

Multidimensional remote sensing based mapping of tropical forests and their dynamics



Loïc Dutrieux

Multidimensional remote sensing based mapping of tropical forests and their dynamics

Loïc Paul Dutrieux

Thesis committee

Promotor

Prof. Dr M. Herold
Professor of Geo-information Science and Remote Sensing
Wageningen University

Co-promotors

Dr L. Kooistra
Assistant Professor, Laboratory of Geo-information Science and Remote Sensing
Wageningen University

Prof. Dr L. Poorter
Personal chair at the Forest Ecology and Forest Management Group
Wageningen University

Other members

Prof. Dr B.J.M. Arts, Wageningen University
Dr N. Pettorelli, Zoological Society of London, United Kingdom
Dr M. Equihua Zamora, Instituto de Ecología, Xalapa, Mexico
Prof. Dr B. Somers, KU Leuven, Belgium

This research was conducted under the auspices of the C.T. de Wit Graduate School
of Production Ecology & Resource Conservation (PE&RC)

Multidimensional remote sensing based mapping of tropical forests and their dynamics

Loïc Paul Dutrieux

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 13 September 2016

at 1.30 p.m. in the Aula.

Loïc Paul Dutrieux

Multidimensional remote sensing based mapping of tropical forests and their dynamics
168 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)

With references, with summary in English

ISBN 978-94-6257-890-6

DOI 10.18174/387682

Acknowledgments

It all started about seven years ago, while sitting in my irrigation class (that's right, I did not initially come to Wageningen to study remote sensing), when the teacher showed a beautiful remote sensing based evapotranspiration map of the Nile delta. At that time I knew nothing about remote sensing, but I immediately knew that it was what I wanted to do. And I did. Then two years later, while finishing my Master thesis defense about remote sensing, I dared to ask the examiners if they knew of a job for me, or at least if they knew of something that could suit me. It turns out this was a good idea for if I had not asked that question, I wouldn't be writing these acknowledgments right now. Prof. Martin Herold called me a few days later to ask if I was still interested and I started almost immediately with a short three month contract that quickly evolved into a PhD. So thank you Martin for believing in my potential, giving me this opportunity to do a PhD, becoming my promotor, and guiding me through the PhD process.

"It's just about writing four paper, how hard can it be?" I asked myself... I had never been so wrong! Although the process went quite well overall, it was far from easy to remain constantly motivated, enthusiastic and focused during these four years. Fortunately many people have directly or indirectly helped me during this period; in staying focus when I had to work, and in enjoying my free time as much as possible and preserving my sanity. I would like to thank all these people.

I first would like to express my gratitude to my supervisors Lourens and Lammert who did a great job in helping me stayed focused and motivated during this long process. Thank you Lammert for your dedication, particularly during the last intense year; it really felt like we were in it together and that helped me a lot. Thank you Lourens for your contagious enthusiasm, for having the patience to listen to long remote sensing stories, and for your always meticulous and pertinent feedback from which I have learnt a lot. I feel privileged to have worked with the two of you.

I had the chance as well to be part of a great group with fantastic colleagues. The atmosphere of the group, back when we were all working in Horapark had in fact been an important aspect that convinced me to accept the PhD position, and stay in Wageningen for four extra years. I had no doubt that it would always remain a pleasure to work in this environment and that turned out to be the case. This is therefore a lot of peo-

ple I wish to thank, hoping I am not forgetting anyone. Antoinette, Truus, Roberto, Richard, Johannes, Valérie, Nandika, Arun, Valerio, Giulia, Sarah, Astrid, Arnold, Sytze, Harm, Jan, Lucasz, Ron, Arend, Erika, Juha, Aldo, Willy, Philip, John, Eskender, Yang, Mathieu, Patrick, Maria, Michael, José, Marston, Konstantin, Eliakim, Qijun, Alvaro, Kalkidan, Sidney, Marcio, Jalal, and Brice; thank you all for the moments we shared, at work, during lunch, at the coffee machine, and also after work for some of you; even when the following mornings were not always so enjoyable. Also, how to forget the fantastic holidays in Canada with Niki and Ben, or the many dinners and parties with Simon; thanks guys! And to my office mates, Titia, Kim, Peter and Johannes; this is not over... I will get my revenge at the office basketball tournament.

During this period, scientific exchange and collaboration went beyond my direct colleagues, and that's something I particularly appreciated. I wish to thank all the students I had the chance to work with; Benjamin, Latifah and Jorn, it was a great pleasure to work with you and I hope you learned as much as I did working with you. I am also grateful to Nataly and Marisol from IBIF, for our collaboration that made the second chapter of this thesis possible. Finally thank you Catarina for the very pleasant collaboration; the work we did together is probably the part of my thesis I enjoyed working on the most.

I had the chance to collaborate with many more people as well since my work was part of the famous ROBIN project; a mega project with about 15 partner institutions in Latin America and Europe, amounting literally hundreds of researchers. Thanks to all the ROBIN people for the scientific exchanges and the good times, and in particular thanks to Masha, Marielos, Gerbert, Boris, Julian, Michael, Melanie, and Maggie for the fun moments in Wageningen, Bolivia, Mexico, Brazil and Spain.

Although the PhD lasted four years (and a few months), Wageningen had been my home for more than seven years. That's a long time, with lots of memories and many people I met along the way, and I will always remember this period. Romain, Sebastien, Nicolas, Gerson, Bérénice, Esther, Maria, Lupita, Linda, Noelia, Carolina, Fritz, the glotzbach family, Francesco, Elena, Meto, Jean-Carlos, Amelie, Coean, Nayara, Gatien, Nelson, André, João, Maíra, Eduardo, Montse, Alejandra and Marlene are probably just a few of the people with whom I share these memories. Big thanks as well to the many housemates I had during that period, Francesco, Jan, Filippo, Stefano, Marta, Alicja, Joseph, Amandine, Arjaan & Saskia, you've all been like family to me.

There are also people I do not know directly, but who's work has helped me a lot, or at least has contributed to keeping me inspired. My first thoughts go to those who have inspired me, there are of course many, but I would like to thank in particular Jean Claude Ameisen for his wonderful science tales that millions of people can enjoy on French radio. I am a partisan of open science and open source; the entire thesis was produced using open source software that many have developed and improved thanks to their hard work and dedication. I am grateful to the entire open source community for building in a

transparent way reliable tools accessible to all, and in particular to Robert, Edzer, Roger, Hadley and Achim. Doing science is a lot about being inspired and creative as well, and music, by filling my ears with joy and fueling my motivation, has often been my muse. So thanks to all the musicians, bands, singers and DJs for all that creative juice.

Finally, I would like to express my deep appreciation to my family who has supported me. To my parents who always encouraged me, first to listen in school and later to pursue higher education studies, and to my brother Cyril (who by the way wrote half of the regex I ever used), my sister in law Aurore, and my niece Ambre.

I have no doubt that this thesis and my time in Wageningen would not have been the same without the contribution and support of you all.

Thank you!

Loïc

Summary

Tropical forests concentrate a large part of the terrestrial biodiversity, provide important resources, and deliver many ecosystem services such as climate regulation, carbon sequestration, and hence climate change mitigation. While in the current context of anthropogenic pressure these forests are threatened by deforestation, forest degradation and climate change, they also have shown to be, in certain cases, highly resilient and able to recover from disturbances. Quantitative measures of forest resources and insights into their dynamics and functioning are therefore crucial in this context of climate and land use change. Sensors on-board satellites have been collecting a large variety of data about the surface of the earth in a systematic and objective way, making remote sensing a tool that holds tremendous potential for mapping and monitoring the earth. The main aim of this research is to explore the potential of remote sensing for mapping forest attributes and dynamics. Tropical South America, which contains the largest area of tropical forest on the planet, and is therefore of global significance, is the regional focus of the research. Different methods are developed and assessed to: (i) map forest attributes at national scale, (ii) detect forest cover loss, (iii) quantify land use intensity over shifting cultivation landscapes, and (iv) measure spectral recovery and resilience of regrowing forests.

Remote sensing data are diverse and multidimensional; a constellation of satellite sensors collects data at various spatial, temporal and spectral resolutions, which can be used to inform on different components of forests and their dynamics. To better map and monitor ecological processes, which are inherently multidimensional, this thesis develops methods that combine multiple data sources, and integrate the spatial, temporal and spectral dimensions contained in remote sensing datasets. This is achieved for instance by assembling time-series to fully exploit the temporal signal contained in the data, or by working with multiple spectral channels as a way to better capture subtle ecological features and processes.

After introducing the general objectives of the thesis in Chapter 1, Chapter 2 presents an approach for mapping forest attributes at national scale. In this chapter, 28 coarse resolution remote sensing predictors from diverse sources are used in combination with *in-situ* data from 220 forest inventory plots to predict nine forest attributes over low-land Bolivia. The attributes include traditional forest inventory variables such as forest

structure, floristic properties, and abundance of life forms. Modelling is done using the random forest approach and reasonable prediction potential was found for variables related to floristic properties, while forest attributes relating to structure had a low prediction potential. This methodological development demonstrates the potential of coarse resolution remote sensing for scaling local *in-situ* ecological measurements to country-wide maps, thus providing information that is highly valuable for biodiversity conservation, resource use planning, and for understanding tropical forest functioning.

Chapter 3 presents an approach to detect forest cover loss from remote sensing time-series. While change detection has been the object of many studies, the novel contribution of the present example concerns the capacity to detect change in environments with strong inter-annual variations, such as seasonally dry tropical forests. By combining Landsat with Moderate Resolution Imaging Spectroradiometer (MODIS) time-series in a change detection framework, the approach provides information at 30 m resolution on forest cover loss, while normalizing for the natural variability of the ecosystem that would otherwise be detected as change. The proposed approach of combining two data streams at different spatial resolutions provides the opportunity to distinguish anthropogenic disturbances from natural change in tropical forests.

Chapter 4 introduces a new method to quantify land use intensity in swidden agriculture systems, using remote sensing time-series. Land use intensity — a parameter known for influencing forest resilience — is retrieved in this case by applying a temporal segmentation algorithm derived from the econometrics field and capable of identifying shifts in land dynamic regimes, to Landsat time-series. These shifts, or breakpoints, are then classified into the different events of the swidden agriculture cycle, which allows to quantify the number of cultivation cycles that has taken place for a given agricultural field. The method enables the production of objective and spatially continuous information on land use intensity for large areas, hence benefiting the study of spatio-temporal patterns of land use and the resulting forest resilience. The results were validated against an independent dataset of reported cultivation frequency and proved to be a reliable indicator of land use intensity.

Chapter 5 further explores the concept of forest resilience. A framework to quantify spectral recovery time of forests that regrow after disturbance is developed, and applied to regrowing forests of the Amazon. Spatial patterns of spectral resilience as well as relations with environmental conditions are explored. Regrowing forests take on average 7.8 years to recover their spectral properties, and large variations in spectral recovery time occur at a local scale. This large local variability suggests that local factors, rather than climate, drive the spectral recovery of tropical forests. While spectral recovery times do not directly correspond to the time required for complete recovery of the biomass and species pool of tropical forests, they provide an indication on the kinetics of the early stages of forest regrowth.

Chapter 6 summarizes the main findings of the thesis and provides additional reflections and prospects for future research. By predicting forest attributes country-wide or retrieving land use history over the 30 years time-span of the Landsat archive, the developed methods provide insights at spatial and temporal scales that are beyond the reach of ground based data collection methods. Remote sensing was therefore able to provide valuable information for better understanding, managing and conserving tropical forest ecosystems, and this was partly achieved by combining multiple sources of data and taking advantage of the available remote sensing dimensions. However, the work presented only explores a small part of the potential of remote sensing, so that future research should intensively focus on further exploiting the multiple dimensions and multi-scale nature of remote sensing data as a way to provide insights on complex multi-scale processes such as interactions between climate change, anthropogenic pressure, and ecological processes. Inspired by recent advances in operational forest monitoring, operationalization of scientific methods to retrieve ecological variables from remote sensing is also discussed. Such transfer of scientific advances to operational platforms that can automatically produce and update ecologically relevant variables globally would largely benefit ecological research, public awareness and the conservation and wise use of natural resources.

Contents

	Page
Acknowledgements	v
Summary	ix
Contents	xiii
Chapter 1 Introduction	1
Chapter 2 Country wide mapping of forest diversity and structure	11
Chapter 3 Monitoring forest cover loss using multiple data streams	35
Chapter 4 Reconstructing land use history from Landsat time-series	63
Chapter 5 Post-disturbance recovery in forest spectral properties across the Amazon	87
Chapter 6 Synthesis	113
References	129
Curriculum Vitae	149
PE&RC Training and Education Statement	151

Chapter 1

Introduction

1.1 When the forest disappeared

Wandering through an ancient Mayan city of Mesoamerica, you may not only witness remnant of pyramids, temples and other Mayan structures, you are also likely to observe luxurious tropical vegetation that most would label as pristine forests. But how could this vegetation be pristine if people have lived in these places? The answer is rather simple; it isn't; these forests regrew and are therefore old secondary forests. About 1300 years ago, during the Mayan late classic Period, Mesoamerica hosted a large population with some of the largest cities of the region exceeding 100,000 inhabitants (Chase & Chase, 1994). The Maya civilization thrived during this period enabling the development of an elaborate and complex society organized around multiple large cities. But with such a high population density, the Mayas also had a large impact on their natural environment (Wahl et al., 2006). Many of these ancient cities were abandoned following the rapid 150 years collapse of the Mayan classic period, between 750 and 900 AD, giving a chance for nature to reconquer these territories. Theories attempting to explain the sudden decline of such a flourishing civilization often highlight the role that deforestation played in the collapse (Oglesby et al., 2010; Cook et al., 2012). Cook et al. (2012) for instance explain that large scale deforestation over Mesoamerica would have amplified a long lasting period of drought resulting in failed harvests, which combined with other political and social factors, may have led to the collapse (Hodell et al., 1995). While the precise reason of the collapse will always remain a mystery (there were no scientific instruments available at that time to record human impact on forests, and its relations to regional climates and population dynamics), it is certain that: (i) The Maya had a large impact on their natural environment which in turn affected their livelihood, and (ii) despite extensive deforestation in the classic period, forests of Mesoamerica recovered. The former highlights the importance of forests in maintaining our environment and ourselves, while the latter is a proof that ecosystems are resilient albeit in this case at temporal scales that are beyond human perception.

Today, tropical forests are once again being put under great pressure by human activities, and it is urgent to gain insights on the state of our forests and on the various processes and dynamics that govern them. Such insights and related policy actions are key to ensure a future with healthy tropical forests and the associated benefits they provide to humankind. The scientific and technological landscape is obviously very different today than it was during the peak of the Maya civilization, and these advances might contribute to gaining the required insight for sustainably managing the forests. It is in this context that the present thesis takes place, the main objectives of the research being to investigate the potential and opportunities of remote sensing techniques to map and monitor forest attributes and dynamics.

1.2 Forests and their dynamics

With an estimated area of 42 million km², forests cover approximately 30% of the global land surface (Bonan, 2008). Not only do forests cover the majority of the land surface, they also play an important role in the provision of ecosystem services. With climate change being at the centre of many international discussions, the role of forests in regulating global climatic patterns is often particularly emphasized (Stocker et al., 2014). The mechanism behind this climate regulation role lies in the carbon stored in the above and below ground biomass of the forests (Saatchi et al., 2011). Removal of this woody biomass, which is a result of deforestation or forest degradation often coincides with atmospheric carbon emissions, which contributes to amplifying the green house effect and fuels global warming (Bonan, 2008). The latest reports estimate that tropical forests store 231 Pg of carbon in their biomass (total biomass of 462 Pg) (Avitabile et al., 2016), and that forest clearing accounts for about 10% of the global CO₂ emissions (Le Quéré et al., 2015). Both biomass stocks and dynamics matter when considering the climate change mitigation role of forests. Growing forests, with positive net primary productivity, are often seen as a carbon sinks (Poorter et al., 2016; Chazdon et al., 2016), using photosynthesis to transform CO₂ into sugar, which is then used to produce leaves, stems and roots. Additionally, forests also contribute to climate regulation at local and regional scale by cycling water through transpiration (Bonan, 2008). However, local to global climate regulation is not the only service forests provide. Tropical forests host a major part of terrestrial biodiversity as shown by the many hotspots they concentrate (Myers et al., 2000), and recent estimates report the probable presence of 40,000 to 53,000 tree species in tropical regions alone (Slik et al., 2015). Local communities also largely benefit from the surrounding forests. The livelihood of riverine Amazonian populations largely rely on their traditional swidden agricultural system — a rotational system that alternate between forest and agriculture. Their use of the secondary forests as a way to enhance land fertility and maintain its productivity is an example of such services provided by forests to local populations (Coomes et al., 2000; Jakovac, 2015). A variety of dynamics govern the current and future state of forests and their associated services. Deforestation, forest degradation, urban and agriculture expansion are part of the drivers that threaten forests and directly result in forest cover loss (Hosonuma et al., 2012). While forests have the potential to regrow and recover their biomass, structure and species composition, the speed of this process varies spatially and is influenced by factors such as climate (Poorter et al., 2016), local environmental conditions, and the intensity of previous disturbances (Jakovac et al., 2015). All these processes that shape the forests, operate at different spatio-temporal scales. For instance forest degradation is by definition a very localized process, as opposed to deforestation which may apply to relatively large areas. Figure 1.1 provides an overview of key forest processes related to forest cover change and forest response to climate and the spatio-temporal scale at which they occur.

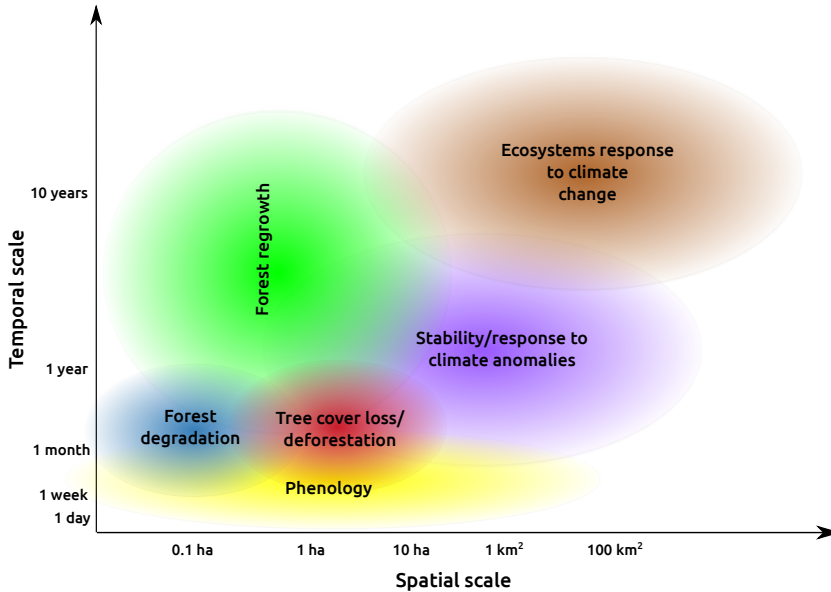


Figure 1.1: Six key forest processes related to forest cover change and response to climate, and the spatial and temporal scales at which they occur.

1.3 Observing, measuring and monitoring forests from space

Remote sensing instruments on-board satellites have been acquiring data of the surface of the earth in an objective and systematic way for more than 40 years (Goward et al., 2006). Because they have been collected by a variety of sensors, the data diversity available is huge, with every dataset having different spatial, temporal and thematic characteristics. Facing the abundance and diversity of data it provides, remote sensing could be thought as the ideal tool for addressing most of the mapping and monitoring activities the scientific communities require to better understand, protect and manage forests. However, sensors on board satellites do not directly collect ecologically meaningful information; for instance optical sensors, measure radiance, which in itself is not of much use for resource use and conservation, forest managers, and policy makers. For remote sensing to be a useful tool for forest mapping and monitoring, there is therefore a need to transform the raw data collected by the remote sensing sensors into meaningful biophysical variables and information about ecological and anthropogenic processes. Given the diversity of remote sensing data available, possibilities to approach every specific mapping objective are numerous, and choosing the appropriate approach and scale largely depends on the

process being observed. I briefly introduce below the process of converting raw remote sensing measurements of vegetation into meaningful biophysical variables, then because a large part of this thesis implements time-series approach to retrieve forest processes, I describe some details of using the temporal dimension of remote sensing data, and finally I elaborate on the multidimensional nature of remote sensing data.

1.3.1 From raw remote sensing data to estimating biophysical variables

The Normalized Difference Vegetation Index (NDVI) is undoubtedly the most notorious vegetation index derived from optical remote sensing and is an example of a simple yet widely used biophysical variable (Tucker, 1979). By taking advantage of the difference in reflectance of green vegetation in the Near Infra-Red (NIR) and red parts of the light spectrum, the NDVI is an indicator of vegetation greenness and health, and has been shown to relate closely to the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) (Myneni & Williams, 1994). Therefore NDVI, which is a simple transformation of the raw data gathered by most optical remote sensing sensors, has nearly become a reference biophysical variable in itself and has acquired a certain legacy among remote sensing scientists and ecologists. However, NDVI is not always sufficient, so that more elaborated methods have been developed for extracting other forms of information from remote sensing data. Remote sensing has gone a long way since NDVI was invented and methods have been developed that are able to retrieve parameters ranging from energy fluxes (Running et al., 2004), to tree level functional traits (Asner et al., 2014b), tree cover dynamics (Hansen et al., 2013), and landscape structure (Broadbent et al., 2008a). Some of the most advanced methods for ecological applications include the work of Asner and his collaborators (Asner et al., 2007; Asner & Martin, 2008; Asner et al., 2014b). Their work, by combining Light Detection And Ranging (LiDAR) with hyperspectral optical data, enables detailed mapping of functional traits at the level of individual trees. Such work currently represents the state of the art in terms of ecological applications of remote sensing. Although global deployment of these technologies remains at the moment prohibitively expensive, these examples provide an outlook on what may become possible in the future. The potential for mapping ecological processes with freely and globally available data is evidently more limited, however, the scientific community has already largely taken advantage of this abundant source of data to develop applications relevant to forest mapping. There is for instance a large body of work on forest cover loss, detection with Landsat time-series, which is presented in the next paragraph.

1.3.2 Time-series approaches for monitoring forest dynamics

Dynamics refer to change in properties over time, so that forest dynamics and processes contain an obvious time dimension, and therefore require the use of time-series approaches

to be measured. Assembling time-series from remote sensing data is possible, and many recent remote sensing advances aimed at mapping of forest dynamics have used time-series assembled from Landsat data for their analysis. Landsat is a constellation of satellite carrying multispectral sensors that have delivered data of the surface of the earth at 30 m resolution for more than 40 years (Roy et al., 2014). The sub-hectare resolution combined with multi-spectral characteristics, the availability of repeated measurements of a same area since the 1970s, and free access makes Landsat data the ideal dataset for investigating forest dynamics from local to global scale (Wulder et al., 2012). More specific uses of Landsat for forest dynamics include deforestation monitoring (Zhu et al., 2012; DeVries et al., 2015b; Dutrieux et al., 2015), temporal trajectories analysis (Huang et al., 2010; Kennedy et al., 2010; Dutrieux et al., 2016) and forest regrowth (DeVries et al., 2015a; Hansen et al., 2013). Today, Landsat data are even taken to the operational level with as an example the recent development of a near real time deforestation alert system (Hansen et al., 2016).

One challenging aspect of investigating changes in vegetation time-series is that vegetation is constantly changing as part of its natural seasonal dynamics, associated with cold periods in the temperate zone, and dry periods in the tropical zone. Whether it is to study trends, shifts in temporal trajectories, or abrupt changes, or other anthropogenic processes, it is therefore important to account for vegetation seasonality. Seasonality was taken into account in this thesis by using time-series regression modelling of the seasonal behaviours (Verbesselt et al., 2010b,a). Although this thesis did not make extensive use of these sensors in a time-series context, potential to study vegetation dynamics and process, particularly at regional and global scale, extends beyond Landsat. Coarse resolution data collected by sensors like the Moderate Resolution Imaging Spectroradiometer (MODIS) or Advanced Very High Resolution Radiometer (AVHRR) have been used to study global vegetation trends (de Jong et al., 2011) or vegetation response to climate change (Jong et al., 2013).

1.3.3 Remote sensing: towards a multi-dimensional approach

Figure 1.1 presents six forest dynamical processes along two, spatial and temporal, axes. Because they occur at different spatial and temporal scales, these forest dynamics contain at least two dimensions. Additionally, the dynamics involve a continuous development of different characteristics of the forests. For instance, when forests regrow, biomass increases, together with changes in horizontal and vertical structure, as well as species composition. Because of these multiple aspects changing in both space and time, forest dynamics can be perceived as a multidimensional process. Remote sensing data are multidimensional too. First, the spatial dimension is ubiquitous in remote sensing data. In the case of optical sensors, radiance is exhaustively sampled in space, at varying resolutions, depending on the sensors characteristics. Secondly, the same areas are measured

repeatedly over time, enabling the data to be assembled in time-series. Remote sensing data therefore contain a temporal dimension, which is particularly required when studying dynamics. Thirdly, optical remote sensing instruments collect data for a range of spectral channels, making the spectral domain the third remote sensing dimension. The structure and chemistry of the observed targets define the amount of light they reflect at various wavelengths, and therefore their spectral signature (Knipling, 1970; Curran, 1989). Lastly, the variety of sensors and technologies that acquire data constitute a fourth remote sensing dimension, which I call here thematic dimension. The different families of sensors (i.e. optical, radar, and Light Detection And Ranging (LiDAR)) inform on different elements of an observed area. In the case of forest for instance, the optical signal is mainly influenced by the structure and the chemistry of the upper canopy elements, while the radar signal penetrates deeper in the canopy and informs on the structure of woody components, and LiDAR retrieves information about the vertical structure of the forest. Remote sensing data are therefore multidimensional, as they are composed of four dimensions, namely the spatial, temporal, spectral, and thematic dimension. Taking advantage of these multiple dimensions and the synergies that exist between them to better capture ecological processes is a general objective of this thesis.

1.4 Problem statement and research objectives

The Amazon contains the largest area of tropical forest on the planet and it plays a major role in carbon sequestration, water cycling, and biodiversity preservation. However, in the current context of climate and land use change, tropical forests and the services they deliver are under great pressure. Forest dynamics, such as deforestation, forest regrowth and forest response to climate change, largely determine the state of the present and future forests and the services they deliver. Deforestation for instance has a direct impact on the carbon cycle and often results in biodiversity loss. Following deforestation, the use that is made of the land determines whether or not a secondary forest will replace the original forest. When given the chance, the capacity of a forest to recover is determined by its resilience, which depends on many natural and anthropogenic factors. In swidden agriculture systems for instance, resilience is largely determined by the frequency of previous events of forest slashing and burning. Because the existence and state of forests is so intertwined with forest dynamics, we must gain further insights on current states of forests and processes that govern them, so that we can ensure successful use, conservation and management of forest resources, as well as effective policy actions aiming at conserving them. Although, remote sensing offers a tremendous potential to map forest resources and monitor the dynamics that govern forest state, a large part of its potential remains un-explored. There is a need to develop innovative remote sensing methods that are able to provide assistance for forest resources accounting by mapping forest state, and that can provide spatio-temporal information on forest dynamics. The main objective of this

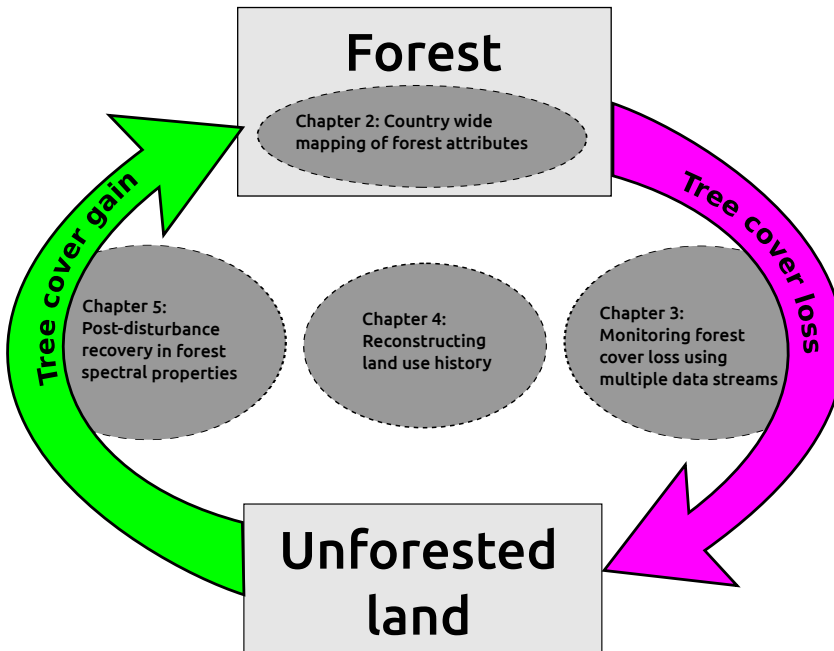


Figure 1.2: Graphical representation of the four core chapters of this thesis relative to forest dynamics. The ellipses indicate the chapters, the boxes represent the two, forested and unforested, states, and the arrows symbolize the dynamic processes involved in the transitions between these states.

this thesis is to explore remote sensing approaches that combine multiple remote sensing dimensions for better mapping of tropical forest resources and monitoring of various forests dynamics. This is done by developing and assessing the potential different innovative approaches that combine *in-situ* and remote sensing data and make use of methods from the machine learning and the econometrics fields.

1.5 Thesis overview

The presentation of the work I conducted is articulated around six chapters, four of which are direct research results presented in the form of scientific publications. As described earlier, the work and methods developed was done for tropical forests, with a regional focus on South America. The outline of the four core chapters is summarized in Figure 1.2.

Chapter 2 concentrates on national scale mapping of forest attributes. I have developed

and tested an approach that combines *in-situ* forest inventory data with coarse resolution remote sensing from multiple sources to predict and map forest attributes. I tested the approach over lowland Bolivia using *in-situ* data from 220 permanent plots from Bolivia's national forest inventory for model training and validation. By combining these 220 plots data with 28 remote sensing derived predictors I assessed the prediction potential for nine forest attributes relating either to forest structure, floristic properties, and abundance of life forms.

Chapter 3 focuses on automatic detection of forest cover loss using remote sensing time-series. More specifically, I developed and assessed the performances of a forest cover change monitoring approach tailored at forests with a large inter-annual variability in their seasonal cycle. Tropical dry forests commonly present such features so that I developed and assessed the approach over a small area of tropical dry forest in lowland Bolivia. The innovativeness of the approach lies in the combined use of two data streams at different resolutions enabling human induced change to be distinguished from natural dynamics.

Chapter 4 extends the work on forest dynamics by not only looking at tree cover loss but by retrieving forest cover trajectories over the last 30 years. Temporal trajectories, which are the result of a temporal segmentation algorithm, are used to quantify the intensity with which the land was used in a swidden agriculture rotational cultivation system. Land use intensity in these systems largely influences the capacity of the forests to regenerate, so that the land use intensity estimates returned by the method provide spatial information relating to forest resilience. The study area used for developing and validating this method is located in the Brazilian Amazon, where swidden agriculture is the cultivation type that dominates the landscape around riverine Amazon villages.

Chapter 5 continues with temporal trajectories and forest resilience. In this chapter I approached regrowing forests of the Amazon basin with a fully spectro-temporal perspective. This approach allows to investigate how spectral recovery dynamics vary spatially and relate to climatic conditions.

Finally the thesis concludes with a synthesis chapter that provides an overview of the main findings and their implications. It provides a case for the need of a multidimensional remote sensing approach, and it gives recommendations and prospects for the future of ecological applications of remote sensing.

Chapter 2

Country wide mapping of forest diversity and structure

This chapter is based on:

Dutrieux, L. P., Poorter, L., Equihua, J., Ascarrunz, N., Herold, M., Peña-Claros, M., Roerink, G., Toledo, M. & Kooistra, L. (2016). Country wide mapping of forest diversity and structure by combining forest inventories with remote sensing. In prep.

Abstract

Species conservation, international initiatives and research in forest ecology can greatly benefit from maps of forest attributes. While national forest inventories have contributed large amount of data covering different areas in the tropics, high costs and various difficulties associated with data collection prevent a wide coverage in space and time. The present study explores the potential of moderate resolution remote sensing to predict forest attributes, with the purpose of producing spatially continuous maps for entire countries or even larger regions. Using a set of 28 globally available remote sensing layers and *in-situ* data from a network of 220 permanent plots, we predicted for every 1 km² covering lowland Bolivia forest inventory variables related to forest structure, floristic composition, and the abundance of functional groups. The whole modelling approach consists in (i) extracting the data contained in pixels overlapping with *in-situ* sites location, (ii) training a Random Forest regression and (iii) predicting the response variable for every pixel covering lowland Bolivia. We assessed model performances for nine response variables and found strong predictive power for floristic composition variables (proxies of species composition and richness from a list of 100 reliably identified species), but limited potential for forest structure variables (e.g. tree density, basal area). We explain the limited potential to predict forest structural properties by a strong disconnection between the attributes measured via optical remote sensing and the response variables considered. Following a variable selection step we found phenology, Leaf Area Index (LAI), and surface reflectance to be highly represented in the list of important predictors, making them key predictors for remote sensing based forest floristic properties modelling. The floristic composition maps present coherent spatial patterns, following the well known rainfall and environmental gradients of lowland Bolivia. Such maps of species composition and diversity and the approach we propose can greatly benefit the ecological understanding of tropical systems as well as contribute to conservation initiatives. Our methodological approach allows the scaling of *in-situ* ecological measurements for national-scale mapping activities and enables the production of information highly valuable to biodiversity conservation and our understanding of tropical forest functioning.

2.1 Introduction

Tropical forests cover about 30% of the continental surface of the earth (Bonan, 2008), concentrate a large part of terrestrial biodiversity (Pimm et al., 2014) including 96% of the global tree species (Fine et al., 2009), and play a major climate regulation role via the carbon they store in their biomass (Stocker et al., 2014). It is crucial, for conservation, resource use, international commitments on green house gases emissions and biodiversity, as well as for understanding mechanisms that drive the organisation of tropical forests, to map and monitor species composition, richness, and forest structure (Watt, 1947; Chave, 2008; Sexton et al., 2015). Forest inventories data have been for a long time the only source to address these questions (Ter Steege et al., 2006; Poorter et al., 2015), however, particularly in tropical regions, these are often difficult and expensive to acquire due to remoteness and low accessibility. Combining forest inventory data with remote sensing might provide a solution to these constraints. Since the 1970s, a variety of remote sensing instruments orbiting the earth have been systematically acquiring large amounts of data, covering every location of the earth. Such data streams therefore provide a great opportunity to map forest variables with large spatial coverage and high spatial detail while maintaining low costs.

However, one important challenge with satellite-based forest attribute mapping is that remote sensing measurements do not directly translate into ecologically meaningful variables (Skidmore et al., 2015). Optical sensors for instance measure reflected light, which in itself is not a useful trait for forest conservation or management. The raw information acquired by the sensors can sometimes be transformed into meaningful biophysical variables (Asner & Martin, 2008; Asner et al., 2014a,b). Asner and colleagues for instance, used a combination of hyper-spectral data and light Detection and Ranging (LiDAR) to produce detailed maps of vegetation types in tropical forest ecosystems (Asner & Martin, 2008), or retrieve plant chemical properties and other functional traits (Asner et al., 2014a,b). Using a similar approach, Fricker et al. (2015) were also able to predict tree species richness for a small area in Panama. While these studies show major scientific and technical advances, they use technologies whose cost currently prevent their use beyond the local scale. Remote sensing products, techniques, and potential vary along with the spatial scale, so that small areas may benefit from detailed and highly accurate measurements (Asner et al., 2014b), while globally available datasets provide different variables and at a resolution reduced in nearly all dimensions (spatial, spectral and thematic). The potential to map similar ecologically relevant variables over large areas therefore remains to be explored (Pereira et al., 2013). Also, facing the great amount and diversity of data accumulated by various remote sensing technologies over time (Justice et al., 2002), it is a challenge to mine that archive and identify which information is relevant for the purpose of characterising forest ecosystems.

In this paper we assess the potential of moderate resolution remote sensing to predict forest attributes at a national scale. We present and assess an empirical approach to predict various forest attributes using multiple remote sensing variables. As an example, we used a forest inventory dataset from Bolivia — a highly diverse tropical country (Navarro & Maldonado, 2002) — in which different forest attributes (forest structure, floristic composition, abundance of functional groups) all having an ecological relevance, were inventoried across the country. Knowledge on floristic properties (species composition and tree diversity) is important for conservation and ecosystems functioning as species affect ecosystem processes through their traits (Finegan et al., 2015; van der Sande et al., 2015). Functional groups of palm, liana, and trees have received increasing attention over the past decade; palm because they are hyperdominant species (Ter Steege et al., 2013) and provide non-timber forest products, lianas because they constitute up to 30% of canopy leaf, compete strongly with trees, and are thought to increase worldwide in abundance (Schnitzer & Bongers, 2011), and emergent trees, because they lock up a disproportional part of forest biomass (Slik et al., 2013) and are especially sensitive to severe droughts, resulting in canopy dieback (McDowell & Allen, 2015). Finally, forest structure is important for timber availability, leaf layering and carbon gains.

We use this *in-situ* collected forest inventory dataset as training data for a Random Forest based modelling and prediction of the forest attributes, and provide a set of essential remote sensing variables for modelling forest attributes. Our methodological approach allows the scaling of *in-situ* ecological measurements for national-scale mapping activities and as such enables the production of information highly valuable to biodiversity conservation and our understanding of tropical forest functioning.

2.2 Material and methods

2.2.1 *In-situ* data

The *in-situ* dataset used for linking the remote sensing covariates to forest attributes originates from the Bolivian network of permanent sample plots (Figure 2.1). The network comprises 220 1 ha plots spread across lowland Bolivia’s forests and distributed along the main climatic and environmental gradients of the area. Stem diameter, height, species name, crown position and liana infestation were recorded for each tree with a stem diameter at breast height (DBH) > 10 cm, and derived information at plot level, such as basal area, tree density and species richness, was computed. A thorough description of the *in-situ* data used can be found in Toledo et al. (2011b) and Toledo et al. (2011a) for structure and floristic composition respectively.

We selected nine variables related to forest structure and floristic composition that are

important for forest functioning (Spies, 1998; Schnitzer & Bongers, 2002; Chave, 2008; Poorter et al., 2015), and biomass stocks (Ketterings et al., 2001). Structural variables are basal area ($m^2.ha^{-1}$), mean stem DBH (cm), and tree density ($\#.ha^{-1}$). Variables that refer to abundance of functional groups are liana infestation (mean relative infestation based on an ordinal scale ranging from 0 to 5), density of emergent trees (trees with their crown above the mean canopy level) ($\#.ha^{-1}$), as well as densities of palm and trees ($\#.ha^{-1}$). Floristic variables are represented by species richness and composition; both variables are derived from a subset of 100 species with reliable taxonomic identification that occurred at least in 5% of the plots. The use of this subset of species allowed identifying the patterns of floristic variation in lowland Bolivia (Toledo et al., 2011a). Toledo et al. (2011a, 2012) used a Detrended Correspondence Analysis (DCA) (Hill & Gauch Jr, 1980) to summarize the species composition of the plots into two continuous variables (or DCA axes). We refer to the two DCA axes retained as species composition first and second axes. See Toledo et al. (2011a) for interpretation of the DCA ordination scores reported in the present study. Although we expect richness derived from this restricted list of species to reflect patterns of total species richness occurring in lowland Bolivia (Toledo et al., 2011a), the variable should be interpreted as a proxy for species richness rather than absolute species richness. We therefore rescaled the variable from absolute species counts to a relative index from 0 to 1, and refer to it as *richness100*.

2.2.2 Study area

Our *in-situ* dataset did not include highland areas, forests from the Gran Chaco or zones with low canopy cover such as savannah (Toledo, 2010), we therefore focused on areas below 500 m above sea level and with a tree cover (as measured by MODIS percentage tree cover (Hansen et al., 2003)) higher than 40%. We excluded the Gran Chaco (south of the country) from the analysis using the map of terrestrial ecoregions of the world (Olson et al., 2001). Such forests are displayed in dark grey in Figure 2.1 and cover an area of 360,000 km² (about 33% of the country).

