Humboldt-Universität zu Berlin – Geographisches Institut

Landsat derived land surface phenology metrics for the characterization of natural vegetation in the Brazilian savanna

DISSERTATION

Zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

im Fach Geographie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M.Sc. Geogr. Marcel Schwieder

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. habil. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter:
Prof. Dr. Patrick Hostert
Prof. Dr. Rasmus Fensholt
PD Dr. hab. Angela Lausch

Eingereicht: 06. März 2018
Tag der Verteidigung: 05. Juli 2018

Acknowledgements

The past four years have been an emotional roller coaster ride on both a personal and a professional level. On the professional side, I can say for sure that the most prominent amplitude in ups- and downs were in the last few weeks. During this time days, even sometimes only hours, in which I was convinced of my work, alternated with times of self-doubt in which I was about to give up. Now, sitting here and writing these last lines of my dissertation, I am more than happy that I did not, and I am fully aware that this is only due to the help and support of some special people, which I herewith would like to thank.

First, there are my supervisors. Thank you, Patrick Hostert, for giving me the opportunity to work in your fabulous lab, for steering me now and then in the right direction, and for always having an open ear for my problems. I am looking forward to our future collaboration.

Thank you, Pedro Leitão, for taking me by the hand through the "jungle" of academia (metaphorically) and the "jungle" of the Cerrado (almost literally). We had exciting times together, in the office as well as in the field and I am grateful for everything I could learn from you.

I know that this thesis would never have happened without the support I got from the whole team of Geomatics and Biogeography Lab of Humboldt's Institute for Geography. Thank you all for being such a nice group of people that make "going to work" actually not feel like "going to work" (at least sometimes ;-)) and for all the good times we shared.

I want to thank a few people of our group that were directly involved in the progress of this dissertation. Thank you, Sebastian van der Linden, for putting your trust in me and letting me be a part of the EnMAP Team. Thank you, Andreas Rabe, for all your programming efforts, for supporting me with codes, ideas, methods and helpful tips and tricks. Thank you Christian Levers, for last-minute text editing. Special thanks to Philippe Rufin for all the times that we shared together, the inspiring meetings and talks, your support in all matters and especially for pushing me to keep going. Thanks to my office mates Leticia Hissa, Florian Gollnow and Cornelius Senf for the good times we had, for helping each other out and for the enjoyable working atmosphere.

I want to thank my friends for always sticking to my side and for your support wherever and whenever needed. I am looking forward to soon being back in "social-life-mode" and to share some quality time with you guys.

Very special thanks to my mum and my sister (with her cool family) for always keeping up the good spirit and that we stay together despite the geographical distance. Thank you for always supporting my decisions by any means.

Finally, yet importantly, there is my own little family, which happened to grow substantially during the last four years. Thank you, Julia for being a wonderful wife, for your patience, your support and for taking care of our precious kids Joon Pepe and Ona Marie, especially during the last months. Thank you little ones for the wonderful time we already had together and for showing me what life is all about.

As the cycle of life goes, we all know that we cannot stay forever. Still, the loss of a loved one is a devastating experience. Especially, if somebody goes way too early, to whom you have always looked up to and who you have admired for his positivity and his great ideas. Thank you, Dad for everything you gave, showed and taught me. This thesis is dedicated to you – Diethard Schwieder (1950 – 2014).

Abstract

The Brazilian savanna, commonly known as the Cerrado, covers around 2 Mkm², or approximately 24 % of Brazil's national area. It is characterized by a unique biodiversity, a high level of endemism and strong gradients in vegetation structure. Due to a weak conservation status and a growing demand for agricultural products, large-scale land conversions have altered almost half of the ecosystem's structure and function. Mapping and monitoring ecological processes in the Cerrado, by means of remote sensing, is a prerequisite to deepen our understanding of its ecosystem dynamics and to adapt policies and conservation strategies based on spatially explicit information. Novel remote sensing sensor systems, open data policies and technological developments regarding data storage and processing performance, facilitate the analysis of unprecedented amounts of remote sensing data. In combination with time series analysis techniques, these developments enabled capturing the seasonal dynamics of vegetation over large extents at high spatial detail. This thesis aimed to analyze the benefits of land surface phenological (LSP) metrics derived from dense Landsat time series, for the spatially explicit characterization of Cerrado vegetation, regarding its structural and phenological diversity, and to assess its relation to above ground carbon. The results revealed that LSP metrics capture the seasonal dynamics of photosynthetically active vegetation, which is amongst others influenced by structural vegetation properties and species composition. LSP metrics were shown to be beneficial for the mapping of vegetation physiognomies, but also revealed limitations of hard classification approaches for mapping vegetation gradients in complex ecosystems. Based on similarities in LSP metrics, which were derived for the whole extent of the Cerrado, LSP archetypes were proposed, which enabled to reveal the spatial patterns of LSP diversity at a 30 m spatial resolution for the first time. The LSP archetype map can enhance current mapping concepts in the Cerrado and offer potential for conservation and biodiversity assessments. Finally, LSP metrics were shown to be relevant input variables for carbon quantification approaches, as strong relationships between above ground carbon (AGC) and LSP metrics could be revealed. LSP metrics enabled the spatially explicit quantification of AGC in three study areas in the central Cerrado and should thus be considered for future carbon estimations. Overall, the insights highlight that LSP metrics derived from dense Landsat time series are beneficial for ecosystem monitoring approaches, which are crucial for sustainable land management strategies that maintain key ecosystem functions and services.

Zusammenfassung

Die Brasilianische Savanne, auch bekannt als der Cerrado, erstreckt sich über eine Fläche von 2 Mkm², oder einem Viertel der Gesamtfläche Brasiliens. Dieses Ökosystem ist von einer einzigartigen Biodiversität mit einem hohen Anteil endemischer Arten und starken Gradienten in der Vegetationsstruktur gekennzeichnet. Auf Grund des lückenhaften Naturschutzes und einer rasant steigenden Nachfrage nach landwirtschaftlichen Produkten wurde beinahe die Hälfte der natürlichen Vegetation des Cerrado in bewirtschaftetes Land umgewandelt. Diese weiträumigen Landnutzungsveränderungen beeinträchtigen unter anderem die Ökosystemfunktion und –struktur des Cerrado. Die fernerkundliche Kartierung und Überwachung ökologischer Prozesse bildet eine Voraussetzung für die Vertiefung unseres Verständnisses der zugrunde liegenden Ökosystemdynamiken und wirkt unterstützend bei der Entwicklung von Maßnahmen und Strategien im Naturschutz des Cerrado. Neuartige Entwicklungen in Sensorsystemen, freier Datenverfügbarkeit sowie Fortschritte im Bereich der Datenspeicherung und – prozessierung ermöglichen heutzutage die Analyse von großen Mengen an Fernerkundungsdaten. In Kombination mit Methoden der Zeitreihenanalyse ermöglichen diese Entwicklungen nun erstmalig die Erfassung von räumlich detaillierten saisonalen Vegetationsdynamiken über große Flächen. In dieser Dissertation wird der Mehrwert von Landsat-basierten Metriken der Landoberflächenphänologie (engl. land surface phenology; LSP) zur räumlich expliziten Charakterisierung der Cerrado Vegetation, in Bezug auf deren strukturelle und phänologische Diversität, sowie zur Schätzung des oberirdischem Kohlenstoffgehaltes, analysiert. LSP Metriken erfassen die saisonalen Dynamiken der photosynthetisch aktiven Vegetation, welche unter anderem durch die strukturelle Beschaffenheit der Vegetation, sowie ihrer Artenzusammensetzung bestimmt sind. Weiterhin sind LSP Metriken nützlich für die Kartierung von Vegetationsphysiognomien, zeigten jedoch auch die Grenzen der Einteilung von Vegetationsgradienten in diskrete Klassen. Die Gruppierung von ähnlichen Ausprägungen in LSP Metriken in sogenannte LSP Archetypen ermöglichte darüber hinaus erstmalig die Erfassung und Darstellung der phänologischen Diversität im gesamten Ökosystem des Cerrado. Die resultierende Kartierung von LSP Archetypen, mit einer räumlichen Auflösung von 30 m, bietet zahlreiche Anwendungsmöglichkeiten, wie zum Beispiel im Bereich der Naturschutzplanung und der Biodiversitätserfassung. Weiterhin wurden LSP Metriken genutzt, um eine räumlich explizite Kohlenstoffschätzung in drei Testgebieten des zentralen Cerrado durchzuführen. Der starke Zusammenhang von LSP Metriken und oberirdischem Kohlenstoffgehalt birgt große Potentiale für zukünftige Kohlenstoffschätzungen. Die Erkenntnisse dieser Dissertation unterstreichen die Vorteile und Nutzungsmöglichkeiten von LSP Metriken im Bereich der Ökosystemüberwachung und haben demnach direkte Implikationen für die Entwicklung und Bewertung nachhaltiger Landnutzungsstrategien.

Contents

List of Figures

List of Tables

List of Supplementary Information

Chapter I:
Introduction

1 Scientific background

1.1 Global environmental change

Carbon is an essential element of life. Through its ability to form a tremendous number of complex compounds with other elements, it forms the basis of all living organisms (EB 2018). During the process of photosynthesis, inorganic carbon is taken from the atmosphere and bound into organic carbon stocks as vegetation biomass, rendering the largest fluxes in the global carbon cycle and highlighting the direct link between the atmosphere and the biosphere (Schlesinger and Bernhardt 2013; Scholes and Smart 2013). During the vegetation's life cycle, carbon is released into the atmosphere and lithosphere through respiration or decomposition. Under the right conditions, vegetation biomass might eventually end up in fossil fuels such as coal, oil or natural gas (EB 2018), which have played a pivotal role in the development of modern societies since the beginning of the industrial revolution around 200 –250 years ago (Steffen et al. 2007). This marked the onset of the Anthropocene, an epoch in which human beings have tremendously altered the global environment and influenced natural cycles of material and energy flows that interlink the main spheres of the Earth system (Crutzen 2002; Steffen et al. 2007).

These alterations are comprehended under the term "global change", which encompasses all planetary scale Earth system changes that affect for example the climate, ocean, and atmospheric circulations, or the water, nitrogen, and carbon cycles (Steffen et al. 2005). A major driver of these global change processes is a growing world population, which is expected to reach around 10 billion people in 2050 (UN 2017), accompanied by an increasing demand for food and natural resources, thereby exerting high pressure on the Earth's ecosystems (Vitousek 1994; Vitousek et al. 1997).

In particular, anthropogenic land conversions, usually from areas dominated by natural vegetation such forests, grasslands or savannas to agricultural land use, have major impacts on the function of ecosystems (MEA 2005) and are one of the main drivers of biodiversity loss (Foley et al. 2005). For example, they cause changes in the surface albedo, the carbon storage capacity, and the provision of habitats. Land use change is, next to fossil fuel combustion, considered as one of the major drivers of climate change (IPCC 2013) and has a strong influence on biodiversity patterns (Haines-Young 2009). Anthropogenic alterations of the land surface began with the Neolithic revolution around 10.000 years ago when humans settled and started to use arable land for their subsistence (Ramankutty et al. 2006).

To date, estimates suggest that humans in one way or another have altered around 50% of the terrestrial surface (Steffen et al. 2005). Land change science has thus become an essential part in global environmental change research (Turner et al. 2007).

Land use change processes have influences on a global scale, but they are not equally distributed around the globe. Especially tropical regions have undergone tremendous land conversions in recent decades with the deforestation of tropical rainforests for agricultural land use, being one of the most prominent examples (Fearnside 2002; Melillo et al. 1996; Morton et al. 2006; Nobre et al. 2002). Research focus on anthropogenic impacts on tropical forests have led to public awareness concerning the destruction of these pristine ecosystems, and ultimately to a range of national and international policies aiming to counteract these developments. Deforestation and land use changes, in general, are a part of broader international policies and treaties such as the Kyoto protocol (UNFCCC 1997), developed under the United Nations Framework Convention on Climate Change (UNFCCC), or the Sustainable Development Goals (SDG; Biermann et al. 2017; Griggs et al. 2013). Both were agreed upon to set a framework in which the challenges that our societies face can be treated along trackable milestones. Concepts like *Reducing Emissions from Deforestation and Forest Degradation* (REDD+; Romijn et al. 2012; Tokola 2015) or the Aichi biodiversity targets (CBD 2010) were developed to set financial incentives to promote the conservation of forests and their carbon storage capabilities or to raise awareness and counteract the loss of biodiversity and species extinction. These international frameworks are intended as guidelines for national policies and incentives, for which one prominent example is the strategy of Brazil to halt deforestation in the Amazon.

The Amazon has become a hotspot of deforestation since the 1970s, which was facilitated through the construction of the Trans-Amazon Highway that opened up the Amazon forest to clear land for agricultural production (Fearnside 2005). During peak times deforestation rates were reported to approach 30,000 km² per year (Arima et al. 2014) and Brazil's tropical rainforest was rapidly diminishing. This was when current national policies, such as the Forest Code, were adapted to counter the recent deforestation trends and defined that 80 % of property land had to be left unchanged as forest reserves. Concrete policies like the Plan to Prevent and Control Deforestation in the Amazon (PPCDAm-I in 2004 (Abdala 2008) and PPCDAM-II in 2008 (Maia et al. 2011)) were set in place, along with deforestation monitoring systems like DETER (INPE 2018a) and PRODES (INPE 2018b) and an extensive protected area system, to influence deforestation processes in the Amazon (Arima et al. 2014).

On the initiative of non-governmental organizations, agreements such as the soy and beef moratorium in 2006 (Gibbs et al. 2015) and 2009 (Tollefson 2015), were made to directly influence the supply chain of agricultural products, with the intention to reduce the carbon footprint of agricultural products. Even though the influence of global agricultural markets on deforestation trends cannot be neglected, there is consent that the implementations of these policies have had a positive impact on the 70 % decrease in deforestation in the Amazon from 2005 to 2013 (Arima et al. 2014; Assunção et al. 2015; Nepstad et al. 2014). However, even if some of these policies apply to the national scale, they have a strong focus on the Amazon biome and are less relevant for other Brazilian biomes. On the contrary, these policies are even thought to cause leakage effects (Arima et al. 2014) and neglect a growing pressure on other important ecosystems such as the Brazilian savanna, also known as the Cerrado (Marris 2005).

1.2 The Cerrado

The Cerrado is the second largest biome of Brazil; it stretches from the equator to 23° south. With an extent of approximately 2 million km², it accounts for roughly 23 % of Brazil's land surface (Figure I-1). It is considered the most biodiverse savanna ecosystem globally (Silva and Bates 2002), with more than 10,000 plant species (Mendonça et al. 2008), of which about half are endemic and make up around 1.5 % of all global plant species (Myers et al. 2000). The Cerrado is mainly located on the Central Brazilian Plateau and within its extent lie four of Brazil's main watersheds (Lima and Silva 2008) as well as the headwaters of several major South American rivers such as the Araguaia, the Tocantins, and the Xingu. Well-drained dystrophic soils with high contents of iron and aluminum dominate the Cerrado (Motta et al. 2002; Reatto et al. 2008). Average annual precipitation and temperature vary across the Cerrado, from 800 mm to 2000 mm and 18 °C to 28 °C, but the climate is in general characterized by a strong seasonality with a wet season from October to April and a dry season from May to September (Oliveira-Filho and Ratter 2002). Alternating dry and wet seasons are the main influence on the phenology of the Cerrado vegetation and this also favors the occurrence of fires (Miranda et al. 2009) that are a common feature in savanna ecosystems (Furley 2010; Miranda et al. 2002).

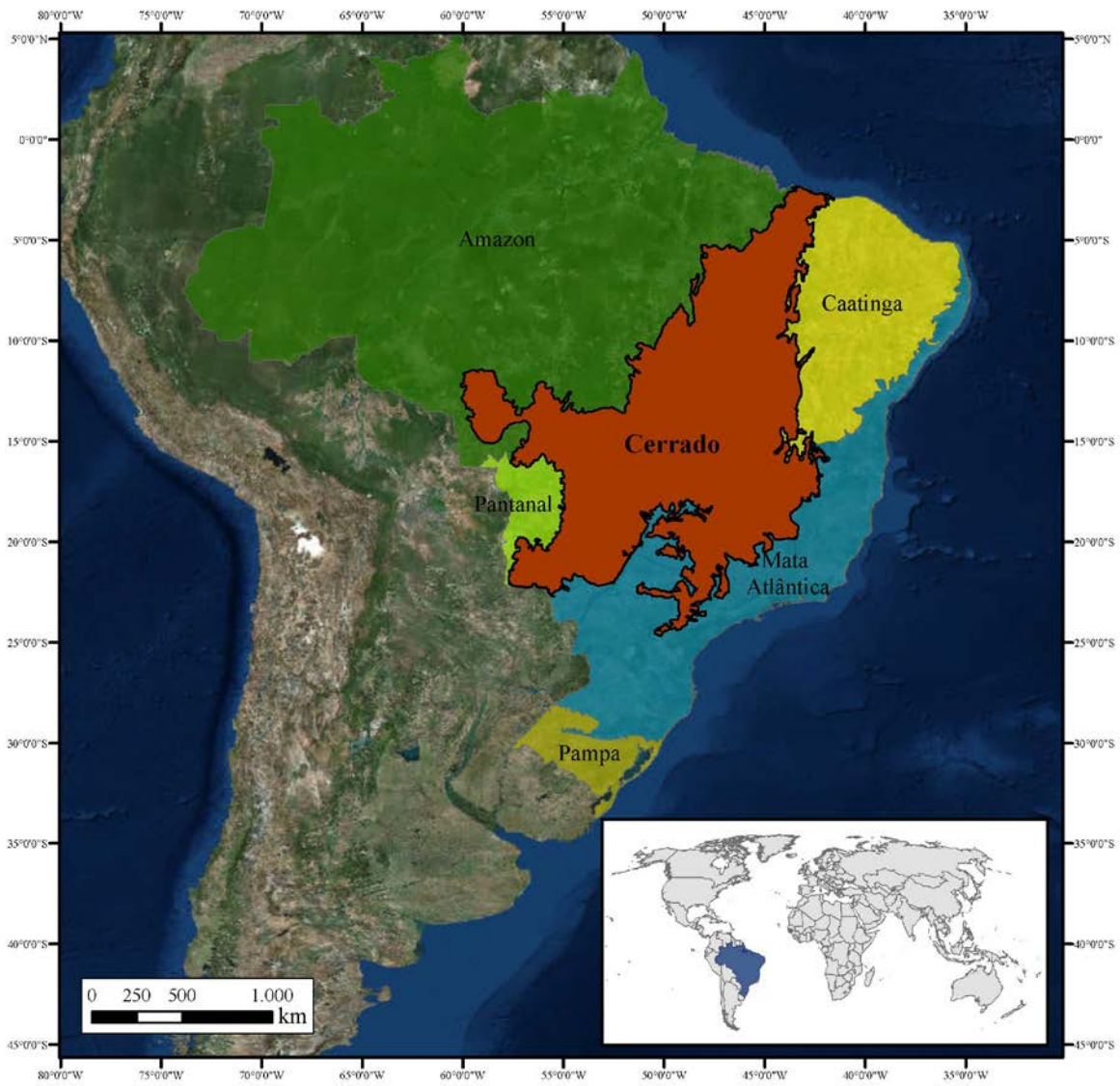


Figure I-1: Brazil's national extent and the individual Brazilian biomes. The colors do not reflect similarities between the biomes, but were selected to enhance differentiation.

The described environmental factors influence the distribution of vegetation in the Cerrado, resulting in a characteristic landscape of vegetation physiognomies (Oliveira-Filho and Ratter 2002) that range from open grasslands, over shrub or tree dominated savanna formations with a grass understory, up to dense forests. Based on structural parameters, such as height and density, the Cerrado vegetation is commonly distinguished in physiognomy classes, with the most prominent ones being *campo limpo* (grasslands), *campo sujo* (grasslands with shrubs), *cerrado sensu stricto* (tree dominated savanna), *cerrado denso* (savanna with dense tree formations) and *cerradão* (dense and high trees) (Ribeiro and Walter 2008). These physiognomies describe the vegetation gradient in the Cerrado and thus

differ in their amounts of biomass and sequestered carbon (Vourlitis and da Rocha 2010), which renders them a useful proxy for carbon reporting in the context of REDD+ (MMA 2017). The Cerrado provides a range of ecosystem services (Lima et al. 2017), which are defined as "benefits people obtain from ecosystems" (MEA 2005). Due to the large extent of the Cerrado, ecosystem services and functions, such as climate regulation, carbon sequestration, or the provision of habitats, are of national and even international importance.

Despite its ecological and societal value, the Cerrado has a very weak conservation status with only 2.2 % of the biome's extent being considered in Brazils protected area network. Caused by its underrepresentation in national policies and conservation strategies, especially in comparison to the Amazon biome (Klink and Machado 2005). In contrast, as a response to a growing demand for agricultural products, governmental programs have aimed at the economic development of rural areas of Brazil. The 'creation' of Brazil's new capital Brasília in the central Cerrado in 1960, was accompanied by the development of transportation systems and infrastructure, opening the Cerrado for settlement and other types of anthropogenic land use (Klink and Moreira 2002). Incentives for new farming activities were set by low-interest loans and advances in agricultural mechanization, while fertilizers enabled cultivation on the nutrient-poor soils of the Cerrado (Klink and Moreira 2002). The combination of these factors led to large-scale land conversion, with land clearing rates that exceeded those of the Amazon (Klink and Machado 2005) resulting in approximately 60 % of remaining natural vegetation (Sano et al. 2010). A growing international demand for agricultural products is considered a direct driver of land use change in Brazil (Lapola et al. 2013), and will have a direct effect on land use, which needs to be steered by appropriate policies to circumvent unsustainable developments (Strassburg et al. 2017).

Developing adequate conservation strategies require assessments of the ecological state of the Cerrado, for example in terms of carbon storage and biodiversity. For this purpose, and to gain a better understanding of large-scale ecological processes in general, it is necessary to map and monitor land surface properties of the Cerrado. Therefore, spatially explicit monitoring approaches that enable to frequently gather data over large extents are a prerequisite. Such monitoring efforts, especially in the heterogeneous landscape of the Cerrado are reliant on data at a sufficient spatial resolution, a temporal resolution that allows for capturing intra-annual dynamics, and a data archive that covers long time frames, making it possible to understand long-term processes.

1.3 Optical remote sensing for terrestrial ecosystem monitoring

The analysis of remote sensing data has become an indispensable tool in a wide range of ecosystem monitoring applications (DeFries et al. 2005). Their capability to frequently provide data over large and even inaccessible extents, enable to gather information about the observed land surface that might reveal changes which are not captured in analysis on a local scale (Smith et al. 2014).

Products derived from remote sensing data can be divided into two groups; namely categorical maps representing the observed surface in discrete thematic classes, for example in land use or land cover classifications, or continuous representations, such as quantities of biophysical or biochemical surface properties, for instance, vegetation carbon stocks. The outcome of remote sensing data analysis is reliant on the sensor characteristics that define the resolution of the data, the platform the sensor operates from (i.e. terrestrial, airborne or spaceborne) and on the availability of reference data. While terrestrial and airborne sensors restrict the analysis to smaller extents, they are usually able to capture land surface properties with high spatial and spectral detail. For example airborne hyperspectral sensors, such as the Hyperspectral Mapper (HyMap; Cocks et al. 1998) or the Advanced Visible Near-Infrared Imaging Spectrometer (AVIRIS; Green et al. 1998), cover the visible (VIS), near-infrared (NIR) and shortwave-infrared (SWIR) regions of the electromagnetic spectrum in hundreds of contiguous bands. Hyperspectral data thus enable to derive valuable information, for example for invasive species mapping approaches (He et al. 2011) or biodiversity assessments (Ghiyamat and Shafri 2010). However, until the advent of forthcoming spaceborne hyperspectral missions, such as EnMAP (Guanter et al. 2015) or HyspIRI (Lee et al. 2015) these approaches are restricted to rather local and regional analyses.

Spaceborne sensors cover large extents of the Earth at the same time but are restricted by the bottleneck of data downstream from space (Iwasaki and Tadono 2014). This leads to trade-offs between the sensor's spatial resolution (in terms of pixel size), the spectral resolution (number of bands) and the temporal resolution (revisit time), which restricts their individual use to specific applications.

Sensors with a high temporal resolution such as AVHRR (Advanced Very High Resolution Radiometer) or MODIS (Moderate-resolution Imaging Spectroradiometer), are for example used to capture the seasonal dynamics of ecosystems (Ehrlich et al. 1994; Zhang et al. 2003). The first sensor of the AVHRR family was launched in 1978 and is thus one of the pioneer Earth observation sensors. The nearly continuous data archive covers the last 40 years, which enables to look back in time and to analyze long-term trends. The two complementary

MODIS sensors on board of the Terra and Aqua satellites started to acquire data from the early 2000's, thus MODIS data are available for the last two decades. Even though AVHRR and MODIS spectral characteristics are not primarily designed to capture vegetation properties, both cover relevant regions of the electromagnetic spectrum. While AVHRR is sensitive in two bands of the visible red and NIR regions, MODIS covers the VIS, NIR, and SWIR in seven spectral bands. However, the outstanding characteristics of both sensors are their frequent revisit times. AVHRR acquires data of the same region twice a day, while the combined use of the two MODIS sensors enables to cover the whole globe every 1 - 2 days, thereby increasing the chances to obtain clear observations even in cloud prone areas. This high rate of data acquisition over large extents (AVHRR and MODIS acquire data with a swath of approximately 2400 km) comes at the cost of spatial detail. MODIS spatial resolution ranges between 250 m and 1000 m across different spectral bands, while AVHRR data have a spatial resolution of around 1.1 km. These rather coarse scale resolutions enable to frequently capture a synoptic view but are insufficient for the analysis of complex and heterogeneous landscapes such as savanna regions.

Multispectral data with a very high spatial resolution as for example acquired by WorldView 1 - 4 sensors (~ 0.3 - 4 m spatial resolution; SIC 2018) are able to overcome this limitation. However, these data are in contrast to AVHRR and MODIS not freely accessible, as they are distributed by commercial data providers. As they are usually acquired upon request, they are a valuable source for small-scale studies and validation efforts for which the spatial detail is essential but do not allow to assess seasonal dynamics over large extents or long time periods.

Data with a medium spatial resolution provide great opportunities to reconcile the trade-offs associated with coarse resolution, but also very high resolution data. Such data are acquired by sensors of Landsat family, which started to collect data with the Multispectral Scanner (MSS) on board of Landsat 1 in 1972. Sometimes referred to as the pioneering sensor of satellite-based optical Earth observation, Landsat 1 enabled for the first time to frequently acquire data of the Earth's surface over large extents (scene size of 170 x 185 km), with a spatial resolution of 80 m on ground (Cohen and Goward 2004). Landsat MSS was the first Earth observation mission that had a global image acquisition strategy and was, until today, followed by 7 successors, which were in the course of time further developed along the current technological state of the art (Markham and Helder 2012). From the beginning, the Landsat sensors covered spectral regions of the electromagnetic spectrum that are important for environmental mapping purposes. While the MSS sensors on board of Landsat 1 and 2

covered the VIS (0.5 – 0.7 µm) and NIR (0.7 – 1.1 µm) spectral region in four bands (with a bandwidth of approximately 0.1 µm), Landsat 3 was additionally equipped with a thermal band being sensitive in the wavelength 10.4 to 12.6 µm. Landsat 4 was launched in 1982 and was the first mission of the Landsat family that had the so-called "Thematic Mapper" (TM) sensor on board. It covered the VIS and NIR spectral regions with five bands and had, next to the thermal band, an additional band in the SWIR (2.08 - 2.35 µm) spectral region. An additional novelty of TM was the improved spatial resolution of 30 m by 30 m in the VIS, NIR, and SWIR, while the spatial resolution of the thermal band was reduced to 120 m. Landsat 5 had similar sensor characteristics as Landsat 4 and had an extraordinary live time of more than 27 years in which it continuously acquired data. Retrospectively, Landsat 5's endurance was a lucky circumstance as its successor Landsat 6 (launched in 1993) never reached its orbit (Cohen and Goward 2004). It was not until 1999 when Landsat 7 was launched with the "Enhanced Thematic Mapper" sensor on board, which additionally gathers data in a panchromatic band with 15 m spatial resolution (Markham and Helder 2012). Besides this novelty, it is equipped with a solar calibration module, which enables an on-board calibration of the sensor. Landsat 7 acquires data until today. Due to major advances in technology, Landsat 8 (launched in 2013) is equipped with a new sensor system the so-called "Operational Land Imager" (OLI). For the reason of continuity, OLI covers the VIS, NIR, SWIR1 (1.57 - 1.65 µm) and SWIR2 (2.11 - 2.29 µm) with seven spectral bands and a spatial resolution of 30 m and has, similar to Landsat 7, a panchromatic band with 15 m spatial resolution. For atmospheric correction, cloud identification and the assessment of water quality, additional bands, such as the aerosol (0.43 - 0.45 µm) and the cirrus band (1.36 - 1.38 µm) were added. Landsat 8 has an independent "Thermal Infrared Sensor" (TIRS) with two bands at 100 m spatial resolution (Roy et al. 2014).

The continuous data acquisition since the launch of Landsat 1 and similar spectral and spatial characteristics of the Landsat sensors have led to a still growing archive of optical Earth observation data that spans today more than 45 years. Even though Landsat data have been widely used for a range of ecological applications and greatly improved our understanding of the state and dynamics of Earth's vegetation (Cohen and Goward 2004), they were through the cost of data acquisition and technical limitations restricted to the analysis of individual images by a specialized research community. However, major changes in data policies in 2008 made all data ever acquired by the Landsat sensors freely accessible (Woodcock et al. 2008). These policies enabled along with advances in pre-processing chains (Schmidt et al. 2013) and cloud detection algorithms (Zhu et al. 2015a; Zhu and Woodcock

2012), the consolidation of data archives (Wulder et al. 2015b) and latest data harmonization approaches (USGS 2018), new opportunities for Landsat data analysis (Wulder et al. 2012). Despite all the unique position features of the Landsat family, one drawback is its temporal resolution with a revisit time of approximately 16 days. Depending on the geographical region of interest and the related cloudiness, frequent image acquisition can be substantially hampered.

The availability of unprecedented amounts of data facilitated to develop approaches that make use of all available Landsat observations, to overcome this limitation. For instance, best pixel composites that provide seamless, cloud free, radiometrically and seasonally consistent images that enable, for example, the analysis of land use change processes over large extents with a high spatial detail (Griffiths et al. 2013; Roy et al. 2010). These approaches require calibration concerning the target day of year, which can be challenging for the analysis of heterogeneous ecosystem over large extents, where seasonal dynamics substantially vary. Recent advances take these spatial variations in seasonal dynamics into account, through enabling to pixel wise calibration of the composites based on phenological information (Frantz et al. 2017), which is, however, not commonly available at a sufficient spatial resolution over large extents. These approaches have to date not extensively been used to derive seasonal information that captures intra-annual dynamics. Spectral temporal metrics that take advantage of the temporal depth of clear Landsat observations throughout a season enable to detect more subtle land surface changes, like agricultural intensification processes (Rufin et al. 2015) or forest regrowth patterns (Müller et al. 2016). These approaches do not directly derive phenological information but rather reflect the annual variations of Landsat spectra in statistical measures and have found to be highly dependent on data availability and distribution (Müller et al. 2015; Rufin et al. 2015). Novel approaches like data pooling, in which vegetation indices of several years are combined, have been shown to be sufficient to asses (long- and short-term) land surface phenological variations on a Landsat spatial resolution for temperate and boreal forest (Melaas et al. 2013; Melaas et al. 2016). Approaches that aim to fill data gaps in Landsat time series, which are due to cloud contamination or sensor errors, allow to derive land surface phenological information that capture intra-annual dynamics and are currently an active field of research (Vuolo et al. 2017; Zhu et al. 2015b). Hence, dense time series of equidistant Landsat data are thought to fulfill the demands that are critical for the spatially explicit characterization of vegetation in heterogeneous ecosystems, such as the Cerrado.

2 Conceptual framework

2.1 Motivation

Pressure on natural ecosystems through human-induced global change processes put a range of provisioning and regulating ecosystem services at risk (MEA 2005). An ecologically relevant example of anthropogenic alterations, which has received comparatively little attention so far, is the Cerrado. A growing demand for land resources and a weak conservation status have led to tremendous land conversions in the Cerrado (Lapola et al. 2013). It is estimated that only around 60 % of its original natural vegetation still exists (Sano et al. 2010). These land use change processes alter, for example, the ecosystem's capability of carbon storage and to release sequestered carbon, directly influencing the global carbon cycle. Further, large-scale land conversions and land use intensification processes have led to a fragmentation of the Cerrado landscape with impacts on its biodiversity. These developments call for adjusted conservation strategies for which large-scale ecosystem assessments are critical.

Robust and frequent mapping approaches by means of remote sensing are thus of utmost importance for a better understanding of ongoing processes and to provide a synoptic basis to inform decision making in conservation policies. With respect to the ecosystem's seasonal dynamics and its heterogeneous landscape characteristic, it is crucial to use data sets with sufficient spatial and temporal resolutions. Facilitated by the emergence of new operational satellite sensor systems along with open data policies (Woodcock et al. 2008) and optimized data homogenization (USGS 2017b), unprecedented amounts of data are now accessible that are of immense value for accurate mapping approaches. This development followed a rapid increase in computational performance and the availability of advanced machine learning algorithms that enabled new ways to analyze the available large remote sensing datasets (Lary et al. 2018). Thus, this thesis aims to further exploit the potential of optical remote sensing data for the mapping and quantification of heterogeneous ecosystem properties in terms of vegetation structure, phenological diversity, and the distribution of above ground carbon.

2.2 Research questions

The overarching goal of this thesis is to analyze the benefits of Landsat based land surface phenological metrics for mapping and quantification applications in heterogeneous ecosystems. In particular, the aims are to provide insights into the spatial patterns of the remaining natural vegetation in the Cerrado and its seasonal behavior. Further, the

relationship between land surface phenological metrics and the distribution of above ground carbon was analyzed in order to improve carbon quantification approaches. The methodological analyses were based on dense 8-day Landsat time series of combined ETM+ and OLI products, in which data gaps were filled using an ensemble of radial basis convolution filters. The following three research questions define the framework of this thesis:

Research Question I: Can land surface phenological metrics be used to differentiate the main vegetation physiognomies in the Cerrado?

The main vegetation physiognomies of the Cerrado describe a gradient in vegetation structure and density. They range from open grasslands to shrublands and scattered tree formations, up to dense forests. Various environmental factors influence their spatial distribution and the species composition. This leads to the hypothesis that these differences can also be observed in the vegetation physiognomies' seasonal response as captured in the spectral signal of remotely sensed vegetation. This research question is the core of chapter II, in which the methodological framework for deriving LSP metrics from the Landsat time series is introduced.

Based on reference data for a study area around Brasília, DF the main objectives are:

i. to analyze which differences in LSP can be revealed between the main Cerrado physiognomies;

ii. and to test if these differences are sufficient to map the spatial distribution of the major Cerrado physiognomies.

Research Question II: What are the spatial patterns of LSP metrics in the natural vegetation across the Cerrado?

Knowledge about the phenology of the natural Cerrado vegetation is still fragmented and there is, to date, no sufficient coverage of phenological information for the whole extent of the Cerrado — this study aims to fill this gap. Chapter III focuses on the Landsat collection 1 enhanced vegetation index LSP metrics that were derived for the Cerrado's entire extent of more than 2 million km² for the year 2014.

The objectives are:

i. to classify the remaining natural vegetation of the Cerrado into groups of similar LSP traits;

ii. to analyze their distribution in the context of defined Cerrado ecoregions.

Research Question III: What is the relationship between Landsat based land surface phenology metrics and above ground carbon?

The remaining natural vegetation in the Cerrado has large amounts of carbon stored in its above- and belowground compartments. Due to the huge extent of the Cerrado, these stocks play a pivotal role in the regional and global carbon cycle and it is, therefore, essential to improve carbon quantification approaches. In chapter IV a random forest regression approach is used, in combination with field data collected in three study areas distributed across the central Cerrado, to analyze the importance of the individual LSP metrics for carbon quantification.

The objectives are:

i. to investigate the potential to model aboveground carbon in a heterogeneous ecosystem based on Landsat derived LSP metrics;

ii. to assess the relationship between these LSP metrics and aboveground carbon;

iii. to use these metrics to map the carbon distribution across different Cerrado landscapes.

2.3 Thesis structure

To frame the findings presented in this thesis within a broader context, an introductory chapter (I) gives an overview of the scientific background and motivation. The research gaps this thesis aims to fill are indicated and guiding research questions are clearly stated. The main body of this thesis is structured in three core chapters (II-IV), which address the defined research questions. Chapter II has been published in a peer-reviewed scientific journal and chapter IV was submitted for publication and is currently under review. Chapter III is currently being prepared for submission. In order to fulfill the journal's requirements concerning the style of publications, redundant information (e.g. in the introductions and methods sections of the individual chapters) could not be avoided. Finally, in chapter V the main findings of this thesis are summarized, the scientific value of the three core chapters as a whole is synthesized, and the implications, as well as limitations, are pointed out. An

additional chapter that describes in detail the spatial allocation of field data to the pixel level, which was a necessity for the analysis of chapter IV, is attached in Appendix A. The following section gives a brief overview of the three core chapters.

Chapter II:	***Marcel Schwieder**, Pedro J. Leitão, Mercedes M. C. Bustamante, Laerte Guimarães Ferreira, Andreas Rabe and Patrick Hostert (2016). Mapping Brazilian savanna vegetation gradients with Landsat time series. International Journal of Applied Earth Observation and Geoinformation, Volume 52, Pages 361–370.*
Chapter III:	***Marcel Schwieder**. Land surface phenological archetypes of the Cerrado.*
Chapter IV:	***Marcel Schwieder**, Pedro J. Leitão, José Roberto R. Pinto Pinto, Ana Magalhaes C. Teixeira, Fernando Pedroni, Maryland Sanchez, Mercedes M. C. Bustamante and Patrick Hostert (in review)[*]. Carbon Balance and Management.*
Appendix A:	*Pedro J. Leitão, **Marcel Schwieder,** Florian Pötzschner, José Roberto R Pinto, Ana Magalhaes C. Teixeira, Fernando Pedroni, Maryland Sanchez, Christian Rogaß, Sebastian van der Linden. Mercedes M. C. Bustamante and Patrick Hostert (in review)[**]. From sample to pixel: using multi-scale remote sensing data for the upscaling of field collected aboveground carbon data in heterogeneous landscapes. Ecosphere.*

[*] A revised version of the manuscript was meanwhile published in: Schwieder, M., Leitão, P.J., Pinto, J.R.R., Teixeira, A.M.C., Pedroni, F., Sanchez, M., Bustamante, M.M., & Hostert, P. (2018). Landsat phenological metrics and their relation to aboveground carbon in the Brazilian Savanna. Carbon Balance and Management, 13, 7.

[**] A revised version of the manuscript was meanwhile accepted for publication in: Leitão, P.J., Schwieder, M., Pötzschner, F., Pinto, J.R.R., Teixeira, A., Pedroni, F., Rogass, C., Sanchez, M., van der Linden, S., da Cunha Bustamante, M.M., & Hostert, P. (in press). From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. Ecosphere.

