MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2016/06.01.14.07-TDI

LAND USE AND LAND COVER MONITORING USING REMOTE SENSING IMAGE TIME SERIES

Victor Wegner Maus

PhD Thesis submitted to the Post Graduation in Earth System Science, supervised by Drs. Gilberto Câmara and Fernando Manoel Ramos, approved in April 29, 2016.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34P/3LQA8K8>

INPE
São José dos Campos
2016

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
Fax: (012) 3208-6919
E-mail: pubtc@inpe.br

COMMISSION OF BOARD OF PUBLISHING AND PRESERVATION OF INPE INTELLECTUAL PRODUCTION (DE/DIR-544):
Chairperson:
Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)
Members:
Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)
Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas (CEA)
Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)
Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial (ETE)
Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)
Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CPT)
Silvia Castro Marcelino - Serviço de Informação e Documentação (SID) **DIGITAL LIBRARY:**
Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Serviço de Informação e Documentação (SID)
DOCUMENT REVIEW:
Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SID)
Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)
ELECTRONIC EDITING:
Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)
André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2016/06.01.14.07-TDI

LAND USE AND LAND COVER MONITORING USING REMOTE SENSING IMAGE TIME SERIES

Victor Wegner Maus

PhD Thesis submitted to the Post Graduation in Earth System Science, supervised by Drs. Gilberto Câmara and Fernando Manoel Ramos, approved in April 29, 2016.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34P/3LQA8K8>

INPE
São José dos Campos
2016

Cataloging in Publication Data

Maus, Victor Wegner.

M448l Land use and land cover monitoring using remote sensing image time series / Victor Wegner Maus. – São José dos Campos : INPE, 2016.

xx + 94 p. ; (sid.inpe.br/mtc-m21b/2016/06.01.14.07-TDI)

Thesis (Doctorate in Earth System Science) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

Guiding : Drs. Gilberto Câmara and Fernando Manoel Ramos.

1. Time series analysis. 2. Dynamic programming. 3. Data mining. 4. Agricultural intensification. 5. Forest degradation. I.Title.

CDU 332.3:528.8

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.

Aluno (a): ***Victor Wegner Maus***

Título: "LAND USE AND ALND COVER MONITORING USING REMOTE SENSING IMAGE TIME SERIES ".

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em

Ciência do Sistema Terrestre

Dra. Silvana Amaral Kampel

Presidente / INPE / SJCampos - SP

Dr. Gilberto Câmara

Orientador(a) / INPE / SJCampos - SP

Dr. Fernando Manoel Ramos

Orientador(a) / INPE / São José dos Campos - SP

Dr. Dalton de Morisson Valeriano

Membro da Banca / INPE / SJCampos - SP

Dra. Maria Isabel Sobral Escada

Membro da Banca / INPE / SJCampos - SP

Dr. Daniel de Castro Victoria

Convidado(a) / EMBRAPA / Campinas - SP

Dr. Márcio Pupin de Mello

Convidado(a) / Boeing Res & Techo / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) maioria simples

(x) unanimidade

São José dos Campos, 29 de Abril de 2016

À memória de Ilo Wegner e Almida Maus

ACKNOWLEDGEMENTS

I thank the institutions who provided infrastructure and financial support during the development of my thesis, especially, the Earth System Science Center at the National Institute for Space Research (Brazil), the Institute for Geoinformatics at the University of Münster (Germany), the International Institute for Applied Systems Analysis (Austria), and the Federal University of Pampa, Campus Itaqui (Brazil).

I would like to express my gratitude to my supervisors Prof. Gilberto Câmara and Prof. Fernando Ramos for the support and time they have given me during these years. Their interest in international collaboration also gave me the possibility for research training abroad. Many thanks for the valuable advice and enthusiasm.

Many thanks to Prof. Edzer Pebesma for his hospitality during my stay at the Institute of Geoinformatics at the University of Münster. His valuable advice was crucial for my thesis.

I wish to express my gratitude to Dr. Ian McCallum and Dr. Aline Mosnier for their advice during the Young Scientists Summer Program (YSSP) hosted at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria. I also thank the YSSP organizing team and all the YSSP colleagues for making the stay at IIASA a great experience.

Many thanks to my friends Alber Sanchez, Marius Apple, Ricardo Cartaxo, and Gilberto Ribeiro, whose directly contribution was crucial for the development of my thesis.

Thanks to all my friends that I have had a pleasure to meet, especially, Vagner Camilotti, Raquel Klein, Aline Soterroni, Diulha Furlan, Rodrogo Wrasse, Deise Michelotti, Paulo Pinho, Renata Pinho, Maria Fernanda Vianna, Alexandre Lopes, Elson Toledo, Luis Paulo Barra, Juliana Gil, David Eitelberg, Merret Buurman, Nanki Sidhu, Meng Lu, Virnei Moreira, Jim Jones, Junior Furlan, and many others. I am grateful for your insights, stimulating discussions, and support in my decisions.

Many thanks to Prof. Afranio Righes for encouraging and helping me since my first steps in science. His enthusiasm for science has always inspired me.

I would like to thank Tiina for all the great shared moments, and for inspiring and helping me on the way.

I really appreciate my family's support, many thanks to my parents, Sérgio and Janice, my sister Débora, and brother Augusto. They have always encouraged me to study and supported me in my decisions.

ABSTRACT

Land system change has a wide range of impacts on Earth system components. Tropical forests in particular have been identified as crucial ecosystems for climate regulation, global biodiversity, and hydrological cycling. The Brazilian Amazon has experienced a high rate of deforestation in the last decade and it is the main source of Brazil's anthropogenic CO_2 emissions. The growing global population will further increase the demand for food and therefore increase the pressure on agricultural systems. High quality, fine resolution, and near-real time land use and land cover monitoring systems play a crucial role in generating information to advance our understanding of human impact on land cover. Earth Observation satellites are the only source that provides a continuous and consistent set of information about the Earth's land. The current large-scale classification systems such as MODIS Land Cover and GLC 2000 have limitations and their accuracy is not sufficient for land change modeling. Therefore, new techniques for improving land system products are urgently needed. The contribution of this thesis to Earth System Science is three-fold. Firstly, the thesis presents a new method for analysis of remote-sensed image time series that improves spatio-temporal land cover data sets and has a substantial potential for contributing to land system change modeling. The developed Time-Weighted Dynamic Time Warping (TWDTW) method is a time-constraint variation of the well-known Dynamic Time Warping (DTW) method, which has in the extensive literature proved to be a robust time series data mining. Secondly, this thesis contributed to open and reproducible science by making the algorithms available for larger audience. TWDTW is implemented in an open source R package called dtwSat available in the Comprehensive R Archive Network (CRAN). Thirdly, this thesis presents an analysis of land cover changes in the Amazon, focusing on the Brazilian state of Mato Grosso that has gone through high rate of deforestation and cropland expansion in the last decade. This study identified and estimated the land cover change using MODIS image time series, contributing to better understand the land dynamics in the Brazilian Amazon. In the study area the pasture is the dominant land use after deforestation, whereas most of the single cropping area comes from pasture, and the cropping system is undergoing intensification from single to double cropping. Moreover, the regenerative secondary forest comes mainly from pasture. The study showed the potential of the TWDTW method for large-scale remote sensing data analysis, which could be extended to other Brazilian biomes to help understand land change in the whole Brazilian territory.

Keywords: Time series analysis. Dynamic programming. Data mining. Crop identification. Agricultural intensification. Deforestation. Forest degradation.

MONITORAMENTO DE MUDANÇAS DE USO E COBERTURA DA TERRA POR ANÁLISE DE SÉRIES TEMPORAIS DE IMAGENS DE SENSORIAMENTO REMOTO

RESUMO

Mudanças na superfície da terra têm uma ampla gama de impactos sobre o sistema terrestre. Florestas tropicais, em particular, são ecossistemas cruciais para regulação climática, manutenção da biodiversidade, a ciclo hidrológico. Na última década a Amazônia brasileira tem experimentado uma alta taxa de desmatamento, sendo a principal fonte de emissões antropogênicas de CO_2 no Brasil. O crescimento da população mundial vai aumentar ainda mais a demanda por alimentos e, portanto, aumentar a pressão sobre agrícultura e pecuária. Dados com alta qualidade, melhor resolução espacial e temporal, e o desenvolvimento de sistemas de monitoramento desempenham um papel crucial na geração de informações para avançar nossa compreensão sobre os impactos humanos na cobertura da terra. Os satélites de observação da Terra são a única fonte que fornece um conjunto contínuo e consistente de informações sobre nosso planeta. Sistemas de classificação em grande escala, como MODIS Land Cover e GLC 2000 têm limitações e sua acurácia não é suficiente para a modelagem de mudanças de use da terra. Portanto, são necessárias novas técnicas para melhoramento dos dados de use e cobertura da terra. Esta tese traz três contribuições para a Ciência do Sistema Terrestre. Primeiramente, esta tese apresenta um novo método para análise de séries temporais de imagens satélite que melhora a classificação de cobertura da terra. O método tem grande potencial para contribuir para a modelagem de mudanças do sistema terrestre. O método desenvolvido, *Time-Weigted Dynamic Time Warping* (TWDTW), é uma adaptação ponderada por tempo do método clássico *Dynamic Time Warping* (DTW), que tem em uma extensa literatura provando ser um método robusto para mineração de dados em séries temporais. Em segundo lugar, esta tese contribuiu para a ciência aberta e reprodutível, tornando algoritmos disponíveis para o público. TWDTW estÃ¡ implementado em um pacote R de código aberto chamado **dtwSat** disponível no *Comprehensive R Archive Network* (CRAN). Em terceiro lugar, esta tese apresenta uma análise as mudanças do uso e cobertura da terra na Amazônia, com foco no estado do Mato Grosso, que passou por alta taxa de desmatamento e expansão agrícola na última década. Este estudo identificou e estimou mudanças de cobertura da terra com séries temporais de imagens MODIS, contribuindo para melhor compreender a dinâmica de ocupação da terra na Amazônia brasileira. Na área de estudo, a pastagem é o uso dominante após o desmatamento, ao passo que a maior parte da área de cultivo com um ciclo anual provem da área de pasto, com o sistema de cultivo passando por intensificação, mudando de cultivo simples para cultivo duplo. Além disso, áreas de regeneração vêm, principalmente, de áreas de pastagem. O estudo mostrou o potencial do método de TWDTW para análise de dados de sensoriamento remoto em grande escala, que poderia ser estendido a outros biomas brasileiros para ajudar a entender as mudanças da terra em todo o território brasileiro.

Keywords: Análise de séries temporais. Programação dinâmica. Mineração de dados. Identificação de culturas. Intensificação da agricultura. Desmatamento. Degradação florestal.

LIST OF FIGURES

Page

1.1 The main research questions of the thesis. 2

2.1 (a) A 3-dimensional array of satellite images, (b) a vegetation index time series I at the pixel location (x, y). The arrows indicate data gaps. 7

2.2 (a) DTW Alignment between two time series with approximately same length, (b) DTW alignments between a pattern whose length is much shorter than the time series. The indexes a and b are starting points and ending points of each interval in the long-term time series, respectively. . 8

2.3 Accumulated cost matrix $\mathbf{D}$ showing three possible alignment of the pattern $\mathbf{U}$ within the long-term time series $\mathbf{V}$. The indexes a are starting points and b ending points of each DTW alignment in $\mathbf{V}$. 10

2.4 Open boundary DTW alignment. Dark and light shades represent the alignments of the patterns U_1 and U_2, respectively. The indexes a_k and b_k represent the starting and ending points of the kth alignment in $\mathbf{V}$ associated with a DTW distance measure δ_k. 12

2.5 Temporal patterns of EVI MODIS 16 days. These patterns are the average based on several ground truth samples of each class. Source: Adapted from Arvor et al. (2011). 14

2.6 Linear and logistic time-weight. The logistic weight has midpoint $\beta = 100\,days$ and steepness $\alpha = 0.1$. 15

2.7 Best matches of forest, pasture, single cropping, and double cropping to an sample time series using DTW without time restriction in (a), and and the time-weighted DTW in (b). 15

2.8 Forest area estimated by the Amazon Monitoring Program PRODES (INPE, 2015) and using the logistic TWDTW based classification for Porto dos Gaúchos. 20

2.9 Total area of double cropping and single cropping in Porto dos Gaúchos estimated by TWDTW and the Brazilian national cropland survey (IBGE, 2014). 21

2.10 Total area of pasture, single cropping, and double cropping from 2001 to 2013 estimated using logistic TWDTW for Porto dos Gaúchos. 22

2.11 Land use/cover maps produced by using the logistic TWDTW classification. Each map shows the classification for an agricultural year (from July to June) in Porto dos Gaúchos. 23

2.12 An example of a classification using the transition rules. This is a sample time series inside of a burned area. This area was degraded in 2011 according to the Detection of Forest Degradation Program (DEGRAD) (INPE, 2014). 25

3.1 Matches of the known temporal pattern to subintervals of the long-term time series 30
3.2 A 3-dimensional array of satellite images (left), an enhanced vegetation index (EVI) time series at the pixel location (x, y) (right). The arrows indicate gaps in the time series. Adapted from Maus et al. (2016). 31
3.3 Example of time series based on MODIS product MOD13Q1 (FRIEDL et al., 2010). The labels of the phenological cycle are shown in the plot. . . . 34
3.4 Temporal patterns of soybean, cotton, and maize based on MODIS product MOD13Q1 (FRIEDL et al., 2010) 34
3.5 Logistic time-weight function `logisticWeight` with steepness `alpha=-0.1` and midpoint `beta=100`. The x axis shows the absolute difference between two dates in days and the y axis shows the time-weight (MAUS et al., 2016). 36
3.6 The four best matches of the "soybean" pattern in the time series using a logistic time-weight 37
3.7 Alignments and dissimilarity measures of the patterns "soybean", "cotton", and "maize" to the subintervals of the long-term time series using a logistic time-weight 38
3.8 Classification of each 6 months periods of the time series using results of the TWDTW analysis with logistic time-weight 39
3.9 Study area in Mato Grosso, Brazil, shown in a © Google Earth image. The area was originally covered by tropical forest that has been removed for agricultural use. 40
3.10 Temporal patterns of forest, cotton-fallow, soybean-cotton, soybean-maize, and soybean-millet based on the ground truth samples 44
3.11 Illustration of the TWDTW dissimilarity from each temporal pattern in 2013 45
3.12 Land use maps for each year from 2008 to 2013 46
3.13 Percentage of area for each land use class from 2008 to 2013 46
3.14 Gains and losses in area from the other classes 47
3.15 TWDTW dissimilarity measure for each pixel over each classified period 48
3.16 Accuracy assessment of the of the TWDTW method for land cover classification 49

4.1 Matches of the temporal patterns of different land cover classes to a long-term EVI time series using the Time-Weighted Dynamic Time Warping (TWDTW) algorithm.
Source: Adapted from Maus et al. (2016). 53
4.2 Land cover transitions rules to classify the time intervals. These rules are used to split: primary forest, degraded forest, and secondary forest regrowth. 55
4.3 Study area covering the Amazon biome at Mato Grosso, Brazil. 56
4.4 A 3-dimensional (3-D) array of satellite images. Each pixel has a set of attributes (e.g. "EVI", "RED", etc) associated with a pixel location over time.
Source: Adapted from Maus et al. (2016). 57
4.5 User's and Producer's Accuracy of the classification based on the TWDTW analysis. The graphic shows the average at confidence interval of 99% computed with a bootstrap simulation of 1000 resampling-with-replacement. 58
4.6 Set of temporal patterns associated with phenological cycle of forest, pasture/grassland, single cropping systems (cotton and soybean), and double cropping system (soybean-maize and soybean-cotton). 60
4.7 Land cover classification in 2001 and 2014 in the Amazon biome at Mato Grosso, Brazil. The complete time series of land cover maps is available in appendix A. 61
4.8 Percentage of area from 2001 to 2014 in the Amazon biome at Mato Grosso, Brazil. 62
4.9 Total area of cash crop agriculture over time in the Amazon biome at Mato Grosso, Brazil. 62
4.10 Total area of pasture, degraded forest, secondary forest regrowth, and deforestation/forest degradation over time in the Amazon biome at Mato Grosso, Brazil. 63
4.11 Example of forest degradation in Porto dos Gaúchos, Brazil, at the latitude -11°54' and longitude -56°40'. 64
4.12 Gains and losses in area from the other classes in the study area. The y axis shows the actual class; the positive direction of x axis shows the gains and the negative direction of x axis shows the losses of the classes indicated in y. The colors indicate from/to which classes the gains/losses belong. 65

4.13 Matrix of changes in the Amazon Biome within Mato Grosso, Brazil. The values are the percentage of area for each class related to the total changes accumulated from 2001 to 2014. 66
4.14 Land cover transitions from 2003 to 2014 in the Amazon biome at Mato Grosso, Brazil. The bars show the area converted from forest to soybean (blue), from pasture to soybean that has been deforested before 2006 (green), from pasture to soybean that has been deforested after 2006 (purple), and from forest to pasture (red). 67

A.1 Legend of the land cover maps. 87
A.2 Land cover classification in 2001 and 2002 in the Amazon biome at Mato Grosso, Brazil. 88
A.3 Land cover classification in 2003 and 2004 in the Amazon biome at Mato Grosso, Brazil. 89
A.4 Land cover classification in 2005 and 2006 in the Amazon biome at Mato Grosso, Brazil. 90
A.5 Land cover classification in 2007 and 2008 in the Amazon biome at Mato Grosso, Brazil. 91
A.6 Land cover classification in 2009 and 2010 in the Amazon biome at Mato Grosso, Brazil. 92
A.7 Land cover classification in 2011 and 2012 in the Amazon biome at Mato Grosso, Brazil. 93
A.8 Land cover classification in 2013 and 2014 in the Amazon biome at Mato Grosso, Brazil. 94

LIST OF TABLES

Page

2.1 Accuracy assessment for each class based on 489 reference samples classified from the Landsat images. 17

2.2 Confusion matrices based on 489 reference samples classified from the Landsat images. 19

2.3 Equivalent classes for comparison between the TWDTW classification and MODIS land cover collection 5, Plant Functional Type (PFT). . . . 19

2.4 Assessment of MODIS collection 5 Plant Functional Type (PFT) and logistic TWDTW based on 489 reference samples classified from the Landsat images. The classes forest, pastureland, and cropland were aggregated according to Table 2.3. 20

3.1 User's and Producer's Accuracy of the land use classification based on TWDTW analysis. μ is the average accuracy, σ the standard deviation, and CI is the confidence interval of 99% using 100 resampling-with-replacement. 50

4.1 Ground truth samples used to create the temporal patterns and to perform the accuracy assessment of the land cover classification. These samples were collected by the Brazilian Agricultural Research Corporation (EMBRAPA) (ARVOR et al., 2011). 57