2.2.3 Statistical modelling approach

The present study predicts various forest attributes measured at forest plots using multivariate statistical modelling applied to multiple remote sensing layers. *In-situ* data are thereafter referred to as response variables while the various remote sensing layers are referred to as predictors. One can distinguish four approaches to model biophysical variables; physically based, parametric regressions, non-parametric regressions, and hybrid approaches (Verrelst et al., 2015). Here we apply a non-parametric approach because it does not require pre-established relationships amongst variables (Rivera Caicedo et al., 2014; Verrelst et al., 2015). We chose the Random Forest approach for modelling the

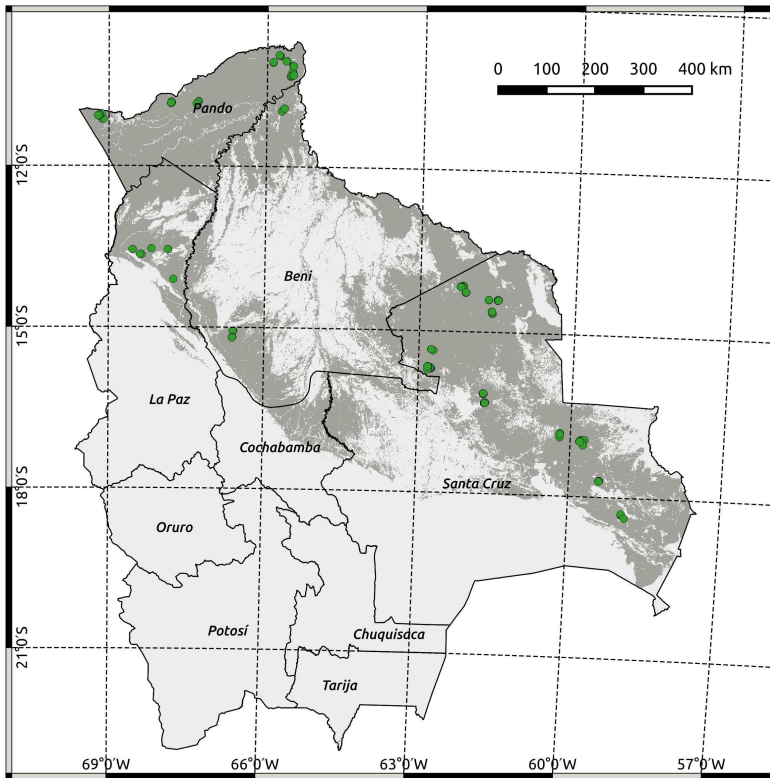


Figure 2.1: Map of Bolivia. The green markers represent the location of the sites for which *in-situ* data were available for this study. 220 forest sites were used in total but spatial clustering of sites result in overlapping markers. Areas in dark grey correspond to forest range considered for prediction of modelled parameters (MODIS tree cover greater than 40% and elevation below 500 m).

response variables based on a set of remote sensing based predictors. Random Forest is an ensemble learning method, developed by Breiman (2001) and widely used in many scientific fields, including the field of remote sensing (Pal, 2005; Avitabile et al., 2012; Laurin et al., 2014; Gessner et al., 2013). It can be used for both classification and regression problems and works by building multiple regression or classification trees from bootstrapped training samples (Breiman et al., 1984). The output of these trees are then aggregated by means of a model averaging in the regression case. This approach is called bootstrap aggregating, or bagging, and aims at increasing the model stability and accuracy (Breiman, 1996, 2001). Random Forest is a good choice for exploratory studies since it is able to cope with datasets that potentially contain many irrelevant predictors, and it performs reasonably well without the need to proceed to a variable pre-selection

(Svetnik et al., 2004). The bootstrapping approach used by Random Forest allows for the production of Out Of Bag (OOB) estimates of error which can be used for performance assessment of the model (Breiman, 2001; Liaw & Wiener, 2002). An additional output of Random Forest is the variable importance, which is computed by permuting the observations for a given predictor and measuring the effect of this permutation on the response variable prediction (Breiman, 2001; Svetnik et al., 2004). Such variable importance can later be used for selecting subsets of important predictors, and understanding of relations between response variables and predictors.

If a given model has a sufficient predictive power, as measured by the performance metrics, we can ultimately use it for predicting the response variable in a spatially continuous way, hence producing maps of the forest attribute of interest.

In addition to investigating the absolute potential to predict certain forest attributes, we aim in this paper to provide a parsimonious approach for such prediction approach applied to questions in forest ecology, management and conservation. While some of the remote sensing layers used in this study can be easily obtained and require little to no pre-processing (e.g., tree height), other predictors can be time and resource consuming to download and prepare. We therefore proceeded to a selection of the most important predictors for such modelling exercise, so that preparatory work can be minimized to the essential inputs for future studies. To select the variables, we used the variable importance returned by Random Forest as starting point and iteratively reduced the number of predictors in increasing variable importance order, re-computing the Mean Squared Error (MSE) for each new reduced model (Svetnik et al., 2004). The rationale of the approach is that removing irrelevant predictors will have little to no effect on the model prediction power while an increase in MSE should be observed once the important predictors start being removed. We therefore applied this procedure for a set of selected response variables, identified points of increase in MSE and derived reduced models for these response variables.

2.2.4 Remote Sensing predictors

We selected a set of layers to use as predictors for modelling the forest variables described above. Although the approach is empirical, we expect the selected set of predictors to either directly relate to some response variables (e.g. tree cover and tree density), or to indirectly have an influence on them (terrain metrics with palm density). As such we selected a set of fluxes, structure related variables and topography metrics as the full set of remote sensing predictors (Table 2.1). Because we aim the approach to be repeatable for different locations in the tropics, we worked only with globally available or continental remote sensing products. Each of the products used is presented below, together with the required pre-processing steps, and the rationales for including it as a predictor are briefly

justified. For consistency across variables, all remote sensing layers were re-sampled to 1 km resolution. While most of the predictors are available for multiple dates (MODIS time-series are 15 years long for instance), we selected 2005 as the year of reference for all remote sensing products because the tree height dataset was only available for that year, and because the *in-situ* measurements occurred between 2002 and 2007 (Toledo, 2010). We do not expect small difference in timing between the *in-situ* measurements and the remote sensing observations to have a significant influence on the resulting predictions; the coarse nature of the predictors likely reflect the general state of the ecosystems rather than their condition at a given time.

MODIS products

A large part of the remote sensing predictors used in the present study are derived from the two MODIS sensors (Terra and Aqua), which collect daily optical data globally about land, oceans and the atmosphere (Justice et al., 2002). Data is collected, depending on the spectral characteristics at a resolution of 250 to 500 m. The moderate spatial resolution, coupled with daily revisit capabilities makes MODIS particularly suited to global scale analysis. Repeated measurement can be assembled in time-series for mapping dynamics while the spatial resolution makes data volumes relatively easy to handle without requiring supercomputing facilities. MODIS has a wide array of applications (Justice et al., 2002), and has been used in many global studies, from modelling biophysical variables (He et al., 2012) to vegetation trends and dynamics studies (Morton et al., 2005; Dutrieux et al., 2012). While the original raw data collected by the MODIS instruments is radiance, a number of higher level products are derived from these data and distributed by the United States Geological Survey (USGS) for public use. In the present study, we are using both the surface reflectance, corrected for Bidirectional Reflectance Distribution Function (BRDF) effects, and a set of higher level MODIS products. MODIS products are generally delivered with quality control information, meaning that every pixel contains not only the values of the variable of interest but also information on the reliability of such value. Data were systematically filtered using this quality control information, retaining good quality data only. This section describes the standard MODIS products produced on an operational basis by the USGS and directly available via their Distributed Active Archive Center (<https://lpdaac.usgs.gov/>).

Surface Reflectance When light illuminates vegetation, it is partitioned into three components; part of the light is absorbed and used to power photosynthesis, another part is transmitted, and the remaining part is reflected (Anderson, 1966). The reflected part of the light partitioning is what optical sensors on-board satellites are measuring. Such light partitioning, which varies along with the wavelengths, depends on the chemistry and the structure of the vegetation (Knipling, 1970; Curran, 1989). Hence differences in chemistry

and structure, which are likely to occur among different tree species associations and forest types, should result in different spectral signatures (Asner et al., 2014b). Surface reflectance measured by MODIS is delivered as a remote sensing product by the USGS in the form of 16 days composites, meaning that 22 measurements of surface reflectance are available for every location throughout the year (Strahler et al., 1999). In order to limit the amount of data usage, we used for each pixel independently, two observations of surface reflectance for each of the seven spectral bands. We used the two extremes of the season (peak of the growing season and lowest point of the season) derived from the HANTS phenology dataset (method described later in section 2.2.4) as reference dates for selecting the two observations of surface reflectance considered. In order to increase the stability of the reflectance values and reduce the probability of cloud cover contamination, we built trimestrial average value composites centered around these two dates for each spectral band. Given that seven spectral bands are available and that two predictors corresponding to the surface reflectance during high and low season are derived for each band, 14 predictors related to surface reflectance are derived from this dataset (Table 2.1). Predictors are named SRn_{min} , SRn_{max} , where *min* and *max* correspond to the low and high seasons (peak and lowest point in photosynthetic activity annual profile) respectively and *n* is the spectral band index.

Net Primary Production (NPP) ($KgC.m^{-2}$) Satellite based annual estimates of NPP are an indicator of forest productivity and carbon assimilation. The USGS processes and delivers annual NPP values following the algorithm described in Running et al. (2004).

Leaf Area Index (LAI) (m_{plant}^2/m_{ground}^2) LAI plays an important role in ecosystem processes such as transpiration, light and rainfall interception or photosynthesis (Maass et al., 1995); it is therefore an important structural characteristic of forest canopies. LAI is provided as a high level MODIS product for every eight day period of the year (Knyazikhin et al., 1999), resulting in 45 yearly LAI estimations. We reduced this LAI time-series to one value by calculating the median LAI value of the year.

Percentage tree cover Percentage tree cover, or Vegetation Continuous Fields (VCS), is the percentage of closed canopy projected on a horizontal plane (Hansen et al., 2003). We therefore expect this predictor to relate to structural attributes such as tree density. Apart from being a predictor, we used the percentage tree cover to constrain the spatial prediction range to forest areas only (Figure 2.1).

Tree height

A lot of attention has been given to satellite-based tree height recently, particularly for the task of forest above ground biomass estimations (Saatchi et al., 2011). Beside being an essential variable in biomass estimations (Ketterings et al., 2001), height plays an important functional role at the individual tree level (Poorter et al., 2006). Here we use an average canopy height dataset, available globally at 1 km resolution, generated by combining measurements from the Geoscience Laser Altimeter System (GLAS) with environmental modelling (Simard et al., 2011). Unlike MODIS derived products, tree height, which requires LiDAR technology to be estimated, is only available for one point in time (2005).

Canopy clumping

Canopy clumping is a direct measure of canopy structure and heterogeneity that quantifies the level of foliage grouping. By taking advantage of the effect of canopy structure on the BRDF of light estimated from multiple MODIS observations, He et al. (2012) were able to retrieve a canopy clumping index at 500 m resolution globally.

Harmonic Analysis of Time-Series (HANTS) phenology

This dataset generated by fitting a Fourier series on a Normalized Difference Vegetation Index (NDVI) MODIS time-series reflects changes and cycles in vegetation photosynthetic activity (Roerink et al., 2003). Here we applied an improved version of the HANTS algorithm that uses inter-annual NDVI composites to generate a reference dataset hence making the method suitable to areas with high cloud coverage. The relevance of phenology and seasonality has been demonstrated in many studies and deciduousness is commonly accepted as a major functional trait of forest ecosystems (Chambers et al., 2007). Ter Steege et al. (2006) also found a concurrence between dry season length and variation in genus composition, suggesting that seasonal dynamics may contribute to the mapping of forest floristic attributes. Rather than one value, several variables, characterizing the intra-annual dynamics of vegetation productivity, can be extracted from the HANTS phenology dataset. Here we use the seasonal amplitude in NDVI (displayed in Figure 2.2 for illustration), the duration of the vegetation production season, the minimum and maximum NDVI values as well as the yearly average NDVI as predictors.

Terrain metrics

Topography may be an important factor to consider when attempting to predict vegetation structure and floristic composition. It is well known for instance that altitude, slope and

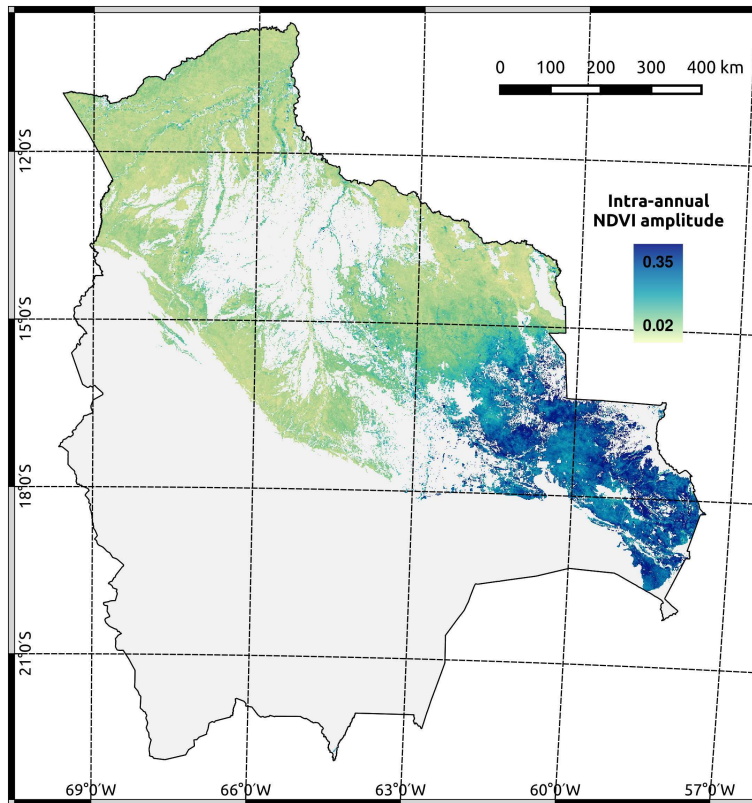


Figure 2.2: Map of NDVI intra-annual amplitude (HANTS phenology dataset) covering the forested areas of lowland Bolivia.

aspect affect micro-climatic conditions such as temperature, amount of incoming radiation or water retention capacity of an area. These micro-conditions may in turn affect the life forms and species composition of a forest and the development of established individuals (Emilio et al., 2014). We computed, using the global elevation dataset from the Shuttle Radar Topographic Mission (SRTM) (Farr et al., 2007), slope, aspect (Horn, 1981), and Terrain Ruggedness Index (TRI) (Wilson et al., 2007), for every 1 km² pixel covering Bolivia.

2.2.5 Variables extraction

To train the Random Forest model, we first need to bring predictors and response variables together. For this data extraction step, we spatially overlaid predictors and response variables and extracted the predictor values corresponding to the pixels overlapping with

Table 2.1: Detail of the eight remote sensing variables used in the present study. Each row corresponds to a variable from which we derived at least one predictor. In addition to the specific pre-processing steps described in the table, all predictors were re-projected to a common projection and re-sampled to 1000 m spatial resolution. NPP and LAI correspond to Net Primary Production and Leaf Area Index respectively.

Variable	Origin	Product name	Original Spatial Resolution (m)	Specific preprocessing	Number of predictors	References
Surface Reflectance	MODIS	MCD43A4	500	average value composite centered around peak and min of season	14	Strahler et al. (1999)
NPP	MODIS	MOD17A3	1000	-	1	Running et al. (2004)
LAI	MODIS	MCD15A2	1000	Compute annual median	1	Knyazikhin et al. (1999)
Tree cover	MODIS	MOD44B	250	-	1	Hansen et al. (2003)
Tree height	GLAS MODIS	-	1000	-	1	Simard et al. (2011)
Canopy clumping	MODIS	-	1000	-	1	He et al. (2012)
Phenology metrics	MODIS	HANTS phenology	250	-	5	Roerink et al. (2003)
Terrain metrics	SRTM	SRTM DEM	90	-	4	Farr et al. (2007); Horn (1981); Wilson et al. (2007)

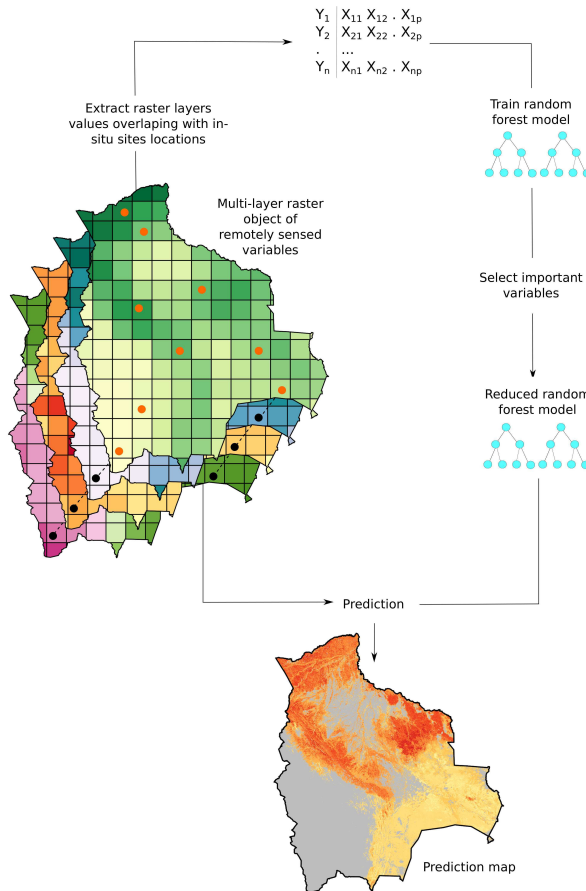


Figure 2.3: Schematic overview of the proposed methodology applied to each response variable.

the *in-situ* site locations (Figure 2.3). When several sites would fall within the same 1 km pixel, we averaged the values of the response variable. This necessity to spatially aggregate neighbouring forest plots reduced the amount of training sites from 220 to 104. Connecting a 1 km² (100 ha) surface to a 1 ha plot requires that an assumption be made about the representativeness of the plot with respect to its surroundings. Such assumption would most certainly be violated in a heterogeneous landscape (Woodcock & Strahler, 1987), in which case alternative approaches using landscape stratification could be used (Baccini et al., 2007). However, in the present case, the way plots were established ensures that they are at least surrounded by a forest matrix.

2.2.6 Measures of accuracy

To evaluate the performance of the Random Forest model, we used a set of measures of model performance. All accuracies are calculated from the estimates of error derived from the Out Of the Bag (OOB) predictions (Breiman, 2001; Liaw & Wiener, 2002). The OOB predictions correspond to the predicted values of the response variables; error rates can therefore be computed by subtracting observed value of the response variable to these OOB predictions.

Following that, the different metrics used can be calculated as follows:

$$RMSE = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n}} \quad (2.1)$$

where \hat{y}_i and y_i are the predicted (OOB) and observed values respectively for the i^{th} sample of the response variable, and n represents the total number of observations.

The Root-Mean-Square Error ($RMSE$) provides an estimate of the prediction error, expressed in the unit of the variable. While it is a reliable metric to compare the predictive power of different models applied to the same response variable, its use for assessing a single model predictive performance is limited. We therefore also computed the normalized version of the $RMSE$, which is defined as:

$$NRMSE = \frac{RMSE}{\sigma_y} \quad (2.2)$$

where σ_y is the standard deviation of the observed variable.

The Normalized $RMSE$ ($NRMSE$) is unit-less and can be interpreted as the proportion of residual variance, which is the variance not explained by the model when predicting new values.

Additionally, we computed Modelling Efficiency (EF) (Mayer & Butler, 1993; Loague & Green, 1991), which is a dimensionless metric that varies between -infinite and 1 and expresses the proportion of variance explained by the 1:1 line. When varying in the 0–1 range, EF can be interpreted similarly to the coefficient of determination (R^2), meaning that higher values reflect better model performance, while negative values suggest very poor prediction performance.

EF is calculated as follow:

$$EF = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.3)$$

Finally, to assess whether systematic patterns of over or under estimation exist in the model predictions we computed for each *in-situ* variable a *Bias*, which is computed as follows:

$$Bias = \hat{y} - \bar{y} \quad (2.4)$$

where \hat{y} and \bar{y} correspond to predicted and observed mean respectively.

For the variable selection step (Svetnik et al., 2004), Mean Squared Error (MSE), calculated using five fold cross validation instead of OOB estimates of error, is used. Svetnik et al. (2004) explains with further details why using OOB estimates of error would be biased for this step specifically. MSE can be calculated using the following equation:

$$MSE = \frac{1}{n} \sum (\hat{y}_i - y_i)^2 \quad (2.5)$$

2.3 Results

2.3.1 Model performance

In order to assess the maximum achievable performances, we evaluated for each of the nine response variables the modelling potential by running a Random Forest model using all 28 predictors. Because these measures are generated with an unrestricted number of predictors we can refer to them as absolute modelling potential. Figure 2.4 presents the performance of the predictions for all response variables. A first observation of the modelling results reveals three levels of performance, associated with variables related to floristic composition, functional groups and structural variables. Floristic variables show the highest potential (with EF values ranging from 0.65 to 0.97), followed by functional groups (EF between 0.16 and 0.66) and structural variables (EF between 0.14 and 0.37). Modelling potential appears to be particularly high for floristic variables suggesting that different species associations result in differences in canopy structure, chemistry and phenology, making them distinguishable using our approach. We selected response variables with EF higher than 0.5 (Table 2.2) for interpretation, selection of essential predictors and further analysis.

2.3.2 Selection of essential predictors for floristic variables and palm density prediction

To identify a set of essential remote sensing variables for predicting the floristic properties and palm density, we carried out a model reduction. An example of the procedure used

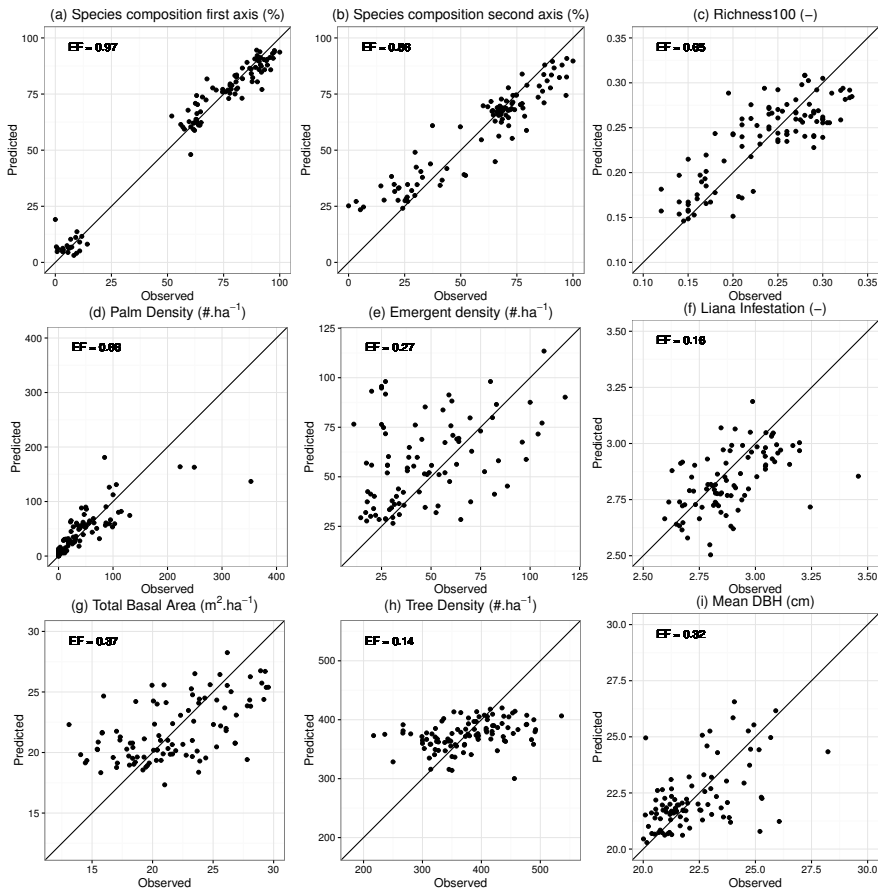


Figure 2.4: Scatter plots of prediction performances for the species composition axes 1 (a) and 2 (b), richness100 (c), palm density (d), emergent density (e), liana infestation (f), basal area (g), tree density (h) and mean DBH (i). Richness100 is a proxy of species richness based on 100 species with reliable taxonomic identification that occurred in at least 5% of the plots. Each dot corresponds to an observed *in-situ* value for which the observation is confronted to the random forest prediction of the corresponding pixel. The diagonals are the 1:1 lines, so that wider spread around these lines indicates lower prediction performances. The top panels refer to floristic variables, the middle panels to functional composition, and the bottom panels to forest structure. All variables apply to trees with a stem diameter at breast height > 10 cm.

is illustrated in Figure 2.5, where the case of increase in MSE with decreasing number of predictors used for richness100 modelling is presented. Starting from a full model (28 predictors, right side of Figure 2.5) the MSE is computed for each reduced model, discarding variable in increasing importance order. In this particular example, a large increase in MSE is observed when reducing from seven to six predictors, resulting in an

Table 2.2: Summary of prediction performances (RMSE, NRMSE, EF, and *Bias*) for all response variables modelled using the Random Forest algorithm and using all remote sensing predictors available. Rows with asterisks (*) are response variables retained for further analysis and prediction.

Variable	RMSE	NRMSE	EF	<i>Bias</i>
Species composition first axis *	5.38	0.17	0.97	0.02
Species composition second axis *	9.33	0.37	0.86	-0.05
Richness100 *	0.04	0.59	0.65	0.00
Palm density *	31.11	0.58	0.66	0.76
Emergent density	37.62	0.85	0.27	1.78
Liana infestation	0.26	0.91	0.16	0.00
Total basal area	3.68	0.79	0.37	0.04
Tree density	62.24	0.92	0.14	-0.72
Mean DBH	1.81	0.82	0.32	0.03

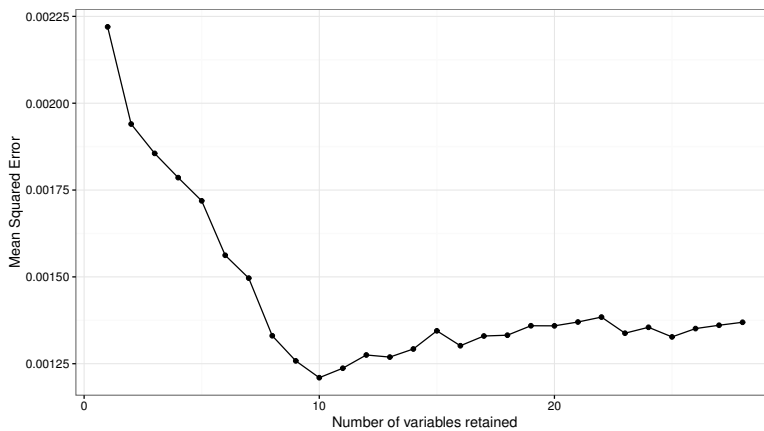


Figure 2.5: Example of decrease in MSE when removing variables ordered by increasing importance. In total 28 variables are considered. This particular example illustrates the case of the richness100 variable, and an inflexion point can clearly be seen at 7 variables.

optimal reduced model containing the seven most important predictors.

The most important predictors for the four response variables retained are presented in Table 2.3. Surface reflectance and phenological metrics account together for 10 out of the 15 essential predictors. HANTS NDVI amplitude and NPP are the two only predictors contributing to the modelling of three out of the four response variables retained; making them key predictors for forest inventory modelling. Two terrain metrics — TRI and slope — are important for modelling palm density. Despite a systematic reduction of more than half of the predictors, the modelling efficiencies (EF) were maintained, indicating

that little information was lost during the model reduction and that a limited number of predictors are relevant.

Table 2.3: Summary table of most important remote sensing variables ranked by importance when modelling species composition (first two axes), richness100, and palm density.

Variable	Species composi- tion first axis	Species composi- tion second axis	Richness100	Palm density
$SR1_{min}$	-	1	-	-
$SR2_{min}$	3	-	7	-
$SR2_{max}$	-	3	-	-
$SR4_{max}$	-	4	-	-
$SR5_{min}$	5	-	6	-
$SR7_{min}$	8	-	-	-
NPP	-	2	5	1
LAI	7	-	-	-
Tree height	6	-	-	2
HANTS $NDVI_{amplitude}$	2	-	1	4
HANTS $NDVI_{mean}$	4	-	2	-
HANTS $NDVI_{min}$	1	-	4	-
HANTS $NDVI_{max}$	-	-	3	-
TRI	-	-	-	3
slope	-	-	-	5
Reduced model efficiency (EF)	0.64	0.97	0.83	0.66

2.3.3 Spatial prediction of forest floristic attributes and palm density

We used the reduced models for predicting the four response variables for every forested 1 km² pixel covering lowland Bolivia (Figure 2.6). Some of the patterns described by Toledo et al. (2011a) can be seen in the resulting prediction maps. For instance the first axis of species composition (Figure 2.6a) presents a strong north–south gradient in floristic composition, which co-varies with the well known south to north rainfall gradient present in the lowland part of the country. The second axis (Figure 2.6b) shows a stronger east–west gradient in floristic composition which corresponds to the east to west gradient in soil fertility. Although richness100 (Figure 2.6c) presents some coherent patterns (higher diversity in the North and East of the country than in the South), it also misses some well known diversity hotspots near the foothill of the Andes.

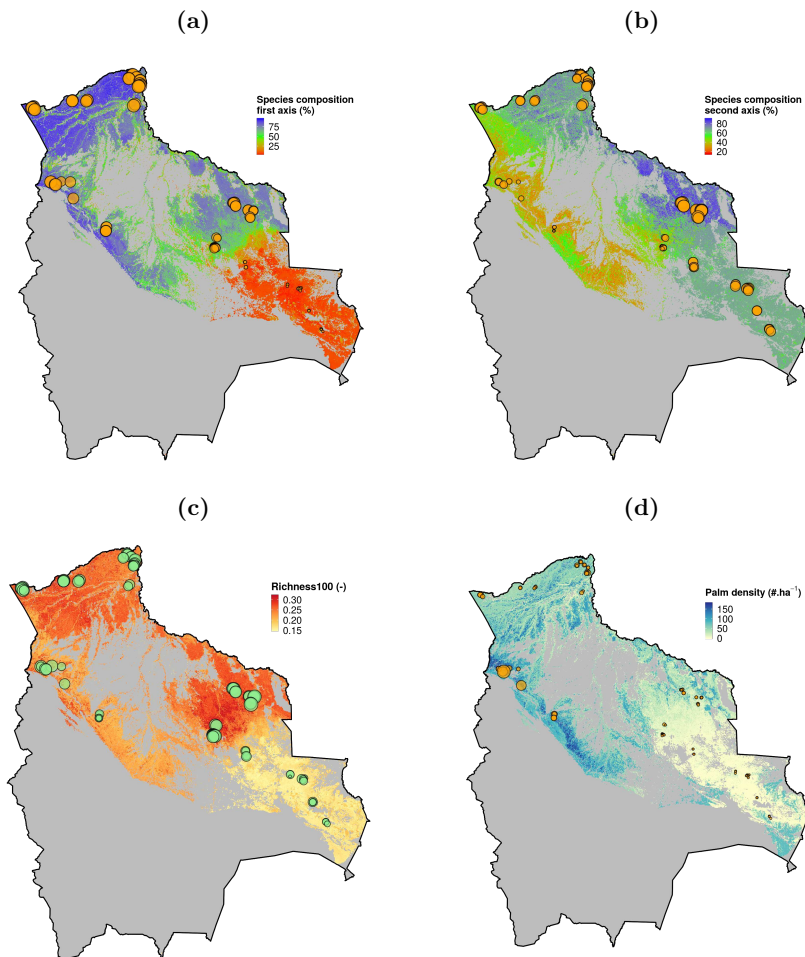


Figure 2.6: Spatial predictions over lowland Bolivia of species composition first and second axes (a–b), richness100 (c), and palm density (d). Predictions were made using the reduced Random Forest models presented in section 2.3.2 for areas with a forest cover > 40% as defined by the forest mask presented in Figure 2.1, so that areas in light grey are considered non-forest. Overlaid scaled markers correspond to observed values, measured from *in-situ* sites.

2.4 Discussion

2.4.1 Moderate resolution remote sensing can predict floristic properties of tropical forests

By combining data from a network of permanent sample plots, multiple remote sensing predictors, and a Random Forest modelling approach we predicted forest attributes across

tropical forests across lowland Bolivia. We found great potential to predict forest floristic properties and palm density, and a more limited potential to predict forest structure. In the case of floristic composition variables and palm density, for which model performance metrics suggest a strong potential for remote sensing based modelling, the relations between predictors and response variables are more direct. For instance, different species associations will likely result in different spectral signatures (Féret & Asner, 2013), and intra-annual phenological patterns (Somers & Asner, 2014), which are two traits partially measured by the remote sensing predictors via surface reflectance and phenology respectively. Species composition could be better modelled than richness100 because of its higher specificity to the local floristic characteristics. Different forest types with non-overlapping species compositions and potentially different structure and functional composition can potentially achieve similar levels of species richness. Given this non-specificity, a statistical modelling approach using the predictors related to canopy structure and chemistry will not be able to capture the entire range of variability in species richness. Additionally, expert opinions on the maps produced suggest that while the North–South gradient pattern of species richness is well captured by our prediction, the high species diversity of forests near the Andes does not appear in the map. The map likely has a high uncertainty in this region which may be explained by a relatively low representation of this particular forest type in the training dataset. Nevertheless, even if richness100 did not achieve the levels of performance observed for species composition ($EF = 0.85\text{--}0.98$), we could surprisingly well model it ($EF = 0.65$). This capacity to produce in a consistent and objective way spatially continuous information on tropical forests floristic properties holds great promises for biodiversity conservation initiatives, and to quantify change in biodiversity over time.

There was limited potential to predict structural properties which can most likely be explained by the lack of remote sensing predictors that are directly comparable to the response variables of interest. With the exception of tree height and terrain, all predictors originate from optical remote sensing and are therefore mainly influenced by reflected light. In these closed forests, the structural components are therefore inferred from the chemistry of the canopy trees alone, which is not a good reflection of the underlying forest biomass or structure (Broadbent et al., 2008b). This disconnection between response variables and predictors, is a limitation for forest structure modelling in general, and also partly explains the large uncertainties that exists in all global biomass mapping efforts attempted to date (Ometto et al., 2015). Unfortunately, with the set of optical remote sensing products currently available at global scale, it is very unlikely that we will be able to overcome this limitation.

2.4.2 Phenology and surface reflectance are essential predictors of floristic properties

Random Forest allowed us to obtain a restricted number of essential remote sensing predictors, based on their contribution in predicting the response variables retained. This is of great relevance as reducing the number of predictors leads to more parsimonious approaches, facilitating the data preparation efforts and the repeatability of such studies for different areas in the tropics. The variable importance also provides a perspective on the mechanistic relations influencing the model. We identified two important groups of predictors from the variable selection process: phenological metrics and surface reflectance (Table 2.3). Both deciduousness and leaf chemistry have been identified as key functional traits in previous studies (Ter Steege et al., 2006; Féret & Asner, 2013; Somers & Asner, 2014), and are important for species identification using remote sensing. Optical sensing offers the capacity to measure both leaf structure and chemistry (via surface reflectance) and phenology, and is therefore well suited for species associations mapping (Lang et al., 2015), as demonstrated in the present study. This importance of surface reflectance also suggests that the use of higher spectral resolution data, which will be more widely available in the future (Kaufmann et al., 2008) could potentially contribute to an even better distinction between species associations. In addition to surface reflectance and phenology, we found that NPP, tree height and LAI also play an important role in the modelling of floristic characteristics of the forest.

Although Bolivia covers a limited spatial extent, it has a large variety of environments, primarily due to its location between the Andes and the Amazon basin where several biomes converge (Amazonian, cloud forest, dry forest, chaco and cerrado). This is an important aspect to consider when interpreting our findings. On the one hand, this variety ensures that the findings are applicable to a wide range of environments, possibly including areas beyond the administrative borders of Bolivia. On the other hand, while we expect the general patterns of predictors importance to be consistent regardless of the tropical region of interest, it is likely that the relative importance of each predictor will vary. We would for instance expect phenology to be less determinant in discriminating floristic properties within homogeneous wet evergreen forests of the Amazonian basin. We therefore recommend, as a way to fully ensure the transferability of the method and its potential, to repeat the present modelling exercise with different *in-situ* datasets covering other regions in the tropics.

2.4.3 Current contribution to ecological questions and resource management

Our work demonstrated the potential of moderate resolution remote sensing to provide spatially continuous information on forest floristic properties. While the focus of the paper

is mostly on the method and its validation rather than on the use of that information, using the full potential of these maps is a logical next step. The potential contribution of these maps is twofold. First the information produced can be used to enhance our mechanistic understanding of processes driving these natural systems. As shown by Watt in 1947, studying spatial patterns of plant communities is an essential component to understanding the processes that drive these communities (Watt, 1947). While it has always been possible to put *in-situ* measurements in a spatial context, the lack of spatial continuum in the data makes it difficult to infer patterns. Using 1 km² spatial resolution maps produced by our method, spatial patterns can easily be identified as well as co-occurrence between environmental variables and floristic properties. Such insight can greatly enhance our understanding of the processes that drive the organization of tropical forests. Additionally, as vegetation dynamics modelling increasingly contributes to the understanding of functional mechanisms that drive forests (Sakschewski et al., 2015), it would be of great interest to refine existing models using floristic data produced by our method as input. Besides, an enhanced understanding of the ecology of these systems also benefits the applied issues listed above as sound management is only possible when the systems considered are well understood. Secondly, the maps are valuable tools for applied issues such as conservation, forest management or international initiatives such as Reducing Emissions from Deforestation and Forest Degradation (REDD+) and the Convention on Biological Diversity (CBD) (Skidmore et al., 2015; Sexton et al., 2015). By providing relevant variables in a spatially explicit, objective and consistent way, they offer a strong basis to delineate areas of interest for conservation, or define region specific management strategies or policies. As an example, such maps, if used jointly with carbon maps, could be used to address biodiversity co-benefits in a REDD+ context.

The uncertainties of our approach also carry important messages for resources management. While the local mismatch between the well known patterns of high species diversity near the foothills of the Andes and the predicted diversity for this region highlights a limitation of our method, it also stresses the importance of well designed national forest inventories. Such method will never be able to accurately predict variables that are beyond the range of the training data; it is therefore crucial, to ensure remote sensing based scalability, that national forest inventories are designed in such a way that all forest types are equally represented and that variables are comparable across sites.

2.4.4 Future prospects for mapping of biodiversity and vegetation structure

The present study demonstrates the potential and limitations of remote sensing to predict nine forest attributes at national-scale. The results suggest a great potential, particularly for mapping variables relating to floristic composition and biodiversity, but also found limitations when attempting to model variables related to forest structure. We have al-

ready discussed that a strong disconnection between the remote sensing predictors — in most cases derived from optical remote sensing — and the response variables, could explain these limited performances. These limitations of optical remote sensing to predict forest structure will always exist and the use of alternative technologies such as radar and LiDAR, that are able to penetrate the upper layers of the canopy and provide information on the structure of the woody parts of the vegetation is a promising way forward (Arcioni et al., 2014). However these technologies, even if they already exist, and were proven to be effective in mapping forest structure (Asner et al., 2012), remain at the moment unavailable at global scale or prohibitively expensive for a realistic global deployment (Mascaro et al., 2014). The future is certainly very promising in this regard, with rapid technological progress and upcoming space-born LiDAR and radar missions targeted at ecosystem structure and dynamics mapping (Le Toan et al., 2011; Dubayah et al., 2014).

Additionally, one alternative approach called regression krigging (Hengl et al., 2004), which would certainly improve the predictions of all response variables, can be mentioned. By taking advantage of spatial autocorrelation contained in the response variables (Sales et al., 2007), regression krigging can potentially compensate for the lack of relationship between the response variables and the predictors. However, such approach can only be considered for areas with a high spatial density of *in-situ* sites — which is often a limiting parameter in remote sensing studies applied to the tropics — and could therefore not be used in the present case.

We explored and assessed in the present case study the potential of the approach for a single point in time, however, as we get further insights on the reliability of such approach by testing it in other areas, we will be able to include a temporal dimension in such studies as well, hence allowing to monitor long term trends in vegetation characteristics, biodiversity and response to climate change.

2.5 Conclusion

We predicted a set of forest attributes from a large number of remote sensing predictors using an empirical statistical modelling approach. Our approach revealed a great potential to map forests floristic properties; we were able to produce high accuracy maps for proxies of species richness and composition, and palm density at 1 km² covering lowland Bolivia. The results also revealed limitations of remote sensing to capture variables related to forest structure. These limitations originate from the nature of remotely sensed data we used and are unlikely to be overcome with technologies currently available and suitable for global mapping. While we used 28 predictors in total, the list of contributive predictors could be reduced to 15, with phenology and surface reflectance playing a particularly important role. The information generated by our approach is spatially continuous and provides an objective and consistent stream of data that can enhance our understanding

of forest functioning, contribute to better zoning and monitoring of resources and benefit biodiversity conservation. As stability of the method gets confirmed, we will be able to move from simple mapping of these variables to monitoring them which is crucial for biodiversity conservation.

Acknowledgements

The research received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement Number 283093 - The Role Of Biodiversity In climate change mitigation (ROBIN). The permanent plots data provided by M. Toledo was funded by the Netherlands Foundation for the Advancement of Tropical Research (WOTRO), the Education For Nature - Russell Train fellowship (EFN/WWF), the International Foundation for Science (IFS) and Wageningen University and Research Centre (WUR).

