The University Of Sheffield.

Remote sensing of tropical forest degradation from selective logging

By:

Matthew G. Hethcoat

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Science
School of Mathematics and Statistics

November 2019

Author Declaration

The material contained within this thesis is my own and has not previously been submitted for a degree at The University of Sheffield or any other institution. In addition to myself (MGH) there were several collaborators in this research; Shaun Quegan (SQ), David P Edwards (DPE), Robert G Bryant (RGB), Joao MB Carreiras (JMBC), Filipe M Franca (FMF), and Carlos A Peres (CAP). Each chapter in this thesis is written in the format of a manuscript for publication and is reproduced in full, with minor formatting alterations for consistency. In all cases the contribution of authors was as follows: MGH, SQ, DEP, RGB, and JMBC conceived the study. MGH performed the analyses and wrote the manuscripts, with contributions from all coauthors.

Chapter 3

This chapter has been published as:

MG Hethcoat, DP Edwards, JMB Carreiras, RG Bryant, FM Franca and S Quegan 2019. A machine learning approach to map tropical selective logging. *Remote Sensing of Environment* 221:569-582.

Chapter 4

This chapter is in revision in *Remote Sensing of Environment* as:

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, and S Quegan Detecting tropical selective logging with SAR data will require a time series approach.

Chapter 5

This chapter is currently in preparation for submission to *Remote Sensing* as:

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, and S Quegan Combining optical and SAR data for monitoring tropical selective logging.

Chapter 6

This chapter is under review in *Environmental Research Letters* as:

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, CA Peres, and S Quegan Mapping the rapid expansion of selective logging in the south-west Brazilian Amazon.

Table of contents

Acknowledgments

I would like to thank my supervisors Shaun Quegan, David Edwards, Robert Bryant, and João Carreiras. Your friendship, guidance, wisdom and support have been invaluable throughout this project and have helped to clarify and improve my thinking and writing. I would also like to express my gratitude for the support from the Mynas group and everyone in Conservation Bites and the Quegan group for the discussions, laughs, coffees, and snacks over the years.

I am extremely grateful for financial support from the Grantham Centre for Sustainable Futures, the University of Sheffield, the School of Mathematics and Statistics, and the Department of Animal and Plant Sciences. In addition, a large number of organizations made this work possible by providing support, data, or expertise: the National Aeronautics and Space Administration, U.S. Geological Survey, the MDA Corporation, the Canadian Space Agency, the Japanese Space Agency, European Space Agency, the National Centre for Earth Observation, the Natural Environment Research Council, the Brazilian Forest Service, AMATA Brazil, Jari Florestal, Google Earth Engine, and Planet Labs.

To my family, Michele, Michael, and Alison, thank you for total and unwavering support for all my endeavors. You have enabled me to pursue and be whatever I want in life and I cannot thank you enough. To my other family, the Taylor-Docherty's, your energy and enthusiasm is always an inspiration. Thank you for being such wonderful grandparents, aunts, and uncles to Teal and supporting the life Teegan and I have made together.

To my daughter Teal, your arrival has fundamentally changed the way I see the world and strengthened my desire to make it a better place for your generation and those still to come. You have deeply enriched my life and I am in awe of you every day.

Finally, to Teegan, I cannot express the depth of my gratitude, admiration, and love for you. You are an incredible mother and scientist and I am constantly overwhelmed by what an amazing partner you are. None of this was possible without your encouragement and support.

This work is dedicated to the activists and environmental defenders globally who put their lives at risk protecting the environment