4.2 User's and Producer's Accuracy of the land use classification based on TWDTW analysis. μ is the averaged accuracy and σ the standard deviation. 59

CONTENTS

Page

1 INTRODUCTION 1

2 A TIME-WEIGHTED DYNAMIC TIME WARPING METHOD FOR LAND-USE AND LAND-COVER MAPPING 5

2.1 Introduction 5
2.2 Methods 7
2.2.1 Step 1: DTW Alignment 9
2.2.2 Step 2: Map building 11
2.3 Experiments 11
2.4 Results 13
2.5 Discussion 22
2.6 Conclusion 25

3 dtwSat: TIME-WEIGHTED DYNAMIC TIME WARPING FOR SATELLITE IMAGE TIME SERIES ANALYSIS IN R . . 27

3.1 Introduction 27
3.2 The Time-Weighted Dynamic Time Warping method 29
3.3 dtwSat package overview 31
3.4 Classifying a time series 32
3.4.1 Input data 32
3.4.2 Detection of time series patterns with TWDTW 35
3.4.3 Visualising the result of the TWDTW algorithm 36
3.4.4 Classifying the long-term time series 37
3.5 Producing a land cover map 38
3.5.1 Input data 39
3.5.2 Creating the time series and the temporal patterns 42
3.5.3 Classifying the image time series 43
3.5.4 Looking at the classification results 45
3.5.5 Assessing the classification accuracy 47
3.6 Conclusions and Discussion 50

4 LAND-USE AND LAND-COVER TRAJECTORIES IN MATO GROSSO BRAZIL 51

4.1 Introduction 51
4.2 Classifying satellite image time series 52
4.2.1 Creating the temporal patterns 52
4.2.2 Applying the TWDTW analysis 53
4.2.3 Creating annual land cover maps 53
4.3 Datasets and validation 56
4.4 Results 57
4.4.1 Accuracy assessment 57
4.4.2 Land cover change 59
4.5 Discussion and Conclusions 63

5 SYNTHESIS 69
5.1 How to deal with irregularly sampled satellite image time series for land-use and land-cover classification? 69
5.2 Contribution to make methods for satellite time series analysis open and reproducible 70
5.3 What can we learn from improved land-use and land-cover data sets? . . 71
5.4 Future work 71
5.5 Conclusion 72

REFERENCES 75

APPENDIX A - LAND COVER MAPS 87

1 INTRODUCTION

Land-use and land-cover change (LUCC) has a wide range of impacts on the Earth system components. For example, vegetation cover transformation affects local to global climate through the albedo change and the emission of greenhouse gases (GHGs) (PIELKE et al., 2002; FOLEY et al., 2005; STOCKER et al., 2013). The agricultural expansion and intensification accompanying by irrigation and fertilization impact water availability and quality (SCANLON et al., 2007). According to Piao et al. (2006) LUCC account for about 50% of the river runoff trend of the last century. Land use and land cover changes also have large global impact on biodiversity reduction (SALA et al., 2000; CHAPIN et al., 2000; CHAPLIN-KRAMER et al., 2015). The LUCC is also causing risks to human well being trough the degradation of ecosystems (MILLENNIUM ECOSYSTEM ASSESSMENT, 2005). In addition, Steffen et al. (2015) argue that land system change, especially deforestation, can contribute to increasing the risk of destabilising the Earth system functioning.

In this context, LUCC is a central issue in the sustainability debate (LAMBIN; MEYFROIDT, 2011). Global population is expected to exceed 8 billion by mid-century (UN, 2015), rising crop and livestock demand and production by around 40% (ALEXANDRATOS et al., 2012). This will increase pressures on land, water, and energy systems (BEDDINGTON, 2009). At the same time we need to mitigate and adapt to climate change (IPCC, 2014). Therefore, more and more, citizens and politicians are pressing scientists to provide qualified information that would allow wise decisions about the future of our planet.

Reliable land-use and land-cover datasets are essential to understand the impact and extent of global land changes (FRITZ et al., 2013; SEE et al., 2015). Earth Observation satellites are the only source that provides a continuous and consistent set of information about the Earth's land. In many areas, remote sensing images are the only data available for this purpose. Recent decisions by major space agencies have made unprecedented amounts of imagery available for research and operations. This brings a unique opportunity to measure the global and local changes in our environment and assess the human impacts on Earth. The question is: how can we make best use of the available satellite data products to improve our knowledge about LUCC?

Open archives of long-term satellite image time series provide opportunities to better quantify global change (LAMBIN; GEIST, 2006). This has lead to the development of automated and semi-automated methods, including multi-image compositing

(GRIFFITHS et al., 2013), detecting forest disturbance (KENNEDY et al., 2010; ZHU et al., 2012; DEVRIES et al., 2015), crop classification (XIAO et al., 2005; WARDLOW et al., 2007; PETITJEAN et al., 2012), planted forest mapping (MAIRE et al., 2014), crop expansion and intensification (GALFORD et al., 2008; SAKAMOTO et al., 2009), detecting trend and seasonal changes (LUNETTA et al., 2006; VERBESSELT et al., 2010b; VERBESSELT et al., 2010a; VERBESSELT et al., 2012), extracting seasonality metrics (JöNSSON; EKLUNDH, 2002; JöNSSON; EKLUNDH, 2004), and synthesizing multi-spectral information from satellite image time series (MELLO et al., 2013).

Satellite image time series will always be contaminated by some degree of residual atmospheric influence, geolocation error, and directional effects (LAMBIN; GEIST, 2006). To make the best use of available satellite data archives, methods for satellite image time series analysis need to deal with data sets that are noisy, irregularly sampled, and in many cases out-of-phase.

In the thesis I address three issues to improve the land-use and land-cover products and better understand the land use dynamics. The first issue is how to deal with irregularly sampled satellite image time series for land-use and land-cover classification. The second issue is to make methods for satellite time series analysis open and reproducible. The third issue is what can we learn from improved spatiotemporal land-use and land-cover data sets. Figure 1.1 presents the main research issues addressed in each chapter of the thesis.

Figure 1.1 - The main research questions of the thesis.

Chapter 2 describes a new method for satellite image time series analysis. This method is a time-weighted variation of the classical Dynamic Time Warping (DTW) method (VELICHKO; ZAGORUYKO, 1970; SAKOE; CHIBA, 1971; SAKOE; CHIBA, 1978; RABINER; JUANG, 1993; BERNDT; CLIFFORD, 1994). Research on time series data mining shows that methods based on DTW have achieved significant results in many applications, such as speech recognition (BERNDT; CLIFFORD, 1994; ESLING; AGON, 2012; RAKTHANMANON et al., 2012). DTW works by comparing a temporal signature of a known event (e.g. a person's speech) to an unknown time series (e.g. a speech record of unknown origin). The algorithm compares two time series and finds their optimal alignment, providing a dissimilarity measure as a result (RABINER; JUANG, 1993). DTW provides a robust distance measure for comparing time series, even if they are irregularly sampled (PETITJEAN et al., 2012) or are out of phase in the time axis (KEOGH; RATANAMAHATANA, 2005). The large range of applications of dynamic time warping for time series analysis motivated the idea of using DTW for remote sensing applications.

Chapter 3 describes the contribution of the **R** (R Core Team, 2015) package **dtwSat** (MAUS, 2015), making methods for satellite time series analysis available to a larger audience. Given the open availability of large image data sets, the research community on Earth Observation would get much benefit from methods that are openly available, reproducible and comparable. However, few of the proposed methods for remote sensing time series analysis are available as open source software, the main exception being the BFAST and BFAST-monitor algorithms for change detection (VERBESSELT et al., 2010b; VERBESSELT et al., 2010a). The **dtwSat** package provides an implementation of Time-Weighted Dynamic Time Warping (TWDTW) for satellite image time series analysis, and is available from the Comprehensive R Archive Network at http://CRAN.R-project.org/package=dtwSat.

Chapter 4 presents a LUCC analysis using MODIS image time series. The area is located in the Brazilian state of Mato Grosso, which had strong deforestation and cropland expansion in the last decade. Understanding land-use and land-cover dynamics in this area is particularly important because agriculture, forestry, and other land use related sectors have been the main source of Brazil's GHGs emission (MCTI, 2014). In addition, increasing human dominance of the forest alters their health and the provision of important ecosystem functions and services (LEWIS et al., 2015). The study case focuses on the deforestation, forest degradation, and agricultural expansion from 2000 to 20014.

2 A TIME-WEIGHTED DYNAMIC TIME WARPING METHOD FOR LAND-USE AND LAND-COVER MAPPING[1]

2.1 Introduction

There is a global increase in food and energy production from agriculture to support 7.3 billion people. To support sustainable practices and find out about unsustainable uses of natural resources, good quality land-use and land-cover datasets are essential (FRITZ et al., 2013). Earth Observation satellites are the only source that provides a continuous and consistent set of information about the Earth's land and oceans. Since remote sensing satellites revisit the same place repeatedly, we can calibrate their images so measures of the same place in different times are comparable. These observation can be organised, so that each measure from sensor is mapped into a three dimensional array in space-time.

From a data analysis perspective, researchers then have access to space-time data sets. This has lead to much recent research on satellite image time series analysis. Algorithms for analysing image time series include methods for time series reconstruction (ROERINK et al., 2000), detecting trend and seasonal changes (LUNETTA et al., 2006; VERBESSELT et al., 2010b; VERBESSELT et al., 2012), extracting seasonality information (JöNSSON; EKLUNDH, 2002), land cover mapping (GRIFFITHS et al., 2013), detecting forest disturbance and recovery (KENNEDY et al., 2010; ZHU et al., 2012; DEVRIES et al., 2015), crop classification (XIAO et al., 2005; WARDLOW et al., 2007; PETITJEAN et al., 2012), planted forest mapping (MAIRE et al., 2014), and crop expansion and intensification (GALFORD et al., 2008; SAKAMOTO et al., 2009).

Research on time series data mining shows that methods based on dynamic time warping (DTW) have achieved significant results in many applications (BERNDT; CLIFFORD, 1994; ESLING; AGON, 2012; RAKTHANMANON et al., 2012). DTW works by comparing a temporal signature of a known event (e.g. a person's speech) to an unknown time series (e.g. a speech record of unknown origin) (VELICHKO; ZAGORUYKO, 1970; SAKOE; CHIBA, 1971; SAKOE; CHIBA, 1978; RABINER; JUANG, 1993; BERNDT; CLIFFORD, 1994). The algorithm compares two time series and finds their optimal alignment, providing a dissimilarity measure as a result (RABINER; JUANG, 1993). DTW provides a robust distance measure for comparing time series,

[1]This chapter is based on the paper: MAUS, V.; CAMARA, G.; CARTAXO, R.; SANCHEZ, A.; RAMOS, F. M.; QUEIROZ, G. R. de. A time-weighted dynamic time warping method for land-use and land-cover mapping. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, PP, n. 99, p. 1-11, 2016.

even if they are irregularly sampled (PETITJEAN et al., 2012) or are out of phase in the time axis (KEOGH; RATANAMAHATANA, 2005). The large range of applications of digital time warping for time series analysis motivated our idea of using DTW for remote sensing applications.

The DTW method works well for shape matching, but is not suited *per se* for remote sensing time series classification. It disregards the temporal range when finding the best alignment between two time series (RABINER; JUANG, 1993; JEONG et al., 2011). The vegetation associated with each land cover class has a distinct phenological cycle that is relevant for space-time classification (REED et al., 1994; ZHANG et al., 2003). Therefore, a good time-series land cover classifier needs to balance between shape matching and temporal alignment. For example, although crops tend to vary their annual phenological cycles, these variations will not be extreme. Consider a set of samples of soybean whose cycles range from 90 to 120 days. A time series with similar shape but with much larger cycle is unlikely to come from a soybean crop. The standard DTW method warps time to match the two series. To avoid such mismatches, we introduce a time constraint that helps to distinguish between different types of land-use and land-cover classes.

Recent papers by Petitjean et al. (2012) and Petitjean and Weber (2014) have used DTW for satellite image time series classification. The method proposed in these papers sets a maximum time delay to avoid inconsistent temporal distortions based on the date of the satellite images. The time series is split in one year segments to match the agricultural phenological cycle in Europe. However, this temporal segmentation reduces the power of the DTW classifier. Crops with phenological cycles longer than one year or taking place in different seasons may not be detected. The time-weighted extension to the DTW algorithm avoids this problem. Temporal segments of a remote sensing time series are classified without splitting them into fixed parts. This method is flexible to account for multiyear crops, single cropping and double cropping. It is also robust to account for other land cover types such as forest and pasture and works with a small amount of training samples.

Our main contribution is to show that a data mining method such as DTW, when used for land-use and land-cover classification of remote sensing time series, benefits from a temporal constraint. This conjecture has been validated in a case study in the Brazilian Amazon, where we compared the result of our proposed method with other time warping classifiers.

2.2 Methods

Since remote sensing satellites regularly cycle the Earth, their data are mappable to three-dimensional arrays in space-time Figure 2.1a. Each pixel location (x, y) in consecutive times, $t_1, ..., t_m$, makes up a satellite image time series, such as the one in Figure 2.1b. From these time series, we can extract land-use and land-cover information. In the example, during the first two years the area was covered by forest. It was deforested in 2002. The area was then used for cattle raising (pasture) for three years. from 2006 to 2008, it was used for crop production.

(a) (b)

Figure 2.1 - (a) A 3-dimensional array of satellite images, (b) a vegetation index time series I at the pixel location (x, y). The arrows indicate data gaps.

Let $\mathbf{V}_{x,y} = (v_1, v_2, ..., v_m)$ be a time series of a pixel location (x, y) in consecutive times, $t_1, ..., t_m$, where v_i is the value of the sensor measure at time t_i. Combining all the satellite's spatial coverage, we get a set of time series $\mathcal{S} = \{\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_s\}$. We assume there is a temporal continuity for each land cover classes, resulting from human actions. A forest area does not change to grassland or to soybeans overnight. Land cover changes take time. Our hypothesis is that it is possible to associate closed intervals of each time series $\mathbf{V}_{x,y}$ to a specific land-use and land-cover type. For example, suppose a ten year period where in the first five years the area was covered by forest. The area was then used for cattle raising (pasture) for two years. After that, it was used for soybean production for three years. We want to associate each of these intervals with one of our land cover classes.

Optical remotely sensed data are affected by cloud cover that introduces a large amount of noise in satellite image time series, as shown in Figure 2.1b. Inter-annual climate variability also changes the phenological cycles of the vegetation, resulting in

time series whose periods and intensities do not match on an year to year basis(REED et al., 1994). To associate intervals of a satellite image time series with land-use and land-cover classes, we need methods suitable for noisy and out-of-phase time series. We chose the Dynamic Time Warping (DTW) algorithm because it is suitable for this problem.

The papers by Petitjean et al. (2012) and Petitjean and Weber (2014) applied the DTW algorithm to classify intervals of satellite image time series, such as in Figure 2.2a. In this case, two time series have approximately the same length and the first and last points in both time series must match. In practice, crop phenological cycles can vary in an year-to-year basis, depending on climate conditions and land management. Examples include shifting the greening and dormancy stages of the vegetation (REED et al., 1994; ZHANG et al., 2003). To avoid possible inconsistent matching of phenological cycles caused by splitting the time series we use an open boundary version of DTW, Figure 2.2b. The open boundary method does not require two time series to be of the same length, and it is suitable to find all possible matches of one pattern within a long-term time series (MüLLER, 2007).

The open boundary DTW algorithm disregards the time dimension and can cause inconsistent phase alignments, e.g. a winter crop template can match the shape of a summer crop. To avoid these temporal inconsistencies, we introduce a temporal constraint. If there is a large seasonal difference between the sample pattern and its match in time series, an extra cost is added to the DTW distance measure. This constraint controls the time warping and makes the time series alignment dependent on the seasons. This is especially useful for detecting temporary crops and for distinguishing pasture from agriculture.

Figure 2.2 - (a) DTW Alignment between two time series with approximately same length, (b) DTW alignments between a pattern whose length is much shorter than the time series. The indexes a and b are starting points and ending points of each interval in the long-term time series, respectively.

Classification using open boundary DTW (MüLLER, 2007) requires matching subsequences of the time series associated with each pixel location to samples of the expected classes. For each class c, we take a set of time series samples $\mathcal{Q}_c = \{\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_q\}$, where $\mathbf{U} = (u_1, ..., u_n)$ is a time series with $n \ll m$ (i.e. the pattern length is much shorter than the sensor time series $\mathbf{V}$). q is the number of patterns for each class. These samples are then used to classify the intervals of the time series $\mathbf{V} \in \mathcal{S}$.

The classification is done for each pixel with two steps. First, the DTW algorithm is applied for each pattern in $\mathcal{Q}$ and each time series $\mathbf{V} \in \mathcal{S}$. This step provides information on how patterns match intervals of the time series. In the second step, the best DTW matches are used to build a sequence of land-use and land-cover maps.

2.2.1 Step 1: DTW Alignment

The DTW alignment starts by computing a n-by-m matrix Ψ, whose elements $\psi_{i,j}$ are the absolute difference between $u_i \in \mathbf{U}\ \forall\ i = 1, ..., n$ and $v_j \in \mathbf{V}\ \forall\ j = 1, ..., m$. From Ψ we compute an accumulated cost matrix $\mathbf{D}$ by a recursive sum of the minimal distances, such that

$$d_{i,j} = \psi_{i,j} + min\{d_{i-1,j}, d_{i-1,j-1}, d_{i,j-1}\}, \tag{2.1}$$

that is subject to the following boundary conditions:

$$d_{i,j} = \begin{cases} \psi_{i,j} & i = 1,\ j = 1 \\ \sum_{k=1}^{i} \psi_{k,j} & 1 < i \leq n,\ j = 1 \\ \sum_{k=1}^{j} \psi_{i,k} & i = 1,\ 1 < j \leq m \end{cases} \tag{2.2}$$

The Figure 2.3 shows an example of the accumulated cost matrix $\mathbf{D}$. Intuitively, the DTW alignment runs along the "valleys" of low cost in the accumulated cost matrix $\mathbf{D}$, that has as many "valleys" as the number of matches between $\mathbf{U}$ and $\mathbf{V}$. The kth low cost path in $\mathbf{D}$ produces an alignment between the pattern and a subsequence $\mathbf{V}|_{a_k}^{b_k}$ with associated DTW distance δ_k, where a_k is the starting point and b_k the ending point of the subsequence k (MüLLER, 2007), as shown in Figure 2.3.

Each minimum point in the last line of the accumulated cost matrix, i.e. $d_{n,j}\ \forall\ j =$

Figure 2.3 - Accumulated cost matrix **D** showing three possible alignment of the pattern **U** within the long-term time series **V**. The indexes a are starting points and b ending points of each DTW alignment in **V**.

