H. C. J.; BEDDINGTON, J. R.; CRUTE, I. R.; HADDAD, L.; LAWRENCE, D.; MUIR, J. F. Food security: the challenge of feeding 9 billion people. **Science**, v. 327, p.812-818, 2010.

HALL, M. A.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; EUTEMANN, P.; WITTEN, I. H. The WEKA data mining software: an update. **SIGKDD Explorations**, v. 11, n. 1, p.10-18, 2009.

HAMUNYELA, E.; VERBESSELT, J.; ROERINK, G.; HEROLD, M. Trends in spring phenology of Western European deciduous forests. **Remote Sensing**, v. 5, n. 12, p.6159-6179, 2013.

HOLDEN, E. C.; CURTIS, E. W. An analysis of Landsat 7 and Landsat 8 under flight data and the implications for time series investigations. **Remote Sensing of Environment**, v. 185, p.16-36, 2016.

HUETE, A.; JUSTICE, C.; VAN LEEUWEN, W. **MODIS Vegetation Index (MOD13) Algorithm Theoretical Basis Document, version 3**. 129p. Available from: http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf.

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.P., GAO, X.; FERREIRA, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote Sensing of Environment**, v. 83, n. 1/2, p.195-213, 2002.

IRISH, R.R.; BARKER, J.L.; GOWARD, S.N.; ARVIDSON, T. Characterization of the Landsat-7 ETM + automated cloud-cover assessment (ACCA) algorithm. **Photogrammetric Engineering and Remote Sensing**, v. 72, n. 10, p.1179-1188, 2006.

JÖNSSON, P.; EKLUNDH, L. Seasonality extraction by function fitting to time-series of satellite sensor data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 40, n. 8, p.1824-1831, 2002.

JÖNSSON, P.; EKLUNDH, L. TIMESAT - a program for analyzing time-series of satellite sensor data. **Computer & Geoscience**, v. 30, n. 8, p.833-845, 2004.

JÖNSSON, P.; EKLUNDH, L. **TIMESAT 3.2 with parallel processing software manual**. Lund, Sweden: Lund University, 2015. 88p. Available from:
http://web.nateko.lu.se/timesat/docs/TIMESAT32_software_manual.pdf.

KIMBALLA, J. S.; MCDONALD, K. C.; RUNNING, S. W.; FROLKING, S. E. Satellite radar remote sensing of seasonal growing seasons for boreal and subalpine evergreen forests. **Remote Sensing of Environment**, v. 90, n. 2, p.243-258, 2004.

KÖRTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. GeoDMA:Geographic Data Mining Analyst. **Computers & Geosciences**, v. 57, p.133-145, 2013.

LAWRENCE, R. L.; WOOD, S. D.; SHELEY, R. L. Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (randomforest). **Remote Sensing of Environment**, v. 100, n. 3, p.356-362, 2006.

LEBOURGEOIS, V.; DUPUY, S.; VINTROU, E.; AMELINE, M.; BUTLER, S.; BÉGUÉ, A. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). **Remote Sensing**, v.9, n. 3, p.259-279, 2017.

LEWIS, J. T. Agriculture in Brazil's one bright spot. **The Wall Street Journal**. Available from: <https://www.wsj.com/articles/agriculture-is-brazils-one-bright-spot-1460157172>. Access in: Apr. 8, 2016.

LIAW, A.; WIENER, M. Classification and regression by randomforest. **R News**, v. 2, n .3, p.18-22, 2002.

LIU, H. Q.; HUETE, A. Feedback based modification of the NDVI to minimize canopy background and atmospheric noise. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, n. 2, p.457-465, 1995.

LUIZ, A. J. B.; SANCHES, I. D.; TRABAQUINI, K.; EBERHARDT, I. D. R.; FORMAGGIO, A. R. Dinâmica agrícola em área de sobreposição de órbitas adjacentes dos satélites Landsat. In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 17., 2015, João Pessoa, Brazil. **Proceedings...** São José dos Campos: INPE, 2015. p. 1308-1315.

LYMBURNER, L.; BOTHA, E.; HESTIR, E.; ANSTEE, J.; SAGAR, S.; DEKKER, A.; MALTHUS, T. Landsat 8: providing continuity and increased precision for measuring multidecadal time series of total suspended matter. **Remote Sensing of Environment**, v. 185, p.108-118, 2016.