Chapter II:
Mapping Brazilian savanna vegetation gradients with Landsat time series

International Journal of Applied Earth Observation and Geoinformation, 2016, Volume 52, Pages 361–370

Marcel Schwieder, Pedro J. Leitão, Mercedes Maria da Cunha Bustamante, Laerte Guimarães Ferreira, Andreas Rabe and Patrick Hostert

DOI: http://dx.doi.org/10.1016/j.jag.2016.06.019
Received 17 February 2016; Revised 15 June 2016; Accepted 21 June 2016

Abstract

Global change has tremendous impacts on savanna systems around the world. Processes related to climate change or agricultural expansion threaten the ecosystem's state, function and the services it provides. A prominent example is the Brazilian Cerrado that has an extent of around 2 million km² and features high biodiversity with many endemic species. It is characterized by landscape patterns from open grasslands to dense forests, defining a heterogeneous gradient in vegetation structure throughout the biome. While it is undisputed that the Cerrado provides a multitude of valuable ecosystem services, it is exposed to changes, e.g. through large scale land conversions or climatic changes. Monitoring of the Cerrado is thus urgently needed to assess the state of the system as well as to analyze and further understand ecosystem responses and adaptations to ongoing changes. Therefore we explored the potential of dense Landsat time series to derive phenological information for mapping vegetation gradients in the Cerrado. Frequent data gaps, e.g. due to cloud contamination, impose a serious challenge for such time series analyses. We synthetically filled data gaps based on Radial Basis Function convolution filters to derive continuous pixel-wise temporal profiles capable of representing Land Surface Phenology (LSP). Derived phenological parameters revealed differences in the seasonal cycle between the main Cerrado physiognomies and could thus be used to calibrate a Support Vector Classification model to map their spatial distribution. Our results show that it is possible to map the main spatial patterns of the observed physiognomies based on their phenological differences, whereat inaccuracies occurred especially between similar classes and data-scarce areas. The outcome emphasizes the need for remote sensing based time series analyses at fine scales. Mapping heterogeneous ecosystems such as savannas requires spatial detail, as well as the ability to derive important phenological parameters for monitoring habitats or ecosystem responses to climate change. The open Landsat and Sentinel-2 archives provide the satellite data needed for improved analyses of savanna ecosystems globally.

1 Introduction

Savanna ecosystems cover an estimated 20% of the global terrestrial surface (Lehmann et al. 2011), providing essential ecosystem goods and services such as food, pollinators and carbon storage (Marchant 2010). They occur in tropical and sub-tropical climate zones across all continents, but are particularly prevalent in Australia, Africa and the Americas (Solbrig 1996). Global land use change processes related to a growing demand for natural resources led to a conversion of approximately 50% of the global savannas with a direct impact on biodiversity and carbon storage (Foley et al. 2011). Climate change, invasive species and fertilizer pollution rapidly impact savanna biodiversity (MEA 2005), while global change is likely to alter the global distribution of savannas and might even lead to a change of biome states (Staver et al. 2011).

The Brazilian savanna, better known as Cerrado, is a prominent example of global change impacts on savanna ecosystems. It is the second largest eco-region in Brazil with an extent of about 2 million km² (Ratter et al. 1997). Landscape formations that range from open grass and shrub dominated lands to dense forests are characteristic (Oliveira-Filho and Ratter 2002) and form a gradient of vegetation structure and biomass, which is often differentiated into discrete structural physiognomy classes (Ribeiro and Walter 2008). The climate of the Cerrado is characterized by a wet and a dry season, which influence the vegetation's spatial and temporal dynamics. It`s harsh environmental conditions led to a high floristic diversity and a variety of phenological adaptation strategies (Ferreira and Huete 2004). The Cerrado is thus considered as the biodiversity-richest savanna globally (Silva and Bates 2002) with approximately 160,000 species of fungi, flora and fauna (Furley 1999). However, a weak land conservation status has led to large-scale conversions from natural to agricultural land that already affected more than 40% of the Cerrado, which is likely to aggravate in the future (Ferreira et al. 2012; Sano et al. 2010). This conversion of primary vegetation threatens the stability of the ecosystem and related services provided, such as carbon sequestration and climate regulation. It further impacts its biodiversity, rendering the Cerrado as an under-researched global biodiversity hotspot in need of in-depth monitoring as basis for profound conservation planning (Myers et al. 2000).Thus, accurate mapping and monitoring of the temporal and spatial dynamics of Cerrado vegetation is essential to understand ecosystem properties and responses to ongoing change processes to support decision makers (Rocha et al. 2011; Sano et al. 2010).

Field based mapping and ecological assessments that rely on established classification schemes are indispensable for a detailed analysis of local processes, but at the same time

costly and thus restricted to relatively small areas. Remote sensing data, on the other hand, have successfully been used to map the Cerrado across large areas and in inaccessible terrain. Sano et al. (2010) created a land cover map of the entire Cerrado based on a mosaic of 170 Landsat scenes from 2002. They differentiated anthropogenic and natural (grass-, shrub- and forestlands) land cover classes by image segmentation and visual interpretation with an overall accuracy of 71%. Spatially less extended studies investigated the usability of combined radar and optical data (Sano et al. 2005) or spectral unmixing of Landsat data (Ferreira et al. 2007) to discriminate Cerrado vegetation physiognomies (structural classes). Other studies have shown that analyzing multi-temporal imagery is advantageous over single-date images for the distinction of spectrally similar vegetation types (e.g. Mesquita Junior 2000; Müller et al. 2015).

Using high temporal resolution satellite data time series not only allows to discriminate vegetation types but also to describe different phenological vegetation phases throughout a season (Zhang et al. 2003). Ferreira and Huete (2004) assessed the seasonal dynamics of the Cerrado vegetation using time series of AVHRR vegetation indices. In spite of the drawbacks inherent to the data (e.g. coarse spatial resolution, sensor uncertainties and broad bandwidths) they were able to derive phenological patterns capable of depicting the seasonal cycle and which allowed to distinguish between savanna formations, pastures, croplands and forests. Ratana et al. (2005) explored the potential of MODIS 16-day composite time series to analyze phenological patterns of different Cerrado physiognomies, revealing their distinct responses to seasonal contrasts with a 250 m spatial resolution. Even though both studies provided valuable insights on phenological differences between Cerrado vegetation formations, they did not aim at mapping their spatial distribution. As the analysis of phenology derived from remote sensing data is based on measures of temporal changes in surface reflectance at the pixel scale (Hanes et al. 2014), it usually relates to a signal mixture of different canopy or understory layers and depicts rather a vegetation community instead of a single species' phenology. The sensor's spatial resolution is therefore critical for the potential detail of the derived information, which is particularly important for analyzing and mapping the complex vegetation gradients in the Brazilian Cerrado. The analysis of Landsat data with a spatial resolution of 30 m is promising, but has mostly been restricted to multi-temporal imagery with low temporal resolution, which is not sufficient to derive continuous phenological information. However, the opening of the extensive data holdings of the Landsat archive (Wulder et al. 2012) allows to combine data from the Landsat Thematic Mapper (TM) with Enhanced Thematic Mapper (ETM+) and Operational Land Imager

(OLI) sensors. This improves the temporal resolution to potentially eight days, rendering Landsat a system with capabilities for detailed phenological information retrieval and offers a great opportunity to observe vegetation gradients over long time periods with sufficient spatial resolution.

In order to deepen the knowledge on Cerrado vegetation, we aim to derive spatially explicit phenological information, using state of the art remote sensing techniques. We focus on the applicability and limitations of the combined use of Landsat time series and an established physiognomy classification scheme (Ribeiro and Walter 2008), which enables the expansion of field based ecological assessments to broader scale studies. Therefore we evaluate how i) phenological parameters can be derived at a 30 m spatial resolution, ii) which phenological differences between the main Cerrado vegetation physiognomies can be revealed based on these parameters and iii) if the derived parameters are beneficial for mapping these physiognomies' spatial distribution.

2 Data and methods

2.1 Study sites

The Brazilian Cerrado has a unique appearance, characterized by a mixture of xeromorphic vegetation formations. Variations in abiotic environmental factors such as soil traits and fertility, long term climate fluctuations, and fire events have shaped the landscapes of the Cerrado. The vegetation is influenced by strong climatic seasonality with a dry season between May and September in which most of the Cerrado plant species flower, renew their leaves and spread their seeds, before they germinate at the beginning of the wet season (de Faria et al. 2012), which lasts from October to April. However, many of the woody plant species in the Cerrado are evergreen with less seasonal variations that develop new leaves at the end of the dry season and reach their peak in the wet season (Ratana et al. 2005). These environmental conditions result in landscape patterns composed of open grass, shrub dominated lands and dense forests, which define a gradient of vegetation density, growth form, vertical structure and biomass (Oliveira-Filho and Ratter 2002; Ribeiro and Walter 2008). Several authors have attempted to define a range of classification schemes that enable to describe the vegetation gradient in distinct physiognomy classes (Oliveira-Filho and Ratter 2002). In this study we follow the classification scheme proposed by Ribeiro and Walter (2008), which distinguishes the physiognomies based on the vegetation height and density (Figure II-1).

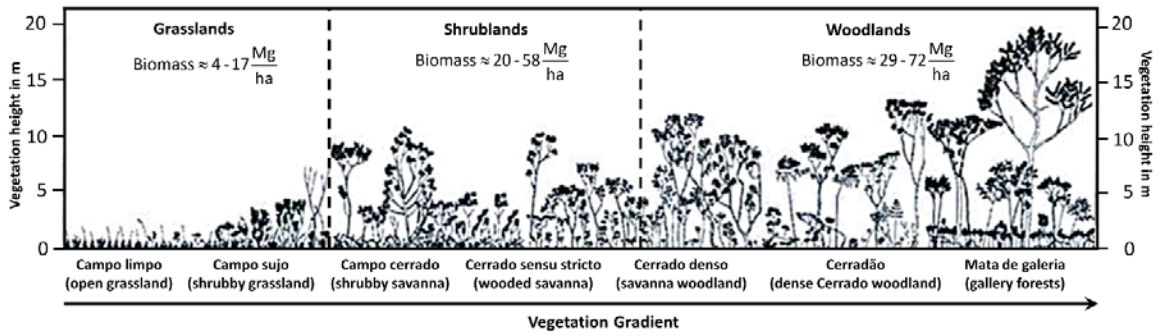


Figure II-1: Cerrado physiognomies describing gradients of vegetation height and density (adapted from Mesquita Junior (2000)). Hereafter only the Portuguese physiognomy names will be used. The above ground biomass values of woody vegetation were comprehended from several studies by *de Miranda et al. (2014)*.

Our analysis focused on the classification of the main Cerrado physiognomies, i.e. *campo limpo* (open grassland), *campo sujo* (grassland with small shrub patches), *campo cerrado* (shrublands with some trees), *cerrado sensu stricto* (wooded savanna with shrubs and trees), *cerrado denso* (woodland with dense shrubs and trees), *cerradão* (dense woodland with high trees) and *mata de galeria* (gallery forests with dense vegetation and high trees along rivers and waterbodies).

Three study sites within one Landsat footprint were analyzed that are located in the central Cerrado around Brasília, Federal District (Figure II-2). All sites lie within protected areas, thus we could ensure that the anthropogenic interference on the land cover is negligible. The first site covers the whole extent (approx. 29,000 ha) of the Brasília National Park (PNB), which encompasses the major savanna formations encountered in the Cerrado, including the transitions from the dominant herbaceous stratum (*campo cerrado*) to the more complex, woody dominated stratum (*cerrado sensu stricto*). The second study site is the Gama Cabeça-de-Veado Environmental Protection Area (APAGCV), which covers around 20,000 ha of the Cerrado south of Brasília. Even though parts of APAGCV's natural vegetation have already been converted, it contains undisturbed areas in the Brazilian Institute of Geography and Statistics (IBGE) Ecological Reserve, Brasília's Botanical Garden and an experimental farm that belongs to the University of Brasília. These areas are characterized by the main Cerrado grassland, shrubland and woodland physiognomies including patches of *cerrado denso* and *cerradão*. The third study site is the Águas Emendadas Ecological Station (ESECAE), a protected area northeast of Brasília with an extent of approximately 8,000 ha. Next to open Cerrado types (mostly shrubland) and denser forest formations, large areas of *cerrado sensu stricto* characterize the study site, which is the most prevalent and complex Cerrado physiognomy, with tree cover ranging from 20% to 70% (Ribeiro and Walter 2008).

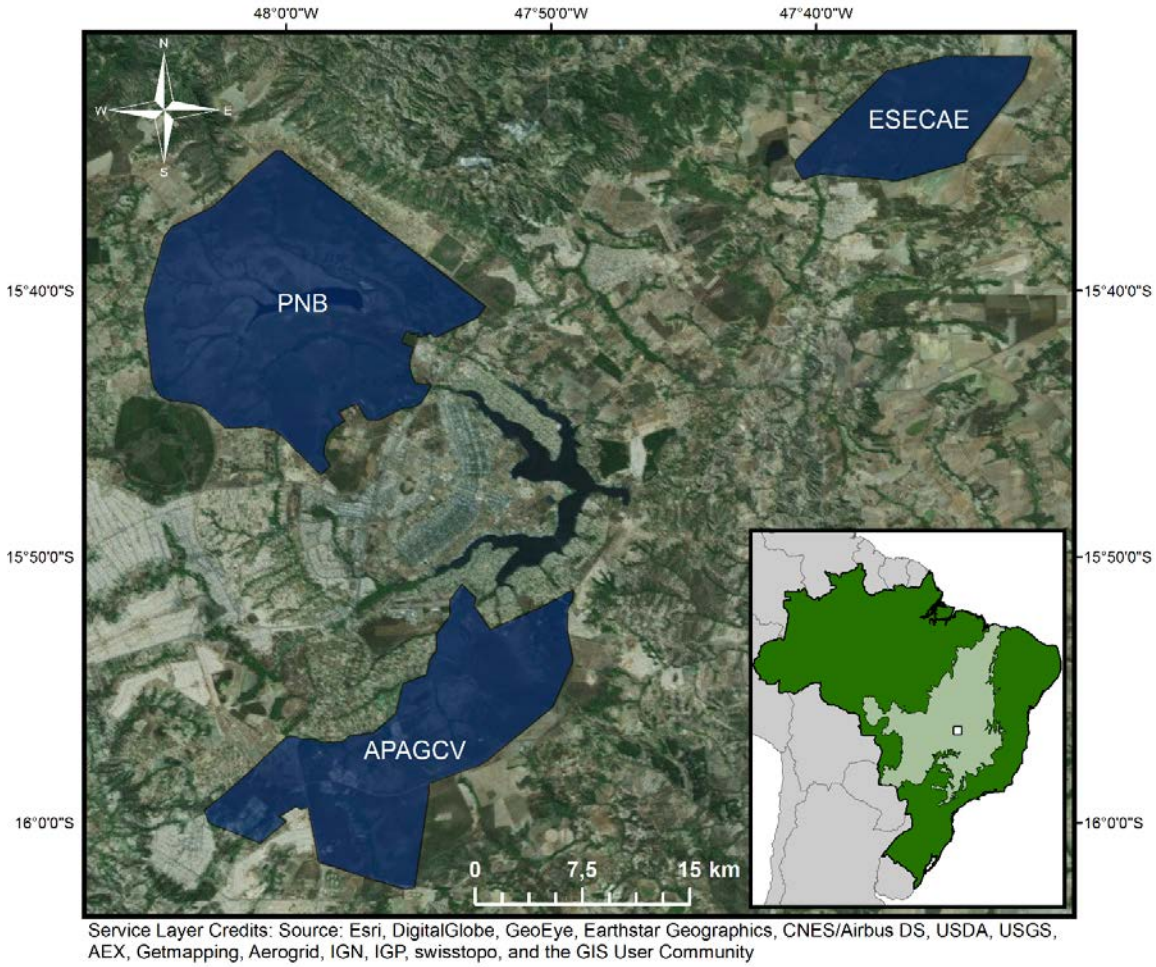


Figure II-2: Study sites surrounding Brasília, Federal District. Brasília National Park (PNB), Águas Emendadas Ecological Station (ESECAE) and Gama Cabeça de Veado Environmental Protection Area (APAGCV) depicted in dark blue. The overview map shows the national territory of Brazil (green) located in South America (grey) as well as the extent of the Brazilian Cerrado (light green). The square in the overview map depicts the location of the three study sites.

2.2 Time series and phenological patterns

In order to create a dense Landsat time series, we combined all available Level 1 Terrain (L1T) geometric corrected TM, ETM+ and OLI data over the study area (path: 221, row: 071) acquired between 2000 and 2014 with a cloud cover of up to 90%, resulting in 431 potential observations in an 8-day interval. However, the climatic seasonality in the Cerrado leads to cloud contamination especially in the wet season and thus to a low clear data availability (Sano et al. 2007). Another limiting factor for a dense time series is enforced by sensor errors, such as the scan line corrector failure in the case of the Landsat 7 ETM+. To overcome these drawbacks, we used a weighted ensemble of Radial Basis Function (RBF) convolution filters to detect outliers and approximate missing data in a Landsat time series. Landsat data were converted to top-of-atmosphere reflectance values and clouds and cloud shadows were masked using the FMASK algorithm (Zhu and Woodcock 2012). The

converted data were transformed to tasseled cap (TC) greenness (TC_{green}), brightness (TC_{bright}) and wetness (TC_{wet}) components using the TC coefficients for top-of-atmosphere reflectance, recommended for applications where atmospheric correction approaches are not feasible (Huang et al. 2002b). The transformation into TC components has been shown to be useful for phenology-based classification approaches (e.g. Dymond et al. 2002).

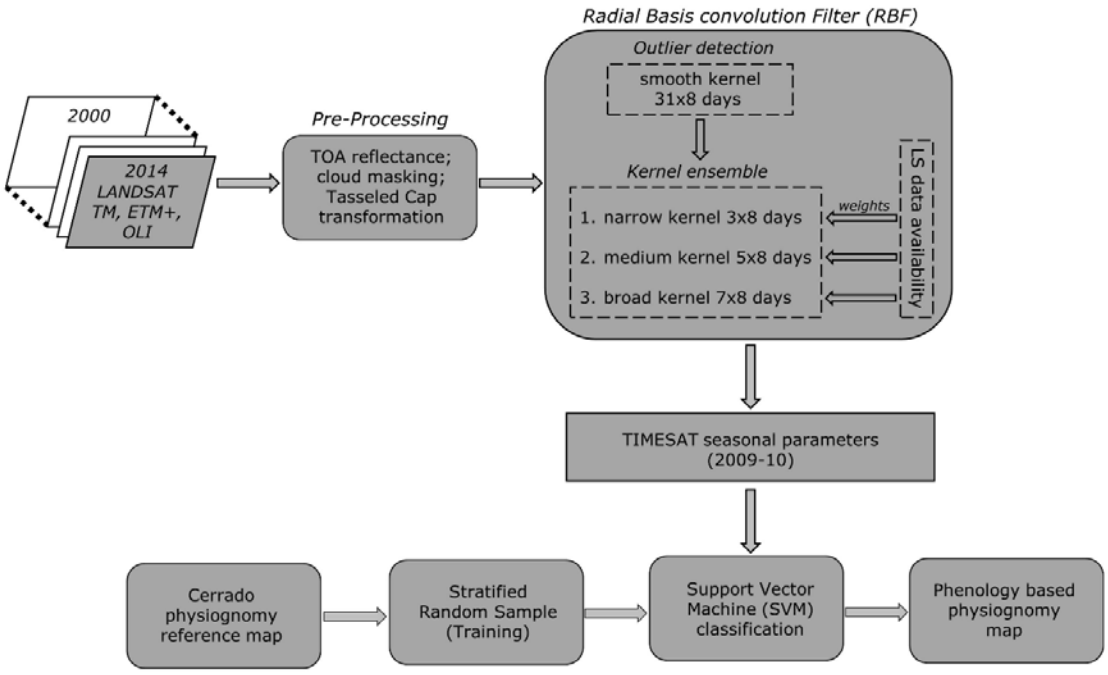


Figure II-3: Schematic workflow of the mapping approach.Landsat pre-processing and TC transformation, followed by RBF convolution filter for outlier detection and a RBF convolution filters ensemble to fill data gaps. Phenological parameters were derived for the 2009/10 season and used as input for a SVM classification.

To approximate the given clear TC observations into a dense 8-days-sampled time series without data gaps, we used a pixel based Gaussian convolution filter approach (Figure II-3). This approach is comparable to a moving window average filter (in time), whereas the weights are given by a Gaussian function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{(Eq. II-1)}$$

with σ defining the shape of the kernel function. After an expert driven visual inspection we used a smooth Gaussian kernel with σ = 5 to detect outliers in our observations, which are mostly related to clouds or cloud shadows. The kernel includes all coefficients of the +/- 3 x σ interval, i.e. an overall size of 31 observations (15 to the left and 15 to the right), resulting in a kernel size of 31 x 8 days = 248 days. All observations that were more distant than one

standard deviation from the fitted function were defined as outliers and eliminated. To derive the final phenological profiles, we used an ensemble of three Gaussian kernel based convolution approximations (σ = 3 x 8, 5 x 8, 7 x 8 days). The final approximation is the weighted average of the three temporal convolution filters, whereas weights are dependent on clear observations within the given kernels. To describe the course of the smoothed continuous TC profiles on a per-pixel basis, we truncated the approximations and derived phenological parameters for the season 2009-2010. For each of the three TC time series 9 seasonal parameters were calculated using the TIMESAT 3.1 software (Jönsson and Eklundh 2004). Parameters included day-of-the-year (DOY) or start, mid, end, and length of season and phenological proxies like peak and base value, seasonal amplitude or rate of increase and decrease. Detailed information on the calculation of TIMESAT parameters are found in Jönsson and Eklundh (2004).

2.3 Mapping vegetation gradients

To map the spatial patterns of Cerrado physiognomies, we applied a support vector machine classification (SVM) algorithm (Vapnik 1998). The SVM is a non-parametric machine learning algorithm, whose underlying principle is to define an optimal hyperplane that separates sample points based on the given training points (Huang et al. 2002a), in our case a set of phenological parameters, and is then applied to the area of interest. The SVM transforms the input data into a high-dimensional feature space by applying a kernel-function to solve a non-linear classification problem. The SVM model was optimized by defining the Gaussian kernel width (γ) and the regularization parameter C via a cross-validated grid search. We used the imageSVM implementation based on LIBSVM (Chang and Lin 2011), as available in the EnMAP-Box Version 2.03 (van der Linden et al. 2015). The SVM model was trained with a set of stratified random sampled pixel, accounting for feature variability in all classes (Figure II-3). A physiognomy map for the PNB site (Ferreira et al. 2007), lastly updated in 2012 by visual interpretation of high resolution imagery from 2009, and a map of the IBGE Ecological Reserve, were used as reference data. For the remaining areas of APAGCV and ESECAE, we used a stratified green peak TC_{green} vegetation map from January 2010 as a proxy for vegetation density (Leitão et al. 2015). All sampled pixels were manually labelled to a physiognomy class by visual interpretation of high resolution Google Earth imagery. Pixel that were partly covered by “non-natural” features such as roads or which could not be adequately labelled e.g. due to missing imagery from that time period were excluded from the training set, which resulted in a total of 1897 trainings pixels (*campo limpo*: 230; *campo sujo*: 451; *campo cerrado*: 425; *cerrado sensu stricto*: 401; *cerrado*

denso: 181; *cerradão*: 64; *mata de galeria*: 145). The classification accuracy was assessed through a 10-fold cross-validation.

3 Results and Discussion

3.1 Phenological profiles

We derived pixel-wise TC_{green}, TC_{bright} and TC_{wet} phenological profiles from the filtered Landsat time series for the period 2009-2010. Figure II-4 shows averaged profiles for each Cerrado physiognomy based on the training pixels and reveal a general seasonal trend, resembling the dry season from May to September and the rainy season from October to April, which was especially prominent in the TC_{green} profiles. In general the TC_{green}, TC_{bright} and TC_{wet} profiles differed in shape and amplitude, but they all showed the same order from high to low values between different physiognomies, with TC_{bright} showing the opposite direction of TC_{green} and TC_{wet}. Physiognomies with sparse vegetation densities i.e. *campo limpo* and *campo sujo* (lighter colors) produced the lowest values (highest in TC_{bright}) and were followed by denser physiognomies, with *cerradão* showing the highest values (lowest in TC_{bright}).

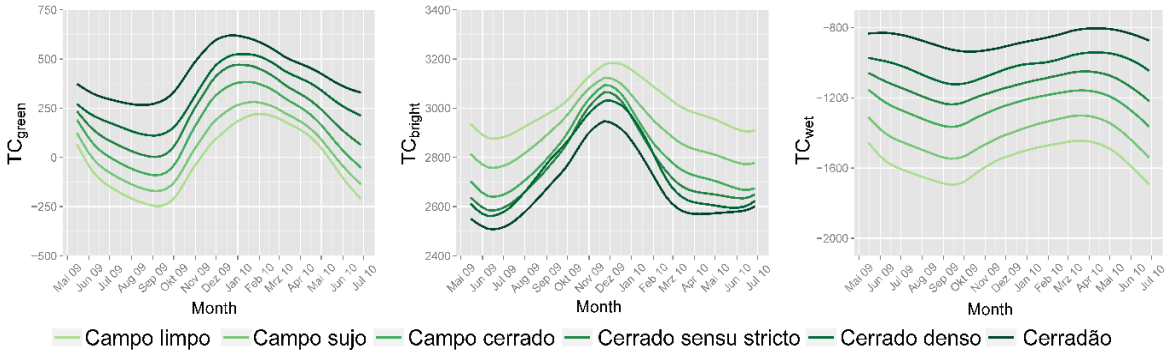


Figure II-4: Averaged TC phenological profiles for each physiognomy in the season 2009-10.

All profiles had a shift of the phenological peaks between different physiognomies. This is particularly evident in the TC_{green} profile, where phenological peaks shifted around 3 months between different physiognomies. Denser physiognomies reached their peak earlier than sparse ones and created flatter profiles, as there was less phenological variation due to e.g. larger degrees of evergreen vegetation. These observations revealed the expected patterns while being in line with findings of other studies using data with coarser spatial resolution (Ferreira and Huete 2004; Ratana et al. 2005). Our fitted temporal profiles allow the

reconstruction of Land Surface Phenology (LSP) in savanna ecosystems at the Landsat spatial resolution, which has to our knowledge not been shown before.

3.2 Phenological metrics

We used the derived profiles to calculate pixel-wise phenological parameters using TIMESAT for the season 2009/2010 to describe the course of the phenological profiles. The distribution of these parameters in each physiognomy stressed their phenological differences. Figure II-5 shows an example of the distribution of selected TC_{green} phenological parameters for the main Cerrado physiognomies based on the set of trainings pixels. In most cases, the physiognomies can be clearly distinguished. The base value, which was defined as the average of the two seasonal minima (Jönsson and Eklundh 2004), revealed a clear trend of generally higher minima with denser vegetation physiognomies. This pattern in the observed LSP signal can be related to less above ground biomass in the open landscape formations such as campo limpo and campo sujo, which naturally increases with denser vegetation physiognomies (de Miranda et al. 2014). Highest values can be observed for cerradão that can consist of up to 15m tall trees with crown covers of around 90% (Ribeiro and Walter 2008). Similar trends were observable for the rates of increase and decrease calculated as the ratio of the difference between 20% and 80% at the beginning (greening-up) and end (browning) of the phenological curve and the related difference in time (Jönsson and Eklundh 2004). The differences between each physiognomy were particularly pronounced in the rates of decrease, which can be related to less amplitude in dense physiognomies stressing the buffer effect of woody vegetation to seasonality (Ratana et al. 2005).

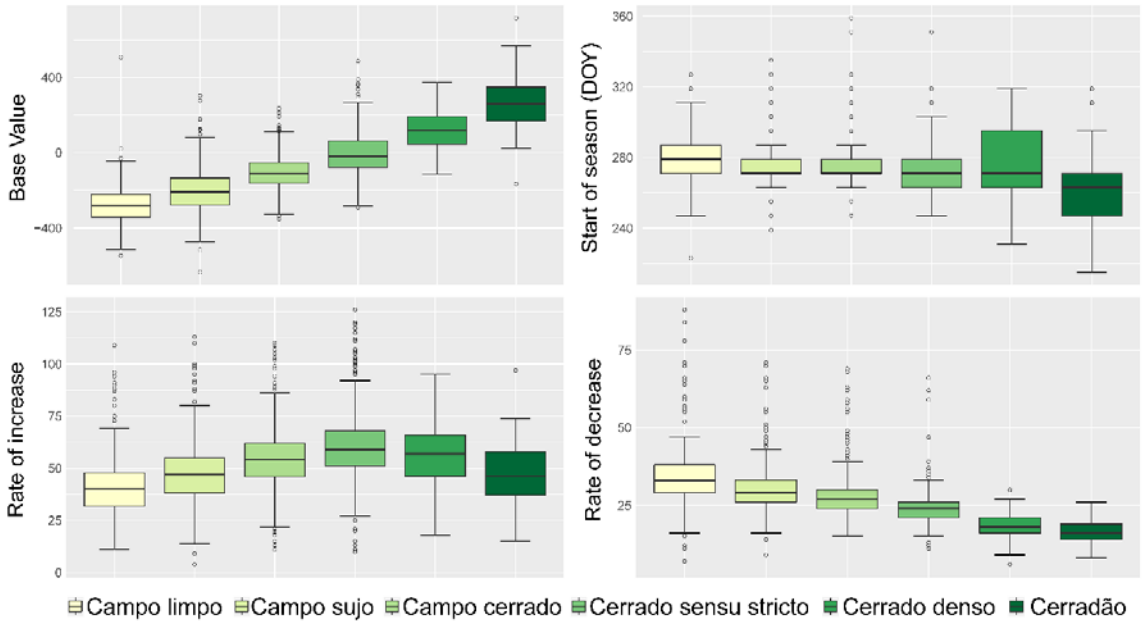


Figure II-5: Boxplots showing the mean, 25 and 75 percentile of selected TC_{green} parameters (base value, end-of-season, rate of increase and decrease) for five different physiognomies based on the training pixels. Whiskers have the length of 1.5 times the inter-quartile range, beyond which all other data are considered outliers.

The TC_{green} start-of-season parameter, defined as the point in time when the phenological curve has reached 10% of its range since the minimum (Jönsson and Eklundh 2004), did not follow this clear trend and showed less differences between the physiognomies (Figure II-5). The woodland physiognomy *cerradão* has the earliest start of the season at around DOY 263 (20th of September 2009), whereas the open landscape *campo limpo* season starts later around DOY 279 (6th of October 2009). The seasonal start of the other physiognomies falls somewhere in between these dates and differs mainly in the interquartile range. The median lies around DOY 271 (28th of September 2009) for all physiognomies. Even though the start-of-season is an important phenological parameter, integrated LSP derived from an 8-daily remote sensing time series are not necessarily in line with in-situ phenological observations (Hanes et al. 2014).

3.3 Vegetation mapping

Based on the set of phenological parameters we trained a SVM classification model to map the spatial distribution of the main Cerrado physiognomies. The broad-scale spatial patterns of all observed physiognomies were well depicted and in line with available reference maps (Figure II-6). Large areas with sparse vegetation physiognomies (from *campo limpo* to *campo cerrado*) were prevalent in PNB with gallery forests following the river flows. Areas with dense physiognomy classes like *cerrado denso* and *cerradão* were almost not present.

The physiognomies in APAGCV showed similar patterns with gallery forests along the water bodies, sparsely vegetated areas in the center and denser physiognomies at the border of the study area. Also small patches of *cerrado denso* and *cerradão* occurred. In contrast to the other two study sites ESECAE was characterized by denser vegetation. Most of the area was classified as *cerrado sensu stricto* and *cerrado denso*, respectively. Areas of sparsely vegetated *campo limpo* and *campo sujo* were located in the center as well as in the western parts of ESECAE. Small patches of *cerradão* were classified within the *cerrado denso* regions.

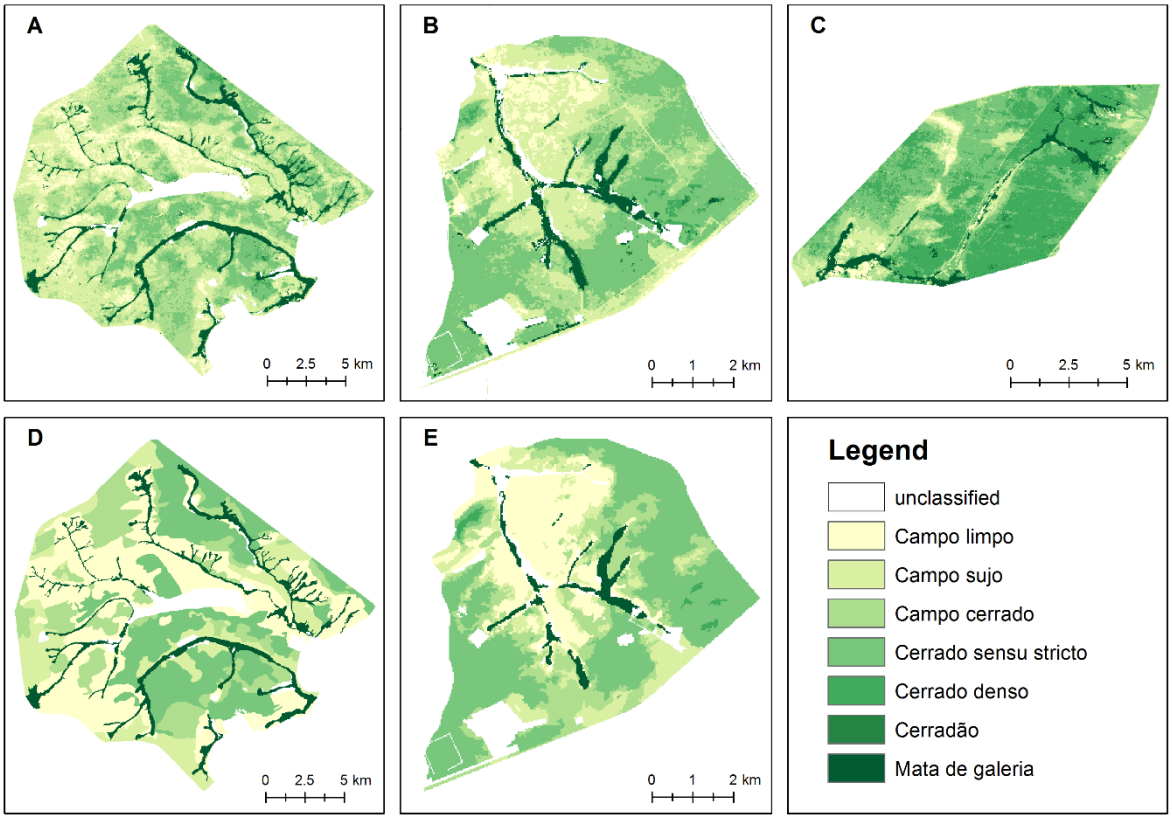


Figure II-6: Classification results based on seasonal phenological parameters for A) PNB, B) IBGE parts of APAGVC for which a reference map was available and C) ESCEA. D) and E) show the physiognomy patterns in the reference maps.

Linear artefacts occurred in all three study sites but were most prominent in ESECAE, predominantly in regions of SLC-off errors. The results suggest that not only the total amount of available cloud and error free data is relevant for our mapping approach, but also its seasonal distribution. This is in line with findings of other studies that also distinguished Cerrado land cover classes based on Landsat time series (e.g. Müller et al. 2015; Rufin et al. 2015). It seems to be rather important that the peak of the season is covered by data, as it directly influences the derivation of phenological parameters, which in turn determine the

classification outcome. However, this effect is likely to be class-dependent, as it did not lead to misclassification throughout all physiognomies.

A detailed look at the classification results revealed that not many regions in all three study sites are classified as pure physiognomies and occasionally include pixels from the adjacent class. This "salt and pepper" effect (Figure II-6; Figure II-7) reflects the heterogeneity within the complex ecosystem and is likely related to mixed pixel between adjacent physiognomy classes, which from a remote sensing perspective are different shares of understory and canopy cover on a pixel level. This effect is particularly prominent when looking at gradients of vegetation density in heterogeneous landscapes at 30 m spatial resolution. Figure II-7 shows examples of spatially-explicit phenological differences and gradual transitions between physiognomies and how those translate into classes according to a standard classification scheme for Cerrado landscapes used in Brazil (Ribeiro and Walter 2008).

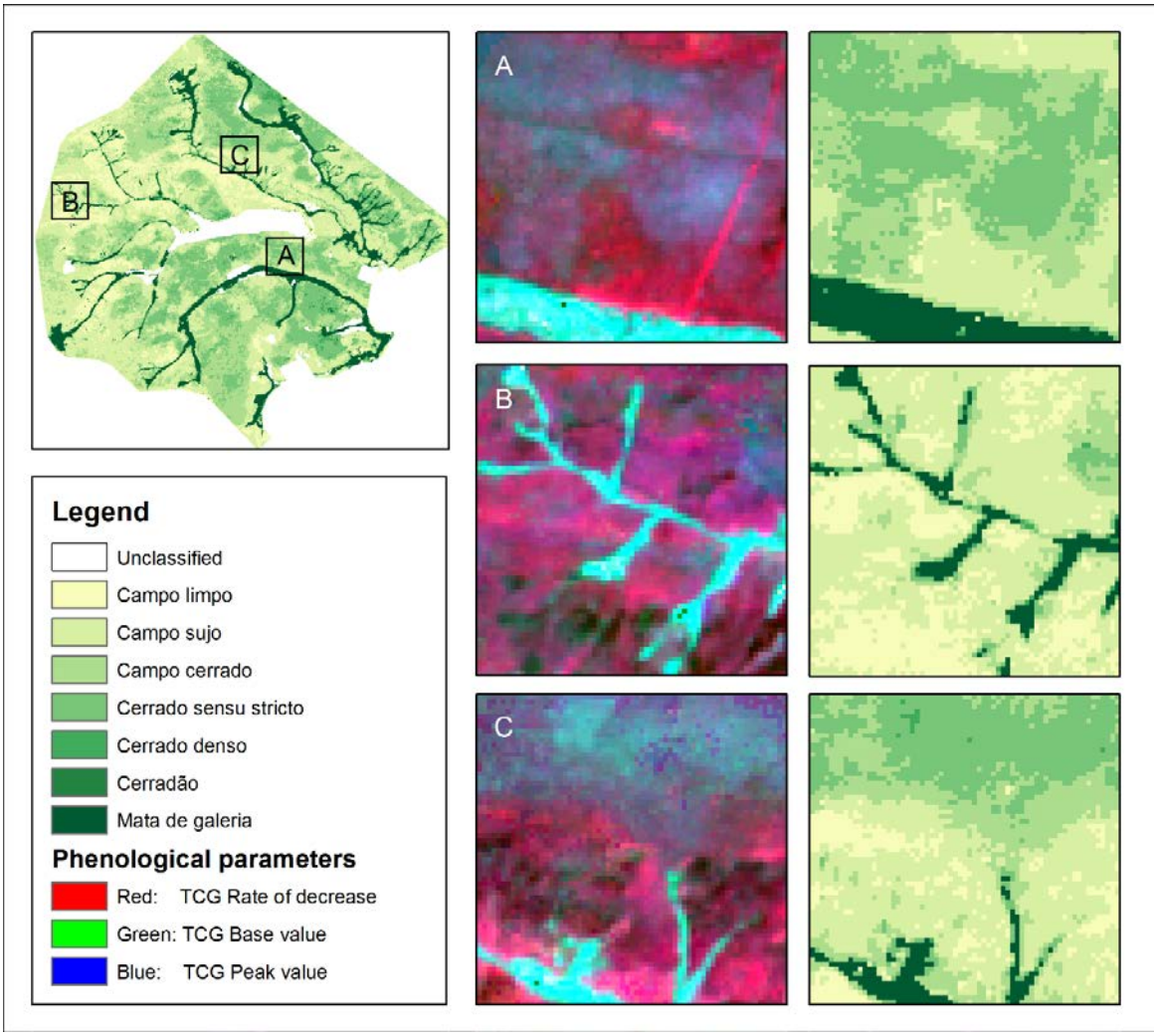


Figure II-7: Close-ups of the PNB classification map and three TC_{green} parameters in RGB that depict gradual transitions between physiognomies.