$1, ..., m$, produces an alignment, with b_k and the δ_k given by,

$$b_k = argmin_k(d_{n,j}), \qquad k = 1, ..., K \tag{2.3}$$

$$\delta_k = d_{n,b_k} \tag{2.4}$$

where K is the number of minimum points in last line of the accumulated cost matrix.

A reverse algorithm, Equation 2.5, maps the warping path $\mathbf{P}_k = (p_1, ..., p_L)$ along the kth low cost "valley" in **D**. The algorithm starts in $p_{l=L} = (i = n, j = b_k)$ and ends when $i = 1$, i.e. $p_{l=1} = (i = 1, j = a_k)$, where L denotes the last point of the alignment. The warping path $\mathbf{P}_k$ contains the matching points between the time series. Note that the backward step in the Equation 2.5 implies the monotonicity condition (SAKOE; CHIBA, 1971; MüLLER, 2007), i.e. the alignment preserves the order of the time series.

$$p_{l-1} = \begin{cases} (i, a_k = j) & if \quad i = 1 \\ (i-1, j) & if \quad j = 1 \\ argmin(\quad d_{i-1,j}, & \\ \qquad\qquad d_{i-1,j-1}, & \text{otherwise} \\ \qquad\qquad d_{i,j-1}) & \end{cases} \tag{2.5}$$

The original DTW algorithm does not account for the phase difference between two time series (JEONG et al., 2011). However, land-use and land-cover types have distinct phenological cycles that are relevant for space-time classification (REED et al., 1994; ZHANG et al., 2003). We introduce a time-weighted extension of DTW (TWDTW),

based on the date of each pixel in the satellite image. This time-weighted version of DTW adds a temporal cost ω to the cost matrix Ψ, whose elements become $\psi_{i,j} = |u_i - v_j| + \omega_{i,j}$. To compute the temporal cost we propose both a linear

$$\omega_{i,j} = g(t_i, t_j) \tag{2.6}$$

and a logistic model with midpoint β, and steepness α, such that

$$\omega_{i,j} = \frac{1}{1 + e^{-\alpha(g(t_i,t_j)-\beta)}}, \tag{2.7}$$

where $g(t_i, t_j)$ is the elapsed time in days between the dates t_i in the pattern and t_j in the time series. We ran many tests using different values of β and α. We then used the best global accuracy performance to set the parameters for the logistic time-weighted DTW.

2.2.2 Step 2: Map building

The DTW algorithm matches each pattern to the input time series independently from the others. Thus, each interval of the time series $\mathbf{V}$ can fit different patterns. To associate an interval of the time series $\mathbf{V}$ to a land-use and land-cover class, we choose the best fitting pattern, i.e. the pattern with the lowest DTW distance in the interval. After finding the best fit, we can produce maps that show a land-use and land-cover classification for a given period.

To compare our results with other land use/cover products, we produced maps matching the agricultural calendar from July to June (gray area in Figure 2.4). We find the pattern that has the lowest DTW distance to a subsequence $\mathbf{V}|_{a_k}^{b_k}$ partly contained in the crop calendar. The Figure 2.4 shows the matching of two patterns, $\mathbf{U}_1$ and $\mathbf{U}_2$, that are partially in the same agricultural year from July 2000 to June 2001. In this case we pick the one with the lowest DTW distance, i.e. the most similar pattern for that period.

2.3 Experiments

In our experiments, we tested the performance of four different DTW methods: *i)* the original DTW algorithm without time constraints (i.e. $\omega = 0$), *ii)* DTW with maximum time delay as proposed by Petitjean et al. (2012), *iii)* linear TWDTW, and *iv)* logistic TWDTW.

We used time series of Enhanced Vegetation Index (EVI) from July 2000 to June

Figure 2.4 - Open boundary DTW alignment. Dark and light shades represent the alignments of the patterns U_1 and U_2, respectively. The indexes a_k and b_k represent the starting and ending points of the kth alignment in $\mathbf{V}$ associated with a DTW distance measure δ_k.

2013 based on Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD13Q1 16 days 250 m. MODIS EVI has improved sensitivity in high biomass regions through a canopy background adjustment and a reduction in the atmosphere influences (HUETE et al., 1997; HUETE et al., 2002).

The EVI time series is subject to atmospheric effects, such as cloud cover and path radiance from aerosols (FREITAS et al., 2011). To reduce the spurious oscillation due to atmospheric effects, we apply a discrete wavelet decomposition (MALLAT, 1998) and then filter the time series by removing the highest wavelet frequency. The wavelet filter preserves the essential temporal variation and is more sensitive to vegetation seasonal changes than filters based on Fourier transform (SAKAMOTO et al., 2005).

An important scientific problem is understanding changes in the Brazilian Amazonian rain forest, which has an area of 4,100,000 km^2. In Amazonia, 720,000 km^2 have been deforested since the 1970s (INPE, 2015). In the Copenhagen Climate Conference in 2009, Brazil pledged to reduce deforestation in Amazonia by 80% relative to the average of the period 1996-2005. Brazil is making good this pledge. Forest cuts in Amazonia fell from 27,700 km^2 in 2004 to 4,900 km^2 in 2012, decreasing by 83%. Given the impact of land changes in Amazonia on global biodiversity, emissions, and ecological services, it is important to understand what causes forest removal (AGUIAR et al., 2006). INPE (Brazil's National Institute for Space Research) and EMBRAPA (Brazil's Agricultural Research Agency) mapped the land use of the

deforested areas in Amazonia up to 2008 (INPE, 2012). Their results show that 63% of the forest cuts are now used for cattle raising. Cattle ranches in Amazonia use extensive practices, with less than 1 head of cattle per hectare. Cash crop agriculture accounts for only 4% of the deforestation. Moreover, more than 20% of the area has been abandoned and is now regrowing as secondary vegetation. To achieve further gains in reducing deforestation and biodiversity loss, we need to understand the different land use trajectories, including the deforestation dynamics, land use intensification, and land abandonment pathways.

We ran a case study in an area in Amazonia that had strong deforestation and cropland expansion in the last decade. We selected the Porto dos Gaúchos municipality, that covers approximately 7,000 km^2 and is located in the state of Mato Grosso, Brazil, inside of the Amazon Biome. In 2013 its total deforested area was 3023.6 km^2, that is 42.9% of the original forest cover (INPE, 2015). The cropland area grew from 59.8 km^2 in 2000 to 580.8 km^2 in 2013 (IBGE, 2014). We chose the most important classes for that area: forest, secondary vegetation, pasture, single cropping, and double cropping. These land cover classes are the most relevant ones for our study on trajectories of change in Amazonia.

Our classification method requires a set of temporal patterns of the chosen land use/cover classes. We defined the temporal patterns of forest, pasture, single cropping, and double cropping based on the paper by Arvor et al. (2011), that presented typical temporal patterns of EVI for different crops types and natural vegetation for the same region of our case study. Arvor et al. (2011) used several ground truth data collections identified through field studies to derive their averaged EVI signal according to the agricultural calendar from July to June. Here we joined some of the temporal patterns from Arvor et al. (2011), such that "soybean" and "cotton" are used as "single cropping", and "soybean-cotton" and "soybean-maize" are "double cropping". We kept the classes "forest" and "pasture". Therefore, each class has one or two patterns shown in Figure 2.5.

2.4 Results

To assess our algorithm, we used 40 random selected spatial locations from that we could classify 489 samples out of 560 in the period from 2001 to 2014. Most of the unclassified samples had cloud contamination during the growing cycles of single and double cropping because the raining season in Mato Grosso state is usually from November to March (AGUIAR et al., 2006). The samples were classified by visual interpretation of Landsat images using the Google Earth Engine (Google, 2014).

Figure 2.5 - Temporal patterns of EVI MODIS 16 days. These patterns are the average based on several ground truth samples of each class.
Source: Adapted from Arvor et al. (2011).

To separate our classes we used a set of images corresponding to the agricultural year from July to June. For each year we used at least 4 images showing different phenological stages of the vegetation that allow us to distinguish: forest, pasture, single cropping, and double cropping.

The logistic TWDTW had the best performance for $\alpha = 0.1$ and $\beta = 100\,days$ (global accuracy 87.32%), meaning a low penalty for time warps smaller then $60\,days$ and significant costs for bigger time warps, Figure 2.6. In the algorithm proposed by Petitjean et al. (2012) we tested maximum time delays ranging from 30 to $130\,days$, and found the best performance when the delay was set to $100\,days$ with global accuracy 84.66%. The linear TWDTW had global accuracy 81.6% and the DTW without time restrictions 70.14%.

Part of the good performance of TWDTW comes from good quality sample patterns. Given a good set of samples, TWDTW uses the length of each pattern as a temporal constraint in its distance measure. The standard version of DTW reduces or enlarges the pattern without temporal restrictions to find the best fit. Unrestricted warping works well for highly variable signals such as speech, but has problems dealing with structured patterns such as land cover signals. To compare DTW without time constraints and TWDTW, see Figure 2.7. In this figure, we show how the best matches for samples patterns of four classes (forest, pasture, single cropping, and double cropping) for the two versions of DTW (with and without time constraints). The DTW without time constraints, Figure 2.7a, overfits the patterns of forest,

Figure 2.6 - Linear and logistic time-weight. The logistic weight has midpoint $\beta = 100\,days$ and steepness $\alpha = 0.1$.

pasture and single cropping. The forest and pasture signals are strongly shortened and the single cropping signal is mapped to the first cycle of a double cropping event. By contrast, TWDTW keeps the temporal consistency for all land classes, as shown in Figure 2.7b.

Figure 2.7 - Best matches of forest, pasture, single cropping, and double cropping to an sample time series using DTW without time restriction in (a), and and the time-weighted DTW in (b).

The Table 2.1 shows the accuracy assessment of the four DTW approaches based on 489 reference samples classified from the Landsat images. In general, the logistic TWDTW had higher accuracy than the other approaches. Although the logistic TWDTW had lower *user accuracy* than the linear TWDTW for double cropping and forest, its *producer accuracy* was higher than the linear TWDTW for these classes (cf. Table 2.1). This means that the logistic TWDTW classified more ground truth pixels as such, but with a slightly lower confidence than the linear TWDTW for pixels classified as double cropping and forest. The logistic TWDTW had the same value of sensitivity for double cropping as the maximum delay DTW (i.e. *producer accuracy* 90.43%), but with larger confidence for this class, *user accuracy* 92.04% in comparison to 88.89%.

Table 2.1 - Accuracy assessment for each class based on 489 reference samples classified from the Landsat images.

Method	Double cropping (%)		Forest (%)		Pasture (%)		Single cropping (%)	
	User	Producer	User	Producer	User	Producer	User	Producer
DTW without time restrictions	74.65	46.09	88.51	72.64	79.53	80.47	50.00	77.78
DTW with maximum delay of 100 *days*	88.89	90.43	93.00	87.74	88.20	84.02	72.82	75.76
Linear TWDTW	96.70	76.52	96.81	85.85	83.54	78.11	60.27	88.89
Logistic TWDTW $\alpha = 0.1$ and $\beta = 100\,days$	92.04	90.43	94.00	88.68	88.41	85.80	75.00	84.85

The confusion matrices of the four DTW approaches are shown in Table 2.2. We see that DTW without time restriction had the worst results, particularly, for double cropping that had 57 pixels classified as single cropping. The linear TWDTW classified 24 pixels of double cropping and 34 pixels of pasture as single cropping, and therefore, its confidence for single cropping was 60.27% (cf. Table 2.1). The logistic TWDW classified 10 pixels of double cropping and 18 pixels of pasture as single cropping, which means a higher confidence than the linear TWDTW classification for single cropping, 75.00%. These results of the logistic TWDW were similar to the results obtained using the maximum time delay DTW, which classified 9 pixels of double cropping and 18 pixels of pasture as single cropping. However, the logistic TWDTW had higher sensitivity than the maximum time delay DTW (84.85% in comparison to 75.76% cf. Table 2.1), that classified 11 pixels as double cropping, 6 as pasture and unclassified other 7 pixels out of 99 pixels of single cropping.

We also compared the accuracy of our classification and the MODIS land cover collection 5, Plant Functional Type (PFT) 500 m (FRIEDL et al., 2010) using the validation points. Mapping from MODIS classes to our classes is shown in Table 2.3. Originally, the study area was covered by forest. Therefore, the other land cover types that appear later result from human activities. We aggregated the MODIS categories of trees to a class called forest. We also assume that MODIS shrubland and grassland classes are used as pastureland for cattle raising, and the categories of cereal crops and broad-leaf crops are aggregated to a class called cropland. Other MODIS classes are less than 0.008% of the pixels in this area, and thus they were not considered in this paper.

The accuracy assessment comparing logistic TWDTW results and MODIS land cover is shown in Table 2.4. The TWDTW algorithm had global accuracy of 91.21%, better than the global accuracy of MODIS (79.36%). TWDTW had higher user's and producer's accuracies than the MODIS classification for all classes. Although, MODIS had high user's accuracy for forest (87.2%) and cropland (89.33%), its producer accuracy for these classes was low, 77.37% and 75.28%, respectively.

We compared our forest area with estimations by the Amazon Monitoring Program PRODES (INPE, 2015). To be able to compare results with the pristine forest area that comes from PRODES, we need to split our "forest" class into "pristine forest" and "secondary vegetation". This requires a land cover transition rule. Areas matching a forest pattern were classified as forest only if they had also been classified as forest in previous years. Otherwise, we classified them as secondary vegetation. For

Table 2.2 - Confusion matrices based on 489 reference samples classified from the Landsat images.

	Reference			
Predicted	Double cropping	Forest	Pasture	Single cropping
DTW without time restrictions				
Double cropping	53	2	4	12
Forest	0	77	7	3
Pasture	5	25	136	5
Single cropping	57	1	19	77
Unclassified	0	1	3	2
DTW with maximum delay of 100 $days$				
Double cropping	104	1	1	11
Forest	0	93	7	0
Pasture	2	11	142	6
Single cropping	9	1	18	75
Unclassified	0	0	1	7
Linear TWDTW				
Double cropping	88	0	0	3
Forest	0	91	3	0
Pasture	3	15	132	8
Single cropping	24	0	34	88
Unclassified	0	0	0	0
Logistic TWDTW for $\alpha = 0.1$ and $\beta = 100\,days$				
Double cropping	104	0	0	9
Forest	0	94	6	0
Pasture	1	12	145	6
Single cropping	10	0	18	84
Unclassified	0	0	0	0

Table 2.3 - Equivalent classes for comparison between the TWDTW classification and MODIS land cover collection 5, Plant Functional Type (PFT).

Aggregated	MODIS PFT	TWDTW
Forest	Evergreen Needleleaf trees, Evergreen Broadleaf trees, and Deciduous Broadleaf trees	Forest, and Secondary Vegetation
Pastureland	Shrub and Grass	Pasture
Cropland	Cereal crops, and Broad-leaf crops	Single cropping and Double cropping

Table 2.4 - Assessment of MODIS collection 5 Plant Functional Type (PFT) and logistic TWDTW based on 489 reference samples classified from the Landsat images. The classes forest, pastureland, and cropland were aggregated according to Table 2.3.

	User (%)		Producer (%)	
Class	MODIS	TWDTW	MODIS	TWDTW
Forest	87.23	94.00	77.36	88.68
Pastureland	67.71	88.41	85.53	85.80
Cropland	89.33	92.00	75.28	96.73

the first year of the time series, the areas matching a forest pattern are classified as forest. There is no secondary vegetation in the first year of our classification. Using this rule, we got a class of "pristine forest" that is comparable to the PRODES dataset.

Since it is difficult to distinguish secondary vegetation from primary forest using visual interpretation of Landsat images, we joined these two classes to forest in the accuracy assessment. The total forest (pristine forest) and the secondary vegetation areas are presented in Figure 2.8. The forest area estimated using the logistic TWDTW is in line with the area estimated by PRODES (INPE, 2015). Most of the deforestation occurred before 2005, which was followed by an increase of the secondary vegetation area in 2007.

Figure 2.8 - Forest area estimated by the Amazon Monitoring Program PRODES (INPE, 2015) and using the logistic TWDTW based classification for Porto dos Gaúchos.

We also compared our estimated cropland area with the yearly Municipal Agricultural Production survey (PAM) from 2001 to 2013 done by the Brazilian Census Bureau (IBGE) (IBGE, 2014). The PAM survey provides the information on planted area, harvested area, amount produced, average yield and production value of permanent and temporary crops by municipality. Since PAM is a sampling survey and not a comprehensive census, some municipalities, especially those in the Brazilian Amazon, can have significant inter-annual variations. We use the PAM because it is the only survey that is available yearly for the period 2000 to 2013. Figure 2.9 shows the area of single cropping and double cropping estimated by using the logistic TWDTW algorithm and the Brazilian national cropland survey (IBGE, 2014) for Porto dos Gaúchos. There is a general agreement between our results and the crop surveys, except in the years 2009 and 2010.

Figure 2.9 - Total area of double cropping and single cropping in Porto dos Gaúchos estimated by TWDTW and the Brazilian national cropland survey (IBGE, 2014).

The total agricultural areas (pasture, single cropping, and double cropping) are shown in Figure 2.10. In the time series, the pasture and single cropping areas were increasing until 2006, while the double cropping area has a growing trend during the whole period. In the last two years of the time series, the double cropping exceeded the single cropping area.

Figure 2.10 - Total area of pasture, single cropping, and double cropping from 2001 to 2013 estimated using logistic TWDTW for Porto dos Gaúchos.

The Figure 2.11 shows the spatial distribution of land-use and land-cover in Porto dos Gaúchos for each second agricultural year from 2001 to 2013. In the last decade, a cropland intensification has happened in the Eastern part of Porto dos Gaúchos while pasture expansion has taken place in the Western part.

2.5 Discussion

Our results show that it pays to have a flexible approach to temporal restrictions when using DTW for land-use and land-cover classification. The original DTW method disregards the temporal range when finding the best alignment between two time series. This precludes an accurate land-use and land-cover classification. The time constraints included in the TWDTW similarity measure should be flexible to handle with the small phase changes related to natural phenological variability.

The maximum time delay, proposed by Petitjean et al. (2012), is flexible for small time warps. However it forces the dynamic algorithm, Equation 2.5, to map the warping path inside of a limiting time window that can preclude the classification of some areas (cf. unclassified samples in Table 2.2).

A large cost for small time warps, as the linear TWDTW method does, harms the classification and reduces its sensitivity. The linear TWDTW had low *producer's accuracy* respectively 78.11%, 76.52%, when classifying pasture and double cropping (cf. Table 2.1).

Figure 2.11 - Land use/cover maps produced by using the logistic TWDTW classification. Each map shows the classification for an agricultural year (from July to June) in Porto dos Gaúchos.