MARUJO, R. F. B.; FONSECA, L. M. G.; KORTING, T. S.; BENDINI, H. N.; QUEIROZ, G. R.; VINHAS, L.; FERREIRA, K. R. Remote sensing

image processing functions in Lua language. **Journal of Computational Interdisciplinary Sciences**, v. 8, n. 3, p.163-172, 2017.

MASEK, J.G.; VERMOTE, E.F.; SALEOUS, N.E.; WOLFE, R.; HALL, F.G.; HUENNERICH, K.F. A Landsat surface reflectance dataset for North America, 1990-2000. **IEEE Geoscience and Remote Sensing Letters**, v. 3, n. 1, p.68-72, 2006.

MAXWELL, S. K.; SCHMIDT, G. L.; STOREY, J. C. A multi-scale segmentation approach to filling gaps in Landsat ETM+ SLC-off images. **International Journal of Remote Sensing**, v. 28, n. 23, p.5339-5356, 2007.

MICIJEVIC, E.; OBAIDUL, H. MD., MISHRA, N. Radiometric calibration updates to the Landsat collection, In: CONFERENCE FOR EARTH OBSERVING SYSTEMS, SPIE OPTICAL ENGINEERING + APPLICATIONS 99720D, 21., 2016, San Diego, United States. **Proceedings...** San Diego, 2016. p.1-12.

MÜLLER, H., RUFIN, P., GRIFFITHS, P., SIQUEIRA, A. J. B., HOSTERT, P. Mining dense Landsat time-series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, p.490-499, 2015.

NATIONAL INSTITUTE FOR SPACE RESEARCH - INPE. **Mapping of land use and vegetation cover in Cerrado**. 2017. Available from: <http://www.dpi.inpe.br/tccerrado/index.php?mais=1>. Access in: Mar. 2017.

NEVES, A. K.; BENDINI, H. N.; KÖRTING, T. S.; FONSECA, L. M. G. Combining time series features and data mining to detect land cover patterns: a case study in northern Mato Grosso State, Brazil. **Brazilian Journal of Cartography**, v. 68, n. 6, p.1133-1142, 2016.

NGUY-ROBERTSON, A.; GITELSON, A.; PENG, Y.; VIÑA, A.; ARKEBAUER, T.; RUNDQUIST, D. Green leaf area index estimation in maize and soybean: combining vegetation indices to achieve maximal sensitivity. **Agronomy Journal**, v. 104, n. 5, p.1336-1347, 2012.

OLIVEIRA, J. C.; TRABAQUINI, K.; EPIPHANIO, J. C. N.; FORMAGGIO, A. R.; GALVÃO, L. S.; ADAMI, M. Analysis of agricultural intensification in a basin with remote sensing data. **GIScience & Remote Sensing**, v. 51, n. 3, p.253-268, 2014.

PAL, M. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, v. 26, n. 1, p.217-222, 2005.

PAN, Z.; HUANG, J.; ZHOU, Q.; WANG, L.; CHENG, Y.; ZHANG, H.; BLACKBURN, G. A.; YAN, J.; LIU, J. Mapping crop phenology using NDVI time-series derived from HJ-1A/B data. **International Journal of Applied Earth Observation and Geoinformation**, v. 34, p.188-197, 2015.

PEÑA, M.A.; BRENNING, A. Assessing fruit tree crop classification from Landsat-8 time-series for the Maipo Valley, Chile. **Remote Sensing of Environment**, v. 171, p. 234-244, 2015.

PINTO, C.; PONZONI, F.; CASTRO, R., LEIGH, L.; MISHRA, N.; AARON, D.; HELDER, D. First in-flight radiometric calibration of MUX and WFI on-Board CBERS-4. **Remote Sensing**, v. 8, n. 5, p.1-22, 2016.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, 2017. Available from: <http://www.R-project.org>. Access in: Apr. 2017.

RATANA, P.; HUETE, A. R.; FERREIRA, L. Analysis of Cerrado physiognomies and conversion in the MODIS seasonal-temporal domain. **Earth Interactions**, v. 9, p.1-22, 2005.

REED, B. C.; BROWN, J. F.; VANDERZEE. Measuring phenological variability from satellite imagery. **Journal of Vegetation Science**, v. 5, n. 5, p.703-714, 1994.