The final map was validated with a 10-fold cross-validation resulting in an overall accuracy of 63% (Table II-1). This relatively weak accuracy measure does not agree with the visual comparison of the reference map and the classification outcome. It is a challenging task to map and validate a classification stemming from a field-based classification scheme with a classification at Landsat resolution. We therefore applied a softer validation scheme to accommodate for spectral-temporal class similarities between neighboring classes along the vegetation gradient (e.g. between campo sujo and campo limpo). The overall accuracy then reached surprisingly high 96% (Table II-1; numbers in parentheses). This pattern was also reflected in the SVM class probabilities, which revealed that approximately 90% of all pixel had the second highest probability in the thematically adjacent class. Emphasizing the difficulties in discrete classifications for mapping vegetation gradients with a high spatial resolution.

Table II-1: Confusion matrix resulting from the 10-fold cross validation, including users (UA), producers (PA) and overall accuracies (OA). The numbers in parentheses show the soft validation results, allowing misclassifications in adjacent classes.

	Reference Class									
Map class	*Campo limpo*	*Campo sujo*	*Campo cerrado*	*Cerrado sensu stricto*	*Cerrado denso*	*Cerradão*	*Mata de galeria*	Sum	UA	*Area (ha)*
Campo limpo	101 (222)	49 (0)	2	1	0	0	1	154 (226)	0.66 (0.98)	5907
Campo sujo	121 (0)	290 (433)	80 (0)	16	0	1	0	508 (450)	0.57 (0.96)	16189
Campo cerrado	5	94 (0)	269 (421)	81 (0)	2	0	2	453 (430)	0.59 (0.98)	12611
Cerrado sensu stricto	2	18	72 (0)	264 (377)	61 (0)	2	7	426 (406)	0.62 (0.93)	12906
Cerrado denso	0	0	2	32 (0)	109 (174)	26 (0)	2	171 (178)	0.64 (0.98)	3748
Cerradão	0	0	0	3	4 (0)	29 (61)	8 (0)	44 (64)	0.66 (0.95)	450
Mata de galeria	1	0	0	4	5	6 (0)	125 (133)	141 (143)	0.89 (0.93)	3823
Sum	230	451	425	401	181	64	145	1897		
PA	0.44 (0.97)	0.64 (0.96)	0.63 (0.99)	0.66 (0.94)	0.6 (0.96)	0.45 (0.95)	0.86 (0.92)	OA	0.63 (0.96)	55634

3.4 Outlook

Our results showed that the proposed approach allows mapping heterogeneous savanna ecosystems. Cloud and error free data are a mandatory prerequisite to generate the dense time series needed to derive phenological information. Recently launched sensors such as OLI on Landsat 8 and the Multispectral Instrument (MSI) onboard Sentinel-2a already

provide high quality data for time series analysis. Their combined use, especially when including future Sentinel-2b data, will potentially allow to generate dense times series with an approximate 5-day temporal resolution, making synthetic data gap filling obsolete. Issues like linear artifacts, as currently present in the classification results, will then be overcome. However, the combination of different data types is still challenging, e.g. due to different spatial and spectral resolutions (Hostert et al. 2015; Wulder et al. 2015a). Especially, the red-edge bands of Sentinel-2 (ca. 705, 740 and 783 nm) with a spatial resolution of 20 m hold potential for vegetation analyses (Verrelst et al. 2012) and further research is required to better comprehend this potential. The assets of narrow spectral bands for the analysis of heterogeneous ecosystems have also been demonstrated in several studies that used for example simulated satellite data based on airborne hyperspectral imagery (e.g. Roberts et al. 2015; Schwieder et al. 2014) or hyperspectral data from experimental satellites such as Hyperion (e.g. Leitão et al. 2015). These data enable the derivation of biophysical vegetation parameters (e.g. Souza et al. 2010) or the spectral unmixing of sub-pixel fractions of photosynthetic and non-photosynthetic active vegetation (e.g. Miura et al. 2003). On the one hand, this additional information can improve classification outputs, e.g. in combination with phenological data. On the other hand, continuous biophysical parameters can be quantified through regression approaches, if appropriate reference data are available. Such strategies allow overcoming the limitations of hard classifications in heterogeneous ecosystems. With the advent of future spaceborne hyperspectral sensors such as EnMAP (Guanter et al. 2015) or HyspIRI (Lee et al. 2015), the potential of regression-based approaches will unfold.

4 Conclusion

Dense Landsat time series were used to analyse phenological differences between the main physiognomies of the Brazilian Cerrado. Data gaps within the time series were synthetically filled using a weighted ensemble of radial basis convolution filters. The available amount of data was sufficient to derive comprehensible phenological profiles and parameters, which enabled depicting differences between the main Cerrado vegetation physiognomies in terms of their response to climatic-dependent phenology. Based on these differences, it was possible to map the main spatial patterns of the Cerrado physiognomies using a SVM classification approach. Linear artefacts within the classification result could be related to sparse data availability (i.e. systematic errors within the data) and its seasonal distribution. This drawback will be overcome in the near future when e.g. Landsat 8 OLI and recently available Sentinel-2 MSI data will be combined. Inaccuracies in the final estimation were

especially prominent between thematically adjacent classes along the vegetation gradient. Incorporating (multi-temporal) spaceborne hyperspectral data, or derived products, in the analysis will further be beneficial to overcome this limitation. This will become even more feasible once upcoming sensors such as EnMAP or HyspIRI are in orbit and routinely and repeatedly provide spectral high-resolution data. Our results emphasize the usefulness of utilizing long-term Landsat time series for the analysis of heterogeneous ecosystems, where spatial detail is as much needed as the ability to derive important phenological parameters. Such time series analyses offer new opportunities e.g. for monitoring trends in habitat changes and related biodiversity or ecosystem responses to climate change. Further research may accordingly focus on the combined use of open data archives and upcoming spaceborne sensor technologies, which enable deepening our understanding of ongoing processes in the Cerrado and comparable heterogeneous ecosystems. Nevertheless, our results also suggest that the use of "hard" classification approaches is not always adequate for mapping vegetation gradients in heterogeneous ecosystems, due to a high structural and spectral class similarity. This was shown to be particularly challenging when distinguishing adjacent (i.e. most similar) classes on the structural gradient along the different Cerrado physiognomies. Depending on the purpose of the derived results and data availability, quantitative approaches avoiding classes, e.g. regression approaches, would offer a solution to overcome shortcomings of classification approaches. Even though these approaches often require a more costly training, continuous results will improve e.g. biomass or carbon retention estimations.

Acknowledgements

This study is part of the research activities of the EnMAP Scientific Advisory Group (EnSAG), and was funded by the German Aerospace Centre (DLR) — Project Management Agency, granted by the Ministry of Economics and Technology (BMWi grant 50EE1309). The authors would like to thank Manuel Ferreira and Betânia Goes for providing the reference maps and Philippe Rufin for inspiring and helpful discussions. This research also contributes to the Global Land Project (http://www.globallandproject.org) and the Landsat Science Team (http://landsat.usgs.gov/Landsat_Science_Team_2012-2017.php)

Chapter III:
Land surface phenological archetypes of the Cerrado

Abstract

Insights into the phenology of ecosystems are crucial to deepen our understanding of ecosystem dynamics and their responses to environmental change. Land surface phenology metrics, derived from remote sensing data, capture the vegetation's seasonal behavior on a pixel level and enable to frequently gather phenological information over large extents in a synoptic manner. Especially in dynamic and heterogeneous ecosystems, such as savannas, the temporal and spatial resolutions of the remote sensing data are critical. The Brazilian savanna, commonly known as the *Cerrado*, is characterized by a gradient of natural vegetation structure and density, along with varying species compositions for which phenological information over large scales remain fragmented. As a weak conservation status along with a growing demand on land resources for agricultural production have already led to large land conversions of the Cerrado's natural vegetation, it is critical to provide land managers and decision makers with spatially explicit information on hotspots of conservation need. To this aim, a gap-filled 8-day Landsat Collection 1 EVI time series (covering 2013 – 2017) was used for the derivation of land surface phenology (LSP) metrics for the 2 Mkm² extent of the Cerrado biome. Eight LSP archetypes of the Cerrado were defined based on the similarities between LSP metrics using a k-means cluster algorithm. The LSP archetypes enabled the phenological patterns throughout the whole Cerrado to be described on an ecoregions level and show the complementary benefits of both concepts e.g. for a spatially explicit design of conservation strategies. The presented results highlight the advantages of LSP metrics derived from dense Landsat time series for the analysis of heterogeneous ecosystems over large extents. This proposed concept of LSP archetypes is beneficial for a range of applications such as carbon quantification, the assessment of biodiversity or climate models.

1 Introduction

Phenology is the study of reoccurring events during a year or season (Lieth 1974). It can be linked to the behavior of animals, such as phases of mating, breeding, or movement and to events such as green-up, bud burst, flowering, or senescence when referring to vegetation, as a response to changing environmental factors throughout a season. To have insights in the vegetation's phenology is a key parameter, for example, to understand vegetation community dynamics in ecosystems (Williams et al. 1999). Further, is the analysis of

phenological patterns in time and space of utmost importance for global and climate change research (Menzel 2002). Indeed, changes in phenological patterns are an indicator of climate change (Menzel et al. 2006; Richardson et al. 2013; Walther et al. 2002) and as such critical input variables, for example, in global circulation models (White et al. 2003). Traditionally, phenological information is derived from repetitive observations of vegetation and protocolling the dates of the vegetation's transition into a different phenophase. Depending on the density of field observations, this approach may be very accurate but at the same time very labor intensive and usually restricted to individual species at local scales (Hanes et al. 2014).

To derive phenological information over large extents, remote sensing based approaches have been widely applied in recent decades (Henebry and de Beurs 2013). These methods are usually based on the seasonal changes of spectral indices (Reed et al. 2009) through the vegetation's reflectance characteristics linked to the state of its biochemical and physical properties (Xiao et al. 2009). Because such approaches are based on pixel level, they do not reflect species-specific phenological observations but rather a mixture of vegetation canopy and background seasonal responses, known as land surface phenology (LSP; Hanes et al. 2014). LSP is recognized as an important spectral trait for the quantification of biodiversity (Lausch et al. 2016) and was proposed as a candidate essential biodiversity variable (EBV) relating to ecosystem function, to monitor progress towards the Aichi Biodiversity Targets (CBD 2010; O'Connor et al. 2015; Skidmore et al. 2015). The detail of the captured LSP is dependent on the spatial and temporal resolution of the sensor used. Commonly, sensors such as AVHRR, MODIS, Spot Vegetation, or PROBA-V have been widely used to obtain LSP (Li and Qu 2013). They have a sufficient temporal resolution of up to one day but a relatively coarse spatial resolution from 250 m to 1 km (Reed et al. 2009), which is problematic, for example, in the analysis of ecosystems with heterogeneous landscapes.

However, this limitation might be overcome by the launch of new operational satellite missions, such as Landsat 8 (Loveland and Irons 2016) and Sentinel-2 (Drusch et al. 2012), along with open data policies (Woodcock et al. 2008; Wulder et al. 2012) and a steadily increasing processing performance. Together these developments enable the analysis of unprecedented amounts of remote sensing data with a sufficient temporal and spatial resolution (Wulder et al. 2015a). Already several studies have made use of the freely accessible Landsat data archive (Wulder et al. 2012), with promising approaches that enable the analysis of time series e.g. for land use classification and intensity mapping (Müller et al. 2015; Rufin et al. 2015) or to derive land surface phenological information on a 30 m

spatial resolution (Melaas et al. 2016; Schwieder et al. 2016). In this study, a time series of Landsat Collection 1 data was analyzed to reveal its potential for deriving LSP metrics in heterogeneous ecosystems and to characterize LSP archetypes.

The regional focus of this study is on the Brazilian savanna, commonly known as the Cerrado, which is the second largest Brazilian biome after the Amazon. The Cerrado is a biodiversity-rich ecosystem with many endemic species (Klink and Machado 2005) and provides a range of crucial ecosystem services (Lima et al. 2017), with implications from local to global scales. However, processes of agricultural expansion and urbanization along with weak conservation strategies have already led to large-scale land conversions resulting in a remaining share of approximately 60 % of natural Cerrado vegetation (Sano et al. 2010). As the expansion of agriculture is not expected to halt (Lapola et al. 2013), there is a need to follow strategies that aim to maintain ecosystem functions and promote a sustainable land use (Strassburg et al. 2017). Mapping of the Cerrado's vegetation properties is hence of crucial importance, to deepen our understanding of ecological processes over large scales and to spatially explicit define conservation priority regions.

The objectives of this study are: i) to derive spatially explicit LSP metrics for the whole extent of the Cerrado, and ii) to propose phenological archetypes for the region, based on LSP similarities. The derived archetypes are further described in the context of Cerrado ecoregions, which were defined to aid biodiversity conservation planning (Arruda 2003; Dinerstein et al. 1995).

2 Data and Methods

2.1 Study area and ancillary data

The Cerrado stretches over more than 20° latitude and covers an extent of approx. 2 million km². Most parts of the Cerrado lie on the Central Brazilian Plateau and make up large parts of the three main watersheds of Brazil (Ferreira et al. 2013). The climate of the Cerrado has a strong seasonality with a wet season from October to April and a dry season from May to September (Oliveira-Filho and Ratter 2002). Soils are mainly acidic oxisols with high aluminum and iron contents (Motta et al. 2002). Shaped by environmental factors, the Cerrado has a characteristic landscape of vegetation physiognomies that exhibit a density gradient from open grasslands, over shrublands and scattered tree formations to dense forests (Oliveira-Filho and Ratter 2002). These physiognomies differ in their structural appearance but also in their species composition (Ribeiro and Walter 2008; Ribeiro and Tabarelli 2002)

and different seasonal responses in phenology (Ratana and Huete 2004; Schwieder 2015). The Cerrado has large amounts of carbon stored in its above- and belowground biomass (de Miranda et al. 2014) and thus has a pivotal role in the global carbon balance (Ribeiro et al. 2011). It further provides habitats for a wealth of species (Francoso et al. 2016) and due to its weak conservation status, it is considered as a biodiversity hotspot (Mittermeier et al. 2011; Myers et al. 2000). However, large-scale land conversions for agricultural production lead to a loss of natural vegetation and are thus threatening ecosystem services and functions (Silva et al. 2006).

To focus the analysis on the remaining natural areas of the Cerrado, the TerraClass Cerrado classification was used to mask out non-natural areas (MMA 2015). TerraClass Cerrado is a mapping project under the coordination of the Brazilian Ministry of the Environment aiming to map the different land uses of the Cerrado on a 30 m spatial resolution. The underlying data for the classification product are Landsat 8 OLI images from 2013 covering the whole extent of the Cerrado. With a minimal mapping unit of 6.25 ha are the different land use types of the Cerrado classified into natural (forests, savanna and grasslands, non vegetated areas and water bodies) and anthropogenic (agriculture, pastures, forestry, mining, urban, land use mosaic, bare soil and others) classes with an overall accuracy of 80,2 % (MMA 2015). Even though the natural vegetation class (forest, savanna, and grasslands combined) has an accuracy of approximately 60 % it is still the most accurate product available that covers the whole extent of the Cerrado and was thus used to mask out areas that were not within the scope of this analysis.

For a better interpretation of the final results, we looked at the ecoregions defined for the Cerrado. Ecoregions are biogeographic units, defined as areas with similar environmental conditions and characteristic species communities (Dinerstein et al. 1995). They represent geographically explicit regions based on natural boundaries (Arruda 2003) and have been developed to assist land management in order to develop appropriate conservation strategies. While large-scale ecoregions have been defined for Latin America based on abiotic factors (Dinerstein et al. 1995), Arruda (2003) refined the proposed ecoregions through the addition of biotic factors, which resulted in 22 specific ecoregions for the Cerrado (Figure III-1; Arruda et al. 2008). As the ecoregions are defined by natural boundaries, their extents do not exactly match the Cerrado extent as defined by the Brazilian Ministry for Environment (MMA) and the Brazilian Institute for Geography and Statistics (IBGE; IBGE 2018; MMA 2018). Thus, the ecoregions borders were clipped to the MMA/IGBE Cerrado extent, leading

to underrepresented ecoregions (such as Chiquitânia) and areas of the Cerrado that are not covered by ecoregions (in the far northeast).

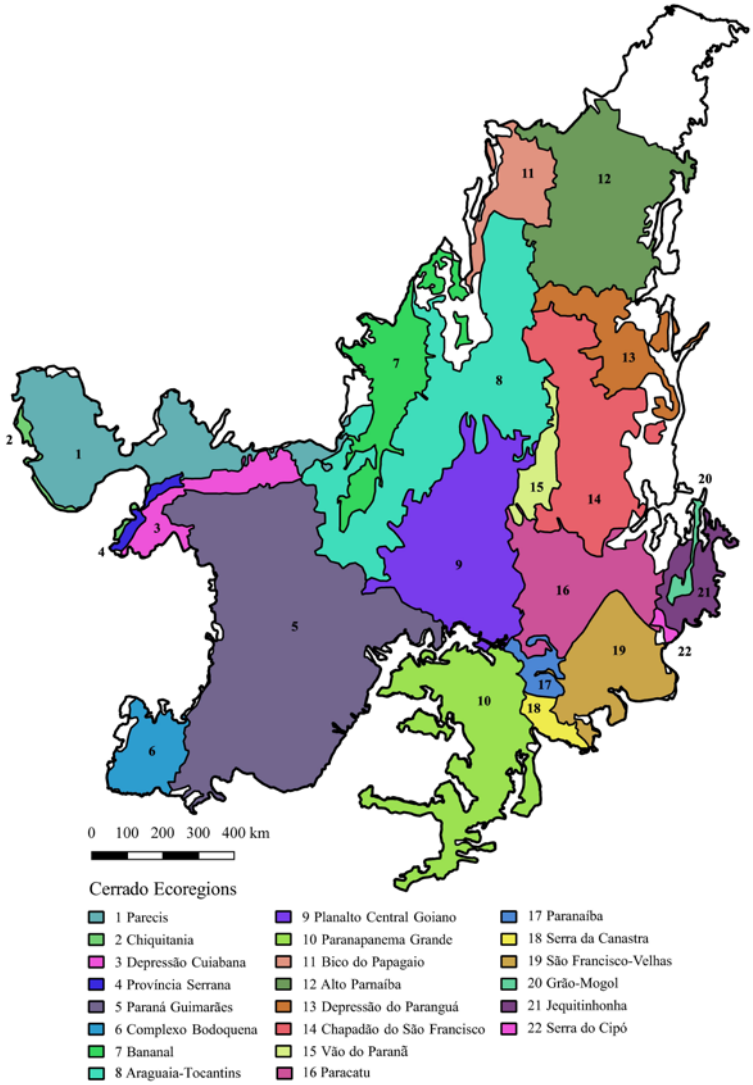


Figure III-1: Map of the 22 ecoregions of the Cerrado (Arruda 2003; Arruda et al. 2008) clipped to the IBGE Cerrado extent (IBGE 2018; MMA 2018).

2.2 Time series and land surface phenological metrics

The extent of the Cerrado is covered by 121 WRS-2 (Worldwide Reference System) Landsat footprints, for which all available Landsat ETM+ and OLI Collection 1 data for the period March 2013 to June 2017 were searched. From this selection the Enhanced Vegetation Index (EVI) (Huete et al. 2002) product was downloaded from the ESPA data archive, using images with a cloud cover of up to 80 %, which totaled approximately 23,000 images. The quality layer provided by ESPA was used to mask out clouds from the individual images.

To generate a wall-to-wall gapless 8-day time series on a per pixel level, we followed the gap-filling approach presented by Schwieder et al. (2016). Three radial basis convolution

filters (RBF) with kernel widths of σ = 8, σ = 16 and σ = 32 were calculated and combined to an ensemble using the original data availability within the respective kernel as a weighting layer. Outliers were excluded beforehand from the time series by deleting observations that were further than two standard deviations away from an RBF with a kernel width of σ = 15, as they were considered as anomalies such as undetected clouds or sensor errors.

Using TIMESAT Version 3.3 (Eklundh and Jönsson 2017) eleven phenological metrics (Table III-1) were derived from the generated gapless time series that describe the seasonal behavior of the observed vegetation (Jönsson and Eklundh 2004). TIMESAT uses a threshold to define the start and end of season which was given as the day in the year when 10 % of the season's amplitude is reached from the seasons left and right minimum values. As a search window, in which a seasonal peak is expected, the timeframe from the beginning of August 2013 to the beginning of September 2014 was defined. Although TIMESAT is capable of deriving phenological metrics for multiple peaks within one season (e.g. double cropping cultivation systems), here only metrics for one season were used as no secondary vegetative peak is expected in natural vegetation.

Table III-1: Description of the derived land surface phenological metrics using TIMESAT, based on (Eklundh and Jönsson 2017).

LSP metric	**Description**
Amplitude [Amp]	Amplitude between EVI Maximum fitted Value and Base Value
Base Value [BV]	Mean EVI value between the minima on the left and right side of the seasonal peak
End of Season Value [EoSVal]	EVI value at the defined end of season (here 10 % of the seasons amplitude)
End of Season [EoS]	8-day temporal index for the timing of the defined end of season
Rate of Increase [RoI]	Difference of the EVI values at 20 % and 80 % of the seasonal amplitude on the left side of the peak divided by the corresponding period of time
Length of Season [LoS]	Number of days between the defined start and end of season
Maximum fitted Value [Mfit]	EVI peak value of the curve
Middle of Season [MoS]	Average of the 8-day temporal indices for which 80 % on the left and on the right side of the seasonal amplitude has been reached

Rate of decrease [Rod]	Absolute values of the difference of the EVI values at 20 % and 80 % of the seasonal amplitude on the right side of the peak divided by the corresponding period of time
Start of Season Value [SoSVal]	EVI value at the defined start of season (here 10 % of the season's amplitude)
Start of Season [SoS]	8-day temporal index for the timing of the defined start of season

2.3 Cluster analysis

In order to identify clusters of vegetation with similar LSP metrics, 500,000 random samples were drawn from the pixels that were classified as natural areas. For these spatially explicit samples, the LSP metrics were derived and outlier (values smaller or equal to 0), as well as no data values, removed. This resulted in a total of 484,164 samples that were scaled prior to the cluster analysis. Based on Lloyd's algorithm (Lloyd 1982) a k-means clustering was performed with a maximum of 300 iterations and 30 initial cluster centers, which were initialized based on the k-means++ procedure (Arthur and Vassilvitskii 2007). In order to determine the number of clusters to be defined, the gap statistic algorithm (Tibshirani et al. 2001) was used. In this approach, the dispersion within each cluster is compared to the expected dispersion under a null distribution. The maximum gap between the derived cluster dispersions indicates a sufficient number of classes for the given data complexity. Due to the random initial cluster selection, this approach does not necessarily lead to constant results, thus the gap statistic calculation was iterated 100 times. The final decision on a useful number of clusters was ultimately taken context-related and not solely based on the underlying data structure. Once the cluster centers were calculated for the random samples, they were applied to the whole extent of the Cerrado to generate a map of phenological archetypes.

3 Results

3.1 Land surface phenological metrics

Eleven LSP metrics were derived from gap-filled Landsat EVI time series for the whole extent of the Cerrado. Descriptive statistics of the LSP metrics for all natural areas of the whole Cerrado extent are shown in Table III-2. On average the season starts around the end of September (day of year 266) and ends around the mid of July (day of year 198) with standard deviations of around 38 and 50 days, respectively. Resulting in a mean season's

length of 293 days (LoS) with a standard deviation of 45 days. Figure III-2 shows the spatially explicit distribution of three selected phenological metrics (Amp, BV, SoS) for the whole extent of the Cerrado in a red, green and blue (RGB) composite. The map reveals regions of similar LSP metrics and distinct patterns with, for example, high base values in the far west, late start of season at the mid-western border, and wide areas of high amplitude and an early start of season distributed in the eastern parts of the Cerrado. However, this visualization only reflects the patterns of three out of eleven LSP metrics and even though some of them are highly correlated (see supplementary material S III-1 for correlation matrix) they can provide valuable insights into the overall LSP patterns.

Table III-2. Mean, standard deviation, 25, 50 and 75 % percentiles of the LSP metrics of all natural areas for the whole Cerrado extent. Values relating to the timing of the season (EoS, SoS, and MoS) are time series index values of the 8-day time series starting from the day of year 114 in 2013 (25th of April).

	Mean	**Stddev**	**Q25**	**Median**	**Q75**
Amp	0.21	0.09	0.14	0.19	0.26
BV	0.28	0.09	0.20	0.26	0.34
EoSVal	0.30	0.10	0.22	0.28	0.37
EoS	56.90	6.30	55	57	60
RoI	0.02	0.01	0.01	0.02	0.03
LoS	293.00	45.00	272	296	320
Mfit	0.49	0.12	0.40	0.50	0.57
MoS	35.50	5.50	32	35	39
RoD	0.01	0.01	0.01	0.01	0.02
SoSVal	0.30	0.09	0.23	0.29	0.36
SoS	20.30	4.70	18	20	22

3.2 Land surface phenological archetypes

The gap statistics algorithm resulted after 100 iterations in an asymptotic curve approaching the gap value of approximately -1.6 with 39 clusters (S III-2) without revealing a clear maximum gap value. Therefore, the amount of cluster was decided on with respect to the phenological complexity of the Cerrado, for which Figure III-2 was used as an indicator, while still allowing interpretability of the results. Hence, eight k-means cluster centers were calculated and assigned to each pixel for all natural areas of the whole Cerrado extent. This resulted in different shares of cluster (Table III-3) with cluster 4 and 5 being the most represented, with 21 % and 23 % of all natural vegetation pixels assigned to them. Clusters

1 to 3 each account for around 13–15 %, while cluster 6 and 8 represent 6 % and 7 % of all pixels analyzed. The smallest cluster, with less than 2 % of all pixels, is cluster 7.

Table III-3: Cluster shares based on all pixels that have been labeled as natural vegetation in the TerraClass classification for the whole extent of the Cerrado

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Share [%]	14.66	14.92	12.70	22.91	20.59	6.96	1.67	5.58

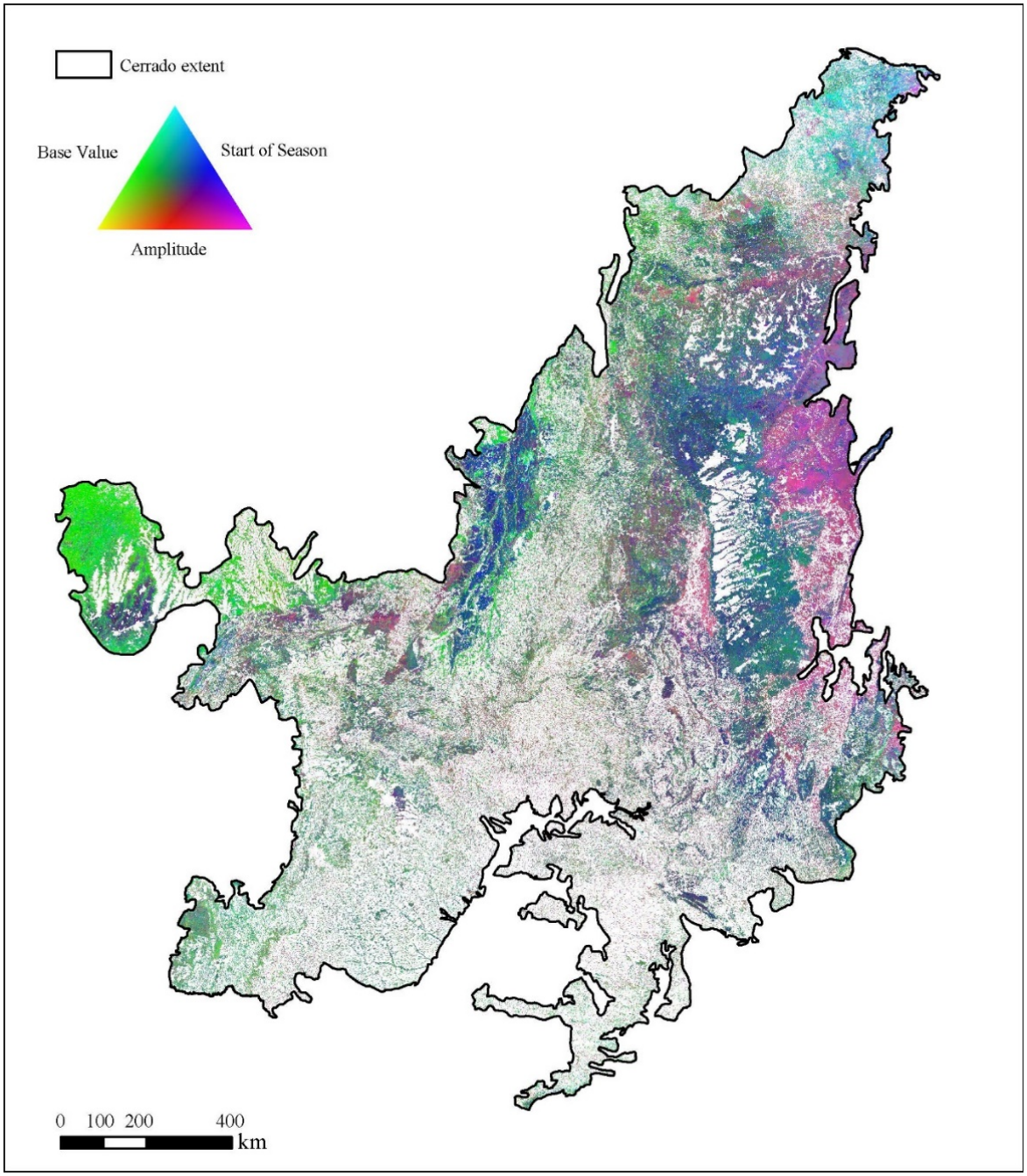


Figure III-2: RGB mosaic of three selected LSP metrics (Amplitude, Base Value, Start of Season) for all natural areas of the whole Cerrado extent.

The graphs in Figure III-3 show the mean values of the distinct LSP metrics in each of the eight clusters, where the size of the segments is scaled to enable the comparison between the individual cluster. Cluster 1 has relatively high mean values in LoS and Mfit and despite small SoS, MoS, and RoD it is characterized by average LSP values. In cluster 2 only LoS and EoS are somewhat more prominent, whereas SoSVal, EoSVal, and BV are low. Cluster 3 is mainly defined by high mean values of EoSVal, BV, SoSVal, Mfit, and LoS, whereas in cluster 4 the temporal metrics LoS and EoS have relatively high mean values and are more prominent than the other metrics. Cluster 5 has high values in LoS and EoS, and despite low RoD it has rather average values. Cluster 6 is characterized by high values of Amp, RoI, RoD, and Mfit, which are more prominent than the temporal metrics. Cluster 7 is defined by high temporal metrics SoS, MoS, EoS and above average values of SoSVal, EoSVal, BV, and Mfit. Cluster 8 is defined with relatively small values in all metrics with only RoD being somewhat more prominent. As the derived clusters are solely based on LSP metrics, they represent the main groups of vegetation with similar seasonal response in term of their LSP. Thus, these clusters are proposed as land surface phenological archetypes.

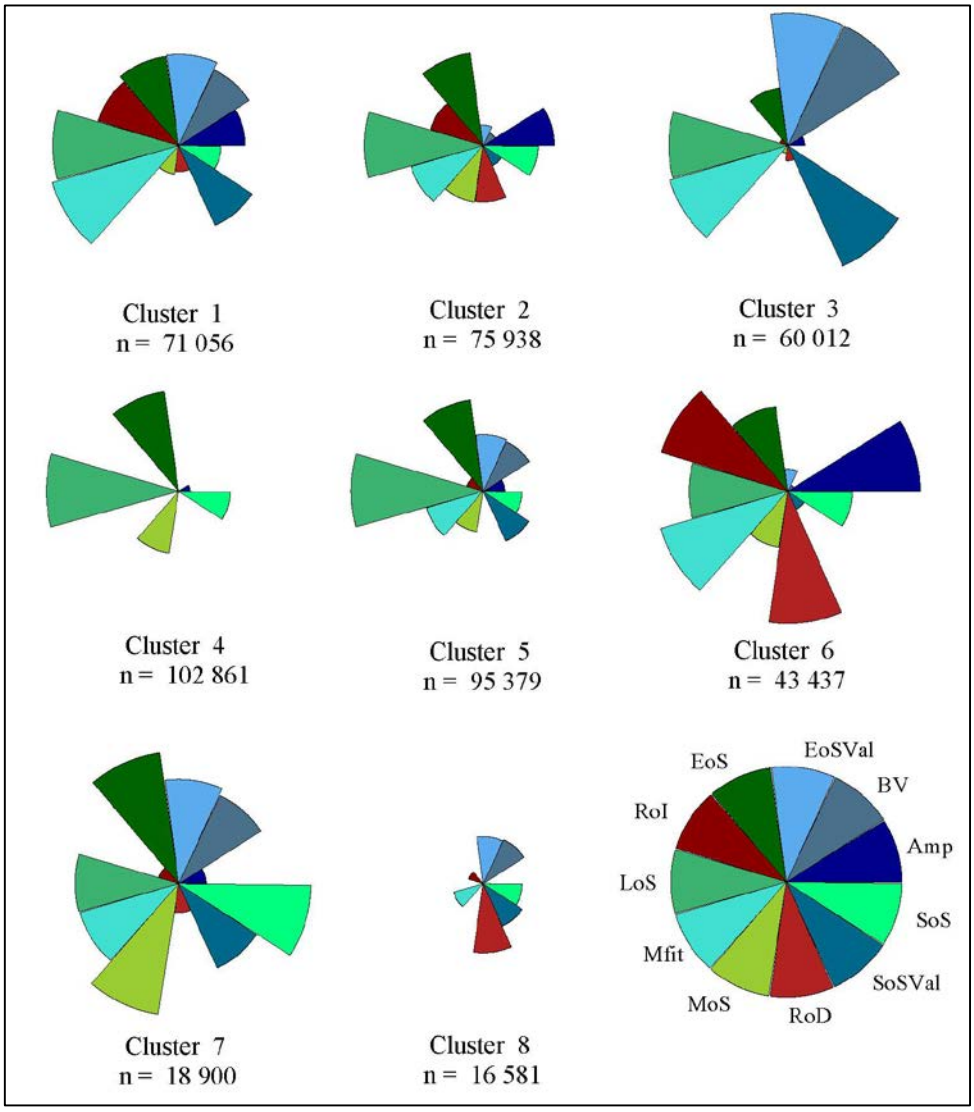


Figure III-3. Relative distribution of the mean values of each LSP metric per cluster based on the initial random sample. Each color represents a specific LSP metric and the size of each segment relates to the scaled mean value of the metric enabling to compare the individual cluster. Below the graphs are the number of samples within each cluster. Explicit values are listed in the table in supplementary material S III-3.

3.3 Spatial patterns of LSP archetypes

The spatial distribution of the LSP archetypes reveals clear patterns of phenological similarities in the natural vegetation of the Cerrado. Most of the unconverted areas can be found in the more western regions of the Cerrado, which is dominated by cluster 3 as well as in the central to northern regions where cluster 4 and 5 are prominent (Figure III-4).

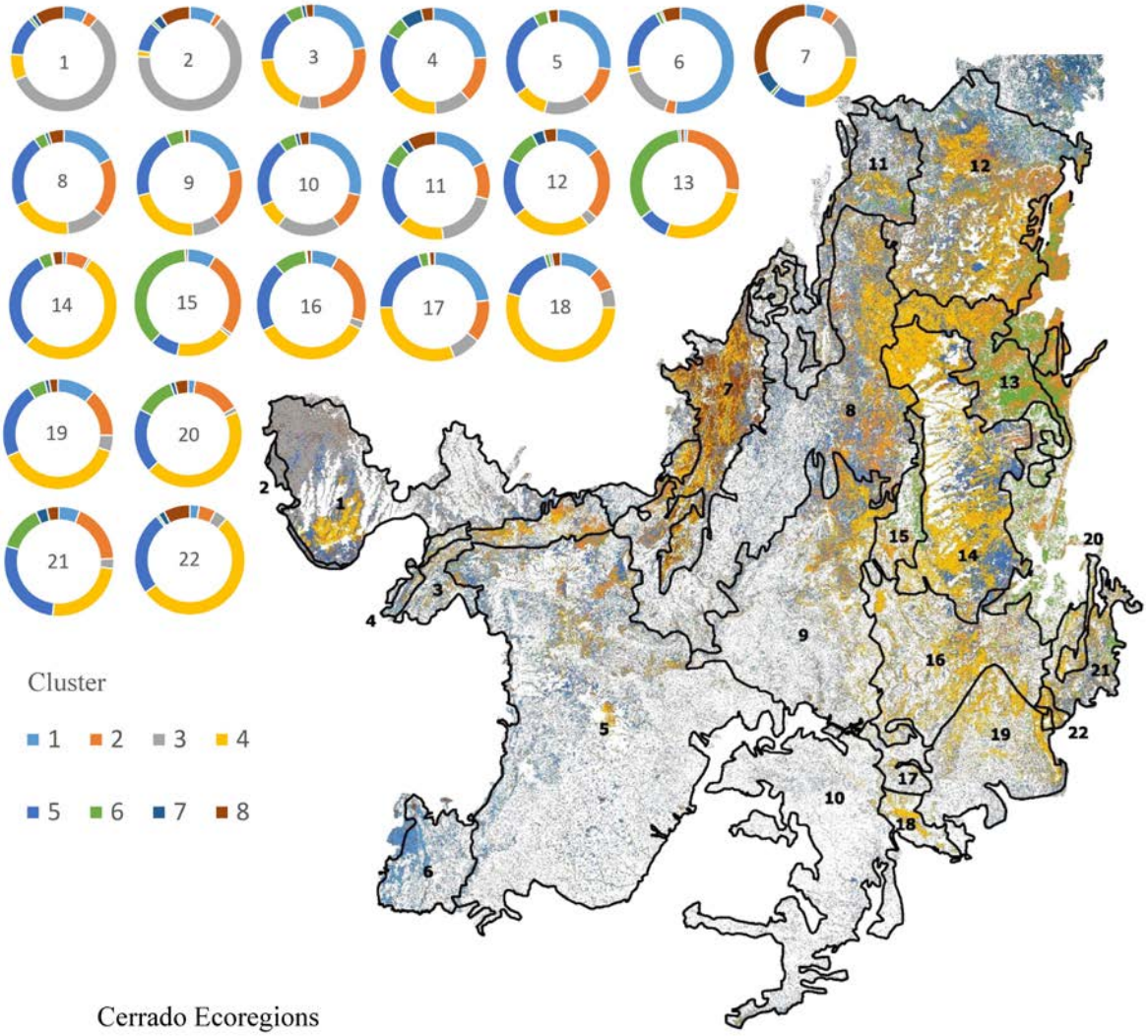


Figure III-4. Spatial distribution of the LSP cluster for the whole Cerrado extent along with the 22 ecoregions. The different colors depict the distinct LSPs. The ring diagrams show the share of clusters in each ecoregion.