Thesis abstract

Remote sensing is the most accurate and cost effective way to monitor forests at large spatial scales. The preceding decade has seen incredible progress in accurate forest monitoring from space, with operationalized deforestation and fire alerts available in near-real-time globally. In contrast, methods for detecting and mapping forest degradation from selective logging have lagged behind; despite recognition that selective logging is a key driver of both deforestation and forest degradation. In this these I develop novel methods that utilize detailed spatial and temporal logging records to train machine learning algorithms to detect and map tropical selective logging. First, I utilized optical satellite data from the Landsat program and show that imagery acquired before the cessation of logging activities (i.e. the final cloud-free image of the dry season during logging) was best for detection, displaying a 90% detection rate (with roughly 20% commission and 8% omission error rates). Next, I tried extending this methodology to the detection of logging with synthetic aperture radar (SAR) data, but poor performance made logging predictions too uncertain. I go on to show that SAR data from Sentinel-1 display a distinct breakpoint in the time series of pixels logged under higher intensities (> 20 m^3 ha^{-1}) and could be used to detect more intensive selective logging within the Amazon. I then assess if combining optical and SAR data improve the detection of logging over the use of either on their own. I show that a combined model performs worse than optical data alone and including SAR data adds uncertainty that lowers model performance. Finally, I refine the optical approach developed in the beginning, generalizing the methodology to facilitate a large spatial and temporal scale assessment of selective logging. We create annual estimates of selective logging between 2000 and 2019 over the Brazilian state of Rondônia. I estimate that 41.0% of the State of Rondônia remained undisturbed forest through 2019, with 3.4% having undergone selective logging and 25.7% being deforested (with 13% Commission Error and 45% Omission Error over the twenty year period). In general, rates of selective logging were twice as high in the first decade relative to the last decade of the period. My results show improved access to data and technologies will enable advances in space-based forest monitoring and reiterate the value of free and open data access policies. Our approach is step in the direction of an operationalized selective logging monitoring system capable of detecting subtle forest disturbances over large spatial scales.

Chapter 1

Introduction

"…one natural feature of [the Amazon], the interest and grandeur of which may be fully appreciated in a single walk: it is the virgin forest. Here no one who has any feeling of the magnificent and the sublime can be disappointed

–Alfred Russel Wallace, 1849

1.1 Background

The value of tropical forests is incalculable. Despite only covering about 10% of Earth's land surface, tropical forests are estimated to host at least two-thirds of terrestrial biodiversity (Gardner et al. 2009). The have immense conservation value and are the reservoirs of functional and phylogenetic diversity. In addition, the goods and services provided by forests links them intimately to human livelihoods and the global economy. Nearly one-third of the world's population relies on wood fuel as their primary energy source, with the proportion rising dramatically in the tropics (FAO 2017). Tropical forests are the last lifeline for communities living on the edge of extreme poverty and provide a means to generate an income, find building materials, and access food and medicines in the absence of wages.

Tropical forests also play a crucial role in Earth's carbon and hydrological cycle. Forests regulate stream flow, filter water, and reduce soil erosion and sedimentation. In addition, evapotranspiration actually induces cloud formation and precipitation, influencing rainfall patterns and seasonality in the tropics (Salati et al. 1979; Wright et al. 2017). While much of the tropics sit on relatively nutrient poor soils (Vitousek 1984), they play a vital role in the global carbon and nitrogen cycles, accounting for nearly 40% of terrestrial net primary productivity and storing about 25% of global biomass (Townsend et al. 2011). The tropical carbon cycle has received considerable attention in recent years, owing to the climate implication of their massive carbon sequestration potential and the emissions associated with their loss (Maxwell et al. 2019).

In addition to anthropocentric value, tropical forests house some of the largest wildernesses and hold immense intrinsic value as wild places. However, a rapidly rising global population and continued loss and degradation of intact forested landscapes are putting incredible pressures on tropical forests globally. The tropics are thought to be nearing a tipping point where fragmentation will begin to dramatically increase (Taubert et al. 2018). Moreover recent work has shown that tropical forests globally are composed of over 50 million forest fragments, encompassing nearly 50 million km of edge (Brinck et al. 2017). At the current pace,

the tropical forests of the future are destined to be considerably higher, steeper, and of lower conservation, economic, and intrinsic value (Betts et al. 2017; Edwards et al. 2019).

Conversion to agriculture is the primary driver of deforestation, accounting for almost 80% of the world's forest losses (Gibbs et al. 2010; Hosonuma et al. 2012; FAO 2018; Curtis et al. 2018). However, the specific drivers and the contribution of particular disturbance types have shifted through time and varies across continents (Rudel et al. 2009; Hosonuma et al. 2012). While the same is true of forest degradation in the tropics, forest management practices and fuelwood extraction collectively account for >75% of tropical forest degradation activities globally (Hosonuma et al. 2012; Pearson et al. 2014). Specifically, production of charcoal is the primary driver of forest degradation in Africa's wooded savannas, while commercial logging operations dominate in Congo and the rest of the tropics (Hosonuma et al. 2012). Selective logging operations in the tropics are often the first anthropogenic disturbance event to impact primary forests. The network of roads facilitates access and promotes additional sources of degradation (e.g. fires, fuel wood extraction, defaunation, illegal logging and mining). Over time many forest tracts are eventually cleared for agriculture or human settlements as the logging frontier shifts to the next region of primary forest.