The DTW without time restriction had the worst results. More than half of the areas of double cropping were classified as single cropping. These errors come from the over warping of single cropping to fit the first growing season of the double cropping

occurrences, cf. Figure 2.7a.

The logistic TWDTW had better results for these land use classes, because of its low penalty for small time warps and its significant costs for large time warps. Its better accuracy derives from its flexibility to find the best match between a pattern and an interval within a long-term time series.

When comparing our cropland estimated area with data from the yearly Municipal Agricultural survey (IBGE, 2014), our results generally match, except for 2009 and 2010 (Figure 2.9). In the PAM, the large variations between 2008 and 2009 and between 2010 and 2011 are difficult to explain. Since this is a region of large-scale crop production, one would not expect such a large variation. This fact indicates that remote sensing time series analysis can complement and add value to cropland surveys such as PAM.

The forest area estimated using the logistic TWDTW was similar to the forest area from the INPE's Amazon Monitoring system (PRODES) (Figure 2.8). However, our algorithm had higher estimates for the forest area until 2006 and lower estimates for subsequent years. The higher forest area estimated by the logistic TWDTW compared to PRODES in the first years of the time series is likely related to their different scale of analysis. While we used MODIS images with 250 m spatial resolution the PRODES project uses 30 m Landsat images. Therefore PRODES is capable of detecting deforestation in small areas that may not be detected at the MODIS resolution.

In the second part of the graphic in Figure 2.8, the lower forest area estimated by our method was caused by the transition rule used in our algorithm to separate the secondary vegetation from the forest. Applying this rule an area that changes from forest to any other land class cannot become forest again. For example, after a degradation event (e.g. by fire) the area is classified as secondary vegetation in our algorithm, cf. Figure 2.12. Therefore, our estimation reduces from the forest area both deforested and degraded areas, whereas PRODES reduces from the forest area only the deforestation by clear-cutting, i.e. it reduces the forest area only when most or all the trees are fully removed.

One current challenge for large-scale application of TWDTW algorithm is its computational time. The implementation of the TWDTW algorithm was developed in `R` (IHAKA; GENTLEMAN, 1996; R Core Team, 2015) using the package `dtw` (GIORGINO, 2009). Our case study in Porto dos Gaúchos has $130,500$ time series,

Figure 2.12 - An example of a classification using the transition rules. This is a sample time series inside of a burned area. This area was degraded in 2011 according to the Detection of Forest Degradation Program (DEGRAD) (INPE, 2014).

each with 300 points. The computation time was around 50 minutes for all DTW variations on a server using 40 cores with 2.6 GHz clock and 256 GB memory. We expect that recent developments on specialized software such as array databases (STONEBRAKER et al., 2013), coupled with hardware advances and better indexing strategies will improve performance considerably.

2.6 Conclusion

This paper presents a version of the Dynamic Time Warping (DTW) algorithm suitable for land-use and land-cover classification of remote sensing time series. Refinements to standard DTW include a temporal restriction that allows for phase-shifts due to seasonal changes of natural and cultivated vegetation types. In a tropical forest area, the method has a high accuracy for mapping classes of single cropping, double cropping, forest, and pasture.

Accuracy assessments show the method compares favourably to other DTW variations for land classification. The logistic TWDTW had better results than the other tested alternatives with a global accuracy of 87.32%. Our classification using the logistic TWDTW has higher accuracy and spatial resolution than the MODIS land cover product. Forest and cropland areas are in line with the Amazon Monitoring Program PRODES and with the Brazilian national cropland surveys, respectively. These results highlight the potential of the TWDTW to improve land-use and land-cover products and contribute to agricultural statistics.

We expect that the TWDTW algorithm will be successful for large-scale land cover classification of remote sensing time series, if some conditions are met. If the spatial

and temporal resolutions of the data are adequate to capture the properties of the landscape, and the samples express the temporal variations of the land cover types, TWDTW has many advantages. Its flexibility for warping a temporal signature is useful to account for natural and cultivated vegetation types even with inter-annual climatic and seasonal variability.

The proposed method is pixel-based. We envisage future versions that include local neighborhoods to reduce border effects and improve classification homogeneity. Given that the DTW algorithm produces a distance measure between each interval of a long-term time series and all the temporal patterns, these measures could be used as a prior probability estimation for a Bayesian post-classification produce that borrows information from the neighbours.

Post-processing rules can improve TDWTW results. In the paper, we show how to use rules to distinguish pristine forest from forest regrowth. Using appropriate rules, it is also possible to apply the method for forest degradation, real-time change detection, and crop condition assessments.

The results on this paper have been obtained using only the EVI time series signal. We expect further improvements using multi-band time series, including the original spectral bands and transformed ones such as NDVI, EVI, and spectral unmixed endmembers.

The TWDTW algorithm is suitable for applications of remote sensing time series where the temporal variation is more important than the spatial variation for classifying remote sensing data sets. These cases include areas of large farms, such as those found in Brazil. For urban areas with less seasonal change or areas with small farms, it is likely that time warping methods need to be combined with object-based image analysis for accurate classification of the landscape.

3 dtwSat: TIME-WEIGHTED DYNAMIC TIME WARPING FOR SATELLITE IMAGE TIME SERIES ANALYSIS IN R[2]

3.1 Introduction

Remote sensing images are the most widely used data source for measuring land-use and land-cover change (LUCC). In many areas, remote sensing images are the only data available for this purpose (LAMBIN; GEIST, 2006; FRITZ et al., 2013). Recently, the opening of large archives of satellite data such as LANDSAT, MODIS and the SENTINELs has given researchers unprecedented access to data, allowing them to better quantify and understand local and global land change. The need to analyse such large data sets has lead to the development of automated and semi-automated methods for satellite image time series analysis. These methods include multi-image compositing (GRIFFITHS et al., 2013), detecting forest disturbance and recovery (KENNEDY et al., 2010; ZHU et al., 2012; DEVRIES et al., 2015), crop classification (XIAO et al., 2005; WARDLOW et al., 2007; PETITJEAN et al., 2012; MAUS et al., 2016), planted forest mapping (MAIRE et al., 2014), crop expansion and intensification (GALFORD et al., 2008; SAKAMOTO et al., 2009), detecting trend and seasonal changes (LUNETTA et al., 2006; VERBESSELT et al., 2010b; VERBESSELT et al., 2010a; VERBESSELT et al., 2012), and extracting seasonality metrics from satellite time series (JöNSSON; EKLUNDH, 2002; JöNSSON; EKLUNDH, 2004). Given the open availability of large image data sets, the research community on Earth Observation would get much benefit from methods that are openly available, reproducible and comparable. However, few of the proposed methods for remote sensing time series analysis are available as open source software, the main exception being the BFAST and BFAST-monitor algorithms for change detection (VERBESSELT et al., 2010b; VERBESSELT et al., 2010a). This paper is a contribution to making methods for satellite time series analysis available to a larger audience.

In this paper we describe the **dtwSat** package, written in R (R Core Team, 2015) and Fortran programming languages, and available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/package=dtwSat`. The package provides an implementation of Time-Weighted Dynamic Time Warping (TWDTW) (MAUS et al., 2016) for satellite image time series analysis.

The TWDTW method is an adaptation of the well-known dynamic time warping

[2]This chapter is based on the paper: MAUS, V.; CAMARA, G.; APPEL, M.; PEBESMA, E. dtwSat: Time-Weighted Dynamic Time Warping for satellite image time series analysis in R. Submitted to the Journal of Statistical Software.

(DTW) method for time series analysis (VELICHKO; ZAGORUYKO, 1970; SAKOE; CHIBA, 1971; SAKOE; CHIBA, 1978; RABINER; JUANG, 1993; BERNDT; CLIFFORD, 1994; KEOGH; RATANAMAHATANA, 2005; MüLLER, 2007) for land-use and land-cover classification. The standard DTW compares a temporal signature of a known event (*e.g.* a person's speech) with an unknown time series. It finds all possible alignments between two time series and provides a dissimilarity measure (RABINER; JUANG, 1993). In contrast to standard DTW, the TWDTW method is sensitive to seasonal changes of natural and cultivated vegetation types. It also considers inter-annual climatic and seasonal variability. In a tropical forest area, the method has achieved a high accuracy for mapping classes of single cropping, double cropping, forest, and pasture (MAUS et al., 2016).

We chose R because it is an open source software that offers a large number of reliable packages. The **dtwSat** package builds upon on a number of graphical and statistical tools in R: **dtw** (GIORGINO, 2009), **proxy** (MEYER; BUCHTA, 2015), **zoo** (ZEILEIS; GROTHENDIECK, 2005), **mgcv** (WOOD, 2000; WOOD, 2003; WOOD, 2004; WOOD, 2006; WOOD, 2011), **sp** (PEBESMA; BIVAND, 2005; BIVAND et al., 2013), **raster** (HIJMANS, 2015), **caret** (KUHN et al., 2016), and **ggplot2** (WICKHAM, 2009). Other R packages that are related and useful for remote sensing and land use analysis include **landsat** (GOSLEE, 2011), **rgdal** (BIVAND; LEWIN-KOH, 2015), **spacetime** (PEBESMA, 2012; BIVAND et al., 2013), **bfast** (VERBESSELT et al., 2010b; VERBESSELT et al., 2010a), **bfastmonitor** (VERBESSELT et al., 2011), **bfastSpatial** (DUTRIEUX; DEVRIES, 2014), **MODISTools** (TUCK et al., 2014), **maptools** (BIVAND; LEWIN-KOH, 2015), and **lucc** (MOULDS et al., 2015). Using existing packages as building blocks, software developers in R save a lot of time and can concentrate on their intended goals.

There is already a R package that implements the standard DTW method for time series analysis: the **dtw** package (GIORGINO, 2009). In the **dtwSat** package, we focus on the specific case of satellite image time series analysis. The analysis method implemented in **dtwSat** package extends that of the **dtw** package; it adjusts the standard DTW method to account for the seasonality of different types of land cover. Our aim is to support the full cycle of land-use and land-cover classification, from selecting sample patterns to visualising and evaluating the final result.

This paper focuses on the motivation and guidance for using the TWDTW method for remote sensing applications. The full description of the method is available in a paper published by the lead author (MAUS et al., 2016). In what follows, section 3.3 gives an overview of the **dtwSat** package. The section 3.2 describes the application

of TWDTW (MAUS et al., 2016) for satellite time series analysis. Then, section 3.4 focuses on the analysis of a single time series and shows some visualisation methods. We then present an example of a complete land use and land cover change analysis for a study area in the Mato Grosso, Brazil in section 3.5.

3.2 The Time-Weighted Dynamic Time Warping method

In this section, we describe the Time-Weighted Dynamic Time Warping (TWDTW) algorithm in general terms. For a detailed technical explanation, refer to Maus et al. (2016). TWDTW is time-constrained version of the Dynamic Time Warping (DTW) algorithm. Although the standard DTW method is good for shape matching (KEOGH; RATANAMAHATANA, 2005), it is not suited *per se* for satellite image time series analysis, since it disregards the temporal range when finding the best matches between two time series (MAUS et al., 2016). When using image time series for land cover classification, one needs to balance between shape matching and temporal alignment, since each land cover class has a distinct phenological cycle associated with the vegetation (ZHANG et al., 2003; REED et al., 1994). For example, soybeans and maize cycles range from 90 to 120 days, whereas sugar-cane has a 360 to 720 days cycle. A time series with cycle larger than 180 days is unlikely to come from soybeans or maize. For this reason, Maus et al. (2016) include a time constraint in DTW to account for seasonality. The resulting method is capable of distinguishing different land-use and land-cover classes.

The inputs to TWDTW are: (a) a set of time series of known temporal patterns (*e.g.* phenological cycles of land cover classes); (b) an unclassified long-term satellite image time series. For each temporal pattern, the algorithm finds all matching subintervals in the long-term time series, providing a dissimilarity measure (cf. Figure 3.1). The result of the algorithm is a set of subintervals, each associated with a pattern and with a dissimilarity measure. We then break the unclassified time series in periods according to our needs (e.g., yearly, seasonality, monthly). For each period, we consider all matching subintervals that intersect with it, and classify them based on the land cover class of the best matching subinterval. In this way, the long-term satellite time series is divided in periods, and each period is assigned a land cover class.

To use TWDTW for land-use and land-cover classification, we need the following data sets:

- A set of remote sensing time series for the study area. For example, a tile

Figure 3.1 - Matches of the known temporal pattern to subintervals of the long-term time series. The solid black line is the long-term time series, the colored lines are the different matches of the same pattern ordered by TWDTW dissimilarity measure, and the gray dashed lines are the matching points.

of a MODIS MOD13Q1 image consists of 4800 x 4800 pixels, covering an area of 10 degrees x 10 degrees at the Equator (FRIEDL et al., 2010). A 15-year (2000-2015) MODIS MOD13Q1 set time series has 23 images per year, with a total of 23 million time series, each with 346 samples.

- A set of time series with land cover information, called *temporal patterns*. Typically, each time series is short and covers one phenological cycle of one land cover type. Examples would be a time series of a soybean crop, or one that describes a mature tropical forest. These temporal patterns can be extracted from the remote sensing image data, if the user knows their spatial and temporal location.

- A set of ground truth points, with spatial and temporal information and land cover classification. These *ground truth* points are used for validation and accuracy assessment.

Based on the information provided by the user about the images to be analysed, our method maps them to a three-dimensional (3-D) array in space-time (Figure 3.2). This array can have multiple attributes, such as the satellite bands (*e.g.* "RED", "NIR", and "BLUE"), and derived indices (*e.g.* "NDVI", "EVI", and "EVI2"). This way, each pixel location is associated to a sequence of measurements, building a satellite image time series. Figure 3.2 shows an example of "EVI" time series for a location in the Brazilian Amazon from 2000 to 2008. In the first two years, the area was covered by forest that was cut in 2002. The area was then used for cattle raising

(pasture) for three years, and then for crop production from 2006 to 2008. Satellite image time series are thus useful to describe the dynamics of the landscape and the land use trajectories.

Figure 3.2 - A 3-dimensional array of satellite images (left), an enhanced vegetation index (EVI) time series at the pixel location (x, y) (right). The arrows indicate gaps in the time series. Adapted from Maus et al. (2016).

3.3 dtwSat package overview

dtwSat provides a set of functions for land cover change analysis using satellite image time series. This includes functions to build temporal patterns for land cover types, apply the TWDTW analysis using different weighting functions, visualise the results in a graphical interface, produce land cover maps, and create spatiotemporal plots for land changes. Therefore, **dtwSat** gives an end-to-end solution for satellite time series analysis, which users can make a complete land change analysis.

For the **dtwSat** package, the user should provide the following inputs:

- A set of time ordered satellite images, all with the same spatial extent. The user should also inform the date of each image. In R the images should use the `RasterBrick` or `RasterStack` class of the **raster** package.
- A list of temporal patterns, each associated to a time series in **zoo** format.
- A list of known ground truth points, each with spatial and temporal information, in a format readable in R, such as CSV or shapefile.

The **dtwSat** package organizes the data in three `S4` classes: `twdtwTimeSeries`, `twdtwMatches`, and `twdtwRaster`. To store time series we use the class

`twdtwTimeSeries`. The objects of class `twdtwTimeSeries` have two slots; the slot called `timeseries` has a list of `zoo` objects; and the slot called `labels` stores the labels of the time series. The class `twdtwMatches` has 3 slots to store inputs and results of the TWDTW analysis. The slots called `timeseries` and `patterns` are objects of the class `twdtwTimeSeries` with the unclassified long-term time series and the temporal patterns, respectively. A third slot called `alignments` has a `list` with detailed information about the matches between the patterns and the unclassified long-term time series. The classes `twdtwTimeSeries` and `twdtwMatches` are used to analyse lists of time series.

The class `twdtwRaster` is used for satellite image time series. This class can store either unclassified raster time series with the satellite raw data, the results of the TWDTW analyis, or a classified raster time series. In both cases, the objects of class `twdtwRaster` have five slots. The slot called `timeseries` is a list of `RasterBrick` or `RasterStack` objects with time ordered satellite images (all with the same temporal and spatial extents); the slot called `timeline` is a vector of class `Date` with dates of the satellite images; the slot called `layers` has the names of satellite bands; the slot called `levels` has levels for the raster values; and the slot called `labels` has labels for the raster values. This class builds upon the R package **raster** to build a multi-attribute 3-D raster in space-time, allowing for multi-band satellite image time series analysis.

3.4 Classifying a time series

This section describes how to classify one time series, using examples that come with the **dtwSat** package. We will show how to match three temporal patterns ("soybean", "cotton", and "maize") to subintervals of a long-term satellite image time series. These time series have been extracted from a set of MODIS MOD13Q1 (FRIEDL et al., 2010) images and include the vegetation indices "ndvi", "evi", and the original bands "nir", "red", "blue", and "mir". In this example, the classification of crop types for the long-term time series is known.

3.4.1 Input data

The inputs for the next examples are time series in **zoo** format. The first is an object of class `zoo` with a long-term time series, referred to as `example_ts`, and the second is a `list` of time series of class `zoo` with the temporal patterns of "soybean", "cotton", and "maize", referred to as `patterns.list`.

From `zoo` objects we construct time series of class `twdtwTimeSeries`, for which we have a set of visualization and analysis methods implemented in the **dtwSat** package. The code below builds two objects of class `twdtwTimeSeries`. The first has the long-term time series and second has the temporal patterns. We use the plot method types `timeseries` and `patterns` to shown the objects `ts` in Figure 3.3 and `patterns_ts` in Figure 3.4, respectively. This plot method uses `ggplot` syntax.

```
# Create twdtwTimeSeries
ts = twdtwTimeSeries(example_ts, labels="Time series")
patterns_ts = twdtwTimeSeries(patterns.list)
# Sequence of phenological cycle in ts
example_ts_labels
```

```
    label       from         to
1 Soybean 2009-09-01 2010-03-01
2  Cotton 2010-03-01 2010-09-01
3 Soybean 2010-09-01 2011-03-01
4  Cotton 2011-03-01 2011-09-01
5 Soybean 2011-09-01 2012-03-01
6   Maize 2012-03-01 2012-09-01
7 Soybean 2012-09-01 2013-03-01
8   Maize 2013-03-01 2013-09-01
```

```
library(dtwSat)
# List of temporal patterns
names(patterns.list)
```

```
[1] "Soybean" "Cotton"  "Maize"
```

```
# Example of time series
head(example_ts, n = 2)
```

```
             ndvi    evi    red    nir   blue    mir
2009-08-05 0.3169 0.1687 0.1167 0.2250 0.0427 0.2193
2009-08-28 0.2609 0.1385 0.1168 0.1993 0.0548 0.2657
```

```
# Plot objects of class twdtwTimeSeries
plot(ts, type = "timeseries") +
```

```
annotate(geom = "text", x = example_ts_labels$from+90,
y = 0.98, label = example_ts_labels$label, size = 2)
```

Figure 3.3 - Example of time series based on MODIS product MOD13Q1 (FRIEDL et al., 2010). The labels of the phenological cycle are shown in the plot.

```
# Plot objects of class twdtwTimeSeries
plot(patterns_ts, type = "patterns")
```

Figure 3.4 - Temporal patterns of soybean, cotton, and maize based on MODIS product MOD13Q1 (FRIEDL et al., 2010)

TWDTW uses both amplitude and phase information to classify the phenological cycles in the long-term time series. The EVI peak of the "soybean" time series has a similar amplitude as that of "cotton". However, the "soybean" series peaks in late December while the "cotton" series peaks in early April. The EVI peak of the

"maize" time series is at the same period as the peak of "cotton". However, the "maize" time series has smaller amplitude than the "cotton" one. Therefore, we can improve the time series classification by combining shape and time information.