The ring diagrams in Figure III-4 show the share of each cluster within the individual ecoregions in relation to the remaining natural vegetation. Based on the TerraClass classification were the shares of remaining natural vegetation per ecoregion derived. The ecoregions *Complexo Bodoquena* (46.74 %), *Planalto Central Goiano* (46.38 %), *São Francisco-Velhas* (44.66 %), *Serra da Canastra* (39.95 %), *Paraná Guimarães* (34.77 %),

Paranaíba (34.24 %), and *Paranapanema Grande* (19.42 %) being the ones in which more than half of the natural areas were converted (S III-4).

Within the different ecoregions vary the shares of LSP archetypes. Most ecoregions have at least a small share of each LSP archetype, and most of the 22 regions are dominated by 1-3 archetypes. The ecoregion, *Complexo Bodoquena*, in the southwest of the Cerrado, shares a border with the Pantanal and is with more than 50 % of all pixels dominated by Cluster 1, which is characterized by high Mfit and LoS values. Cluster 2 with high EoS and LoS values is most prominent in the ecoregions *Planalto Central Goiano*, *Alto Parnaíba*, *Depressão do Paranguá*, *Vão do Paranã* and *Paracatu*, which are located in the central to northeastern parts of the Cerrado. *Parecis* and *Chiquitania* are dominated by Cluster 3, which is defined by relatively high values of BV, Mfit, SoSVal, and EoSVal. Cluster 4 and 5 are dispersed throughout the whole extent of the Cerrado but with some larger contiguous regions in the ecoregions *Araguaia-Tocantins*, *Planalto Central Goiano*, *Alto Parnaíba*, *Chapadão do São Francisco* in the northeastern Cerrado. Cluster 6 with high values of RoI and RoD, is dominant in the ecoregions *Depressão do Paranguá* and *Vão do Paranã*. Cluster 7 and 8 are the two smallest clusters and are mainly located in *Parecis*, *Chiquitania,* and the *Bananal*. The *Bananal* is the only ecoregion that has a considerable share (>25 %) of Cluster 8 with clear spatial patterns along the river formations. Cluster 8 has in general relatively small LSP metric values, with the exception of high RoD values indicating a rapid decrease of EVI values at the end of the season.

4 Discussion

Dense time series of Landsat collection 1 EVI products enabled the derivation of LSP metrics for the whole extent of the Cerrado. Because the EVI is known to decouple the vegetation canopy from the background signal (Huete et al. 2002) it enabled the analysis of LSP in a heterogeneous ecosystem, especially with a 30 m spatial resolution. The gap filled 8-day time series of combined Landsat ETM+/OLI data captured the seasonal dynamics of the Cerrado vegetation and allowed to derive LSP metrics. Based on the similarities of these metrics, eight LSP archetypes were proposed for the Cerrado, which allow describing the phenological diversity of the Cerrado. However, the proposed concept of LSP archetypes is not restricted to the selected amount of clusters, which might be adjusted depending on the intended application.

As there was a relationship expected between the LSP archetypes and Cerrado vegetation physiognomies (Ratana et al. 2005; Schwieder et al. 2016), their patterns were visually compared to high resolution Google Earth imagery. In combination with the distribution of LSP metrics within the clusters (Figure III-3), it can be concluded that for example Cluster 1 and 3 broadly relate to high biomass vegetation, indicated through relatively high LSP values of Mfit and BV. These archetypes might thus be related to vegetation physiognomies such as gallery forests and *cerradão*. In contrast Cluster 2, 4 and 5 are characterized by long seasons and a late end of season, with low values in BV, Mfit, SoSVal, and EoSVal, describing a seasonal dynamic that is rather related to grassland physiognomies such as *campo limp*o and *campo sujo* (Schwieder et al. 2016). However, LSP archetypes reflect the seasonal dynamic of the observed vegetation as captured in dense Landsat EVI time series. They are thus related to variations in photosynthetic activity over time, which are on the one hand determined by vegetation structure, but on the other hand, also influenced by species composition and environmental factors that influence water and nutrient availability. Even though there is a relationship between these two concepts a direct assignment of LSP archetypes to vegetation physiognomies is not straight forward, as classes in the latter approach are defined by structural vegetation parameters (Ribeiro and Walter 2008). The additional information in LSP archetypes is critical especially for large-scale analyses of ecosystems with large extents and therefore strong variations in environmental factors, as in the Cerrado. These variations might also explain the spatial patterns of LSP archetypes, with some of them being more prominent in specific regions of the Cerrado.

To interpret the general patterns of LSP archetypes they were compared with the Cerrado ecoregions. It could be observed that the patterns match in some cases the shapes of the ecoregions such as *Parecis*, *Vão do Paranã* and *Bananal*. The *Bananal* is a very prominent example in which the LSP archetypes render the extent of the ecoregion. It is the only ecoregion that is dominated by cluster 4 and 8, which might be caused by the influence of the high frequency of inundation. The ecoregion is characterized by plain surfaces, interfluvials, and river connecting channels (Arruda 2003) and within its extent is the largest river island of the world which is seasonally flooded (Borma et al. 2009). Silva et al. (2006) revealed even similarities between the *Bananal* and the Pantanal, which is world's largest wetland area and commonly considered as an independent biome of Brazil. However, in other regions of the Cerrado contiguous areas of the same LSP archetype cross the extent of the defined ecoregions (e.g. *Depressão do Paranguá* and *Chapadão do São Francisco*). The additional spatial detail of the LSP archetypes can be seen as complementary information,

as spatially explicit phenological information was not considered for the definition of Cerrado ecoregions (Arruda et al. 2008). Both concepts are thus not mutually exclusive and might in combination be useful for spatially adapted conservation strategies. Especially, as recent studies have revealed significant relationships between phenological and floristic similarity of vegetation (Viña et al. 2016; Viña et al. 2012) the presented results may have the potential to improve biodiversity assessments.

5 Conclusion

Phenological archetypes of natural vegetation were proposed for the whole extent of the Cerrado. They were defined by similarities in land surface phenological metrics of the natural Cerrado vegetation, as derived from dense Landsat time series. For the first time land surface phenological patterns were revealed for the whole extent of the Cerrado with a spatial resolution of 30 m x 30 m. The presented results highlight the benefits of the free accessibility to the archive of homogenized Landsat data for the analysis of land surface phenology patterns in heterogeneous ecosystems. The LSP archetypes could, to some extent, be related to the main vegetation physiognomies of the Cerrado, but in contrast were not restricted to structural properties of the observed vegetation. As LSP archetypes are derived from dense time series they reflect the seasonal dynamic of the observed vegetation. They are thus not only related to the vegetation structure but also to species composition and other influencing environmental factors. However, at this stage it remains unclear to what extent the LSP archetypes are determined for example by structural parameters or species composition. This should be assessed in follow-up studies in combination with field reference data, for which Landsat's spatial resolution is sufficient. This finding is supported by the somewhat matching patterns of LSP archetypes and Cerrado ecoregions extents, which were defined by a combination of environmental variables that not only relate to structural vegetation properties. The two concepts are not mutually exclusive but can rather be seen as complementary approaches to develop spatially explicit conservation strategies. Future research should further focus on potential applications of the proposed approach, for example, in biodiversity assessments, or spatially explicit carbon quantifications.

Acknowledgements

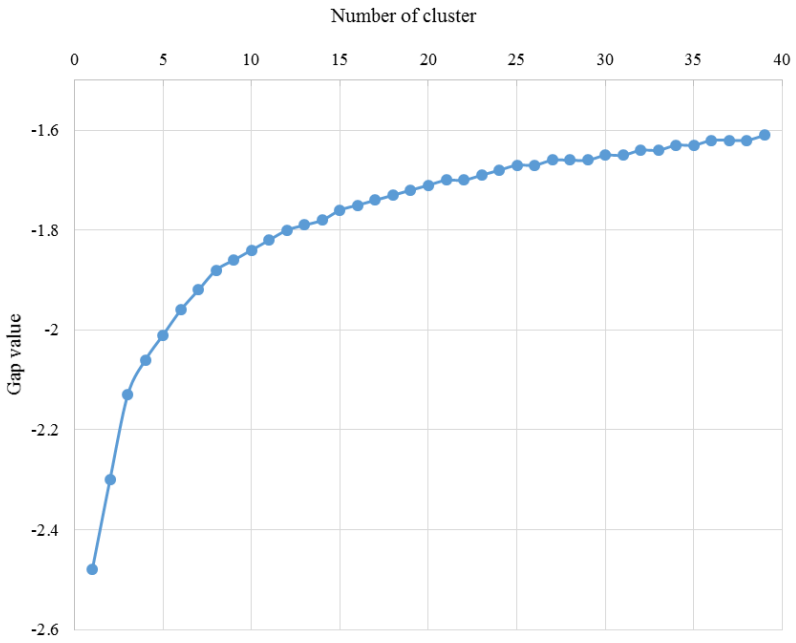
I would like to thank my colleague Andreas Rabe for his programming support that enabled the large scale analysis of this chapter, as well as Pedro Leitão and Philippe Rufin for

valuable meetings, comments and suggestions during the development and finalization of this chapter.

Supplementary Information

	Amp	BV	EoSVal	EoS	RoI	LoS	Mfit	MoS	RoD	SoSVal	SoS
Amp	1.00	-0.05	0.03	0.33	0.71	0.22	0.68	0.27	0.60	0.06	0.31
BV	-	1.00	0.97	0.22	0.01	0.29	0.70	0.11	-0.04	0.97	0.00
EoSVal	-	-	1.00	0.22	0.10	0.27	0.74	0.14	-0.01	0.90	0.04
EoS	-	-	-	1.00	0.21	0.86	0.39	0.82	0.07	0.26	0.69
RoI	-	-	-	-	1.00	0.08	0.52	0.06	0.37	0.05	0.29
LoS	-	-	-	-	-	1.00	0.38	0.55	-0.05	0.34	0.23
Mfit	-	-	-	-	-	-	1.00	0.28	0.40	0.75	0.22
MoS	-	-	-	-	-	-	-	-	0.27	0.13	0.79
RoD	-	-	-	-	-	-	-	-	1.00	0.05	0.20
SoSVal	-	-	-	-	-	-	-	-	-	1.00	0.02
SoS	-	-	-	-	-	-	-	-	-	-	1.00

S III-1: Correlation matrix showing Pearsons correlation coefficient for all pairs of the phenological metrics based on the initial random sample. Coefficients above the threshold of 0.70 are highlighted in red.



S III-2: Mean gap statistic plot after 100 iterations. The number of cluster on the x-axis and the derived gap values on the y-axis.

Cluster	Amp	BV	EoSVal	EoS	RoI	LoS	Mfit	MoS	RoD	SoSVal	SoS	Samples
1	0.26	0.34	0.37	57	321	0.03	0.60	33	118	0.36	20	71056
2	0.27	0.22	0.24	58	253	0.03	0.49	37	160	0.25	21	75938
3	0.16	0.43	0.44	51	141	0.03	0.58	30	103	0.44	14	60012
4	0.14	0.19	0.20	60	122	0.03	0.34	38	82	0.21	21	102861
5	0.17	0.29	0.31	58	162	0.03	0.46	35	83	0.31	19	95379
6	0.40	0.21	0.24	57	436	0.03	0.60	37	265	0.25	23	43437
7	0.18	0.37	0.39	66	171	0.03	0.54	48	122	0.37	32	18900
8	0.12	0.28	0.29	39	158	0.02	0.40	29	178	0.29	19	16581

S III-3: Mean values of LSP metrics per cluster. Derived from a random sample of natural vegetation pixel according to the TerraClass Cerrado classification.

Ecoregion	Natural [km²]	Non-natural [km²]	Sum [km²]	Natural areas [%]
1 Parecis	61896	18049	79945	77
2 Chiquitania	27289	113205	140493	19
3 Depressão Cuiabana	5480	1056	6536	84
4 Província Serrana	5250	7893	13144	40
5 Paraná Guimarães	20101	10831	30932	65
6 Complexo Bodoquena	118880	37447	156327	76
7 Bananal	12751	10159	22910	56
8 Araguaia-Tocantins	51585	47501	99085	52
9 Planalto Central Goiano	23384	14836	38220	61
10 Paranapanema Grande	74790	86469	161259	46
11 Bico do Papagaio	77766	47845	125611	62
12 Alto Parnaíba	90506	39717	130222	70
13 Depressão do Paranguá	18207	20743	38950	47
14 Chapadão do São Francisco	2432	452	2883	84
15 Vão do Paranã	50284	6773	57057	88
16 Paracatu	138118	91088	229206	60
17 Paranaíba	34534	42791	77324	45
18 Serra da Canastra	32058	11394	43452	74
19 São Francisco-Velhas	4772	2745	7517	63
20 Grão-Mogol	4170	2372	6542	64
21 Jequitinhonha	5196	9978	15174	34
22 Serra do Cipó	129372	242743	372115	35
	(sum) 988821	(sum) 866087	(sum) 1854908	53

S III-4: Pixel count per Cerrado ecoregion and the share of natural/non-natural areas based on the TerraClass Cerrado classification.

Chapter IV:
Landsat phenological metrics and their relation to aboveground carbon in the Brazilian Savanna

(In review) Submitted to Carbon Balance and Management (12/2017)[*]

Marcel Schwieder, Pedro J. Leitão, Jose Roberto Pinto, A. Teixeira, Fernando Pedroni, Maria Sanchez, Mercedes Maria da Cunha Bustamante and Patrick Hostert

[*] A revised version of the manuscript was meanwhile published in: Schwieder, M., Leitão, P.J., Pinto, J.R.R., Teixeira, A.M.C., Pedroni, F., Sanchez, M., Bustamante, M.M., & Hostert, P. (2018). Landsat phenological metrics and their relation to aboveground carbon in the Brazilian Savanna. Carbon Balance and Management, 13, 7.

Abstract

The quantification and spatially explicit mapping of carbon stocks in terrestrial ecosystems is important to better understand the global carbon cycle and to monitor and report change processes, especially in the context of international policy mechanisms such as REDD+ or the implementation of Nationally Determined Contributions (NDCs) and the UN Sustainable Development Goals (SDGs). Especially in heterogeneous ecosystems, such as savannas, accurate carbon quantifications are still lacking, where highly variable vegetation densities occur and a strong seasonality hinders consistent data acquisition. In order to account for these challenges we analyzed the potential of land surface phenological metrics derived from gap-filled 8-day Landsat time series for carbon mapping. We selected three areas located in different subregions in the central Brazil region, which is a prominent example of a savanna with significant carbon stocks that has been undergoing extensive land cover conversions. Here phenological metrics from the season 2014/2015 were combined with aboveground carbon field samples of cerrado sensu stricto vegetation using Random Forest regression models to map the regional carbon distribution and to analyze the relation between phenological metrics and aboveground carbon.

The gap filling approach enabled to accurately approximate the original Landsat ETM+ and OLI EVI values and the subsequent derivation of annual phenological metrics. Random Forest model performances varied between the three study areas with RMSE values of 1.64 t/ha (mean relative RMSE 30%), 2.35 t/ha (46%) and 2.18 t/ha (45%). Comparable relationships between remote sensing based land surface phenological metrics and aboveground carbon were observed in all study areas. Aboveground carbon distributions could be mapped and revealed comprehensible spatial patterns.

Phenological metrics were derived from 8-day Landsat time series with a spatial resolution that is sufficient to capture gradual changes in carbon stocks of heterogeneous savanna ecosystems. These metrics revealed the relationship between aboveground carbon and the phenology of the observed vegetation. Our results suggest that metrics relating to the seasonal minimum and maximum values were the most influential variables and bear potential to improve spatially explicit mapping approaches in heterogeneous ecosystems, where both spatial and temporal resolutions are critical.

1 Introduction

Terrestrial ecosystems play a pivotal role in providing regulating ecosystem services related to global and climate change (Scholes and Smart 2013). Photosynthesis and respiration processes of vegetation are the direct link between biosphere and atmosphere, stressing terrestrial ecosystems' importance in the global carbon cycle (Sabine et al. 2004). As natural or anthropogenic disturbances such as fires and land cover conversions alter ecosystem functions and eventually can turn carbon sinks into sources, it is crucial to monitor change processes and map related carbon stocks and the changes thereof. A better understanding of the carbon cycle is especially important due to the uncertainties of how vegetation will respond to a changing climate (Bonan 2008; Mitchard et al. 2013). In addition, accurate quantification of carbon stocks and related changes is essential for measuring and reporting schemes within the context of international climate policies such as the Reducing Emissions from Deforestation and Forest Degradation (REDD+) mechanism of the United Nations Framework Convention on Climate Change (UNFCCD) (Mitchard et al. 2013; Rodríguez-Veiga et al. 2017). National and regional stakeholders also need up-to-date information to support the NDCs and the UN SDGs as a mechanism to finance climate change adaptation related policies (Biermann et al. 2017). In order to map carbon stocks over large extents remote sensing data have been shown to be mandatory (Goetz et al. 2009; Houghton and Goetz 2008). During the last decades several approaches using active (Asner et al. 2012; Mitchard et al. 2009; Saatchi et al. 2007), passive (Avitabile et al. 2012) or both (Clark et al. 2011) remote sensing data types have proven sufficient accuracies for carbon quantification. This development has been catalyzed by the broad availability of improved remote sensing datasets and the evolution of cutting-edge data mining techniques for remote sensing data analysis (Goetz et al. 2009; Grace et al. 2014). The majority of these studies has focused on dense forest ecosystems, however, a large share of the terrestrial surface is rather characterized by ecosystems with gradual transitions in vegetation density, such as savannas. Globally, savannas cover approximately 20% of the land area (Lehmann et al. 2011) and even though they usually contain less carbon than dense forest ecosystems they are important carbon sinks (de Miranda et al. 2014) and cannot be neglected in global carbon cycle analyses (González-Roglich and Swenson 2016; Grace et al. 2014). This is specifically true considering recent trends of land conversions in savanna regions (e.g. Stephanie et al. 2014). A prominent example of these ecosystems is the Brazilian savanna, known as the Cerrado, which covers approximately 2 million km² or ca. 23% of Brazil's surface area (Ratter et al. 1997). It is characterized by diverse vegetation structure and density, strong seasonality and

fire events (Beerling and Osborne 2006; Oliveira-Filho and Ratter 2002). The Cerrado is has a high biodiversity with many endemic species (Francoso et al. 2016) and is, due to a weak conservation status and subsequently a loss of habitat, considered as one of the global biodiversity hotspots (Mittermeier et al. 2011; Myers et al. 2000). Large areas of the natural vegetation have already undergone tremendous land cover changes, leading to a share of approximately 60% of remaining natural vegetation, which is expected to further decline in the future (Ferreira et al. 2012). The combination of these factors directly impacts the link between the land surface and the atmosphere (Arantes et al. 2016), which is e.g. through processes such as photosynthesis and respiration reflected in the vegetation's phenology (Richardson et al. 2013). It emphasizes the role of the Cerrado in the carbon cycle, the need for accurate carbon quantifications (Ribeiro et al. 2011; Sano et al. 2010) and for a better understanding of phenology - carbon relations.

Similar to other savanna regions, the main challenges for remote sensing based carbon quantification in the Cerrado are related to the strong seasonality of rainfall, as cloud cover in the wet season hinders the frequent acquisition of optical imagery (Sano et al. 2007). During the last decades, several studies have shown the potential of multi-temporal remote sensing approaches and time series analysis to capture land surface phenology, based on high temporal resolution data from sensors such as e.g. AVHRR (Ferreira and Huete 2004; Franca and Setzer 1998) or MODIS (Arantes et al. 2016; Ratana et al. 2005) and also discussed its benefits for biomass estimation (Zhang and Ni-meister 2014). However, approaches based on high temporal resolution data usually lack the spatial resolution that is necessary to monitor heterogeneous and fragmented ecosystems, where reflectance measures are composed of spectral properties from several land cover types (Baccini et al. 2007; Melaas et al. 2013) and aboveground carbon might change at finer spatial scales than captured in spatial coarse resolution data (Avitabile et al. 2012). Recently, it has been shown that using Landsat data with its spatial resolution of 30 m can help to overcome this shortcoming. Avitabile et al. (2012) demonstrated Landsat's potential for aboveground biomass estimation in Uganda and their results suggest that adding phenological information from multi-temporal imagery could improve model performance by better discriminating vegetation types. In aboveground biomass models of Sudano-Sahelian woodlands, Karlson et al. (2015) identified the median of a dry season Landsat NDVI time series as one of the three most important variables. However, Landsat's relatively low temporal resolution with a revisit time of 16 days challenges deriving annual land surface phenology (LSP) that captures the whole growing season, especially in cloud prone areas (Zhang et al. 2017). At the same time,

the huge amount of freely accessible, archived data holds unexplored potential for ecosystem mapping and monitoring (Pasquarella et al. 2016).

First promising approaches to analyze dense Landsat time series exist, which are e.g. based on data pooling (Fisher et al. 2006; Melaas et al. 2013; Melaas et al. 2016) or gap filling approaches (Schwieder et al. 2016; Zhu et al. 2015b). We here aim to further exploit the approach proposed by Schwieder et al. (2016), based on the hypothesis that a link between annual dynamics of vegetation as captured in LSP metrics, the productivity of plants and the aboveground carbon stored in vegetation can be established. Our objectives thus are to i) investigate the potential to model aboveground carbon in a heterogeneous ecosystem based on Landsat-derived LSP metrics ii) assess the relation between these phenology metrics and aboveground carbon and iii) use these metrics to map the carbon distribution across different Cerrado landscapes.

2 Methods

2.1 Study areas and field data

The Cerrado stretches from around 2° to 25° South. Its elevation ranges from sea level to 1,800 m above sea level (Ratter et al. 1997), with most of the Cerrado being part of the Brazilian Central Plateau. Average annual precipitation ranges from 1,300 to 1,600 mm, with distinct dry (May to September) and wet seasons (October to April). With a mean temperature of 20.1°C, the Cerrado is classified as Aw climate after Köppen-Geiger, which is typical for savanna regions (Alvares et al. 2013; Hill et al. 2010). The well drained soils of the Cerrado are mainly dystrophic with rather high aluminum and iron contents (Ratter et al. 1997). These environmental factors can vary widely over the vast extent of the Cerrado, adding to the heterogeneity of the biome. Fire occurrence and long term climatic fluctuations further increase vegetation variability creating strong gradients in vegetation structure and density over space and time (Oliveira-Filho and Ratter 2002). The resulting mosaic of landscape formations ranges from open grasslands over shrub-dominated areas and scattered tree formations with grassland understory to dense forest patches. This mosaic is therefore classified in distinct physiognomy classes based on their respective vegetation height and tree cover (Ribeiro and Walter 2008). Different physiognomies are accordingly characterized by different biomass and thus also differ in their shares of stored above- and belowground carbon (de Miranda et al. 2014; Ottmar et al. 2001).

This study focuses on three areas in the Brazilian Savanna that are characterized as cerrado sensu stricto, which is the most abundant physiognomy in the remaining natural Cerrado vegetation (Sano et al. 2010). The cerrado sensu stricto may feature 20 to 50 % of tree cover and individual tree heights range from 3 to 6 m (Ribeiro and Walter 2008). All three sites lie within protected areas, i.e. anthropogenic land cover changes can be neglected for this analysis. Their spatial extents were defined by available field data. The most western of our three study areas is located near the border of the Brazilian federal states of Goiás and Mato Grosso close to the city of Barra do Garças (Figure IV-1 a). It lies within the borders of the Parque Estadual da Serra Azul (PESA). According to data from the Shuttle Radar Topography Mission (SRTM) the elevation in our study area ranges between 413 m and 771 m asl. The lower elevations are covered by dense semi-deciduous forests, whereas the surrounding higher areas are cerrado sensu stricto. The second study area covers parts of the Parque Estadual de Terra Ronca (PETR) (Figure IV-1 b). It is located near the city of São Domingos at the border of the states Goiás and Bahia (Teixeira et al. 2016). Here elevations vary between 617 m to 1013 m asl. The third study area is located in the North of Goiás state, close to the city of Alto Paraíso de Goiás within the Parque Nacional da Chapada dos Veadeiros (PNCV; Figure IV-1c). Elevations vary from 1,068 m to 1,267 m asl, with rocky outcrops at higher altitudes. Cerrado sensu stricto areas dominate in PNCV with a few gallery forest patches along water bodies (Pinto et al. 2009).

In all three study areas, field plots were established following sampling protocols based on the procedures for the permanent plots of the PPBio program (Magnusson et al. 2005) with adaptations to the Cerrado biome. Two parallel 5 km tracks (1 km apart) were defined, with five equally spaced 250 m x 40 m (10,000 m²) plots being staked out along each route. The longer dimension of the plots followed the contour of the terrain in order to avoid the possible effects of variations in altitude on its characteristics. Within these plots all trees with a minimum diameter at a height of 30 cm above the ground level of 5 cm were sampled between 2012 and 2014 (taxa, height, diameter). However, we excluded trees with a diameter smaller than 10 cm from the carbon analysis to make sure that we focus on vegetation with relative stable carbon stocks. Field plots that covered vegetation physiognomies other than cerrado sensu stricto were excluded from the analysis, resulting in 8 field plots each for PESA and PETR and 6 for PNCV (Figure IV-1 a-c). Aboveground carbon values for each sampled tree were calculated with a specific allometric equations (Rezende et al. 2006), which is considered representative for the cerrado sensu stricto physiognomy as it is based on a broad variety of 174 individuals sampled in Brasília, DF. The resulting carbon values

were spatially allocated to the 30 m spatial resolution of the Landsat grid following the approach proposed by Leitão et al. (submitted). In the first step the polygons that were sampled in the field are intersected with the Landsat grid. Polygons that fully lie within a Landsat pixel were randomly separated and used to estimate linear regression coefficients based on their individual carbon values measured in the field and a high resolution (5 m x 5 m) RapidEye vegetation index layer (REVI; Peng and Gitelson 2012). The best performing model coefficients were selected based on cross validation. Finally, the REVI layer and the estimated regression coefficients were used to spatially allocate and extrapolate, the carbon values measured in the field, to the unsampled areas of the intersecting Landsat pixel (Leitão et al. submitted).

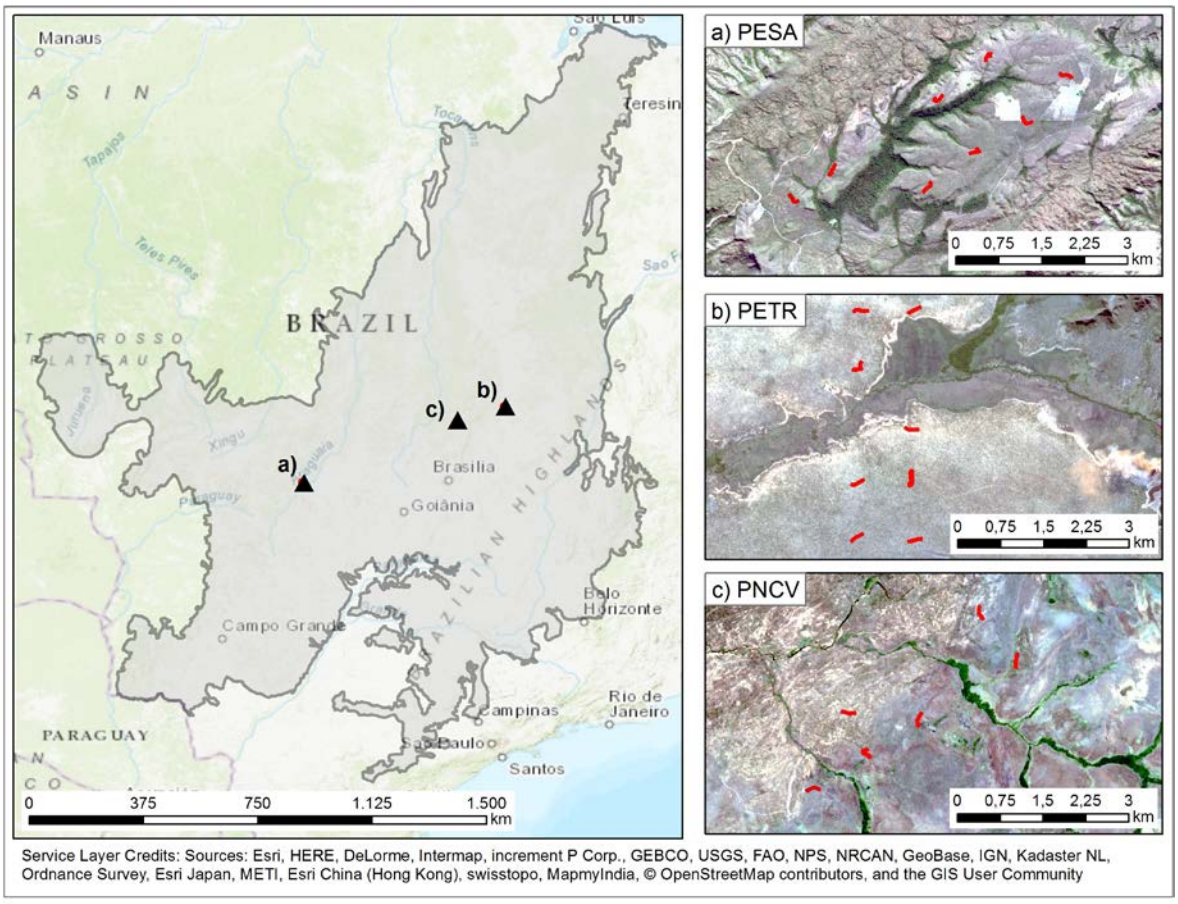


Figure IV-1: Locations of the three study areas, which are located within a) Parque Estadual da Serra Azul (PESA), b) Parque Estadual de Terra Ronca (PETR) and c) Parque Nacional da Chapada dos Veadeiros (PNCV) in the Brazilian Cerrado. The red polygons show the location of the field transects within the study areas with underlying true color Rapid Eye imagery.

2.2 Phenological metrics

Land surface phenology and related phenological metrics were derived from a combined Landsat ETM+ and OLI 8-day time series. Therefore, all available L1T corrected Landsat ETM+ and OLI surface reflectance data (path/row: PESA 224/071; PNCV 221/070; PETR

220/069) that were acquired between the beginning of 2014 and the end of 2015 (cloud cover < 90%) were downloaded along with their respective cfmask product (USGS 2017a). The enhanced vegetation index (EVI) was calculated based on equation (Eq. IV-1), as it is known to decouple the canopy background signal and to reduce atmospheric influences, while still being sensitive to high biomass / carbon values (Huete et al. 2002):

$$EVI = 2.5 * \frac{\rho NIR - \rho RED}{\rho NIR + 6 * \rho RED - 7.5 * \rho BLUE + 1} \quad \text{(Eq. IV-1)}$$

where ρ relates to the surface reflectance values in the respective Landsat bands covering the near infrared (NIR), red (RED) and blue (BLUE) wavelengths of the electromagnetic spectrum. Following Schwieder et al. (2016), a weighted ensemble of three radial basis convolution filters (RBF) with varying kernel widths (σ) was used to fill temporal data gaps in a vegetation index time series at pixel level. The RBF ensemble was applied, using temporal bins of eight days, to the period from the first of January 2014 to the 31 of December 2015, resulting in a total of 92 potential original Landsat observations within two years, to which available original data were assigned based on their respective acquisition dates. Outliers were excluded from the time series if they were more distant than one standard deviation to a convolution filter function with a kernel width of σ = 20 (Schwieder et al. 2016). Three convolution filters with kernel widths of σ = 8, σ = 16 and σ = 32 were subsequently used to fill the data gaps. The final time series profile is the combination of the three, whereas each filter is weighted based on the original data availability (Figure IV-2). To assess the deviation of the fitted RBF values from the original Landsat ETM+ and OLI EVI values, the RMSE was calculated for each sample pixel. Finally, the gap-filled time series were further processed in TIMESAT (Jönsson and Eklundh 2004) to derive LSP metrics. As it was not expected to detect more than one phenological season in the natural vegetation of the Cerrado, the seasonality parameter was set to 1 and the start/end of season were defined as the day of year when 20% of the seasons amplitude was reached. Additional to the Start of season (SoS) and End of season (EoS), the Mid of season (MoS) and Length of Season (LoS) were derived as day(s) of year from the time series. Further, the phenological metrics Base value (BV), Maximum fitted value (MfV), Amplitude (Amp) and Rate of increase (RoI) / decrease (RoD) were derived as EVI values. Details on the derived metrics are explained in Jönsson and Eklundh (2004).

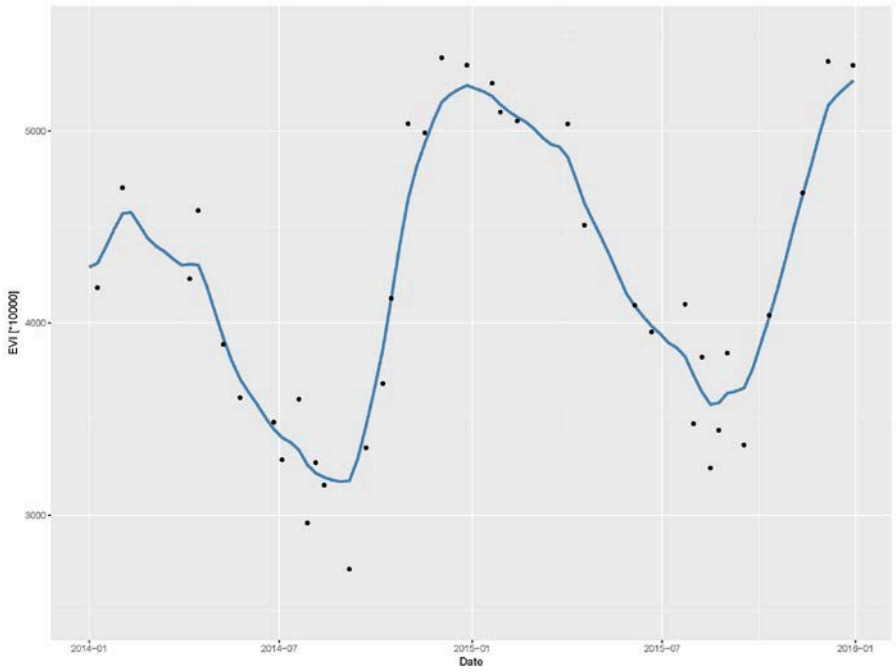


Figure IV-2: Phenological pixel profile after outlier detection and RBF fitting. The black points represent the original Landsat EVI values and the blue line the fitted RBF ensemble values within 8-day temporal bins.

2.3 Carbon models

To map carbon distributions as well as to analyze the relation between phenological parameters and aboveground carbon values, we used a Random Forest regression (RFR). RFR has been shown to adequately estimate carbon from remote sensing data that is usually not linearly related to metrics from satellite imagery (Lu et al. 2016). RFR is an ensemble approach, which is based on the Classification and Regression Tree algorithm (Breiman et al. 1984). Best splits of the training data are derived at each node using a subset of the input features. As single trees are assumed to be prone to errors, RFR builds many regression trees (i.e. a forest) from random subsets of the input data and validates the results on the withheld data. The final result of the regression is the averaged outcome of all regression trees (Breiman 2001).

The carbon model for each study area was iterated 1,000 times with randomly drawn subsets of 70% training and 30% validation data to derive performance measures and to obtain statistically robust results. Model performance was assessed using the Root Mean Square Error (RMSE), the relative root mean square error (relRMSE), defined as the ratio between RMSE and the mean trainings pixel carbon values, as well as the coefficient of determination (R^2). To derive the optimum number of sample pixels, considering the area of the pixel that was actually measured in the field a sensitivity analysis was performed for each study area (Leitão et al. submitted). Therefore, we executed our carbon models with subsets of the original sample data set based on the percentage of pixel area sampled in the field in 10 %

steps. The individual sample sets for each study area were subsequently filtered for further analysis, using the derived optimal thresholds. The final maps are based on the mean carbon predictions after 1,000 iterations. To further assess the relation between the phenological metrics and aboveground carbon, we evaluated the influence of the individual variables based on the RFR variable importance. It is derived by calculating the difference between the cross-validated model performance (out-of-bag mean square error; MSE) using all variables as model input and the performance of a model with permutated values within the respective variable, which enables a ranking of the most important variables by increase in MSE. The measures are scaled based on their respective standard errors (Liaw and Wiener 2002). Partial dependency plots (PDP) of the phenological metrics were created using the R-package pdp (Greenwell 2017). These plots allow analyzing the influence of each input variable on the response, by individually evaluating the Random Forest model based on the variations within one selected variable, while all other variables are fixed to their respective mean. However, as PDP's are useful to analyze the relation between carbon distribution and LSP metrics, but do not reveal the relations among the input variables, a principle component analysis (PCA) was performed based on the correlation matrix of the phenological metrics. Then the carbon values were plotted within the new feature space. Both analyses were derived for three RFR models based on all available samples of each study area, considering the respective sensitivity analysis threshold. The carbon models were built in the R environment (R Core Team 2017) using the tuneRF function of the randomForest R-package (Liaw and Wiener 2002) for automated model parameter optimization.

3 Results

During the season of interest (2014/15) a total of 46 Landsat ETM+ and OLI observations were acquired at 8-day intervals. Cloud coverage and sensor errors led to a reduced effective observation density in our study areas, which greatly differed between the dry and the wet season. On average, 23 (PESA), 21 (PNCV) and 26 (PETR) observations of our sample pixels were available for the whole season. During the dry season, an average of 13 (PESA), 16 (PNCV) and 17 (PETR) from 20 potential observations were available, in contrast to 10 (PESA), 5 (PNCV) and 9 (PETR) from 26 during the wet season (Table IV-1). The deviations between the fitted and the original EVI values resulted in an average RMSE of 0.018 in the PESA sample pixels, where fitted EVI values ranged around a mean of 0.344. Based on the PNCV samples the average RMSE was 0.010 and the fitted RBF EVI values had a mean of 0.238. In PETR the RMSE was 0.013 with a fitted EVI mean of 0.271. The mean allocated

aboveground carbon values were 5.47 t/ha in PESA, 3.66 t/ha in PNCV and 4.73 t/ha in PETR (Table IV-1). From these dense 8-days Landsat time series pixel-wise seasonal phenological metrics were derived for each study area, whenever TIMESAT recognized a full season. These metrics, which describe the course of the phenological profiles, enabled a standardized interpretability of the results and reduced the amount of model input variables from 46 EVI values to 9 phenological metrics per pixel.

Table IV-1: Average data availability from Landsat ETM+ and OLI observations within the sample pixels for the dry (May-September 2014) and wet (October 2014 – April 2015) season, relative to the amount of potentially available original observations. The mean RMSE values are based on the deviations between the fitted and the original EVI values for each study area.

	No. of samples	*Data availability dry season [%]*	*Data availability wet season [%]*	*Mean RMSE (min; max)*	*Mean RBF EVI (min; max)*	*Mean allocated Carbon [t/ha] (min; max)*
PESA	198	63	39	0.018 (0.012; 0.024)	0.344 (0.204; 0.447)	5.47 (0; 15.21)
PETR	207	85	36	0.013 (0.007; 0.018)	0.271 (0.176; 0.350)	4.73 (0; 20.56)
PNCV	165	81	18	0.010 (0.005; 0.016)	0.238 (0.183; 0.299)	3.66 (0; 14.63)

The spatial allocation approach, which was used to match field polygons and the phenological metrics pixel grid, led to a total of 198 sample pixels in PESA and 207 in PETR to be used as input for the carbon models. As some of the data in the 165 PNCV pixel were too noisy to derive phenological parameters, 15 pixels were excluded from the carbon models. The regression coefficients used for spatial allocation were 2.810 (PESA), 3.973 (PNCV) and 2.695 (PETR).