Chapter 3

Monitoring forest cover loss using multiple data streams

This chapter is based on:

Dutrieux, L. P., Verbesselt, J., Kooistra, L., & Herold, M. (2015). Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing*, *107*, 112-125.

Abstract

Automatically detecting forest disturbances as they occur can be extremely challenging for certain types of environments, particularly those presenting strong natural variations. Here, we use a generic structural break detection framework (BFAST) to improve the monitoring of forest cover loss by combining multiple data streams. Forest change monitoring is performed using Landsat data in combination with MODIS or rainfall data to further improve the modelling and monitoring. We tested the use of the Normalized Difference Vegetation Index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) with varying spatial aggregation window sizes as well as a rainfall derived index as external regressors. The method was evaluated on a dry tropical forest area in lowland Bolivia where forest cover loss is known to occur, and we validated the results against a set of ground truth samples manually interpreted using the TimeSync environment. We found that the addition of an external regressor allows to take advantage of the difference in spatial extent between human induced and naturally induced variations and only detect the processes of interest. Of all configurations, we found the 13 by 13 km MODIS NDVI window to be the most successful, with an overall accuracy of 87%. Compared with a single pixel approach, the proposed method produced better time-series model fits resulting in increases of overall accuracy (from 82% to 87%), and decrease in omission and commission errors (from 33% to 24% and from 3% to 0% respectively). The presented approach seems particularly relevant for areas with high inter-annual natural variability, such as forests regularly experiencing exceptional drought events.

3.1 Introduction

Getting spatially and timely accurate information about land cover change is essential as change in forest cover is likely to affect aspects of the biosphere such as carbon cycle and biodiversity (Malhi et al., 2008). Remote sensing, especially in recent years, with large amount of data becoming freely available offers great opportunities to monitor forest at relatively high spatial resolution and in a systematic and objective way. However, there is a challenge that consists in extracting the desired information from large amounts of spatio-temporal data containing natural variability (seasonality and exceptional events such as droughts) and noise. The scientific community has reacted to this growing need and a variety of methods and products has appeared in the last few years, with a trend moving from bi/multi-temporal to hyper-temporal approaches (Lu et al., 2014). This trend can be explained by the opening of the Landsat archive, which allowed researchers to take full advantage of all the Landsat data that have been acquired since the beginning of the program in the 70s. In addition to being widely available at no cost, Landsat data, thanks to their spatial (30 m resolution) and temporal characteristics (16 days revisit period) are well suited to monitor processes such as deforestation and forest degradation (Asner et al., 2004a,b; de Wasseige & Defourny, 2004; Souza et al., 2005). Hence, by mining the entire Landsat archive for a given area, Zhu et al. (2012) were able to detect forest disturbances in temperate forests of the western United States. In a following paper Zhu & Woodcock (2014) extended their method to include land cover mapping following disturbance detection. Using super geo-computing facilities, Hansen et al. (2013) were able to globally map annual forest cover loss. From a methodological point of view, the existing multi-temporal methods differ slightly, depending on their intended use. For instance, in order to investigate land trajectories, Kennedy et al. (2010) developed LandTrendr, a temporal segmentation method capable of reconstructing recent land use history. LandTrendr works by identifying, using a time-series of annual image composites, segments of stable land use trajectories (stable forest, agriculture, regrowing forest, etc). For producing annual estimates of forest cover loss, Hansen et al. (2013) also worked on a year by year basis, computing differences between forest cover percentages produced annually from regression trees calibrated with Very High Resolution (VHR) images, such as Quickbird imagery. However, aiming at producing forest cover loss information with precise timing of disturbances requires approaches that work directly on the entire time-series. Working on the entire time-series, which means using all data available has two main advantages. First it allows to capture not only inter-annual differences, but also intra-annual variation. For instance vegetation, particularly in temperate regions is well known for having an intra-annual seasonal signal, which we call phenology. A time-series approach that uses full time-series can account for the phenology and identify changes in the phenological patterns (Verbesselt et al., 2010a). Second, using all data allows a higher precision in the timing of the events being detected. An example of method which

uses the full time-series is the Continuous Monitoring of Forest Disturbances Algorithm (CMFDA), proposed by Zhu et al. (2012), which uses Fourier series and a threshold based deviation from the model mechanism to detect abnormal behaviours in newly acquired images. CMFDA requires three consecutive observations exceeding the threshold for a change event to be confirmed, which in an optimal setting with two Landsat satellites (8 days revisit period) can be as short as 16 days (Zhu et al., 2012). Similarly, approaches proposed by Brooks et al. (2014), Verbesselt et al. (2012), and DeVries et al. (2015b) use all pixels available within the time-series and identify change based on statistical hypotheses of stability and confidence intervals. An intermediate approach, using all data from cloud free Landsat scenes, rather than all pixels, is proposed by Huang et al. (2010) for their Vegetation Change Tracker, which uses spectral distances to a set of reference pixels to derive forest likelihood. All these methods have been successful in detecting on a pixel basis events of forest cover loss soon after they occur in temperate forests at Landsat scales (Zhu et al., 2012; Brooks et al., 2014; DeVries et al., 2015b) and forest disturbances and droughts from Moderate resolution time-series (Verbesselt et al., 2010a,b, 2012). However, an additional challenge, which has not been much investigated yet, consists in detecting in a similar way forest cover loss in highly variable forest environments. Seasonality in some environments may vary in amplitude from year to year, making the use of traditional seasonality models inappropriate since they cannot model differences among years. The challenge therefore consists in discriminating what is natural from what is anthropogenic in a temporal signal.

We identified two mechanisms potentially capable of discriminating human induced from anthropogenic variations. The first consists in taking advantage of the spatial context, since natural variations occur at regional scales, while human induced change (excluding climate change) tend to be much more localized. The second mechanism is to use an additional independent variable capable of predicting the observed variable. When interested in vegetation, or more specifically the Normalized Difference Vegetation Index (NDVI), time-series, the two above mentioned strategies can be translated as (1) including an extra NDVI time series at a different spatial resolution and (2) using a climatic index, since climatic conditions are known to be an important driver of vegetation greenness (Jong et al., 2013; Nemani et al., 2003). Previous studies taking advantage of contextual information or external data sources in a change detection context include Huang & Friedl (2014) and Kleynhans et al. (2011). Huang & Friedl (2014) developed a method capable of detecting sub MODIS pixel land cover change with context based calibration, while Kleynhans et al. (2011) developed a change metric at MODIS pixel level, using information from a 3×3 pixels window, hence using spatial context as a way to discriminate anthropogenic change from natural variability. Although the above mentioned studies illustrate ways of using the spatial context for change detection the level of spatial detail they provide is not sufficient for tracking small forest disturbances.

The proposed approach is an extension of the Break detection For Additive Season and

Trend (BFAST) concept. BFAST has been developed as a generic change detection framework for disturbance detection. It is fully statistically based and has been validated for forest change detection (Verbesselt et al., 2010a), phenological change (Verbesselt et al., 2010b) and optimised for near real-time detection (Verbesselt et al., 2012). The framework has recently been evaluated for forest change monitoring in the tropics on Landsat time series for continuous forest change tracking in Ethiopia (DeVries et al., 2015b), and for disturbance monitoring using fused radar-Landsat time-series (Reiche et al., 2015). In this study we further test and improve the BFAST framework by including extra data sources to improve the discrimination of anthropogenic changes from natural variability. Monitoring is performed using Landsat data. We tested two candidate variables to be used as external regressors; NDVI from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Standardized Precipitation Index (SPI). The proof of concept is done over a tropical dry forest of lowland Bolivia and the accuracy of the method is assessed using a set of visually interpreted Landsat time-series in combination with recent Very High Resolution imagery. Further insights on the approach performances and applicability are gained by experimenting with the different parameters combinations of the method, as well as via qualitative assessment of spatio-temporal patterns and single time-series profiles.

3.2 Material and methods

3.2.1 Change detection algorithm

In the BFAST framework, techniques for break detection (Chu et al., 1996; Leisch et al., 2000; Zeileis et al., 2005, 2010) which have been optimised for linear regression frameworks by Zeileis (2005), are combined with a seasonal-trend model for detecting change in vegetation dynamics (Verbesselt et al., 2012). The method consists in fitting a model to the data by Ordinary Least Square (OLS) fitting on a period defined as stable history, and to check for stability of that same model during a period defined as monitoring period. Discrepancy between the model predictions and the data during the monitoring period is estimated using a moving sums of residuals (MOSUM) approach (Chu et al., 1996), and when observed data significantly deviate from the model, a break is detected.

Seasonal-trend model with an external regressor

The additive seasonal trend model used by Verbesselt et al. (2012) to detect change within satellite image time-series is the following:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \epsilon_t \quad (3.1)$$

where the dependent variable y at a given time t is expressed as the sum of an intercept α_1 , a slope α_2 for potential temporal trend in the data, a sum of different frequency harmonic components representing seasonality ($\sum_{j=1}^k \gamma_j \sin(\frac{2\pi j t}{f} + \delta_j)$), and an error term ϵ_t . For the harmonic component of the model, $j = 1$ corresponds to the 1 year cycle, k is the chosen harmonic order, γ_j and δ_j correspond respectively to the amplitude and phase of the harmonic order j , and f is the known frequency of the time-series (i.e.: number of observations per year).

Here, we add an extra external explanatory variable (hereafter referred to as external regressor), x_t , to equation 3.1:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \alpha_3 x_t + \epsilon_t \quad (3.2)$$

where x_t corresponds to observation of the external regressor at time t . This addition, we hypothesize, should allow for better model fits and consequently more realistic predictions during the monitoring period. In this study two external regressors will be evaluated; MODIS NDVI and SPI as introduced in section 3.2.3.

Monitoring process

The breakpoint detection performed over the monitoring period works by estimating the deviation from the model of the newly acquired observations using a MOSUM window. From the model presented in equation 3.1, a standard linear regression can be formulated:

$$y_t = x_t^T \beta + \epsilon_t \quad (3.3)$$

where x_t^T is the vector of regressors, β a vector of regression coefficients and ϵ_t the residual term. From this equation, $\hat{\beta}$ which we define as the vector of coefficients for the stable history period, can be estimated by OLS. The monitoring process then consists in testing whether $\hat{\beta}$ remains a valid predictor through time for the incoming data during the monitoring period. The MOSUM used for that purpose can be expressed as follows:

$$MO_t = \frac{1}{\hat{\sigma} \sqrt{n}} \sum_{i=1-h+1}^t (y_s - x_s^T \hat{\beta}) \quad (3.4)$$

where h is the width of the MOSUM window. The hypothesis of structural stability is rejected when MO_t significantly deviates from 0, and crosses a boundary defined by the functional central limit theorem (Leisch et al., 2000). See Zeileis et al. (2005) for further details about the boundary used.

Stability of the history period

To facilitate reliable monitoring, the history period needs be free of disturbances (i.e. referred to as stable history). A stability test, part of the BFAST framework, selects the stable history by moving backwards in time while assessing the Reversed-ordered-cumulative sum of residuals (ROC) (Pesaran & Timmermann, 2002; Verbesselt et al., 2012). Having a sufficiently long stable history period for model fitting is critical for an accurate detection of change. For instance Verbesselt et al. (2012) suggest a stable history period of at least two years to accurately monitor change when using MODIS time-series with a 16 days temporal resolution. Although Landsat data are irregularly spaced in time, making such definition of a minimum length impossible, the temporal resolution being equal or lower than MODIS 16 days composites suggests a minimum length of history period equal or longer than two years for change monitoring with BFAST.

3.2.2 Study area

In order to test and validate the approach proposed in this paper, we chose a case study area situated in lowland Bolivia in the department of Santa Cruz (Figure 3.1). The area is a subset of the Landsat scene path/row 230/71 covering about 40 by 50 km of a forested area nearby the town of Concepción, where forest cover loss is known to occur. The type of forest of the study area, which can be categorized as tropical dry forest presents interesting challenges for change detection. Although it is in a tropical environment, numerous species in that forest type are deciduous and unlike tropical rainforests for which seasonality in the spectral signal is still being debated (Huete et al., 2006; Xiao et al., 2006; Saleska et al., 2007; Myneni et al., 2007; Samanta et al., 2010, 2012; Silva et al., 2013; Morton et al., 2014), a clear seasonality appears in the vegetation indices time-series for that area (Figure 3.2). In addition, a strong inter-annual variability can be observed in the phenological signal, with very low NDVI values reached during the dry season of certain years. Figure 3.2 depicts the temporal profiles of rainfall, average NDVI from MODIS and SPI which is a moisture stress metric. Large drops in SPI, visible in Figure 3.2, which appear to coincide with dips in the NDVI profile correspond to abnormally dry years. Details about SPI and NDVI can be found in a subsequent section of the present paper. Analysing these profiles in parallel, low SPI values, which correspond to negative rainfall anomalies, appear to coincide with the exceptional drops in NDVI, which may infer that greenness dynamics in these environments are largely driven by water availability. Such

strong inter-annual natural variations in NDVI make the change detection monitoring process particularly challenging. Pixel based approaches, for which a statistical model is fitted to the data and change detection based on stability through time are likely to detect change in these exceptionally dry years, which can be attributed to the nature of the time-series but not necessarily to anthropogenic disturbances. For a more thorough description of these forest environments and their ecology, please refer to Mostacedo et al. (2009).

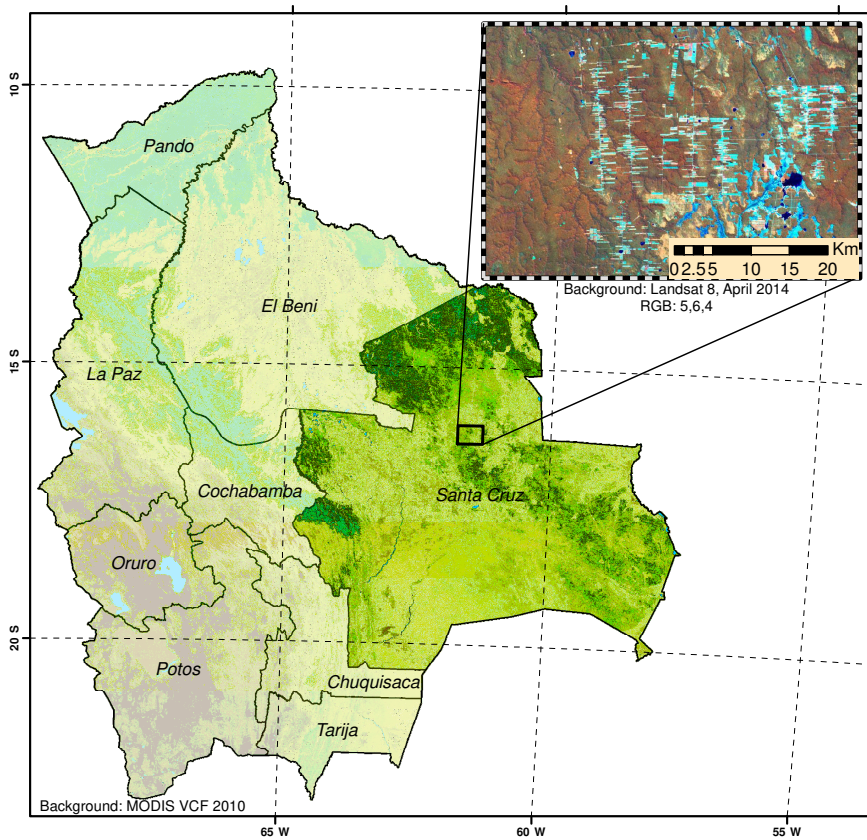


Figure 3.1: Map of the study area, subset of Landsat path/row 230/71 located in the region of Santa-Cruz, lowland Bolivia

3.2.3 Datasets and indices

In the following section, we present the datasets used to demonstrate the approach and evaluate its performances. When mention is made of VHR, we refer to the sub 5 m spatial

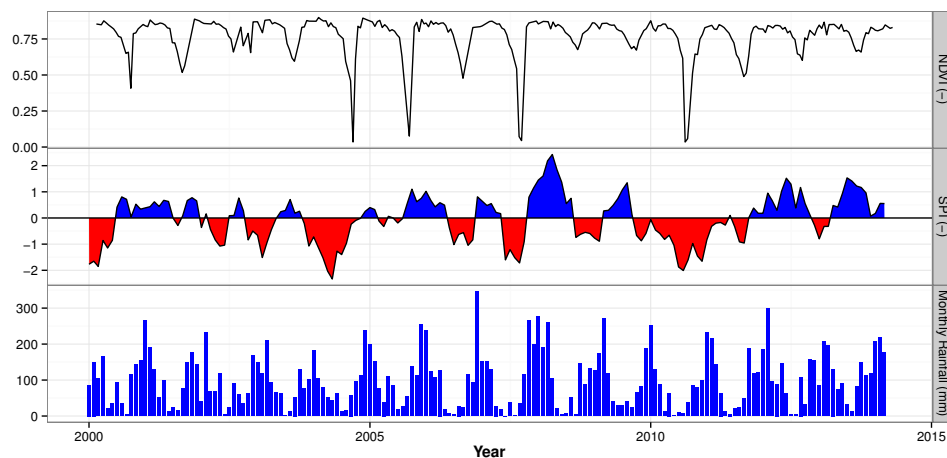


Figure 3.2: Temporal profiles of a) NDVI from MODIS spatially averaged over the study area. b) SPI computed from monthly TRMM data, with a 6 month time scale. c) Monthly rainfall from TRMM 3B43

resolution data, while MODIS types of data (250–1000 m range) are considered moderate resolution.

Landsat data

NDVI calculated from Landsat surface reflectance data at 30 m resolution is used as response variable in the change detection process. NDVI, which is a measure of vegetation “greenness” has been used and described extensively in remote sensing studies relating to vegetation (Tucker, 1979; Jackson & Huete, 1991). We chose NDVI over other metrics such as the Enhanced Vegetation Index (EVI) (qing Liu & Huete, 1995) or other more advanced indices, considering the robustness of NDVI and the fact that it can be calculated in a comparable manner from various data sources. Although EVI was shown to perform well over forested areas, and be less sensitive to saturation effects for high photosynthetic activity (Huete et al., 2002), EVI is also known to be more sensitive to varying atmospheric conditions. In addition, a recent paper by Morton et al. (2014) describes NDVI as being less sensitive than other indices to varying sun-sensor geometry, hence making a NDVI a safe and robust option for time-series analyses. All Landsat (ETM+ and TM) data acquired were downloaded already processed to surface reflectance from the United States Geological Survey (USGS) archives and only scenes with LT1 terrain correction were selected. In total, this resulted in 328 scenes spanning the 2000–2014 range for the study area. The USGS uses the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006) to perform automatic atmospheric correction

of the Landsat archive, while the *fmask* algorithm (Zhu & Woodcock, 2012) is used for cloud detection. We used the *fmask* cloud mask in order to ensure that all pixels included in the change detection process were clear land pixels.

Forest mask

For additional Landsat data preparation, we used a forest/non forest mask as a way to exclude non forested pixels from the process. This step has two objectives which are: (i) to focus only on pixels of interest, which in this case are forested pixels, and (ii) to decrease processing requirements by decreasing the total number of pixels. For the needs of this paper, we used a forest mask derived from the global year 2000 Landsat based percent tree cover (Hansen et al., 2013), and considered as forest every pixel having a tree cover greater than or equal to 50%. The Landsat based percent tree cover product was produced using a decision tree classifier trained with interpreted very high resolution data (Hansen et al., 2011). It is to be noted that this is a simple way to produce a forest mask and potentially not the most accurate mask. However, the method presented in this paper is not bound to any forest mask, so that potential users may use any masks of various origins to perform this non-forest masking step.

External regressors

We experimented with two different datasets to be used in the approach as external regressor; NDVI derived from the MODIS sensors, with a varying window size around the Landsat pixel of interest, and SPI from the Tropical Rainfall Measuring Mission (TRMM) satellite (Figure 3.2). First pre-processing of MODIS NDVI is briefly described, then SPI and the data from which it was computed are introduced.

MODIS NDVI One external regressor tested in this study is NDVI derived from the MODIS sensors. The product name more specifically is MOD13A2, which is the NDVI product generated at 1 km resolution from the MODIS sensor on-board the Terra satellite. In addition to the regular cleaning of the MODIS data using pixel reliability information to remove low quality and cloud contaminated observations (only pixels labelled as “good quality” were kept in that case), we filtered for potential unstable pixels through time. For that we used the MODIS based Vegetation Continuous Field (VCF) of Land Cover product (Hansen et al., 2003), product name MOD44B and only kept MODIS NDVI pixel time-series with percentage tree cover consistently higher than a certain threshold for further analysis. Such a step ensures that MODIS pixel highly affected by forest cover loss during the change detection process are not included as external regressor. As a way to assess the sensitivity of this threshold, hereinafter referred to as *VCF filtering*

threshold, as well as to gain further insights on the transferability of the method, we tested different combinations of this parameter, ranging from a 20% VCF filtering threshold to 70%. A low filtering threshold means that most MODIS NDVI pixels are included, potentially comprising non forested pixels and pixels that have undergone change through the considered period, while a high VCF filtering threshold restricts the MODIS NDVI pixels included as external regressor to the most densely forested pixels only.

SPI from TRMM The other variable tested in this study for use as external regressor is the Standardized Precipitation Index (SPI). SPI expresses relative water conditions, and can be interpreted as the probability of occurrence of a given amount of precipitation as compared to historical data (McKee et al., 1993). In simpler terms SPI reflects the cumulative precipitation anomalies; negative values correspond to abnormally dry periods while positive values correspond to abnormally wet periods. The index requires a time scale parameter for its computation, which corresponds to the temporal scale at which anomalies are cumulated. In this way the index integrates a temporal dimension about water stress status, hence automatically capturing the well known temporal lag of vegetation response to rainfall (Nicholson et al., 1990; Wang et al., 2003; Vicente-Serrano et al., 2013). For this reason SPI is thought to be a suitable variable to be used as external regressor in the change detection process. Vicente-Serrano et al. (2013) showed that the timing of vegetation response to droughts varies across biomes so that the time scale parameter of the SPI is highly context specific. We selected the time scale parameter by fitting a linear model using equation 3.2 for all possible time scales ranging from 1 to 12 months, $\alpha_2 t$ being MODIS NDVI at time t averaged over the study area, and we selected the time scale resulting in the best model fit. Six months was retained as the time scale resulting in the best model fit and therefore used in all subsequent analyses, which is in agreement with the findings of Vicente-Serrano et al. (2013). We computed SPI at 0.25° spatial resolution using rainfall estimates from TRMM, product name 3b43. 3b43 is a monthly product that combines rainfall estimates from TRMM and interpolated rain gauge measurement for correction. Adeyewa & Nakamura (2003); Dinku et al. (2007); Nicholson et al. (2003), and Franchito et al. (2009) assessed the product accuracy and generally found good agreement between in-situ rainfall measurements and TRMM estimates. The temporal profile of SPI with time scale of 6 months for the study area is depicted in Figure 3.2.

Validation dataset

A spatio temporal validation is essential in order to assess the performance of the method. For this study we chose to build a validation dataset by performing a visual interpretation of one VHR image (from Bing map (<http://www.bing.com/maps/>)) combined with the Landsat temporal profile of the associated pixel. In the absence of ground validation

dataset, visually interpreting the Landsat time-series is the only alternative to obtain a dataset that matches the temporal precision of the algorithm (Zhu et al., 2012). This interpretation was facilitated by the use of the TimeSync environment (Cohen et al., 2010). TimeSync is an environment developed in order to facilitate human interpretation of temporal trajectories in remote sensing time series. Initially developed jointly with the LandTrendR algorithm (Kennedy et al., 2010) for its validation, it can greatly benefit the validation of any land dynamic detection method. It is important to note that most of the forest cover loss happening in the present study area consists of a one-way land cover change, without regrowth of the forest, but with agriculture as a follow-up land use instead (Müller et al., 2012; Killeen et al., 2007). Considering this scenario, a spatial interpretation of a recent very high resolution image can suffice to perform most of the spatial validation of omission and commission error. Using this dataset in combination with the associated pixel Landsat time-series allows the interpreter to determine the timing of change accurately and with a precision close to the Landsat data availability frequency. The validation dataset was generated by visually interpreting a set of 300 samples, divided in two strata (assumed change and assumed no change). The two strata were generated from the Global forest cover loss product (Hansen et al., 2013). Bing map was used instead of Google Earth as the environment for VHR interpretation for it contained the most recent imagery, the whole area being covered with VHR images acquired in 2012. After interpretation and cleaning of samples with large uncertainties (e.g. proximity to forest edges), 258 validation points remained out of the original 300 samples.

3.2.4 Implementation

Model parameters

From equation 3.2, one important parameter to be set is the harmonic order j . While the study area is known for having a simple, single vegetation production season per year, also visible in Figure 3.2, it is interesting to test different harmonic orders. Higher orders may provide better fits and increase the change detection accuracy. We therefore tested, the three harmonic orders from $k = 1$ to $k = 3$ and investigated the effect of varying this parameter on the accuracy of the method. Another important parameter is the width of the MOSUM window used during the monitoring process; similarly to Verbesselt et al. (2012) we chose a value of $0.25 \times n$, with n being the size of the history sample.

Temporal component

In order to run the regression model, the two datasets need to have matching observations at similar dates. Since these do not systematically occur, we created daily time-series for

the external regressors by linearly interpolating values between successive dates. We subsequently only kept values from dates with matching Landsat observations. An other option would consist in interpolating the Landsat time-series instead, but because of more irregular observations and higher noise in the Landsat time-series we decided to interpolate the already smoother moderate resolution time-series.

MODIS NDVI residuals extraction

Considering equation 3.1, which is a nested part of the model presented in equation 3.2, seasonality and trend are already being modelled. Given that MODIS NDVI is also a variable with seasonality, which should follow similar patterns than Landsat NDVI, we decided to de-seasonalize and de-trend it before including it in the model. This step consists in extracting the residuals from a linear model fit using equation 3.1 and $k = 1$. The removal of the trend may also be able to compensate for a potential degradation of the external regressor over time due to increasing land cover heterogeneity at sub MODIS pixel level. Such step does not apply to SPI, which already represent anomalies and consequently does not have seasonality directly comparable to the one of Landsat NDVI.

Spatial component

Following the idea to take advantage of the difference in spatial extent between natural and anthropogenic processes, we investigated various spatial configurations for the use of MODIS NDVI. The aim of this exercise is to find an optimal size of spatial aggregation window for MODIS NDVI. The configurations differed in size (8 levels from 1 to 225 (15^2) MODIS pixels) as well as in the aggregation function used (mean and median). Larger aggregation windows are expected to be less sensitive to forest cover loss, while NDVI from smaller windows should better correlate with Landsat NDVI. A similar spatial aggregation step does not apply to the case of SPI. The interest of including SPI as an external regressor lies in the variable itself, assumed to be a driver of vegetation greenness and therefore being a potentially good predictor of NDVI. Consequently SPI was simply used as external regressor on a pixel basis. Table 3.1 presents the different levels of the 4 method parameters tested; in total this resulted in 288 combinations.

Table 3.1: Summary of the different parameters tested in the method implementation

VCF filtering (%)	Harmonic order (j)	Aggregation window size (1 km^2 pixels)	Aggregation function
20	1	1	mean
30	2	9 (3^2)	median
40	3	25 (5^2)	
50		49 (7^2)	
60		81 (9^2)	
70		121 (11^2)	
		169 (13^2)	
		225 (15^2)	

3.3 Results and discussion

3.3.1 Overview

The following section presents the results of the change detection process performed over the study area. We tested all combinations of window size, harmonic order, aggregation function and VCF filtering (3.1). Accuracy figures are produced using the TimeSync validation samples. For all configurations, monitoring started in January 1st 2008 and extended until the date of the most recently acquired Landsat scenes (May 4th 2014). The change map, with timing of change, resulting from the configuration with highest accuracy (accuracy assessment presented in section 3.3.2) is depicted in Figure 3.3. An initial observation of the map reveals that the prediction of change is rather conservative, with little isolated pixels being flagged as change but rather distinct spatial patterns coherent with deforestation patterns occurring in the area. Although this is not systematic, isolated pixels often reveal false detections, while human induced changes tend to present more geometric features (e.g., fish bone structure with deforestation strips perpendicular to the roads) (de Filho & Metzger, 2006). While most of the roads in the present case appear to be properly detected, a few seem to be missing from the change detection results. These omissions are likely to be attributed either to the spatial characteristics of the roads, too narrow to be detected by 30 m Landsat pixels, or to the time at which these roads were constructed. Any feature already present prior to the period under investigation, will not appear in the change detection output, and that is certainly the case for many of the undetected roads. Regarding the timing of change, it is encouraging to note that changes detected around similar dates are spatially clustered, which is coherent with the switch from forest to agriculture process, where large strips of forest are removed at once and transformed into agricultural fields. We also noticed a few isolated pixels surrounding the

dark organic shaped features in the south eastern corner of the study area. These zones are non forested, and these detections, we assume, are mostly the results of inaccuracies in the forest mask used. From these observations it is clear that the quality of the forest mask is reflected in the output of the change detection process. Although we considered that thresholding the global Landsat based percent tree cover product (Hansen et al., 2013) was sufficient for the purpose of this study, we encourage users of the approach who require the highest quality output to carefully consider their choice of forest mask. Alternatives to obtain forest masks include local, supervised and object based classifications methods (Dorren et al., 2003) or approaches based on multitemporal data, as suggested by Zhu et al. (2012).

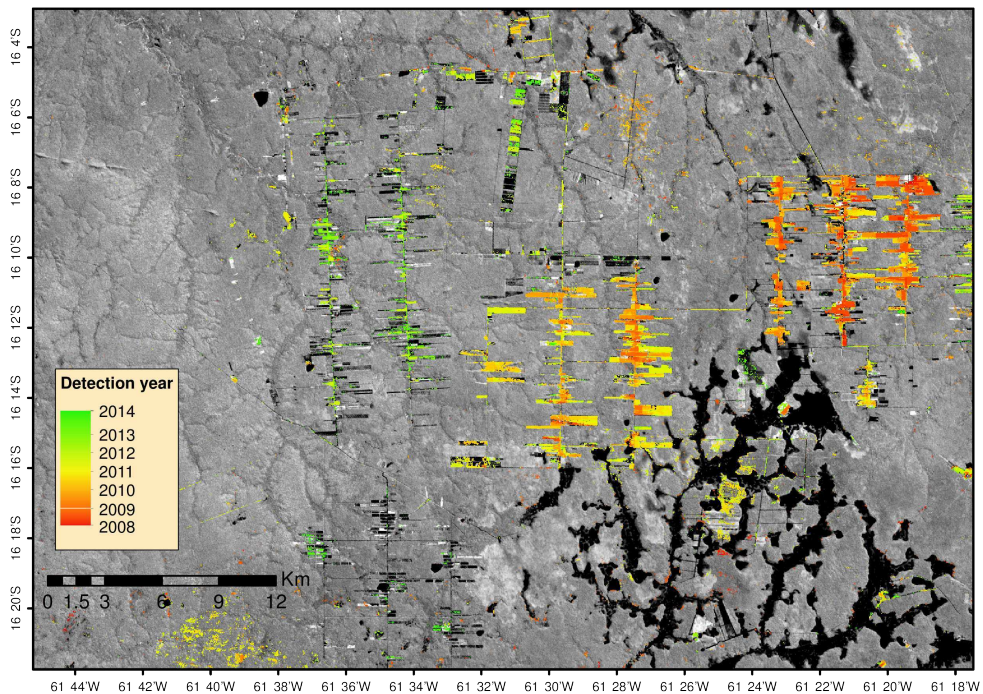


Figure 3.3: Change map of the study area after 2008 produced using the method with external regressor yielding the highest overall accuracy (169 MODIS pixel, mean aggregating function, harmonic order 2 and VCF filtering threshold of 60% (section 3.3.2)). NDVI from Landsat 8 acquired on April 26th 2014 is used as background layer. The continuous colours indicate timing at which change was detected by the algorithm.

3.3.2 Accuracy assessment

Accuracies calculated using confusion matrices between the TimeSync validation samples and the results for the different configurations (Table 3.1) are displayed in Figure 3.4 and Figure 3.5. Although the main objective of the study is to investigate the effect of including an external regressor, hence taking advantage of the spatial context, we also present in the section below the impact on the accuracy of varying some other parameters of the method. These additional comparisons and assessments can potentially inform on the limits of the method and help in understanding its behaviour. Accuracy for the configuration using SPI as external regressor was very low and is not presented in any figure. This configuration resulted in a commission error of 28%, an omission error of 37% and an overall accuracy of 67%. Well below all the other configurations, including the single pixel approach.

We decided to only display a single aggregation function (mean) on Figure 3.4 and Figure 3.5, since differences in the accuracy of using mean or median as aggregation function were marginal.

First of all it is important to note that in all cases, the overall accuracy benefited from including MODIS NDVI as external regressor (Figure 3.4). The maximum overall accuracy with the single pixel approach is achieved with $k = 2$, resulting in an accuracy value of 83%, while the highest accuracy recorded for all combinations has an overall accuracy of 87% and is the result of the combination with a VCF filtering threshold set to 60%, an aggregation window size of 169 km², a mean aggregation function and a second harmonic order ($k = 2$). The resulting map from this combination is presented in Figure 3.3.

Harmonic order

While the results show that increasing the harmonic order of the model has a strong effect on the accuracy of the single pixel method, particularly between $k = 1$ and $k = 2$, with an increase in overall accuracy from 74% to 83% between $k = 1$ and $k = 2$ respectively, varying the harmonic order has a smaller effect on the overall accuracy of the method with external regressor. This difference in effect of the harmonic order can possibly be explained by a more dynamic fit in higher harmonic orders, better able to cope with intra-annual variability in the single pixel case, while this more dynamic fit is directly provided by the external regressor itself in the case of the combinations with external regressors.

We also computed commission error estimates (not presented in the figure), which also seem to react mostly to changes in harmonic order. Commission error for $k = 1$ are in nearly all cases with external regressor equal to 0% and vary between 2% and 4% for third harmonic order. This observation may suggest a more conservative approach when

using low harmonic orders (e.g.: $k = 1$) and higher probability of obtaining false positive in higher harmonic orders. Although encouraging these low commission error results have to be considered with care, for the type of validation we opted for is likely to under represent the no change stratum, which despite the large proportion of change within the study area remains much larger than the change stratum. Considering that an increasing harmonic order has the effect of increasing the commission error, but on the other hand slightly reduces overall accuracy by increasing the omission error, the choice of harmonic order appears to be trade off between a more conservative approach and a more sensitive approach.

VCF filtering and aggregation window size

To understand the effect of varying the VCF filtering threshold, it has to be considered together with the varying window size effect. First it is interesting to note that low VCF filtering levels have little to no effect on the accuracy. The VCF thresholding was initially thought as a mechanism to make the method more robust in heterogeneous landscapes, potentially containing large proportions of non forested areas. We were consequently expecting lower accuracies for low VCF filtering threshold due to a contamination of the external regressor by non forested pixels. This indicates that filtering or not for unforested pixels in the external regressor does not affect the performances of the method. However, as a precaution, we still recommend excluding a maximum of non-forested pixel in the external regressor. There is a clear drop in overall accuracy, and a jump in omission error, between 30% and 40% filtering thresholds for small aggregation window size, and between 60% and 70% for all window sizes. We can explain this pattern by an exclusion of nearly all MODIS pixels by the filtering threshold making the method unable to produce results and consequently decreasing the accuracy. This happens later for larger windows for the probability of encountering pixels that satisfy a certain condition becomes higher when the number of pixel increases. It is interesting to note that increasing window size seems to result in slightly better accuracies as well, although the range of variation is small (overall accuracy improvement from 84% to 87% for the two extreme cases). Overall, best accuracies were obtained when aggregating MODIS NDVI residuals from a 13×13 km (169 km^2) window. We therefore recommend using large aggregation windows since it appears to robustify the method in spatially heterogeneous environments and provides a slight increase in accuracy. However, it is predictable that more heterogeneous landscape will present a greater challenge for the use of MODIS NDVI and the optimisation of window size and aggregation function parameters (Hansen et al., 2008). Accordingly a transfer of the proposed approach to different study areas, particularly if they present strong landscape heterogeneity should be preceded by a careful investigation of its appropriate spatial parameters. In order to facilitate such transfer, a consideration for further research would be to develop a mechanism leading to a more dynamic external regressor, with real

time land cover change information informing the reliability of the external regressor, or an external regressor weighting based on neighbouring pixels reliability.

SPI

We found the method to perform poorly when SPI was used as external regressor. Reasons for the low performances of SPI are not clear. We hypothesize that the forests in that area are not sufficiently water limited to strongly react to dry conditions, hence making water status a poor predictor of forest greenness. Additionally, it is possible that SPI is not the ideal way to include moisture conditions in that environment. Limitations of the index, which considers rainfall anomalies alone without including evaporative demand have been discussed by Trenberth et al. (2014). Another explanation could come from a non linear response of vegetation to water status; with droughts not necessarily affecting greenness (Huete et al., 2006) and excess rainfall increasing trees vulnerability to large scale natural disturbances. Facing these uncertainties, and the potential benefits of including a variable that is fully independent from Landsat NDVI, the use of climatic variables as external regressor is a topic of great interest that will require to be further investigated.

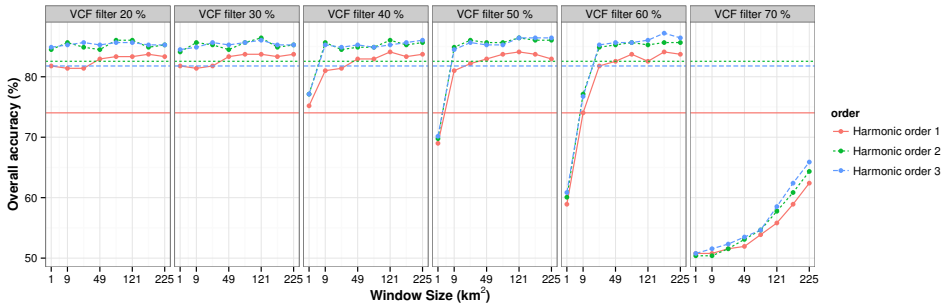


Figure 3.4: Overall accuracy of the method with varying aggregation window size, harmonic order and VCF filtering threshold. The largest window size (farthest points to the right) are 225 km^2 , which corresponds to a $15 \times 15 \text{ km}$ window. Horizontal lines correspond to the accuracies for the single pixel configuration.

3.3.3 Qualitative assessments

In order to get further insights on the performances of the method presented and to better understand the mechanisms leading to higher accuracies, we performed two extra qualitative analyses. First we created maps of agreement using the validation samples, for the single pixel and the configuration yielding the highest accuracy. We expect from these maps that spatial patterns of agreement may reveal performance differences related to the

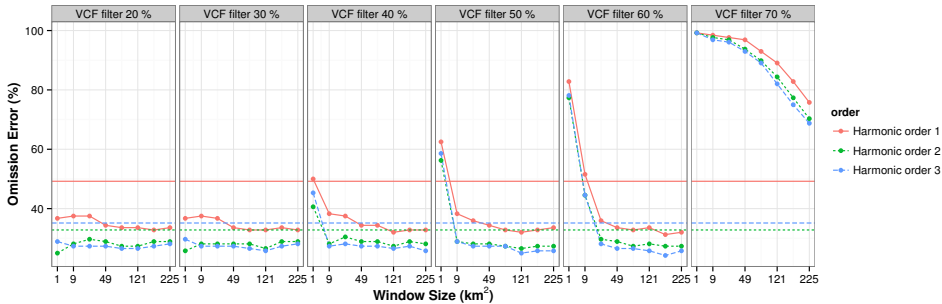


Figure 3.5: Omission error of the method with varying aggregation window size, harmonic order and VCF filtering threshold. The largest window size (farthest points to the right) are 225 km^2 , which corresponds to a $15 \times 15 \text{ km}$ window. Horizontal lines correspond to the omission errors for the single pixel configuration.

spatial context. Secondly we extracted pixels for which results differed and observed the behaviour of the algorithm for these selected pixels.

Spatial accuracy

The maps of agreement are presented in Figure 3.6. A first observation concerns the 3 samples flagged as false positive (blue dots) in the south-western corner of the area for the *pixel* configuration (upper panel). These three false detections, which were properly omitted in the method with external regressor, appear to occur in an area with a lot of noise, where pixels flagged as disturbance do not follow geometrical patterns. This type of false detection, which we could characterize as speckle appear to be partly reduced when using the method with MODIS as external regressor. The reasons for the occurrence of this spatial cluster of noise remain un-explained. Regarding omission error, if one compares Figure 3.3 with areas of reduced omission error, it is interesting to note that most differences between the two configuration appear to occur in areas that experienced recent changes (vertically aligned samples in the top centre of the study area and eastern samples). This pattern may indicate a higher sensitivity of the proposed method with MODIS as external regressor as time increases since the beginning of the monitoring period. However, these areas of recent deforestation still concentrate a large part of the omission error. This apparent lower sensitivity of the method towards the end of the monitoring period can be explained by the MOSUM values boundary used (see Zeileis et al. (2005), equation 7) for the monitoring process. These boundaries are diverging from 0 as a function of time, meaning that larger magnitude disturbances need to occur for the hypothesis of structural stability to be rejected as time from the start of the monitoring period increases. Implementation of sequential monitoring as proposed by DeVries et al.