3.4.2 Detection of time series patterns with TWDTW

Each subinterval of the long-term time series in `ts` has a known phenological cycle. We will now compare the known information with the result of the TWDTW algorithm. We use the function `twdtwApply` that returns an R object of class `twdtwMatches` with all matches of each temporal pattern in the time series.

```
log_weight = logisticWeight(alpha = -0.1, beta = 100)
matches = twdtwApply(x = ts, y = patterns_ts,
          weight.fun = log_weight, keep=TRUE)
slotNames(matches)
```

```
[1] "timeseries" "patterns"   "alignments"
```

```
show(matches)
```

```
An object of class "twdtwMatches"
Number of time series: 1
Number of Alignments: 16
Patterns labels: Soybean Cotton Maize
```

To retrieve the complete information of the matches we set `keep=TRUE`. We need this information for the plot methods of the class `twdtwMatches`. The argument `weight.fun` defines the time-weight to the dynamic time warping analysis (MAUS et al., 2016). By default the time-weight is zero, meaning that the function will run a standard dynamic time warping analysis. The arguments `x` and `y` are objects of class `twdtwTimeSeries` with the unclassified long-term time series and the temporal patterns, respectively. For details and other arguments see `?twdtwApply`.

In our example we use a logistic weight function for the temporal constraint of the TWDTW algorithm. This function is defined by `logisticWeight`. The **dtwSat** package provides two in-built functions: `linearWeight` and `logisticWeight`. The `linearWeight` function with slope `a` and intercept `b` is given by

$$\omega = a \cdot g(t_1, t_2) + b,$$

and the `logisticWeight` with midpoint `beta`, and steepness `alpha`, given by

$$\omega = \frac{1}{1 + e^{\alpha(g(t_1,t_2)-\beta)}}.$$

The function g is the absolute difference in days between two dates, t_1 and t_2. The linear function creates a strong time constraint even for small time differences. The logistic function has a low weight for small time warps and significant costs for bigger time warps, cf. Figure 3.5. In our previous studies (MAUS et al., 2016) the logistic-weight had better results than the linear-weight for land cover classification. Users can define different weight functions as temporal constraints in the argument `weight.fun` of the `twdtwApply` method.

Figure 3.5 - Logistic time-weight function `logisticWeight` with steepness `alpha=-0.1` and midpoint `beta=100`. The x axis shows the absolute difference between two dates in days and the y axis shows the time-weight (MAUS et al., 2016).

3.4.3 Visualising the result of the TWDTW algorithm

dtwSat provides five ways to visualise objects of class `twdtwMatches` through the plot types: `matches`, `alignments`, `classification`, `path`, and `cost`. The plot type `matches` shows the matching points of the patterns in the long-term time series; the plot type `alignments` shows the alignments and dissimilarity measures; the plot type `path` shows the low cost paths in the TWDTW cost matrix; and the plot type `cost` allows the visualisation of the cost matrices (local cost, accumulated cost, and time cost); and the plot type `classification` shows the classification of the long-term time series based on the TWDTW analysis. The plot methods for class `twdtwMatches` return a `ggplot` object, so that users can further manipulate the result using the **ggplot2** package. For more details on visualisation functions, please

refer to the **dtwSat** documentation in the CRAN (MAUS, 2015).

We now describe the plot types `matches` and `alignments`. The code bellow shows how to visualise the matching points of the four best matches of "soybean" pattern in the long-term time series, cf. Figure 3.6.

```
plot(matches, type="matches", patterns.labels="Soybean", k=4)
```

Figure 3.6 - The four best matches of the "soybean" pattern in the time series using a logistic time-weight. The solid black line is the long-term time series; the coloured lines are the temporal patterns; and the grey dashed lines are the respective matching points.

The next example (Figure 3.7) uses the plot type `alignments` to show the alignments of the temporal patterns. We set the threshold for the dissimilarity measure to be lower than 3.0. This is useful to display the different subintervals of the long-term time series that have at least one alignment whose dissimilarity is less than the specified threshold.

```
plot(matches, type="alignments", attr = "evi", threshold = 3.0)
```

3.4.4 Classifying the long-term time series

Using the matches and their associated dissimilarity measures, we can classify the subintervals of the long-term time series using `twdtwClassify`. To do this, we need to define a period for classification and the minimum overlap between the period and the alignments that intersect with it. We use the plot type `classification` to show the classification of the subintervals of the long-term time series based on

Figure 3.7 - Alignments and dissimilarity measures of the patterns "soybean", "cotton", and "maize" to the subintervals of the long-term time series using a logistic time-weight. The solid black line is the EVI time series, and the coloured lines are the alignments of the patterns that have dissimilarity measure lower than three.

the TWDTW analysis. For this example, we set classification periods of 6 months from September 2009 to September 2013, and a minimum overlap of 50% between the alignment and the classification period. This means that at least 50% of the alignment has to be contained inside of the classification period.

```
ts_classification = twdtwClassify(x = matches,
      from = as.Date("2009-09-01"), to = as.Date("2013-09-01"),
      by = "6 month", overlap = 0.5)
plot(ts_classification, type="classification")
```

Comparing the results of the classified time series in Figure 3.8 with the crop cycles in Figure 3.3 we see that the algorithm has classified correctly all the eight subintervals from 2009 to 2013, which are, respectively: “Soybean”, “Cotton”, “Soybean”, “Cotton”, “Soybean”, “Maize”, “Soybean”, “Maize”.

3.5 Producing a land cover map

In this section we present an application of TWDTW for land use and land cover change analysis using satellite image time series. Our input is a set of images, each

Figure 3.8 - Classification of each 6 months periods of the time series using results of the TWDTW analysis with logistic time-weight. The solid lines are the attributes of the time series, the background colours indicate the classification of the periods.

covering the same geographical area at different times. Each pixel location is then associated to an unclassified satellite image time series. We assume to have done field work in the area; for some pixel locations and time periods, we know what is the land cover. We then will show how to obtain a set of template patterns, based on the field samples and how to apply these patterns to land cover classification of the set of images. In the end of this section we show how to perform land cover change analysis and how to do accuracy assessment. The satellite images and the field samples used in the examples come with **dtwSat** package.

Our method is not restricted to cases where the temporal patterns are obtained from the set of images. The patterns for the TWDTW analysis can be any time series with same bands or indices as the unclassified images, such as in the examples of Section 3.4 above.

3.5.1 Input data

The inputs are: *a)* the satellite images for a given geographical area, organised as a set of georeferenced raster files in GeoTIFF format, each containing all time steps of a spectral band or index; and *b)* a set of ground truth samples. The satellite images are extracted from the MODIS product MOD13Q1 collection 5 (FRIEDL et al., 2010) and include vegetation indexes “ndvi”, “evi”, and original bands “nir”, “red”, “blue”,

and “mir”. This product has 250 x 250 m spatial and 16 day temporal resolution.

The region is a tropical forest area in Mato Grosso, Brazil of approximately 5300 km^2 with images from 2007 to 2013 (Figure 3.9). This is a sequence of 160 images with 999 pixels each for 6 years. We also have a set of 603 ground truth samples of the following classes: “forest”, “cotton-fallow”, “soybean-cotton”, “soybean-maize”, and “soybean-millet”.

Figure 3.9 - Study area in Mato Grosso, Brazil, shown in a © Google Earth image. The area was originally covered by tropical forest that has been removed for agricultural use.

The data files for the examples that follow are in the **dtwSat** installation folder *lucc_MT/data/*. The *tif* files include the time series of “blue”, “red”, “nir”, “mir”, “evi”, “ndvi”, and “doy” (day of the year); the text file *timeline* has the dates of the satellite images; the CSV file *samples.csv* has the `longitude`, `latitude`, `from`, `to`, and `label` for each field sample; and the text file *samples_projection* contains information about the cartographic projection of the samples, in the format of coordinate reference system used by `sp::CRS`.

```
data_folder = system.file("lucc_MT/data", package = "dtwSat")
dir(data_folder)
```

```
 [1] "blue.tif"           "doy.tif"            "evi.tif"
 [4] "mir.tif"            "ndvi.tif"           "nir.tif"
 [7] "red.tif"            "samples_projection" "samples.csv"
[10] "timeline"
```

In this example, we have stored all the time series for each band in one single file. In this way, we can use the function `raster::brick` to read the satellite images. The algorithm also works when the time steps for each band are split in many files. In this case, the user should call the function `raster::stack` with the appropriate parameters. Because of processing performance, we suggest that interested users group their images in bricks and follow the procedures given below.

```
blue = brick(paste(data_folder,"blue.tif", sep = "/"))
red  = brick(paste(data_folder,"red.tif", sep = "/"))
nir  = brick(paste(data_folder,"nir.tif", sep = "/"))
mir  = brick(paste(data_folder,"mir.tif", sep = "/"))
evi  = brick(paste(data_folder,"evi.tif", sep = "/"))
ndvi = brick(paste(data_folder,"ndvi.tif", sep = "/"))
day_of_year  = brick(paste(data_folder,"doy.tif", sep = "/"))
dates = scan(paste(data_folder,"timeline", sep = "/"),
             what = "dates")
```

The set of ground truth samples in the CSV file *samples.csv* has a total of 603 samples divided in five classes: 68 "cotton-fallow", 138 "forest", 79 "soybean-cotton", 134 "soybean-maize", and 184 "soybean-millet". Reading this CSV file, we get a `data.frame` object, with the spatial location (`latitude` and `longitude`), starting and ending dates (`from` and `to`), and the `label` for each sample.

```
field_samples = read.csv(paste(data_folder,"samples.csv", sep = "/"))
head(field_samples, 2)
```

```
  longitude  latitude       from         to         label
1 -55.98819 -12.03646 2011-09-01 2012-09-01 Cotton-fallow
2 -55.99118 -12.04062 2011-09-01 2012-09-01 Cotton-fallow
```

```
table(field_samples[["label"]])
```

```
 Cotton-fallow Forest Soybean-cotton Soybean-maize Soybean-millet
            68    138             79           134            184
```

```
proj_str = scan(paste(data_folder,"samples_projection", sep = "/"),
                what = "character")
proj_str
```

```
[1] "+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
```

3.5.2 Creating the time series and the temporal patterns

After reading our data, we need to create the time series for analysis. For this purpose, **dtwSat** provides the constructor `twdtwRaster` that builds a multi-band satellite image time series. The inputs of this function are `RasterBrick` objects with the same temporal and spatial extents, and a vector (`timeline`) with the acquisition dates of the images in the format `"YYYY-MM-DD"`. The argument `doy` is optional. If `doy` is not declared, the function builds a `RasterBrick` object using the dates given by `timeline`. This function produces an object of class `twdtwRaster` with the time series of multiple satellite bands.

```
raster_timeseries = twdtwRaster(blue, red, nir, mir, evi, ndvi,
                    timeline = dates, doy = day_of_year)
```

We now need to identify the temporal patterns. Usually, this can be done using the collected field samples. In the next example we use the function `getTimeSeries` to get the time series of each field sample from our raster time series. The arguments of the function are a set of raster time series, a `data.frame` with spatial and temporal information about the fields samples (as in the object `field_samples` given above), and a `proj4string` with the projection information. The projection should follow the `sp::CRS` format. The result is an object of class `twdtwTimeSeries` with one time series for each field sample.

```
field_samples_ts = getTimeSeries(raster_timeseries,
                  y = field_samples, proj4string = proj_str)
field_samples_ts
```

```
An object of class "twdtwTimeSeries"
Slot "timeseries" length: 603
Slot "labels": [1] Cotton-fallow Cotton-fallow Cotton-fallow
5 Levels: Cotton-fallow Forest Soybean-cotton ... Soybean-millet
```

After obtaining the time series associated to the field samples, we need to create the template patterns for each class. For this purpose, **dtwSat** provides the function `createPatterns`. This function fits a Generalized Additive Model (GAM) (HASTIE;

TIBSHIRANI, 1986; WOOD, 2011) to the field samples and retrieves a smoothed temporal pattern for each band (*e.g.* "blue", "red", "nir", "mir", "evi", and "ndvi"). We use the GAM because of its flexibility for non-parametric fits, with less rigorous assumptions on the relationship between response and predictor. This potentially provides better fit to satellite data than purely parametric models, due to the data's inter- and intra-annual variability.

To produce the set of template patterns using the function `createPatterns`, we need to set the temporal frequency of the resulting patterns and the smoothing function for the GAM model. In the example below, we set `freq=8` to get temporal patterns with a frequency of 8 days. We also set the GAM smoothing formula to be `formula = y  s(x)`, where function `s` sets up a spline model, with `x` the time and `y` a satellite band (for details see `?mgcv::gam` and `?mgcv::s`).

```
temporal_patterns =
  createPatterns(field_samples_ts, freq = 8, formula = y ~ s(x))
```

We use the plot method `type="patterns"` to show the results of the `createPatterns` in Figure 3.10.

```
plot(temporal_patterns, type="patterns") +
     theme(legend.position = c(.8,.25))
```

After obtaining the template patterns for each land cover class, it is useful to perform a pre-classification analysis to assess their quality and their informational content. Ideally, one would need template patterns that, when applied to the set of unknown time series, produce consistent results. For this reason, it is advisable that the user performs a pre-classification step, along the lines of the individual analysis described in Section 3.4. In this way, the users would assess how good their patterns are before classifying a large data set.

3.5.3 Classifying the image time series

After obtaining a consistent set of temporal patterns, we use the function `twdtwApply` to run the TWDTW analysis for each pixel location in the raster time series. The input raster time series in the object `twdtwRaster` should be longer or have approximatly the same length as the temporal patterns. This function retrieves an object of class `twdtwRaster` with the TWDTW dissimilarity measure of the

Figure 3.10 - Temporal patterns of forest, cotton-fallow, soybean-cotton, soybean-maize, and soybean-millet based on the ground truth samples.

temporal patterns for each time period. The arguments `overwrite` and `format` are passed to `raster::writeRaster`. The arguments `weight.fun` and `overlap` are described in section 3.4. In the next example, we classify the raster time series using the temporal patterns in `temporal_patterns` obtained as described above. The result is a `twdtwRaster` with five layers; each layer contains the TWDTW dissimilarity measure for one temporal pattern over time. We use the plot type `distance` to illustrate the TWDTW dissimilarity for each temporal pattern in 2013, cf. Figure 3.11 .

```
log_fun = logisticWeight(alpha=-0.1, beta=50)
twdtw_dist =
        twdtwApply(x = raster_timeseries, y = temporal_patterns,
        overlap = 0.5, weight.fun = log_fun, overwrite=TRUE,
        format="GTiff")


plot(x = twdtw_dist, type="distance", time.levels = 6)
```

The results of the example above can be used to create categorical land cover maps. The function `twdtwClassify` selects the most similar pattern for each time period and retrieves a `twdtwRaster` object with the time series of land use maps. The

Figure 3.11 - Illustration of the TWDTW dissimilarity from each temporal pattern in 2013. The blue areas are more similar to the pattern and the red areas are less similar to the pattern.

resulting object includes two layers, the first has the classified categorical maps and the second has the TWDTW dissimilarity measure.

```
land_use_maps = twdtwClassify(twdtw_dist,
                              format="GTiff", overwrite=TRUE)
```

3.5.4 Looking at the classification results

The classification results can be visualised using the `plot` methods of the class `twdtwRaster`, which supports four plot types: “maps”, “area”, “changes”, and “distance”. The `type="maps"` shows the land cover classification maps for each period, cf. Figure 3.12.

```
plot(x = land_use_maps, type = "maps")
```

The next example shows the accumulated area for each class over time, using `type="area"`, cf. Figure 3.13.

Figure 3.12 - Land use maps for each year from 2008 to 2013.

```
plot(x = land_use_maps, type = "area")
```

Figure 3.13 - Percentage of area for each land use class from 2008 to 2013.

Users can also view the land cover transition for each time period, by setting `type="changes"`. For each land cover class and each period, the plot shows gains and losses in area from the other classes. This is the visual equivalent of a land

transition matrix, cf. Figure 3.14.

```
plot(x = land_use_maps, type = "changes")
```

Figure 3.14 - Gains and losses in area from the other classes. The y axis shows the actual class; the positive direction of x axis shows the gains and the negative direction of x axis shows the losses of the classes indicated in y. The colors indicate from/to which classes the gains/losses belong.

We can also look at the dissimilarity of each classified pixel setting `type="distance"`. This plot can give a measure of the uncertainty of the classification of each pixel for each time period, cf. Figure 3.15.

```
plot(x = land_use_maps, type="distance")
```

3.5.5 Assessing the classification accuracy

In this section we show how to assess the accuracy of the TWDTW method for land cover classification. To do this, we split the ground truth samples into training and validation sets, using the function `splitDataset` from the package **dtwSat**. This function splits set of time series in the object `twdtwTimeSeries` for training and validation. The argument `p` defines the percentage used for training and the argument `times` gives the number of different partitions to create. This is a a stratified sampling with a simple random sampling within each stratum, see

Figure 3.15 - TWDTW dissimilarity measure for each pixel over each classified period. The blue areas have high confidence and the red areas have low confidence in the classification.