The sensitivity analysis revealed that carbon model performance generally decreased and was less stable during 1000 iterations, when fewer samples were included in the model. In terms of relative RMSE the models performed best with thresholds of around 0.1. Thus all pixels in which less than 10 % was sampled in the field were excluded from further analysis.

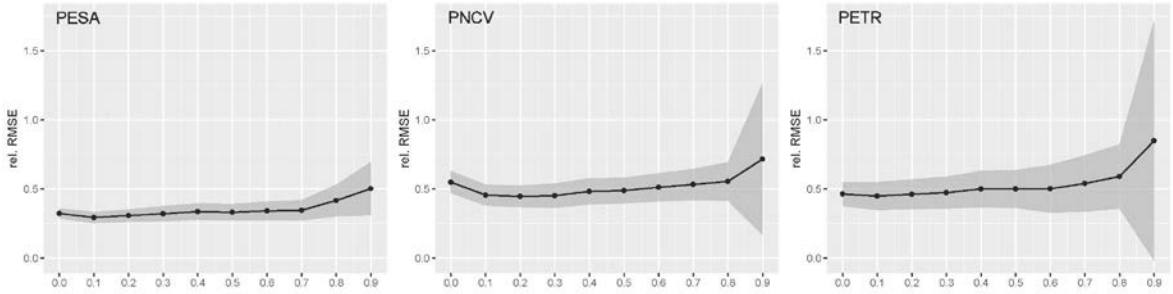


Figure IV-3: Results of the regression model sensitivity analysis for each study area. Relative RMSE after 1000 model runs are shown for each threshold, while the grey ribbons relate to +/- one standard deviation.

Carbon model performances differed between the three study areas with averaged R^2 values of 0.69 for PESA, 0.43 for PETR and 0.36 for PNCV and a similar trend in RMSE values of 1.65 t/ha (relative RMSE 0.30) in PESA, 2.18 t/ha (relative RMSE 0.45) in PETR and 2.35 t/ha (relative RMSE 0.46) in PNCV (Table IV-2).

Table IV-2: Averaged model performance measures (R^2 and RMSE) and related standard deviations after 1,000 iterations, along with the average descritptive statistics of the carbon measures (t/ha).

	Threshold	*Number of samples*	*Mean R^2*	*R^2std*	*Mean RMSE*	*RMSE std*	*Mean relRMSE*	*Carbon min*	*Carbon mean*	*Carbon max*
PESA	0.1	145	0.70	0.06	1.64	0.20	0.29	0.44	5.59	14.92
PETR	0.1	163	0.44	0.14	2.18	0.44	0.45	1.07	4.88	19.71
PNCV	0.1	90	0.36	0.11	2.34	0.34	0.46	0.42	5.16	14.22

Despite the differences in model performance, the variable importance ranking was largely stable across models, with BV and MfV being ranked as first or second most influential variables across all study areas (Figure IV-4). Ranks of further variables varied between the study areas, with higher standard deviations in PNCV and PETR. Especially in PETR the individual variable importance was small with comparably high standard deviations (Figure IV-4).

The partial dependency plots of important phenological metrics (Figure IV-5) highlight the relation between the phenological metrics and carbon distributions. BV and MfV are positively correlated to carbon values throughout all study areas. While BV steadily increases with increasing carbon values even in data sparse regions, saturations occur in MfV when carbon values approximate around 5.7 to 6.7 t/ha. Even though carbon value distributions vary among the study areas, the relation between BV and carbon is similar. The relation between MoS and carbon is comparable in PESA and PNCV with lower carbon values being associated to a later peak of season. Even though the trend is not as clear in PETR, it remains comparable to the other regions until mid of season around DOY 54.

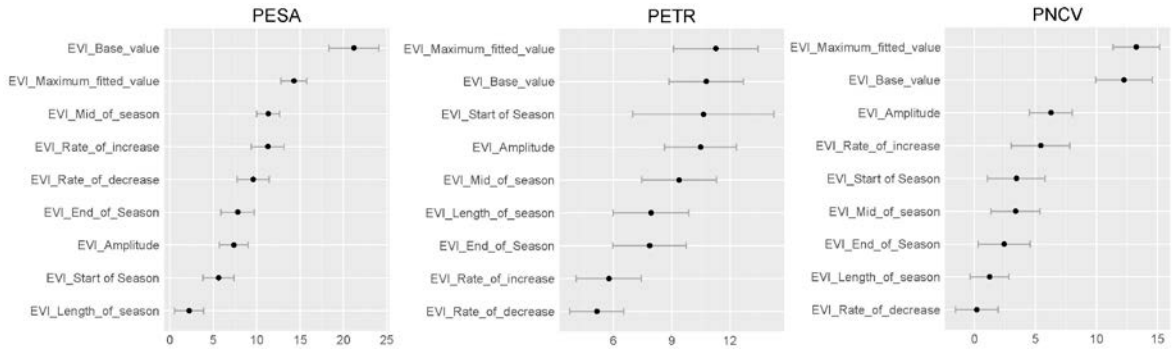


Figure IV-4: Mean variable importance measures for the three study areas after 1,000 model iterations. The values are scaled using their respective standard errors. Horizontal bars indicate standard deviations.

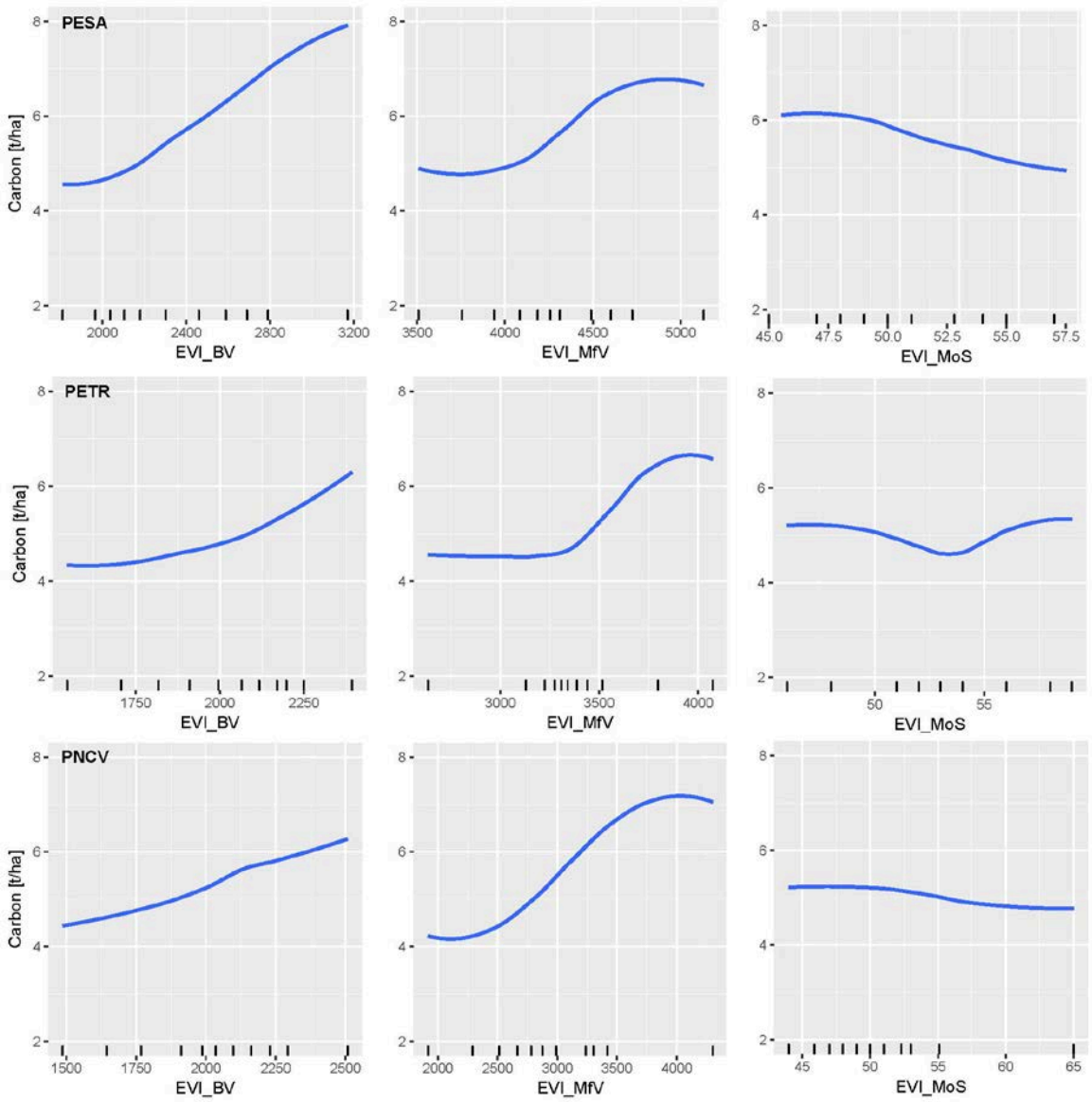


Figure IV-5: Selected smoothed partial dependency plots (PDP) for each study area. Plots are derived by validating a RFR model using all available samples of each study area. The additional ticks on the x-axis mark the min/max and decile values of the input variable. BV and MfV are shown in EVI * 10000, MoS are shown as 8-day temporal bins starting from 01/01/2014. PDP's of all variables are shown in the supplementary material (S IV-1-3).

Additional to the analysis of the relations between carbon and individual phenological metrics, the principal component analysis (PCA) reveals the distribution of carbon values within an uncorrelated variable space (Figure IV-6). In PESA the first two axes explain 66% of the variance within the phenological metrics. The first axis is defined by EVI-related values such as BV and MfV, as opposed to metrics related to the timing of the season such as MoS and SoS. The latter are related to smaller carbon values, whereas BV and MfV are associated with larger values. The second axis is defined by the negative correlation between temporal metrics (EoS and LoS) and EVI related values (RoD and Amp). In PNCV the first

two axes of the PCA explain 67% of the metrics' variation. Here the negative correlation between EVI metrics (Amp, BV, MfV) and seasonal timing (SoS) is revealed, where at a cluster of rather large carbon values is oriented towards MfV, BV and Amp. The second axis is defined by the negative correlation between LoS, EoS and RoD, where average carbon values cluster. In the case of PETR, the first two axes of the PCA explain 69 % of the variance. The first axis is defined by the temporal metrics MoS and SoS and their negative correlation with RoI and MfV, with the latter being associated to rather larger carbon values. The second axis is defined by RoD on one side and BV on the other side, with rather equally distributed carbon values at both ends.

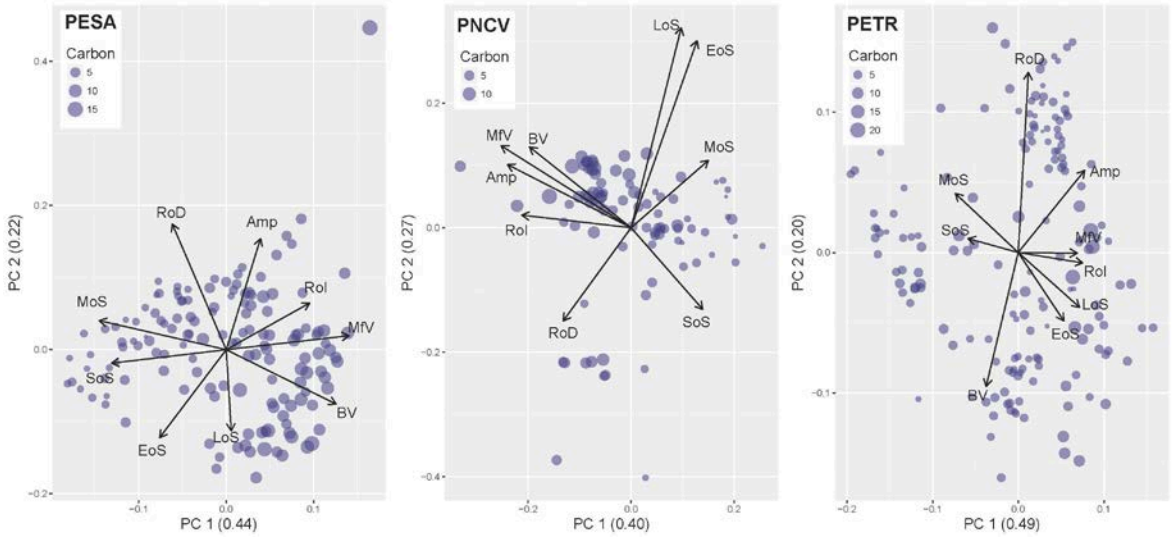


Figure IV-6: Plot of the first two axes of the PCA of the phenological metrics for each study area (numbers in brackets report the explained variance within the respective principal component). The angle between the arrows depict approximately the correlation between the phenological metrics. The points mark the carbon values within the new variable ordination space, while their size refers to the original carbon values.

The carbon distribution was mapped based on the mean of 1,000 model predictions throughout the selected study areas (Figure IV-7). The final RF models explained 68 % (PESA), 37 % (PETR) and 49 % (PNCV) of the withheld variance on average. The predicted carbon values range between 1.8 and 11.8 t/ha in PESA, between 2.4 and 13.1 t/ha in PETR and between 1.3 and 9.6 t/ha in PNCV. Standard deviations of up to 1.7 (PESA), 2.7 (PETR) and 2.2 (PNCV) mainly relate to predictions for areas with high carbon values.

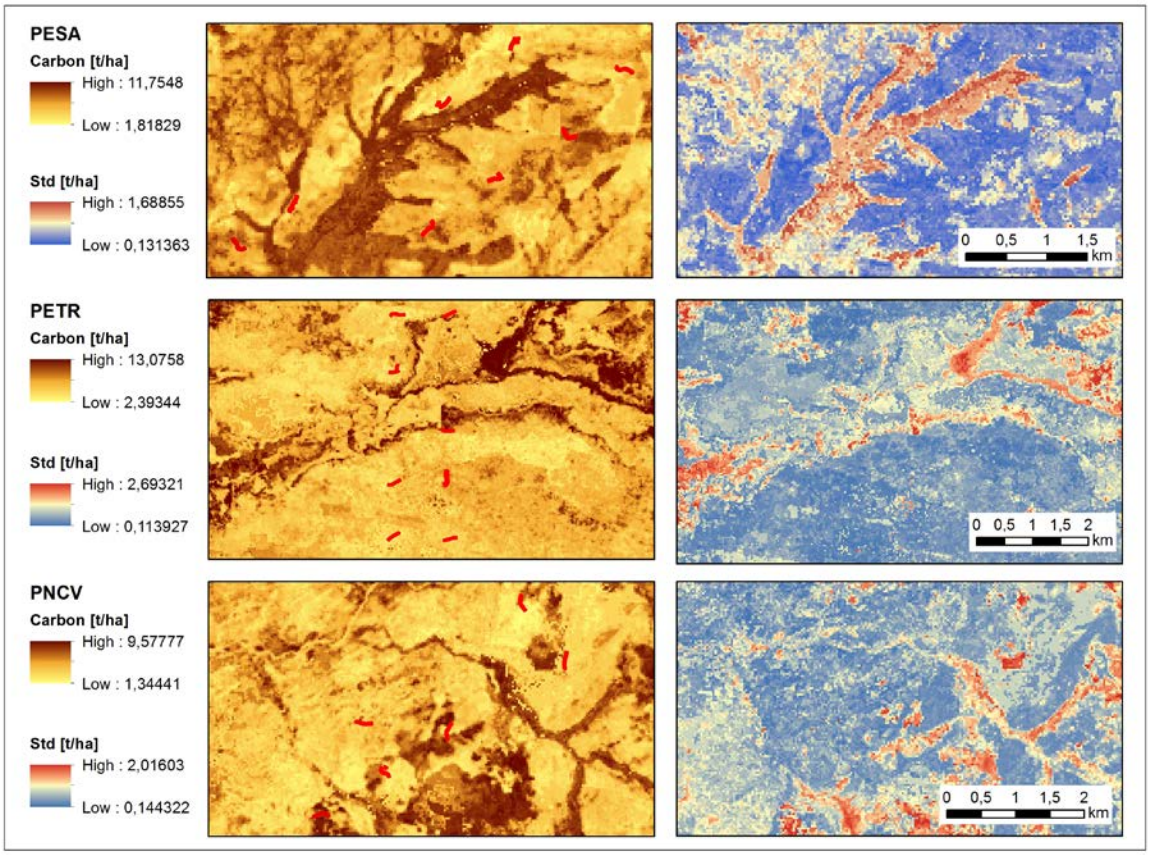


Figure IV-7: Left: Carbon maps for the study areas PESA, PNCV and PETR based on the mean predictions of 1,000 individual Random Forest regression models for each study area along with the sampling transects in red. Right: Corresponding standard deviation maps.

4 Discussion

Spatially explicit quantification of aboveground carbon distributions is essential to monitor and understand ongoing changes in carbon stocks. In heterogeneous landscapes, such as typical for the Cerrado, the spatial and temporal resolution at which we monitor the ecosystem play a pivotal role for our ability to map carbon. We therefore analyzed the potential of annual LSP metrics, derived from dense 8-day Landsat time series with a spatial resolution of 30 m x 30 m, for the spatially explicit quantification of aboveground carbon. Using all available satellite observation from Landsat enabled us to assign field based carbon values to LSP metrics from Landsat time series that describe the seasonal changes of the monitored vegetation as captured in the EVI.

Our results indicate that the RBF filter ensemble was able to capture the seasonal profile of the original observations, with minor deviations from the original EVI values. Using all available cloud-free data from Landsat provides a maximum number of observations as the basis for fitting the RBF and accordingly data gap fill values represent the best possible temporal interpolation. However, the regression model performances reveal a clear relation

between the LSP metrics derived from RBF-fitted values and the carbon distributions, with similar behavior for the three study areas. It could be observed that along with a decrease in the number of samples led to weak and less robust model performances, i.e. an increase in relRMSE and the related standard deviations. Our results show that the spatially allocated carbon values differ between the three study areas with highest values in PESA, followed by PETR and PNCV. This trend is also reflected in the distribution of EVI values and the respective phenological metrics and follows the gradient from overall lower (PESA) to higher (PNCV) elevations above sea level between the observed areas.

Although the model performances vary between the study areas they are still within a range that is comparable to results from studies dealing with the mapping of carbon in similar ecosystems. For example, González-Roglich and Swenson (2016) used, among other variables, products derived from Landsat data to map carbon distributions in Argentinian savanna regions. Their spatially explicit predictions, based on field samples ranging around a mean of 27 t/ha, had a mean prediction error of 9.6 t/ha (relRMSE 35%) at 60 m spatial resolution. Karlson et al. (2015) reported model performances with an R^2 of 0.57 and RMSE of 17.6 t/ha (relRMSE 66%) when mapping aboveground biomass in Sudano-Sahelian woodlands using multi-temporal (7 observations) Landsat OLI products. Along with the abovementioned findings our results stress the complexity and challenges of carbon mapping approaches in heterogeneous savanna systems. Here, a critical issue is the influence of understory vegetation (grasses and shrubs) on the spectral signal throughout a season. On the one hand, we accounted for these variations by interpreting LSP metrics that describe a full annual season, as EVI values are known to be sensitive to structural variations in vegetation canopies as the canopy background is decoupled from the signal (Huete et al. 2002). But even though the RBF ensemble approach accounts for data availability through adjusted weights, the variations in available observations between the dry and wet season still have an influence on the phenological profiles and ultimately the derived metrics. This is a limitation that in the future is likely to be overcome by the integration of additional data from other sensors, such as Sentinel-2 (Drusch et al. 2012). On the other hand, our samples only considered woody vegetation with a diameter of at least 10 cm. Thus, the carbon values lack shares of smaller trees and the non-woody vegetation layer, which would require a more frequent field sampling design to capture their dynamics. Reported carbon values are therefore conservative, in the sense that we will rather under- than overestimate carbon stocks.

Even though the employed allometric model for deriving carbon values from the field measurements is representative for the most abundant plant families within our study areas (*Vochysiaceae* and *Fabaceaea*), we found differences between the three study areas in species composition and abundance. Especially, species that were prominently abundant in PNCV and PETR (e.g. *vellozia squamanta*, *virtella ciliata*) were not considered for deriving the allometric model presented in Rezende et al. (2006). This might influence model performances and stresses the importance for refined allometric equations in further research. Despite these restrictions in our models, we showed to our best knowledge for the first time the benefits of using LSP metrics with a 30 m x 30 m spatial resolution for carbon modelling, which in contrast to e.g. raw time series of vegetation indices simplifies further analysis of the relation between aboveground carbon and LSP.

Especially metrics that are directly related to the amount of vegetation, such as base value and maximum fitted value, have the potential to explain carbon variations, as they were ranked as first or second most important in our regression models. This pattern is also revealed by the PCAs that on the one hand show correlations among the LSP metrics but also reveal relationships between them and carbon values. In all three study areas, but especially prominent in PESA, are clusters of lower carbon values associated to high values of variables related to seasonality, such as SoS or MoS. This suggests that vegetation with higher carbon densities might be related to an earlier start of season, which is here defined as the point in time when 20% of the ascending part of the phenological profile is reached, as well as an earlier peak of season. A possible explanation might be the leaf producing strategy of some of the (semi-) deciduous species, for which a high activity of leaf production could be observed at the end of the dry season (Lenza and Klink 2006; Pirani et al. 2009), causing an earlier green-up in the phenological profile. Similar phenological patterns have been observed in the western part of the Sudanian Savanna, where a later start of season was observed in LSP for areas with higher shares of herbaceous than woody vegetation (Gessner et al. 2015).

The final carbon maps show comprehensive spatial patterns with e.g. high carbon values along the riparian vegetation (gallery forests) and the dense forest patches in the lower elevations in PESA. Vegetation patterns in very high resolution imagery from RapidEye suggest that the region's heterogeneity is very well reflected in the spatial variations of the mapped carbon patterns. The spatially explicit quantification captures the landscape composition and highlights the benefits of Landsat's spatial resolution for estimating carbon across different study areas in the Cerrado.

The range of mapped carbon values is in line with values found in the literature. Carbon values for cerrado sensu stricto are for example summarized in Ribeiro et al. (2011), as well as in Vourlitis and da Rocha (2010) and range between 3.3 and 32.5 t/ha (mean 8.5 t/ha) or 5.0 to 15.9 t/ha (mean 9.7 t/ha), respectively, depending on the regional focus and the methods used. However, due to a lack of additional reference data, the maps could not be independently validated and especially estimates for physiognomies that were not included in the training samples (such as grasslands and dense forests areas) need to be regarded with caution. Especially, the carbon-dense areas, e.g. gallery forests and forest patches (e.g. seasonal forest and Cerradão) are associated with the highest standard deviations, as they are model extrapolations and our models will most likely underestimate carbon stocks in these areas. However, as the carbon maps reveal comprehensible patterns, it is highlighted that phenological metrics derived from freely available remote sensing data, are a valuable contribution to carbon mapping approaches, providing spatially explicit knowledge for environmental managers and policy makers in support of sustainable development policies related to REDD+ or the UN SDG's.

5 Conclusions

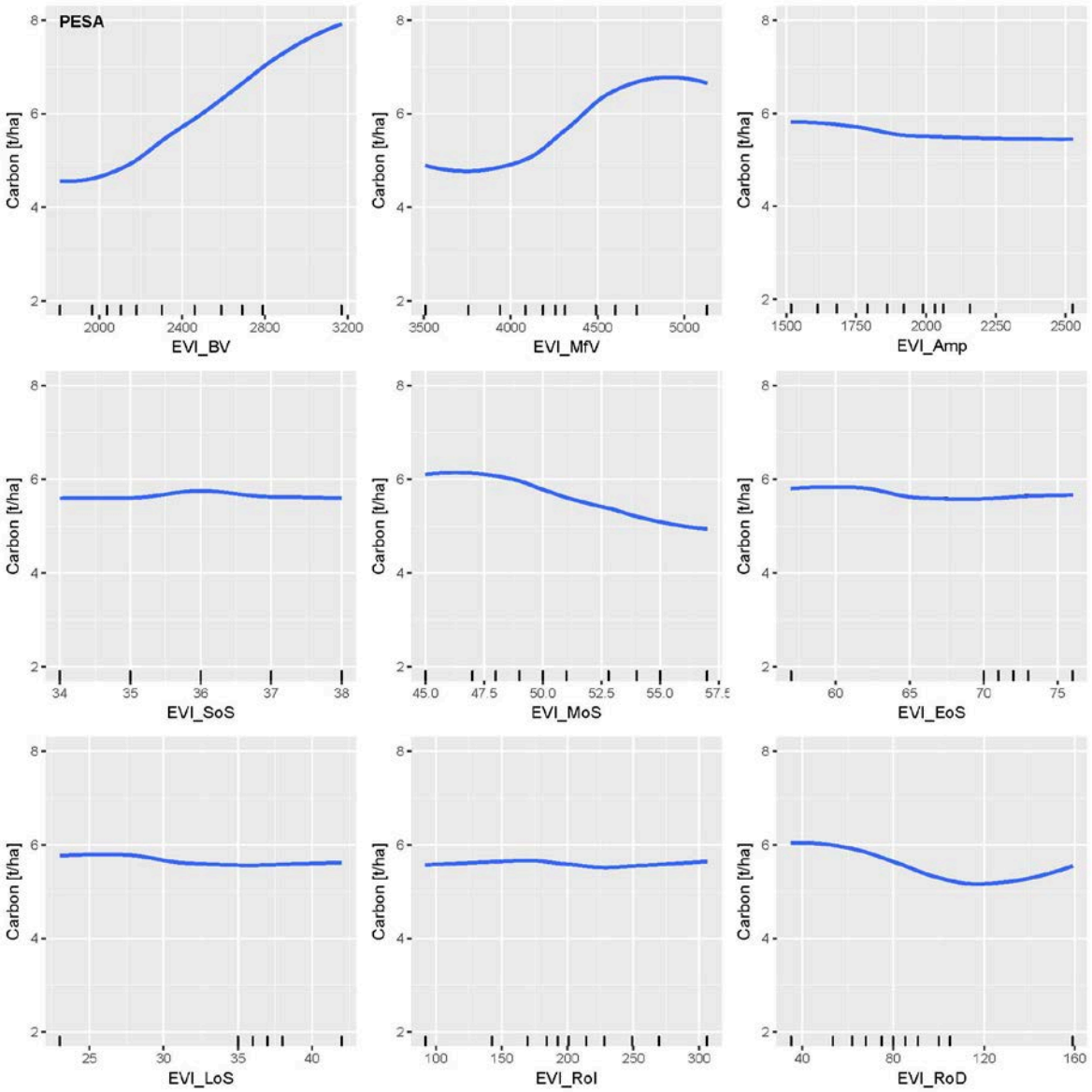
Based on gap-filled 8-day Landsat EVI time series, we derived annual land surface phenology (LSP) metrics for cerrado sensu stricto vegetation. The derived metrics enabled us to reduce the amount of input data for a following Random Forest regression analysis, while preserving the information needed to approximate LSP. They further facilitated determining a relationship to the aboveground carbon distribution, with an adequate spatial resolution for mapping gradual vegetation transitions in heterogeneous savanna ecosystems, such as the Cerrado. Our results are comparable to those of similar studies and we successfully identified the relation between the seasonal behavior of cerrado sensu stricto vegetation and its carbon distribution. Metrics that are instantly linked to amounts of vegetation such as Base value and Maximum fitted value have been shown to be important for such a mapping approach. Metrics relating to the timing of phenological events, such as start or mid of season showed a weaker relation and were not consistently relevant for carbon mapping. We were able to map carbon distributions within the selected study areas, whereat higher uncertainties were identified in physiognomies and related carbon values that were not well represented by the field sampling. We verified that Landsat based annual LSP metrics are beneficial variables to analyze carbon – phenology relations and for the spatially explicit quantification of aboveground carbon in heterogeneous ecosystems such as the

Cerrado. In order to improve large-scale carbon mapping efforts, our findings stress the need for representative sampling strategies, along with subsequent improvement of allometric equations, which together reflect the variability within the observed savanna vegetation gradient. Further research should investigate the potential of mapping approaches that synergistically combine phenological metrics with variables related to the vertical structure of vegetation (such as lidar or radar) and analyze the influence of additional data (e.g. Sentinel-2) on the accuracy of the derived phenological metrics.

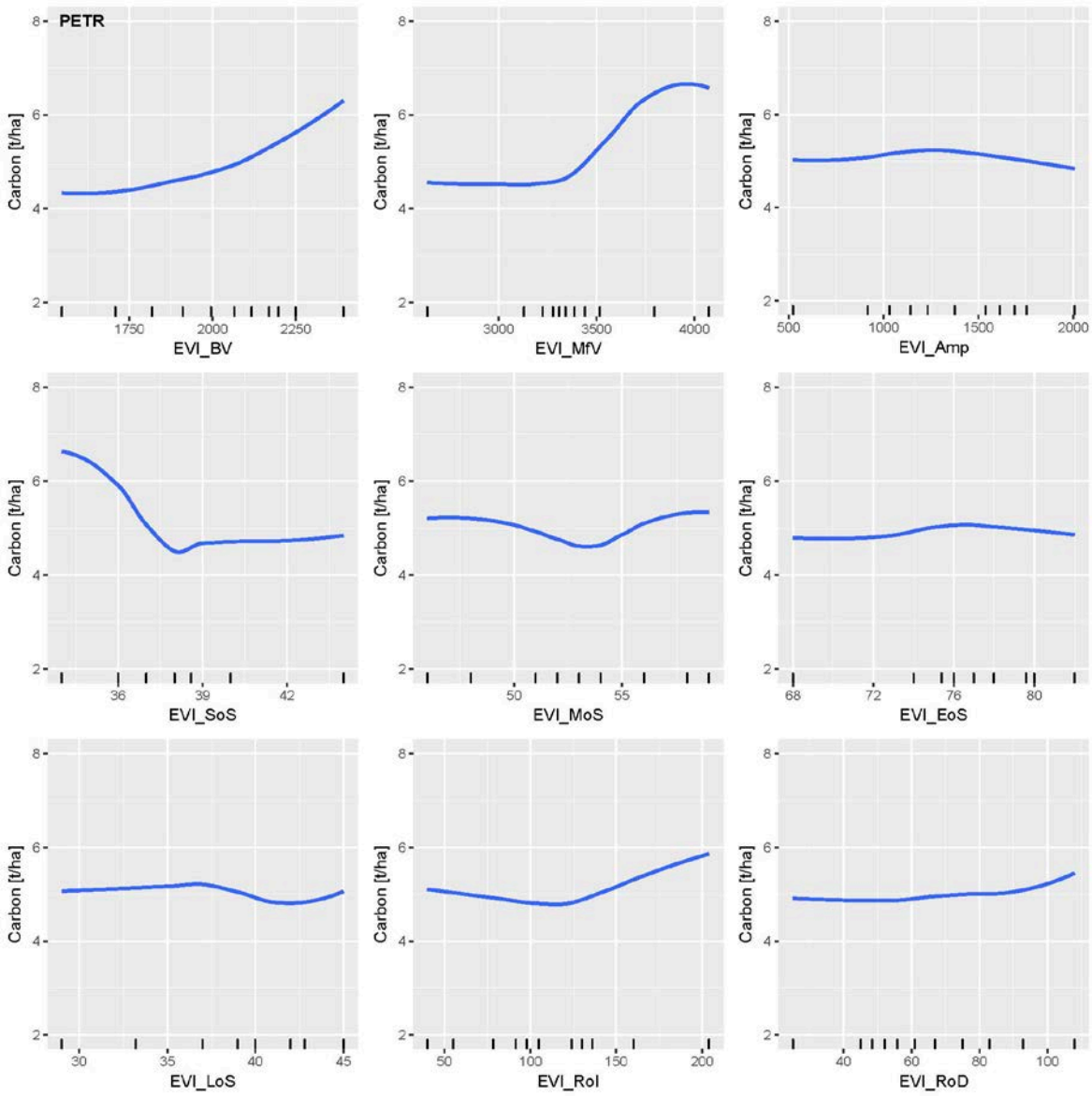
Acknowledgements

We thank Andreas Rabe for optimizing and translating the RBF convolution filter code into Python and for his help during remote sensing data processing. Florian Poetzschner supported the field data consolidation and preparing the spatial allocation code. Special thanks to Philippe Rufin for very helpful suggestions and inspiring conversations regarding the presented research. This research contributes to the 2012-2017 Landsat Science Team (https://landsat.usgs.gov/2012-2017-science-team).

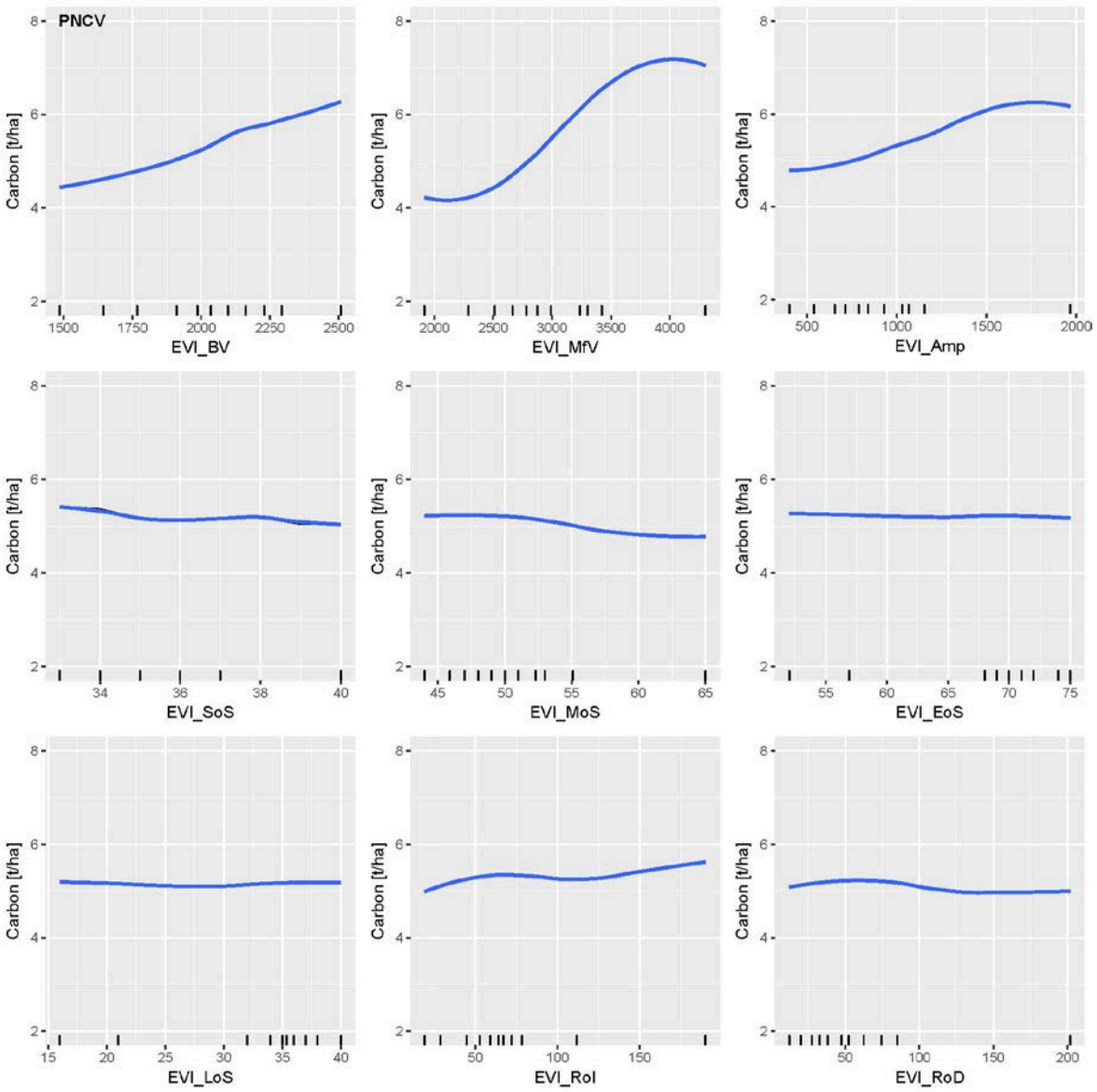
Supplementary Information



S IV-1: All partial dependency plots for PESA (Serra Azul State Park, Brazil) for RFR models based on all available samples using the threshold 0.1. Metrics that relate to index values are shown in EVI * 10000. Metrics related to time are shown as 8-day temporal bins starting from 01/01/2014.



S IV-2: All partial dependency plots for PETR (Terra Ronca State Park, Brazil) for RFR models based on all available samples using the threshold 0.1. Metrics that relate to index values are shown in EVI * 10000. Metrics related to time are shown as 8-day temporal bins starting from 01/01/2014.



S IV-3: All partial dependency plots for PNCV (Chapada dos Veadeiros National Park, Brazil) for RFR models based on all available samples using the threshold 0.1. Metrics that relate to index values are shown in EVI * 10000. Metrics related to time are shown as 8-day temporal bins starting from 01/01/2014.

Chapter V:
Synthesis

1 Summary

The beginning of the Anthropocene marks the start of an era in which the impacts of humans on environmental systems reached a planetary scale (Crutzen 2002; Steffen et al. 2007). Approximately half of the Earth's terrestrial surface has been altered by humans, with severe impacts on Earth system processes, such as the global carbon, water, or nitrogen cycles (Steffen et al. 2005). Some of the planetary boundaries, which were proposed to set a safe operating space for human societies, have already been crossed or are in a zone of uncertainty, one of which is *Land-system change* (Rockström et al. 2009; Steffen et al. 2015). A hotspot of land use change processes is the Cerrado, one of Brazil's largest biomes expanding over more than 2 Mkm² and accounting for approximately 24 % of Brazil's terrestrial surface. It provides a wide range of ecosystem services (Lima et al. 2017) and functions, such as carbon sequestration and the provision of habitats, which are of national and even international importance due to the Cerrado's very rich biodiversity (Mendonça et al. 2008) and it's huge extent. A growing demand for land resources, technological advances, a weak conservation status, and political incentives for the development of the Cerrado (Klink and Machado 2005; Klink and Moreira 2002), have led to large scale land conversions resulting in a remaining share of around 60% of the Cerrado's natural vegetation (Sano et al. 2010). Expected growing demands on agricultural products, which is considered as a direct driver of land use change in Brazil (Lapola et al. 2013), stress the need to design policies and conservation strategies that allow a sustainable land use, fulfilling the demands for agricultural products, while at the same time maintaining key ecosystem services (Strassburg et al. 2017). Spatially explicit information of the Cerrado's vegetation and its seasonal behavior is therefore necessary to deepen our understanding of the ecosystems dynamics and their response to global change processes. Remote sensing based approaches enable to frequently map ecosystems over large extents and even inaccessible terrain and are thus an essential tool for ecosystem monitoring (DeFries et al. 2005). The advent of new operational spaceborne sensors, along with open data policies and technical developments in terms of data processing and storage capacities, facilitate the analysis of unprecedented amounts of remote sensing data (Wulder et al. 2015b). In combination with state-of-the-art time series techniques these data have the potential to be beneficial for the analysis of heterogeneous ecosystems, where spatial, spectral and temporal resolution matter.

The main goal of this thesis was to analyze the benefits of dense Landsat time series to capture the annual seasonal response patterns of natural vegetation. Aiming to reveal benefits and limitations of the approach for mapping natural vegetation and its phenological patterns across large scales, as well as for carbon quantification approaches, in heterogeneous ecosystems. The main methodological approach underlying all analyses is a radial basis convolution filter (RBF), which enables to fill data gaps in time series of combined Landsat ETM+ and OLI vegetation indices. This resulted in gap free 8-day Landsat time series, which were further used to derive land surface phenological (LSP) metrics that describe the seasonal response of the observed vegetation.