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P.; RIBEIRO, J. F. (Eds.). **Cerrado: ecologia e flora**. Brasília: Embrapa Cerrados, 2008. p.153-212.

RISSO, J.; RUDORFF, B. F. T.; ADAMI, M.; AGUIAR, A. P. D.; FREITAS, R. M. MODIS time series for land use change detection in fields of the Amazon soy moratorium. In: INTERNATIONAL ARCHIVES OF THE PHOTOGRAMMETRY, REMOTE SENSING AND SPATIAL INFORMATION SCIENCES, 23., Melbourne, Australia. **Proceedings...** 2012. p.339-344.

RODRIGUEZ-GALIANO, V. F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL-SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 67, p.93-104, 2012.

RUBINSTEIN, R.; KROESE, D. **Simulation and the Monte Carlo Method**. 3.ed. Hoboken, US: Wiley-Interscience, 2008. 340p.

RUDORFF, B. F. T.; AGUIAR, D. A.; SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo State (Brazil) using Landsat data. **Remote Sensing**, v. 2, n. 4, p.1057-1076, 2010.

RUFIN, P.; MULLER, H.; PFLUGMACHER, D.; HOSTERT, P. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 41, p.1-10, 2015.

SAKAMOTO, T.; YOKOZAWA, M.; TORITANI, T.; SHIBAYAMA, M.; ISHITSUKA, N.; OHNO, H. A crop phenology detection method using

time-series MODIS data. **Remote Sensing of Environment**, v. 96, n. 3/4, p.366-374, 2005.

SANCHES, I. D. A.; FEITOSA, R. Q.; DIAZ, P. M. A.; SOARES, M. D.; LUIZ, A. J. B.; SCHULTZ, B.; MAURANO, L. E. P. Campo Verde database: seeking to improve agricultural remote sensing of tropical areas. **IEEE Geoscience And Remote Sensing Letters**, v. 15, n. 3, p. 369-373, 2018.

SANO, E. E.; ROSA, R.; BRITO, J. L.; FERREIRA, L. G. Land cover mapping of the tropical savanna region in Brazil. **Environment Monitoring and Assessment**, v. 166, n.1/4, p.113-124, 2010.

SCHWIEDER, M.; LEITÃO, P. J.; BUSTAMANTE, M. M. C.; FERREIRA, L. G.; RABEA, A.; HOSTERT, P. Mapping brazilian savanna vegetation gradients with Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 52, p.361-370, 2016.

SHANG, R.; LIU, R.; XU, M.; LIU, Y.; ZUO, L.; GE, Q. The relationship between the threshold-based and the inflexion-based approaches in extraction of land surface phenology. **Remote Sensing of Environment**, v. 199, p.167-170, 2017.

SHAO, Y.; LUNETTA, R. S.; WHEELER, B.; IIAMES, J. S.; CAMPBELL, J. B. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. **Remote Sensing of Environment**, v. 174, p. 258 – 265, 2016.

SHEN, H.; LI, X.; CHENG, Q.; ZENG, C.; YANG, G.; LI, H.; ZHANG, L. Missing information reconstruction of remote sensing data: a technical review. **IEEE Geoscience and Remote Sensing Magazine**, v. 3, n. 3, p.61-85, 2015.

SPERA, S. A.; GALFORD, G. L.; COE, M. T.; MACEDO, M. N.; MUSTARD, J. F. Land use change affects water recycling in Brazil's last agricultural frontier. **Global Change Biology**, v. 22, n. 10, p.3405-3413, 2016.

SPERA, S. Agricultural Intensification can preserve the Brazilian Cerrado: applying lessons from Mato Grosso and Goiás to Brazil's last agricultural frontier. **Journal of Tropical Conservation Science**, v. 10, n. 1, p. 1-7, 2017.

STRASSBURG, B. B. N.; BROOKS, T.; FELTRAN-BARBIERI, R.; IRIBARREM, A.; CROUZEILLES, R.; LOYOLA, R.; LATAWIEC, A.E.; OLIVEIRA FILHO, F.J.B.; SCARAMUZZA, C.A.D.M.; SCARANO, F.R.; SOARES-FILHO, B.; BALMFORD, A. Moment of truth for the Cerrado hotspot. **Nature Ecology & Evolution**, v. 1, n. 99, p.1-3., 2017.