`?createDataPartition` for details. In the next example we create 100 different partitions of the data. Each partition uses 10% of the data for training and 90% for validation. The output is a list with 100 different data partitions; each partition has the temporal patterns based on the training samples and a set of time series for validation.

```
set.seed(1)
partitions = splitDataset(field_samples_ts, p=0.1, times=100,
                          freq = 8, formula = y ~ s(x, bs="cc"))
```

For each data partition we run the TWDTW analysis to classify the set of validation time series using the trained temporal patterns. The result is a list of `twdtwMatches` objects with the classified set of time series for each data partition. To compute the *User's Accuracy* (UA) and *Producer's Accuracy* (PA) of the classified time series we use the function `dtwSat::twdtwAssess` that retrieves a `data.frame` with the accuracy assessment for all data partitions.

```
log_fun = logisticWeight(alpha=-0.1, beta=50)
twdtw_res = lapply(partitions, function(x){
    res = twdtwApply(x = x$ts, y = x$patterns,
                     weight.fun = log_fun, n=1)
    twdtwClassify(x = res)
})
assessment = twdtwAssess(twdtw_res)
head(assessment, 5)
```

```
  resample          label        UA       PA
1        1  Cotton-fallow 1.0000000 0.953125
2        1         Forest 1.0000000 1.000000
3        1 Soybean-cotton 0.8732394 1.000000
4        1  Soybean-maize 1.0000000 0.952381
5        1 Soybean-millet 1.0000000 1.000000
```

Figure 3.16 shows the average μ and standard deviation σ of *user's* and *producer's accuracy* based on a bootstrap simulation of 100 different data partitions using resampling-with-replacement. The *user's accuracy* gives the confidence and the *producer's accuracy* gives the sensitivity of the method for each class. In our analysis all classes had high *user's* and *producer's accuracy*, meaning that TWDTW has high confidence and sensitivity for the classes included in the analysis. The average, standard deviation, and the 99% confidence interval is also shown in Table 3.1.

Figure 3.16 - User's Accuracy (UA) and Producer's Accuracy (PA) of the TWDTW method for land cover classification. The plot shows the averages and their confidence interval for 99%.

Table 3.1 - User's and Producer's Accuracy of the land use classification based on TWDTW analysis. μ is the average accuracy, σ the standard deviation, and CI is the confidence interval of 99% using 100 resampling-with-replacement.

Class	User's Accuracy (UA) %			Producer's Accuracy (PA)%		
	μ	σ	CI	μ	σ	CI
Cotton-fallow	99.93	(0.46)	[99.80-100.00]	95.78	(1.29)	[95.42-96.07]
Forest	100.00	(0.00)	[100.00-100.00]	100.00	(0.00)	[100.00-100.00]
Soybean-cotton	88.92	(1.81)	[88.45-89.35]	98.58	(5.37)	[96.90-99.70]
Soybean-maize	99.58	(0.95)	[99.33-99.78]	93.51	(4.79)	[92.10-94.58]
Soybean-millet	97.12	(6.09)	[95.33-98.50]	99.67	(0.70)	[99.47-99.82]

3.6 Conclusions and Discussion

Nowadays, there are large open archives of Earth Observation data, but few open source methods for analysing them. With this motivation, this paper provides guidance on how to use the Time-Weighed Dynamic Time Warping (TWDTW) method for remote sensing applications. As we have discussed in a companion paper (MAUS et al., 2016), the TWDTW method is well suited for land cover change analysis of satellite image time series.

The main goal of **dtwSat** package is to make TWDTW accessible for researchers. The package supports the full cycle of land cover classification using image time series, ranging from selecting temporal patterns to visualising and evaluating the results. The current version of the **dtwSat** package provides a pixel-based time series classification method. We envisage that future versions of the package could include local neighborhoods to reduce border effects and improve classification homogeneity.

To aim for maximum usage by the scientific community, the **dtwSat** package described in this paper works with well-known R data classes such as provided by packages **zoo** and **raster**. We are planning improvements, so that **dtwSat** can be combined with array databases, such as SciDB (STONEBRAKER et al., 2013). We believe that combining array databases with image time series analysis software such as presented here is one way forward to scaling the process of information extracting to very large Earth Observation data.

4 LAND-USE AND LAND-COVER TRAJECTORIES IN MATO GROSSO BRAZIL[3]

4.1 Introduction

Brazil intends to reduce its greenhouse gas (GHG) emissions by 37% in 2025 and by 43% in 2030 below 2005 levels (BRAZIL, 2015). Agriculture, forestry, and other land use related sectors have been the main source of Brazil's GHG emissions. These sectors together contributed to 78% of Brazil's emissions in 2005 and 52% in 2012 (MCTI, 2014). The decreasing rate of deforestation was the main sector responsible for this reduction, dropping its contribution to GHG emissions from 58% to 20% in the same period. The cuts in the Brazilian Amazon rain forest, for example, decreased by 79% from 27,772 km^2 in 2004 to 5,831 km^2 in 2015 (INPE, 2015). Meanwhile, the agricultural sector has become a more important driver Brazil's GHG emissions, increasing its contribution to GHG emissions from 15% to 37% (MCTI, 2014). Therefore, the land use related sectors are crucial for near-term GHG mitigation in Brazil.

The Amazon rain forest has a crucial role on biodiversity, climate stability, and other ecological services (LEWIS et al., 2015). Therefore, it is important to understand the land use dynamics and the causes of the forest removal (LAMBIN et al., 2003; AGUIAR et al., 2006). For this reason, the Brazilian National Institute for Space Research (INPE) and the Brazilian Agricultural Research Corporation (EMBRAPA) have mapped the land use after the deforestation in the Brazilian Amazon for every two years since 2008 (INPE, 2012). Their results show that in 2012 60% of the cut forest areas were used for cattle raising whereas regeneration areas (secondary vegetation) accounted for 23%. Cattle ranches in Amazon use extensive practices, with less than one head of cattle per hectare. Cash crop agriculture accounted for only 6% of the deforestation. To achieve further gains in reducing deforestation and biodiversity loss, we need to understand the different land use trajectories, including deforestation dynamics, land use intensification, and land abandonment pathways.

In this paper we explore the potential of spatio temporal mapping to study land cover trajectories. For that, we used the Time-Weighted Dynamic Time Warping (TWDTW) method (MAUS et al., 2016) to classify the Amazon biome at Mato Grosso, Brazil, from 2001 to 2014. This region has been largely deforested, accounting for 139,824 km^2 since 1988 (INPE, 2015). The deforestation rate in Mato

[3]This is an initial version of the paper to be submitted to the journal Remote Sensing of Environment.

Grosso has been reduced during the last years, from 11,814 km^2 in 2004 to 1,508 km^2 in 2015, decreasing by 87% (INPE, 2015). Meanwhile, land use intensification process is taking place through multiple cropping systems (GALFORD et al., 2008; RUFIN et al., 2015). Therefore, some interesting questions about the land cover trajectories in this region are: Which are the main land use after deforestation? Which are the trajectories of agricultural intensification? What are the policy impacts on deforestation? Answering these questions we can contribute to better understand the land-use and land-cover change (LUCC) in the Brazilian Amazon and inform policy-makers.

4.2 Classifying satellite image time series

We use the TWDTW method (MAUS et al., 2016) to create a sequence of annual maps from 2001 to 2014. The inputs to TWDTW are: (a) a set of time series of known temporal patterns, e.g., the phenological cycle of the vegetation associated with a land cover class, (b) an unclassified long-term satellite image time series. For each temporal pattern, the algorithm finds all matching subintervals in the long-term time series, providing a dissimilarity measure. The result of the algorithm is a set of subintervals, each associated with a pattern and with a dissimilarity measure. In a second step the algorithm breaks the unclassified time series in periods according to our needs (e.g., yearly, seasonality, monthly) and classifies them based on the best matching land cover class in each subinterval. This way, the long-term satellite time series is divided into periods where each period is assigned a land cover class. In what follows we describe the inputs and outputs of the TWDTW algorithm and how the results are combined with transition rules to classify the satellite image time series.

4.2.1 Creating the temporal patterns

The first step before applying the TWDTW analysis is to define the temporal patterns for each land cover class in the study area. We used a Generalized Additive Model (GAM) (HASTIE; TIBSHIRANI, 1986) to create a smoothed temporal pattern for each class and each band of the satellite image, e.g. “BLUE”, “RED”, “NIR”, “MIR”, “EVI”, and “NDVI”. The GAM is flexible for irregularly sampled time series and non-parametric fits, which mean less rigorous assumptions on the relationship between response and predictor. This potentially provides a better fit to satellite data than purely parametric models, due to the data’s inter- and intra-annual variability.

4.2.2 Applying the TWDTW analysis

The TWDTW method is an adaptation of the well-known Dynamic Time Warping (DTW) method used for time series analysis (VELICHKO; ZAGORUYKO, 1970; SAKOE; CHIBA, 1971; SAKOE; CHIBA, 1978; RABINER; JUANG, 1993; BERNDT; CLIFFORD, 1994; KEOGH; RATANAMAHATANA, 2005; MüLLER, 2007). TWDTW is a robust algorithm that can handle with auto-of-phase and irregularly sampled satellite image time series (MAUS et al., 2016). This algorithm compares a well known temporal pattern to an unclassified long-term satellite image time series and finds their best matches by taking into account both amplitude and phase, Figure 4.1. In a tropical forest area, the method has achieved a high accuracy for mapping classes of single cropping, double cropping, forest, and pasture (MAUS et al., 2016).

Figure 4.1 - Matches of the temporal patterns of different land cover classes to a long-term EVI time series using the Time-Weighted Dynamic Time Warping (TWDTW) algorithm.
Source: Adapted from Maus et al. (2016).

The TWDTW algorithm compares the temporal patterns with the satellite image time series for each pixel location in the study area and find which subintervals are associated with one of the pre-defined patterns. This set of subintervals associated with a pattern and a dissimilarity measure are used to create categorical land cover maps as we describe below.

4.2.3 Creating annual land cover maps

As a third step, the algorithm breaks the unclassified time series into 12-month periods according to the Brazilian agricultural calendar, from July to next July. For each period, we consider all the TWDTW matches that have more than 50%

overlap with the period, meaning that at least 50% of the match has to be inside of the classification period. We then classify each period based on the best matching, i.e. the lowest time warping cost, among the land cover classes in the interval. This way, the long-term satellite image time series is divided into yearly periods, and a land cover class is assigned to each period.

In our classification we also split degraded forests and regrowing vegetation from primary forest. The temporal patterns of primary forests and secondary forests are similar over the year, therefore, they are not easily distinguished in the MODIS spatial resolution of 250 m. For this reason, we used land cover transition rules based on the underlying causes that create conditions for the appearance of secondary forest Figure 4.2. The typology adopted here was adapted from Chokkalingam and Jong (2001) and distinguish: primary forest, degraded forest, and secondary forest regrowth.

The "degraded forest" follows a disturbance event that can be natural and/or human-initiated. Both causes are included because they can have significant contribution in the formation of secondary forests. Chokkalingam and Jong (2001) argues that, often natural disturbances are initiated or fuelled by human activities, and it is difficult to split the source of the event. Therefore, "degraded forest" includes regenerating forests after catastrophic natural and/or anthropogenic activities, such as post-fire and post-extraction secondary forests. The sequence below shows the "degraded forest" formation

$$\textit{forest} \implies \textit{disturbance} \implies \textit{degraded forest}$$

We also assume that several disturbance events can occur over the same area, i.e. the disturbance event can be recurrent. Therefore, a second possible pathway to "degraded forest" is

$$\textit{degraded forest} \implies \textit{disturbance} \implies \textit{degraded forest}$$

In our analysis we also estimate "secondary forest regrowth", which is a regeneration after an agricultural use. This secondary forest exists only if the land has been used for agricultural activities at any time from the begin of the time series until the present moment. It includes regrowth due to land abandonment and/or

rehabilitation efforts on degraded lands, but it does not distinguish natural from planted regenerating forests. The sequence below shows the "secondary forest regrowth" formation

$$Agriculture \implies abandonment \implies secondary\ forest\ regrowth$$

To split "primary forest", "degraded forest", and "secondary forest regrowth" we use the transition rules in Figure 4.2. Areas matching a forest pattern are classified as "primary forest" only if they have been classified as "primary forest" in the previous year. Otherwise, we classified them as secondary forest. For the first year of the time series, the areas matching a forest pattern are classified as "primary forest". The two main implications of these rules are: a) there is no secondary forest in the first year of the classification, and b) the primary forest area decreases asymptotically.

Figure 4.2 - Land cover transitions rules to classify the time intervals. These rules are used to split: primary forest, degraded forest, and secondary forest regrowth.

4.3 Datasets and validation

Our study area is the Amazon biome at Mato Grosso, Brazil, which covers an area of 489,700 km^2 Figure 4.3. To create the land maps we used multiple-bands time series extracted from the MODIS product MOD13Q1 collection 5. Bisides the real date of each pixel in the image, we used the vegetation indices "NDVI", "EVI", and the original bands "NIR", "RED", "BLUE", and "MIR". This product has 250 x 250 m spatial and 16 day temporal resolution using Maximum Value Composition (MVC) (FRIEDL et al., 2010). Not that we used the real day of each pixel, therefore, our time series are not regularly sampled.

Figure 4.3 - Study area covering the Amazon biome at Mato Grosso, Brazil.

We used 6 MODIS tiles to cover the whole study area, including all images from 2000 to 2014. This means 2070 MODIS images where each image has 23,040,000 pixels. To handle this big data set, we used the array database system SciDB (STONEBRAKER et al., 2013), which can store satellite images in a 3-dimensional (3-D) array, Figure 4.4. This storage system allows designing spacetime data analytic methods in a generic and sensor independent way.

In addition to the satellite data, we also used a set of 2593 ground truth samples divided into six land cover classes: "forest", "pasture", "soybean", "cotton", "soybean-cotton", "soybean-maize", Table 4.1. These samples were collected by the Brazilian Agricultural Research Corporation (EMBRAPA), for details see Arvor et

Figure 4.4 - A 3-dimensional (3-D) array of satellite images. Each pixel has a set of attributes (e.g. "EVI", "RED", etc) associated with a pixel location over time. Source: Adapted from Maus et al. (2016).

al. (2011). We used these data set to create the temporal patterns of the phenological cycle associated with each land cover class. This data was also used to assess the accuracy of the classification. Using a bootstrap algorithm we simulated 1000 resampling-with-replacement to estimate the averaged (μ) at the confidence interval of 99% (CI) of the User's and Producer's accuracy. Each simulation used 10% of the ground truth samples for training and 90% for validation.

Table 4.1 - Ground truth samples used to create the temporal patterns and to perform the accuracy assessment of the land cover classification. These samples were collected by the Brazilian Agricultural Research Corporation (EMBRAPA) (ARVOR et al., 2011).

	Class	Aggregated class	Number of sample
1	Forest	Forest	493
2	Pasture	Pasture	421
3	Soybean	Single cropping	653
4	Cotton	Single cropping	198
5	Soybean-Maize	Double cropping	561
6	Soybean-Cotton	Double cropping	267

4.4 Results

4.4.1 Accuracy assessment

Figure 4.5 shows the average of the *user's* and *producer's accuracy* based on a bootstrap simulation with 1000 data partitions. The figure also shows the confidence

interval at the level of 99%. In our analysis all classes had high *user's* and *producer's accuracy*, meaning a high confidence and sensitivity for the samples data set. The average (μ), standard deviation (σ), and the confidence interval (CI) are also shown in Table 4.2.

Figure 4.5 - User's and Producer's Accuracy of the classification based on the TWDTW analysis. The graphic shows the average at confidence interval of 99% computed with a bootstrap simulation of 1000 resampling-with-replacement.

Table 4.2 - User's and Producer's Accuracy of the land use classification based on TWDTW analysis. μ is the averaged accuracy and σ the standard deviation.

Class	User's Accuracy (UA) %			Producer's Accuracy (PA) %		
	μ	σ	CI*	μ	σ	CI*
Cotton-Fallow	95.79	(2.98)	[95.54-96.02]	91.49	(1.27)	[91.39-91.59]
Forest	99.78	(0.10)	[99.77-99.78]	99.88	(0.11)	[99.87-99.89]
Pasture	99.86	(0.29)	[99.84-99.89]	96.67	(0.75)	[96.61-96.74]
Soybean-Cotton	82.46	(1.89)	[82.30-82.61]	99.34	(0.70)	[99.28-99.40]
Soybean-Fallow	97.06	(0.86)	[96.99-97.13]	98.93	(0.87)	[98.86-99.00]
Soybean-Maize	99.75	(0.20)	[99.73-99.76]	93.90	(1.17)	[93.80-93.99]

* Confidence interval at the level of 99% computed with a bootstrap simulation of 1000 resampling-with-replacement.

4.4.2 Land cover change

Using our field samples we built the set of temporal patterns in Figure 4.6. These patterns show the phenological cycle of the most important land cover classes in the study area. Their amplitude and phase information can be used to distinguish between each other. For example, the single cropping systems with cotton and soybean have only one peak but with different phase. The double cropping systems with soybean-maize and soybean-cotton have phenological events happening approximately at the same time but with different shape and amplitude. These temporal patterns were used as input for the TWDTW algorithm to classify the study area.

Using the TWDTW method for land cover classification, we built a map for each agricultural year from 2000 to 2014. In Figure 4.7 we present the result for 2001 and 2014, showing a great expansion of the human activities over the whole area. This trend can be also visualized in Figure 4.8 that shows the percentage of total area. The complete time series of land cover maps is available in appendix A.

Figure 4.9 shows the time series of cash crop area in the study area. There is a growing trend in area of double cropping systems (soybean-maize and soybean-cotton), whereas the area of single cropping systems (soybean and cotton) has reduced in the last years.

Pasture is the most dominating human activity in the study area, Figure 4.8. The last two years of the time series show a reduction of pasture area, as we can see

Figure 4.6 - Set of temporal patterns associated with phenological cycle of forest, pasture/grassland, single cropping systems (cotton and soybean), and double cropping system (soybean-maize and soybean-cotton).

in Figure 4.10. This figure also shows the area of degraded forest, secondary forest regrowth, and deforestation/forest degradation.

Figure 4.11 shows an example of degradation next to human activities in the municipality of Porto dos Gaúchos, Brazil. In the center of this area there are two events of degradation over time, the first in 2008 and the second in 2011.

Figure 4.12 shows the annual changes among the different land cover classes, i.e. the gains and losses in area of each land cover class. In Figure 4.14 we can see the accumulated change from each land cover class to all the other classes form 2001 to 2014.

Figure 4.14 shows the land cover transitions related to pasture and soybean. The transitions from forest to soybean (blue bars) show a growing trend until 2005 that is reduced to almost zero after 2006. The changes from pasture to soybean from areas deforested after 2006 (purple bars) are low compared to the changes from

Figure 4.7 - Land cover classification in 2001 and 2014 in the Amazon biome at Mato Grosso, Brazil. The complete time series of land cover maps is available in appendix A.

Figure 4.8 - Percentage of area from 2001 to 2014 in the Amazon biome at Mato Grosso, Brazil.

Figure 4.9 - Total area of cash crop agriculture over time in the Amazon biome at Mato Grosso, Brazil.

Figure 4.10 - Total area of pasture, degraded forest, secondary forest regrowth, and deforestation/forest degradation over time in the Amazon biome at Mato Grosso, Brazil.

areas deforested before 2006 (green bars). The red bars show the pasture area that came from deforestation over the time. This kind of transition shows some peaks in 2003, 2006, 2009, and 2012.