The following section summarizes the main findings of the thesis's core chapters (II – IV) alongside the three main research questions, which are introduced in more detail in chapter I.

Research Question I: Can land surface phenological metrics be used to differentiate the main vegetation physiognomies in the Cerrado?

The focus of chapter II was to introduce and test a gap filling approach for its applicability in differentiating the main vegetation physiognomies of the Cerrado. These differ in their structural composition, their vegetation density and hence in their amount of stored carbon, rendering them an appropriate proxy for carbon estimations. To compare their seasonal response patterns concerning LSP metrics, Landsat top-of-atmosphere images that cover three study areas around Brasília, DF, were downloaded and transformed to tasseled cap greenness, brightness and wetness values. After removing clouds and outliers from the time series, the resulting data gaps have been filled based on radial basis convolution filters. This approach is comparable to a moving average filter, whereas the resulting filtered value is weighted based on a Gaussian distribution with a specific kernel width, rather than a simple average. An ensemble of three kernel widths (integrating 19, 31 or 43 potential observations of the time series) was used, to derive values that are representative for dense as well as for sparse data situations. The final values are the average of the three kernels, weighted by the data density of clear observations within each kernel. Based on reference maps and high resolution Google Earth imagery, a set of training samples was derived for each of the main Cerrado vegetation physiognomies. Direct comparisons of the phenological profiles derived from the time series already revealed trends in the seasonal response of the distinct physiognomies. The main result of this chapter was that, on average, dense physiognomies with high trees and more closed canopies, like *cerradão* or *cerrado denso,* tend to have an earlier start of season and less variation throughout the season in comparison to

physiognomies such as *cerrado sensu stricto* or *campo sujo*. Similar seasonal response patterns could be observed in all three tasseled cap components greenness, brightness and wetness. From the tasseled cap time series land surface phenology (LSP) metrics were derived, which enabled a more direct interpretation and comparison of the seasonal response. Again clear trends between the physiognomies seasonal response, in terms of the distribution of LSP metrics, were recognizable. Especially metrics such as Base value and Rate of decrease showed clear differences between the physiognomies for example with lower Base values for sparse and higher for dense physiognomies. The derived LSP metrics were further used to train and apply a Support Vector Classification. The resulting map showed comprehensible spatial patterns of the main physiognomies in comparison to available reference maps. However, a quantitative accuracy assessment showed that the main confusion was between adjacent classes, due to mixed pixel problems and non-independent physiognomy classes. Leading to the conclusion, that hard classification approaches are problematic in heterogeneous ecosystems, where the vegetation describes a rather continuous gradient in density and structure.

After this initial test concerning the applicability of Landsat derived LSP metrics and their use to reveal differences in the vegetation's seasonal response, chapter III investigated the spatial distribution of LSP metrics across the entire extent of the Cerrado.

Research Question II: What are the spatial patterns of LSP metrics in the natural vegetation across the entire Cerrado?

To gain insights into the spatial distribution of vegetation with a similar seasonal response, LSP metrics were analyzed for the whole extent of the Cerrado. These metrics were derived from a dense gap filled 8-day Landsat ETM+/OLI time series of EVI Collection 1 products for the season 2014 - 2015. A total of approximately 23,000 EVI products in 121 WRS-2 tiles were acquired and processed. This was for the first time that spatially explicit LSP metrics were derived for the whole extent of the Cerrado with a 30 m spatial resolution. In order to focus the analysis on areas of the remaining natural vegetation, all other land use classes were masked from the LSP map based on the TerraClass Cerrado classification (MMA 2015). A map of selected LSP metrics (Amplitude, Base Value, Start of Season) revealed regions of vegetation with similar seasonal response across the Cerrado. However, for the interpretation of all LSP metrics they were clustered based on an unsupervised k-means algorithm. This resulted in a total of eight clusters of LSP similarities, which were proposed as LSP archetypes of Cerrado vegetation. Based on previous findings from chapter

II, relations between the LSP archetypes and Cerrado vegetation physiognomies were assumed. Even though a visual comparison of the LSP archetypes with high resolution Google Earth imagery revealed somewhat matching patterns, directly assigning vegetation physiognomies to LSP archetypes is not straightforward. While the physiognomy classes are defined by vegetation structural parameters (Ribeiro and Walter 2008), LSP metrics reflect the seasonal dynamics of the observed vegetation as captured in EVI variations. Thus, they do not solely relate to the vegetation's structure but also to the species composition and other environmental factors that determine the vegetation's seasonal response.

A first assessment of the archetype patterns was performed by comparing the revealed patterns with Cerrado ecoregions. These were defined in order to set conservation priorities and their extents are defined by natural boundaries, determined by various biotic and abiotic factors. Some of the ecoregions showed a good match with regions of similar LSP archetypes, while in other regions areas of contiguous archetypes crossed the boarders of several ecoregions. One outstanding example was the Bananal ecoregion, which is characterized by flat plains, rivers and interfluvials, in which the spatial distribution of LSP archetypes revealed more than relations to structural parameters. As ecoregions were defined without spatial explicit information on the vegetation's phenology, the two concepts might complement each other for the spatial explicit adaptation of conservation and land management strategies.

However, at this stage it cannot be assessed to which extent LSP archetypes are determined by vegetation structure, species composition or other influencing factors. This should be investigated in follow-up studies with adequate (field) reference data, for which the spatial resolution of 30 m is sufficient. As a first step in this direction the relationship between LSP metrics and above ground carbon (AGC) was assessed in chapter IV.

Research Question III: What is the relationship between Landsat based land surface phenological metrics and above ground carbon?

In the fourth chapter of this thesis, LSP metrics were derived from gap-filled L1T corrected Landsat ETM+/OLI EVI time series for three study areas distributed in the central Cerrado. Based on the hypothesis that EVI time series reflect the seasonal dynamics of photosynthetic biomass, the relationship between LSP metrics and AGC was assessed. To overcome the limitations of hard classifications, a regression approach was used in order to improve the accuracy of spatially explicit carbon quantifications. Extensive fieldwork was carried out in the three selected study areas that are characterized by the *cerrado sensu stricto*

physiognomy. Parameters that enable the estimation of AGC stored in vegetation were measured, along defined field plot transects. Using allometric equations, AGC values were calculated on an individual tree level and aggregated to each field plot with a size of approximately 10 m x 20 m. To spatially allocate the field plot measures to the pixel level they were overlaid with a 30 m x 30 m grid and an additional high resolution (RapidEye) vegetation index layer. Based on derived regression coefficients, AGC values for unsampled pixel regions were extrapolated (see Appendix A). Random Forest regression models were used to analyze the relation between LSP metrics and AGC, for each study region individually. Overall model performances varied between the three study areas, but were still comparable to accuracies reported in other studies, stressing the complexity of AGC estimations. The results revealed that metrics that are directly linked to vegetation density such as Maximum fitted value and Base value were the most important variables to describe AGC throughout all study areas, whereas the importance of metrics related to the timing of the seasons, differed between the areas. The regression models allowed mapping the spatial distribution of AGC in the three study areas, which revealed comprehensible patterns when compared to high resolution imagery. Higher uncertainties could be observed in areas that were not represented in the field samples. Overall, the analysis led to the conclusion that LSP metrics derived from dense Landsat time series are valuable variables for the spatially explicit quantification of AGC.

2 Main conclusions

This thesis aimed at assessing the benefits of land surface phenological metrics, derived from dense Landsat time series, for the characterization of vegetation in a heterogeneous savanna ecosystem, in terms of its phenological diversity and patterns of above ground carbon. The results presented in this thesis allow to draw the following main conclusions.

Combining acquisitions from the Landsat ETM+ and OLI sensors with state of the art gap filling techniques enable large scale analyses of seasonal vegetation dynamics in the Cerrado.

The overall Landsat data availability was sufficient to generate equidistant time series for the Cerrado, even though the data acquisition during the wet season is frequently hampered due to clouds (chapter IV). Data gaps, which were due to cloud contamination and sensor errors could be filled by applying an ensemble of radial basis convolution filters. This gap-filling approach allowed to generate equidistant time series at 8 - day intervals and 30 m

spatial resolution covering the entire extent of the Cerrado. This was facilitated by the Landsat archive consolidation (Wulder et al. 2015b) and especially the between sensor data harmonization (USGS 2018). The latter includes precise geometric correction and across sensor calibration in Collection 1 data (USGS 2017b), enabling the combined usage of data from different Landsat sensors for time series analysis (USGS 2018). The results presented in the three core chapters did not provide evidence for lacking inter-sensor calibration or major differences between the vegetation indices derived from both sensors. Further, the results indicated that not only the bare amount of clear observations is important to appropriately capture the vegetation dynamics but that observations distributed around the seasonal peak are vital for characterizing Cerrado vegetation.

Deriving land surface phenology (LSP) metrics from dense time series enables the interpretation of intra-annual vegetation dynamics.

Dense Landsat time series reflect the seasonal dynamics of the observed vegetation in individual vegetation index values in 8 - day intervals, complicating their direct interpretation, especially over large extents. Even though the fitted time series values may be used as input for classification or regression approaches, they do not allow to reveal and interpret relationships between the dependent and independent variables. Deriving LSP metrics, however, facilitates the comparison and interpretation of resulting spatial patterns, as they describe the seasonal response of the observed vegetation. Hence, LSP metrics enable further contextualized analyses in an ecologically meaningful way.

Discrete classification approaches of vegetation physiognomies pose challenges in complex ecosystems such as the Cerrado.

The results of chapter II show that LSP metrics as derived from Landsat time series are beneficial for the mapping of physiognomies on a 30 m spatial resolution. However, the class-wise accuracy assessment indicates problems arising from non-independent “hard” classes in a vegetation gradient, defined by rather arbitrary boundaries (Oliveira-Filho and Ratter 2002). In an ecosystem, which is characterized by environmental gradients, this leads to mixed pixel problems, distorting the classification results with the main confusion being between adjacent classes. Landsat LSP metrics are thus, to a certain level of thematic detail, beneficial for the mapping of natural vegetation patterns with a 30 m spatial resolution, but map accuracies are strongly dependent on the defined class boundaries.

The LSP archetypes offer potential to enhance existing vegetation mapping concepts

Based on the results from chapter II and the resulting spatial patterns of the LSP archetypes presented in chapter III, the assumption can be made that LSP metrics derived from Landsat capture the environmental gradient with higher thematic and spatial detail than is considered in current physiognomy maps. While physiognomy classes mostly relate to the structural diversity of the vegetation, LSP archetypes capture the general phenological diversity of the observed vegetation communities. Even though they relate to the vegetation's structure, LSP metrics describe the phenological response of the vegetation, which is also influenced by species composition as well as other environmental factors, such as soil conditions or terrain.

LSP metrics are beneficial for spatially explicit carbon quantifications

Seasonal dynamics of photosynthetic active vegetation are captured in LSP metrics. The results of chapter IV revealed the relationship between LSP metrics and AGC, which enabled the spatially explicit estimation of AGC. In particular metrics that relate to vegetation index values at the start, peak, and end of season had the most influence in the regression models. LSP metrics should thus be considered as valuable input variables for direct carbon quantification approaches, which are not restricted by pre-defined physiognomy classes that do not allow to account for intra-class variability.

3 Implications

The proposed analysis of dense Landsat time series revealed a large potential for enhancing ecosystem monitoring approaches in the Cerrado. Currently, maps of Cerrado vegetation physiognomies are widely used, for example, to assess conservation status (Pinto et al. 2009) as well as to gain knowledge of vegetation dynamics and related impacts on biogeochemical cycles (Sano et al. 2010). Even though these maps are very accurate, they are largely based on image segmentation and visual interpretation. In comparison to Landsat based classification approaches, they lack the spatial detail that accounts for the heterogeneity of the Cerrado landscape, even within the individual vegetation physiognomies (MMA 2017).

Combining Landsat ETM+ and OLI data with a gap-filling approach and TIMESAT enabled the spatially explicit analysis of LSP patterns and facilitated to analyze their potential to benefit vegetation monitoring and carbon quantifications on test sites in the Cerrado. The software package TIMESAT was originally developed for the analysis of coarse scale AVHRR data (Jönsson and Eklundh 2002) and later adapted for the use of other remote

sensing data sets with high and equidistant observation frequency, such as MODIS. However, TIMESAT is still not widely used for the analysis of data with a medium spatial resolution such as Landsat, due to its relatively low temporal resolution. The gap-filling approach used, enabled to overcome this limitation and revealed for the first time the potential of Landsat data to capture phenological diversity, as expressed by LSP metrics, with a 30 m spatial resolution. This has substantial implications for the future use of time series of medium spatial resolution satellite imagery, as it allows for the analysis of seasonal vegetation dynamics over large extents. Due to the temporal depth of the Landsat data archive these dynamics can even be derived for the past decades, enabling the analysis of long-term LSP trends. However, the availability and distribution of clear observations is a limiting factor of the approach, which needs to be kept in mind when analyzing long-term trends.

The proposed method to create spatially explicit phenological information from Landsat data is potentially beneficial for a range of applications, such as biodiversity assessments, for which the use of phenological information is widely discussed. Even though phenological field observations, that directly relate to events such as bud burst, flowering, seeding, etc., differ substantially from LSP measures, the latter are considered as an important spectral trait for the quantification of biodiversity (Lausch et al. 2016). LSP has also been proposed as a candidate essential biodiversity variable relating to ecosystem functioning, in order to track progress towards Aichi Biodiversity Targets (O'Connor et al. 2015; Skidmore et al. 2015). Recent results presented by Viña et al. (2016) revealed for example significant relationships between LSP and floristic similarities and underpinned the use of Landsat spatial resolution over coarse scale approaches using MODIS for biodiversity monitoring. Their findings also highlight that the vegetation index used should be appropriate for the intended application. The radial basis convolution filter approach is a generic algorithm that can be used to fill data gaps independent of spectral bands or indices of interest and is thus broadly applicable and offers a strong transferability potential.

Since this thesis focused, among others, on improving carbon mapping approaches, the EVI was used as it is robust towards saturation in high biomass vegetation. EVI based LSP metrics reflect the seasonal dynamics of photosynthetic activity, resembling a direct link between the biosphere and the atmosphere, and were hence considered as an appropriate variable for improving carbon estimation approaches. Usually, official carbon or biomass estimations that are used for Brazil´s national reports in the framework of REDD+ or National GHG Inventories of the UNFCCC, are based on vegetation physiognomy maps to

which rather static carbon values are assigned (MCT 2016; MMA 2017). These estimates are thus generalizations that do not account for detailed spatial heterogeneity, which is particularly relevant in the Cerrado. The results of chapter IV revealed the relation between LSP metrics and AGC as well as the benefits for its spatially explicit quantification. Even though model performances varied between the study areas, LSP metrics enabled to quantify AGC without being reliant on physiognomy classes and associated restrictions. They are thus appropriate variables to refine spatially explicit carbon estimations, an important prerequisite for future mapping approaches.

4 Outlook

LSP metrics derived from dense Landsat time series have been shown to be valuable for characterizing vegetation properties in heterogeneous ecosystems. The robustness of the derived LSP metrics was assessed through the validation of the individual mapping and regression approaches, but was not compared to actual in-situ phenological patterns. The discrepancy between LSP and on-ground phenology is a well-known problem and an active field of research (Coops et al. 2012; Rodriguez-Galiano et al. 2015). It has thus to be kept in mind that the patterns and dynamics captured in LSP do not reflect the actual phenological behavior of individual plants, but rather the seasonal dynamics of vegetation communities. Follow-up studies should further investigate, which of the environmental factors that influence the vegetation's phenology, are actually reflected in the individual LSP metrics.

Despite this drawback, the approach has the potential to be transferred to other savanna regions, which cover 10-30% of the global terrestrial surface (Viergever et al. 2008) and act as major carbon sources and sinks (Pellegrini et al. 2016). However, capturing the intra-annual dynamics of vegetation requires sufficient data density as well as an appropriate seasonal distribution of clear observations, which might be hampered in some regions of the world and is thus a limitation of the approach. Integrating data from additional optical sensors to the time series can help to overcome such limitations, for instance low observation density in the wet season due to cloud contamination. For example, the combined use of Landsat and Sentinel - 2 a/b (Drusch et al. 2012) data leads to an improved temporal resolution of 2 – 4 days (Claverie et al. 2016). Harmonization that allows for combining data from the Landsat and Sentinel - 2 sensors is currently an active field of research (e.g. Flood 2017), as the data need for example to be corrected for spatial misregistrations (Storey et al. 2016), BRDF effects and band pass differences (Claverie et al. 2016; NASA 2018).

Future research should further analyze the effects of the combined use of LSP metrics with complementary remote sensing data beyond the optical domain, to fully exploit the potential of the entire suite of operational spaceborne sensors (Wulder et al. 2015a). For example, data or derived products from Radar or Lidar sensors have the ability to capture the vertical structure of vegetation, which is beneficial for carbon estimation approaches (eg. Asner et al. 2012; Bouvet et al. 2018; Mitchard et al. 2009). Freely accessible Radar data from Sentinel - 1 (Torres et al. 2012) and the forthcoming missions BIOMASS (Le Toan et al. 2011), or the spaceborne Lidar ICESAT - 2 (Abdalati et al. 2010), will facilitate the combined use with LSP metrics.

Georeferenced field data that account for the variability in vegetation and environmental conditions are necessary, to enable spatially explicit carbon estimations over large extents. The provided LSP archetypes map of the Cerrado could be used as a valuable tool for an appropriate field sampling design. There is further a need to refine allometric equations, which are used for deriving vegetation carbon contents on an individual tree level, and hence are critical for the accuracy of carbon estimations (Chave et al. 2004). These challenges for improving mapping and monitoring of ecosystems over large extents, stress the need for research collaborations between ecologists and remote sensing scientists. Therefore, the exchange among the research domains and open access to field data as well as remote sensing data must be enabled (Bustamante et al. 2016). Recent trends in data policies and the advent of peer reviewed journals that strive to enhance the communication among research communities indicate a positive trend in this direction (Pettorelli et al. 2014).

Joint efforts will increase the robustness of carbon quantification approaches, as well as biodiversity assessments (Bustamante et al. 2016). Both are critical to deepen our understanding of ecosystem dynamics and responses to global change processes. This thesis contributed to these efforts by highlighting the benefits and limitations of LSP metrics for the spatially explicit characterization of Cerrado vegetation. Facilitated through technological advances and the availability of unprecedented amounts of high quality remote sensing data, the approach can be transferred to other regions of the world. It may thus be used to support the design of sustainable land use policies and conservation strategies that fulfill the demands of a growing world population, while at the same time maintaining key ecosystem services and functions to stay within safe planetary boundaries.

References

Abdala, G. (2008). Plano de Ação para Prevenção e Controle do Desmatamento na Amazônia Legal (PPC-DAm) Documento de avaliação 2004-2007. *Brasília, DF: Ministério do Meio Ambiente*

Abdalati, W., Zwally, H.J., Bindschadler, R., Csatho, B., Farrell, S.L., Fricker, H.A., Harding, D., Kwok, R., Lefsky, M., & Markus, T. (2010). The ICESat-2 laser altimetry mission. *Proceedings of the IEEE, 98*, 735-751

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Goncalves, J.L.D., & Sparovek, G. (2013). Koppen's climate classification map for Brazil. *Meteorologische Zeitschrift, 22*, 711-728

Arantes, A.E., Ferreira, L.G., & Coe, M.T. (2016). The seasonal carbon and water balances of the Cerrado environment of Brazil: Past, present, and future influences of land cover and land use. *ISPRS Journal of Photogrammetry and Remote Sensing, 117*, 66-78

Arima, E.Y., Barreto, P., Araújo, E., & Soares-Filho, B. (2014). Public policies can reduce tropical deforestation: Lessons and challenges from Brazil. *Land Use Policy, 41*, 465-473

Arruda, M.B. (2003). Representatividade ecológica com base na biogeografia de biomas e ecoregiões continentais do Brasil. O caso do bioma Cerrado. In, *Instituto de Ciências Biológicas* (p. 178). Brasília, DF, Brazil: Universidade de Brasília

Arruda, M.B., Proença, C.E.B., Rodrigues, S.C., Campos, R.N., Martins, R.C., & Martins, E.S. (2008). Ecorregiões, unidades de conservação e representatividade ecológica do bioma Cerrado. In S.M. Sano, S.P. de Alemeida, & J.F. Ribeiro (Eds.), *Cerrado: ecologia e flora* (pp. 230-272). Brasilia, DF, Brazil: Embrapa Cerrados

Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In, *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035): Society for Industrial and Applied Mathematics

Asner, G.P., Mascaro, J., Muller-Landau, H.C., Vieilledent, G., Vaudry, R., Rasamoelina, M., Hall, J.S., & van Breugel, M. (2012). A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia, 168*, 1147-1160

Assunção, J., Gandour, C., & Rocha, R. (2015). Deforestation slowdown in the Brazilian Amazon: prices or policies? *Environment and Development Economics, 20*, 697-722

Avitabile, V., Baccini, A., Friedl, M.A., & Schmullius, C. (2012). Capabilities and limitations of Landsat and land cover data for aboveground woody biomass estimation of Uganda. *Remote Sensing of Environment, 117*, 366-380

Baccini, A., Friedl, M.A., Woodcock, C.E., & Zhu, Z. (2007). Scaling field data to calibrate and validate moderate spatial resolution remote sensing models. *Photogrammetric Engineering and Remote Sensing, 73*, 945-954

Beerling, D.J., & Osborne, C.P. (2006). The origin of the savanna biome. *Global Change Biology, 12*, 2023-2031

Biermann, F., Kanie, N., & Kim, R.E. (2017). Global governance by goal-setting: the novel approach of the UN Sustainable Development Goals. *Current Opinion in Environmental Sustainability, 26–27*, 26-31

Blackard, J.A., Finco, M.V., Helmer, E.H., Holden, G.R., Hoppus, M.L., Jacobs, D.M., Lister, A.J., Moisen, G.G., Nelson, M.D., Riemann, R., Ruefenacht, B., Salajanu, D., Weyermann, D.L., Winterberger, K.C., Brandeis, T.J., Czaplewski, R.L., McRoberts, R.E., Patterson, P.L., & Tymcio, R.P. (2008). Mapping US forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment, 112*, 1658-1677

Boisvenue, C., Smiley, B.P., White, J.C., Kurz, W.A., & Wulder, M.A. (2016). Improving carbon monitoring and reporting in forests using spatially-explicit information. *Carbon Balance and Management, 11*, 23

Bonan, G.B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science, 320*, 1444-1449

Borma, L.d.S., Da Rocha, H., Cabral, O., Von Randow, C., Collicchio, E., Kurzatkowski, D., Brugger, P., Freitas, H., Tannus, R., & Oliveira, L. (2009). Atmosphere and hydrological controls of the evapotranspiration over a floodplain forest in the Bananal Island region, Amazonia. *Journal of Geophysical Research: Biogeosciences, 114*

Bouvet, A., Mermoz, S., Le Toan, T., Villard, L., Mathieu, R., Naidoo, L., & Asner, G.P. (2018). An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR. *Remote Sensing of Environment, 206*, 156-173

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, 5–32

Breiman, L., Friedman, J., Stone, C.J., & Olshen, R.A. (1984). *Classification and Regression Trees*. Monterey, Calif., USA: Wadsworth, Inc.

Burivalova, Z., Bauert, M.R., Hassold, S., Fatroandrianjafinonjasolomiovazo, N.T., & Koh, L.P. (2015). Relevance of Global Forest Change Data Set to Local Conservation: Case Study of Forest Degradation in Masoala National Park, Madagascar. *Biotropica, 47*, 267-274

Bustamante, M.M.C., Roitman, I., Aide, T.M., Alencar, A., Anderson, L.O., Aragão, L., Asner, G.P., Barlow, J., Berenguer, E., Chambers, J., Costa, M.H., Fanin, T., Ferreira, L.G., Ferreira, J., Keller, M., Magnusson, W.E., Morales-Barquero, L., Morton, D., Ometto, J.P.H.B., Palace, M., Peres, C.A., Silvério, D., Trumbore, S., & Vieira, I.C.G. (2016). Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. *Global Change Biology, 22*, 92-109

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature, 486*, 59-67

CBD (2010). Decision X/2, The Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets. Nagoya, Japan, 18 to 29 October 2010

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology, 2*, 1–27

Chave, J., Condit, R., Aguilar, S., Hernandez, A., Lao, S., & Perez, R. (2004). Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 359*, 409-420

Chavez Jr., P.S. (1996). Image-based atmospheric corrections - revised and improved. *Photogrammetric Engineering & Remote Sensing, 62*, 1025-1036

Chen, C., Zhao, S.H., Duan, Z., & Qin, Z.H. (2015). An Improved Spatial Downscaling Procedure for TRMM 3B43 Precipitation Product Using Geographically Weighted Regression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*, 4592-4604

Clark, M.L., Roberts, D.A., Ewel, J.J., & Clark, D.B. (2011). Estimation of tropical rain forest aboveground biomass with small-footprint lidar and hyperspectral sensors. *Remote Sensing of Environment, 115*, 2931-2942

Claverie, M., Masek, J.G., & Ju, J. (2016). Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide. Online available. Access date: 04.03.2018. https://nex.nasa.gov/nex/static/media/publication/HLS.v1.0.UserGuide.pdf

Cocks, T., Jenssen, R., Stewart, A., Wilson, I., & Shields, T. (1998). The HyMapTM airborne hyperspectral sensor: the system, calibration and performance. In, *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy* (pp. 37-42): EARSeL

Cohen, W.B., & Goward, S.N. (2004). Landsat's Role in Ecological Applications of Remote Sensing. *BioScience, 54*, 535-545

Coops, N.C., Hilker, T., Bater, C.W., Wulder, M.A., Nielsen, S.E., McDermid, G., & Stenhouse, G. (2012). Linking ground-based to satellite-derived phenological metrics in support of habitat assessment. *Remote Sensing Letters, 3*, 191-200

Crutzen, P.J. (2002). Geology of mankind. *Nature, 415*, 23

Datt, B., McVicar, T.R., Van Niel, T.G., Jupp, D.L.B., & Pearlman, J.S. (2003). Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Transactions on Geoscience and Remote Sensing, 41*, 1246-1259

de Faria, R.A.P.G., Coelho, M.d.F.B., de Figueiredo e Albuquerque, M.C., & de Azevedo, R.A.B. (2012). *Phenology of Species in the Cerrado of Mato Grosso State, Brazil-Brosimum Gaudichaudii Trécul (Moraceae)*. INTECH Open Access Publisher

de Miranda, S.d.C., Bustamante, M., Palace, M., Hagen, S., Keller, M., & Ferreira, L.G. (2014). Regional Variations in Biomass Distribution in Brazilian Savanna Woodland. *Biotropica, 46*, 125-138

DeFries, R., Pagiola, S., Adamowicz, W.L., Akcakaya, R.H., Arcenas, A., Babu, S., Balk, D., Confalonieri, U., Cramer, W., Falconi, F., Fritz, S., Green, R., Gutierrez-Espeleta, E., Hamilton, K., Kane, R., Latham, J., Matthews, E., Ricketts, T., & Xiang Yue, T. (2005). Analytical Approaches for Assessing Ecosystem Condition and Human Well-being. In

R.M. Hassan (Ed.), *Ecosystems and human well-being: current states and trends* (pp. 37–71). Washington, D.C, USA: Island Press

Dinerstein, E., Olson, D.M., Graham, D.J., Webster, A.L., Primm, S.A., Bookbinder, M.P., & Ledec, G. (1995). *A conservation assessment of the terrestrial ecoregions of Latin America and the Caribbean*. Washington, DC, USA: World Bank, World Wildlife Fund

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment, 120*, 25-36

Dymond, C.C., Mladenoff, D.J., & Radeloff, V.C. (2002). Phenological differences in Tasseled Cap indices improve deciduous forest classification. *Remote Sensing of Environment, 80*, 460-472

EB (2018). Encyclopædia Britannica. Carbon. Access date: 01.02.2018. https://www.britannica.com/science/carbon-chemical-element

Ehrlich, D., Estes, J.E., & Singh, A. (1994). Applications of NOAA-AVHRR 1 km data for environmental monitoring. *International Journal of Remote Sensing, 15*, 145-161

Eicher, C.L., & Brewer, C.A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science, 28*, 125-138

Eklundh, L., & Jönsson, P. (2017). Timesat 3.3 Software Manual. *Lund and Malmö University, Sweden*

Fearnside, P.M. (2005). Deforestation in Brazilian Amazonia: History, Rates, and Consequences; Deforestación en la Amazonía Brasileña: Historia, Tasas y Consecuencias. *Conservation Biology, 19*, 680-688

Fearnside, P.M.A., Deforestation of (2002). Amazonia, Deforestation. In A.S. Goudie, & D.J. Cuff (Eds.), *Encyclopedia of Global Change: Environmental Change and Human Society* (pp. 31-38). New York, USA: Oxford University Press

Ferreira, L.G., & Huete, A.R. (2004). Assessing the seasonal dynamics of the Brazilian Cerrado vegetation through the use of spectral vegetation indices. *International Journal of Remote Sensing, 25*, 1837-1860

Ferreira, M.E., Ferreira, L.G., Latrubesse, E.M., & Miziara, F. (2013). Considerations about the land use and conversion trends in the savanna environments of Central Brazil under a geomorphological perspective. *Journal of Land Use Science, 11*, 33-47

Ferreira, M.E., Ferreira, L.G., Miziara, F., & Soares-Filho, B.S. (2012). Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation. *Journal of Land Use Science, 8*, 403-421

Ferreira, M.E., Ferreira, L.G., Sano, E.E., & Shimabukuro, Y.E. (2007). Spectral linear mixture modelling approaches for land cover mapping of tropical savanna areas in Brazil. *International Journal of Remote Sensing, 28*, 413-429

Fisher, J.I., Mustard, J.F., & Vadeboncoeur, M.A. (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment, 100*, 265-279

Fisher, P.F., & Langford, M. (1995). Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A, 27*, 211-224

Flood, N. (2017). Comparing Sentinel-2A and Landsat 7 and 8 Using Surface Reflectance over Australia. *Remote Sensing, 9*, 659

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. (2005). Global consequences of land use. *Science, 309*, 570-574

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., & Zaks, D.P. (2011). Solutions for a cultivated planet. *Nature, 478*, 337-342

Franca, H., & Setzer, A.W. (1998). AVHRR temporal analysis of a savanna site in Brazil. *International Journal of Remote Sensing, 19*, 3127-3140

Francoso, R.D., Haidar, R.F., & Machado, R.B. (2016). Tree species of South America central savanna: endemism, marginal areas and the relationship with other biomes. *Acta Botanica Brasilica, 30*, 78-86

Frantz, D., Röder, A., Stellmes, M., & Hill, J. (2017). Phenology-adaptive pixel-based compositing using optical earth observation imagery. *Remote Sensing of Environment, 190*, 331-347

Furley, P. (2010). Tropical savannas: Biomass, plant ecology, and the role of fire and soil on vegetation. *Progress in Physical Geography, 34*, 563-585

Furley, P.A. (1999). The nature and diversity of neotropical savanna vegetation with particular reference to the Brazilian cerrados. *Global Ecology and Biogeography, 8*, 223-241

GCOS (2018). The GCOS Essential Climate Variable (ECV) Data Access Matrix. Access date: 02.03.2018. https://www.ncdc.noaa.gov/gosic/gcos-essential-climate-variable-ecv-data-access-matrix

Gessner, U., Knauer, K., Kuenzer, C., & Dech, S. (2015). Land Surface Phenology in a West African Savanna: Impact of Land Use, Land Cover and Fire. In C. Kuenzer, S. Dech, & W. Wagner (Eds.), *Remote Sensing Time Series: Revealing Land Surface Dynamics* (pp. 203-223). Cham, Switzerland: Springer International Publishing

Ghiyamat, A., & Shafri, H.Z.M. (2010). A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing, 31*, 1837-1856

Gibbs, H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., Soares-Filho, B., Barreto, P., Micol, L., & Walker, N.F. (2015). Brazil's Soy Moratorium. *Science, 347*, 377-378

Gitelson, A.A., Merzlyak, M.N., & Lichtenthaler, H.K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J Plant Physiol, 148*, 501-508

Goetz, S., Baccini, A., Laporte, N., Johns, T., Walker, W., Kellndorfer, J., Houghton, R., & Sun, M. (2009). Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management, 4*, 1-7

Gomes, M.F., & Maillard, P. (2015). Using spectral and textural features from RapidEye images to estimate age and structural parameters of Cerrado vegetation. *International Journal of Remote Sensing, 36*, 3058-3076

González-Roglich, M., & Swenson, J.J. (2016). Tree cover and carbon mapping of Argentine savannas: Scaling from field to region. *Remote Sensing of Environment, 172*, 139-147

Grace, J., Mitchard, E., & Gloor, E. (2014). Perturbations in the carbon budget of the tropics. *Global Change Biology, 20*, 3238-3255

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., & Solis, M. (1998). Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment, 65*, 227-248

Greenwell, M.B. (2017). pdp: An R Package for Constructing Partial Dependence Plots. *The R Journal, 9*, 421-436

Griffiths, P., van der Linden, S., Kuemmerle, T., & Hostert, P. (2013). A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 6*, 2088-2101

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M.C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). Sustainable development goals for people and planet. *Nature, 495*, 305-307

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., & Sang, B. (2015). The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing, 7*, 8830-8857

Haines-Young, R. (2009). Land use and biodiversity relationships. *Land Use Policy, 26*, 178-186

Hanes, J., Liang, L., & Morisette, J. (2014). Land Surface Phenology. In J.M. Hanes (Ed.), *Biophysical Applications of Satellite Remote Sensing* (pp. 99-125). Berlin, Heidelberg, Germany: Springer

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R.G. (2013). High-resolution global maps of 21st-century forest cover change. *Science, 342*, 850-853

He, H.S., Mladenoff, D.J., Radeloff, V.C., & Crow, T.R. (1998). Integration of gis data and classified satellite imagery for regional forest assessment. *Ecological Applications, 8*, 1072-1083

He, K.S., Rocchini, D., Neteler, M., & Nagendra, H. (2011). Benefits of hyperspectral remote sensing for tracking plant invasions. *Diversity and Distributions, 17*, 381-392

Henebry, G.M., & de Beurs, K.M. (2013). Remote Sensing of Land Surface Phenology: A Prospectus. In M.D. Schwartz (Ed.), *Phenology: An Integrative Environmental Science* (pp. 385-411). Dordrecht, Netherlands: Springer

Hill, M.J., Román, M.O., & Schaaf, C.B. (2010). Biogeography and Dynamics of Global Tropical and Subtropical Savannas. *Ecosystem Function in Savannas* (pp. 3-37). Boca Raton, FL, USA: CRC Press

Hostert, P., Griffiths, P., van der Linden, S., & Pflugmacher, D. (2015). Time Series Analyses in a New Era of Optical Satellite Data. In C. Kuenzer, S. Dech, & W. Wagner (Eds.), *Remote Sensing Time Series* (pp. 25-41). Cham, Netherlands: Springer

Houghton, R., & Goetz, S. (2008). New satellites offer a better approach for determining sources and sinks of carbon. *Eos Transactions of the American Geophysical Union, 43*, 417 - 418

Huang, C., Davis, L.S., & Townshend, J.R.G. (2002a). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*, 725-749

Huang, C., Wylie, B., Yang, L., Homer, C., & Zylstra, G. (2002b). Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing, 23*, 1741-1748

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., & Ferreira, L.G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*, 195-213

IBGE (2018). Instituto Brasileiro de Geografia e Estatística. Mapa dos biomas. Access date: 24.02.2018. https://ww2.ibge.gov.br/english/geociencias/recursosnaturais/mapas/mapas_doc1.shtm

INPE (2018a). DETER - Sistema de Detecção de desmatamentos em Tempo Real. Access date: 22.02.2018. http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/deter/deter

INPE (2018b). PRODES - Monitoramento da Floresta Amazônica Brasileira por Satélite. Access date: 22.02.2018. http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes

IPCC (2013). Summary for Policymakers. In T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, B. V., & P.M. Midgley (Eds.), *Climate*

Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge United Kingdom and New York, NY, USA: Cambridge University Press

Iwasaki, A., & Tadono, T. (2014). Future Prospect on Space Communication Infrastructure for Remote Sensing Satellites. In, *International Conference on Space Optical Systems and Applications (ICSOS)* Kobe, Japan

Jönsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *Geoscience and Remote Sensing, IEEE Transactions on, 40*, 1824-1832

Jönsson, P., & Eklundh, L. (2004). TIMESAT - a program for analyzing time-series of satellite sensor data. *Computers & Geosciences, 30*, 833-845

Kalacska, M., Sanchez-Azofeifa, G.A., Rivard, B., Caelli, T., White, H.P., & Calvo-Alvarado, J.C. (2007). Ecological fingerprinting of ecosystem succession: Estimating secondary tropical dry forest structure and diversity using imaging spectroscopy. *Remote Sensing of Environment, 108*, 82-96

Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., & Mattsson, E. (2015). Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sensing, 7*, 10017-10041

Klink, C.A., & Machado, R.B. (2005). Conservation of the Brazilian Cerrado. *Conservation Biology, 19*, 707-713

Klink, C.A., & Moreira, A.G. (2002). Past and current human occupation, and land use. In P.S. Oliveira, & R.J. Marquis (Eds.), *The cerrados of Brazil: ecology and natural history of a neotropical savanna* (pp. 69-88). New York, USA: Columbia Press

Langford, M. (2006). Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers Environment and Urban Systems, 30*, 161-180

Lapola, D.M., Martinelli, L.A., Peres, C.A., Ometto, J.P.H.B., Ferreira, M.E., Nobre, C.A., Aguiar, A.P.D., Bustamante, M.M.C., Cardoso, M.F., Costa, M.H., Joly, C.A., Leite, C.C., Moutinho, P., Sampaio, G., Strassburg, B.B.N., & Vieira, I.C.G. (2013). Pervasive transition of the Brazilian land-use system. *Nature Climate Change, 4*, 27-35

Lary, D.J., Zewdie, G.K., Liu, X., Wu, D., Levetin, E., Allee, R.J., Malakar, N., Walker, A., Mussa, H., & Mannino, A. (2018). Machine Learning Applications for Earth Observation. *Earth Observation Open Science and Innovation* (pp. 165-218). Cham, Switzerland: Springer

Lausch, A., Bannehr, L., Beckmann, M., Boehm, C., Feilhauer, H., Hacker, J.M., Heurich, M., Jung, A., Klenke, R., Neumann, C., Pause, M., Rocchini, D., Schaepman, M.E., Schmidtlein, S., Schulz, K., Selsam, P., Settele, J., Skidmore, A.K., & Cord, A.F. (2016). Linking Earth Observation and taxonomic, structural and functional biodiversity: Local to ecosystem perspectives. *Ecological Indicators, 70*, 317-339

Le Toan, T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., & Ulander, L. (2011). The BIOMASS

mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment, 115*, 2850-2860

Lee, C.M., Cable, M.L., Hook, S.J., Green, R.O., Ustin, S.L., Mandl, D.J., & Middleton, E.M. (2015). An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment, 167*, 6-19

Lefsky, M.A., Turner, D.P., Guzy, M., & Cohen, W.B. (2005). Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modeled forest productivity. *Remote Sensing of Environment, 95*, 549-558

Lehmann, C.E.R., Archibald, S.A., Hoffmann, W.A., & Bond, W.J. (2011). Deciphering the distribution of the savanna biome. *New Phytologist, 191*, 197-209