TILMAN, D.; CASSMAN, K.; MATSON, P.; NAYLOR, R.; POLASKY, S. Agricultural sustainability and intensive production practices. **Nature**, v. 418, n. 6898, p.671-677, 2002.

TOMÁS, J. C.; FARIA, F. A.; ESQUERDO, J. C. D. M.; COUTINHO, A. C.; MEDEIROS, C. B. SiRCub – Brazilian Agricultural Crop Recognition System. In: BRAZILIAN SYMPOSIUM ON REMOTE SENSING, 17., 2015, João Pessoa, Brazil. **Proceedings...** São José dos Campos: INPE, 2015. p. 6273-6280.

UNITED STATES GEOLOGICAL SURVEY - USGS. **Landsat surface reflectance level-2 science products**. 2016. Available from: <https://landsat.usgs.gov/landsat-surface-reflectance-data-products>. Access in: Apr. 2017.

VERMOTE, E.; TANRE, D.; DEUZE, J.; HERMAN, M.; MORCETTE, J.J. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 3, p.675-686, 1997.

VERMOTE, E.; JUSTICE, C.; CLAVERIE, M.; FRANCH, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. **Remote Sensing of Environment**, v. 185, p.46-56, 2016.

WARDLOW, B. D.; EGBERT, S. L. KASTENS; J. H. Analysis of time-series MODIS 250m vegetation index data for crop classification in the US central Great Plains. **Remote Sensing of Environment**, v. 108, n. 3, p.290-310, 2007.

WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3. ed. San Francisco, US: Morgan Kaufmann, 2011. 622p.

WULDER, M. A.; MASEK, J. G.; COHEN, W. B.; LOVELAND, T. R.; WOODCOCK, C. E. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. **Remote Sensing of Environment**, v. 118, p.127-139, 2012.

XU, H. A study on information extraction of water body with the Modified Normalized Difference Water Index (MNDWI). **Journal of Remote Sensing**, v. 9, n. 5, p.511-517, 2005.

ZHANG, X.; FRIEDL, M. A.; SCHAAF, C. B. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, n. 3, p.471-475, 2003.

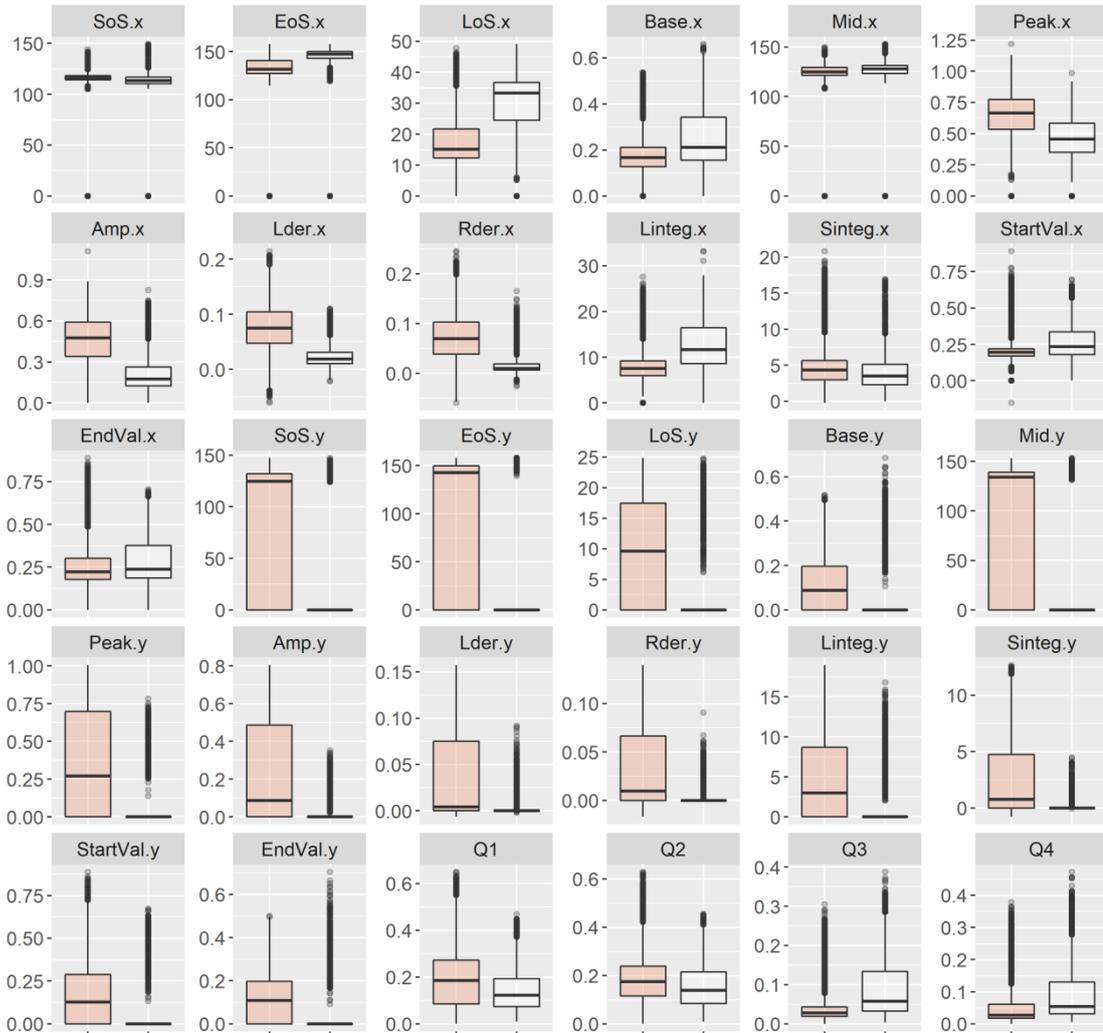
ZHENG, B.; MYINT, S. W.; THENKABAIL, P. S.; AGGARWAL, R. M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. **International Journal of Applied Earth Observation and Geoinformation**, v. 34, p.103-112, 2015.