4.5 Discussion and Conclusions

The expansion of the human activities over our study area Figure 4.7 is mostly dominated by pasture, cf. Figure 4.8. This result is in line with the results of the TerraClass project (INPE, 2012). The pasture has been the main direct land use after deforestation, 16.03% of the accumulated land use change over the 14 years, Figure 4.14. Other studies have reported the creation of pasture for cattle raising as the main driver of deforestation (GEIST; LAMBIN, 2002). In our results we also observed that pasture has given space to cash crop activities, 7.42% to single and 1.46% to double cropping over all changes. However, the main changes from pasture are to secondary forest regrowth, 12.73%. This shows that many areas of pasture have been abandoned, corroborating other studies (SPERA et al., 2014).

In our study the accumulated change from single cropping to double cropping reached 4.95% during the 14 years, Figure 4.14. Other studies have also found that single cropping is giving space to double cropping (GALFORD et al., 2008; RUFIN et al., 2015; SPERA et al., 2014). In our study we also identified that the most common crop system in areas of intensification is soybean-maize (double cropping), which

Figure 4.11 - Example of forest degradation in Porto dos Gaúchos, Brazil, at the latitude -11°54' and longitude -56°40'.

Figure 4.12 - Gains and losses in area from the other classes in the study area. The y axis shows the actual class; the positive direction of x axis shows the gains and the negative direction of x axis shows the losses of the classes indicated in y. The colors indicate from/to which classes the gains/losses belong.

From \ To	Primary forest	Deforestation/Forest degradation	Degraded forest	Secondary forest regrowth	Pasture/grassland	Single cropping	Double cropping	Unclassified
Primary forest		100%	0%	0%	0%	0%	0%	0%
Deforestation/Forest degradation	0%		30.4%	9.5%	55.2%	2.9%	0.6%	1.5%
Degraded forest	0%	100%		0%	0%	0%	0%	0%
Secondary forest regrowth	0%	100%	0%		0%	0%	0%	0%
Pasture/grassland	0%	0%	0%	53.5%		31.2%	6.2%	9.1%
Single cropping	0%	0%	0%	2.5%	54.2%	6.9%	34.9%	1.5%
Double cropping	0%	0%	0%	1.9%	19%	45.4%	33.6%	0.2%
Unclassified	0%	0%	4.4%	1.6%	81%	11.9%	1.1%	

Figure 4.13 - Matrix of changes in the Amazon Biome within Mato Grosso, Brazil. The values are the percentage of area for each class related to the total changes accumulated from 2001 to 2014.

Figure 4.14 - Land cover transitions from 2003 to 2014 in the Amazon biome at Mato Grosso, Brazil. The bars show the area converted from forest to soybean (blue), from pasture to soybean that has been deforested before 2006 (green), from pasture to soybean that has been deforested after 2006 (purple), and from forest to pasture (red).

increased by 94%, from $890\,km^2$ in 2001 to $1,725\,km^2$ in 2014, cf. Figure 4.9. The crop system with soybean-maize also overcame the single cropping systems of soybean in the last three years of the time series.

The secondary forest regrowth comes mainly from pasture, 12.73%, considering all the changes from 2001 to 2014, Figure 4.10. Meanwhile, the accumulated change from degraded forest to deforestation/forest degradation reached 5.74%, which point towards recurrence of degradation events in areas previously degraded, cf. Figure 4.14.

Our results can also help to better understand the land cover change and its drivers. The reduction on the soybean expansion over the forest areas after 2006 (cf. Figure 4.14) is an effect of soy moratorium. Another impact of the soybean moratorium is that areas of pasture deforested after 2006 have not been converted to soybean. However, areas of pasture deforested before 2006 have been converted to soybean. Meanwhile, the deforestation is still happening in order to open new pasture for cattle raising. This can be an effect of the soybean taken place over the

pasture, which is then taking the forest area.

Our classification showed a systematic overestimation of the pasture coming from deforestation in 2003, 2006, 2009, and 2012, cf. Figure 4.14. This problem is related to the transition rules that were not adequate to separate natural inter-annual osculations from degradation/deforestation events. The transitions rules still need to be improved in order to better separate degradation/deforestation from natural oscillations. Areas of flood plain, for example, have a great inter-annual variability and were misclassified as degradation and consequently as degraded forest. In some cases these areas were also misclassified as pasture.

In this study we identified and estimated the land change in the Amazon region using MODIS image time series. The results contribute to better understanding of land cover dynamics in the Brazilian Amazon. Our method has the advantage of being an automated classification process that can be used to produce annual maps for large areas. We believe that the method can be extended to other Brazilian biomes to help understand land change in the whole Brazilian territory.

5 SYNTHESIS

This thesis shows that making a contribution to advance Earth System Science sometimes requires a substantial background effort before producing actual results that advance our scientific understanding. The motivation for this work is the need to improve on the land-use and land-cover classification of big Earth observation data sets. I now revise the main contributions of this thesis and then discuss further research topics.

5.1 How to deal with irregularly sampled satellite image time series for land-use and land-cover classification?

Open archives of long-term satellite image time series provide opportunities to better quantify global change, which has lead to the development of automated and semi-automated methods for satellite image time series analysis. The challenge is that satellite image time series will always be contaminated by some degree of residual atmospheric influence, geolocation error, and directional effects. Therefore, to make the best use of available satellite data archives, methods for satellite image time series analysis need to deal with data sets that are noisy, irregularly sampled, and in many cases out-of-phase.

To tackle the challenge, this thesis proposed a solution based on the Dynamic Time Warping method. Methods based on DTW have achieved significant results in time-series data mining. The large range of applications of dynamic time warping for time series analysis motivated the idea of using DTW for remote sensing applications. The original DTW method works well for shape matching, but is not suited for remote sensing time-series classification. This is because it disregards the temporal range when finding the best alignment between two time series. Since the vegetation associated to each land cover class has a specific phenological cycle, a good time-series land cover classifier needs to balance between shape matching and temporal alignment.

In this thesis, I adjusted the original DTW method to include a temporal weight that accounts for seasonality of the vegetation associated to land cover types. The algorithm compares two time series and finds their optimal alignment, providing a dissimilarity measure as a result. DTW provides a robust distance measure for comparing time series, even if they are irregularly sampled or are out of phase in the time axis.

In a tropical forest area, the method showed a high accuracy for mapping classes of single cropping, double cropping, forest, and pasture. Accuracy assessments showed that the method compares favourably to other DTW variations for land classification. The logistic TWDTW had better results than the other tested alternatives with a global accuracy of 87.32%. Our classification using the logistic TWDTW has higher accuracy and spatial resolution than the MODIS land cover product. Forest and cropland areas are in line with the Amazon Monitoring Program PRODES and with the Brazilian national cropland surveys, respectively. These results highlight the potential of the TWDTW to improve land-use and land-cover products and contribute to agricultural statistics.

5.2 Contribution to make methods for satellite time series analysis open and reproducible

Given the open availability of large image data sets, such as LANDSAT, MODIS, and SENTINELs, the research community on Earth Observation would get much benefit from methods that are openly available, reproducible and comparable. However, few of the proposed methods for remote sensing time series analysis are available as open source software, the main exception being the BFAST and BFAST-monitor algorithms for change detection.

This thesis contributed to open and reproducible science by developing an R package called **dtwSat** that provides an implementation of the Time-Weighed Dynamic Time Warping (TWDTW) method (see chapter 3) for satellite image time series analysis. The package is freely available from the Comprehensive R Archive Network. In addition, end-to-end (step-to-step) guidance of how to implement the TWDTW method by using the dtwSat package is currently submitted to the Journal of Statistical Software (JSS).

The **dtwSat** package allows researchers to use the TWDTW method for full cycle of land cover classification using image time series, including 1) selecting temporal patters, 2) visualizing the results, and 3) evaluating the result. To aim for maximum usage by the scientific community, the **dtwSat** package works with well-known R data classes such as provided by packages **zoo** and **raster**.

5.3 What can we learn from improved land-use and land-cover data sets?

I used the TWDTW method and MODIS image time series to analyse LUCC in the Amazon biome, focusing on the Brazilian state of Mato Grosso. This area has gone through high rate of deforestation and cropland expansion in the last decade. This study showed that pasture is the dominant land use after deforestation, whereas most of the single cropping area comes from pasture, and the cropping system is undergoing intensification from single to double cropping. Moreover, the regenerative secondary forest comes mainly from pasture. In the case study most of the land use intensification through double cropping system comes from the soybean-maize cultivation.

The results contribute to better understanding of land-use and land-cover dynamics in the Brazilian Amazon. The method used in this study has the advantage of being an automated classification process that can be used to produce annual maps. This study shows the potential of the TWDTW method for large-scale remote sensing data analysis. I believe that the method can be extended to other Brazilian biomes to help understand land change in the whole Brazilian territory.

5.4 Future work

Based on my experience and the results of this thesis, I now describe some specific points for future work:

- The proposed method is pixel-based. Future versions including local neighborhoods can reduce border effects and improve classification homogeneity. Given that the DTW algorithm produces a distance measure between each interval of a long-term time series and all the temporal patterns, these measures could be used as a prior probability estimation for a Bayesian post-classification produce that borrows information from the neighbours.
- Post-processing rules can improve TDWTW results. This thesis showed how to use rules to distinguish pristine forest from forest regrowth. Using appropriate rules, it is also possible to apply the method for forest degradation, real-time change detection, and crop condition assessments.
- The results have been obtained using only the MODIS product MOD13Q1. Further improvements are expected by using data sets from multi-band

sensors with higher spatial resolution, such as the new SENTINELs.

- The TWDTW algorithm is suitable for applications of remote sensing time series where the temporal variation is more important than the spatial variation for classifying remote sensing data sets. These cases include areas of large farms, such as those found in Brazil. For urban areas with less seasonal change or areas with small farms, it is likely that time warping methods need to be combined with object-based image analysis for accurate classification of the landscape.

- I expect that the TWDTW algorithm will be successful for large-scale land cover classification of remote sensing time series, if some conditions are met. If the spatial and temporal resolutions of the data are adequate to capture the properties of the landscape, and the samples express the temporal variations of the land cover types, TWDTW has many advantages. Its flexibility for warping a temporal signature is useful to account for natural and cultivated vegetation types even with inter-annual climatic and seasonal variability.

- The current version of the **dtwSat** package provides a pixel-based time series classification method. Future versions of the package could include local neighborhoods to reduce border effects and improve classification homogeneity.

- Combining array databases with image time series analysis software such as presented here is one way forward to scaling the process of information extracting to very large Earth Observation data.

5.5 Conclusion

The methods developed in this these have a substantial potential for contribution on LUCC modelling. Extensive results in the literature point out that the DTW method is a robust way to do time series data mining. The time constraint introduced in DTW improves the algorithm's potential for use with remote sensing data. As it is the case of all data mining methods, the actual performance of TWDTW will depend on the quality of the input samples and how these samples can manage to distinguish between the different land classes. In general, good samples will support a good classification.

The next steps will be to develop an infrastructure for large-scale remote sensing data

analysis and to apply TWDTW to the classification of land change in all Brazilian biomes. Some adjustments will be required, but the resulting data sets can provide important new information.

REFERENCES

AGUIAR, R. G.; RANDOW, C. V.; FILHO, N. P.; MANZI, A. O.; AGUIAR, L. J. G.; CARDOSO, F. L. Fluxos de massa e energia em uma floresta tropical no sudoeste da amazônia. **Brazilian Journal of Meteorology**, v. 21, n. 3b, p. 248–257, 2006. 12, 13, 51

ALEXANDRATOS, N.; BRUINSMA, J. et al. **World agriculture towards 2030/2050**: the 2012 revision. Rome: FAO/ESA Working paper No. 12-03, 2012. 1

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. xiii, xvii, 13, 14, 57

BEDDINGTON, J. Food security: contributions from science to a new and greener revolution. **Philosophical Transactions of the Royal Society of London B: Biological Sciences**, The Royal Society, v. 365, n. 1537, p. 61–71, 2009. 1

BERNDT, D. J.; CLIFFORD, J. **Using Dynamic Time Warping to Find Patterns in Time Series**. Menlo Park, California: AAAI Press, 1994. 359-370 p. 3, 5, 28, 53

BIVAND, R.; LEWIN-KOH, N. **maptools**: tools for reading and handling spatial objects. [S.l.], 2015. R package version 0.8-37. Available from: <CRAN.R-project.org/package=maptools>. 28

BIVAND, R. S.; PEBESMA, E.; GOMEZ-RUBIO, V. **Applied spatial data analysis with R**. 2. ed. New York: Springer, 2013. 28

BRAZIL. Intended nationally determined contribution: towards achieving the objective of the United Nations Framework Convention on climate change. In: PARIS CLIMATE CONFERENCE (COP21). Paris: UN, 2015. Available from: <www4.unfccc.int/submissions>. 51

CHAPIN, F. S.; ZAVALETA, E. S.; EVINER, V. T.; NAYLOR, R. L.; VITOUSEK, P. M.; REYNOLDS, H. L.; HOOPER, D. U.; LAVOREL, S.; SALA, O. E.; HOBBIE, S. E.; MACK, M. C.; DIAZ, S. Consequences of changing biodiversity. **Nature**, Macmillan Magazines Ltd., v. 405, n. 6783, p. 234–242, may 2000. ISSN 0028-0836. 1

CHAPLIN-KRAMER, R.; SHARP, R. P.; MANDLE, L.; SIM, S.; JOHNSON, J.; BUTNAR, I.; CANALS, L. Milá i; EICHELBERGER, B. A.; RAMLER, I.; MUELLER, C.; MCLACHLAN, N.; YOUSEFI, A.; KING, H.; KAREIVA, P. M. Spatial patterns of agricultural expansion determine impacts on biodiversity and carbon storage. **Proceedings of the National Academy of Sciences of the United States of America**, National Academy of Sciences, v. 112, n. 24, p. 7402–7407, jun 2015. ISSN 1091-6490. 1

CHOKKALINGAM, U.; JONG, W. de. Secondary forest: a working definition and typology. **The International Forestry Review**, JSTOR, p. 19–26, 2001. 54

DEVRIES, B.; VERBESSELT, J.; KOOISTRA, L.; HEROLD, M. Robust monitoring of small-scale forest disturbances in a tropical montane forest using landsat time series. **Remote Sensing of Environment**, v. 161, n. 0, p. 107 – 121, 2015. 2, 5, 27

DUTRIEUX, L.; DEVRIES, B. **bfastSpatial**: set of utilities and wrappers to perform change detection on satellite image time-series. 2014. R package version 0.6.2. Available from: <https://github.com/dutri001/bfastSpatial>. 28

ESLING, P.; AGON, C. Time-series data mining. **ACM Comput. Surv.**, v. 45, n. 1, p. 12:1–12:34, dec 2012. ISSN 0360-0300. 3, 5

FOLEY, J. A.; DEFRIES, R.; ASNER, G. P.; BARFORD, C.; BONAN, G.; CARPENTER, S. R.; CHAPIN, F. S.; COE, M. T.; DAILY, G. C.; GIBBS, H. K.; HELKOWSKI, J. H.; HOLLOWAY, T.; HOWARD, E. A.; KUCHARIK, C. J.; MONFREDA, C.; PATZ, J. A.; PRENTICE, I. C.; RAMANKUTTY, N.; SNYDER, P. K. Global Consequences of Land Use. **Science**, v. 309, n. 5734, p. 570–574, 2005. 1

FREITAS, R. M. de; ARAI, E.; ADAMI, M.; FERREIRA, A. S.; SATO, F. Y.; SHIMABUKURO, Y. E.; ROSA, R. R.; ANDERSON, L. O.; RUDORFF, B. F. T. Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over south america. **Journal of Computational Interdisciplinary Sciences**, v. 2, n. 1, p. 57–68, 2011. 12

FRIEDL, M. A.; SULLA-MENASHE, D.; TAN, B.; SCHNEIDER, A.; RAMANKUTTY, N.; SIBLEY, A.; HUANG, X. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. **Remote Sensing of Environment**, v. 114, n. 1, p. 168 – 182, 2010. ISSN 0034-4257. xiv, 18, 30, 32, 34, 39, 56

FRITZ, S.; SEE, L.; YOU, L.; JUSTICE, C.; BECKER-RESHEF, I.; BYDEKERKE, L.; CUMANI, R.; DEFOURNY, P.; ERB, K.; FOLEY, J.; GILLIAMS, S.; GONG, P.; HANSEN, M.; HERTEL, T.; HEROLD, M.; HERRERO, M.; KAYITAKIRE, F.; LATHAM, J.; LEO, O.; MCCALLUM, I.; OBERSTEINER, M.; RAMANKUTTY, N.; ROCHA, J.; TANG, H.; THORNTON, P.; VANCUTSEM, C.; VELDE, M. van der; WOOD, S.; WOODCOCK, C. The need for improved maps of global cropland. **Eos, Transactions American Geophysical Union**, v. 94, n. 3, p. 31–32, 2013. ISSN 2324-9250. 1, 5, 27

GALFORD, G. L.; MUSTARD, J. F.; MELILLO, J.; GENDRIN, A.; CERRI, C. C.; CERRI, C. E. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in brazil. **Remote Sensing of Environment**, v. 112, n. 2, p. 576–587, 2008. 2, 5, 27, 52, 63

GEIST, H. J.; LAMBIN, E. F. Proximate causes and underlying driving forces of tropical deforestation: tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. **BioScience**, v. 52, n. 2, p. 143–150, 2002. 63

GIORGINO, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. **Journal of Statistical Software**, v. 31, n. 7, p. 1–24, 2009. 24, 28

Google. **Google Earth Engine**. 2014. Available from: <https://earthengine.google.org/>. 13

GOSLEE, S. Analyzing remote sensing data in R: the landsat package. **Journal of Statistical Software**, v. 43, n. 1, p. 1–25, 2011. ISSN 1548-7660. 28

GRIFFITHS, P.; LINDEN, S. van der; KUEMMERLE, T.; HOSTERT, P. A pixel-based landsat compositing algorithm for large area land cover mapping. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 6, n. 5, p. 2088–2101, Oct 2013. ISSN 1939-1404. 2, 5, 27

HASTIE, T.; TIBSHIRANI, R. Generalized additive models. **Statistical Science**, Institute of Mathematical Statistics, v. 1, n. 3, p. 297–310, 1986. 43, 52

HIJMANS, R. J. **raster**: geographic data analysis and modeling. [S.l.], 2015. R package version 2.5-2. Available from: <CRAN.R-project.org/package=raster>. 28

HUETE, A.; DIDAN, K.; MIURA, T.; RODRIGUEZ, E.; GAO, X.; FERREIRA, L. 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. ISSN 0034-4257. 12

HUETE, A.; LIU, H.; BATCHILY, K.; LEEUWEN, W. van. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote Sensing of Environment**, v. 59, n. 3, p. 440 – 451, 1997. 12