Leitão, P.J., Schwieder, M., Pötzschner, F., Pinto, J.R.R., Teixeira, A., Pedroni, F., Rogass, C., Sanchez, M., van der Linden, S., da Cunha Bustamante, M.M., & Hostert, P. (submitted). From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. *Ecosphere*

Leitão, P.J., Schwieder, M., Suess, S., Okujeni, A., Galvão, L., Linden, S., & Hostert, P. (2015). Monitoring Natural Ecosystem and Ecological Gradients: Perspectives with EnMAP. *Remote Sensing, 7*, 13098-13119

Lenza, E., & Klink, C.A. (2006). Comportamento fenológico de espécies lenhosas em um cerrado sentido restrito de Brasília, DF. *Brazilian Journal of Botany, 29*, 627-638

Li, M., & Qu, J.J. (2013). Satellite applications for detecting vegetation phenology. In J. Qu, A. Powell, & M.V.K. Sivakumar (Eds.), *Satellite-based applications on climate change* (pp. 263-276). Dordrecht, Netherlands: Springer

Liaw, A., & Wiener, M. (2002). Classifiaction and Regression by randomForest. *R News, 2*, 18–22

Lieth, H. (1974). Purposes of a phenology book. In H. Lieth (Ed.), *Phenology and seasonality modeling* (pp. 3-19). Berlin, Heidelberg, Germany: Springer

Lima, J.E.F.W., de Gois Aquino, F., Chaves, T.A., & Lorz, C. (2017). Development of a spatially explicit approach for mapping ecosystem services in the Brazilian Savanna–MapES. *Ecological Indicators, 82*, 513-525

Lima, J.E.F.W., & Silva, E.M.d. (2008). Recursos hídricos do bioma Cerrado. In S.M. Sano, S.P. de Almeida, & J.F. Ribeiro (Eds.), *Cerrado: ecologia e flora* (pp. 89-106). Brasilia, DF, Brazil: Embrapa Cerrados

Lloyd, S. (1982). Least squares quantization in PCM. *Ieee Transactions on Information Theory, 28*, 129-137

Loveland, T.R., & Irons, J.R. (2016). Landsat 8: The plans, the reality, and the legacy. *Remote Sensing of Environment, 185*, 1-6

Lu, D., Chen, Q., Wang, G., Liu, L., Li, G., & Moran, E. (2016). A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *International Journal of Digital Earth, 9*, 63-105

Magnusson, W.E., Lima, A.P., Luizão, R., Luizão, F., Costa, F.R.C., Castilho, C.V.d., & Kinupp, V.F. (2005). RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. *Biota Neotropica, 5*, 19-24

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J., & Watt, A.D. (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends in Ecology & Evolution, 25*, 574-582

Maia, H., Hargrave, J., Gómez, J.J., & Röper, M. (2011). Avaliação do Plano de Prevenção e Controle do Desmatamento na Amazônia Legal: PPCDAm: 2007-2010

Marchant, R. (2010). Understanding complexity in savannas: climate, biodiversity and people. *Current Opinion in Environmental Sustainability, 2*, 101-108

Markham, B.L., & Helder, D.L. (2012). Forty-year calibrated record of earth-reflected radiance from Landsat: A review. *Remote Sensing of Environment, 122*, 30-40

Marris, E. (2005). Conservation in Brazil: The forgotten ecosystem. *Nature, 437*, 944-945

McRoberts, R.E., & Tomppo, E.O. (2007). Remote sensing support for national forest inventories. *Remote Sensing of Environment, 110*, 412-419

MCT (2016). Terceira Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima – Sumário executivo/Ministério da Ciência, Tecnologia e Inovação. *Brasília, DF, Brasil*

MEA (2005). *Ecosystems and human well-being. Millenium Ecosystem Assessment.* Washington, DC, USA: Island Press

Melaas, E.K., Friedl, M.A., & Zhu, Z. (2013). Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM plus data. *Remote Sensing of Environment, 132*, 176-185

Melaas, E.K., Sulla-Menashe, D., Gray, J.M., Black, T.A., Morin, T.H., Richardson, A.D., & Friedl, M.A. (2016). Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat. *Remote Sensing of Environment, 186*, 452-464

Melillo, J.M., Houghton, R.A., Kicklighter, D.W., & McGuire, A.D. (1996). Tropical deforestation and the global carbon budget. *Annual Review of Energy and the Environment, 21*, 293-310

Mendonça, R.C., Felfili, J.M., Walter, B.M.T., Da Silva Junior, M.C., Rezende, A., V., Filgueiras, T.S., & Nogueira, P.E. (2008). Flora vascular do bioma Cerrado. In S.M. Sano, S.P. de Alemeida, & J.F. Ribeiro (Eds.), *Cerrado: ecologia e flora* (pp. 423-442). Brasilia, DF, Brazil: Embrapa Cerrados

Menzel, A. (2002). Phenology: Its Importance to the Global Change Community. *Climatic Change, 54*, 379-385

Menzel, A., Sparks, T.H., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kübler, K., Bissolli, P., Braslavská, O.G., Briede, A., Chmielewski, F.M., Crepinsek, Z., Curnel, Y.,

Dahl, Å., Defila, C., Donnelly, A., Filella, Y., Jatczak, K., Måge, F., Mestre, A., Nordli, Ø., Peñuelas, J., Pirinen, P., Remišová, V., Scheifinger, H., Striz, M., Susnik, A., Van Vliet, A.J.H., Wielgolaski, F.-E., Zach, S., & Zust, A.N.A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology, 12*, 1969-1976

Mesquita Junior, H., N. (2000). NDVI measurements of neotropical savanna physiognomies a gradient of: biomass, structure and phenology changes. *International Archives of Photogrammetry and Remote Sensing, 33*, 93-100

Miglani, A., Ray, S.S., Vashishta, D.P., & Parihar, J.S. (2011). Comparison of Two Data Smoothing Techniques for Vegetation Spectra Derived From EO-1 Hyperion. *Journal of the Indian Society of Remote Sensing, 39*, 443-453

Miranda, H.S., Bustamante, M.M., Miranda, A.C., Oliveira, P., & Marquis, R. (2002). The fire factor. In P.E.M. Oliveira, & R.J. Marquis (Eds.), *The Cerrados of Brazil. Ecology and Natural History of a Neotropical Savanna* (pp. 51-68). New York, USA: Columbia Univerity Press

Miranda, H.S., Sato, M.N., Neto, W.N., & Aires, F.S. (2009). Fires in the cerrado, the Brazilian savanna. *Tropical Fire Ecology* (pp. 427-450). Berlin, Heidelberg, Germany: Springer

Mitchard, E.T.A., Saatchi, S.S., Baccini, A., Asner, G.P., Goetz, S.J., Harris, N.L., & Brown, S. (2013). Uncertainty in the spatial distribution of tropical forest biomass: a comparison of pan-tropical maps. *Carbon Balance and Management, 8*, 10

Mitchard, E.T.A., Saatchi, S.S., Woodhouse, I.H., Nangendo, G., Ribeiro, N.S., Williams, M., Ryan, C.M., Lewis, S.L., Feldpausch, T.R., & Meir, P. (2009). Using satellite radar backscatter to predict above-ground woody biomass: A consistent relationship across four different African landscapes. *Geophysical Research Letters, 36*

Mittermeier, R.A., Turner, W.R., Larsen, F.W., Brooks, T.M., & Gascon, C. (2011). Global Biodiversity Conservation: The Critical Role of Hotspots. In F.E. Zachos, & J.C. Habel (Eds.), *Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas* (pp. 3-22). Berlin, Heidelberg, Germany: Springer

Miura, T., Huete, A.R., Ferreira, L.G., & Sano, E.E. (2003). Discrimination and biophysical characterization of cerrado physiognomies with EO-1 Hyperspctral Hyperion. In R.E. Green (Ed.), *12th JPL Airborne Earth Science Workshop* (pp. 207-212). NASA, Jet Propulsion Laboratory, California Institute of Technology. Passadena, CA

MMA (2015). Ministério de Meio Ambiente. Mapeamento do Uso e Cobertura da Terra do cerrado – Projeto TerraClass cerrado 2013. Access date: 08.02.2018. http://www.dpi.inpe.br/tccerrado/

MMA (2017). Ministério de Meio Ambiente. Brazil's Forest Reference Emission Level for Reducing Emissions from Deforestation in the Cerrado biome for Results-based Payments for REDD+ under the United Nations Framework Convention on Climate Change. In (p. 76)

MMA (2018). Ministério do Meio Ambiente. Geoprocessamento Access date: 24.02.2018. http://www.mma.gov.br/governanca-ambiental/geoprocessamento

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Arai, E., del Bon Espirito-Santo, F., Freitas, R., & Morisette, J. (2006). Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. *Proceedings of the National Academy of Sciences, 103*, 14637-14641

Motta, P.E., Curi, N., & Franzmeier, D.P. (2002). Relation of soils and geomorphic surfaces in the Brazilian Cerrado. In P.E.M. Oliveira, & R.J. Marquis (Eds.), *The Cerrados of Brazil*. New York, USA: Columbia University Press

Müller, H., Rufin, P., Griffiths, P., de Barros Viana Hissa, L., & Hostert, P. (2016). Beyond deforestation: Differences in long-term regrowth dynamics across land use regimes in southern Amazonia. *Remote Sensing of Environment, 186*, 652-662

Müller, H., Rufin, P., Griffiths, P., Siqueira, A.J.B., & Hostert, P. (2015). Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment, 156*, 490-499

Muukkonen, P., & Heiskanen, J. (2007). Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: A possibility to verify carbon inventories. *Remote Sensing of Environment, 107*, 617-624

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature, 403*, 853-858

NASA (2018). Harmonized Landsat Sentinel-2. Access date: 01.03.2018. https://hls.gsfc.nasa.gov/

Nepstad, D., McGrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., Bezerra, T., DiGiano, M., Shimada, J., Seroa da Motta, R., Armijo, E., Castello, L., Brando, P., Hansen, M.C., McGrath-Horn, M., Carvalho, O., & Hess, L. (2014). Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *Science, 344*, 1118-1123

Nobre, C.A., Artaxo, P., Assunção, M., Dias, F.S., Victoria, R.L., Nobre, A.D., & Krug, T. (2002). The Amazon Basin and Land-Cover Change: A Future in the Balance? In (pp. 137-141). Berlin, Heidelberg: Springer

O'Connor, B., Secades, C., Penner, J., Sonnenschein, R., Skidmore, A., Burgess, N.D., & Hutton, J.M. (2015). Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets. *Remote Sensing in Ecology and Conservation, 1*, 1 -10

Oliveira-Filho, A.T., & Ratter, J.A. (2002). Vegetation Physiognomies and Woody Flora of the Cerrado Biome. In P.E.M. Oliveira, & R.J. Marquis (Eds.), *The Cerrados of Brazil - Ecology and Natural History of a Neotropical Savanna*. New York, USA: Columbia University Press

Ottmar, R.D., Vihnanek, R.E., Miranda, H.S., Sata, M.N., & Andrade, S.M. (2001). Stereo photo series for quantifying cerrado fuels in Central Brazil - volume I. In, *General Technical Report PNW-GRT-519*. Northwest Research Station, Portland: USDA-FS

Pasquarella, V.J., Holden, C.E., Kaufman, L., & Woodcock, C.E. (2016). From imagery to ecology: leveraging time series of all available Landsat observations to map and monitor ecosystem state and dynamics. *Remote Sensing in Ecology and Conservation, 2*, 152-170

Pearlman, J., Carman, S., Segal, C., Jarecke, P., Clancy, P., & Browne, W. (2001). Overview of the Hyperion Imaging Spectrometer for the NASA EO-1 mission. In, *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)* (pp. 3036-3038 vol.3037)

Pellegrini, A.F.A., Socolar, J.B., Elsen, P.R., & Giam, X. (2016). Trade-offs between savanna woody plant diversity and carbon storage in the Brazilian Cerrado. *Global Change Biology, 22*, 3373-3382

Peng, Y., & Gitelson, A.A. (2012). Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. *Remote Sensing of Environment, 117*, 440-448

Pereira, H.M., Navarro, L.M., & Martins, I.S. (2012). Global Biodiversity Change: The Bad, the Good, and the Unknown. *Annual Review of Environment and Resources, Vol 37, 37*, 25-50

Petrou, Z.I., Manakos, I., & Stathaki, T. (2015). Remote sensing for biodiversity monitoring: a review of methods for biodiversity indicator extraction and assessment of progress towards international targets. *Biodiversity and Conservation, 24*, 2333-2363

Pettorelli, N., Nagendra, H., Williams, R., Rocchini, D., & Fleishman, E. (2014). A new platform to support research at the interface of remote sensing, ecology and conservation. *Remote Sensing in Ecology and Conservation*, 1-3

Pezzini, F.F., Melo, P.H.A.d., Oliveira, D.M.S.d., Amorim, R.X.d., Figueiredo, F.O.G.d., Drucker, D.P., Rodrigues, F.R.d.O., Zuquim, G., Emilio, T., Costa, F.R.C., Magnusson, W.E., Sampaio, A.F., Lima, A.P., Garcia, A.R.d.M., Manzatto, A.G., Nogueira, A., Costa, C.P.d., Barbosa, C.E.d.A., Bernardes, C., Castilho, C.V.d., Cunha, C.N.d., Freitas, C.G.d., Cavalcante, C.d.O., Brandão, D.O., Rodrigues, D.d.J., Santos, E.C.d.P.R.d., Baccaro, F.B., Ishida, F.Y., Carvalho, F.A., Moulatlet, G.M., Guillaumet, J.-L.B., Pinto, J.L.P.V., Schietti, J., Vale, J.D.d., Belger, L., Verdade, L.M., Pansonato, M.P., Nascimento, M.T., Santos, M.C.V.d., Cunha, M.S.d., Arruda, R., Barbosa, R.I., Romero, R.L., Pansini, S., & Pimentel, T.P. (2012). The Brazilian Program for Biodiversity Research (PPBio) Information System. *Embrapa Roraima-Artigo em periódico indexado (ALICE)*

Pinto, J.R.R., Sano, E.E., Reino, C.M., & Pinto, C.A.S. (2009). Parques Nacionais do Cerrado e os tipos de formações vegetacionais preservados. *Natureza & Conservacao, 7*, 57-71

Pirani, F.R., Sanchez, M., & Pedroni, F. (2009). Fenologia de uma comunidade arbórea em cerrado sentido restrito, Barra do Garças, MT, Brasil. *Acta Botanica Brasilica, 23*, 1096-1109

Psomas, A., Kneubühler, M., Huber, S., Itten, K., & Zimmermann, N.E. (2011). Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing, 32*, 9007-9031

R Core Team (2017). R: A language and environment for statistical computing. In. Vienna, Austria. URL https://www.R-project.org/: R Foundation for Statistical Computing

Ramankutty, N., Graumlich, L., Achard, F., Alves, D., Chhabra, A., DeFries, R.S., Foley, J.A., Geist, H., Houghton, R.A., & Goldewijk, K.K. (2006). Global land-cover change: Recent progress, remaining challenges. *Land-use and land-cover change* (pp. 9-39). Berlin, Heidelberg, Germany: Springer

Ratana, P., & Huete, A. (2004). Seasonal dynamics of native and converted cerrado physiognomies with MODIS data. *Igarss 2004: Ieee International Geoscience and Remote Sensing Symposium Proceedings, Vols 1-7*, 4336-4339

Ratana, P., Huete, A.R., & Ferreira, L. (2005). Analysis of cerrado physiognomies and conversion in the MODIS seasonal-temporal domain. *Earth Interactions, 9*, 1-22

Ratter, J.A., Ribeiro, J.F., & Bridgewater, S. (1997). The Brazilian Cerrado Vegetation and Threats to its Biodiversity. *Annals of Botany, 80*, 223-230

Reatto, A., Correia, J.R., Spera, S.T., & Martins, E.S. (2008). Solos do bioma Cerrado: aspectos pedológicos. In S.M. Sano, S.P.d. Almeida, & J.F. Ribeiro (Eds.), *Cerrado: ambiente e flora* (pp. 107-149). Brasilia, DF, Brazil: Embrapas Cerrado

Reed, B.C., Schwartz, M.D., & Xiao, X. (2009). Remote Sensing Phenology: Status and the Way Forward. In A. Noormets (Ed.), *Phenology of Ecosystem Processes: Applications in Global Change Research* (pp. 231-246). New York, NY, USA: Springer

Rezende, A.V., Vale, A.T., Sanquetta, C.R., Figueiredo Filho, A., & Felfili, J.M. (2006). Comparison of mathematical models to volume, biomass and carbon stock estimation of the woody vegetation of a cerrado sensu stricto in Brasília, DF. *Scientia Forestalis, 71*, 65-76

Ribeiro, J.F., & Walter, B.M.T. (2008). As principais fitofisionomias do bioma Cerrado. In S.M. Sano, S.P. de Almeida, & J.F. Ribeiro (Eds.), *Cerrado: ecologia e flora* (pp. 153-212). Brasilia, DF, Brazil: Embrapa Cerrados

Ribeiro, L.F., & Tabarelli, M. (2002). A structural gradient in cerrado vegetation of Brazil: changes in woody plant density, species richness, life history and plant composition. *Journal of Tropical Ecology, 18*, 775-794

Ribeiro, S.C., Fehrmann, L., Soares, C.P.B., Jacovine, L.A.G., Kleinn, C., & Gaspar, R.D. (2011). Above- and belowground biomass in a Brazilian Cerrado. *Forest Ecology and Management, 262*, 491-499

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., & Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology, 169*, 156-173

Roberts, D.A., Dennison, P.E., Roth, K.L., Dudley, K., & Hulley, G. (2015). Relationships between dominant plant species, fractional cover and Land Surface Temperature in a Mediterranean ecosystem. *Remote Sensing of Environment, 167*, 152-167

Rocha, G.F., Ferreira, L.G., Ferreira, N., C., & Ferreira, M.E. (2011). Detecção De Desmatamentos No Bioma Cerrado Entre 2002 E 2009: Padrões, Tendências E Impactos. *Revista Brasileira de Cartografia, 63*, 341-349

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V.J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., & Foley, J.A. (2009). A safe operating space for humanity. *Nature, 461*, 472–475

Rodriguez-Galiano, V.F., Dash, J., & Atkinson, P.M. (2015). Intercomparison of satellite sensor land surface phenology and ground phenology in Europe. *Geophysical Research Letters, 42*, 2253-2260

Rodríguez-Veiga, P., Wheeler, J., Louis, V., Tansey, K., & Balzter, H. (2017). Quantifying Forest Biomass Carbon Stocks From Space. *Current Forestry Reports, 3*, 1-18

Rogass, C., Guanter, L., Mielke, C., Scheffler, D., Boesche, N.K., Lubitz, C., Brell, M., Spengler, D., & Segl, K. (2014a). An automated processing chain for the retrieval of georeferenced reflectance data from hyperspectral EO-1 Hyperion acquisitions. In Z. B., M. Kycko, & R. Reuter (Eds.), *34th EARSeL Symposium*. Warsaw, Poland: EARSeL

Rogass, C., Mielke, C., Scheffler, D., Boesche, N.K., Lausch, A., Lubitz, C., Brell, M., Spengler, D., Eisele, A., Segl, K., & Guanter, L. (2014b). Reduction of uncorrelated striping noise - applications for hyperspectral pushbroom acquisitions. *Remote Sensing, 6*, 11082-11106

Romijn, E., Herold, M., Kooistra, L., Murdiyarso, D., & Verchot, L. (2012). Assessing capacities of non-Annex I countries for national forest monitoring in the context of REDD+. *Environmental Science & Policy, 19–20*, 33-48

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., & Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment, 114*, 35-49

Roy, D.P., Wulder, M.A., Loveland, T.R., C.E, W., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., & Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154-172

Rufin, P., Müller, H., Pflugmacher, D., & Hostert, P. (2015). Land use intensity trajectories on Amazonian pastures derived from Landsat time series. *International Journal of Applied Earth Observation and Geoinformation, 41*, 1-10

Running, S.W., Baldocchi, D.D., Turner, D.P., Gower, S.T., Bakwin, P.S., & Hibbard, K.A. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment, 70*, 108-127

Saatchi, S.S., Halligan, K., Despain, D.G., & Crabtree, R.L. (2007). Estimation of Forest Fuel Load From Radar Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing, 45*, 1726-1740

Sabine, C.L., Heimann, M., Artaxo, P., Bakker, D.C., Chen, C.-T.A., Field, C.B., Gruber, N., Le Quéré, C., Prinn, R.G., & Richey, J.E. (2004). Current status and past trends of the global carbon cycle. *Scope-scientific committee on problems of the environment international council of scientific unions, 62*, 17-44

Sano, E.E., Ferreira, L.G., Asner, G.P., & Steinke, E.T. (2007). Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical savanna. *International Journal of Remote Sensing, 28*, 2739-2752

Sano, E.E., Ferreira, L.G., & Huete, A.R. (2005). Synthetic aperture radar (L band) and optical vegetation indices for discriminating the Brazilian savanna physiognomies: A comparative analysis. *Earth Interactions, 9*, 1-15

Sano, E.E., Rosa, R., Brito, J.L., & Ferreira, L.G. (2010). Land cover mapping of the tropical savanna region in Brazil. *Environmental Monitoring and Assessment, 166*, 113-124

Savitzky, A., & Golay, M.J.E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry, 36*, 1627-1639

Scheffler, D., & Karrasch, P. (2014). Destriping of hyperspectral image data: an evaluation of different algorithms using EO-1 Hyperion data. *Journal of Applied Remote Sensing, 8*, 083645-083645

Schlesinger, W.H., & Bernhardt, E.S. (2013). Chapter 11 - The Global Carbon Cycle. *Biogeochemistry (Third Edition)* (pp. 419-444). Boston, USA: Academic Press

Schmeller, D.S., Julliard, R., Bellingham, P.J., Bohm, M., Brummitt, N., Chiarucci, A., Couvet, D., Elmendorf, S., Forsyth, D.M., Moreno, J.G., Gregory, R.D., Magnusson, W.E., Martin, L.J., McGeoch, M.A., Mihoub, J.B., Pereira, H.M., Proenca, V., van Swaay, C.A.M., Yahara, T., & Belnap, J. (2015). Towards a global terrestrial species monitoring program. *Journal for Nature Conservation, 25*, 51-57

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (2013). Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. In: US Geological Survey

Scholes, R.J., & Smart, K. (2013). 4.09 - Carbon Storage in Terrestrial Ecosystems A2 - Pielke, Roger A. *Climate Vulnerability* (pp. 93-108). Oxford, England: Academic Press

Schwieder, M., Leitão, P.J., da Cunha Bustamante, M.M., Ferreira, L.G., Rabe, A., & Hostert, P. (2016). Mapping Brazilian savanna vegetation gradients with Landsat time series. *International Journal of Applied Earth Observation and Geoinformation, 52*, 361-370

Schwieder, M., Leitão, P.J., Suess, S., Senf, C., & Hostert, P. (2014). Estimating Fractional Shrub Cover Using Simulated EnMAP Data: A Comparison of Three Machine Learning Regression Techniques. *Remote Sensing, 6*, 3427-3445

Schwieder, M., Leitão, P.J., Rabe, A., Bustamante, M., Ferreira, L.G., Patrick Hostert (2015). Mapping Cerrado physiognomies using Landsat time series based phenological profiles. In, *Brazilian Symposium on Remote Sensing*. Joao Pessoa, Brazil

SIC (2018). Satellite Imaging Coroperation. Satellite Sensors. Access date: 02.03.2018. https://www.satimagingcorp.com/satellite-sensors/

Silva, J.F., Farinas, M.R., Felfili, J.M., & Klink, C.A. (2006). Spatial heterogeneity, land use and conservation in the cerrado region of Brazil. *Journal of Biogeography, 33*, 536-548

Silva, J.M.C., & Bates, J.M. (2002). Biogeographic Patterns and Conservation in the South American Cerrado: A Tropical Savanna Hotspot. *BioScience, 52*, 225-234

Skidmore, A., Pettorelli, N., Coops, N.C., Geller, G.N., Hansen, M., Lucas, R., Mücher, C.A., O'Connor, B., Paganini, M., Pereira, H.M., Schaepman, M.E., Turner, W., Wang, T., & Wegmann, M. (2015). Agree on biodiversity metrics to track from space. *Nature, 523*, 403-405

Smith, A.M.S., Kolden, C.A., Tinkham, W.T., Talhelm, A.F., Marshall, J.D., Hudak, A.T., Boschetti, L., Falkowski, M.J., Greenberg, J.A., Anderson, J.W., Kliskey, A., Alessa, L., Keefe, R.F., & Gosz, J.R. (2014). Remote sensing the vulnerability of vegetation in natural terrestrial ecosystems. *Remote Sensing of Environment, 154*, 322-337

Solbrig, O.T. (1996). The Diversity of the Savanna Ecosystem. In O.T. Solbrig, E. Medina, & J.F. Silva (Eds.), *Biodiversity and Savanna Ecosystem Processes* (pp. 1-27). Berlin Heidelberg, Germany: Springer

Souza, A.A., Galvao, L.S., & Santos, J.R. (2010). Relationships between Hyperion-derived vegetation indices, biophysical parameters, and elevation data in a Brazilian savannah environment. *Remote Sensing Letters, 1*, 55-64

Staver, A.C., Archibald, S., & Levin, S.A. (2011). The Global Extent and Determinants of Savanna and Forest as Alternative Biome States. *Science, 334*, 230-232

Steffen, W., Crutzen, P.J., & McNeill, J.R. (2007). The Anthropocene: Are Humans Now Overwhelming the Great Forces of Nature. *AMBIO: A Journal of the Human Environment, 36*, 614–621

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., & Sörlin, S. (2015). Planetary boundaries: Guiding human development on a changing planet. *Science, 347*

Steffen, W.L., Sanderson, R.A., Tyson, P.D., Jäger, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.-J., Turner, B.L., & Wasson, R.J. (2005). *Global change and the Earth system : a planet under pressure*. Berlin ; New York: Springer

Stephanie, A.S., Avery, S.C., Leah, K.V., Jack, F.M., Bernardo, F.R., Joel, R., & Marcos, A. (2014). Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. *Environmental Research Letters, 9*, 064010

Storey, J., Roy, D.P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2016). A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment, 186*, 121-122

Strassburg, B.B.N., Brooks, T., Feltran-Barbieri, R., Iribarrem, A., Crouzeilles, R., Loyola, R., Latawiec, A.E., Oliveira Filho, F.J.B., Scaramuzza, C.A.d.M., Scarano, F.R., Soares-Filho, B., & Balmford, A. (2017). Moment of truth for the Cerrado hotspot. *Nature Ecology &Amp; Evolution, 1*, 0099

Su, M.D., Lin, M.C., Hsieh, H.I., Tsai, B.W., & Lin, C.H. (2010). Multi-layer multi-class dasymetric mapping to estimate population distribution. *Science of the Total Environment, 408*, 4807-4816

Teixeira, A.M.C. (2015). Florística e estrutura da vegetação em Cerrado sentido restrito no Parque Estadual de Terra Ronca, Goiás: método RAPELD. In, *Instituto de Biologia, Departamento de Botânica* (p. 96). Brasília, Brasil.: Universidade de Brasília

Teixeira, A.M.C., Pinto, J.R.R., Amaral, A.G., & Munhoz, C.B.R. (2016). Angiosperm species of "Cerrado" sensu stricto in Terra Ronca State Park, Brazil: floristics, phytogeography and conservation. *Brazilian Journal of Botany, 40*, 225-234

Thessler, S., Ruokolainen, K., Tuomisto, H., & Tomppo, E. (2005). Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. *Global Ecology and Biogeography, 14*, 315-325

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*, 411-423

Tokola, T. (2015). Remote Sensing Concepts and Their Applicability in REDD+ Monitoring. *Current Forestry Reports, 1*, 252-260

Tollefson, J. (2015). Battle for the Amazon. *Nature, 520*, 20

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P., & Kennedy, P. (2002). Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment, 82*, 156-171

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I.N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., Huchler, M., & Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment, 120*, 9-24

Tuominen, S., Eerikäinen, K., Schibalski, A., Haakana, M., & Lehtonen, A. (2010). Mapping biomass variables with a multi-source forest inventory technique. *Silva Fennica, 44*, 109-119

Turner, B.L., Lambin, E.F., & Reenberg, A. (2007). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences, 104*, 20666-20671

UN (2017). United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP/248

UNFCCC (1997). Kyoto Protocol to the United Nations Framework Convention on Climate Change adopted at COP3 in Kyoto, Japan, on 11 December 1997

USGS (2017a). Data available from the U.S. Geological Survey

USGS (2017b). *Department of the Interior U.S. Geological Survey. Landsat Collection 1 Level 1 Product definition*. Sioux Falls, South Dakota, USA: EROS

USGS (2018). Landsat Collections. Access date: 27.02.2018. https://landsat.usgs.gov/landsat-collections

van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitão, P.J., Okujeni, A., Schwieder, M., Suess, S., & Hostert, P. (2015). The EnMAP-Box—A Toolbox and Application Programming Interface for EnMAP Data Processing. *Remote Sensing, 7*, 11249-11266

Vapnik, V. (1998). *Statistical Learning Theory*. New York, USA: Wiley

Verkerk, P.J., Levers, C., Kuemmerle, T., Lindner, M., Valbuena, R., Verburg, P.H., & Zudin, S. (2015). Mapping wood production in European forests. *Forest Ecology and Management, 357*, 228-238

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment, 118*, 127-139

Viergever, K.M., Woodhouse, I.H., & Stuart, N. (2008). Monitoring the World's Savanna Biomass by Earth Observation. *Scottish Geographical Journal, 124*, 218-225

Viña, A., Liu, W., Zhou, S.Q., Huang, J.Y., & Liu, J.G. (2016). Land surface phenology as an indicator of biodiversity patterns. *Ecological Indicators, 64*, 281-288

Viña, A., Tuanmu, M.-N., Xu, W., Li, Y., Qi, J., Ouyang, Z., & Liu, J. (2012). Relationship between floristic similarity and vegetated land surface phenology: Implications for the synoptic monitoring of species diversity at broad geographic regions. *Remote Sensing of Environment, 121*, 488-496

Vitousek, P.M. (1994). Beyond Global Warming: Ecology and Global Change. *Ecology, 75*, 1861-1876

Vitousek, P.M., Mooney, H.A., Lubchenco, J., & Melillo, J.M. (1997). Human Domination of Earth's Ecosystems. *Science, 277*, 494-499

Vourlitis, G., & da Rocha, H. (2010). Flux Dynamics in the Cerrado and Cerrado-Forest Transition of Brazil. *Ecosystem Function in Savannas* (pp. 97-116). Boca Raton, FL, USA: CRC Press

Vuolo, F., Ng, W.-T., & Atzberger, C. (2017). Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data. *International Journal of Applied Earth Observation and Geoinformation, 57*, 202-213

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature, 416*, 389-395

White, M.A., Brunsell, N., & Schwartz, M.D. (2003). Vegetation Phenology in Global Change Studies. In M.D. Schwartz (Ed.), *Phenology: An Integrative Environmental Science* (pp. 453-466). Dordrecht, Netherlands: Springer

Williams, R.J., Myers, B.A., Eamus, D., & Duff, G.A. (1999). Reproductive Phenology of Woody Species in a North Australian Tropical Savanna. *Biotropica, 31*, 626-636

Wisz, M.S., Hijmans, R.J., Peterson, A.T., Graham, C.H., Guisan, A., & NCEAS Predicting Species Distributions Working Group (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions, 14*, 763-773

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., & Team, L.S. (2008). Free access to Landsat imagery. *Science, 320*, 1011-1011

Wulder, M.A., Hilker, T., White, J.C., Coops, N.C., Masek, J.G., Pflugmacher, D., & Crevier, Y. (2015a). Virtual constellations for global terrestrial monitoring. *Remote Sensing of Environment, 170*, 62-76

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., & Woodcock, C.E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment, 122*, 2-10

Wulder, M.A., White, J.C., Fournier, R.A., Luther, J.E., & Magnussen, S. (2008). Spatially explicit large area biomass estimation: Three approaches using forest inventory and remotely sensed imagery in a GIS. *Sensors, 8*, 529-560

Wulder, M.A., White, J.C., Loveland, T.R., Woodcock, C.E., Belward, A.S., Cohen, W.B., Fosnight, E.A., Shaw, J., Masek, J.G., & Roy, D.P. (2015b). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment, 185*, 271-283

Xiao, X., Zhang, J., Yan Huimin, Wu, W., & Biradar, C. (2009). Convergence of Satellite and CO2 Eddy Flux Observations. In A. Noormets (Ed.), *Phenology of ecosystem processes. Applications in global change research*. Dordrecht, Netherlands: Springer

Zandler, H., Brenning, A., & Samimi, C. (2015). Potential of Space-Borne Hyperspectral Data for Biomass Quantification in an Arid Environment: Advantages and Limitations. *Remote Sensing, 7*, 4565-4580

Zhang, X., Friedl, M.A., Schaaf, C.B., Strahler, A.H., Hodges, J.C.F., Gao, F., Reed, B.C., & Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment, 84*, 471-475

Zhang, X., & Ni-meister, W. (2014). Remote Sensing of Forest Biomass. In J.M. Hanes (Ed.), *Biophysical Applications of Satellite Remote Sensing* (pp. 99-125). Berlin Heidelberg, Germany: Springer

Zhang, X., Wang, J., Gao, F., Liu, Y., Schaaf, C., Friedl, M., Yu, Y., Jayavelu, S., Gray, J., Liu, L., Yan, D., & Henebry, G.M. (2017). Exploration of scaling effects on coarse resolution land surface phenology. *Remote Sensing of Environment, 190*, 318-330

Zheng, G., Chen, J.M., Tian, Q., Ju, W.M., & Xia, X.Q. (2007). Combining remote sensing imagery and forest age inventory for biomass mapping. *Journal of Environmental Management, 85*, 616-623

Zhu, Z., Wang, S., & Woodcock, C.E. (2015a). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment, 159*, 269-277

Zhu, Z., & Woodcock, C.E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment, 118*, 83-94

Zhu, Z., Woodcock, C.E., Holden, C., & Yang, Z. (2015b). Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time. *Remote Sensing of Environment, 162*, 67-83

Appendix A:

From sample to pixel: using multi-scale remote sensing data for the upscaling of field collected aboveground carbon data in heterogeneous landscapes

(In review) Submitted to Ecosphere (09/2017)[*]

Pedro J. Leitão, Marcel Schwieder, Florian Pötzschner, Jose Roberto Pinto, A. Teixeira, Fernando Pedroni, Maria Sanchez, Christian Rogass, Sebastian van der Linden, Mercedes Maria da Cunha Bustamante and Patrick Hostert

[*] A revised version of the manuscript was meanwhile accepted for publication in: Leitão, P.J., Schwieder, M., Pötzschner, F., Pinto, J.R.R., Teixeira, A., Pedroni, F., Rogass, C., Sanchez, M., van der Linden, S., da Cunha Bustamante, M.M., & Hostert, P. (in press). From sample to pixel: multi-scale remote sensing data for upscaling aboveground carbon data in heterogeneous landscapes. Ecosphere.

Abstract

In times of rapid global change, ecosystem monitoring is of utmost importance. Combined field and remote sensing data enable large-scale ecosystem assessments, while maintaining local relevance and accuracy. In heterogeneous landscapes, however, the integration of field-collected data with remote sensing image pixels is not a trivial matter. Indeed, much of the uncertainty in models that rely on remote sensing to map larger areas depends on field data integration. In this study we propose to use fine spatial resolution remote sensing data as auxiliary information for upscaling field-sampled aboveground carbon data to target (meso-scale) image pixels. We also aim at assessing the effects of field data disaggregation, extrapolation and their joint effects, on its upscaling, with and without auxiliary data. We test this on three study sites in heterogeneous landscapes of the Brazilian Savanna. We thus compare two methods that use auxiliary data - surface method, which uses a weighting layer; and regression method, which applies a regression model - with one method without auxiliary data - cartographic method. To evaluate our results, we compared observed vs. estimated aboveground carbon values (for known samples) at the pixel level. Additionally, we fitted a random forest regression model with the assigned carbon estimates and the target satellite imagery, and assessed the influence of the fraction of extrapolated vs. sampled carbon values on model performance. We observed that, in heterogeneous landscapes, the use of fine spatial resolution remote sensing data improves the upscaling of field-based aboveground carbon data to coarser image pixels. Our results also show that a surface method is more suitable for spatial disaggregation, while a regression approach is preferable for extrapolating non-sampled pixel fractions. In our study, larger datasets, which included a higher proportion of estimated values, generally delivered better models of aboveground carbon than smaller datasets that are assumed to more reliably reflect reality. Our approach enables to link field and remote sensing data, which in turn enables the detailed mapping of aboveground carbon in heterogeneous landscapes over large areas through the optimized integration of field-data and multi-scale remote sensing data.

1 Introduction

In times of rapid global change, with implications on ecosystem functioning and the services provided (Cardinale et al. 2012; Foley et al. 2005), the monitoring of ecosystems is of utmost importance. Indeed, only through monitoring it is possible to assess the degree and patterns of change to further develop adequate mitigation and adaptation strategies (Turner et al.

2007). International programs towards mitigating the effects of climate change and halting biodiversity loss, such as the United Nations Reducing Emissions from Deforestation and Forest Degradation (REDD+) program or the Aichi Biodiversity Targets set by the Convention on Biological Diversity require frequent monitoring of carbon stocks and biodiversity at a global scale (Running et al. 1999; Schmeller et al. 2015), including the definition of Essential Climate Variables (ECV; GCOS 2018) and Essential Biodiversity Variables (EBV) (Pereira et al. 2012).

Field-based monitoring schemes form the basis of our knowledge on e.g. stored carbon of ecosystems. Implementation costs render field-based assessments being best suited for local studies, while broad-scale monitoring, e.g. for nation-wide or even larger assessments, is usually unfeasible (Bustamante et al. 2016). The use of remote sensing data, on the other hand, allows cost-effective ecosystem monitoring for large areas (Hansen et al. 2013; Petrou et al. 2015), although with limited applicability at local scale (Burivalova et al. 2015). Continuous large-scale ecosystem monitoring therefore requires permanent monitoring plots distributed over large-areas, such as the Long Term Ecological Research Network (Magnusson et al. 2005; Magurran et al. 2010) or National Forest Inventories (Blackard et al. 2008). Data from these monitoring schemes, providing that the location of the plots is precisely recorded, can then be integrated with remote sensing data for broad-scale assessments of e.g. carbon stock or biodiversity (Bustamante et al. 2016; McRoberts and Tomppo 2007). Combined remote sensing and field survey data can thus address our needs for large-scale ecosystem assessments, while maintaining local relevance and accuracy (Boisvenue et al. 2016; Zheng et al. 2007).

The upscaling of field data (e.g. forest inventory data) to remote sensing pixels in managed and homogeneous environments is usually done by recurring to e.g. (tree) density measures at the plot level which can then be assigned to the image pixels (Tuominen et al. 2010; Wulder et al. 2008). In natural, heterogeneous landscapes, it is however unfeasible to accurately assign density values to heterogeneous plots (Thessler et al. 2005), and the combination of field and remote sensing data becomes a difficult task which usually involves integrating point or polygon-based datasets with image pixels (He et al. 1998). On the one hand, assigning field data to a target image pixel may require the disaggregation and interpolation between field-based samples (He et al. 1998; Zheng et al. 2007), an area of active research - most particularly in demographic and climatological studies (Chen et al. 2015; Langford 2006). On the other hand, when the sample units are smaller than the image pixels, the proportions of pixels not fully covered by the field data need to be extrapolated.