ZHU, Z.; WOODCOCK, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. **Remote Sensing of Environment**, v.118, p.83–94, 2012.

ZHU, Z.; WANG, S.; WOODCOCK, C. E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. **Remote Sensing of Environment**, v. 159, p.269-277, 2015.

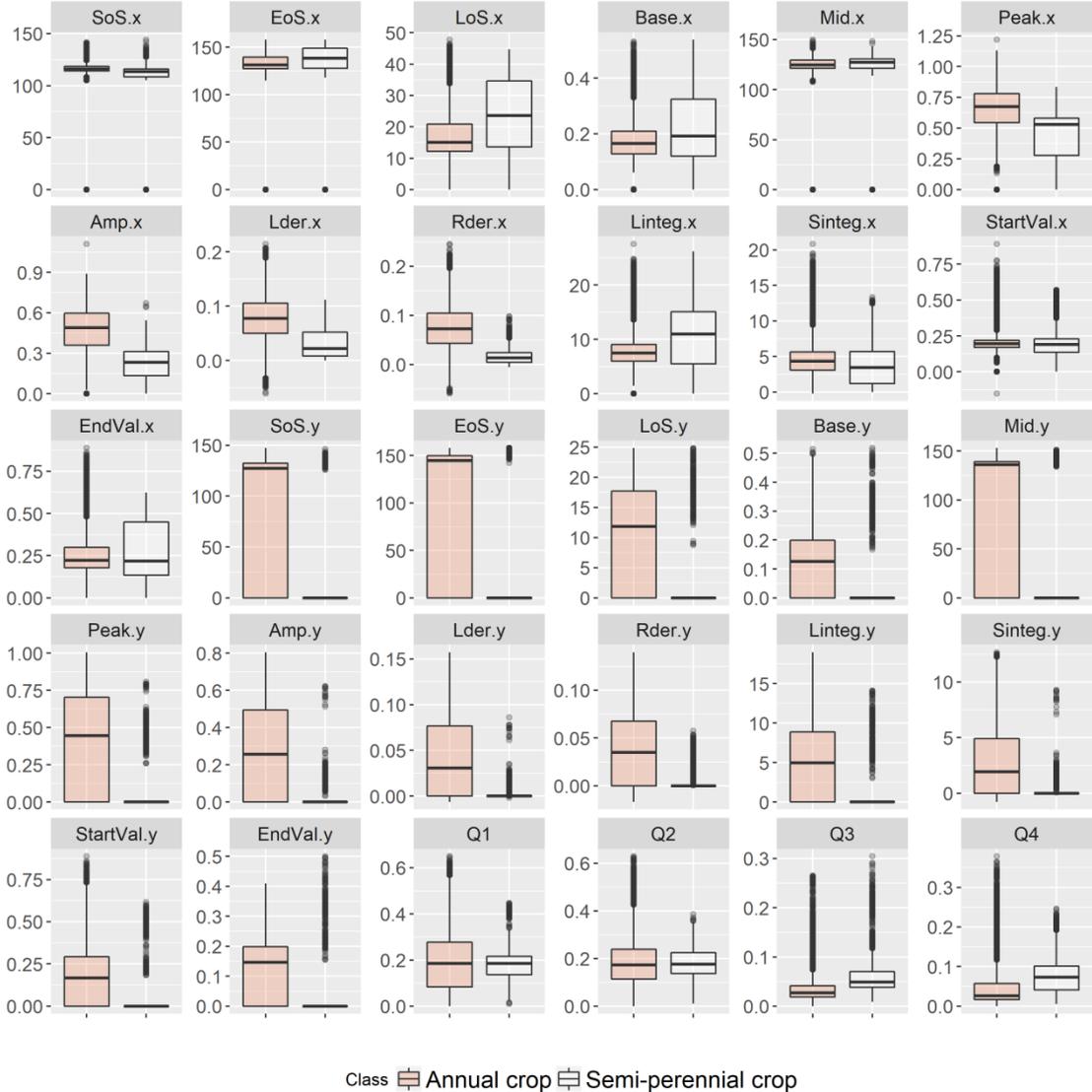
Appendix A

Figure A1. Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 1 classification (Annual crop and Semi-perennial crop; Perennial crop and Non crop) based on the training pixels.



Class  Annual crop and Semi-perennial crop  Perennial crop and Non crop

Figure A2. Boxplots with the mean, 25 and 75 percentiles of EVI phenological parameters for level 2 classification (Annual crop and Semi-perennial crop) based on the training pixels.



Appendix B

Table B1. Mean Confusion Matrix of classification Level 1 (Annual crop and Semi-perennial crop) resulting from the Monte Carlo process, with 1000 simulations.

Prediction	Reference	
	<i>Annual crop and Semi-perennial crop</i>	<i>Perennial crop and Non crop</i>
<i>Annual crop and Semi-perennial crop</i>	13847,31	113,11