(2015b) provides an alternative to that decreasing sensitivity. The principle of sequential monitoring consists in working with only short monitoring period at a time (e.g.: 1 to 2 years) and subsequently extending the stable history period for stable pixels before monitoring again for 1 or 2 extra years. With that technique the sensitivity of the method remains consistent through time. Although this approach was not implemented in the present study, sequential monitoring is perfectly suited to the proposed approach and if implemented would certainly increase the sensitivity towards the end of the monitoring period, hence further enhancing the accuracy of the method.

Individual pixels assessment

A set of three individual time-series for which results of the single pixel and the configuration with a 121 km² aggregation window differed was investigated. The comparisons are presented in figures 3.7, 3.8 and 3.9. Figures 3.7 and 3.8 are two examples in which commission error is reduced thanks to the addition of an extra regressor, while Figure 3.9 presents the case for which an omission error is avoided by using the proposed method. A common pattern for these three time-series is that the model with MODIS regressor always provides a better, more adaptive fit, that is able to predict inter-annual variability and exceptional variations much better than the seasonal trend model. This is true both during the history period, in which case the ROC stability test benefits from this better fit (Figure 3.7), and during the monitoring period. In the latter case, exceptionally low NDVI values may trigger false detections (Figure 3.8). Finally, benefits of the method from the omission error reduction side also appear to be provided by better fit during the history period, increasing its length and consequently the sensitivity during the monitoring period, as illustrated by Figure 3.9. In fact, following Verbesselt et al.'s (2012) recommendations to use a minimum history period length of 2 years in the case of a regular MODIS time-series, a time-series like the one shown in Figure 3.7, with a temporal density lower than a MODIS time-series, cannot be accurately monitored. In practice, given that it does not meet a requirement of the monitoring framework, such a pixel should be unclassified in a post processing step, and if not, is likely to be wrongly classified. The inclusion of the external regressor in Figure 3.7 case therefore allows such pixel to meet the minimum 2 year requirements to monitor and detect change.

3.3.4 Temporal accuracy

The method performs continuous monitoring over the defined monitoring period and therefore produces timing of disturbance with precision equalling the data acquisition frequency. Having gathered information about timing of disturbance in the TimeSync validation sample, we were able to produce some visual estimates of accuracy for the timing of change outputted by the method. The metric under investigation in that case

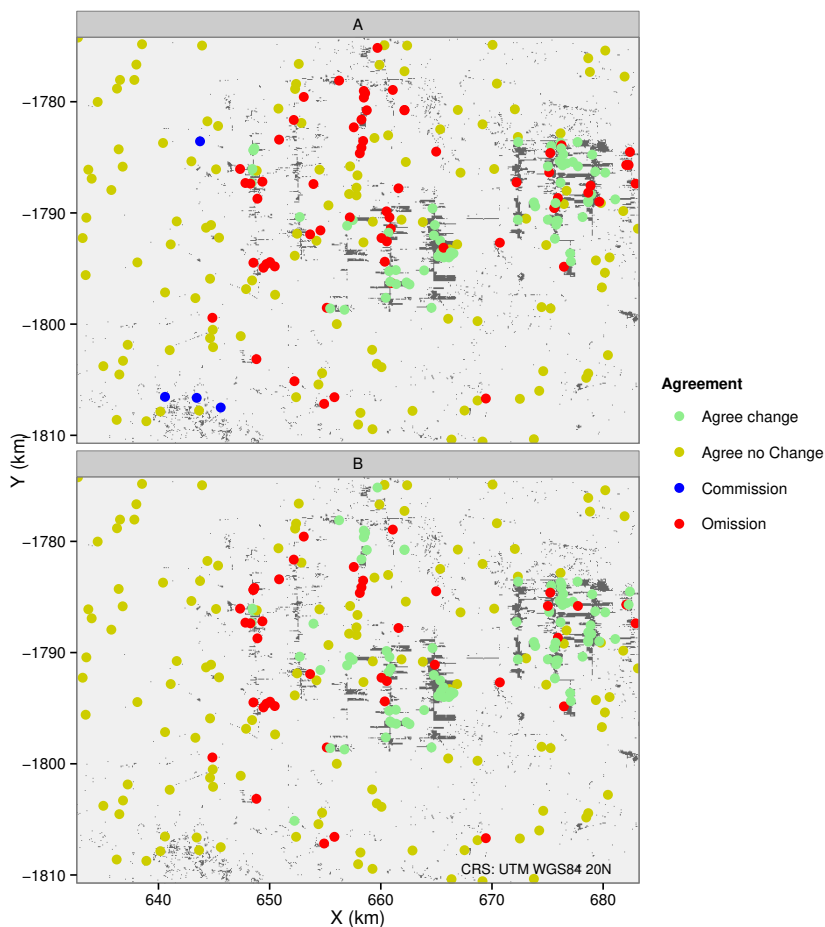


Figure 3.6: Map of agreement between the validation samples and the corresponding pixels. The upper panel (A) represents the pixel based approach without external regressor, while lower panel (B) presents the agreement with the method using MODIS residuals as external regressor. For both panels the background layer is the resulting change map from the configuration with external regressor.

to assess the temporal accuracy of the method is the time lag between the moment a disturbance has occurred (assessed in TimeSync) and the detection time by the algorithm. Figure 3.10 depicts this detection lag for the single pixel and the configuration with external regressor. It is interesting to note that a large amount of samples present a negative lag. This could be interpreted as an error, however, it is more likely to be due to the way samples are interpreted in the TimeSync environment. TimeSync time-

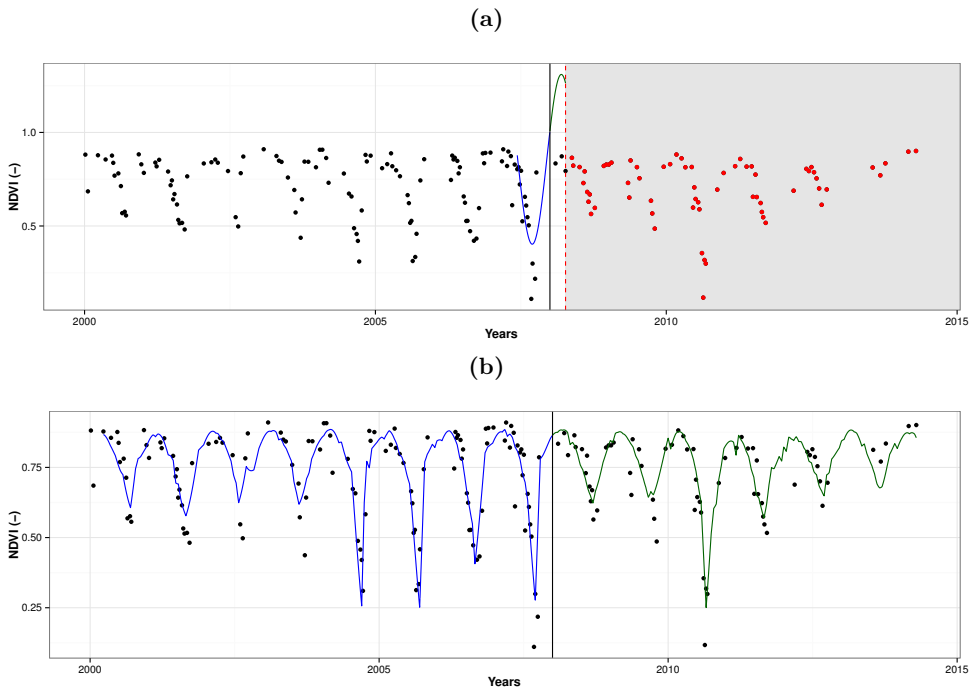


Figure 3.7: Example of a falsely detected break when applying the single pixel approach method (A), as opposed to the same pixel properly modelled when using MODIS NDVI residuals as external regressor (B). False detection in that case is to be attributed to instability in the history period not resulting in a realistic model and prediction. Green line, blue line, vertical dashed red line and vertical black line correspond respectively to stable history period model, model predictions during the monitoring period, detection of a break-point and beginning of the monitoring period.

series interpretation consists in identifying break points surrounding disturbance events in the investigated pixel's temporal profile (Cohen et al., 2010). The pixel's temporal profile appears as a broken line made of vertices connected by segments; each vertex corresponds to a cloud free observation and there is one vertex per year displayed on the temporal profile. The timing of disturbance we recorded for producing Figure 3.10 corresponds to the date of the first post change vertex identified. While it is possible in the TimeSync environment to refine vertices selection to be as close as possible to the time of the event observed, we did not put much emphasis on accurate timings when interpreting the samples. Consequently, in our validation dataset, part of the timing of disturbances may very well have occurred half a year or more before the labelled vertex. Considering this potential artifact and the distribution of the lag histogram of the configuration with external regressor, we can say that a reasonable proportion of

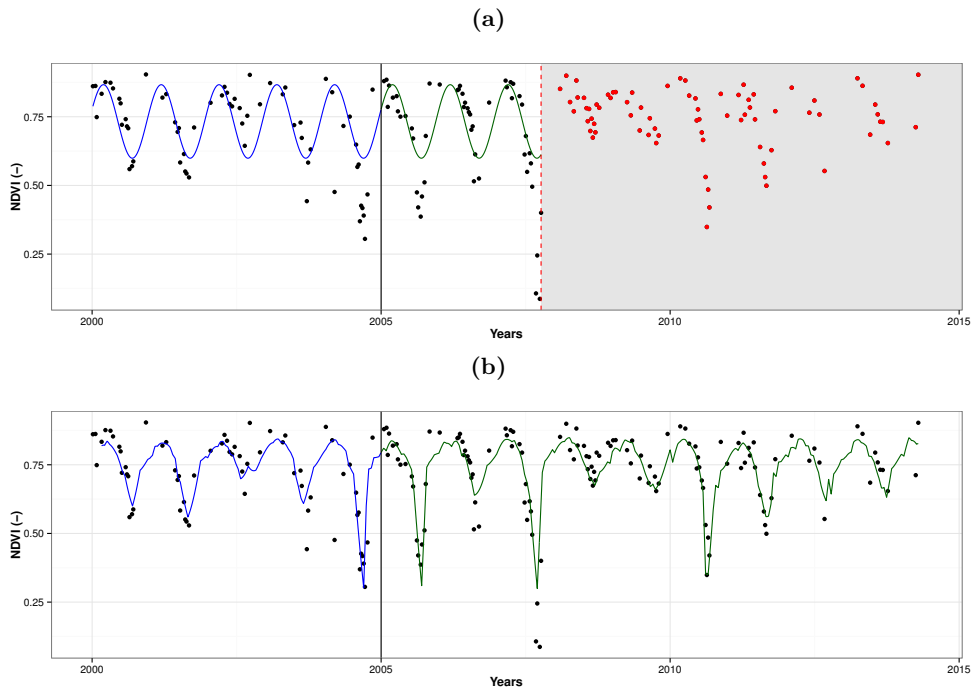


Figure 3.8: Example of an undisturbed pixel for which a break was detected using the single pixel approach due to strong inter-annual variations (A). The use of MODIS NDVI residuals as extra regressor (B) was able to compensate for this effect. For details about the different elements of the figure, refer to the caption of Figure 3.7.

disturbances are detected by the method within a year. Looking more specifically at the single pixel time lag distribution, one can see a few samples which have been detected long before the actual disturbance occurred. These detections are likely to be false positives, which happened to experience a disturbance at a later time. The rest of the single pixel configuration distribution is relatively spread with a peak slightly in the positive part of the time lag, meaning that disturbances are likely detected with a small delay. The external regressor time lag distribution has a slightly different shape, much narrower and centred on 0 days. Considering the temporal precision and accuracy of the validation dataset, as discussed above, it is difficult to conclude on the absolute temporal accuracy of one method or another. However, given that the same samples are used, with the same potential bias or systematic error, a relative comparison of one configuration against the other remains possible. Then, comparing the two distributions, corresponding to the single pixel and the external regressor configurations in Figure 3.10, we can conclude that the configuration with external regressor tends to detect disturbances earlier than the single pixel configuration. This aspect can be at least partly explained by the better model fits

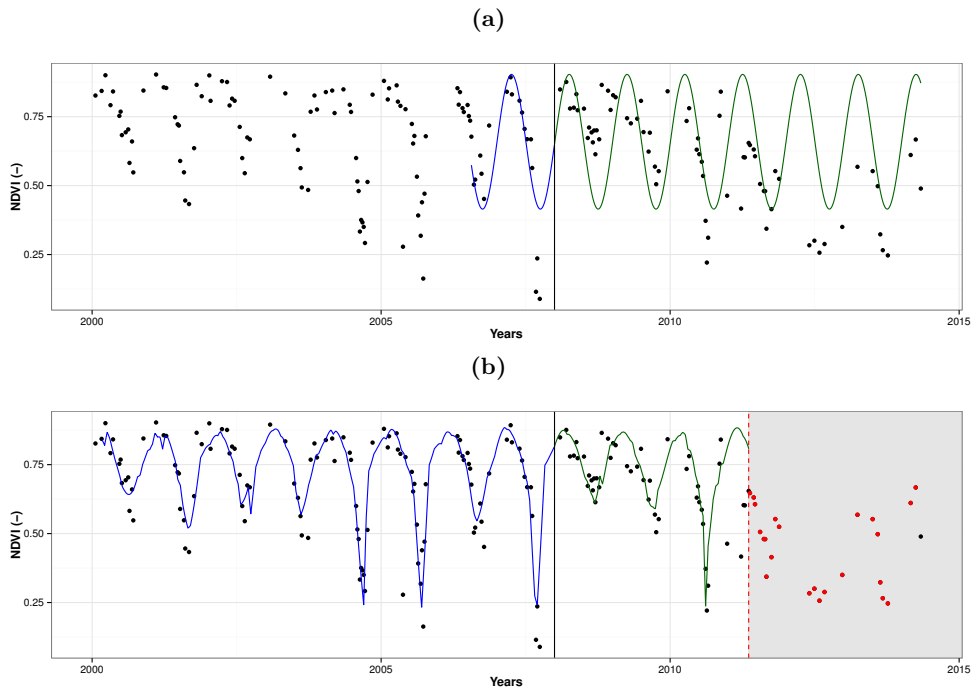


Figure 3.9: Example of a time-series for which a disturbance remains undetected when using the simple pixel based method (A), while it is detected in the configuration with external regressor. For details about the different elements of the figure, refer to the caption of Figure 3.7.

as illustrated in the previous section. The fact that conclusions on absolute temporal accuracy are difficult to draw despite our validation dataset containing some estimate of disturbance timing reveals the complexity to obtain such complete and accurate spatio-temporal validation dataset and consequently the difficulties associated with validation of near real time change detection method. A more thorough human interpretation of Landsat time-series is a promising way, and in that regard, environments such as TimeSync are of great value (Cohen et al., 2010). While spatial thematic accuracy assessments are now well developed, which can at least partly be attributed to land cover mapping efforts (Foody, 2002), it is now time to develop robust frameworks for spatio-temporal accuracy assessments. Zhu et al. (2012) propose an interesting way to present temporal accuracy in a near real time detection context by using proportions of samples for which disturbance detection occurred before or at the same time than the reference.

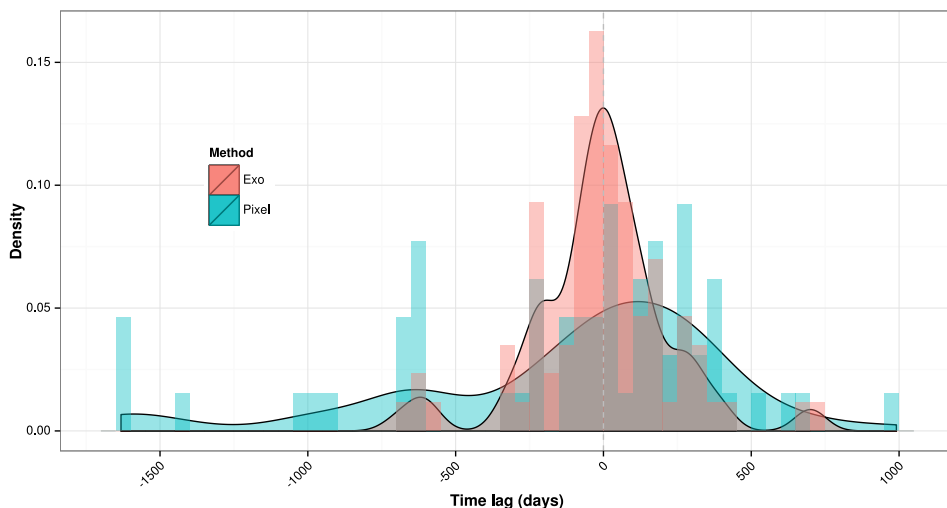


Figure 3.10: Histograms and associated smoothed density estimates presenting the time difference (in days) between observed times of disturbance (validation samples) and predicted times from the single pixel approach and the configurations with external regressor.

3.3.5 General considerations

We evaluated how external regressors can improve the forest cover change monitoring using the BFAST framework (Verbesselt et al., 2012). Such approach differs from a data fusion approach in the way that the added data source is used externally in order to better explain variations in the observed variable, and not fused into a single time-series. The hypotheses were that by taking advantage of the different characteristics of different datasets, change detection results could be improved. The two hypothetical improvements formulated were that: (i) by including NDVI at a different spatial scale we can discriminate natural phenomena from anthropogenic disturbances and (ii) including SPI will result in better Landsat NDVI predictions also resulting in an offset of natural variability. The method was demonstrated using Landsat, MODIS and TRMM data, and only the first hypothesis could be confirmed, provided by the results of the Landsat-MODIS configuration. Although demonstrated with MODIS and Landsat, the approach is not bound to the use of specific datasets, but rather introduces a generic spatial implementation framework to guide the use of external regressors in multi-temporal regression based change detection methods. This non specificity is particularly relevant in the current context where an increasing amount and variety of remotely sensed data are becoming available. The upcoming launch of the Sentinel 2 constellation for instance holds great promise to enhance forest cover change monitoring capabilities (Drusch et al., 2012). However, for further improvements to be achieved, the potential synergies existing between datasets

have to be fully exploited (De Sy et al., 2012). This requires using and further improving innovative methods accordingly (Lu et al., 2014). As such, considerations for further research regarding the proposed approach were mentioned in the previous paragraphs. Those include: (i) more sophisticated spatial context considerations, such as definition of objects for spatial aggregation rather than windows and assessment of neighbouring pixels reliability to increase performance of the method in heterogeneous environments, (ii) further investigations on ways to use climatic data as external regressor, and (iii) a more thorough temporal validation of the method. Finally, while the proposed approach was successful in improving change detection accuracy for the study area, we also acknowledge that it adds a bit of complexity to the overall change detection framework. Different near real time approaches can co-exist and should be used depending on user requirements and context. While a pixel based approach such as the one proposed by Verbesselt et al. (2012); Zhu et al. (2012); Brooks et al. (2014), and DeVries et al. (2015b) might yield good results in many stable environments with sufficient data available, more variable environments may benefit from the addition of an external regressor as presented here, and a fusion with radar time-series will be more suitable in presence of high cloud cover (Reiche et al., 2015).

3.4 Conclusion

In this paper we evaluate how external regressors can improve the forest cover change monitoring using the BFAST framework (Verbesselt et al., 2012). Advantages of adding an additional external variable are that the spatial context can be included, hence taking advantage of the scale difference between natural and anthropogenic phenomena. The approach was tested for a dry tropical forest area in Bolivia. Validation performed against a set of sample pixels interpreted using the TimeSync environment revealed accuracy improvements of the proposed method, when compared with the single pixel approach. Overall accuracy increased from 83% to 87% when including MODIS as external regressor in its best configuration. These improvements are explained by a better model fit to the time-series, particularly during extreme natural events such as droughts, hence resulting in more realistic predictions of NDVI values during the monitoring period and a positive impact on the sensitivity of the method. Additionally, qualitative accuracy assessment of the timing of break detection suggests that the temporal accuracy of the method also benefit from the inclusion of the external regressor, and disturbances are detected with a shorter temporal lag. The combination of Landsat NDVI as dependent variable for observing the process, and MODIS aggregated from a 13×13 km moving window was found to result in the best algorithm performances for the study area. The method proposed is particularly relevant for environments that present large inter-annual variability. Such development is a step forward in the development of global automated near real time monitoring systems, for generation of forest cover change information.

The BFAST framework is available for use as a R package (R Core Team, 2014). Further information illustrating how to apply the method on satellite image time-series are available at <http://bfast.R-Forge.R-project.org/>.

Acknowledgements

The authors wish to thank Z. Yang (Oregon State University) and W.B. Cohen (US Forest Service) for their support with TimeSync. Additionally, the research received partial funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant agreement Number 283093 — The Role Of Biodiversity In climate change mitigatioN (ROBIN).

Chapter 4

Reconstructing land use history from Landsat time-series

This chapter is based on:

Dutrieux, L. P., Jakovac, C. C., Latifah, S. H., & Kooistra, L. (2016). Reconstructing land use history from Landsat time-series: Case study of a swidden agriculture system in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, *47*, 112-124.

Abstract

We developed a method to reconstruct land use history from Landsat images time-series. The method uses a breakpoint detection framework derived from the econometrics field and applicable to time-series regression models. The Breaks For Additive Season and Trend (BFAST) framework is used for defining the time-series regression models which may contain trend and phenology, hence appropriately modelling vegetation intra and inter-annual dynamics. All available Landsat data are used for a selected study area, and the time-series are partitioned into segments delimited by breakpoints. Segments can be associated to land use regimes, while the breakpoints then correspond to shifts in land use regimes. In order to further characterize these shifts, we classified the unlabelled breakpoints returned by the algorithm into their corresponding processes. We used a Random Forest classifier, trained from a set of visually interpreted time-series profiles to infer the processes and assign labels to the breakpoints. The whole approach was applied to quantifying the number of cultivation cycles in a swidden agriculture system in Brazil (state of Amazonas). Number and frequency of cultivation cycles is of particular ecological relevance in these systems since they largely affect the capacity of the forest to regenerate after land abandonment. We applied the method to a Landsat time-series of Normalized Difference Moisture Index (NDMI) spanning the 1984–2015 period and derived from it the number of cultivation cycles during that period at the individual field scale level. Agricultural fields boundaries used to apply the method were derived using a multi-temporal segmentation approach. We validated the number of cultivation cycles predicted by the method against in-situ information collected from farmers interviews, resulting in a Normalized Residual Mean Squared Error (NRMSE) of 0.25. Overall the method performed well, producing maps with coherent spatial patterns. We identified various sources of error in the approach, including low data availability in the 90s and sub-object mixture of land uses. We conclude that the method holds great promise for land use history mapping in the tropics and beyond.

4.1 Introduction

Land use and land use dynamics affect elements of the biosphere with local to global impacts (Foley et al., 2005). Forests are of particular importance in this system for the role they play in maintaining biodiversity levels and delivering ecosystem services such as climate regulation and water supplies (Foley et al., 2005; Myers et al., 2000). While old-growth forests are often considered first when accounting for these services, secondary forests, with an estimated 2010 area of 165,230 km² for the Brazilian Amazon alone (TerraClass, 2011), cannot be ignored since they too play an increasingly important role in the provision of ecosystem services (Bongers et al., 2015). However, multiple factors need to be taken into account when dealing with secondary forests. Secondary forests can succeed to a variety of land use histories such as logging, agriculture, or cattle ranching and these different histories will in turn impact the characteristics of the present forests. It has been shown that the vegetation structure, species composition and resilience of secondary forests are strongly affected by previous land uses (Jakovac et al., 2015; Lawrence et al., 2010; Mesquita et al., 2001). The longer the areas is kept under land use and, in the case of shifting cultivation, the higher the frequency of use, the lower the recovery rate and species diversity of the succeeding secondary forests (Zarin et al., 2005; Jakovac et al., 2015). Intensive land use can also result in a shift in species composition (Mesquita et al., 2001; Jakovac et al., 2016a) and can ultimately hinder the secondary succession (Longworth et al., 2014). Examples of previous land uses include swidden agriculture—also known as slash and burn agriculture, which is a type of shifting cultivation widespread in the tropics and from which a large part of today’s secondary forests originate (Van Vliet et al., 2012; Mertz, 2009). Swidden agriculture has a very particular rotation cycle which consists in cutting and burning the forest, cultivating the land for a period of one to three years and leaving the land fallow until the next rotation five to sometimes more than 20 years later (Coomes et al., 2000; Jakovac et al., 2015). For these systems specifically, research carried out at field level has shown that the frequency of land usage as well as the total number of cultivation cycles are important determinants of the current structure and function of the resulting secondary forest (Lawrence et al., 2010; Jakovac et al., 2015).

Considering the strong connection that secondary forests have with their previous land uses, land use history is an important aspect to take into account when measuring and modelling the current state of secondary forest ecosystems. However, despite its importance such information is not always straightforward to obtain. Interviewing local stakeholders can provide accurate information, but such effort can only be carried for a few punctual locations, while information on the remaining part of the landscape would still be lacking. There is therefore a need to develop standardized methods capable of deriving quantitative, spatial informations about past land uses in secondary forest systems (Van Vliet et al., 2012). Remote sensing techniques offer great promises to contribute

to this effort of studying the dynamics and evolution of these ecosystems in a spatial context. Satellites have been acquiring images of the earth for more than 40 years. This represents large amounts of data which have been collected in an objective, systematic and spatially continuous way. However, exploiting the remote sensing data for ecological purposes contains challenges as well. While variables measured in field inventories have a direct ecological meaning, optical remote sensing measures light reflected by the elements of the earth surface. Reflected light is a biophysical variable which does not have any direct ecological meaning and therefore does not provide useful information to the ecologist or the decision maker. There is therefore a challenge that consists in translating raw satellite measurements into useful and meaningful variables. One way of achieving this is by taking advantage of the temporal dimension of the remote sensing measurements. The earth has been continuously monitored by the various sensors on board the Landsat satellites since roughly the 80s (Goward et al., 2006). By assembling these repeated measurements, land dynamics can be extracted and used to provide useful information about the history and dynamics of the land (Kennedy et al., 2014).

A variety of methods has been developed to investigate change using remote sensing time-series (Lu et al., 2014). The wide range of techniques developed responds to various monitoring needs both in terms of spatial extent and dynamics observed. The land dynamics can be either gradual or abrupt and concern changes about the intra-annual or the inter-annual dynamics (Kennedy et al., 2014). Using simple linear trends, Dutrieux et al. (2012) and Fensholt & Proud (2012) have investigated gradual change regionally to globally using Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) coarse resolution time-series. More sophisticated approaches, such as de Jong et al. (2013) and Forkel et al. (2013) used the Breaks For Additive Season and Trend (BFAST) framework (Verbesselt et al., 2010b,a, 2014) to investigate shifts in vegetation trends globally. Studies at higher spatial resolution, nearly always based on Landsat data, include methods for near real time deforestation monitoring (Zhu et al., 2012; Brooks et al., 2014; DeVries et al., 2015b; Dutrieux et al., 2015; Reiche et al., 2015), as well as more general tools for vegetation dynamics and land trajectories monitoring (Kennedy et al., 2010; Huang et al., 2010; Zhu & Woodcock, 2014; DeVries et al., 2015a).

Here we are interested in applying a method to the case of swidden agriculture with the objective to map land use intensity defined as the number of times the land has been cultivated. Swidden agriculture dynamics are fast and complex, which adds additional challenges to the time-series approach we need to develop. This complexity originates firstly from the location of these systems. Because they are in the tropics where vegetation regrowth happens immediately after disturbances, the optical signal related to the swidden agriculture events fades rapidly making the temporal window to detect events very short (Asner et al., 2004a,b). An additional challenge comes from the repeated cycles, all containing gradual changes, such as the periods of forest regrowth, and abrupt changes

triggered by the burning events. Finally the high cloud coverage usually present above tropical forest regions further restricts data availability making the observation of dynamics even more challenging (Asner, 2001). These constraints call for a hyper-temporal approach capable of capturing both gradual and abrupt changes.

Here we propose to use a statistically based breakpoint detection method derived from the econometrics literature to retrieve land use history. The method is applicable to Landsat time-series hence producing information on past land uses at medium spatial resolution. We applied the method to the case of swidden agriculture, trying to quantify the land use intensity defined as the number of times an area has been cultivated for two municipalities of the Brazilian Amazon where swidden agriculture is the predominant cultivation practice. Performances of the method are assessed and discussed by comparing method's output with a ground truth dataset about agricultural management collected via farmer's interviews.

4.2 Material and methods

4.2.1 Time-series segmentation

Detecting change in time-series

As described in the introduction of this paper, detecting change may take several forms, depending on whether the change observed is gradual, abrupt, occurs on intra-annual dynamics, or between years (Kennedy et al., 2014). Here we propose an approach to perform full segmentation of the time-series. Abrupt changes, referred to as breakpoints are detected from changes in the coefficients of a time-series regression model (Bai, 1994). Periods between breakpoints, which we call temporal segments can be characterized in different ways, based on their duration, trend, mean values or intra-annual characteristics.

Breakpoint detection—theoretical background

In order to detect abrupt changes in the remote sensing time-series, we use a breakpoint detection method derived from the econometrics literature. The method developed to detect breakpoints in time-series regression models was initially introduced by Bai (1994), and later extended to the detection of multiple breakpoints (Bai, 1997a,b; Bai & Perron, 1998). Examples of applications are given by Zeileis et al. (2003), and Bai & Perron (2003). Given a regression model (Equation 4.1), the method tests the hypothesis that the regression coefficients remain constant over time against the alternative hypothesis that at least one of the coefficients changes. To test this hypothesis, a triangular Residual

Sum of Squares (RSS) matrix, which gives the RSS for each possible temporal segment in the time-series, is computed. The optimal number of partitions is then obtained by minimizing the Bayesian Information Criterion (BIC), while the position of the breakpoints is determined by minimizing the RSS among all possible partitioning schemes given by the RSS matrix (Zeileis et al., 2003; Bai & Perron, 2003). The un-partitioned time-series regression (Equation 4.1) can therefore be re-written for each potential partitioning following Equation 4.2.

The initial time-series regression model can be written as follows:

$$y_i = x_i^\top \beta_i + u_i \quad (i = 1, \dots, n), \quad (4.1)$$

where y_i is the observation at time i of the dependent variable, x_i a vector of regressors, and β_i the vector of regression coefficients, and u_i the error term. After partitioning, Equation 4.1 can be re-written as:

$$y_i = x_i^\top \beta_j + u_i \quad (i = i_{j-1} + 1, \dots, i_j, j = 1, \dots, m + 1), \quad (4.2)$$

where j is the segment index for a time-series regression containing $m + 1$ segments within which the regression coefficients (β_j) are constant. By convention $i_0 = 0$ and $i_{m+1} = n$ (Zeileis et al., 2003).

Here, because the method is being applied to vegetation, which is known for having seasonality due to its phenological cycles, we chose to allow the use of a seasonal-trend regression model (Verbesselt et al., 2010b,b; Roerink et al., 2003). The model we use can be written as follow:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \epsilon_t \quad (4.3)$$

where the dependent variable y at a given time t is expressed as the sum of an intercept α_1 , a slope α_2 for potential temporal trend in the data, a sum of different frequency harmonic components representing seasonality ($\sum_{j=1}^k \gamma_j \sin(\frac{2\pi j t}{f} + \delta_j)$), and an error term ϵ_t . For the harmonic component of the model, $j = 1$ corresponds to the one year cycle, k is the chosen harmonic order, γ_j and δ_j correspond respectively to the amplitude and phase of the harmonic order j , and f is the known frequency of the time-series (i.e., number of observations per year).

Breakpoint detection—implementation

An important parameter for the implementation of the breakpoint detection algorithm is the minimum segment size relative to the sample size, referred to as h . In that case sample size refers to the number of observation in the given time-series. While a large h value may increase the stability of the model, it may result in short segments being missed, given that the cultivation period is only two years in average. On the other hand, a too small value for h may result in the detection of spurious breakpoints due to the noise that might be contained in the data.

Based on recommendations from Bai & Perron (2003), and empirical experiments we chose a h value of 0.07, which corresponds to approximately 24 samples in our case. The regression model used is another important parameter of the implementation. Although we designed the method to work with a seasonal trend model, we decided based on a priori knowledge of the system and visual investigation of randomly chosen pixel time-series to use a simple trend model without seasonal component, meaning that we set k to 0 in Equation 4.3. This decision is justified by the presence of essentially evergreen tree species in the area.

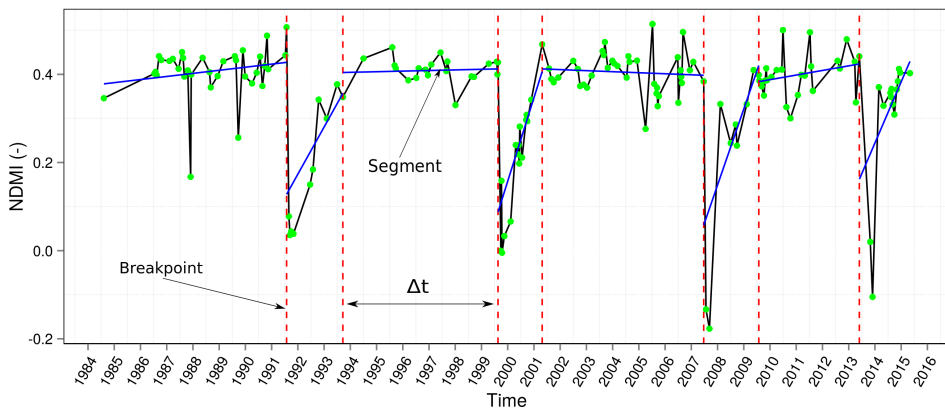


Figure 4.1: Illustration of a temporal Normalized Difference Moisture Index (NDMI) profile with breakpoints detected. Blue lines correspond to the fitted regression model for each segment while green dots represent the observations.

4.2.2 Breakpoint classification

In the present case study we are interested in quantifying land use intensity in a slash and burn agriculture system. The breakpoint detection algorithm introduced in Section 4.2.1 returns all breakpoints regardless of the change process associated to them. We therefore

need to characterize the segmented profiles in order to retrieve the variables of interest. Following Jakovac et al. (2015), we define land-use intensity as the number of cultivation cycles that the land has undergone in the past. The objective of this characterization therefore becomes to identify and count the number of cultivation cycles from the segmented temporal profiles. After investigating a set of segmented time-series, it appeared that the typical segmented profiles would present a succession of burning and stabilization breakpoints (Figure 4.1). Burning breakpoints correspond to the slashing and burning events in the swidden agricultural system and can be identified on the temporal profile by a dramatic drop in Normalized Difference Moisture Index (NDMI, see description in Section 4.2.4) (1st, 3rd, 5th and 7th breakpoints in Figure 4.1). The stabilization breakpoints occur generally two years after burning events when NDMI stabilizes to its saturation level following a gradual increase period (2nd, 4th and 6th breakpoints in Figure 4.1). While this appears to be the transition from a regrowing forest to a stable forest, we know that two years is not a sufficient period for forest regrowth and the NDMI stabilization reflects more a limitation of the index to discriminate the later stages of forest regrowth than a real ecological transition. Because we were interested in the number of cultivation cycles, we decided to use the number of breakpoints related to the burning process as a the proxy for the number of cycles. We classified the breakpoints for being able to count for each segmented time-series the total number breakpoints associated to the burning process.

While it is relatively easy for a human to assign classes to the breakpoints by jointly considering several elements of the temporal context, it would be very difficult to programmatically define a set of rules and thresholds that would assign the right class to the breakpoints. We therefore decided to use a Random Forest classifier in order to automatically classify breakpoint classes based on the characteristics of the surrounding segments of each breakpoint (slope, mean, begin and end NDMI values). Random Forest is an ensemble learning method, developed by Breiman (2001) and widely used in many scientific fields, including the field of remote sensing (Pal, 2005; Avitabile et al., 2012). It can be used for both classification and regression problems and works by building multiple regression or classification trees (Breiman et al., 1984) from bootstrapped training samples. The output of these trees are then aggregated. In the present case, by learning from the patterns of the surrounding segments, the classifier can assign a class to all the breakpoints of every individual time-series. Random Forest requires a training dataset in order to be able to perform supervised classification. We assembled this training dataset by visually interpreting a set of randomly selected partitioned NDMI profiles and manually assigning classes to all breakpoints. In total we considered three classes; the *burning* and *stabilization* breakpoints, as illustrated in Figure 4.1 as well as an additional *undefined* class, for breakpoints from which the process could not be clearly identified from the temporal patterns.

4.2.3 Study area

We developed and tested the method in an area of Brazil, state of Amazonas, where swidden agriculture is used for the cultivation of manioc. The area is covered by two municipalities (Tefé and Alvarães) and comprises 22 villages (Figure 4.2). The area is relatively flat with an elevation comprised between 25 and 80 m above sea level. The river (*Rio Tefé*—a tributary of the Amazon river) is central to the organization, water ways being the main transportation network in the area, and all villages have been established along the rivers. The agricultural fields stretch out of the villages from the immediate proximity to a few kilometres inland and have an average size of 0.5–1 ha. Old-growth forests represent the remaining part of the landscape. Villagers cultivate the land essentially for manioc and rely on a rotation system that consists in slashing and burning the forest to prepare the land for a cropping period of two years, after which the forest is left to regrow until the next rotation (Jakovac et al., 2015). The time between two rotations vary in most cases between two and seven years (Jakovac et al., 2015). Information on the land history and use intensity were collected in the area by means of farmers interviews, conducted in 2012 and 2013. A complete description of the field data set used in this study can be found in Jakovac et al. (2015).

4.2.4 Data sets and data processing

Landsat data

The present study uses data collected by the Landsat sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI). These data have a ground resolution of 30 m and span the 30 years period of 1984–2015. Early steps of the pre-processing such as geo-referencing, atmospheric correction as well as cloud and cloud shadows detection are directly performed by the United States Geological Survey (USGS), from where the data were downloaded. The USGS uses the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006) to perform automatic atmospheric correction of the Landsat archive, while the *fmask* algorithm (Zhu & Woodcock, 2012) is used for cloud detection. LEDAPS is a highly standardized pre-processing framework which ensures the comparability of reflectance values among the different sensors on-board Landsat satellites (Masek et al., 2006), while *fmask* is an advanced object based cloud and cloud shadow detection algorithm that takes advantage of both surface reflectance and temperature (Zhu & Woodcock, 2012). Only terrain corrected data (L1T), which corresponds to the highest level of spatial accuracy, were used. In total this resulted in 350 Landsat scenes covering the study area. Scenes with cloud cover were kept but pixels contaminated by clouds were excluded by using the cloud mask provided by the *fmask* algorithm. By not excluding scenes partially contaminated by clouds, we take advantage of all observation available, making the use

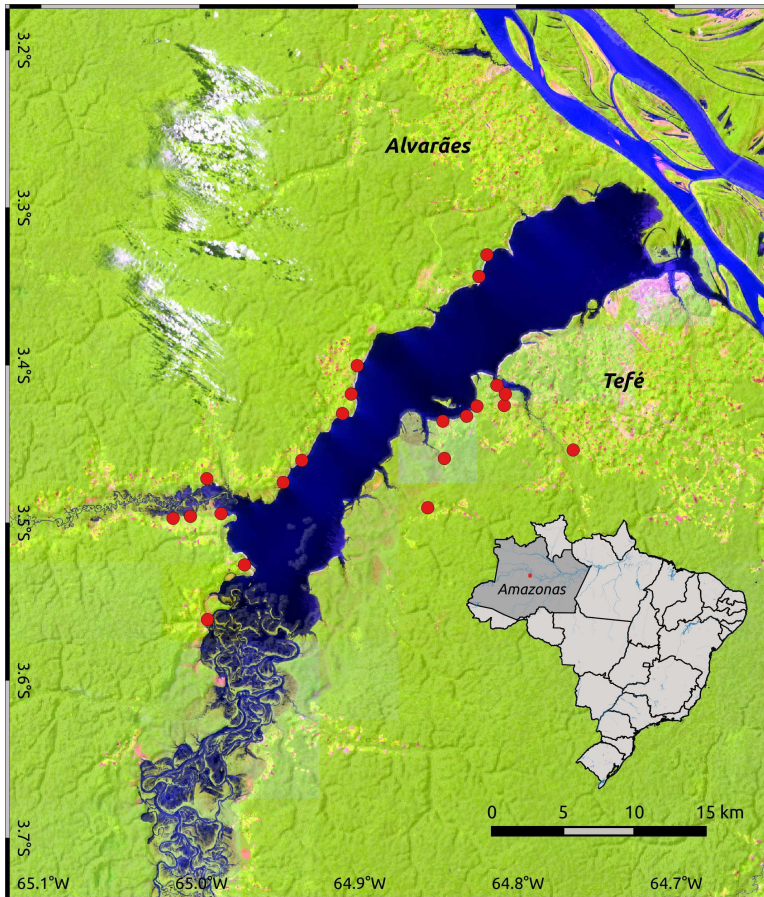


Figure 4.2: Map of the study area, red dots correspond to the location of the different villages. An Operational Land Imager (OLI) image (band combination 6/5/2) acquired on September 29th 2014 is used as background

of methods requiring entire time-series possible. This is analogue to many other recently developed approaches (DeVries et al., 2015b,a; Dutrieux et al., 2015; Zhu et al., 2012; Zhu & Woodcock, 2014), that use all data available, in order to perform computation at the pixel level. These methods are often referred to as multi-temporal, or hyper-temporal approaches (Brooks et al., 2014).