BRAZILIAN INSTITUTE OF GEOGRAPHY AND STATISTICS (IBGE). **Municipal Agricultural Production - PAM**. 2014. Available from: <www.sidra.ibge.gov.br/bda>. Access: 5th March 2015. xiii, 13, 21, 24

IHAKA, R.; GENTLEMAN, R. R: a language for data analysis and graphics. **Journal of computational and graphical statistics**, Taylor & Francis, v. 5, n. 3, p. 299–314, 1996. 24

NATIONAL INSTITUTE FOR SPACE RESEARCH, BRAZIL (INPE). **Amazon land use database - TerraClass**. 2012. Available from: <www.inpe.br/cra/projetos_pesquisas/terraclass2008>. Access: 27th July 2015. 13, 51, 63

_____. **Detection of forest degradation - (DEGRAD)**. 2014. Available from: <www.obt.inpe.br/degrad>. Access: 15th December 2015. xiv, 25

_____. **Amazon Deforestation Monitoring Project - PRODES**. 2015. Available from: <www.obt.inpe.br/prodes>. Access: 5th December 2015. xiii, 12, 13, 18, 20, 51, 52

IPCC. **Climate Change 2014**: synthesis report, contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate changes. [S.l.], 2014. 1

JEONG, Y.-S.; JEONG, M. K.; OMITAOMU, O. A. Weighted dynamic time warping for time series classification. **Pattern Recognition**, v. 44, n. 9, p. 2231 – 2240, 2011. ISSN 0031-3203. Computer Analysis of Images and Patterns. 6, 10

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. 1824–1832, 2002. 2, 5, 27

_____. Timesat–a program for analyzing time-series of satellite sensor data. **Computers & Geosciences**, v. 30, n. 8, p. 833 – 845, 2004. ISSN 0098-3004. 2, 27

KENNEDY, R. E.; YANG, Z.; COHEN, W. B. Detecting trends in forest disturbance and recovery using yearly landsat time series: LandTrendr – temporal segmentation algorithms. **Remote Sensing of Environment**, v. 114, n. 12, p. 2897–2910, 2010. ISSN 0034-4257. 2, 5, 27

KEOGH, E.; RATANAMAHATANA, C. A. Exact indexing of dynamic time warping. **Knowledge Information Systems**, v. 7, n. 3, p. 358–386, 2005. 3, 6, 28, 29, 53

KUHN, M.; WING with contributions from J.; WESTON, S.; WILLIAMS, A.; KEEFER, C.; ENGELHARDT, A.; COOPER, T.; MAYER, Z.; KENKEL, B.; TEAM the R. C.; BENESTY, M.; LESCARBEAU, R.; ZIEM, A.; SCRUCCA, L.; TANG, Y.; CANDAN, C. **caret**: classification and regression training. [S.l.], 2016. R package version 6.0-64. Available from: <CRAN.R-project.org/package=caret>. 28

LAMBIN, E. F.; GEIST, H. **Land-Use and Land-Cover Change**: local processes and global impacts. [S.l.]: Springer, 2006. (Global Change – The IGBP Series). 1, 2, 27

LAMBIN, E. F.; GEIST, H. J.; LEPERS, E. Dynamics of land-use and land-cover change in tropical regions. **Annual Review of Environment and Resources**, v. 28, n. 1, p. 205–241, 2003. 51

LAMBIN, E. F.; MEYFROIDT, P. Global land use change, economic globalization, and the looming land scarcity. **Proceedings of the National Academy of Sciences**, v. 108, n. 9, p. 3465–3472, 2011. 1

LEWIS, S. L.; EDWARDS, D. P.; GALBRAITH, D. Increasing human dominance of tropical forests. **Science**, v. 349, n. 6250, p. 827–832, 2015. 3, 51

LUNETTA, R. S.; KNIGHT, J. F.; EDIRIWICKREMA, J.; LYON, J. G.; WORTHY, L. D. Land-cover change detection using multi-temporal MODIS NDVI data. **Remote Sensing of Environment**, v. 105, n. 2, p. 142 – 154, 2006. 2, 5, 27

MAIRE, G. le; DUPUY, S.; NOUVELLON, Y.; LOOS, R. A.; HAKAMADA, R. Mapping short-rotation plantations at regional scale using MODIS time series:

case of eucalypt plantations in Brazil. **Remote Sensing of Environment**, v. 152, n. 0, p. 136 – 149, 2014. 2, 5, 27

MALLAT, S. **A Wavelet Tour of Signal Processing**. 2. ed. San Diego, California, USA: Academic Press An imprint of Elsevier, 1998. 637 p. 12

MAUS, V. **dtwSat**: Time-Weighted Dynamic Time Warping for remote sensing time series analysis. [S.l.], 2015. R package version 0.1.0. 3, 37

MAUS, V.; CâMARA, G.; CARTAXO, R.; RAMOS, F. M.; SANCHEZ, A.; RIBEIRO, G. Q. Open boundary dynamic time warping for satellite image time series classification. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2015. **Proceedings of the IEEE International Geoscience and Remote Sensing Symposium**. Milan, Italy: IEEE, 2015. p. 3349–3352.

MAUS, V.; CAMARA, G.; CARTAXO, R.; SANCHEZ, A.; RAMOS, F. M.; QUEIROZ, G. R. de. A time-weighted dynamic time warping method for land-use and land-cover mapping. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, PP, n. 99, p. 1–11, 2016. ISSN 1939-1404. xiv, xv, 27, 28, 29, 31, 35, 36, 50, 51, 52, 53, 57

MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION, BRAZIL (MCTI). **Estimativas Anuais de Emissões de Gases de Efeito Estufa no Brasil**. 2. ed. Brasília, 2014. 3, 51

MELLO, M. P.; VIEIRA, C. A. O.; RUDORFF, B. F. T.; APLIN, P.; SANTOS, R. D. C.; AGUIAR, D. A. STARS: A new method for multitemporal remote sensing. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 4, p. 1897–1913, April 2013. ISSN 0196-2892. 2

MEYER, D.; BUCHTA, C. **proxy**: distance and similarity measures. [S.l.], 2015. R package version 0.4-15. Available from: <`CRAN.R-project.org/package=proxy`>. 28

MILLENNIUM ECOSYSTEM ASSESSMENT. **Ecosystems and human well-being**. Washington, DC: Island press, 2005. 1

MOULDS, S.; BUYTAERT, W.; MIJIC, A. An open and extensible framework for spatially explicit land use change modelling: the lulcc R package. **Geoscientific Model Development**, v. 8, n. 10, p. 3215–3229, 2015. 28

MüLLER, M. **Information retrieval for music and motion**. London: Springer, 2007. 64-84 p. ISBN 9783540740483. 8, 9, 10, 28, 53

PEBESMA, E. spacetime: apatio-temporal data in R. **Journal of Statistical Software**, v. 51, n. 1, p. 1–30, 2012. ISSN 1548-7660. 28

PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R News**, v. 5, n. 2, p. 9–13, November 2005. 28

PETITJEAN, F.; INGLADA, J.; GANCARSKI, P. Satellite image time series analysis under time warping. **IEEE Transactions on Geoscience and Remote Sensing**, v. 50, n. 8, p. 3081–3095, 2012. ISSN 0196-2892. 2, 3, 5, 6, 8, 11, 14, 22, 27

PETITJEAN, F.; WEBER, J. Efficient satellite image time series analysis under time warping. **IEEE Geoscience and Remote Sensing Letters**, v. 11, n. 6, p. 1143–1147, June 2014. 6, 8

PIAO, S.; FRIEDLINGSTEIN, P.; CIAIS, P.; NOBLET-DUCOUDR´, N. de; LABAT, D.; ZAEHLE, S. Changes in climate and land use have a larger direct impact than rising co(2) on global river runoff trends. **Proceedings of the National Academy of Sciences of the United States of America**, National Academy of Sciences, v. 104, n. 39, p. 15242–15247, oct 2006. ISSN 1091-6490. 1

PIELKE, R. A.; MARLAND, G.; BETTS, R. A.; CHASE, T. N.; EASTMAN, J. L.; NILES, J. O.; NIYOGI, D. d. S.; RUNNING, S. W. The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. **Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, The Royal Society, v. 360, n. 1797, p. 1705–1719, 2002. ISSN 1364-503X. 1

R Core Team. **R**: a language and environment for statistical computing. Vienna, Austria, 2015. Available from: <https://www.R-project.org/>. 3, 24, 27

RABINER, L.; JUANG, B.-H. **Fundamentals of speech recognition**. [S.l.]: Prentice-Hall International, Inc., 1993. 3, 5, 6, 28, 53

RAKTHANMANON, T.; CAMPANA, B.; MUEEN, A.; BATISTA, G.; WESTOVER, B.; ZHU, Q.; ZAKARIA, J.; KEOGH, E. Searching and mining trillions of time series subsequences under dynamic time warping. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND

DATA MINING. **Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining**. New York, USA: ACM, 2012. p. 262–270. 3, 5

REED, B. C.; BROWN, J. F.; VANDERZEE, D.; LOVELAND, T. R.; MERCHANT, J. W.; OHLEN, D. O. Measuring phenological variability from satellite imagery. **Journal of Vegetation Science**, v. 5, n. 5, p. 703–714, 1994. 6, 8, 10, 29

ROERINK, G. J.; MENENTI, M.; VERHOEF, W. Reconstructing cloudfree NDVI composites using Fourier analysis of time series. **International Journal of Remote Sensing**, Taylor & Francis, v. 21, n. 9, p. 1911–1917, 2000. 5

RUFIN, P.; MüLLER, 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. ISSN 0303-2434. 52, 63

SAKAMOTO, T.; VAN, P. C.; KOTERA; NGUYEN, K. D.; YOKOZAWA, M. Analysis of rapid expansion of inland aquaculture and triple rice-cropping areas in a coastal area of the Vietnamese Mekong Delta using MODIS time-series imagery. **Landscape and Urban Planning**, v. 92, n. 1, p. 34–46, 2009. 2, 5, 27

SAKAMOTO, T.; YOKOZAWA, M.; TORITANI, H.; 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. ISSN 0034-4257. 12

SAKOE, H.; CHIBA, S. A dynamic programming approach to continuous speech recognition. In: INTERNATIONAL CONGRESS ON ACOUSTICS. **Proceedings of the Seventh International Congress on Acoustics**. Budapest: Akadémiai Kiadó, 1971. v. 3, p. 65–69. 3, 5, 10, 28, 53

_____. Dynamic programming algorithm optimization for spoken word recognition. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 26, n. 1, p. 43–49, feb 1978. 3, 5, 28, 53

SALA, O. E.; CHAPIN, F. S.; III; ARMESTO, J. J.; BERLOW, E.; BLOOMFIELD, J.; DIRZO, R.; HUBER-SANWALD, E.; HUENNEKE, L. F.; JACKSON, R. B.; KINZIG, A.; LEEMANS, R.; LODGE, D. M.; MOONEY, H. A.; OESTERHELD, M.; POFF, N. L.; SYKES, M. T.; WALKER, B. H.;

WALKER, M.; WALL, D. H. Global biodiversity scenarios for the year 2100. **Science**, v. 287, n. 5459, p. 1770–1774, 2000. 1

SALAME, C. W.; QUEIROZ, J. C. B.; ROCHA, G. d. M.; AMIN, M. M.; ROCHA, E. P. da. Use of spatial regression models in the analysis of burnings and deforestation occurrences in forest region, Amazon, Brazil. **Environmental Earth Science**, 75, n. 3, Feb 2016.

SCANLON, B. R.; JOLLY, I.; SOPHOCLEOUS, M.; ZHANG, L. Global impacts of conversions from natural to agricultural ecosystems on water resources: quantity versus quality. **Water Resources Research**, Wiley Online Library, v. 43, n. 3, 2007. 1

SEE, L.; SCHEPASCHENKO, D.; LESIV, M.; MCCALLUM, I.; FRITZ, S.; COMBER, A.; PERGER, C.; SCHILL, C.; ZHAO, Y.; MAUS, V.; SIRAJ, M. A.; ALBRECHT, F.; CIPRIANI, A.; VAKOLYUK, M.; GARCIA, A.; RABIA, A. H.; SINGHA, K.; MARCARINI, A. A.; KATTENBORN, T.; HAZARIKA, R.; SCHEPASCHENKO, M.; VELDE, M. van der; KRAXNER, F.; OBERSTEINER, M. Building a hybrid land cover map with crowdsourcing and geographically weighted regression. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 103, n. 0, p. 48 – 56, 2015. ISSN 0924-2716. Global Land Cover Mapping and Monitoring. 1

SPERA, S. A.; COHN, A. S.; VANWEY, L. K.; MUSTARD, J. F.; RUDORFF, B. F.; RISSO, J.; ADAMI, M. Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. **Environmental Research Letters**, v. 9, n. 6, p. 064010, 2014. 63

STEFFEN, W.; RICHARDSON, K.; ROCKSTRöM, J.; CORNELL, S. E.; FETZER, I.; BENNETT, E. M.; BIGGS, R.; CARPENTER, S. R.; VRIES, W. de; WIT, C. A. de; FOLKE, C.; GERTEN, D.; HEINKE, J.; MACE, G. M.; PERSSON, L. M.; RAMANATHAN, V.; REYERS, B.; SöRLIN, S. Planetary boundaries: guiding human development on a changing planet. **Science**, v. 347, n. 6223, 2015. 1

STOCKER, B. D.; ROTH, R.; JOOS, F.; SPAHNI, R.; STEINACHER, M.; ZAEHLE, S.; BOUWMAN, L.; XU-RI; PRENTICE, I. C. Multiple greenhouse-gas feedbacks from the land biosphere under future climate change scenarios. **Nature Clim. Change**, Nature Publishing Group, v. 3, n. 7, p. 666–672, jul 2013. 1

STONEBRAKER, M.; BROWN, P.; ZHANG, D.; BECLA, J. SciDB: a database management system for applications with complex analytics. **Computing in Science & Engineering**, v. 15, n. 3, p. 54–62, 2013. 25, 50, 56

TUCK, S. L.; PHILLIPS, H. R.; HINTZEN, R. E.; SCHARLEMANN, J. P.; PURVIS, A.; HUDSON, L. N. MODISTools – downloading and processing MODIS remotely sensed data in R. **Ecology and Evolution**, v. 4, n. 24, p. 4658–4668, 2014. ISSN 2045-7758. 28

UN. **World Population Prospects**: the 2015 revision. 2015. Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241. 1

VELICHKO, V.; ZAGORUYKO, N. Automatic recognition of 200 words. **International Journal of Man-Machine Studies**, v. 2, n. 3, p. 223–234, 1970. ISSN 0020-7373. 3, 5, 28, 53

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 1, p. 106–115, 2010. ISSN 0034-4257. 2, 3, 27, 28

VERBESSELT, J.; HYNDMAN, R.; ZEILEIS, A.; CULVENOR, D. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 12, p. 2970 – 2980, 2010. ISSN 0034-4257. 2, 3, 5, 27, 28

VERBESSELT, J.; ZEILEIS, A.; HEROLD, M. **Near Real-Time Disturbance Detection in Terrestrial Ecosystems Using Satellite Image Time Series**: drought detection in somalia. [S.l.], 2011. 28

_____. Near real-time disturbance detection using satellite image time series. **Remote Sensing of Environment**, v. 123, n. 0, p. 98 – 108, 2012. 2, 5, 27

WARDLOW, B. D.; EGBERT, S. L.; KASTENS, J. H. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. central great plains. **Remote Sensing of Environment**, v. 108, n. 3, p. 290 – 310, 2007. 2, 5, 27

WICKHAM, H. **ggplot2**: elegant graphics for data analysis. [S.l.]: Springer-Verlag New York, 2009. ISBN 978-0-387-98140-6. 28

WOOD, S. **Generalized Additive Models**: an introduction with R. [S.l.]: Chapman and Hall/CRC, 2006. 28

WOOD, S. N. Modelling and smoothing parameter estimation with multiple quadratic penalties. **Journal of the Royal Statistical Society (B)**, v. 62, n. 2, p. 413–428, 2000. 28

_____. Thin-plate regression splines. **Journal of the Royal Statistical Society (B)**, v. 65, n. 1, p. 95–114, 2003. 28

_____. Stable and efficient multiple smoothing parameter estimation for generalized additive models. **Journal of the American Statistical Association**, v. 99, n. 467, p. 673–686, 2004. 28

_____. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. **Journal of the Royal Statistical Society (B)**, v. 73, n. 1, p. 3–36, 2011. 28, 43

XIAO, X.; BOLES, S.; LIU, J.; ZHUANG, D.; FROLKING, S.; LI, C.; SALAS, W.; III, B. M. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. **Remote Sensing of Environment**, v. 95, n. 4, p. 480 – 492, 2005. 2, 5, 27

ZEILEIS, A.; GROTHENDIECK, G. zoo: S3 infrastructure for regular and irregular time series. **Journal of Statistical Software**, v. 14, n. 6, p. 1–27, 2005. 28

ZHANG, X.; FRIEDL, M. A.; SCHAAF, C. B.; STRAHLER, A. H.; HODGES, J. C.; GAO, F.; REED, B. C.; HUETE, A. Monitoring vegetation phenology using MODIS. **Remote Sensing of Environment**, v. 84, n. 3, p. 471 – 475, 2003. 6, 8, 10, 29

ZHU, Z.; WOODCOCK, C. E.; OLOFSSON, P. Continuous monitoring of forest disturbance using all available landsat imagery. **Remote Sensing of Environment**, v. 122, n. 0, p. 75–91, 2012. Landsat Legacy Special Issue. 2, 5, 27

APPENDIX A - LAND COVER MAPS

This appendix shows the land cover maps from 2001 to 2014 for the Amazon biome at Mato Grosso, Brazil. The methodology of the classification is described in chapter 4. Figure A.1 presents the legend of the land cover maps in Figures A.2-A.8.

Figure A.1 - Legend of the land cover maps.

Figure A.2 - Land cover classification in 2001 and 2002 in the Amazon biome at Mato Grosso, Brazil.

Figure A.3 - Land cover classification in 2003 and 2004 in the Amazon biome at Mato Grosso, Brazil.

Figure A.4 - Land cover classification in 2005 and 2006 in the Amazon biome at Mato Grosso, Brazil.

Figure A.5 - Land cover classification in 2007 and 2008 in the Amazon biome at Mato Grosso, Brazil.

Figure A.6 - Land cover classification in 2009 and 2010 in the Amazon biome at Mato Grosso, Brazil.

Figure A.7 - Land cover classification in 2011 and 2012 in the Amazon biome at Mato Grosso, Brazil.

Figure A.8 - Land cover classification in 2013 and 2014 in the Amazon biome at Mato Grosso, Brazil.

PUBLICAçÃES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Programas de Computador (PDC)

São a seqência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.