<i>Perennial crop and Non crop</i>	229,49	2055,08

Table B2. Mean Confusion Matrix of classification Level 2 (Annual crop and Semi-perennial crop) resulting from the Monte Carlo process, with 1000 simulations.

Prediction	Reference	
	<i>Annual crop</i>	<i>Semi-perennial crop</i>
<i>Annual crop</i>	13235.04	32.34
<i>Semi-perennial crop</i>	68.25	623.37

Table B3. Mean Confusion Matrix of classification Level 3 (First crop / Second crop, First crop / Winter crop, Single crop, and Single / Non-commercial crop) resulting from the Monte Carlo process, with 1000 simulations.

Prediction	Reference			
	<i>First crop / Second crop</i>	<i>First crop / Winter crop</i>	<i>Single crop</i>	<i>Single / Non-commercial crop</i>
<i>First crop / Second crop</i>	6999,33	0,04	24,68	55,04
<i>First crop / Winter crop</i>	2,69	485,21	0,64	2,30
<i>Single crop</i>	38,20	0,686	3818,37	92,73
<i>Single / Non-commercial crop</i>	144,89	0,16	114,38	1361,67

Table B4. Mean Confusion Matrix of classification Level 4 (Crop Rotation) resulting from the Monte Carlo process, with 1000 simulations.