The sample-to-pixel data integration allows subsequent analysis at the pixel level and hence at the full area covered by the image. Such analyses include, e.g., modeling the aboveground carbon of a particular region by fitting the pixel-allocated sample data to the respective remote sensing data.

In this study, we propose the use of fine spatial resolution remote sensing data as auxiliary information for upscaling field-sampled aboveground carbon data to target, meso-scale, image pixels in heterogeneous landscapes. We also aim at assessing the effects of data disaggregation, extrapolation and their joint effects, on the respective data integration, with and without auxiliary data. We illustrate our approach with a case-study where we upscale aboveground carbon data derived from field-based vegetation inventory data, collected in small sample plots of up to 10 x 10m2, to the pixel grid of spaceborne hyperspectral data. These data, by systematically describing the Earth's surface in a very detailed manner, has great potential for ecosystem monitoring and aboveground carbon mapping (Leitão et al. 2015). Indeed, pioneer studies have made use of spectral indices derived from hyperspectral Hyperion data for modeling aboveground biomass in both grassland and woody vegetation (Psomas et al. 2011; Zandler et al. 2015), or estimating forest structure and diversity parameters (Kalacska et al. 2007). While spaceborne hyperspectral programs are underway (Guanter et al. 2015; Lee et al. 2015), their planned meso-scale pixel (30 x 30 m2) pose problems for sample to pixel allocation, particularly in heterogeneous environments. In this study, we use data from the EO-1 Hyperion sensor (Pearlman et al. 2001) as a precursor of future hyperspectral missions (currently discontinued).

Here we focus in three study sites in the Brazilian Neotropical savannah (Cerrado), a highly heterogeneous system, which consists on a mosaic of different vegetation physiognomies (Schwieder et al. 2016). Our hypothesis is that the use of auxiliary data from a high spatial resolution sensor (such as RapidEye) improves the upscaling of field-sampled aboveground carbon data to the image pixels, particularly in such heterogeneous landscapes, ultimately enabling its use for carbon mapping across larger regions. Indeed, a recent study by González-Roglich and Swenson (2016) also used high spatial resolution satellite data for estimating tree cover in a savannah in Argentina, which was later related to carbon values. High spatial resolution satellite data, although not available to an extent that allows for large area mapping, could be used for the upscaling field samples. We thus tested several methods of incorporating auxiliary information for integrating these data. Finally, we investigate the importance of data quality for the resulting model performance by using random forest

regression models to fit different sets of pixel-based aboveground carbon data to hyperspectral data.

2 Data and Methods

2.1 Remote sensing data

To improve the upscaling of field-based aboveground carbon values to the target pixel level, we used auxiliary data derived from high spatial resolution RapidEye data (Figure A-1). The RapidEye program is a constellation of five identical spaceborne optical sensors. It covers the visible (VIS) and near infrared (NIR) portion of the electromagnetic spectrum in 5 bands with bandwidths ranging from 40 to 90 nm. A particularly interesting feature in RapidEye data is the so called “red edge band” that covers the spectral region between 690 and 730 nm and makes the data especially valuable for characterizing vegetation condition and structure(Gitelson et al. 1996; Gomes and Maillard 2015). RapidEye data are processed and delivered as individual tiles (25 x 25 km2) with a ground sampling distance (spatial resolution) of 5 m. In this study the Level 3a ortho-rectified product was used, which was already corrected for radiometric, geometric and sensor specific effects. For all three study sites we used nearly cloud free data, which were acquired close to the dates of the field surveys (Table A-1).

These data were radiometrically corrected by applying a dark object subtraction (Chavez Jr. 1996), for subsequent analysis. Finally, we derived the Red Edge Normalized Difference Vegetation Index (RENDVI; Peng and Gitelson 2012), following the formula:

$$\mathrm{RENDVI} = \frac{\rho\mathrm{NIR} - \rho\mathrm{RE}}{\rho\mathrm{NIR} + \rho\mathrm{RE}} \qquad \text{(Eq. A-1)}$$

where NIR are the reflectance values on the Near Infrared spectral band (band 5: 760 - 850 nm) and RE are the reflectance values on the Red Edge band (band 4: 690 - 730 nm). This spectral index retrieves information relevant for describing vegetation productivity (Peng and Gitelson 2012), being thus suitable for use as auxiliary data for the spatial allocation of vegetation and above ground carbon, respectively.

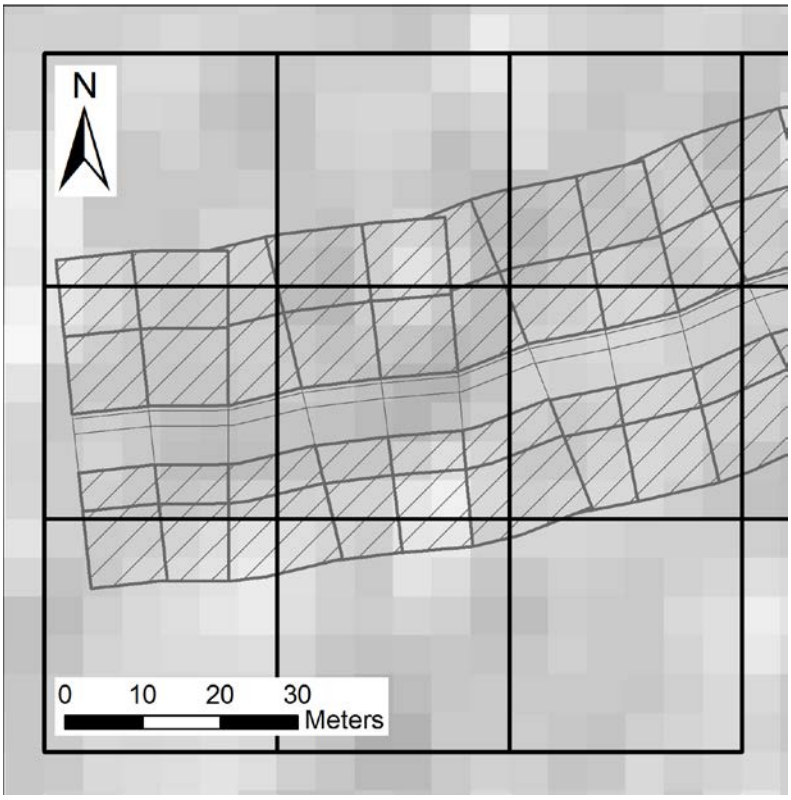


Figure A-1: Spatial allocation of the field-based data to the target (Hyperion) pixels in one of the sample transects. The sample polygons along one transect are represented by the grey polygons, the dashed polygons are those where tree allometric data were collected and the hollow polygons are those where no data were collected. The target (Hyperion) pixels are represented in black. The weighting layer (in this case the Red Edge Normalized Difference Vegetation Index calculated from RapidEye data) is represented in the background in grayscale.

Table A-1: Overview of the remote sensing and field data used in this study along with their acquisition dates.

Study Site	*Field data*	*RapidEye*	*Hyperion*
PESA	02/2012 - 05/2014	21/07/2014	27/06/2014
PNCV	08/2014 - 05/2015	07/09/2014	15/04/2015
PETR	05/2014	27/06/2014	29/06/2014

We used spaceborne hyperspectral data from the Hyperion sensor on board of the Earth Observing-1 (EO-1) platform as target remote sensing data (Figure A-1), to which the aboveground carbon data should be registered for subsequent modeling (Table A-1). The EO-1 satellite was launched as a scientific demonstrator in 2001, and while it was originally planned for a lifetime of one year it has recorded multispectral and hyperspectral data until March 2017 (Pearlman et al. 2001). This doesn't come without problems and the near "end-of-life" Hyperion data came with many different issues, such as data striping, pixel shift, and a low Signal-to-Noise ratio, etc. (Scheffler and Karrasch 2014). The Hyperion data was radiometrically corrected, including correction for pixel shifts, striping, keystone and smile,

as well as atmospheric effects. The visible and near infrared (VNIR, 400 – 1,000 nm) and the shortwave infrared (SWIR, 1000 - 2500 nm) detectors, which separately record electromagnetic radiation in their respective wavelength ranges, were co-registered (Datt et al. 2003; Rogass et al. 2014a; Rogass et al. 2014b). Data were spatially subsetted to the respective study regions (Figure A-2) and co-registered using precision terrain-corrected (L1T) Landsat OLI scenes for spatial consistency across all study areas. Erroneous or noisy spectral bands were interactively screened and excluded. Data were spectrally smoothed with a Savitzky-Golay filter (Miglani et al. 2011; Savitzky and Golay 1964). This resulted in a total of 83 spectral bands per Hyperion image (from the original 242), covering the visible, near and shortwave infrared portions of the electromagnetic spectrum.

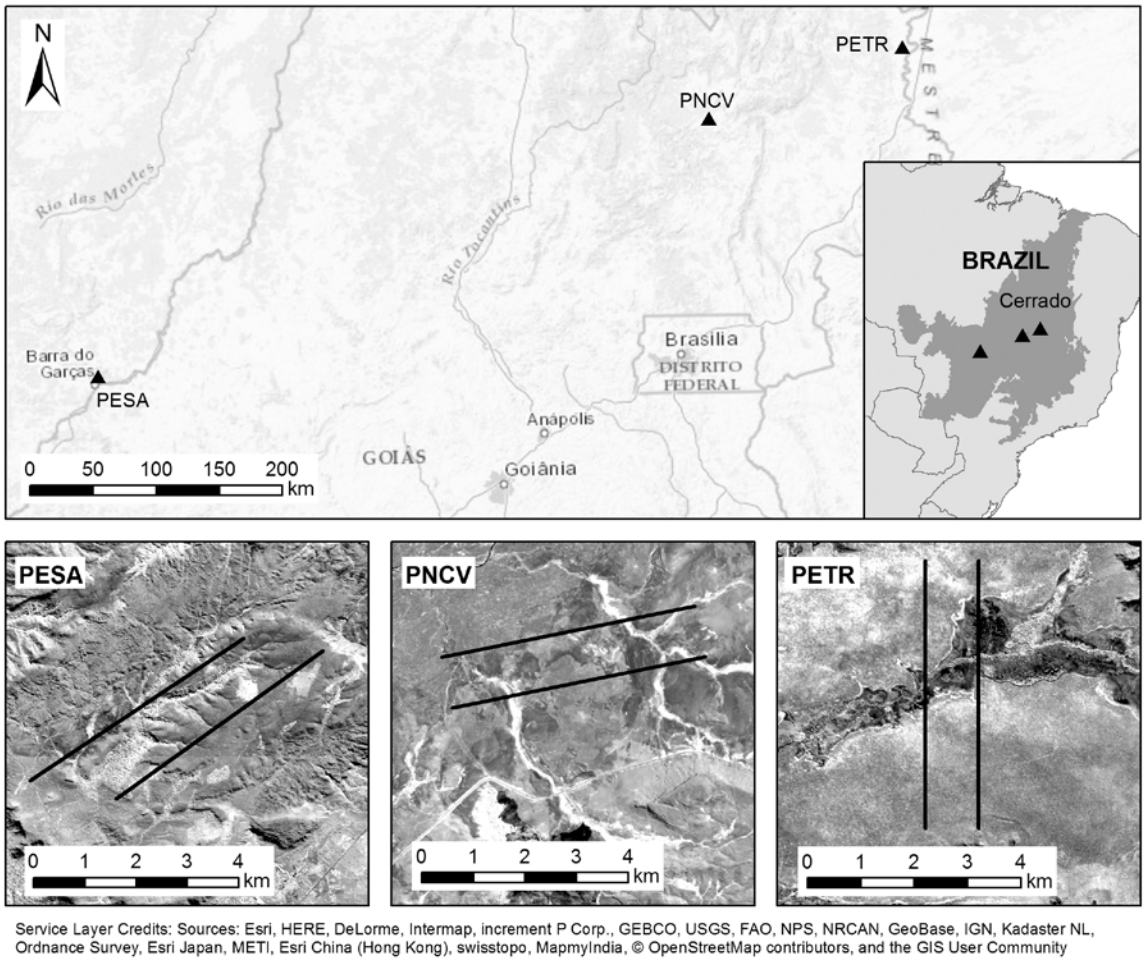


Figure A-2: This study is located in three sites in the Brazilian Cerrado: Parque Estadual da Serra Azul (PESA), Parque Nacional da Chapada dos Veadeiros (PNCV) and Parque Estadual de Terra Ronca (PETR). Above, the Cerrado is depicted in dark grey (right), and the study sites located with a triangle. Below are represented the three study sites, with the respective parallel trails that define the sampling scheme, overlaid on the near Infrared spectral band (b5) of the respective RapidEye image.

2.2 Study sites and field data

Our study is located in three sites in the Brazilian neotropical savannah (Cerrado; Figure 2): Parque Estadual de Terra Ronca (PETR), Parque Nacional da Chapada dos Veadeiros (PNCV) and Parque Estadual da Serra Azul (PESA). These sites can be considered characteristic of the typical savannahs of the central Cerrado, while representing well its variability, ranging from a low altitude sandy savannah (PETR) to upland savannahs on rock (PNCV) or deep soil substrates (PESA). The Cerrado covers ca. 20% of Brazil's land surface and it holds the richest biodiversity of all of the world's savannahs (Francoso et al. 2016). This system is, however, mostly unprotected and highly threatened thus constituting a global biodiversity hotspot (Klink and Machado 2005; Myers et al. 2000) which requires monitoring. For all sites, vegetation inventory data were collected, following a common scheme within the Program for Biodiversity Research (PPBio)(Pezzini et al. 2012), using the RAPELD principle designed for sampling the Cerrado (Magnusson et al. 2005; Teixeira 2015). This methodology implements an integrated sampling of vegetation biomass and multi-taxa biodiversity data in a standardized manner, thus allowing research on the linkages between carbon and biodiversity (Bustamante et al. 2016; Teixeira 2015).

The data were collected on a system of trails and plots, following a systematic scheme, as follows. It consists of two 5 km long parallel trails with a distance of 1 km between them, placed in a manner that fit fully within natural vegetation areas. Along each trail five plots were located 1 km apart from each other (at marks 500 m to 4500 m, counting from the beginning of the trail). Each sampling plot consists of a 250 m center line that follows the elevation contour, with a varying width, according to the taxon sampled (Pezzini et al. 2012; Teixeira 2015). Each plot was further segmented in subplots of ca. 10 m length, in a total of 25 per plot. According to the sampling design, trees with a diameter at breast height of 10 cm or more were sampled in two 10 m wide adjacent sections (sample polygons) on each side of the central line, for each subplot (Figure A-1; Figure A-2). In the cases where the plot crossed an obstacle (such as a road or river stream), or when adjacent subplots had too much overlap (depending on topography) some subplots were excluded and compensated at the end of the plot line, thus guaranteeing the sampling of 25 subplots per plot (Teixeira 2015). In this study we considered only data collected in plots covering savannah vegetation, although with varying density (Table A-1). Also, data with missing spatial reference was excluded. In total we considered 8 sampling plots in PESA, 6 in PNCV and 8 in PETR. The floristic inventory data were converted into aboveground carbon, following general allometric equations for the region (Rezende et al. 2006).

2.3 Disaggregation and extrapolation of aboveground carbon data

With the aim of testing the use of auxiliary data, from fine resolution remote sensing data, for upscaling field-collected aboveground carbon data into the meso-scale pixel grid cells, these data had to be initially disaggregated and then extrapolated into this grid. Spatial data disaggregation (interpolation) and extrapolation methods can be divided into three main types (Fisher and Langford 1995): cartographic, surface and regression methods. A simple cartographic method is the areal weighting method. Spatial interpolation or extrapolation is achieved by overlaying the target zones (e.g. pixels) on the source zones (e.g. field sample units or sample polygons) and determining the areas of intersection. Last, the values of the target zones are derived from the sum of the component portion of the source zone values. Another cartographic method is the dasymetric method, which assigns known weights (e.g. tree density) to different zones of the area of analysis that are considered to be homogeneous (Eicher and Brewer 2001) (Su et al. 2010). Surface methods. In comparison, are based on the mathematical assumption that the target values follow a continuously varying probability distribution. These methods thus require the input of auxiliary data in the form of a surface grid, to be used as weighting layer for the spatial disaggregation or extrapolation of the source values. Surface methods can be considered an extension of dasymetric mapping to a continuous weighting grid, rather than discrete classes with specific weights. Finally, regression methods make use of proxies to determine the spatial allocation of the target values. The model can take a linear or nonlinear form, as well as it can be a global or a local model (Chen et al. 2015; Verkerk et al. 2015). Both the surface and regression methods can benefit from using the spectral variation in high spatial resolution remote sensing imagery to spatially disaggregate a field sample or to infer the non-sampled portion of a target pixel.

To test our hypothesis that using auxiliary data from high spatial resolution satellite data improves the spatial allocation of the field collected floristic inventory data to the target image pixel, and to investigate which method for data integration should be used, we performed three tests: one on spatial data disaggregation (interpolation); one on spatial extrapolation; and one on the joint effects of both. For these tests we selected all the sample polygons that laid fully within each of the target pixels, for which we have known (sampled) aboveground carbon values (Figure A-3). In each test (and for each study site), we applied the referred methods of data integration - cartographic, surface and regression. The cartographic method refers to the area-weighting approach. As it does not require any auxiliary information, it assumes homogeneity on the spatial distribution of the target values, and served as control for testing our hypothesis. Both the surface and the regression methods

used the auxiliary data as respectively a weighting or a predictor layer for the spatial data integration. The regression method used was defined as a linear model.

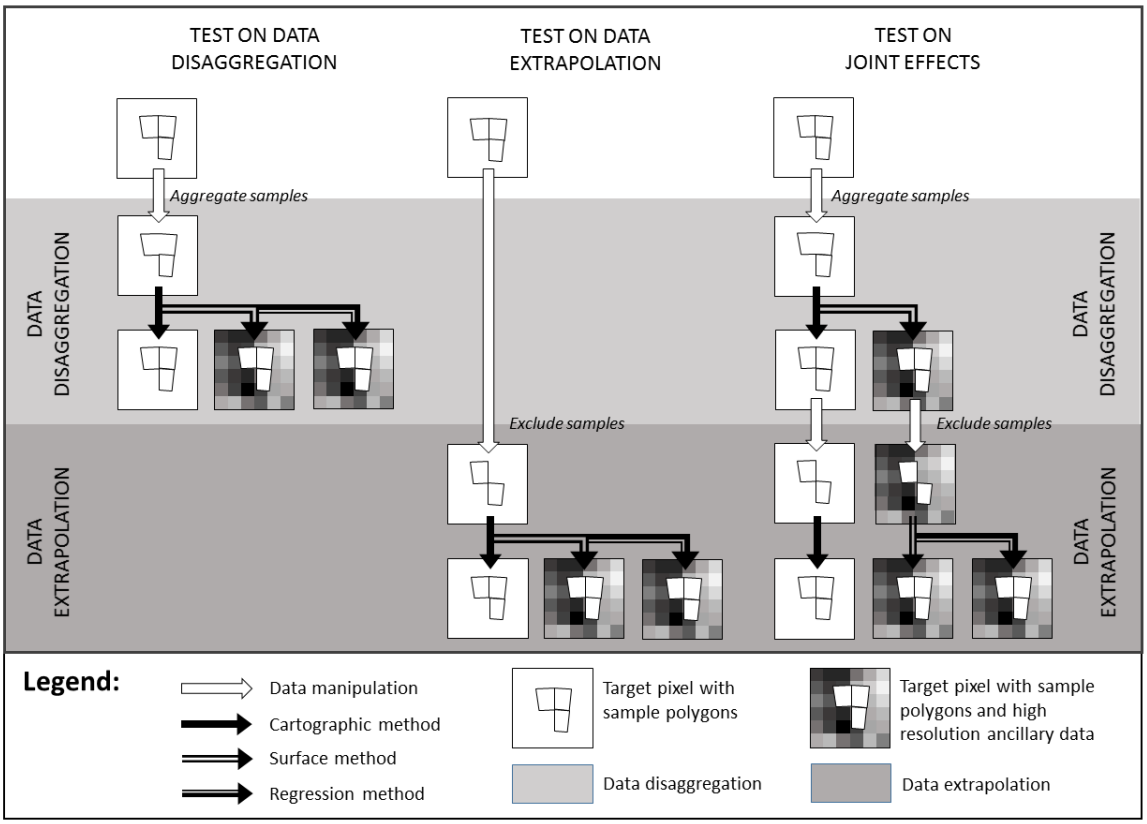


Figure A-3: Representation of the tests on the carbon data spatial disaggregation (a), extrapolation (b) and their joint effects (c), for the Hyperion pixels shown in Figure A-2. The sample polygons selected for these tests are those which are fully within a target (Hyperion) pixel. All tests used three different methods for data aggregation and extrapolation: Cartographic (which does not make use of high resolution auxiliary data from RapidEye); surface (uses RapidEye auxiliary data as a weighting layer); and regression (builds a regression model on the auxiliary data). In the test on spatial disaggregation (a), two polygons within each pixel are randomly merged, to be then estimated using the three methods described. In the test on spatial extrapolation (b), within each pixel one random polygon is deleted, to then be estimated based on the information from the remaining polygons within the same pixel. The test on the joint effects (c) starts with the results of the test on spatial disaggregation (i.e. with both measured and estimated polygons) to then be tested for spatial extrapolation.

The first test on the effects of using auxiliary data on the spatial disaggregation of field data related to one sample polygon falling in two or more pixels, i.e. whose carbon values had to be disaggregated to the respective pixels. In this case, the sample polygons were merged (aggregated) in random pairs and the respective aboveground carbon values summed. These merged polygons were subsequently disaggregated following the described methods. Here the regression method assumed the carbon values (Ct) as a function of the respective area multiplied (At) by the weighting / auxiliary layer (Wt):

$$Ct \sim At * Wt \qquad \text{(Eq. A-2)}$$

Model intercept was kept fixed at zero, to ensure that only are and weighting layer influence the response value. The resulting regression model was used to disaggregate the merged polygons back to the original ones.

The second test, on data extrapolation relates to situations where the target pixels were not fully sampled, and the respective carbon values need to be extrapolated to represent the full image pixel. In this test, one random polygon was excluded per target pixel to be subsequently estimated following the different methods. Here the regression model assumed the total carbon value for the pixel (Ct) to be the sum of the known (sampled) carbon value (Cs) and a function of the unknown (non-sampled) area (Au) multiplied by the weighting / auxiliary layer (Wu):

$$Cs \sim Cs + Au * Wu \qquad \text{(Eq. A-3)}$$

Again, model intercept was fixed at zero, while in this case the coefficient for the first predictor variable (Cs) was fixed at one. This way we constrained the regression so that the response value is a function of the known (sampled) carbon value added to the area and weight layer values. The resulting regression model was used to estimate the carbon value for all pixels.

The third test investigated the joint effects of data interpolation and extrapolation. This test most commonly approximates the reality, where the field samples (polygons) may fall within more than one image pixel and where the samples do not cover full pixels. In this case, two random polygons within each pixel were merged and the respective carbon value summed. They were disaggregated again to the original polygons and the respective carbon values estimated. Subsequently one random polygon was excluded for each pixel, and the respective carbon value estimated according to each method. In this test, and following our results from the test on spatial data disaggregation, the regression method was not applied for the disaggregation (spatial interpolation) of the merged polygons, but only for the extrapolation. This means that when testing the joint effects of data disaggregation and extrapolation using auxiliary data with a regression approach, we used the interpolated polygons resulting from the surface method, and not those from the regression method.

Finally, the aboveground carbon values were estimated for all pixels at least partially covered by the field sample polygons, following the best performing approach. For this, we calculated the pixel-based aboveground carbon by applying the surface method for disaggregating polygons which fell between different pixels, and by applying the regression

model resulting from the polygon-based test on the joint effects, always using the RENDVI calculated from the RapidEye imagery as auxiliary layer.

2.4 Validation and evaluation

We validated the spatial disaggregation and extrapolation tests at the pixel level, which means that the sum of the resulting carbon estimates per pixel were compared to the actually sampled values, by the root mean square error (RMSE), relative RMSE (RMSErel equals the RMSE divided by the mean input carbon value) and the squared Pearson correlation coefficient (R^2) between predicted and observed validation samples.

The number of pixels available for use in a regression model of aboveground Carbon depends on the assigned threshold of minimum pixel coverage by field samples. We performed a sensitivity analysis on the resulting regression models to evaluate the trade-off between having a high number of pixels (which may cover a larger portion of the variability within the study region) and a low share of estimated samples. We defined equidistant 10% thresholds ranging from 0% (all pixels partially sampled in the field) to 90% (pixels which were at least 90% sampled in the field) leading to a decreasing number of input samples with an assumed increasing reliability (Figure A-4), based on the fraction of pixels actually sampled in the field. For each threshold, we iterated the data splitting into 70% training and 30% validation data 1000 times and fitted the aboveground carbon values to the hyperspectral Hyperion data using a Random Forest (RF) regression model (Breiman 2001). RF is a machine learning approach based on the Classification and Regression Tree (CART) algorithm (Breiman et al. 1984). The algorithm trains a decision tree with a randomly drawn subset of the given input data and internally evaluates its performance with the leftover data. An ensemble of many decision trees (a forest) is trained reflecting that every single tree can be erroneous. Results are then averaged into the final model. Model performance was evaluated based on mean RMSE, RMSErel and R^2 between predicted and observed validation samples, from the withheld 30% validation data, on all iterations. All processing was performed in (R Core Team 2017) using the randomForest-package (Liaw and Wiener 2002).

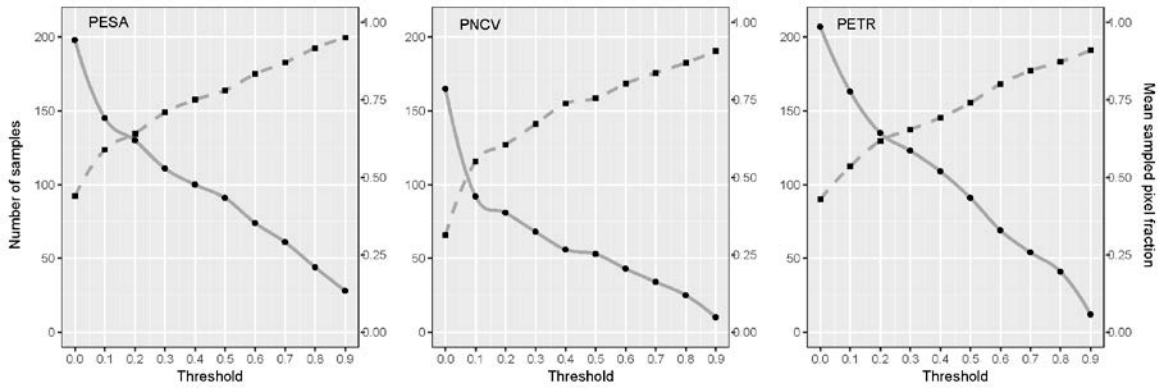


Figure A-4: Number of samples used for the analysis with varying thresholds depicted as black dots and the solid grey line. The black squares and the dashed grey line relate to the averaged share of area sampled within all pixel with the respective threshold (second y-axis).

3 Results

3.1 Disaggregation and extrapolation of aboveground carbon data

Our results were consistent across all study sites and performance measures (Table A-2). The tests on disaggregation of aboveground carbon data showed consistently better performances when using auxiliary data in a surface interpolation method (RMSErel values ranging between 0.479 and 0.655 and R^2 values ranging between 0.750 and 0.826). Using the auxiliary data in a regression approach, however, performed worse than using no auxiliary data, through the cartographic method, and this approach was abandoned (also on the test for joint effects).

Table A-2: Results of the tests on the integration of field data to the pixel level, using a cartographic (area-weighted), surface and linear regression method: test on spatial disaggregation; spatial extrapolation; and their joint effects. Both the surface and regression methods recur to auxiliary data derived from high spatial resolution RapidEye data.

		RMSE			***RMSE$_{rel}$***			***r2***				
		cartographic	*surface*	*regression*	*cartographic*	*surface*	*regression*	*cartographic*	*surface*	*regression*	*Coefficient*	*n*
	Disaggregation	1.066	**0.930**	1.593	0.751	**0.655**	1.097	0.661	**0.750**	0.260	2.693 ± 0.001	
PETR	*Extrapolation*	1.049	0.979	**0.723**	0.739	0.689	**0.509**	0.750	0.769	**0.843**	2.711 ± 0.012	106
	Joint effects	1.106	1.040	**0.778**	0.944	0.887	**0.663**	0.668	0.687	**0.771**	2.696 ± 0.012	
	Disaggregation	1.003	**0.819**	1.516	0.733	**0.599**	1.091	0.717	**0.821**	0.335	4.082 ± 0.003	
PNCV	*Extrapolation*	0.567	0.571	**0.539**	0.414	0.418	**0.394**	0.907	0.907	**0.914**	3.915 ± 0.019	70
	Joint effects	0.672	0.679	**0.634**	0.632	0.639	**0.596**	0.808	0.806	**0.815**	3.973 ± 0.018	
	Disaggregation	0.922	**0.752**	1.308	0.588	**0.479**	0.794	0.724	**0.826**	0.473	2.728 ± 0.001	
PESA	*Extrapolation*	0.648	0.639	**0.587**	0.409	0.403	**0.370**	0.870	0.873	**0.896**	2.805 ± 0.007	87
	Joint effects	0.689	0.682	**0.614**	0.548	0.542	**0.488**	0.777	0.780	**0.818**	2.810 ± 0.007	

When extrapolating the aboveground carbon data, the use of auxiliary information was always beneficial. Their use in a regression approach achieved the best results (RMSErel ranging between 0.370 and 0.509, and R^2 between 0.843 and 0.914), while the surface method was not so well performing. The extrapolation of aboveground carbon data without the use of auxiliary data performed least well.

The results of the test on the joint effects of data disaggregation and extrapolation were consistent with those of the test on data extrapolation, with the best results being achieved with the use of auxiliary information. Indeed, the best performing approach used the auxiliary data with a surface method for the carbon data disaggregation, followed by its use in a regression for the respective extrapolation (RMSErel between 0.488 and 0.663 and R^2 between 0.771 and 0.818). Using the auxiliary data with surface method also for the extrapolation of the aboveground carbon data performed less well than the previously described approach. The lack of use of auxiliary information for both the disaggregation and extrapolation of the aboveground carbon data consistently resulted in the worst performing results.

The coefficients obtained in the regression models (for the carbon data extrapolation) were specific to each study site, though consistent across all tests (between 2.728 and 2.810 for PESA; 3.915 and 4.082 for PNCV; and 2.693 and 2.711 for PETR), and with little variation between iterations.

3.2 Test on data quality

The carbon model performances differed between the three study sites, although it was possible to identify general trends related to the data thresholds used. Best model performances were derived in the PESA study site with averaged relative RMSE values ranging from 0.34 to 0.68. In PNCV the RMSErel values ranged from 0.42 to 0.72 and in PETR from 0.52 to 1.09 (Figure A-5).

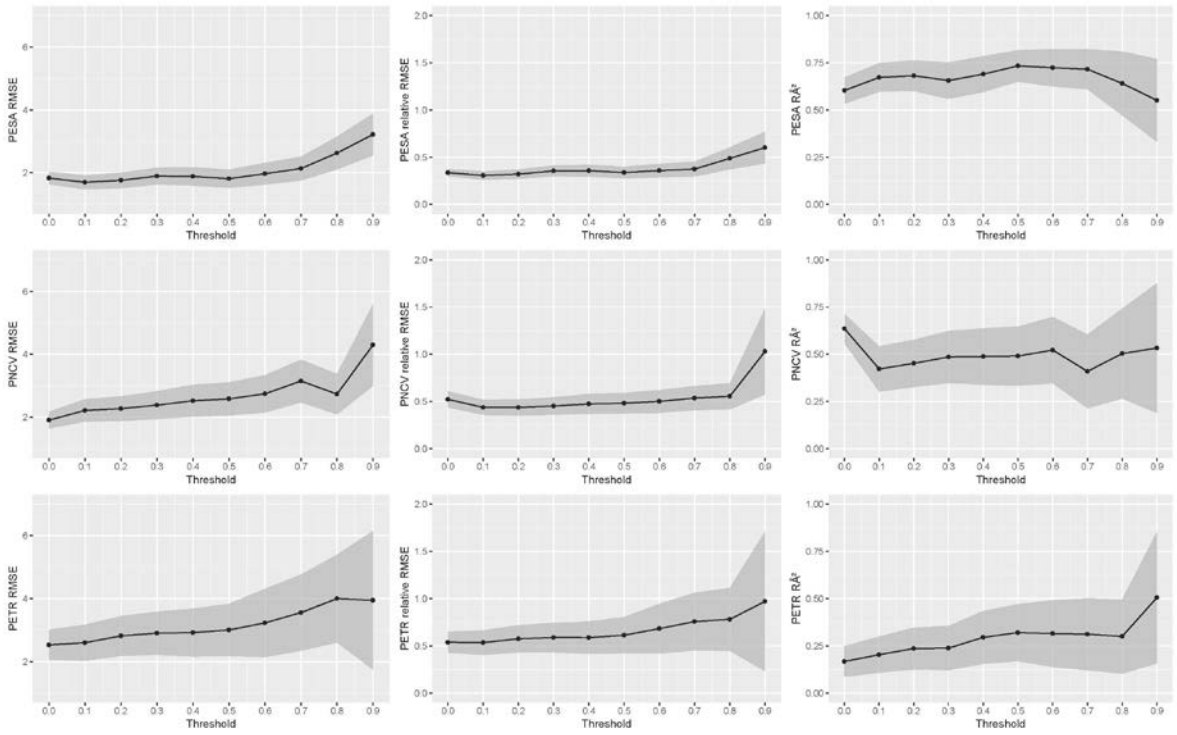


Figure A-5: Averaged carbon model results in terms of RMSE, relative RMSE and R^2 after 1000 iteration for all three study sites. The grey shaded area along the curves show +/- one standard deviation around the mean performance measures.

Generally, the decreasing number of training pixels led to less robust results with a higher standard deviation in the performance measures along with an overall performance loss (increase in RMSE and RMSErel). In terms of R^2, the trend of performance loss with decreasing sample size was not always observed, although smaller samples always resulted in higher variation between models (higher standard deviation).

4 Discussion

Our hypothesis was that, in heterogeneous landscapes, the use of auxiliary information from high spatial resolution remote sensing data improves the upscaling of field-sampled aboveground carbon data to meso-scale remote sensing image pixels, thus enabling their use for carbon mapping, potentially over large areas.

Indeed, we observed that the use of auxiliary data did improve the data integration, although the choice of the method used can be influential. On the one hand, the lack of auxiliary information (with a cartographic method) never delivered the best results in upscaling the field-collected data. Its assumption of homogeneity in the distribution of aboveground carbon is not adequate in such heterogeneous systems. On the other hand, when disaggregating the field samples into different target pixels, the use of auxiliary data in a regression approach for data disaggregation resulted in the poorest results for data

disaggregation. This also agrees with what has been previously found by e.g. Fisher and Langford (1995). As regression models are fitted globally, estimates from locally fitted methods should adapt better to heterogeneous environments. When extrapolating the field samples into the full (partially non-sampled) pixels, however, the regression method was the best performing approach.

We found that the best approach for upscaling field-collected aboveground carbon data makes use of auxiliary data in a (local) surface method for data disaggregation, and in a regression for data extrapolation.

Our conclusions on the use of high spatial resolution as auxiliary data agrees with findings in a similar system, the Argentinean savannahs (González-Roglich and Swenson 2016), which suggests the generality of our approach for savannahs and other heterogeneous systems. In the referred study, the authors used fine resolution satellite imagery for scaling up field data, to assess tree cover at the meso (Landsat) scale, which was later related to carbon. In homogeneous environments, however, such an approach is not necessary and the co-registration of field samples and image pixels is commonly done through the estimation of plot-level tree density values in a dasymetric approach (McRoberts and Tomppo 2007; Tuominen et al. 2010).

The use of multi-scale remote sensing imagery for vegetation monitoring has been widely used for e.g. generating maps of forest biomass or productivity over large areas (Lefsky et al. 2005; Muukkonen and Heiskanen 2007; Tomppo et al. 2002). The choice of the higher-resolution data to be used as auxiliary information in this approach is critical for successfully upscaling field data, as it needs to relate to the field measured variable - in our case, the vegetation's aboveground carbon. Here, we used a spectral index based on the Red Edge spectral of RapidEye imagery, known to relate to vegetation structure and therefore biomass in the Cerrado (Gomes and Maillard 2015). Further research is still required to learn about the best possible data to be used as a weighting layer. This, however, falls outside the scope of this study and would raise issues related to data availability constraints. Ultimately, the data used as auxiliary layer will determine the estimated regression coefficient used in this approach. Also, the approach presented here could potentially be used to integrate field data with multi-scale systems, such as that of Sentinel-2, which collects large amounts of data across the globe on a high frequency (Drusch et al. 2012). In this case, e.g. the 10m data could be used for the upscaling of field data to the 20m pixels.

Our analyses also showed that, while a more restrictive (high) threshold on the share of sampled (versus estimated) data should ensure better data reliability, it also results in fewer training pixels which in turn generate less good models. Indeed, smaller sample sizes usually mean that a smaller proportion of the system's variability is captured, thus resulting in less generalizable estimations (Wisz et al. 2008).

Upscaling field samples to target pixels enables the use of remote sensing imagery for carbon mapping and ecosystem monitoring in heterogeneous environments. Through this approach it is possible to e.g. do wall-to-wall mapping of carbon over large areas with time-series of widely available multi-spectral imagery (Wulder et al. 2015a), or characterize particular areas with high detail with spaceborne hyperspectral imagery (Leitão et al. 2015) (Guanter et al. 2015). Ultimately, this will have deep implications for global carbon mitigation programs such as REDD+, by allowing the detailed calculation of above-ground carbon in a spatially explicit manner.

5 Conclusion

Using high spatial resolution remote sensing imagery as auxiliary data is beneficial for the spatial allocation of field sampled data to a larger target pixel. This is particularly relevant in heterogeneous environments, where it is not possible to define homogeneous plots of known vegetation density. The method for integrating the auxiliary data in the analysis is however not trivial and can have a great influence in its overall performance. While local, surface approaches are preferable for the spatial disaggregation (interpolation) of field samples to the target pixel grid, the extrapolation of the data into the full pixel extent is better done with a global, regression model. This approach enables the spatial allocation of field data to larger image pixels, thus allowing the use of remote sensing imagery for ecosystem monitoring and carbon mapping on heterogeneous areas.

Acknowledgements

This study is part of the research activities of the EnMAP Scientific Advisory Group (EnSAG), and was funded by the German Aerospace Centre (DLR) — Project Management Agency, granted by the Ministry of Economics and Technology (BMWi grant 50EE1309). Underlying RapidEye data has been contributed on behalf of the German Aerospace Center through funding of the Federal Ministry of Economy and Energy (RESA ID 00186). The field data were collected within the project CNPq 457497/2012-2 and Edital Sisbiota - CNPq nº 47/2010 (CNPq 563134/2010-0, Projeto Diversidade biológica do Cerrado: estrutura e padrões). J.R.R.Pinto benefitted from a Productivity fellowship (CNPq 307701/2014-0).

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne Verwendung unerlaubter Hilfe angefertigt zu haben. Die aus fremden Quellen direkt oder indirekt übernommenen Inhalte sind als solche kenntlich gemacht. Die Dissertation wird erstmalig und nur an der Humboldt-Universität zu Berlin eingereicht. Weiterhin erkläre ich, nicht bereits einen Doktortitel im Fach Geographie zu besitzen. Die dem Verfahren zu Grunde liegende Promotionsordnung ist mir bekannt.

Marcel Schwieder

Berlin, den 06.03.2018