Although the Landsat satellites have a 16 days revisit period (8 days on a two satellites configuration), the temporal resolution of useful data is not constant over the entire period investigated. Several reasons can explain this difference between satellite revisit period and temporal resolution of useful data. First, sensors are not systematically acquiring the

data; this is particularly pronounced over the 90s, period during which little data were acquired outside of the United States (Goward et al., 2006). Even when data are acquired by the sensor and successfully transferred to a receiving station, they can be contaminated by clouds or affected by sensor malfunctioning. An example of sensor malfunctioning is the ETM+ Scan Line Corrector failure in 2003, which has resulted in systematic data loss ever since (Goward et al., 2006).

Vegetation index

While Landsat data, with six spectral bands, contain a lot of information for different regions of the light spectrum, the breakpoint detection algorithm we are using has not been designed to work with multiple time-series. We are consequently required to reduce these six surface reflectance time-series into a single vegetation index time-series. Vegetation indices often emphasize a certain trait at the expenses of another potentially useful trait, so that choosing a vegetation index often comes as a trade-off. Many indices can be computed from Landsat surface reflectance data. Here we are mostly interested in indices relating to the properties of the vegetation cover such as greenness or photosynthetic activity. The most notoriously used vegetation index is undoubtedly the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979). NDVI has the advantages of being robust and easily interpretable (Jackson & Huete, 1991), but it also tend to saturate at high biomass levels (Huete et al., 2002). A number of alternatives have been proposed to the NDVI, including the Enhanced Vegetation Index (EVI) (Huete et al., 2002), however, this latter was also found to being largely affected by varying sun-sensor geometry characteristics, resulting in potential artefacts in the data when used in time-series (Morton et al., 2014; Brede et al., 2015). Another alternative to NDVI originates from the family of indices which consider both Short Wave Infra-Red reflectance (SWIR) and Near Infra-Red (NIR) such as the Normalized Difference Moisture Index (NDMI) (Vogelmann & Rock, 1988), or other metrics derived from the Tasseled Cap transformation (Crist, 1985). By taking advantage of the NIR and SWIR spectral regions which are respectively sensitive to photosynthetic activity and moisture, these indices have been shown to be particularly appropriate for discriminating among forest age classes (Horler & Ahern, 1986; Jin & Sader, 2005; Wilson & Sader, 2002). Here we are interested in capturing the process of forest recovery after land abandonment, we therefore decided to use the NDMI (Vogelmann & Rock, 1988). The index is calculated using the equation below (Equation 4.4).

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} \quad (4.4)$$

where *NIR* corresponds to the Near Infra-Red (770–900 nm) part of the light spectrum and *SWIR* to the Short-Wave Infra-Red (1550–1750 nm). *NIR* and *SWIR* correspond

respectively to the bands 4 and 5 of the TM and ETM+ sensors and to the bands 5 and 6 of the OLI sensor.

Because it is a simple normalized ratio, NDMI is easy to compute and does not depend on other reference pixels or end members, and varies by design between -1 and 1. Low NDMI values are expected for bare soils and thin regrowing forest canopies while higher values correspond to thicker, fully developed forest canopies (Wilson & Sader, 2002).

4.2.5 Spatial implementation

In our approach we analyze the time-series profiles at the object level. Objects are clusters of pixels grouped according to their resemblance and spatial proximity. Working with objects rather than pixels has two advantages for our purpose of analysing land dynamics in the swidden agriculture mosaic landscape. First, when working with objects features being mapped correspond to elements of the landscape, which results in easier interpretation of the outputs of the analysis (Blaschke, 2010). Secondly, creating the object layer simultaneously allows for a pre-selection of the areas of interest and tremendously reduces the number of time-series that need to be processed, hence reducing computing time and efforts. A singular characteristic of the swidden agriculture land use is the constant shift in fields footprints. While farmers often return to the same location when re-opening a piece of land for a new cultivation cycle, they may decide to only re-open part of the previous field (Metzger, 2003). Considering such characteristic of the system, a simple multi-spectral image segmentation would be unsuitable to delineate the objects of interest as different land use histories could be present within the same object. Here we use a multi-temporal spatial segmentation, so that instead of being defined according to spectral resemblance, the segments cluster pixels based on their spectral-temporal similarity. The resulting segmentation hence delineates objects having undergone similar land use temporal trajectories. This approach is adapted from a method initially proposed by Desclée et al. (2006), further improved by Verhegghen et al. (2010) and used for object based deforestation mapping in central Africa by Duveiller et al. (2008) and Ernst et al. (2013).

We used a mean shift segmentation algorithm, as implemented in the Orfeo ToolBox (www.orfeo-toolbox.org) (version 4.2.0) to perform the multi-temporal segmentation (Fukunaga & Hostetler, 1975; Inglada & Christophe, 2009). A NDMI time-series, which we assembled by building annual mean value composites resulting in one average NDMI value per year, was used as input data. Since spatial segmentation can only be performed on gap free data, we visually screened the annual composites for remaining clouds or gaps and discarded non gap free layers. The resulting NDMI stack used as input for the segmentation contains 25 layers spanning the 1986-2014 period. The mean shift segmentation algorithm requires as input parameters a range radius, a spatial radius and

a minimum object size. We chose for the range radius, which is the similarity (Euclidean distance) measure for a pair of pixel profiles, a value of 0.14. Such value exceeds the natural variability between two similar pixels while successfully differentiating two pixels with different land use trajectories. A too small range radius value has the effect of over-segmenting the area, while larger values have the opposite effect. Based on a-priori knowledge of the system, we set the spatial radius to three pixels and the minimum object size to three pixels (Jakovac et al., 2015).

This initial segmentation step results in many segments, some of which are not swidden agriculture. Since we are only interested in the swidden agriculture fields, we applied a set of rules to filter and keep only those polygons that delineate areas of swidden agriculture. First, we discarded objects larger than 15 ha, which is larger than the average field size of 1 ha, usually found in the area (Jakovac et al., 2015). Following that we used the NDMI time-series associated to the segments in order to filter out remaining segments of stable forest (NDMI never falling below 0.3), and urban areas, permanent agriculture and wetlands (NDMI never exceeding 0.3). A graphical representation of the overall approach is presented in Figure 4.3.

4.2.6 Validation

In order to assess the performances of the method in quantifying land use intensity, we complemented this analysis by a validation exercise. We compared the number of cultivation cycles predicted by the proposed method with a reference dataset providing information on past land uses (Jakovac et al., 2015). The reference dataset was collected in 2012 and 2013 by means of farmer's interviews and contains the number of times the individual fields have been used for manioc cultivation. In total 35 data points, each corresponding to a single field were available for validation. We used the Normalized Root Mean Square Error (*NRMSE*), Normalized Mean Absolute Error (*NMAE*) and *Bias* to assess model performances. *NRMSE* and *NMAE* inform on the relative amount of residual variance, which is the variance in the observed variable that is not explained by the model (Mayer & Butler, 1993). *Bias* indicates whether the model predictions over or under-estimates the variable when predicting new values. The performance metrics can be computed using the following equations:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \quad (4.5)$$

Where $y_{max} - y_{min}$ is the range of the observed variable and *RMSE* the Root Mean Square Error expressed by,

$$RMSE = \sqrt{\frac{\sum (\hat{y}_i - y_i)^2}{n}} \quad (4.6)$$

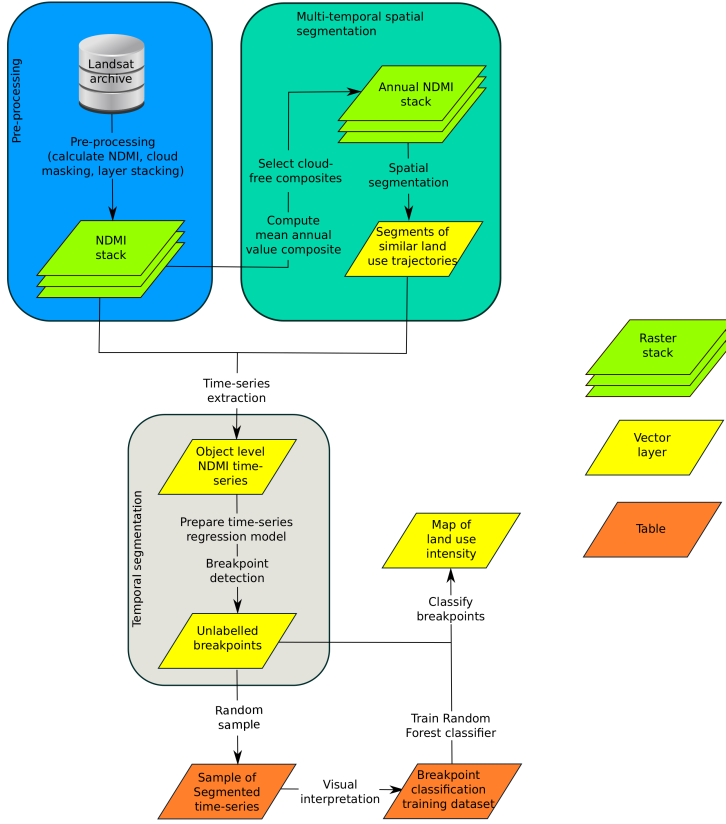


Figure 4.3: Schematic overview of the proposed methodology

In the above equation, y_i is the i^{th} observation of the observed variables, while \hat{y}_i is the corresponding predicted value for that observation. n is the total number of observations used for validation.

Similarly to $NRMSE$, $NMAE$ is a normalized version of the Mean Absolute Error (MAE).

$$NMAE = \frac{MAE}{y_{max} - y_{min}} \quad (4.7)$$

Where MAE is expressed by,

$$MAE = \frac{\sum |\hat{y}_i - y_i|}{n} \quad (4.8)$$

By design, $NMAE$ will always be smaller than or equal to $NRMSE$ (Mayer & Butler, 1993).

Finally $Bias$ is simply the difference between the predicted mean \hat{y} and the observed mean \bar{y} .

$$Bias = \hat{y} - \bar{y} \quad (4.9)$$

4.3 Results and discussion

4.3.1 Multi-temporal segmentation

Figure 4.4 presents the resulting multi-temporal segmentation applied to a subset of the study area. The transparent polygons are overlaid with a colour composite of NDMI at three different years. Although not representing the whole time-span considered, the resulting colour composite reflects some variability in land use trajectories. One can see in Figure 4.4 polygons appropriately separating groups of pixels with different colours. Delineations are particularly clear at the periphery of the cultivation area, where fields also tend to be larger. These fields at the periphery likely correspond to recently acquired land, opened and cultivated for the first time. This characteristic makes their spatio-temporal signature relatively simple and distinct from the surrounding forest matrix and easily delineated by the segmentation algorithm. Segments are smaller in the core of the cultivation area. These correspond to areas that have been cultivated multiple times, with fields footprints potentially shifting over time, making the spatio-temporal dynamics more complex and resulting in smaller, more numerous objects. In total 65,000 segments, corresponding to 50,000 ha (11% of the study area shown in Figure 4.2), were delineated. The average segment size is 0.77 ha, which is comprised in the field size range (0.5–1 ha) reported by Jakovac et al. (2015).

4.3.2 Quantifying land use intensity

We applied the developed approach to a swidden agriculture area, aiming at predicting the number of cultivation cycles each land parcels had gone through. As described in Section 4.2, retrieving the number of cultivation cycles from the time-series involves several steps. The first step, which is the core of the method presented here, and also the most computationally demanding step, is the temporal segmentation using the breakpoint detection algorithm. We ran the algorithm on 65,000 time-series corresponding to the Landsat spatially aggregated values for all the swidden agriculture polygons identified in the multi-temporal segmentation. It took approximatively 25 CPU hours to run the

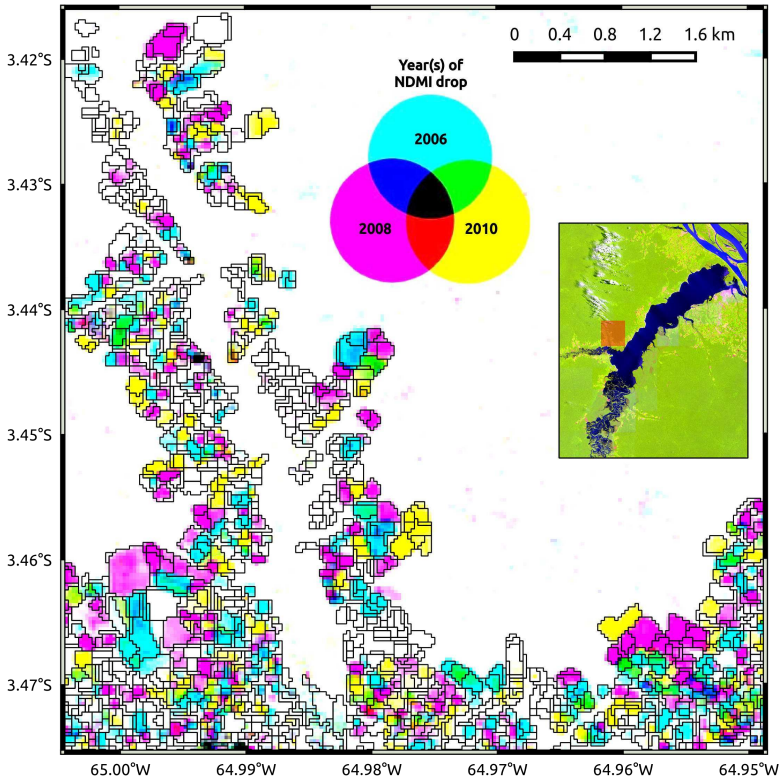


Figure 4.4: Example of segmentation results for an area bordering a secondary river. Segments with transparent filling are overlaid with a RGB composite with red, green and blue channels corresponding to NDMI in years 2006, 2008 and 2010 respectively. White areas correspond to stable forest with little to no variation in NDMI among the three years, and different colours reflect various NDMI temporal trajectories.

temporal segmentation algorithm on all the time-series for the 1984–2015 period, resulting in an average processing time of 1.4 s per polygon. For the second step, which consists in classifying the unlabelled breakpoints returned by the temporal segmentation algorithm, we visually interpreted 100 randomly selected temporal profiles and manually classified 700 breakpoints into the three classes; *Burning break*, *Stabilization break* and *Undefined*. This training dataset was then used for training the Random Forest classifier. From the Out Of Bag (OOB) estimates of error, the *Burning Breakpoint* class, which is the most important class to identify for later retrieving the number of cultivation cycles, had a prediction accuracy of 88%. Following the breakpoint classification, we were able to produce the land use intensity map presented in Figure 4.5, which presents the predicted number of cultivation cycles over part of the study area. We predicted that, over the

study area presented in Figure 4.2, the area with at least one cultivation cycle was about 36,000 ha. A first observation of the map reveals that cultivated areas occur within a 2 km zone surrounding waterways. The mapped variable also appears to be structured in space, with most fields with only one cultivation cycle located in the outer bound of the cultivated area and more intensively used land parcels closer to villages. Such patterns were expected since agricultural expansion grows organically away from the villages or main transportation networks (Jakovac et al., 2016b).

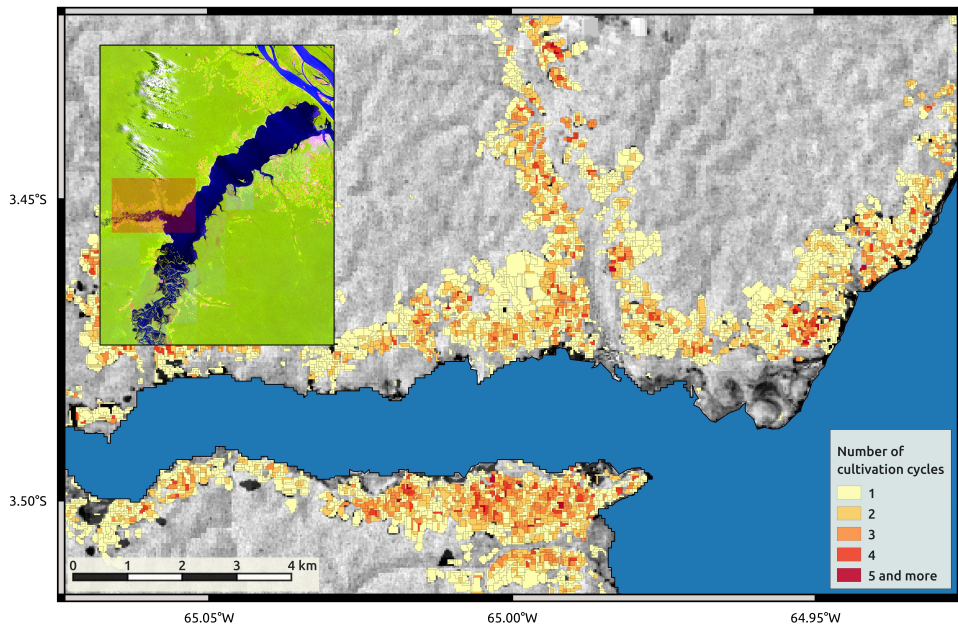


Figure 4.5: Map depicting the predicted number of cultivation cycles for a subset of the study area. NDMI from Landsat 8 acquired on September 29th, 2014 is used as background layer.

4.3.3 Method performance assessment

We assessed the performance of the method by confronting the method's predictions to an observed dataset and computed performance metrics. This ground truth dataset is based on 35 farmer's interviews who reported the number of times their land has been cultivated in the past. The resulting method performance assessment is represented in Figure 4.6. The NRMSE and NMAE inform on the relative amount of residual variance unexplained by the model. With values of 0.25 and 0.17 for NRMSE and NMAE respectively, we are

confident that the method can provide reliable information in this context, and that the number of cultivation cycles predicted reflects the actual management intensity of the land. However, it is clear from the validation that certain time-series did not produce the expected number of cultivation cycles and deviated from the information reported by farmers. From the spread of the validation samples in Figure 4.6, and confirmed by the *Bias* of -0.23 cycles, there is no strong systematic pattern of the method to either overestimate or underestimate the number of cultivation cycles. In order to further explain these deviations from the reference number of cultivation cycles, we examined three time-series profiles from the validation data set (Figure 4.7).

The time-series profiles, together with their associated model predictions and reference values are presented in Figure 4.7. Figure 4.7a, represents a time-series for which the number of cultivation cycles predicted by the method is much smaller than the reference number of cycles, with two and four cultivation cycles for the prediction and reference respectively. One striking characteristic of this profile, when compared to a typical NDMI profile as shown for instance in Figure 4.1, is the large variability, or noise in the NDMI values. With such high variability in the time-series, the optimal partitioning found by the breakpoint detection algorithm consists of a single partition for the 1984–2007 period. One hypothesis for explaining such high variability is the presence of several land use trajectories within the object of interest. Such sub-object mixture could originate from a limitation of the multi-temporal segmentation to appropriately discriminate land use trajectories between various individual pixels temporal profiles, resulting in spatial segments composed of several land use histories. However, considering the small size of agricultural fields (0.5–1 ha), which can be just a few pixels in the case of 30 m Landsat data, such mixture is likely to occur and reflects more a physical boundary of the method related the spatial resolution of the input data rather than a shortcoming of the multi-temporal segmentation. Also, such mixture is likely to occur more frequently in the core of the area, where objects are small and spatio-temporal patterns more complex. Figure 4.7b is an example of temporal profile for which the prediction over-estimates the actual number of cultivation cycles. While breakpoints are detected in the NDMI time-series, transition patterns between consecutive temporal segments are unclear and the underlying transition processes cannot easily be inferred. These unclear patterns result in miss-classifications by the Random Forest classifier in charge of labelling the breakpoints. The rapid dynamics of the system of interest are forcing us to set a small minimum segment size for the breakpoint detection step of the method, hence increasing the sensitivity to noise and the possibility of detecting breakpoints which are not related to actual change processes. This source of error, and limitation, is very specific to the current application of the method and we are confident that the robustness of the method would increase in environments with slower dynamics, such as temperate forests for instance. The last source of deviation from the reference dataset we identified, presented in Figure 4.7c, appears to originate from inaccuracies in the reference dataset itself. While three cultivation cycles were reported

for that particular field, the visual interpretation of the NDMI profile, as well as the method's prediction, suggests four cycles. When interpreting the method's performance, it is important to consider the origin of the reference dataset, and the corresponding accuracy. Although this dataset is undoubtedly the most reliable reference information available, it was collected from interviews and therefore relies on individuals accurately remembering the past. Considering that fact, inaccuracies in the reference dataset are likely to occur, particularly regarding events that occurred more than 15–20 years ago. This however, can only be acknowledged and is hardly quantifiable.

In addition to the three examples presented above, other limitations of the current approach can be discussed, as they are likely to occur. First of all, the length of the Landsat observations period is obviously limiting the retrieval of dynamics that occurred before the 80s. Landsat is among the satellites that have the longest data record and we will not be able to overcome this limitation. While for the present application retrieving land use history of the past three decades is sufficient to infer the pressure of the intensification on the secondary forest's capacity to regenerate (Jakovac et al., 2016b), other mapping activities may require longer land use history data in order to retrieve useful information. In addition to the limitations of the dataset, other limitations inherent to the method also exist. An important parameter of the method is the minimum segment size (h), defined as a proportion of the total number of observations available in the time-series. This parameter of the breakpoint detection algorithm further limits the temporal extent of the period during which potential breakpoints can be detected. This period which can be referred to as the effective monitoring period stretches between 1988 and 2014 in our case (represented by the interval between the two red dashed lines in Figure 4.8). Additionally, given that data density varies with time, the sensitivity to detecting rapid dynamics evolves alongside and can be largely different between two time periods. As shown in Figure 4.8, which represents the minimum segment duration assuming all Landsat acquisitions are clear observations, the mid 90s particularly suffers from a data shortage, and two consecutive breakpoints during that period need to be at least four years apart when using a minimum segment size of 0.07. Although it is method specific, this limitation also needs to be considered relatively to the input data as denser time-series and time-series with reduced noise will help more robust temporal segmentations. Finally, we acknowledge that working with a single vegetation index might be a limitation to properly capture regrowing forest dynamics. In all cases presented in Figure 4.7 we can see that the base NDMI level corresponding to a fully grown forest is reached relatively shortly after the burning breakpoint. Choosing a single vegetation index constitutes a trade-off which may restrict the time span during which we are able to detect the regrowing phase of the forest. Developing change detection approaches that take advantage of all the spectral information available should therefore constitute a priority for future work.

Many of the limitations listed above are directly or indirectly related to temporal density of the data. This holds great promises for an undoubtedly data richer future, thanks to

the launch of the Sentinel 2 constellation and an increased data acquisition capacity of the recent and upcoming Landsat sensors (Drusch et al., 2012; Roy et al., 2014). Other regions of the world with greater data availability than the Brazilian Amazon may also hold great promises for the method to retrieve information about past land uses. Testing and validating the method we propose in different ecosystems, including environments outside of the tropical zone with various seasonality patterns, will be part of our future studies.

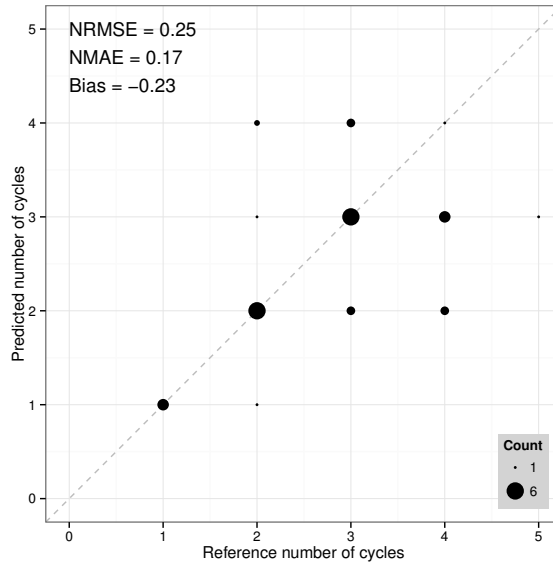


Figure 4.6: Scatter plot of predicted number of cultivation cycles against information reported by farmers.

4.3.4 General considerations

We developed and tested an approach for reconstructing land use history from Landsat time-series. The method uses all data available in a pixel time-series regardless of the cloud cover present in the rest of the scene to partition a vegetation index time-series into segments and abrupt changes. The breakpoint detection algorithm that finds abrupt changes in time-series regression models (Bai & Perron, 1998; Zeileis et al., 2003), is the core of the method. We tested the method capacity to retrieve land use intensity, defined as the number of cultivation cycles, in a swidden agriculture context. In that specific context the implementation of the method involved an additional multi-temporal spatial segmentation step as well as a post-partitioning breakpoint classification. In its present application the method was able to predict the number of cultivation cycles with

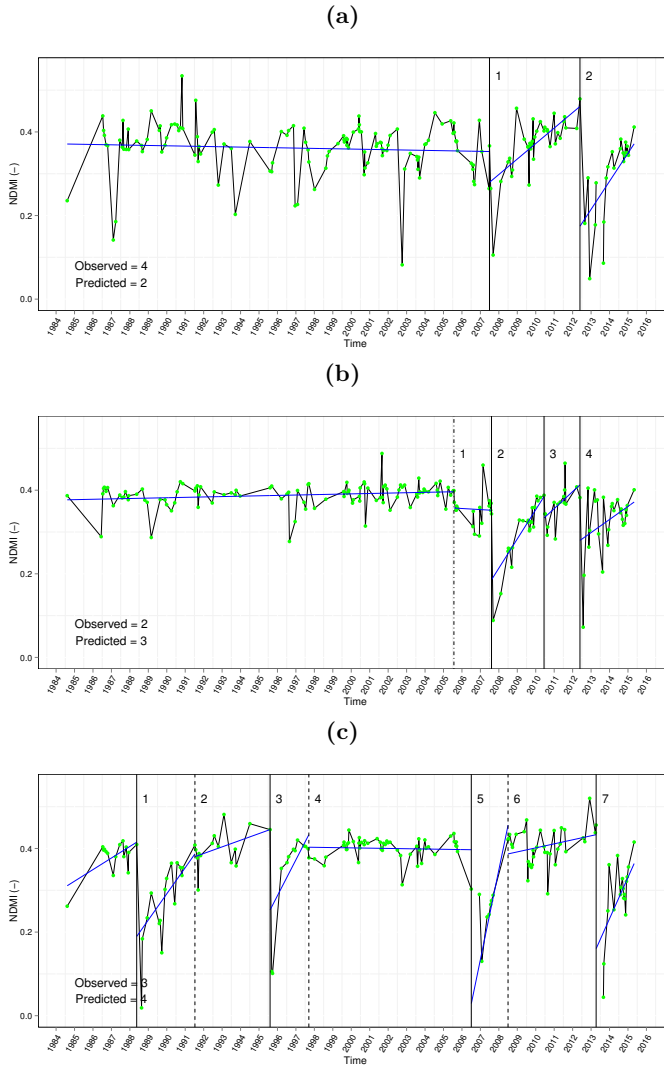


Figure 4.7: Illustration of some sources of variability in prediction performances from three temporal profiles (a, b and c) for which the number of predicted cultivation cycles diverges from the reference number of cycles. *Burning*, *stabilization* and *undefined* breakpoints are represented by solid, dashed and dot-dashed vertical lines respectively. (a) corresponds to a profile with high variability in the NDMI signal, resulting in undetected breakpoints. (b) is an example of a temporal profile for which the method overestimates the number of cultivation cycles due to classification errors. (c) illustrates a case of potential inaccuracy in the reference dataset, where visual interpretation of the NDMI profile suggests four cultivation cycles while only three were reported.

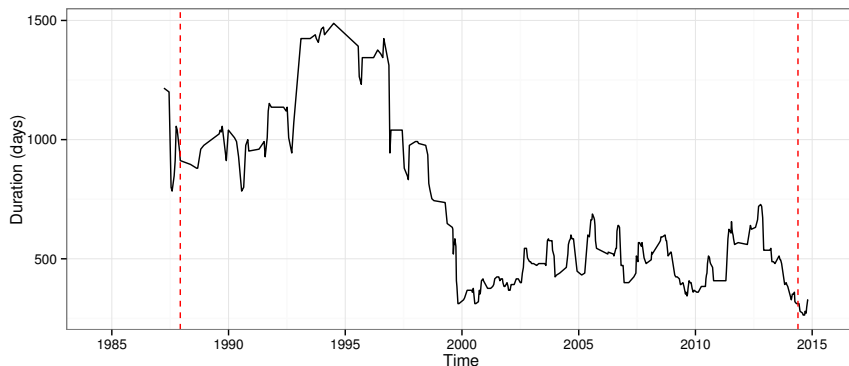


Figure 4.8: Minimum segment duration based on the density of data available and a minimum segment size h of 0.07. The horizontal scale corresponds to the date at the centre of the segment and the dashed red lines represent the earliest and latest possible breakpoint detection dates.

a NRMSE of 0.25 and a NMAE of 0.17 (Figure 4.6). Further exploitation of the results with an ecological perspective are presented in Jakovac et al. (2016b), where the method was successful in evaluating changes in land use dynamics through time. The method can potentially retrieve land use history in various environments by detecting changes in trajectories including potential changes in phenological patterns over time. From an application point of view, since they both have been demonstrated with Landsat time-series and provide similar types of outputs, the method we propose can be seen as analogue to LandTrendR (Kennedy et al., 2010). The method we propose brings one difference with respect to LandTrendR, which is the use of all data available in the time-series. As we demonstrated in the present article, this characteristic allows tracking of rapid vegetation dynamics, which often is a feature of tropical environments (Asner et al., 2004a,b). Since it was introduced in 2010, LandTrendR has achieved a certain legacy and the present method is absolutely not meant to replace it. The use of all data rather than annual composites likely makes the present method more sensitive to noise in the time-series. However, the method we are proposing is particularly relevant when tracking fast dynamics or when change in intra-annual dynamics over time is relevant to consider.

As discussed earlier, the current application of the method revealed some limitations, most of which relate to the temporal density of the input data. Periods of low data availability are likely to result in missing breakpoints when attempting to detect fast dynamics. The 90s are notorious over many regions of the world for having poor data availability (Goward et al., 2006). Potential users should be aware of this limitation and temporal resolution of the input data should be taken into consideration when interpreting results.

At the moment we were only able to demonstrate part of the potential of the method by

applying it to an area of swidden agriculture for the quantification of land use intensity. Such variable is of great interest in that particular agricultural system since it largely influences the capacity of the secondary forest to regenerate following land abandonment (Jakovac et al., 2015). Although we developed a proof of concept for swidden cultivation systems applied to manioc, and we acknowledge that further testing is required due to the use of a single case study, this method will likely be useful for retrieving land use history in other contexts and can be seen as a generic land use history reconstruction method. It could for instance be used—considering minor parameters adjustments—for other rotational systems such as banana and pineapple plantations, which are also common in the Amazon, and therefore comprise a similar sequence of cutting break, cultivation and regrowing forest. Future research will look into further testing the method in other environments and for different applications, such as carbon modelling and secondary forest recovery. Such application would undoubtedly require a more thorough characterization of the segmented temporal profiles than was required for the present case study. Improvements to the current application will also be considered in further research, such as multi-spectral approaches as a way to overcome trade-off and limitations related to choosing a single vegetation index.

4.4 Conclusion

With this paper we propose a method capable of reconstructing past land use history by using Landsat time-series. Individual time-series are partitioned into segments corresponding to land use regimes, and breakpoints, which can be thought as abrupt changes between regimes. Although the present case study did not require its use, because it detects changes in time-series regression models, the method can be used in combination with a seasonal model, hence potentially detecting changes over time in phenological patterns. The method was tested in the context of swidden agriculture for a chosen study area in Brazil, with the aim of quantifying the number of cultivation cycles fields had undergone in the past three decades. Such land use intensity metric is relevant in swidden cultivation systems since it largely affects the capacity of the forest to regenerate following land abandonment. For this case study specifically we applied the method at the object level rather than for individual pixels, using objects from a multi-temporal spatial segmentation step. A simple characterization of the segmented temporal profiles was applied by classifying breakpoints and counting the number of burning breakpoints. We validated the resulting land use intensity against a set of reference data collected from farmer's interviews and obtained accuracies of 0.25 and 0.17 of NRMSE and NMAE respectively. While the method requires further testing for different areas and in various environments for its potential to be fully confirmed, we already demonstrated a use case for which it was able to provide useful information to ecologists and we are confident that the method holds greater potential for retrieving land use history in different environments of the

planet.

Acknowledgements

The present research was conducted and can be reproduced using free and open source software only, we wish to thank the open source community for their time and efforts spent on making in a transparent way reliable tools accessible to all. The research received partial funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement Number 283093 - The Role Of Biodiversity In climate change mitigation (ROBIN). C.C.J. received a scholarship from The Netherlands Fellowship Program (NFP-NUFFIC).

Chapter 5

Post-disturbance recovery in forest spectral properties across the Amazon

This chapter is based on:

Dutrieux, L. P., Kooistra, L., Herold, M. & Poorter, L. (2016). Post-disturbance recovery in forest spectral properties across the Amazon. In prep.

Abstract

Secondary forests constitute a large part of today's existing forests and play a major role in climate regulation; yet little is known on how forest resilience varies spatially or what drives the recovery of previously disturbed forests. The present study is twofold; we first propose a framework able to quantify time necessary for a disturbed and abandoned forest to recover its spectral properties. We then investigate spatial patterns and potential drivers of spectral recovery time across the Amazon basin. The framework, applicable to multi-spectral time-series data such as those collected by the sensors on-board the Landsat satellites, is largely derived from the econometrics field and combines change detection, temporal trajectory identification, and structural stability monitoring. We applied it to a set of 3596 Landsat time-series sampled from regrowing forests across the Amazon basin, thus producing estimates of recovery time in spectral properties, which we call spectral resilience. On average, it took 7.8 years, with a large variability ($sd = 5.3$ years) for disturbed forests to recover their spectral properties. We expected spectral resilience to vary across forests and increase with water availability. Instead, we found that spectral resilience does not show any spatial structure and appears to be independent from water availability, as indicated by annual rainfall and climatic water deficit (CWD). The high local variability in spectral recovery suggests that spectral resilience largely depends on local conditions which we could not identify. We did not possess sufficient elements, particularly regarding the relation between spectral resilience and resilience in other forest attributes (biomass, species richness, etc) to explain whether the patterns observed are likely specific to the spectral properties of forest or can be generalized to other forest attributes. Facing this uncertainty we suggest additional research to be conducted, with the aim of better understanding this newly introduced spectral resilience metric.

5.1 Introduction

Secondary forests constitute over half of the area of tropical forests, will increase in proportion in the future and sequester large amounts of carbon (Bongers et al., 2015; Brown et al., 1990; Poorter et al., 2016). They are therefore a potentially important missing piece of the global carbon balance and their potential role for biodiversity conservation is underestimated. Despite their tremendous relevance in the current context of climate change and environmental degradation, large blocks of knowledge are missing concerning them. While recent studies brought great insights on potential drivers of forest resilience (Poorter et al., 2016; Jakovac et al., 2015), much variation remains unexplained. There is therefore a fundamental need to better understand forest recovery, resilience and their underlying drivers.

Here we propose to approach forest resilience from a remote sensing perspective. Relatively long time-series of remotely sensed data can be assembled (Goward et al., 2006), allowing dynamic processes to be observed and studied in a spatially explicit way. The dynamic process we are interested in is the resilience in spectral properties. We define resilience in that context as the time necessary for spectral properties to recover to their initial levels following a sequence of forest disturbance and abandonment, hence adopting the engineering resilience definition proposed by Pimm (1984) (see also Holling (1996)).

Previous studies have used remote sensing to describe global resilience of natural vegetation (Hirota et al., 2011; De Keersmaecker et al., 2015). Both have approached resilience using the ecological definition of the term, which corresponds to the magnitude of disturbance that a system can absorb before switching to a different stable state (Pimm, 1984; Holling, 1996). Hirota et al. (2011) used static data from Moderate Resolution Imaging Spectroradiometer (MODIS) to infer resilience, while De Keersmaecker et al. (2015) have developed a framework using Normalized Difference Vegetation Index (NDVI) time-series, which they then applied to the 8 km spatial resolution Global Inventory Modelling and Mapping Studies (GIMMS) dataset. In contrast we are interested in quantifying resilience of forests following the engineering definition of the term. The engineering resilience of a system is characterized by the speed of recovery in its properties, through regrowth in the case of a forest systems. We decided to address this dynamic process using a time-series approach. Considering the scale at which forest disturbances and regrowth occur, Landsat data, with a spatial resolution of 30 m provide an ideal data source to observe and quantify these processes. Numerous studies have used Landsat time-series to monitor land cover change (Zhu et al., 2012; Brooks et al., 2014; DeVries et al., 2015b; Dutrieux et al., 2015; Huang et al., 2010; Kennedy et al., 2010; Dutrieux et al., 2016; DeVries et al., 2015a; Hansen et al., 2013). The most recent developments of that field resulted in the implementation of a near real time deforestation alert system (Hansen et al., 2016).

The framework we propose is largely derived from the econometrics literature and combines break detection (Bai, 1994, 1997b,a; Bai & Perron, 1998, 2003; Zeileis et al., 2003), trajectories identification, and structural stability monitoring (Leisch et al., 2000; Zeileis et al., 2005; Zeileis, 2005; Zeileis et al., 2010). Most time-series approaches focus on single vegetation indices, but here we use a full multi-spectral time-series approach to take full advantage of the spectral dimension contained in the Landsat data, and capture subtle differences in spectral signatures between secondary and old growth forests canopies.

We first introduce a framework to quantify forest recovery time in spectral properties from multi-spectral remote sensing time-series. We then investigate spatial patterns and drivers of spectral recovery time across the Amazon basin.

5.2 Material and methods

5.2.1 Study area and sampling of forest regrowth pixel

The objective of the study is to characterize the spectral recovery of regrowing forests and how it is influenced by environmental variables. We therefore sampled pixels of regrowing forests across different environments. The study stretches from -70 to -50 degrees longitude, and from -18 to 2 degrees latitude (Figure 5.1). This area covers a large part of the Brazilian Amazon as well as lowland Bolivia, contains a large variety of environments, forest types and climatic conditions, and is globally important for carbon sequestration and biodiversity conservation. Because we are only interested in areas having undergone forest regrowth, we selected pixels based on the global map of forest cover change (Hansen et al., 2013). This product includes a forest cover gain category for the 2000–2014 period and provides information on tree cover percentage and change in tree cover percentage at 30 m resolution, globally. To decrease the risk of sampling pixels which are the result of errors in the forest cover change product, we first identified connected pixels of forest cover increase with an area equal to or greater than 1 ha. Isolated pixels are more likely to be artefacts, while there is greater confidence that larger connected areas are actual areas of forest regrowth. This initial filtering step resulted in a total area of regrowing forest during the 2000–2014 period over the study area of 6,318 km². To keep data volumes reasonably manageable, we then randomly sampled 0.05% of these remaining pixels, resulting in a sampled area of 324 ha (3,596 30 m pixels) (Figure 5.1). Individual pixels time-series of surface reflectance for the 6 spectral bands originating from the TM and ETM+ sensors were extracted using the Google Earth Engine platform. Surface reflectance is processed from sensor radiance values directly on the platform by using the LEDAPS algorithm (Masek et al., 2006). LEDAPS also produces a pixel cloudiness information which we used to discard observations likely to be contaminated by clouds from the time-series.