Prediction	Reference																					
	Maize	Maize / Beans	Maize / Carrot	Maize / Onion	Maize / Potato	Maize / Regrowing	Maize / Soy	Maize / Weed	Cotton	Cotton / Millet	Millet	Sorghum	Soy	Soy / Bean	Soy / Braquiaria	Soy / Maize	Soy / Cotton	Soy / Millet	Soy / Potato	Soy / Sorghum	Soy / Soy	Soy / Weed
Maize	649,52	0,00	0,00	0,00	0,00	0,00	0,00	0,00	4,20	0,29	0,00	0,00	73,57	0,00	0,00	2,25	0,02	0,16	0,00	0,00	0,00	0,01
Maize / Beans	0	93,35	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,34	0,40	0,32	0,19	0,00	0,00	0,00	0,00
Maize / Carrot	0	0,01	26,00	0,00	0,00	0,001	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,31	0,00	0,00	0,00
Maize / Onion	0	0,00	0,00	30,31	0,00	0,31	0,00	0,00	0,00	0,12	0,00	0,00	0,024	0,00	0,01	0,01	0,00	0,00	0,02	0,00	0,00	0,00
Maize / Potato	0,00	0,00	0,00	0,00	61,00	0,00	0,00	0,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Maize / Regrowing	0,21	0,00	0,00	0,77	0,00	40,63	0,00	0,00	0,01	0,01	0,00	0,00	0,35	0,00	0,28	0,02	0,10	0,01	0,00	0,00	0,00	0,00
Maize / Soy	0,00	0,00	0,00	0,00	0,00	0,00	122,81	0,00	0,34	0,00	0,00	0,00	0,05	0,00	0,00	0,29	0,00	0,00	0,00	0,00	0,00	0,00
Maize / Weed	0,00	0,00	0,00	0,00	0,27	0,00	0,00	7,23	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	0,00
Cotton	0,61	0,00	0,00	0,00	0,00	0,00	0,00	0,00	889,96	0,08	0,00	0,07	5,80	0,00	0,78	11,71	19,06	2,40	0,00	0,00	0,00	0,39
Cotton / Millet	0,23	0,00	0,00	0,01	0,00	0,01	0,00	0,00	0,80	13,27	0,02	0,00	0,45	0,00	1,52	0,04	0,10	0,69	0,01	0,00	0,00	1,26
Millet	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,03	0,89	4,26	0,00	1,11	0,00	0,21	0,17	0,10	0,37	0,00	0,00	0,00	0,14
Sorghum	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,00	4,01	5,03	0,00	0,00	1,75	0,02	0,04	0,00	0,00	0,00	0,00
Soy	30,09	0,00	0,00	0,00	0,00	0,00	0,00	0,00	14,50	0,00	0,00	0,00	2387,22	0,00	3,42	10,28	0,81	0,52	0,69	0,00	0,00	1,09
Soy / Bean	0,04	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,33	2,64	0,00	0,65	0,21	0,00	0,00	0,25	0,00	0,03
Soy / Braquiaria	0,15	0,00	0,01	0,00	0,00	0,00	0,00	0,00	2,911	0,11	0,01	0,00	0,67	0,00	121,08	12,28	3,74	1,52	0,03	0,00	0,00	0,05
Soy / Maize	3,39	0,00	0,00	0,00	0,00	0,05	0,00	0,00	9,10	0,34	0,00	0,09	9,36	0,00	8,68	3458,51	68,12	1,73	0,11	0,02	0,00	1,97
Soy / Cotton	0,74	0,00	0,00	0,00	0,00	0,00	0,00	0,00	15,61	0,25	0,00	0,00	0,85	0,01	1,68	126,84	2879,94	3,86	0,00	0,42	0,00	1,81
Soy / Millet	0,38	0,01	0,00	0,00	0,00	0,00	0,00	0,00	4,03	0,74	0,01	0,00	1,71	0,00	2,54	14,97	4,00	68,56	0,00	0,00	0,00	0,69
Soy / Potato	0	0,01	0,04	0,05	0,00	0,00	0,29	0,00	0,00	0,00	0,00	0,00	0,81	0,00	0,31	0,01	0,00	0,00	267,75	0,00	0,00	0,00
Soy / Sorghum	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,25	0,00	0,00	0,00	0,09	0,27	0,00	24,60	3,11	0,00	0,00	19,06	0,00	0,15
Soy / Soy	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,02	0,00	0,00	0,00	15,05	0,00
Soy / Weed	1,23	0,00	0,00	0,03	0,00	0,00	0,01	0,00	1,25	0,14	0,01	0,00	6,1	0,00	0,29	11,22	0,83	2,30	0,08	0,00	0,00	143,90