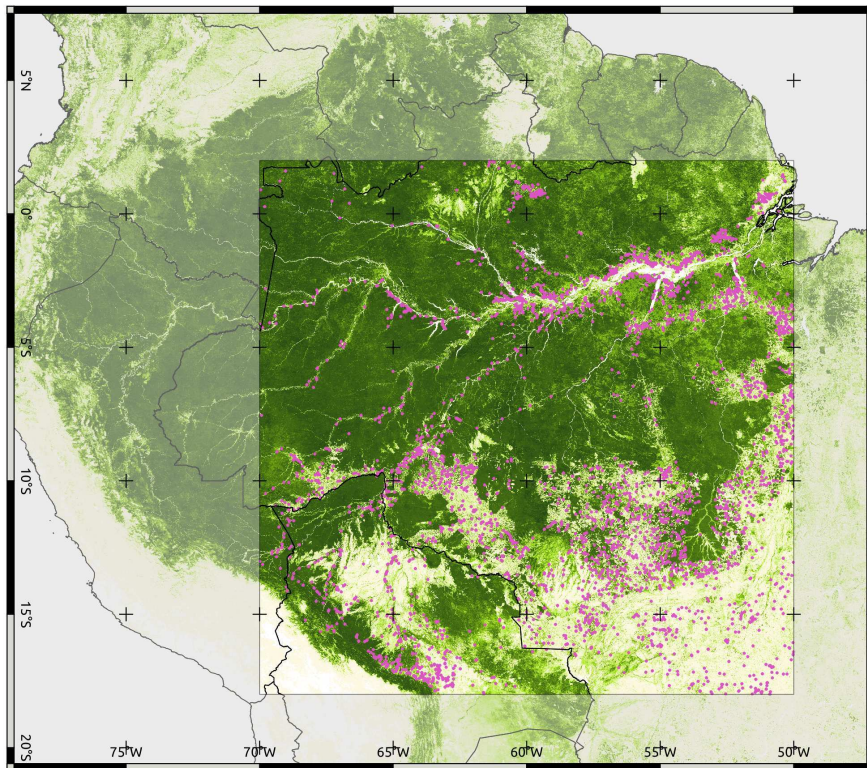


Figure 5.1: Map representing the study area from which Landsat pixels of regrowing forest were sampled. Background colours represent tree cover derived from MODIS (Hansen et al., 2003). Magenta markers indicate the location of sampled pixels.

5.2.2 Quantifying recovery time in spectral properties

We developed a framework combining temporal segmentation, classification of temporal dynamics, and structural stability monitoring to quantify recovery time in spectral properties, and applied it to every sampled time-series. The different steps of the framework consist in: (i) partitioning the time-series into segments of homogeneous temporal dynamics, (ii) classifying each temporal segment into its corresponding land dynamic class (stable forest, regrowing forest, etc), (iii) identifying and retaining time-series that contain a succession of stable forest followed by a period of regrowth, and (iv) monitoring along the regrowing period for structural stability taking the preceding stable period as reference. Details of the framework can be found in Appendix 5.4.

5.2.3 Variables and indices

Choosing a vegetation index for a particular application often comes as a trade-off since different families of indices emphasize different specific traits at the expense of other potentially useful ones. While the framework we developed (Appendix 5.4) contains one step that requires the use of a single vegetation index (for temporal segmentation of the time-series), the actual structural stability monitoring can be performed on any time-series metric. We therefore decided to fully explore the spectral dimension of the input data by working with each spectral region available in TM and ETM+ data since different bands contain different information on vegetation attributes (Table 5.1). From the wide range of vegetation indices available, we chose the Normalized Difference Moisture Index (NDMI) (Equation 5.1) as the single variable required for temporal segmentation. While NDVI is undoubtedly the most popular index, with many advantages such as being easily interpretable and easy to compute (Jackson & Huete, 1991), it suffers from saturation effects for forests with dense canopies as it is the case in many tropical forests. This makes NDVI suitable for certain applications (e.g. detecting abrupt events in tropical forests such as deforestation (DeVries et al., 2015b; Dutrieux et al., 2015)), but limits its use for tracking more subtle processes like forest regeneration. While the Enhanced Vegetation Index (EVI) (Huete et al., 2002) claims to address this saturation limitation, it is also known for being largely affected by varying sun-sensor geometry characteristics, resulting in potential artefacts in the data when used in time-series (Morton et al., 2014; Brede et al., 2015). We chose to use NDMI (Vogelmann & Rock, 1988), which is particularly sensitive to differences in canopy attributes of young secondary forests (Horler & Ahern, 1986; Jin & Sader, 2005; Wilson & Sader, 2002; Dutrieux et al., 2016; DeVries et al., 2015a).

The NDMI can be computed using the following equation:

$$NDMI = \frac{NIR - SWIR1}{NIR + SWIR1} \quad (5.1)$$

where NIR corresponds to the Near Infra-Red (770–900 nm) part of the light spectrum and SWIR to the Short-Wave Infra-Red (1550–1750 nm). NIR and SWIR correspond respectively to the bands 4 and 5 of the TM and ETM+ sensors. Low NDMI values are expected for bare soils and thin regrowing forest canopies while higher values correspond to thicker, fully developed forest canopies (Wilson & Sader, 2002).

Concerning the structural stability monitoring step, fully exploring the spectral dimension of the data implies that each time-series returns a recovery time for each of the six spectral channels (Table 5.1) as well as NDMI, which we kept for comparison, and an additional multi-spectral variable described in the next section.

Table 5.1: Spectral channels available in TM and ETM+ data and the vegetation related attributes they indicate (Horler & Ahern, 1986)

Band name	Spectral region	Wavelength (nm)	Indicator of
B1	Blue	450 – 520	Surface brightness and soil vegetation differentiation
B2	Green	520 – 600	Vegetation greenness and plant vigour/health
B3	Red	630 – 690	Chlorophyll absorption and photosynthetic activity
B4	Near Infra Red (NIR)	770 – 900	Canopy biomass
B5	Short Wave Infra Red (SWIR1)	1550 – 1750	Canopy moisture content
B7	Short Wave Infra Red (SWIR2)	2090 – 2350	Canopy disturbance and fire

5.2.4 Combining all spectral information into one index

The evaluated spectral channels will likely respond differently to the forest recovery process (Vieira et al., 2003). We expect for example NIR to be generally more contributive to differentiating young secondary forests from old growth forests than Blue. These specific spectral channel responses may in turn vary per forest type. In an attempt to combine the potential benefits of each spectral band in a flexible way, we developed a multi-spectral index using an euclidean distance approach. A distance can only be calculated relatively to a point of origin, and several remote sensing studies have developed similar distance based indices using neighbouring pixels or pixels in the image labelled as intact forest for reference (Huang et al., 2008). Because we are working with time-series we could define a dynamic origin for every location independently by using the modelled temporal behaviour of the preceding stable forest segment. For every time-series independently the euclidean distance was computed as follows: for every spectral channel, the temporal extent of the stable forest segment preceding regrowth was used to fit a harmonic model (Equation 5.2)

$$y_t = \alpha_1 + \sum_{j=1}^3 \gamma_j \sin\left(\frac{2\pi jt}{f} + \delta_j\right) + \epsilon_t \quad (5.2)$$

where the dependent variable y at a given time t is expressed as the sum of an intercept α_1 , a sum of different frequency harmonic components representing seasonality ($\sum_{j=1}^3 \gamma_j \sin(\frac{2\pi jt}{f} + \delta_j)$), and an error term ϵ_t . For the harmonic component of the model,

j corresponds to the harmonic order, 1 being the annual cycle, γ_j and δ_j correspond respectively to the amplitude and phase of the harmonic order j , and f is the known frequency of the time-series (i.e., number of observations per year).

New values are then predicted for each spectral band and every observation of the time-series using the respective fitted model (Equation 5.2), allowing to compute an euclidean distance between the predicted and observed reflectance with the following formula:

$$D_t = \sqrt{\sum_{i=1}^k (\hat{y}_{it} - y_{it})^2} \quad (5.3)$$

Where D_t is the euclidean distance at time t computed between the predicted (\hat{y}) and the observed reflectance (y) using all k spectral bands (i).

Structural stability monitoring, which tests for a time-series whether new values significantly diverge from a modelled stable history, can subsequently be done with the euclidean distance in a similar way as has been done for each spectral band (Appendix 5.4). The only difference is the use of a single intercept regression model in the structural stability test instead of the harmonic model used when testing individual spectral channels (Equation 5.2). The reason for not using a seasonal model with euclidean distance is that such model was already used for computing the distance, hence already removing seasonality from the temporal signal.

To better understand spectral response of regrowing forests, we analysed recovery time for the six spectral channels of Landsat (Table 5.1), the euclidean distance (thereafter referred to as Dist) and NDMI. We refer to these different combinations as time-series metrics; the term encompassing both single spectral channels and derived indices.

5.2.5 Spatial analysis and explanatory factors

Our method provides for a set of locations spread across the Amazon basin how much time is needed to recover their spectral properties to pre-disturbance values (Figure 5.1). We hypothesize that water availability, expressed either as the total annual rainfall or as climatic water deficit (CWD) (obtained from Chave et al. (2014)), would explain a large part of the variations in recovery time, and that recovery time would be structured in space in a similar way than water availability. We therefore analysed variations in recovery time by comparing it to water availability, and by investigating the spatial structure with a variogram. Figure 5.2 presents the clear diagonal gradient in water availability with drier regions located at the South-East of the study area, while the North-West corner represents the wetter end of the gradient.

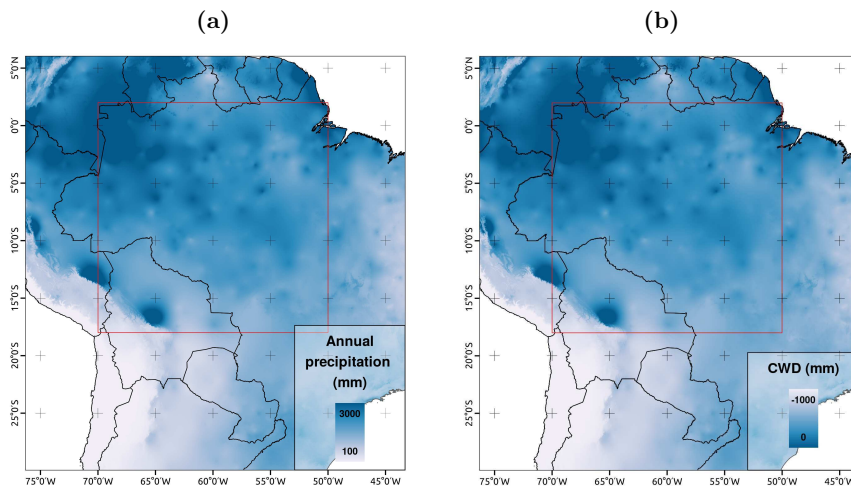


Figure 5.2: Maps of rainfall and Climatic Water Deficit (CWD). Rainfall and CWD data were obtained from Hijmans et al. (2005) and Chave et al. (2014), respectively. The red squares in both panels indicate the boundaries of the study area.

5.3 Results and discussion

5.3.1 Overall method performance

We sampled a total of 3596 time-series across the study area to understand spectral recovery dynamics of regrowing forests. Because forest clearing prevails in the southern states of the Brazilian Amazon (Mato Grosso, Rondônia) making these landscapes very dynamic, these states concentrate a large proportion of the samples even though they only represent a small part of the study area. Another large cluster of sample is located along the Amazon river representing the dynamic agricultural systems of riverine Amazonia (Van Vliet et al., 2012; Jakovac et al., 2015). The final number of observations analysed is reduced by an order of 70% compared to the initial dataset, unsuitable time-series being discarded at different steps along the process. From the initial 3596 sampled observations, 1982 remained after segment classification as they contained an appropriate sequence of stable forest followed by deforestation and then regrowth. For the remaining part of the samples, the segmentation and classification algorithms did not find appropriate successions of stable forest followed by deforestation. Observations were sampled from a simple layer of forest cover change where areas undergoing a forestation process are labelled as increasing tree cover regardless of their previous land covers. The uncertainty about the identity of previous land cover in the sampled dataset may partially explain why the stable-regrowth sequence was not systematically found in all profiles. An additional

explanation may originate from a limitation of the algorithm to isolate fast changing dynamics such as those occurring in swidden agriculture areas where a succession of multiple different land uses may occur over relatively short periods of time (two to seven years) (Jakovac et al., 2015; Dutrieux et al., 2016). Structural stability was tested for each time-series metric for each sample. A reduction in the amount of samples occurs once again at this stage due to certain time-series metrics never reaching structural stability. The proportions of observations reaching structural stability for every time-series metric is reported in Table 5.2. While time-series metrics from the visible part of the light spectrum (B1–B3) reach structural stability in most cases (90–96%), it is a lot more common for Short Waves Infra-red metrics to not reach stability before the end of the time-series (58% in the case of B5). Differences in the amount of observations reaching structural stability among time-series metric can be influenced by several factors. In certain cases, a given metric does not return to its initial state because the conditions have changed and spectral properties of the new land cover, even under stable circumstances differ from those of the previous land cover taken as reference by the method to define stable conditions (Figure 5.A3b). This might for instance occur for certain plantations, for which a different species composition than the previously established forest results in a different spectral signature, and certain time-series metrics never reach stability. In many other cases, the time-series metrics did not reach structural stability because the time-series ended before the on-going spectral recovery process had time to complete. The existence of such boundary effect implies that time-series with long recovery period are likely under-represented compared to fast recovering ones.

Table 5.2: Number and proportion of time-series metrics that have reached structural stability. Proportion are computed relatively to the number of time-series for which a suitable sequence of stable forest followed by regrowth was identified.

Metric	Number of time-series	Proportion of time-series
B1	1912	0.96
B2	1842	0.93
B3	1785	0.90
B4	1732	0.87
B5	1141	0.58
B7	1490	0.75
NDMI	1493	0.75
Dist	1259	0.63

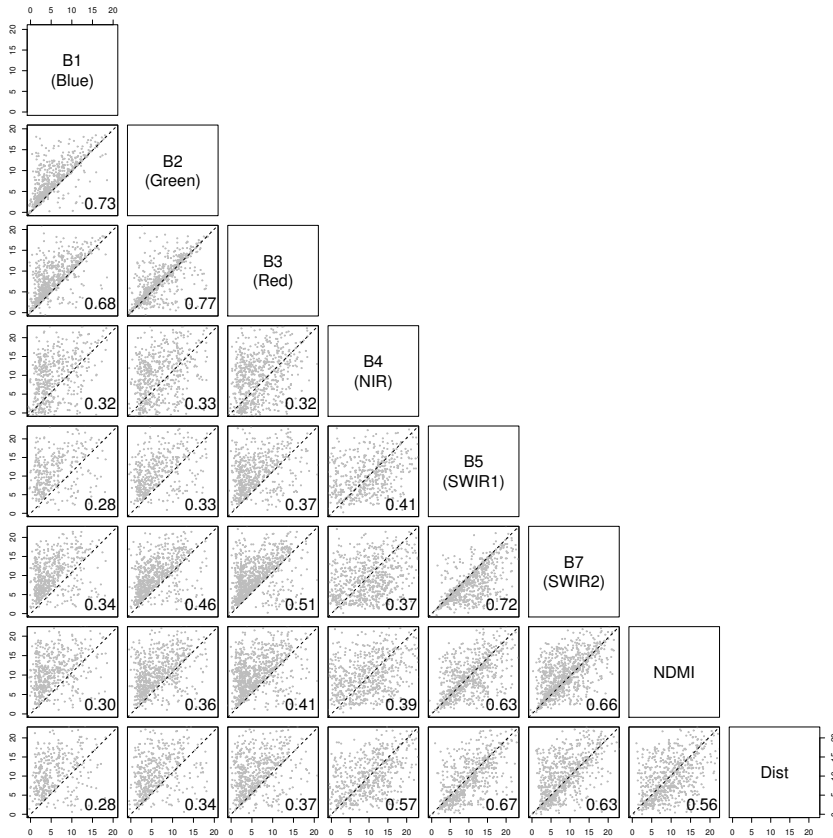


Figure 5.3: Pair-wise relations and Pearson correlations in recovery time (in years) between the 6 spectral bands of TM and ETM+, NDMI and the euclidean distance metric (Dist). The diagonals across each panel correspond to the 1:1 lines.

5.3.2 Contribution and differences between spectral regions to detect regrowing forests

Because different spectral channels are sensitive to different traits of forest canopies (Table 5.1), we expect each spectral region to respond differently to regrowing forests. Figure 5.3 presents pair-wise relations and correlations in recovery time between all metrics used for all the samples of the study area that reached structural stability. High correlations in recovery time between two bands, as it is for instance the case between B2 and B3, indicates that both bands reach structural stability roughly around the same time during regrowth and therefore respond similarly to canopy dynamics of regrowing

forests. There is therefore a certain redundancy between bands for which recovery times are highly correlated. High correlation can be observed among channels of the visible domain (B1–B3). Low pair-wise correlations can take two forms; observations can be either symmetrically distributed across both sides of the 1:1 line, or distributed with an unbalance to one side of the line. In the latter case, the unbalance suggests that one band systematically have longer recovery times than the other one. In the former case the two spectral regions complement each others, one still being able to differentiate a regrowing canopy from a stable forest, while the other one has already fully recovered, and *vice versa*. The combination of B4 and B5 is an example where both channels complement each others, as indicated by the wide symmetrical scatter in recovery time and the low correlation value. Vegetation indices that combine such uncorrelated bands should be successful in discriminating regrowing forests from stable forests under many circumstances. One band, B4, displays such low correlation with an even scatter for all the bands it is associated with, making it a crucial spectral region to use in association with additional bands. A large unbalance is present in most cases in pair-wise relations involving B1, systematically out of favour for B1. For that reason, and also because it is known for being largely influenced by atmospheric scattering, we decided to exclude B1 from the calculation of the integrated euclidean distance index (Equation 5.3). It is also interesting to note that while Dist combines information from all other spectral bands, recovery time in this metric is not systematically longer than recovery time of single bands or NDMI (Figure 5.3). Differences in recovery time among the eight time-series metrics presented in Figure 5.4 confirm the relative importance of the SWIR region of the light spectrum (particularly B5) in capturing the regrowth process.

5.3.3 Spatial patterns of recovery in spectral properties

It takes 7.8 years on average for forests of the Amazonian basin to recover their spectral properties following a disturbance. However, spectral recovery time has a large local spatial variability, observations distant from each others by only a few kilometres having a spectral recovery time that varies from about 5 to 15 years (Figure 5.5b). This large local variability suggests that spectral recovery time is largely influenced by local conditions such as intensity and frequency of land use and the cause of the disturbance. However, a large local variability does not necessarily imply an absence of spatial structure; there can be underlying regional effect influencing the variation over large distances. To test for such potential underlying spatial structure, we built a semi-variogram for the spectral recovery time. This analysis resulted in very low levels of spatial autocorrelation, that disappeared for distance ranges larger than 9 km. This finding is confirmed by Figure 5.5a in which regional variations do not appear to present any spatial structure. The lack of spatial structure might appear contradictory to previous studies that have found that resilience is largely driven by water availability (Hirota et al., 2011; Poorter

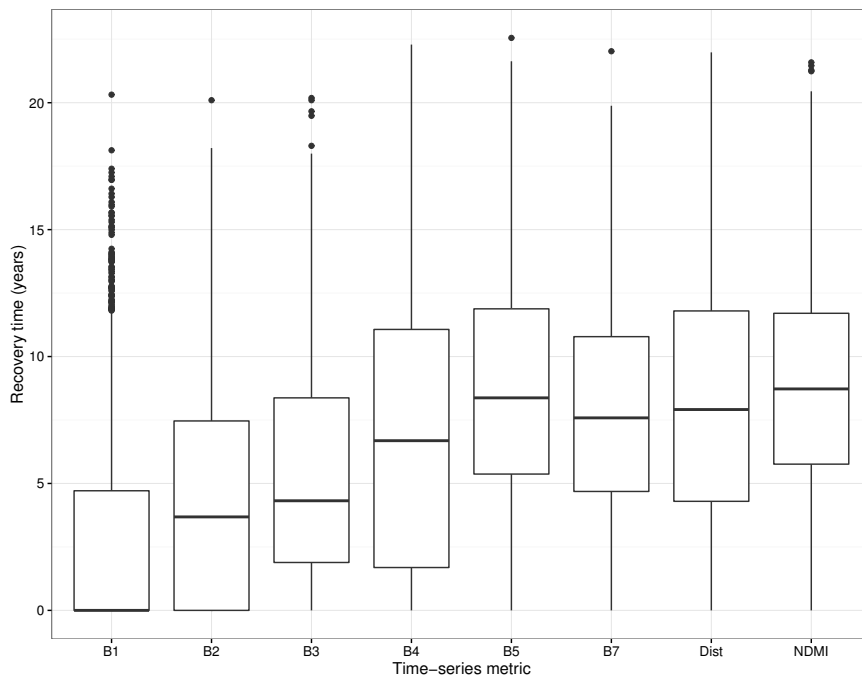


Figure 5.4: Boxplot of recovery time for the eight time-series metrics used.

et al., 2016) and therefore presents spatial patterns similar to those of water availability distribution. The strength of the relation between resilience in spectral properties and resilience of other forest attributes (biomass, structure, species composition) is not known, so that we lack elements to fully interpret the absence of spatial patterns or the influence of water availability. While local factors such as previous land uses, landscape context or disturbance intensity possibly out-weight more regional, climatic effects on forest resilience in general, the absence of spatial structure in our result can potentially also be explained by an independence between spectral resilience and more general concepts of forest resilience.

5.3.4 Water availability does not explain variations in spectral resilience

All spectral regions show a recovery time that varies between 1 and 20 years (Figure 5.4). Here we try to find regional or climatic effects that can explain this large variability. Water availability, within the rainfall and CWD ranges present in the study area, does not appear to have a strong effect on recovery time (Figure 5.6). The negative trend in recovery time with increasing water availability particularly pronounced for visible bands indicates that spectral properties in the visible part of the light spectrum have a longer

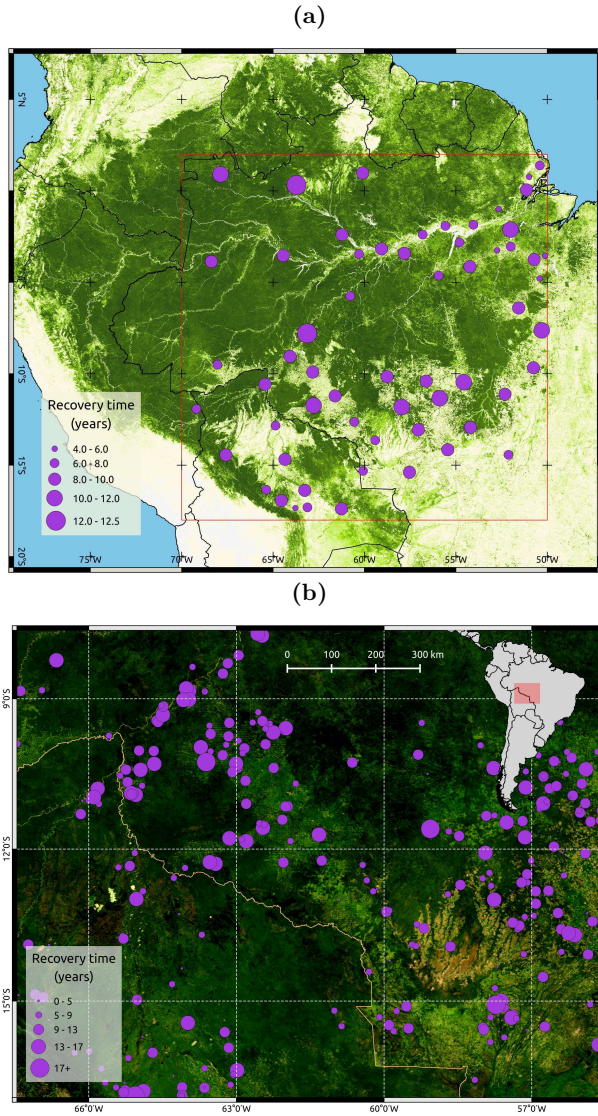


Figure 5.5: Maps displaying spatial variability in spectral recovery time for the entire area (a), and for a subset of the study area located across the border between Bolivia and Brazil (b). For the entire area map (a), data were spatially aggregated into 60 clusters using the k-means clustering algorithm (Hartigan & Wong, 1979). Background imagery of the subset (b) is provided by MapQuest (<http://www.mapquest.com/>) and contributors.

recovery time in more water limited environments, such as dry forests. However, NIR,

SWIR metrics, NDMI, and the euclidean distance (Dist) were not significantly related to water availability. The large variability in spectral recovery time ($sd = 5.3$ years) therefore appears to be rather driven by other factors than water availability, such as disturbance intensity and the availability of seed trees and dispersal agents in the surrounding areas. The mechanism in the case of biomass is that water availability enables longer growing seasons hence contributing to faster recovery. This positive effect of water availability on biomass recovery speed is particularly pronounced for water limited environments such as dry forests with a CWD inferior than 600 mm (Poorter et al., 2016). Because the study area only includes a few samples with such dry conditions, potential patterns of spectral recovery response to water availability may be hidden beyond the range of our dataset. Also we do not yet know how recovery in spectral properties relates to recovery in structure, biomass or species richness as observed in field studies (Jakovac et al., 2015; Poorter et al., 2016). It is therefore too early to derive ecological implications from the apparent independence we found between spectral recovery speed and water availability.

5.3.5 Opportunities for remote sensing and secondary forests mapping

We developed a novel approach using remote sensing to quantify post-disturbance recovery of forests in terms of their spectral properties. The approach combines change detection and structural stability monitoring, applied to time-series of multi-spectral Landsat data. In the present exploratory study we applied the method to time-series sampled over a large area covering the Amazonian basin. It takes 7.8 years on average for forests of the Amazonian basin to recover their spectral properties following a disturbance. We are certain that this relatively short period does not directly reflect the time required for a disturbed forest to recover all of its properties, which may take hundreds of years in certain cases (Brown et al., 1990; Poorter et al., 2016). Our findings imply that until 7.8 years after the recovery process has started, secondary forests remain spectrally different from their neighbouring old growth forests. This provides a large window of opportunities for remote sensing to assist mapping and further characterizing secondary forests. For instance, there is currently no existing map of young secondary forests at global scale. In the present study we used information on tree cover change to infer potential location of secondary forests (Hansen et al., 2013). This tree cover change information is relatively non-specific and does not for instance differentiate plantations from natural forest regeneration (Tropek et al., 2014). Such map of young secondary forests would be of great use for many purposes, including carbon sequestration accounting, and as shown in the present and in other studies (Fagan et al., 2015; Chazdon et al., 2016), the technology at our disposal would allow such mapping to be done.

The method we propose moves beyond the simple use of vegetation indices by taking an almost fully spectro-temporal approach. Despite the additional information provided by

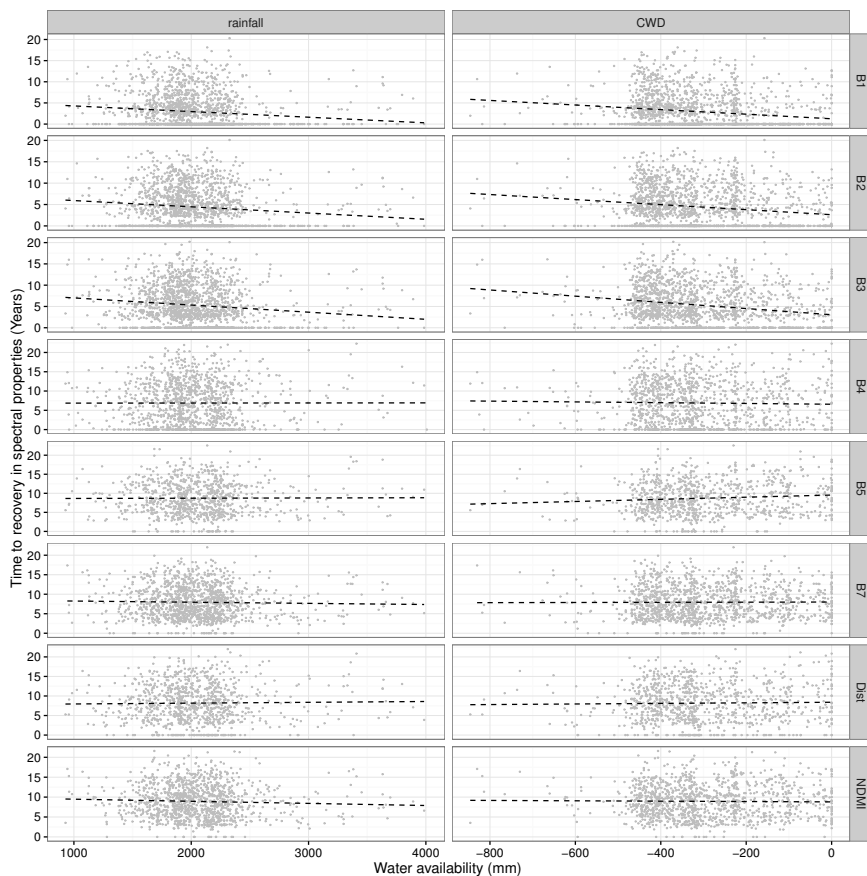


Figure 5.6: Relation between water availability (defined as mean annual rainfall and Climatic Water Deficit) and recovery time in the 8 time-series metrics investigated.

the spectral dimension, we could find very few examples of Landsat studies that integrated spectral and temporal dimensions (Fagan et al., 2015). In a time of data abundance and diversity we believe that it is crucial to take maximum advantage of all dimensions contained in the data so that the information extracted from them can be maximized. Beside using all dimensions of a single dataset, such as we did in the present case, fusion of multiple data streams could potentially lead to additional useful information. Specifically to forest regrowth mapping, combining radar with optical, which can be achieved similarly to the way we combined the different spectral channels, has great potential to enhance the sensitivity of the method to distinguish secondary from old growth forests. While the optical signal alone is only sensitive to the chemistry and the structure of the upper layers of the canopy, radar would provide an additional stream of information informing on the

structure of the woody part of the forest (Reiche et al., 2016). The recent launch of the sentinel 1 sensor has greatly benefited the availability and accessibility of radar data so that a fused optical-radar product could soon become a reality. Remote sensing data are spatially continuous and have been collected in a systematic way for several decades (since the beginning of the 80s in the case of Landsat TM). This archive therefore provides an invaluable source of information to assist in assessing, measuring and mapping ecological processes. The repeated measurement can be assembled in time-series which makes the data particularly suited to the monitoring of dynamic processes. We attempted to apply these principles to the case of forest recovery over the tropical region of the Amazon Basin. Success of the method to appropriately capture or proxy forest resilience from remote sensing would enable spatially continuous mapping of forests capacity to regrow, regardless of the presence of *in-situ* sites, which are generally scarce in the tropics.

The results of this preliminary study focus more on the spectral characterization of the regrowth process than on the connection with resilience in other forest attributes (Poorter et al., 2016). It is therefore too early to draw conclusions about the potential of the method to produce ecologically relevant estimates of resilience which would in turn inform nature conservation actions and policies. One important observation about the results is that the patterns of spectral resilience do not, as expected, follow the patterns of biomass resilience reported by Poorter et al. (2016). While Poorter et al. (2016) found water availability to be a major driver of biomass resilience, our results appear to be largely influenced by local conditions and no effect of water availability on spectral resilience was found. Several factors relating both to the recovery process (influence of type and intensity of previous disturbances, spatial context), the method we developed (variable temporal data density, inability to capture fast dynamics), and the limited water availability range of our study area can explain these differences. However, while remote sensing is able to retrieve variations and patterns, it has little power to infer causality from them. Better understanding how spectral resilience relates to ecological resilience by linking it to *in-situ* ecological measurements and ecosystem processes (flux data) should therefore be a priority for further studies.

5.4 Conclusion

The principal objectives of this study were to develop a framework to quantify resilience in spectral properties of regrowing forests and perform a preliminary investigation of spatial patterns and drivers of spectral resilience. By combining change detection, trajectories identification, and structural stability monitoring and applying it to multi-spectral time-series data from Landsat, we were able to determine the time necessary for forests to recover their spectral properties following disturbance and abandonment. The framework allows recovery time to be quantified for individual spectral bands as well as for the overall

spectral properties, expressed as an euclidean distance index. We applied the framework to a set of time-series sampled from regrowing forests across the Amazon basin. Relations between recovery time of the different spectral bands confirm the well known importance of the infra-red region (NIR and SWIR) of the light spectrum (represented by TM and ETM+ bands 4, 5 and 7) for distinguishing subtle differences in canopy attributes. Spatial variation in spectral resilience was not related to regional variations in water availability, but rather presented large local variability. This suggests that land use history and forest fragmentation in the surrounding areas may explain forest resilience. However, while we could observe this large local variability, we did not have sufficient elements to explain it; we therefore suggest further research to be carried out, primarily focusing on (i) better understanding the connection between recovery in spectral properties and recovery in other forest attributes as observed in the field, and (ii) investigating potential local drivers of recovery time variability (landscape context, disturbance intensity and land use history).

Acknowledgements

The research received partial funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement Number 283093 - The Role Of Biodiversity In climate change mitigatioN (ROBIN).

5.A Appendix

This appendix presents the framework we developed to quantify recovery time in spectral properties from remote sensing time-series.

5.A.1 Time-series temporal segmentation

The first step of our framework for characterizing recovery of spectral properties of regrowing forest is to identify the different phases of the forest across time. We used a breakpoint detection method derived from the econometrics literature capable of detecting abrupt changes in time-series regression model to partition each time-series into segments of homogeneous land dynamics (Bai, 1997a; Zeileis et al., 2003; Dutrieux et al., 2016). Given a time-series regression model (Equation 5.4), the algorithm tests whether the regression coefficients of the model remain constant over time and detects a breakpoint if at least one of the coefficients changes. To test this hypothesis, a triangular Residual Sum of Squares (RSS) matrix, which gives the RSS for each possible temporal segment in the time-series, is computed. The optimal number of partitions is then obtained by minimizing the Bayesian Information Criterion (BIC), while the position of the breakpoints is determined by minimizing the RSS among all possible partitioning schemes given by the RSS matrix (Zeileis et al., 2003; Bai & Perron, 2003). The initial time-series regression model can be written as follows:

$$y_i = x_i^\top \beta_i + u_i \quad (i = 1, \dots, n), \quad (5.4)$$

where y_i is the observation at time i of the dependent variable, x_i a vector of regressors, and β_i the vector of regression coefficients, and u_i the error term.

After partitioning model Equation 5.4 can be re-written as:

$$y_i = x_i^\top \beta_j + u_i \quad (i = i_{j-1} + 1, \dots, i_j, j = 1, \dots, m + 1), \quad (5.5)$$

where j is the segment index for a time-series regression containing $m + 1$ segments within which the regression coefficients (β_j) are constant. By convention $i_0 = 0$ and $i_{m+1} = n$ (Zeileis et al., 2003).

One important tuning parameter of the breakpoints detection method is the h value. h corresponds to the minimum segment size — expressed in percentage of the total number of observations in the time-series — allowed by the algorithm, and can be seen as a trade-off between sensitivity and stability. While small h values make the method more sensitive to fast dynamics, they also increase its sensitivity to noise and the probability

of detecting breakpoints which do not reflect any land dynamic process occurring on the ground. We decided to work with a h value of 0.2, which is sufficiently large to avoid spurious breaks to be detected, but short enough so that short lasting land dynamics regimes can be captured. Given that we are working with time-series of approximately 30 years, assuming an homogeneous spread of the observations across time, a h value of 0.2 corresponds to a minimum segment duration of 6 years. However, the assumption of an homogeneous spread of the observations across the Landsat observation period is rarely true, so that data rich periods (post 2000) will likely have shorter minimum segment durations than data poor periods (90s).

Here because we are interested in applying this method to the case of vegetation, which is known for having an annual cyclic behaviour usually referred to as phenology, we used a seasonal-trend regression model for Equation 5.4. Such regression model which approximates both intra and inter-annual behaviours of vegetation (Verbesselt et al., 2010b; Roerink et al., 2003), can be written as follows:

$$y_t = \alpha_1 + \alpha_2 t + \sum_{j=1}^k \gamma_j \sin\left(\frac{2\pi j t}{f} + \delta_j\right) + \epsilon_t \quad (5.6)$$

where the dependent variable y at a given time t is expressed as the sum of an intercept α_1 , a slope α_2 for potential temporal trend in the data, a sum of different frequency harmonic components representing seasonality ($\sum_{j=1}^k \gamma_j \sin(\frac{2\pi j t}{f} + \delta_j)$), and an error term ϵ_t . For the harmonic component of the model, $j = 1$ corresponds to the one year cycle, k is the chosen harmonic order, γ_j and δ_j correspond respectively to the amplitude and phase of the harmonic order j , and f is the known frequency of the time-series (i.e., number of observations per year).

By using this phenological trend model, a change in either intra-annual dynamics or temporal trend will trigger a breakpoint, resulting in a time-series containing multiple partitions. Because these partitions (or temporal segments) correspond to period with an invariable phenology or trend, they can be translated into periods of stable land dynamics regimes; such as a stable forest or a homogeneous forest regrowth following a disturbance. Unlike the rest of our framework this particular step requires that we work with a single vegetation index; we chose to use the Normalized Difference Moisture Index (NDMI) and justify this decision in section 5.2.3.

An example of time-series partitioning using the breakpoints detection algorithm with a seasonal trend model is presented in Figure 5.A1.

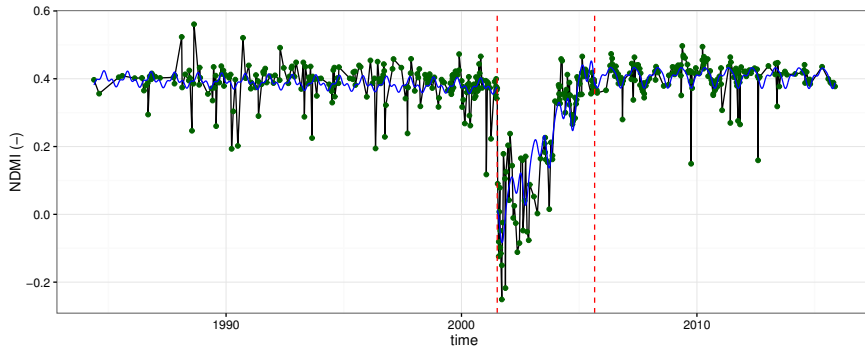


Figure 5.A1: Example a time-series partitioning. Green dots represent the observations of the variable (NDMI in that case), blue curves are the modelled temporal behaviour of each segment and the vertical red lines correspond to the breakpoints separating the segments. This partitioned profile example is a case of a stable forest (first segment to the left) that experienced a disturbance (first breakpoint) immediately followed by a period of regrowth (second segment) up to full recovery of the NDMI variable.

5.A.2 Classification of temporal segments

The temporal segmentation algorithm partitions the time-series into unlabelled segments of stable vegetation dynamics regimes. To be able to identify period of stable forest and forest regrowth we had to classify the segments into their corresponding classes. While it is relatively easy for a human — by simultaneously considering multiple elements of the spatio-temporal context — to assign classes to segments from visual interpretation (Cohen et al., 2010), it would be very challenging to define the appropriate set of thresholds required to automatically label the segments. We therefore trained a random forest classifier from a random set of visually interpreted partitioned time-series and used it to predict segment classes for the entire dataset. To aid the visual interpretation of segments we built an interface allowing the experimenter to quickly visualize the result of the temporal segmentation and assign labels to each segment. We then used this dataset to train a random forest classifier with the following segment properties used as predictors: mean, slope, phenological amplitude, duration of the segment and intercept values at the beginning and end of the segment. A preview of the interface used to interpret the temporal segments is shown in Figure 5.A2. Additionally to the temporal patterns present in the partitioned time-series, the experimenter can visualize the location of the sample over a very high resolution satellite image, and infer the probable dynamic from the spatial context, hence allowing to refine the interpretation of the segments. Although only a static preview of the spatial context is available, it did greatly benefit interpretation of unusual profiles such as in areas bordering rivers, where sedimentation have led to establishment of trees on areas previously covered by water. Although we did not make

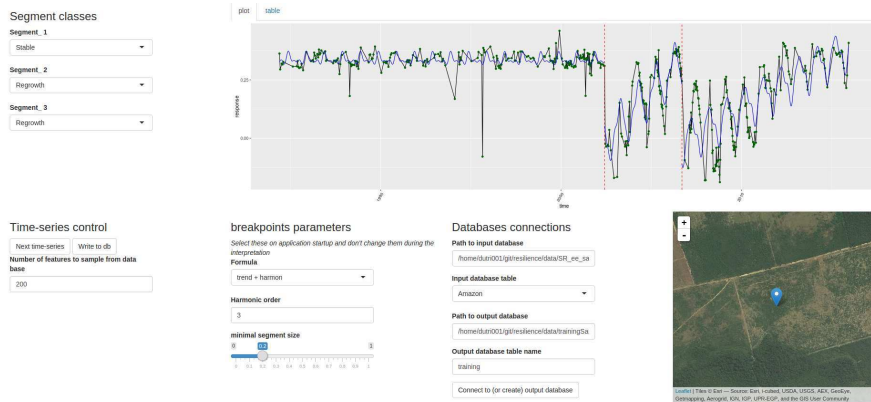


Figure 5.A2: Preview of the interface used for visual interpretation of the segments

use of this information, plantation could also be identified from the very high resolution images.

5.A.3 Monitoring spectral recovery dynamics – theory

Following the temporal segments classification we retained time-series that contained a sequence of stable forest, directly or indirectly followed by a period of forest regrowth. The characterization of recovery in spectral properties can only be done relatively to a stable period taken as reference; it is therefore essential that each time-series retained contains both a stable and a regrowth period. This step therefore implies a reduction of the initial number of samples, time-series for which a sequence of stable forest, regrowth could not be identified being discarded.

To monitor the recovery in spectral properties for each spectral band individually we used a structural stability test derived from the econometrics literature (Chu et al., 1996; Leisch et al., 2000; Zeileis et al., 2005, 2010). Structural stability monitoring consists in testing whether observations of a defined monitoring period significantly deviate from a stable behaviour as modelled over a stable period. The different steps of that approach consist in (i) defining a stable period from which stable behaviour of the time-series can be modelled, (ii) fitting a linear model to the stable period, (iii) monitoring for structural stability over the monitoring period by testing whether the results of a Moving Sum of Residuals (MOSUM) significantly deviate from zero or not. A structural stability boundary derived from the central limit theorem (Leisch et al., 2000) is used to test whether values are significantly different from zero; the hypothesis of structural stability is accepted when the MOSUM values (Equation 5.7) are inferior to that boundary (meaning that they are not significantly different from zero) and is rejected otherwise.

The MOSUM can be computed for any observation at time t using the following formula:

$$MO_t = \frac{1}{\hat{\sigma}\sqrt{n}} \left| \sum_{i=t-h+1}^t (y_i - \hat{y}_i) \right| \quad (5.7)$$

where h is the width of the MOSUM window defined as a proportion of the stable period, $\hat{\sigma}$ is the predicted variance estimated on the stable period, n is the number of observations of the stable period, and $y - \hat{y}$ are the residuals. h has to be set by the user, larger values decrease the sensitivity to potential noise contained in the time-series while increasing the time-lag between actual events and their confirmation by the statistical test, and lower values decrease that time-lag but make the method more sensitive to noise. We chose a value of 0.25 — meaning that the width of the MOSUM window equals 0.25 times the number of observations of the stable period — as a compromise between detection lag and sensitivity.

A typical disturbance-regrowth monitoring of a variable will display an increase in MOSUM values shortly after the disturbance and a first boundary crossing followed by a decrease of the MOSUM values until there is a second boundary crossing; the variable is then considered stable again. The $\frac{1}{\hat{\sigma}\sqrt{n}}$ component of Equation 5.7 makes MOSUM values sensitive to the quality of the model fit and the duration of the stable period. Better statistical fit and longer stable period will therefore result in higher MOSUM values — more likely to deviate from zero —, while a noisy stable period decreases the sensitivity to detect structural instability.

5.A.4 Monitoring spectral recovery dynamics – implementation

Our aim is to characterize recovery in spectral properties of regrowing forests. Theoretically, one could say that the spectral properties of a regrowing forest have not recovered until all individual spectral band or derived index have not reached structural stability. We therefore applied the structural stability monitoring to each spectral band available in the Landsat 5 and 7 sensors as well as NDMI and a combined variable we introduce in Section 5.2.4.

While the structural stability framework presented above is valid for any time-series, applying it to the case of regrowing forest requires some specific parameters adjustments. Firstly, since we are once again working with vegetation time-series, we chose to use a seasonal model for defining the reference stable behaviour of the forests spectral signal. A seasonal model is a reduced version of the seasonal trend model presented in Equation 5.6, that does not contains the trend component ($\alpha_2 t$). We decided not to use trend in the model because (i) these temporal segments supposedly represent stable forests, and (ii)

modelling even a slight temporal trend during the stable period could result in predicting unrealistic high values for the monitoring period.

Secondly, the stable period must be set by the user. Having already identified periods of stable forest in the segment classification step, we used these segments to define the stable period for each time-series independently, and tested for structural stability on every subsequent observation. In cases when a stable forest segment is not immediately followed by a period of regrowth (alternative land use in between); we started counting the regrowth time from the beginning of the regrowth segment. Thirdly, spurious crossings of the structural stability boundary, which do not necessarily reflect an actual recovery of spectral properties are likely to occur during the monitoring period. In order to better reflect the ecological reality of regrowing forests and similarly to DeVries et al. (2015a), we complemented the structural stability test by a confirmation period of one year, meaning that MOSUM values need to remain below the boundary for a minimum period of one year for a recovery to be confirmed. Several reasons such as noise in observations or an inversion in the residuals can trigger these spurious crossings of the boundary value.

Figure 5.A3 presents the concept of the structural stability monitoring framework applied to two variables derived from the same time-series (also used to illustrate the temporal segmentation process in Figure 5.A1). The bottom panel in Figure 5.A3a clearly depicts a first crossing of the boundary shortly following the end of the stable period, at which point the hypothesis of structural stability is rejected, until the MOSUM values experience a second crossing. The time during the end of the stable forest segment and the second boundary crossing determines the recovery time (of NDMI in that case). Figure 5.A3b presents a case where the values of the variable (SWIR 2 in that case) never return to their original level.

As illustrated in Figure 5.A3 the MOSUM values experience a first crossing of the boundary shortly after the end of the stable period, which corresponds to the point of rejection of the structural stability hypothesis and a second crossing when the time-series becomes stable again. In the present case we are mostly interested in the second crossing, which determines the duration required for the recovery.

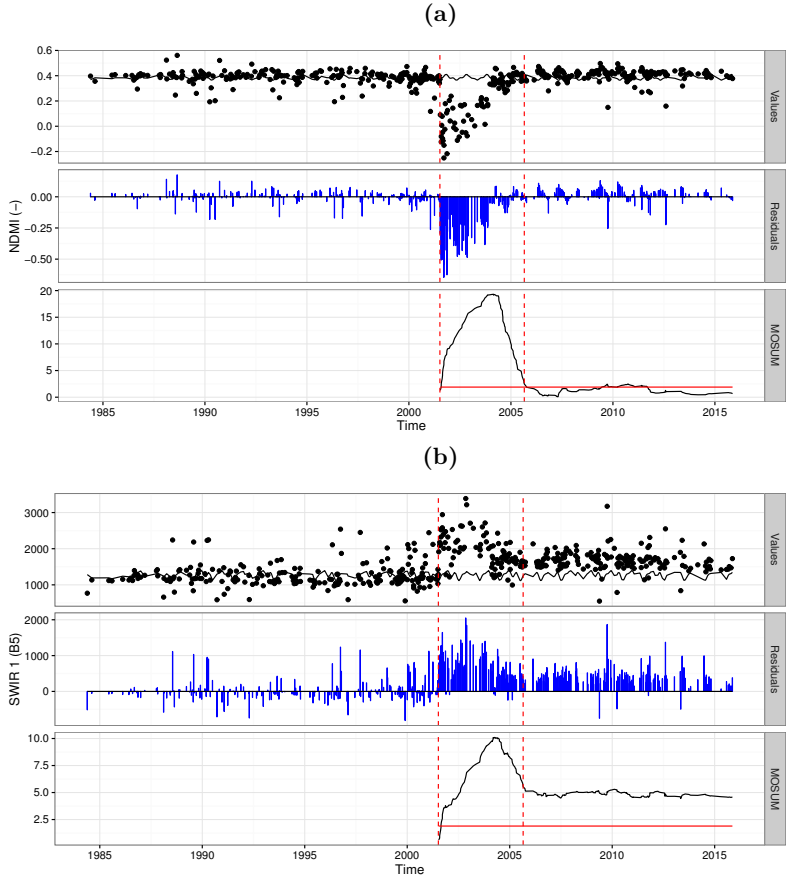


Figure 5.A3: Example of monitoring of structural stability with MOSUM for two time-series metrics (NDMI (a) and SWIR 2 (b)) of a same pixel. Vertical dashed red lines correspond to regime shifts identified in the temporal segmentation step. The monitoring period, during which new observation are compared to predictions from a modelled stable history, begins after the first regime shift. The horizontal red line corresponds to the structural stability boundary; when, during the monitoring period, MOSUM value are above the boundary the hypothesis of structural stability is rejected.

Chapter 6

Synthesis

6.1 Introduction

Adequate and accurate information on ecosystems, their characteristics and their dynamics is crucial in the context of climate change and biodiversity loss the world is currently experiencing. Information such as how much biodiversity does a forest host, how is this biodiversity spread spatially, what are the most vulnerable forests or where are they more threatened, is essential for resource conservation and management. Sensors on-board satellites are thought to be a potentially valuable source of information for mapping ecosystems and their dynamics, because they acquire data globally, consistently, and in a systematic way. However, these sensors do not provide direct measurements of biodiversity, resilience or other variables of interest to ecologists. The raw data (surface reflectance in the case of optical sensors) must be transformed into information, so that it can eventually become useful for a specific mapping need. Facing the amount, diversity and multidimensional nature of remote sensing data available, options to extract relevant information out of these data are nearly infinite, including statistical and physically based modelling, analysis of spatial patterns, and time-series analyses. The main objective of this thesis was to investigate options to retrieve ecologically relevant variables on forest structure, composition, diversity and dynamics from remote sensing.

Natural environments are complex and multidimensional systems in which different ecological processes operate at non-independent spatial and temporal scales (Peterson et al., 1998). Getting a complete understanding of the processes at stake therefore requires a multi-scale and multi-dimensional approach. The spatial dimension (in some cases along with the temporal and spectral dimensions) is ubiquitous in remote sensing datasets, making remote sensing a potentially well suited tool to analyse multi-scale and multidimensional systems and processes. This parallel between ecological processes and remote sensing dimensions was exploited in different ways along this thesis, in which different methods were developed and tested using Amazonian tropical forests as a case study. In the present chapter, I first summarize the main findings of the different chapters, I then discuss the consequences of scale and dimension for the ecological application of remote sensing, illustrating the benefits of multidimensional approaches for mapping and monitoring forest dynamics with the results of this thesis. I then discuss these scientific advances in the broader context of mapping biodiversity and ecological processes using remote sensing. Finally, the chapter concludes with some visions and future directions concerning the increasing contribution of remote sensing to answer ecological questions.

6.2 Specific contributions of this thesis

This thesis focused on mapping forest state and dynamics. The mapping of forest state was illustrated in chapter 2, in which I developed and evaluated a method to produce

coarse resolution maps of forest attributes by combining multiple remote sensing datasets and *in-situ* forest inventory data. The mapping of dynamical attributes was elaborated in chapters 3, 4 and 5, focusing on abrupt changes for deforestation monitoring (chapter 3), and forest resilience mapping (chapter 4 and 5).

6.2.1 Moderate resolution remote sensing can assist country-wide mapping of forest attributes

Chapter 2 presents a simple method to map forest attributes at country scale using coarse resolution (1000 m) remote sensing products. Variables were modelled over Bolivia using *in-situ* national forest inventory plots for training of a random forest regression framework. The approach performed relatively well for predicting and mapping forest floristic properties (e.g. proxies of species richness, composition and palm trees density) but relatively poorly for predicting and mapping forest structure (e.g. tree density, basal area, diameter at breast height). The main conclusions and implications drawn from these results are that: (i) there is added value in combining a variety of data-sources in an empirical modelling approach, (ii) coarse resolution remote sensing offers the potential to generate information on forest attributes that reflects the main known patterns of species richness and composition, (iii) dynamic predictors like phenology are particularly important for modelling static response variables (Somers & Asner, 2012; DeFries & Townshend, 1994), and (iv) there is a strong disconnection between variables measured by most remote sensing sensors and forest structure. Although the approach performed well in predicting floristic variables spatially, it cannot act as a replacement for forest inventories and local expert knowledge. The predicted species richness did not represent well some known biodiversity hotspots, emphasizing the dependence on sufficient *in-situ* data, representative of all environments considered and the importance of local expert knowledge in the validation and interpretation of such results.

Overall this chapter demonstrates the potential of remote sensing for coarse, large-scale mapping of forest attributes. The ability to produce spatially continuous estimates of floristic properties over tropical forests brings us a step closer to operational remote sensing based biodiversity monitoring. Such empirical approach developed and implemented over Bolivia is relatively simple, hence providing a cost efficient way to obtain spatial data relative to forest attributes. It can potentially be repeated for any area provided that it contains sufficient *in-situ* data for model calibration.

6.2.2 Forest dynamics and remote sensing

Chapters 3, 4 and 5 approach forest dynamics from a remote sensing perspective. Chapter 3 addresses the relatively simple dynamic of tree cover loss, while chapters 4 and 5 develop methods to analyse and quantify forest resilience. In all three cases time-series approaches

were used, taking full advantage of the temporal dimension contained in the the remote sensing data.

Deforestation monitoring

Forest cover loss is a conceptually simple dynamic process, but because forests are constantly changing due to seasonal dynamics and response to climatic events like droughts, detecting forest cover loss in an automated way can be challenging. In Chapter 3, a multi-scale approach was developed and tested to improve existing automated deforestation detection methods. The main scientific advance brought by this thesis in term of deforestation monitoring is related to the ability to detect deforestation in environments with highly irregular seasonal dynamics, such as tropical dry forests. This approach builds upon methods that use the deviation from a modelled seasonal greenness cycle to infer tree cover loss (Verbesselt et al., 2012; Zhu et al., 2012; DeVries et al., 2015b). When inter annual irregularities caused by natural variability are present in the seasonal cycle, climate related anomalies can be confused with anthropogenic disturbances. Because climatic anomalies occur at a regional scale while anthropogenic forest disturbances are more local processes, using the spatial dimension offers an opportunity to distinguish them. The method that I developed in Chapter 3 improves the detection accuracy of remote sensing based deforestation monitoring by taking advantage of this spatial scale difference between natural dynamics and human-induced changes. By tackling forest cover change using a multi-scale approach, this work, together with Hansen et al. (2008) has emphasized the benefits of working at different resolutions simultaneously, hence allowing processes that occur at different spatial scales to be distinguished from each other.

Recently there has been rapid developments in methods for forest cover change detection. The main developments were improvements in: (i) detection accuracy (sensitivity, robustness, versatility), (ii) spatial resolution, and (iii) timing of detection. The most recent advances in the field have led to the deployment of an operational deforestation alert system (globalforestwatch.org/) that provides updated deforestation estimates every 8 days at 30 m resolution (Popkin, 2016). Such platforms play important roles for resources conservation, outreach and public awareness. It benefits resource conservation by providing a tool for local forest managers and legal authorities to track current forest changes; it benefits outreach by providing easily accessible information about forest cover change to non-remote sensing specialists; and it benefits public awareness thanks to the capacity it provides for anyone with an internet connection and a computer to visualize in near real time changes occurring to our forests. Such platforms also play an important scientific role by providing quantitative forest change estimates for forest resources monitoring and carbon accounting.

Proxies of forest resilience

Chapters 4 and 5 extend the contribution of this thesis on remote sensing based forest dynamics by focusing on forest regrowth and resilience. Two different approaches were developed with the aim to obtain spatially continuous estimates of forest resilience. Chapter 4 presents a novel method to estimate land use intensity — a known proxy of the capacity of secondary forests to recover in swidden agriculture systems. By applying a temporal segmentation algorithm derived from the econometrics field to 30 years long Landsat time-series covering a swidden agriculture area of Brazil, the different events and stages of the swidden agriculture cycle (forest removal, agriculture, abandonment, forest regrowth) could be identified. The number and frequency of cultivation cycles provide quantitative estimates of land use intensity which largely determines the capacity of forests to recover. By quantifying land use intensity from remote sensing time-series, the method developed and tested in Chapter 4 provides spatially continuous and objective estimates of forest resilience for swidden agriculture systems.

Chapter 5 investigates a method to quantify recovery time in spectral properties of regrowing forests. The method developed as part of Chapter 5 applies to Landsat time-series, and combines temporal segmentation and structural stability monitoring. Periods of forest regrowth are identified using the temporal segmentation, and the structural stability monitoring allows to quantify the time required for each Landsat spectral channel to recover to its pre-disturbance values. The framework is applied to 3,596 sampled locations across the Amazon and spectral recovery time produced are compared to water availability. On average, it took 7.8 years, with a large variability ($sd = 5.3$ years) for disturbed Amazonian forests to recover their spectral properties. Spectral recovery time did not appear to be structured in space or influenced by water availability, suggesting that local conditions (e.g. type and intensity of disturbance, surrounding forest matrix) likely affect spectral resilience.

Resilience is a dynamic process, and the two approaches take full advantage of the temporal dimension available in the Landsat archive that contains repeated observations since the 1980s. By extracting information from Landsat based temporal trajectories and linking it to ecological variables, these contributions extend the work started in 2010 with the development of the LandTrendR algorithm (Kennedy et al., 2010) and the Vegetation Change Tracker (Huang et al., 2010). Additionally the two Chapters demonstrate how, by mining the information contained in temporal trajectories, remote sensing is capable of delivering insights that go beyond simple land cover (an important cornerstone of remote sensing applications), and derive information related to land use instead. Finally, by delivering ecologically relevant information, both contributions are a proof of the large amount of information contained in the Landsat archive. A lot of this information has to be mined from the temporal dimension using time-series approaches. Further exploring temporal trajectories from Landsat time-series will undoubtedly result in great benefits

for remote sensing based ecological applications.

6.3 Reflection

6.3.1 Potential of remote sensing for ecology

"At different scales, different sets of mutually reinforcing ecological processes leave their imprint on spatial, temporal and morphological patterns." This quote from Peterson et al.'s 1998 review, summarizes well the complexity and multidimensional nature of ecosystems.

To illustrate this notion of scale interaction in a more concrete and simplified way, we can take the example of the process of tree cover change — a dynamic process that played a central role in three out of the four core chapters presented in this thesis. First of all, tree cover change is a dynamic process, which means that it obviously involves the temporal dimension. The change in tree cover can be either negative, in which case there is a loss in tree cover — often induced by human deforestation —, or positive, when forests regrow. However, there is also a strong spatial scale component that affects tree cover dynamics, and a dependence exists between events occurring at the local scale and services delivered at the scale of the landscape. A simplified conceptual representation of that spatial interaction is presented in Figure 6.1, and can be described as follows. Individual trees make up a landscape which delivers services that are beneficial to these individual trees themselves. When tree cover loss occurs, individual trees disappearing, it reduces landscape integrity, and in turn affects the capacity of the ecosystem to deliver its services. One important service that was also investigated in this thesis, is the capacity of components of the system to regenerate. This capacity to regenerate, which is also being referred to as "engineering resilience", is influenced by conditions (micro-climate, seed sources and non-eroded top soils) that are largely determined by the integrity of the surrounding forest matrix (Chazdon et al., 2007).

This simplified example therefore already comprises two non-independent dimensions (spatial and temporal) and its main processes interact at two different spatial scales (single tree and landscape). Understanding such a system, and determining its vulnerability to disturbance intensity requires a multi-dimensional approach in which different spatial and temporal scales are integrated.

Remote sensing offers the potential to capture processes at different spatial scales. The spatial dimension is ubiquitous in remote sensing datasets, and as shown in the different chapters of this thesis, remote sensing offers the capability to retrieve information for large areas while maintaining the local realism of human scale processes. To provide a more complete and practical overview of the different scales of ecological processes in relation

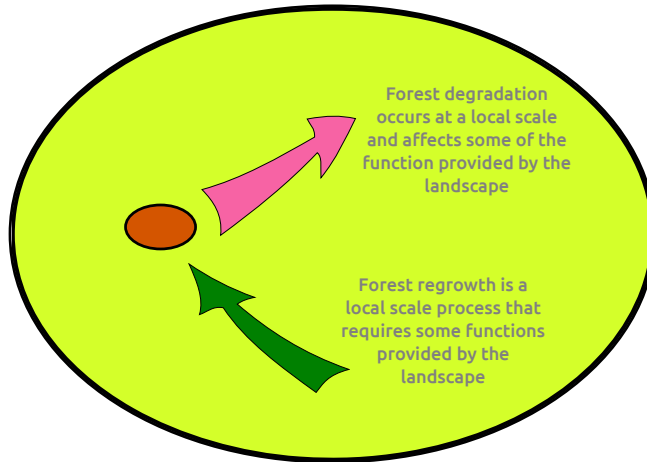


Figure 6.1: Simplified conceptual representation of a feedback loop interaction between spatial scale and tree cover dynamics. The large green ellipse represents a forested landscape while the inner patch represents a local scale change in forest cover.

to remote sensing capabilities, I present in Figure 6.2 a conceptual diagram in which I position common optical remote sensing technologies in relation with the dimensions and scale of ecological processes of interest.

However, while processes that operate at different spatial scales can be captured individually using remote sensing techniques, fully understanding the complexity of processes and their interactions requires that different scales be integrated. An example of how integrating multiple scales can be achieved is given below.

Table 6.1: List of acronyms used in Figure 6.2.

Sensor acronym	Full name
AVHRR	Advanced Very High Resolution Radiometer
ETM+	Enhanced Thematic Mapper Plus
MODIS	Moderate Resolution Imaging Spectroradiometer
OLI	Operational Land Imager
PROBA-V	PROBA Vegetation
SPOT	<i>Satellite Pour l'Observation de la Terre</i>
TM	Thematic Mapper

I have approached in this thesis the topic of deforestation, which can now be monitored operationally using Landsat data (Hansen et al., 2016), and other authors have recently

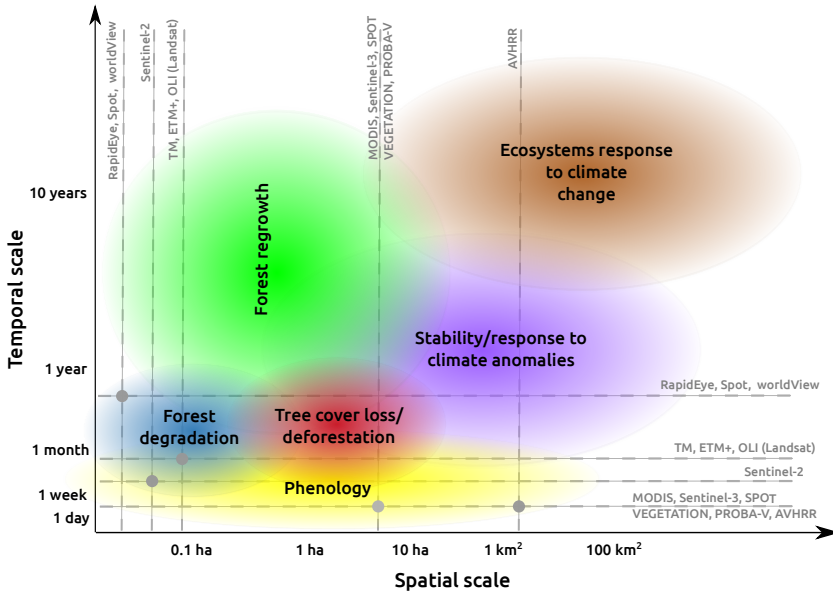


Figure 6.2: Conceptual diagram showing the correspondence between spatio-temporal scales of six ecological processes and 13 available optical sensors. Spatio-temporal suitability of a sensor to map a process is given by the position of the intersecting lines relative to the spatio-temporal scale of the process. A sensor is suitable for a given process when its spatio-temporal resolution is higher than the scale of the processes (lines intersecting at the bottom left of the ellipse). Acronyms are explained in Table 6.1. Reported scales of ecological processes are hypothetical.

developed methods to estimate ecological resilience from coarse resolution remote sensing time-series (De Keersmaecker et al., 2015). Ecological resilience can be perceived as an ecosystem property, while deforestation is a local process that, if integrated over a large area, can inform on the degree of disturbance the ecosystem has gone through. While there is to my knowledge no empirical evidence that a connection exist between the degree of disturbance within a landscape and its ecological resilience, it is highly plausible that the two are not independent. Thanks to remote sensing and its intrinsic spatial dimension, we now have the ability to quantify both the local scale process and the ecosystem property; but rather than using these two sources of information independently, it is necessary to combine them. While this is simply one example among many other possible benefits of combining analyses at different spatial scales, investigating the relation between degree of disturbance and ecological resilience would be highly relevant, as it could for instance inform conservation actions on critical levels of disturbance and result in well guided policy actions.

6.3.2 On the use of multidimensional remote sensing approaches to map, monitor, and analyse ecological processes

The previous example about tree cover change dynamics did not only present the existence of interactions between scales in ecological processes, it also revealed their multidimensionality. The case of tree cover change presents two obvious dimensions (the spatial and the temporal dimension), however, if one would observe the process at a fixed spatial scale, that person would certainly notice that various aspects of the forest are changing as well. A regrowing forest for instance will at least display a development in structure and species composition, resulting in a different canopy structure and chemistry. These different attributes that evolve during forest regeneration can be perceived as additional ecological dimensions. Similarly to ecological processes, remote sensing datasets are often multidimensional too. Common dimensions of remote sensing datasets are the spatial, temporal, spectral and thematic dimensions. I refer here to the thematic dimension as the potential combination of data-sources from different remote sensing technologies. Each dimension has potential to retrieve different types of information. By taking advantage of the spatial relations that exist among the elements at the surface of the earth, the spatial dimension can improve mapping accuracies; the temporal dimension is particularly suited for investigating dynamic processes; the spectral dimension allows subtle differences in chemistry or structure to be distinguished; and the thematic dimension enables retrieving information on different independent properties of an object (structure of woody components, chemistry).

While each dimension potentially provides complementary information, the greatest benefits can be achieved by combining them. Such synergistic use of multiple dimensions can be illustrated by the example of remote sensing based land cover mapping. Not only land cover is an output that perfectly suits the portfolio of remote sensing capabilities, it is also tightly coupled with life forms and ecological processes and is therefore of great interest to ecological studies. It is therefore no surprise that land cover mapping has been an ongoing research topic for many years already (Justice, 1978). Land cover mapping nearly always make use of the spectral dimension; a common approach — as taught in remote sensing introductory classes — is to establish a statistical relation between the reflectance in multiple wavelengths of each pixel and land cover. While this simple uni-temporal approach in which pixels are considered independently from each others often yield good results for highly contrasted land covers, pioneers of land cover mapping have discovered a long time ago that land cover maps could be improved by extending the dimensionality of the input data to the spatial and temporal domains (Gurney & Townshend, 1983; Tucker et al., 1985; DeFries & Townshend, 1994). More recent developments have integrated both spatial and temporal dimensions in a synergistic way, using temporal similarity of neighbouring pixels to define spatio-temporal objects (Lhermitte et al., 2008). However, synergies between dimensions are not limited to land cover mapping; additional examples

of these synergies include Desclée et al. (2006) who developed a fully spatio-temporal approach to detect deforestation in the Congo basin, and Somers & Asner (2012) who combined spectral and temporal dimensions in a hyper-spectral time-series approach that helped to enhance mapping accuracy of invasive plant species.

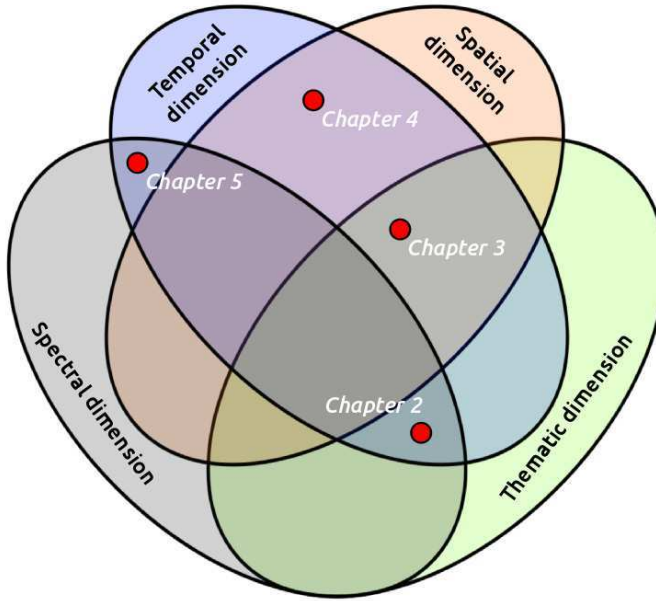


Figure 6.3: The four chapters of this thesis placed according to the four common remote sensing dimensions (spatial, temporal, spectral and thematic). Chapter 2 combined multiple data sources for forest attributes mapping, Chapter 3 took advantage of the spatial scale difference between natural and anthropogenic processes in a time-series approach to improve deforestation detection, Chapter 4 mapped land use intensity from Landsat time-series at the landscape level for individual swidden agriculture fields, and Chapter 5 investigated temporal dynamics in spectral properties of regrowing forests of the Amazon.

As illustrated in the scheme of Figure 6.3, the present thesis also made extensive use of multiple dimensions and synergies among them. Combining multiple data sources in Chapter 2 enabled country wide mapping of forest attributes. A large diversity of variables including MODIS (Moderate Resolution Imaging Spectroradiometer) derived surface reflectance, leaf area index and phenology, terrain metrics, and tree height were used as predictors in a random forest modelling framework. Phenology and surface reflectance were particularly contributive in modelling floristic properties, highlighting the benefits of combining temporal and spectral dimensions (which in itself exploits the thematic dimension) for such mapping efforts. Chapter 3 took advantage of the spatial difference between climate related and anthropogenic processes. Combining MODIS with Landsat time-series

helped distinguishing the deforestation events from seasonal irregularities and therefore revealed that integrating the spatial dimension in a time-series approach is possible and can enhance deforestation detection. Chapter 4 identified agricultural field boundaries in a swidden agriculture landscape using temporal similarity of neighbouring pixels, and retrieved land use history at the level of individual fields by mining their temporal signature. This chapter therefore used both spatial and temporal dimensions, and fully exploiting the information of the temporal signature was particularly contributive to quantifying the land use intensity variable. Finally, a full spectral-temporal approach for quantifying spectral recovery of tropical forests was developed in Chapter 5 enabling the detection of subtle differences in canopy development during the forest regrowth process.

While there are already many examples of remote sensing applications that benefited from using different dimensions, there is still room for improvement. I showed in Chapter 4, about swidden agriculture, that remote sensing based temporal trajectories can be used to derive past land use intensity, which is a specific type of land use information. Building upon this method, and further mining the information contained in temporal trajectories potentially holds great potential for characterizing land use parameters beneficial to answering ecological questions at various scales. Thorough characterization of land use history could therefore be used as an input in carbon modelling studies applied to secondary forests. But the temporal dimension is not the only one that can be further explored. The thematic dimension holds great potential as well, and methods that aim at combining multiple data-sources of different nature will certainly help maximizing the information we can retrieve from remote sensing. I discussed for instance that optical remote sensing signal over dense forests is mainly influenced by canopy structure and leaf chemistry, while radar on the other hand is insensitive to chemistry but largely influenced by structure of the woody components. A combination of optical and radar remote sensing can therefore highly benefit mapping efforts such as for biomass and carbon accounting (Baccini et al., 2012). Another example of radar-optical combination is given by Reiche et al. (2015) who successfully combined Landsat and radar data in a time-series approach to improve deforestation detection. Finally, Asner et al. (2007) show that making synergistic use of hyper-spectral and LiDAR data — two already highly valuable sources of information when used independently — could result in an almost unrivalled mapping potential, the system and its associated method being able to identify species of individual trees.

The diversity of available remote sensing data will keep on increasing in the coming years with the launch of new sensors designed for ecosystems structure and dynamics mapping (Le Toan et al., 2011; Dubayah et al., 2014), and combing these new datasets with existing datastreams will ensure that all information at our disposal is used to its full potential (Reiche et al., 2016; Pettorelli et al., 2016).

Finally, hyperspectral data which until now have been mainly collected using sensors

onboard air-planes will become increasingly available in a near future thanks to the launch of the Environmental Mapping and Analysis Program (EnMAP) sensor (Kaufmann et al., 2008). While Asner et al. (2014b) has demonstrated that hyperspectral remote sensing has a great potential for detailed vegetation mapping and monitoring, the spectral dimension remains a largely unexplored dimension. Figure 6.2 for instance illustrates this knowledge gap. Because we evolve in these dimensions every day, our knowledge and ability to perceive the spatio-temporal domain is relatively advanced and it was therefore easy to position ecological processes along the space and time axes. Our perception of the spectral dimension is not as intuitive as for space and time, and although a similar graphical representation including the spectral dimension would be very useful, considering our current knowledge it would be challenging to position ecological processes along a scale of spectral contrasts.

Table 6.2 shows the potential contribution of the identified remote sensing dimensions to capture the six ecological processes reported in Figure 6.2. Because all six processes are dynamic, hence containing a strong temporal component, it is no surprise that the temporal dimension has an important role in capturing them all. The spectral dimension is particularly contributive to capturing subtle processes such as forest degradation or forest regrowth, the thematic dimension often offers complementary information such as structure of woody components or vertical structure when optical remote sensing is combined with LiDAR or radar remote sensing.

Table 6.2: Potential contribution of the common remote sensing dimensions in detecting and monitoring six ecological processes shown in Figure 6.2. The number of pluses indicates the hypothesized potential contribution of the given dimension to contributes to quantifying or mapping the associated process.

	Spatial dimension	Temporal dimension	Spectral dimension	Thematic dimension
Tree cover loss / Deforestation	++	+++	+	++
Forest degradation	++	+++	++	++
Phenology	+	+++	+	+
Forest regrowth	+	+++	++	++
Response to climate anomalies	++	+++	+	+
Response to climate change	+	+++	+	+

6.4 Future prospects

The remote sensing landscape has evolved continuously over the past 40 years with an increase in data availability and accessibility. Today, remote sensing data are not only becoming more abundant and accessible, but also more diverse and at a higher spatial and temporal resolution. This evolution towards higher resolution data is also reflected in the semantics. AVHRR for example, which at the time it was launched at the end the 1970s was certainly a major technological advance, was labelled "very high resolution". Today, sensors that collect data at 1 km spatial resolution would enter the coarse resolution category, the term "very high resolution" being only used for the sub 5 m domain. Science has successfully taken advantage of these technological advances and today, remote sensing not only provides valuable insights to the scientific community, it also provides near real time information that is accessible to all. Taking into account the increasing abundance of high resolution data, as well a shift toward operational remote remote sensing, I provide an outlook on possible future developments below.

6.4.1 Toward a global operational remote sensing based mapping of ecosystem dynamics and biodiversity

The newly developed alert system for near real time deforestation monitoring perfectly exemplifies a successful transfer of technological innovation to an operational system. Today, any citizen of the earth can access (given he or she has access to the internet) and visualize in a simple and understandable way information that was once only accessible to specialists. Reaching this stage will undoubtedly have large impacts in terms of awareness and in turn benefit conservation initiatives. Contemplating the success of operational remote sensing based deforestation mapping, can we expect the production of other variables to follow the same path? Can maps of forest resilience, ecosystem structure or species richness be produced globally, in an operational way and regularly updated? While this would be desirable and highly beneficial for environmental conservation causes, biodiversity and the various ecological processes I have studied in this thesis are much more complex than tree cover, and therefore present different challenges. Ecosystems are complex, multidimensional, and can be studied at various scales so that even the definition of terms such as biodiversity are still subject to controversies (Pettorelli et al., 2016). For ecosystem monitoring to achieve the operational level of deforestation monitoring, it is important that a discussion takes place among all parties at stake (ecologists, policy makers/governments, conservationists, space agencies, the remote sensing community, and local communities involved) (Pereira et al., 2013; Skidmore et al., 2015) in order to reach agreements and frame biodiversity variables in a global observation context. This discussion process has in fact already started, involving international experts under the umbrella of the GEO-BON (the Group on Earth Observations - Biodiversity Observation Network), and has already

reached encouraging preliminary results. For instance the network has largely contributed to framing global biodiversity monitoring and reviewed some known potential of existing technologies (Pereira et al., 2013; Pettorelli et al., 2016). Additionally, in line with the Global Forest Watch (globalforestwatch.org) — a platform developed by the World Resources Institute to provide information on forest cover change — another important recent contribution of the GEO-BON network is the set-up of a global platform for sharing and analysing biodiversity data named *BON in a Box* (boninabox.geobon.org). By providing easily accessible information to many stakeholders and scientific communities, these platforms play a crucial role in accelerating the pace of scientific and technological development. They facilitate collaboration and mutual feedback mechanisms among communities, and hence the application of the obtained information for resource conservation and use.

6.4.2 Remote sensing of ecology in the new era of data science and artificial intelligence

Extracting valuable information from the abundance and diversity of remote sensing data that will become available in a near future will require use of new and innovative approaches. One of the methods that can play an important role in retrieving valuable information in this context is computer vision. Computer vision is a sub-field of artificial intelligence particularly targeted at detecting and identifying objects in digital images. To better understand what is the added value of computer vision, we can take as an analogy the way the human brain interpret images. When we, humans, look at an image, we automatically and unconsciously integrate the spatial context in our interpretation of the elements of the image; our eyes see groups of pixels but our brain immediately sees objects, identifies their nature, and perceives dependences and connections among them. Similarly to the way the human brain identifies objects in images, computer vision techniques are able to detect and identify the nature of objects such as trees, cars, people or animals in digital images. Computer vision has largely benefited the field of medical imagery already and is ubiquitous in most modern internet picture services. A transfer of these existing and already operational methods, would greatly benefit ecological applications of remote sensing by transforming raw data into properties relevant to the ecological domain.

6.5 Concluding remarks

The remote sensing landscape has evolved continuously over the past 40 years. Today a large variety of remote sensing sensors with different spectral and thematic capabilities deliver data globally at various spatial and temporal resolutions. This current availability

of remote sensing data provides great opportunities for regional to global forest dynamics mapping and retrieval of other variables relevant to ecological applications. To ensure that the available data are exploited to their full potential, innovative methods that can (i) combine multiple data sources, (ii) take advantage of the spatial, temporal, spectral and thematic dimensions contained in remote sensing datasets, and (iii) make synergistic use of *in-situ* and remote sensing data, must be developed and assessed. By developing and testing methods to map forest state and dynamics while combining all these requirements, this thesis demonstrates the potential of remote sensing for tropical forest mapping and monitoring applications. The amount and diversity of data will keep on increasing, both from the remote sensing and *in-situ* sides, and it is important to work on further integrating all these data streams for successful global forest and biodiversity mapping and monitoring in the future. The technology we have at our disposal and its capacity to monitor the earth is a great advantage that the Maya did not have, and we must make proper use of it to ensure conservation and wise use of these important forest resources.

References

- Adeyewa, Z. D., & Nakamura, K. (2003). Validation of TRMM radar rainfall data over major climatic regions in Africa. *Journal of Applied Meteorology*, *42*, 331–347.
- Anderson, M. C. (1966). Stand structure and light penetration. II. A theoretical analysis. *Journal of Applied Ecology*, (pp. 41–54).
- Arcioni, M., Bensi, P., Fehringer, M., Fois, F., Heliere, F., Lin, C.-C., & Scipal, K. (2014). The Biomass mission, status of the satellite system. In *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International* (pp. 1413–1416). IEEE.
- Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, *22*, 3855–3862.
- Asner, G. P., Keller, M., Pereira, R., Jr, Zweede, J. C., & Silva, J. N. (2004a). Canopy damage and recovery after selective logging in Amazonia: field and satellite studies. *Ecological Applications*, *14*, 280–298.
- Asner, G. P., Keller, M., & Silva, J. N. (2004b). Spatial and temporal dynamics of forest canopy gaps following selective logging in the eastern Amazon. *Global Change Biology*, *10*, 765–783.
- Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Jones, M. O., Martin, R. E., Boardman, J., & Field, C. B. (2007). Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, *1*, 013536–013536.
- Asner, G. P., & Martin, R. E. (2008). Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, *7*, 269–276.
- Asner, G. P., Martin, R. E., Carranza-Jiménez, L., Sinca, F., Tupayachi, R., Anderson, C. B., & Martinez, P. (2014a). Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *New Phytologist*, *204*, 127–139.
- Asner, G. P., Martin, R. E., Tupayachi, R., Anderson, C. B., Sinca, F., Carranza-Jiménez, L., & Martinez, P. (2014b). Amazonian functional diversity from forest canopy chemical assembly. *Proceedings of the National Academy of Sciences*, *111*, 5604–5609.

- 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.
- Avitabile, V., Baccini, A., Friedl, M. A., & Schmillius, 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.
- Avitabile, V., Herold, M., Heuvelink, G., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N. et al. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global change biology*, .
- Baccini, A., Friedl, M., Woodcock, C., & Zhu, Z. (2007). Scaling field data to calibrate and validate moderate spatial resolution remote sensing models. *Photogrammetric Engineering & Remote Sensing*, *73*, 945–954.
- Baccini, A., Goetz, S., Walker, W., Laporte, N., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P., Dubayah, R., Friedl, M. et al. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, *2*, 182–185.
- Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, *15*, 453–472.
- Bai, J. (1997a). Estimating multiple breaks one at a time. *Econometric Theory*, *13*, 315–352.
- Bai, J. (1997b). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, *79*, 551–563.
- Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, (pp. 47–78).
- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, *18*, 1–22.
- Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, *65*, 2–16.
- Bonan, G. B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *science*, *320*, 1444–1449.
- Bongers, F., Chazdon, R., Poorter, L., & Peña-Claros, M. (2015). The potential of secondary forests. *Science (New York, NY)*, *348*, 642–643.
- Brede, B., Verbesselt, J., Dutrieux, L. P., & Herold, M. (2015). Performance of the enhanced vegetation index to detect inner-annual dry season and drought impacts on Amazon forest canopies. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *40-7/W3*, 337–344.

- Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*, 123–140.
- Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5–32.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. *The Wadsworth statistics/probability series*, .
- Broadbent, E. N., Asner, G. P., Keller, M., Knapp, D. E., Oliveira, P. J., & Silva, J. N. (2008a). Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biological conservation*, *141*, 1745–1757.
- Broadbent, E. N., Asner, G. P., Peña-Claros, M., Palace, M., & Soriano, M. (2008b). Spatial partitioning of biomass and diversity in a lowland Bolivian forest: Linking field and remote sensing measurements. *Forest Ecology and Management*, *255*, 2602–2616.
- Brooks, E. B., Wynne, R. H., Thomas, V. A., Blinn, C. E., & Coulston, J. W. (2014). On-the-fly massively multitemporal change detection using statistical quality control charts and Landsat data. *Geoscience and Remote Sensing, IEEE Transactions on*, *52*, 3316–3332.
- Brown, S., Lugo, A. E. et al. (1990). Tropical secondary forests. *Journal of tropical ecology*, *6*, 1–32.
- Chambers, J. Q., Asner, G. P., Morton, D. C., Anderson, L. O., Saatchi, S. S., Espírito-Santo, F. D., Palace, M., & Souza Jr, C. (2007). Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. *Trends in Ecology & Evolution*, *22*, 414–423.
- Chase, A. F., & Chase, D. Z. (1994). Details in the archaeology of caracol, Belize: an Introduction. *Studies in the Archaeology of Caracol, Belize*, (pp. 1–11).
- Chave, J. (2008). Spatial variation in tree species composition across tropical forests: pattern and process. In *Tropical Forest Community Ecology* (pp. 11–30). Wiley-Blackwell Chichester, UK.
- Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B., Duque, A., Eid, T., Fearnside, P. M., Goodman, R. C. et al. (2014). Improved allometric models to estimate the aboveground biomass of tropical trees. *Global change biology*, *20*, 3177–3190.
- Chazdon, R. L., Broadbent, E. N., Rozendaal, D. M., Bongers, F., Zambrano, A. M. A., Aide, T. M., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H. et al. (2016). Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances*, *2*, e1501639.
- Chazdon, R. L., Letcher, S. G., Van Breugel, M., Martínez-Ramos, M., Bongers, F., & Finegan, B. (2007). Rates of change in tree communities of secondary Neotropical forests following major disturbances. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*, 273–289.

- Chu, C.-S. J., Stinchcombe, M., & White, H. (1996). Monitoring structural change. *Econometrica: Journal of the Econometric Society*, (pp. 1045–1065).
- Cohen, W. B., Yang, Z., & Kennedy, R. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. *Remote Sensing of Environment*, *114*, 2911–2924.
- Cook, B., Anchukaitis, K., Kaplan, J., Puma, M., Kelley, M., & Gueyffier, D. (2012). Pre-Columbian deforestation as an amplifier of drought in Mesoamerica. *Geophysical Research Letters*, *39*.
- Coomes, O. T., Grimard, F., & Burt, G. J. (2000). Tropical forests and shifting cultivation: secondary forest fallow dynamics among traditional farmers of the Peruvian Amazon. *Ecological Economics*, *32*, 109–124.
- Crist, E. P. (1985). A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment*, *17*, 301–306.
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote sensing of Environment*, *30*, 271–278.
- De Keersmaecker, W., Lhermitte, S., Tits, L., Honnay, O., Somers, B., & Coppin, P. (2015). A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover. *Global Ecology and Biogeography*, *24*, 539–548.
- De Sy, V., Herold, M., Achard, F., Asner, G. P., Held, A., Kelldorfer, J., & Verbesselt, J. (2012). Synergies of multiple remote sensing data sources for REDD+ monitoring. *Current Opinion in Environmental Sustainability*, *4*, 696–706.
- DeFries, R., & Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*, *15*, 3567–3586.
- Desclée, B., Bogaert, P., & Defourny, P. (2006). Forest change detection by statistical object-based method. *Remote Sensing of Environment*, *102*, 1–11.
- DeVries, B., Decuyper, M., Verbesselt, J., Zeileis, A., Herold, M., & Joseph, S. (2015a). Tracking disturbance-regrowth dynamics in tropical forests using structural change detection and Landsat time series. *Remote Sensing of Environment*, *169*, 320–334.
- DeVries, B., Verbesselt, J., Kooistra, L., & Herold, M. (2015b). Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sensing of Environment*, *161*, 107–121.
- Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S., & Ropelewski, C. (2007). Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, *28*, 1503–1526.
- Dorren, L. K., Maier, B., & Seijmonsbergen, A. C. (2003). Improved Landsat-based forest mapping in steep mountainous terrain using object-based classification. *Forest Ecology*

- and Management*, 183, 31–46.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. et al. (2012). Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, 120, 25–36.
- Dubayah, R., Goetz, S., Blair, J., Fatoyinbo, T., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S. et al. (2014). The Global Ecosystem Dynamics Investigation. In *AGU Fall Meeting Abstracts* (p. 07). volume 1.
- Dutrieux, L. P., Bartholomeus, H., Herold, M., & Verbesselt, J. (2012). Relationships between declining summer sea ice, increasing temperatures and changing vegetation in the Siberian Arctic tundra from MODIS time series (2000–11). *Environmental Research Letters*, 7, 044028.
- Dutrieux, L. P., Jakovac, C. C., Latifah, S. H., & Kooistra, L. (2016). Reconstructing land use history from Landsat time-series: Case study of a swidden agriculture system in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 47, 112–124.
- Dutrieux, L. P., Verbesselt, J., Kooistra, L., & Herold, M. (2015). Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107, 112–125.
- Duveiller, G., Defourny, P., Desclée, B., & Mayaux, P. (2008). Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. *Remote Sensing of Environment*, 112, 1969–1981.
- Emilio, T., Quesada, C. A., Costa, F. R., Magnusson, W. E., Schietti, J., Feldpausch, T. R., Brienen, R. J., Baker, T. R., Chave, J., Álvarez, E. et al. (2014). Soil physical conditions limit palm and tree basal area in Amazonian forests. *Plant Ecology & Diversity*, 7, 215–229.
- Ernst, C., Mayaux, P., Verhegghen, A., Bodart, C., Christophe, M., & Defourny, P. (2013). National forest cover change in Congo Basin: deforestation, reforestation, degradation and regeneration for the years 1990, 2000 and 2005. *Global change biology*, 19, 1173–1187.
- Fagan, M. E., DeFries, R. S., Sesnie, S. E., Arroyo-Mora, J. P., Soto, C., Singh, A., Townsend, P. A., & Chazdon, R. L. (2015). Mapping Species Composition of Forests and Tree Plantations in Northeastern Costa Rica with an Integration of Hyperspectral and Multitemporal Landsat Imagery. *Remote Sensing*, 7, 5660–5696.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L. et al. (2007). The shuttle radar topography mission. *Reviews of geophysics*, 45.

- Fensholt, R., & Proud, S. R. (2012). Evaluation of earth observation based global long term vegetation trends—Comparing GIMMS and MODIS global NDVI time series. *Remote sensing of Environment*, *119*, 131–147.
- Féret, J.-B., & Asner, G. P. (2013). Tree species discrimination in tropical forests using airborne imaging spectroscopy. *Geoscience and Remote Sensing, IEEE Transactions on*, *51*, 73–84.
- de Filho, F. J. B. O., & Metzger, J. P. (2006). Thresholds in landscape structure for three common deforestation patterns in the Brazilian Amazon. *Landscape Ecology*, *21*, 1061–1073.
- Fine, P. V., Ree, R. H., & Burnham, R. J. (2009). The disparity in tree species richness among tropical, temperate and boreal biomes: the geographic area and age hypothesis. *Tropical Forest Community Ecology*, (pp. 31–45).
- Finegan, B., Peña-Claros, M., Oliveira, A., Ascarrunz, N., Bret-Harte, M. S., Carreño-Rocabado, G., Casanoves, F., Díaz, S., Eguiguren Velepucha, P., Fernandez, F. et al. (2015). Does functional trait diversity predict above-ground biomass and productivity of tropical forests? Testing three alternative hypotheses. *Journal of Ecology*, *103*, 191–201.
- 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. et al. (2005). Global consequences of land use. *science*, *309*, 570–574.
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote sensing of environment*, *80*, 185–201.
- Forkel, M., Carvalhais, N., Verbesselt, J., Mahecha, M. D., Neigh, C. S., & Reichstein, M. (2013). Trend change detection in NDVI time series: Effects of inter-annual variability and methodology. *Remote Sensing*, *5*, 2113–2144.
- Franchito, S. H., Rao, V. B., Vasques, A. C., Santo, C. M., & Conforte, J. C. (2009). Validation of TRMM precipitation radar monthly rainfall estimates over Brazil. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *114*.
- Fricker, G. A., Wolf, J. A., Saatchi, S. S., & Gillespie, T. W. (2015). Predicting spatial variations of tree species richness in tropical forests from high resolution remote sensing. *Ecological Applications*, *25*, 1776–1789.
- Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, *21*, 32–40.
- Gessner, U., Machwitz, M., Conrad, C., & Dech, S. (2013). Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. *Remote Sensing of Environment*, *129*, 90–102.

- Goward, S., Arvidson, T., Williams, D., Faundeen, J., Irons, J., & Franks, S. (2006). Historical Record of Landsat Global Coverage. *Photogrammetric Engineering & Remote Sensing*, *72*, 1155–1169.
- Gurney, C. M., & Townshend, J. R. (1983). The use of contextual information in the classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, *49*, 55–64.
- Hansen, M. C., DeFries, R., Townshend, J., Carroll, M., Dimiceli, C., & Sohlberg, R. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, *7*, 1–15.
- Hansen, M. C., Egorov, A., Roy, D. P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., & Loveland, T. R. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters*, *2*, 279–288.
- Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., & Moore, R. (2016). Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, *11*, 034008.
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T. et al. (2013). High-resolution global maps of 21st-century forest cover change. *science*, *342*, 850–853.
- Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., & Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, *112*, 2495–2513.
- Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*, 100–108.
- He, L., Chen, J. M., Pisek, J., Schaaf, C. B., & Strahler, A. H. (2012). Global clumping index map derived from the MODIS BRDF product. *Remote Sensing of Environment*, *119*, 118–130.
- Hengl, T., Heuvelink, G. B., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, *120*, 75–93.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*, *25*, 1965–1978.
- Hill, M. O., & Gauch Jr, H. G. (1980). Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, *42*, 47–58.
- Hirota, M., Holmgren, M., Van Nes, E. H., & Scheffer, M. (2011). Global resilience of tropical forest and savanna to critical transitions. *Science*, *334*, 232–235.

- Hodell, D., Curtis, J., & Brenner, M. (1995). Possible Role of Climate in the Collapse of Classic Maya Civilization. *Nature*, *375*, 391–394.
- Holling, C. S. (1996). Engineering resilience versus ecological resilience. *Engineering within ecological constraints*, *31*, 32.
- Horler, D., & Ahern, F. (1986). Forestry information content of Thematic Mapper data. *International Journal of Remote Sensing*, *7*, 405–428.
- Horn, B. K. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*, 14–47.
- Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A., & Romijn, E. (2012). An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters*, *7*, 044009.
- Huang, C., Goward, S. N., Masek, J. G., Thomas, N., Zhu, Z., & Vogelmann, J. E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*, 183–198.
- Huang, C., Song, K., Kim, S., Townshend, J. R., Davis, P., Masek, J. G., & Goward, S. N. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, *112*, 970–985.
- Huang, X., & Friedl, M. A. (2014). Distance metric-based forest cover change detection using MODIS time series. *International Journal of Applied Earth Observation and Geoinformation*, *29*, 78–92.
- 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.
- Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyyra, L. R., Yang, W., Nemani, R. R., & Myneni, R. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*.
- Inglada, J., & Christophe, E. (2009). The Orfeo Toolbox Remote Sensing Image Processing Software. In *IGARSS (4)* (pp. 733–736).
- Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, *11*, 185–200.
- Jakovac, C. C. (2015). *Resilience of Amazonian Landscapes to agricultural intensification*. Ph.D. thesis Wageningen University Wageningen, The Netherlands.
- Jakovac, C. C., Bongers, F., Kuyper, T. W., Mesquita, R. C. G., & Peña-Claros, M. (2016a). Land use as a filter for species composition in Amazonian secondary forests, . In review.
- Jakovac, C. C., Peña-Claros, M., Kuyper, T. W., & Bongers, F. (2015). Loss of secondary-

- forest resilience by land-use intensification in the Amazon. *Journal of Ecology*, *103*, 67–77.
- Jakovac, C. C., Siti, L., Dutrieux, L. P., Peña-Claros, M., & Bongers, F. (2016b). Spatial and temporal dynamics of swidden cultivation in the middle-Amazonas river: expansion and intensification, . In prep.
- Jin, S., & Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote sensing of Environment*, *94*, 364–372.
- de Jong, R., de Bruin, S., de Wit, A., Schaepman, M. E., & Dent, D. L. (2011). Analysis of monotonic greening and browning trends from global NDVI time-series. *Remote Sensing of Environment*, *115*, 692–702.
- Jong, R., Schaepman, M. E., Furrer, R., Bruin, S., & Verburg, P. H. (2013). Spatial relationship between climatologies and changes in global vegetation activity. *Global change biology*, *19*, 1953–1964.
- de Jong, R., Verbesselt, J., Zeileis, A., & Schaepman, M. E. (2013). Shifts in global vegetation activity trends. *Remote Sensing*, *5*, 1117–1133.
- Justice, C. (1978). An examination of the relationships between selected ground properties and Landsat MSS data in an area of complex terrain in Southern Italy, .
- Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., & Morisette, J. (2002). An overview of MODIS Land data processing and product status. *Remote sensing of Environment*, *83*, 3–15.
- Kaufmann, H., Segl, K., Guanter, L., Hofer, S., Foerster, K.-P., Stuffer, T., Mueller, A., Richter, R., Bach, H., Hostert, P. et al. (2008). Environmental Mapping and Analysis Program (EnMAP)-Recent advances and status. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International* (pp. IV–109). IEEE volume 4.
- Kennedy, R. E., Andréfouët, S., Cohen, W. B., Gómez, C., Griffiths, P., Hais, M., Healey, S. P., Helmer, E. H., Hostert, P., Lyons, M. B. et al. (2014). Bringing an ecological view of change to Landsat-based remote sensing. *Frontiers in Ecology and the Environment*, *12*, 339–346.
- Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, *114*, 2897–2910.
- Ketterings, Q. M., Coe, R., van Noordwijk, M., Palm, C. A. et al. (2001). Reducing uncertainty in the use of allometric biomass equations for predicting above-ground tree biomass in mixed secondary forests. *Forest Ecology and Management*, *146*, 199–209.
- Killeen, T. J., Calderon, V., Soria, L., Quezada, B., Steinger, M. K., Harper, G., Solórzano, L. A., & Tucker, C. J. (2007). Thirty years of land-cover change in Bolivia.

- AMBIO: A journal of the Human Environment*, 36, 600–606.
- Kleynhans, W., Olivier, J. C., Wessels, K. J., Salmon, B. P., Van den Bergh, F., & Steenkamp, K. (2011). Detecting land cover change using an extended Kalman filter on MODIS NDVI time-series data. *Geoscience and Remote Sensing Letters, IEEE*, 8, 507–511.
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1, 155–159.
- Knyazikhin, Y., Glassy, J., Privette, J., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J., Votava, P., Myneni, R. et al. (1999). MODIS leaf area index (LAI) and fraction of photosynthetically active radiation absorbed by vegetation (FPAR) product (MOD15) algorithm theoretical basis document. *Theoretical Basis Document, NASA Goddard Space Flight Center, Greenbelt, MD, 20771*.
- Lang, C., Costa, F. R. C., Camargo, J. L. C., Durgante, F. M., & Vicentini, A. (2015). Near Infrared Spectroscopy Facilitates Rapid Identification of Both Young and Mature Amazonian Tree Species. *PLoS one*, 10, e0134521.
- Laurin, G. V., Chan, J. C.-W., Chen, Q., Lindsell, J. A., Coomes, D. A., Guerriero, L., Del Frate, F., Miglietta, F., & Valentini, R. (2014). Biodiversity mapping in a tropical West African forest with airborne hyperspectral data, .
- Lawrence, D., Radel, C., Tully, K., Schmook, B., & Schneider, L. (2010). Untangling a decline in tropical forest resilience: Constraints on the sustainability of shifting cultivation across the globe. *Biotropica*, 42, 21–30.
- Le Quéré, C., Moriarty, R., Andrew, R., Canadell, J., Sitch, S., Korsbakken, J., Friedlingstein, P., Peters, G., Andres, R., Boden, T. et al. (2015). Global Carbon Budget 2015. *Earth System Science Data*, 7, 349–396.
- Le Toan, T., Quegan, S., Davidson, M., Balzter, H., Paillou, P., Papathanassiou, K., Plummer, S., Rocca, F., Saatchi, S., Shugart, H. et al. (2011). The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote sensing of environment*, 115, 2850–2860.
- Leisch, F., Hornik, K., & Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16, 835–854.
- Lhermitte, S., Verbesselt, J., Jonckheere, I., Nackaerts, K., van Aardt, J. A., Verstraeten, W. W., & Coppin, P. (2008). Hierarchical image segmentation based on similarity of NDVI time series. *Remote Sensing of Environment*, 112, 506–521.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2, 18–22.
- qing Liu, H., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *Geoscience and Remote Sensing, IEEE*

- Transactions on*, 33, 457–465.
- Loague, K., & Green, R. E. (1991). Statistical and graphical methods for evaluating solute transport models: overview and application. *Journal of Contaminant Hydrology*, 7, 51–73.
- Longworth, J. B., Mesquita, R. C., Bentos, T. V., Moreira, M. P., Massoca, P. E., & Williamson, G. B. (2014). Shifts in dominance and species assemblages over two decades in alternative successions in Central Amazonia. *Biotropica*, 46, 529–537.
- Lu, D., Li, G., & Moran, E. (2014). Current situation and needs of change detection techniques. *International Journal of Image and Data Fusion*, 5, 13–38.
- Maass, J., Vose, J. M., Swank, W. T., & Martínez-Yrizar, A. (1995). Seasonal changes of leaf area index (LAI) in a tropical deciduous forest in west Mexico. *Forest Ecology and Management*, 74, 171–180.
- Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W., & Nobre, C. A. (2008). Climate change, deforestation, and the fate of the Amazon. *science*, 319, 169–172.
- Mascaro, J., Asner, G. P., Davies, S., Dehgan, A., & Saatchi, S. (2014). These are the days of lasers in the jungle. *Carbon Balance and Management*, 9, 1–3.
- Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., Gao, F., Kutler, J., & Lim, T.-K. (2006). A Landsat surface reflectance dataset for North America, 1990–2000. *Geoscience and Remote Sensing Letters, IEEE*, 3, 68–72.
- Mayer, D., & Butler, D. (1993). Statistical validation. *Ecological modelling*, 68, 21–32.
- McDowell, N. G., & Allen, C. D. (2015). Darcy's law predicts widespread forest mortality under climate warming. *Nature Climate Change*, .
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (pp. 179–183). American Meteorological Society Boston, MA volume 17.
- Mertz, O. (2009). Trends in shifting cultivation and the REDD mechanism. *Current Opinion in Environmental Sustainability*, 1, 156–160.
- Mesquita, R. C., Ickes, K., Ganade, G., & Williamson, G. B. (2001). Alternative successional pathways in the Amazon Basin. *Journal of Ecology*, 89, 528–537.
- Metzger, J. P. (2003). Effects of slash-and-burn fallow periods on landscape structure. *Environmental Conservation*, 30, 325–333.
- Morton, D., Nagol, J., Carabajal, C., Rosette, J., Palace, M., Cook, B., Vermote, E., Harding, D., & North, P. (2014). Amazon forests maintain consistent canopy structure and greenness during the dry season. *Nature*, 506, 221.
- Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Del Bon Espirito-Santo, F., Hansen, M., & Carroll, M. (2005). Rapid assessment of annual deforestation

- in the Brazilian Amazon using MODIS data. *Earth Interactions*, 9, 1–22.
- Mostacedo, B., Putz, F. E., Fredericksen, T. S., Villca, A., & Palacios, T. (2009). Contributions of root and stump sprouts to natural regeneration of a logged tropical dry forest in Bolivia. *Forest ecology and management*, 258, 978–985.
- Müller, R., Müller, D., Schierhorn, F., Gerold, G., & Pacheco, P. (2012). Proximate causes of deforestation in the Bolivian lowlands: an analysis of spatial dynamics. *Regional Environmental Change*, 12, 445–459.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.
- Myneni, R., & Williams, D. (1994). On the relationship between FAPAR and NDVI. *Remote Sensing of Environment*, 49, 200–211.
- Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Juárez, R. I. N., Saatchi, S. S. et al. (2007). Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Sciences*, 104, 4820–4823.
- Navarro, G., & Maldonado, M. (2002). *Geografía ecológica de Bolivia: Vegetación y ambientes acuáticos*. Centro de Ecología Simón I. Patiño, Departamento de Difusión.
- Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., & Running, S. W. (2003). Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *science*, 300, 1560–1563.
- Nicholson, S. E., Davenport, M. L., & Malo, A. R. (1990). A comparison of the vegetation response to rainfall in the Sahel and East Africa, using normalized difference vegetation index from NOAA AVHRR. *Climatic Change*, 17, 209–241.
- Nicholson, S. E., Some, B., McCollum, J., Nelkin, E., Klotter, D., Berte, Y., Diallo, B., Gaye, I., Kpabeba, G., Ndiaye, O. et al. (2003). Validation of TRMM and other rainfall estimates with a high-density gauge dataset for West Africa. Part ii: Validation of TRMM rainfall products. *Journal of Applied Meteorology*, 42, 1355–1368.
- Oglesby, R. J., Sever, T. L., Saturno, W., Erickson, D. J., & Srikishen, J. (2010). Collapse of the Maya: Could deforestation have contributed? *Journal of Geophysical Research: Atmospheres*, 115.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C. et al. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51, 933–938.
- Ometto, J. P., Aguiar, A. P., Assis, T., Soler, L., Valle, P., Tejada, G., Lapola, D. M., & Meir, P. (2015). Amazon forest biomass density maps: tackling the uncertainty

- in carbon emission estimates. In *Uncertainties in Greenhouse Gas Inventories* (pp. 95–110). Springer.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*, 217–222.
- Pereira, H. M., Ferrier, S., Walters, M., Geller, G., Jongman, R., Scholes, R., Bruford, M. W., Brummitt, N., Butchart, S., Cardoso, A. et al. (2013). Essential biodiversity variables. *Science*, *339*, 277–278.
- Pesaran, M. H., & Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, *9*, 495–510.
- Peterson, G., Allen, C. R., & Holling, C. S. (1998). Ecological resilience, biodiversity, and scale. *Ecosystems*, *1*, 6–18.
- Pettorelli, N., Wegmann, M., Skidmore, A., Múcher, S., Dawson, T. P., Fernandez, M., Lucas, R., Schaepman, M. E., Wang, T., O'Connor, B. et al. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing in Ecology and Conservation*, .
- Pimm, S. L. (1984). The complexity and stability of ecosystems. *Nature*, *307*, 321–326.
- Pimm, S. L., Jenkins, C. N., Abell, R., Brooks, T. M., Gittleman, J. L., Joppa, L. N., Raven, P. H., Roberts, C. M., & Sexton, J. O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, *344*, 1246752.
- Poorter, L., Bongers, F., Aide, T. M., Zambrano, A. M. A., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H., Broadbent, E. N., Chazdon, R. L. et al. (2016). Biomass resilience of Neotropical secondary forests. *Nature*, *530*, 211–214.
- Poorter, L., Bongers, L., & Bongers, F. (2006). Architecture of 54 moist-forest tree species: traits, trade-offs, and functional groups. *Ecology*, *87*, 1289–1301.
- Poorter, L., Sande, M., Thompson, J., Arets, E., Alarcón, A., Álvarez-Sánchez, J., Ascarunz, N., Balvanera, P., Barajas-Guzmán, G., Boit, A. et al. (2015). Diversity enhances carbon storage in tropical forests. *Global Ecology and Biogeography*, *24*, 1314–1328.
- Popkin, G. (2016). Satellite alerts track deforestation in real time. *Nature*, *530*, 392.
- R Core Team (2014). *R: A Language and Environment for Statistical Computng*. R Foundation for Statistical Computing Vienna, Austria.
- Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., Kellndorfer, J. M., Rosenqvist, A., Lehmann, E. A., Woodcock, C. E. et al. (2016). Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, *6*, 120–122.
- Reiche, J., Verbesselt, J., Hoekman, D., & Herold, M. (2015). Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sensing of Environment*, *156*,

- 276–293.
- Rivera Caicedo, J., Verrelst, J., Munoz-Mari, J., Moreno, J., & Camps-Valls, G. (2014). Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 7, 1249–1259.
- Roerink, G., Menenti, M., Soepboer, W., & Su, Z. (2003). Assessment of climate impact on vegetation dynamics by using remote sensing. *Physics and Chemistry of the Earth, Parts A/B/C*, 28, 103–109.
- Roy, D. P., Wulder, M., Loveland, T., Woodcock, C., Allen, R., Anderson, M., Helder, D., Irons, J., Johnson, D., Kennedy, R. et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, 54, 547–560.
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S. et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108, 9899–9904.
- Sakschewski, B., Bloh, W., Boit, A., Rammig, A., Kattge, J., Poorter, L., Peñuelas, J., & Thonicke, K. (2015). Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. *Global Change Biology*, 21, 2711–2725.
- Sales, M. H., Souza, C. M., Kyriakidis, P. C., Roberts, D. A., & Vidal, E. (2007). Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil. *Ecological Modelling*, 205, 221–230.
- Saleska, S. R., Didan, K., Huete, A. R., & Da Rocha, H. R. (2007). Amazon forests green-up during 2005 drought. *Science*, 318, 612–612.
- Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., Nemani, R. R., & Myneni, R. B. (2010). Amazon forests did not green-up during the 2005 drought. *Geophysical Research Letters*, 37.
- Samanta, A., Ganguly, S., Vermote, E., Nemani, R. R., & Myneni, R. B. (2012). Why is remote sensing of Amazon forest greenness so challenging? *Earth Interactions*, 16, 1–14.
- van der Sande, M., Arets, E., Peña-Claros, M. et al. (2015). Soil fertility and species traits, but not species diversity, drive productivity and biomass stocks in a tropical rainforest, . Submitted for publication.
- Schnitzer, S. A., & Bongers, F. (2002). The ecology of lianas and their role in forests. *Trends in Ecology & Evolution*, 17, 223–230.

- Schnitzer, S. A., & Bongers, F. (2011). Increasing liana abundance and biomass in tropical forests: emerging patterns and putative mechanisms. *Ecology letters*, *14*, 397–406.
- Sexton, J. O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., Anand, A., Huang, C., Channan, S., Pimm, S. L. et al. (2015). Conservation policy and the measurement of forests. *Nature Climate Change*, .
- Silva, F. B., Shimabukuro, Y. E., Aragão, L. E., Anderson, L. O., Pereira, G., Cardozo, F., & Arai, E. (2013). Large-scale heterogeneity of Amazonian phenology revealed from 26-year long AVHRR/NDVI time-series. *Environmental Research Letters*, *8*, 024011.
- Simard, M., Pinto, N., Fisher, J. B., & Baccini, A. (2011). Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, *116*.
- Skidmore, A. K., Pettoirelli, N., Coops, N. C., Geller, G. N., Hansen, M., Lucas, R., Mùcher, C. A., O'Connor, B., Paganini, M., Pereira, H. M. et al. (2015). Environmental science: agree on biodiversity metrics to track from space. *Nature*, *523*, 403–405.
- Slik, J., Paoli, G., McGuire, K., Amaral, I., Barroso, J., Bastian, M., Blanc, L., Bongers, F., Boundja, P., Clark, C. et al. (2013). Large trees drive forest aboveground biomass variation in moist lowland forests across the tropics. *Global Ecology and Biogeography*, *22*, 1261–1271.
- Slik, J. F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L. F., Ashton, P., Balvanera, P., Bastian, M. L., Bellingham, P. J., Van Den Berg, E. et al. (2015). An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences*, *112*, 7472–7477.
- Somers, B., & Asner, G. P. (2012). Hyperspectral time series analysis of native and invasive species in Hawaiian rainforests. *Remote Sensing*, *4*, 2510–2529.
- Somers, B., & Asner, G. P. (2014). Tree species mapping in tropical forests using multi-temporal imaging spectroscopy: Wavelength adaptive spectral mixture analysis. *International Journal of Applied Earth Observation and Geoinformation*, *31*, 57–66.
- Souza, C. M., Roberts, D. A., & Cochrane, M. A. (2005). Combining spectral and spatial information to map canopy damage from selective logging and forest fires. *Remote Sensing of Environment*, *98*, 329–343.
- Spies, T. A. (1998). Forest structure: a key to the ecosystem. *Northwest science*, *72*, 34–36.
- Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., & Midgley, P. M. (2014). *Climate change 2013: The physical science basis: Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press Cambridge, UK, and New York.

- Strahler, A. H., Muller, J., Lucht, W., Schaaf, C., Tsang, T., Gao, F., Li, X., Lewis, P., & Barnsley, M. J. (1999). MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0. *MODIS documentation*, .
- Svetnik, V., Liaw, A., Tong, C., & Wang, T. (2004). Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In *Multiple Classifier Systems* (pp. 334–343). Springer.
- Ter Steege, H., Pitman, N. C., Phillips, O. L., Chave, J., Sabatier, D., Duque, A., Molino, J.-F., Prévost, M.-F., Spichiger, R., Castellanos, H. et al. (2006). Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature*, *443*, 444–447.
- Ter Steege, H., Pitman, N. C., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., Phillips, O. L., Castilho, C. V., Magnusson, W. E., Molino, J.-F. et al. (2013). Hyperdominance in the Amazonian tree flora. *Science*, *342*, 1243092.
- TerraClass (2011). *Sumário de informações de uso e cobertura da terra na Amazônia*. Technical Report, EMBRAPA and INPE.
- Toledo, M. (2010). *Neotropical lowland forests along environmental gradients*. Ph.D. thesis Wageningen University Wageningen, The Netherlands.
- Toledo, M., Peña-Claros, M., Bongers, F., Alarcón, A., Balcázar, J., Chuviña, J., Leano, C., Licona, J. C., & Poorter, L. (2012). Distribution patterns of tropical woody species in response to climatic and edaphic gradients. *Journal of Ecology*, *100*, 253–263.
- Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Chuviña, J., Leano, C., Licona, J. C., ter Steege, H., & Bongers, F. (2011a). Patterns and determinants of floristic variation across lowland forests of Bolivia. *Biotropica*, *43*, 405–413.
- Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Leano, C., Licona, J. C., & Bongers, F. (2011b). Climate and soil drive forest structure in Bolivian lowland forests. *Journal of Tropical Ecology*, *27*, 333–345.
- Trenberth, K. E., Dai, A., van der Schrier, G., Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, *4*, 17–22.
- Tropek, R., Sedláček, O., Beck, J., Keil, P., Musilová, Z., Šímová, I., & Storch, D. (2014). Comment on “High-resolution global maps of 21st-century forest cover change”. *Science*, *344*, 981–981.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, *8*, 127–150.
- Tucker, C. J., Townshend, J. R., & Goff, T. E. (1985). African land-cover classification using satellite data. *Science*, *227*, 369–375.
- Van Vliet, N., Mertz, O., Heinemann, A., Langanke, T., Pascual, U., Schmook, B., Adams,

- C., Schmidt-Vogt, D., Messerli, P., Leisz, S. et al. (2012). Trends, drivers and impacts of changes in swidden cultivation in tropical forest-agriculture frontiers: a global assessment. *Global Environmental Change*, *22*, 418–429.
- Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010a). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114*, 106–115.
- Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010b). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, *114*, 2970–2980.
- Verbesselt, J., Zeileis, A., & Herold, M. (2012). Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, *123*, 98–108.
- Verbesselt, J., Zeileis, A., & Hyndman, R. (2014). *bfast: Breaks For Additive Season and Trend (BFAST)*. R package version 1.5.7.
- Verhegghen, A., Ernst, C., Defourny, P., & Beuchle, R. (2010). Automated land cover mapping and independent change detection in tropical forest using multi-temporal high resolution data set. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, GEOBIA, Ghent, Belgium*, *38*, C7.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, .
- Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., Beguería, S., Trigo, R., López-Moreno, J. I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J. et al. (2013). Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences*, *110*, 52–57.
- Vieira, I. C. G., de Almeida, A. S., Davidson, E. A., Stone, T. A., de Carvalho, C. J. R., & Guerrero, J. B. (2003). Classifying successional forests using Landsat spectral properties and ecological characteristics in eastern Amazonia. *Remote Sensing of Environment*, *87*, 470–481.
- Vogelmann, J. E., & Rock, B. N. (1988). Assessing forest damage in high-elevation coniferous forests in Vermont and New Hampshire using Thematic Mapper data. *Remote Sensing of Environment*, *24*, 227–246.
- Wahl, D., Byrne, R., Schreiner, T., & Hansen, R. (2006). Holocene vegetation change in the northern Peten and its implications for Maya prehistory. *Quaternary Research*, *65*, 380–389.
- Wang, J., Rich, P., & Price, K. (2003). Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *International Journal of Remote*

- Sensing*, 24, 2345–2364.
- de Wasseige, C., & Defourny, P. (2004). Remote sensing of selective logging impact for tropical forest management. *Forest Ecology and Management*, 188, 161–173.
- Watt, A. S. (1947). Pattern and process in the plant community. *Journal of Ecology*, 35, 1–22.
- Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80, 385–396.
- Wilson, M. F., O’Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multi-scale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30, 3–35.
- Woodcock, C. E., & Strahler, A. H. (1987). The factor of scale in remote sensing. *Remote sensing of Environment*, 21, 311–332.
- 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.
- Xiao, X., Hagen, S., Zhang, Q., Keller, M., & Moore III, B. (2006). Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images. *Remote Sensing of Environment*, 103, 465–473.
- Zarin, D. J., Davidson, E. A., Brondizio, E., Vieira, I. C., Sá, T., Feldpausch, T., Schuur, E. A., Mesquita, R., Moran, E., Delamonica, P. et al. (2005). Legacy of fire slows carbon accumulation in Amazonian forest regrowth. *Frontiers in Ecology and the Environment*, 3, 365–369.
- Zeileis, A. (2005). A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals. *Econometric Reviews*, 24, 445–466.
- Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, 44, 109–123.
- Zeileis, A., Leisch, F., Kleiber, C., & Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, 20, 99–121.
- Zeileis, A., Shah, A., & Patnaik, I. (2010). Testing, monitoring, and dating structural changes in exchange rate regimes. *Computational Statistics & Data Analysis*, 54, 1696–1706.
- 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. (2014). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144, 152–171.

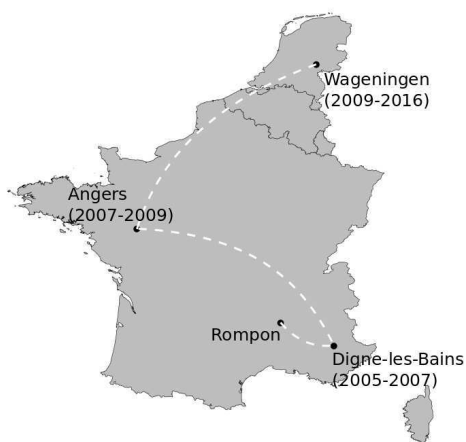
Zhu, Z., Woodcock, C. E., & Olofsson, P. (2012). Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sensing of Environment*, 122, 75–91.

Curriculum Vitae

Loïc (born 1987) grew up in Rompon, a small village of the Ardèche district in southern France. Growing up in a rural context, in proximity with the nature, he developed an interest for natural environments from a young age. After finishing the agricultural high school in Aubenas in 2005, he joined a two years university program in Digne-les-Bains. There, he learned about a wide range of topics relating to biology and environmental engineering. The mediterranean-alpine context of the region, allowed him to prepare his mountain guide license which he took four years to complete, in parallel to his study.



Upon completion of his university degree Loïc chose to continue his studies and was admitted at the *Ecole Supérieure d'Agriculture* in Angers. He stayed there for two years, studying agronomy and landscape ecology, and traveling to the tropics for the first time, in the frame of an internship in Belize.



While still enrolled in Angers, he joined Wageningen University in 2009 as a double degree student with the aim to specialize in irrigation and water management. It is during this period that he discovered remote sensing and decided to articulate his career around that tool. After spending four months in Mali to conduct research with rice farmers for his major thesis, and working on Arctic vegetation dynamics for his minor thesis, he graduated from his two masters in 2011.

Early 2012 he started a PhD at the Laboratory of Geo-Information Sciences and Remote Sensing (Wageningen University). Under the umbrella of a large international interdisciplinary research project on biodiversity, he

developed and tested approaches to map biodiversity and dynamics of tropical forests using remote sensing. Latin America being the regional focus of the project, Loïc traveled to Bolivia, Mexico and Brazil to carry out field work and collaborate with the project partners. The thesis resulted in four scientific articles, two of which have already been published. Additionally to his thesis he was actively involved in education, co-developing and teaching MSc level courses, and in the development of open source software to work with satellite data. Loïc also presented his work at international conferences, including the American Geophysical Union fall meeting in San Francisco, where he won the Outstanding Student Paper Award in the biogeosciences category.

Peer reviewed publications

Dutrieux, L.P., Jakovac, C.C., Latifah, S.H. and Kooistra, L., 2016. Reconstructing land use history from Landsat time-series: Case study of a swidden agriculture system in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 47, pp.112-124.

Dutrieux, L.P., Verbesselt, J., Kooistra, L., & Herold, M. (2015). Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107, 112-125.

Poorter, L., Sande, M.T., Thompson, J., Arets, E.J.M.M., Alarcón, A., ÁlvarezSánchez, J., Ascarrunz, N., Balvanera, P., BarajasGuzmán, G., Boit, A., Bongers, F., **et al.** 2015. Diversity enhances carbon storage in tropical forests. *Global Ecology and Biogeography*, 24(11), pp.1314-1328.

Brede, B., Verbesselt, J., **Dutrieux, L.P.** and Herold, M., 2015. Performance of the Enhanced Vegetation Index to Detect Inner-annual Dry Season and Drought Impacts on Amazon Forest Canopies. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(7), p.337.

Dutrieux, L.P., Bartholomeus, H., Herold, M. and Verbesselt, J. (2012). Relationships between declining summer sea ice, increasing temperatures and changing vegetation in the Siberian Arctic tundra from MODIS time series (2000–11). *Environmental Research Letters*, 7(4), p.044028.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review of literature (6 ECTS)

- Remote sensing based monitoring of tropical forest biodiversity: examples from research and opportunities

Writing of project proposal (4.5 ECTS)

- Spatio-temporal data analysis for monitoring and assessment of biodiversity and forest dynamics in tropical South-America

Post-graduate courses (3 ECTS)

- REDD + Science + Governance, opportunities and challenges; PE&RC (2012)
- Linear models; PE&RC (2013)
- Generalized linear models; PE&RC (2013)

Laboratory training and working visits (2 ECTS)

- Biodiversity proxies using coarse resolution remote sensing; CONABIO, Mexico (2013)

Invited review of (unpublished) journal manuscript (5 ECTS)

- Remote Sensing: arctic vegetation trends (2012)

-
- Remote Sensing of Environment: spatio-temporal remote sensing data gap filling (2014)
 - ISPRS: multi-resolution data fusion (2014)
 - ISPRS: operational drought monitoring (2015)
 - CAGEO: satellite data processing tool (2015)

Deficiency, refresh, brush-up courses (3 ECTS)

- Advanced forest ecology and management; FEM (2012)

Competence strengthening / skills courses (3.6 ECTS)

- PhD Carousel; PE&RC (2015)
- Mental coaching; WUR (2015)
- Career orientation; WUR (2015)
- Efficient writing strategies; WUR (2015)

PE&RC Annual meetings, seminars and the PE&RC weekend (1.8 ECTS)

- Firsts years PE&RC weekend (2012)
- PE&RC Day (2014)
- Last years PE&RC weekend (2015)

Discussion groups / local seminars / other scientific meetings (6 ECTS)

- Ecological theory and application (2012-2015)
- ROBIN Project annual meeting (2012-2015)

International symposia, workshops and conferences (7 ECTS)

- NAC11; Veldhoven, the Netherlands (2012)
- Living planet symposium; Edinburgh, UK (2013)
- GV2M; Avignon, France (2014)
- AGU; San Francisco, USA (2015)

Lecturing / supervision of practical's / tutorials (3 ECTS)

- Geo-scripting (2014-2015)

Supervision of MSc students (9 ECTS)

- Performance of EVI for detecting intra and inter annual vegetation dynamics in the Amazon
- Quantifying land use intensity from Landsat time series in slash and burn systems
- Effect of landscape structure on forest recovery

This research received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant agreement Number 283093 — The Role Of Biodiversity In climate change mitigatioN (ROBIN).