Improvements in data and technologies have increased the accuracy and speed of forest monitoring systems on a global scale (Hansen et al. 2013; Gorelick et al. 2017). Near real-time deforestation alerts are now possible from a variety of sources, like Global Forest Watch (e.g. FORMA and GLAD) and the Brazilian Space Agency (e.g. DETER). In addition, the ability to map fires over large spatial and temporal scales has been aided by platforms like Google Earth Engine (Gorelick et al. 2017), though the carbon implications are not well understood. In contrast, however, monitoring and mapping of selective logging activities has lagged behind, despite the recognition of the role it plays in driving both deforestation and fires (Asner et al. 2009; Hosonuma et al. 2012).

Satellites are considered the most accurate and cost effective way to monitor forests at large spatial scales. The same tools and technologies that have advanced deforestation can be brought to bear in the efforts to map selective logging. Now that the scientific community has, to some extent, cracked the deforestation problem there is increasing attention on improving abilities to monitor forest degradation. The principle aim of this thesis is to contribute to that body of work. I provide a brief review of tropical forest degradation and forest management practices, with an emphasis on Brazil, and discuss the impacts on the global carbon cycle, biodiversity and the other key ecosystem processes. I review and discuss technologies and methodological advancements to monitor tropical forests globally. Finally, I outline the aims and objectives of this thesis in regards to the major research needs and the key knowledge gaps.

1.2 Forest degradation

1.2.1 Characterizing degradation

While there is no internationally agreed definition of forest degradation, there is general agreement that it embodies disturbances within a forest that persists as a forest (Simula 2009; Ghazoul et al. 2015). This ambiguity has made generalizing the impacts of forest degradation difficult, because it can include forests subject to varying intensities of selective logging, fire, mining, fuelwood extraction, hunting pressure, infestation of invasive species, etcetera. This lack of consensus has, in part, also hampered the development of coordinated international forest policies to track and monitor forest degradation (Sasaki and Putz 2009; Herold et al. 2011; Ghazoul et al. 2015). However, there is growing agreement that the Intergovernmental Panel on Climate Change (IPCC) guidelines on how to report and monitor forest degradation under the United Framework Convention on Climate Change (UNFCCC) will form the basis for an internationally agreed framework under the Reducing Emissions from Deforestation and forest Degradation (REDD+) mechanism (FAO 2011; Herold et al. 2011). Thus, global action on forest degradation will largely be linked to climate mitigation potential, with the anticipation of achieving secondary benefits for biodiversity and human livelihoods.

1.2.2 Tropical forest degradation and the global carbon budget

Under REDD+, forest degradation represents a loss of carbon stocks within forested landscapes (UN-REDD 2018). Activities associated with forest degradation are thought to impact more than 100 million hectares annually (Herold et al. 2011), quadruple the area deforested every year (Hansen et al. 2013). From the perspective of the global carbon budget, forest degradation is thought to be a major source of carbon emissions, comprising up to an additional 50% of emissions from forest losses alone (Asner et al., 2005, 2010; Grace et al., 2014; Bustamante et al., 2016). However, quantifying and monitoring the carbon implications of tropical forest degradation remains a major technical challenge and large uncertainties remain in the estimates of carbon emissions (Bustamante et al. 2016; de Andrade et al. 2017; Mitchell et al. 2017). Indeed, the emissions estimates from tropical land use are currently lumped into a single net value (comprising degradation, deforestation, and forest regrowth) in the global carbon budget and represent the difference between the sum of all other components to balance the budget (Le Quéré et al. 2018). About half of the anthropogenic emissions remain in the atmosphere, with the rest being taken up by the land and ocean sinks (Le Quéré et al. 2018, Figure 1.1). While tropical forests are thought to make an approximately neutral contribution to the global carbon cycle (Mitchard 2018; Sellers et al. 2018), there is growing evidence that they can easily become net carbon emitters if not properly managed (Grace et al. 2014; Baccini et al. 2017; Maxwell et al. 2019).

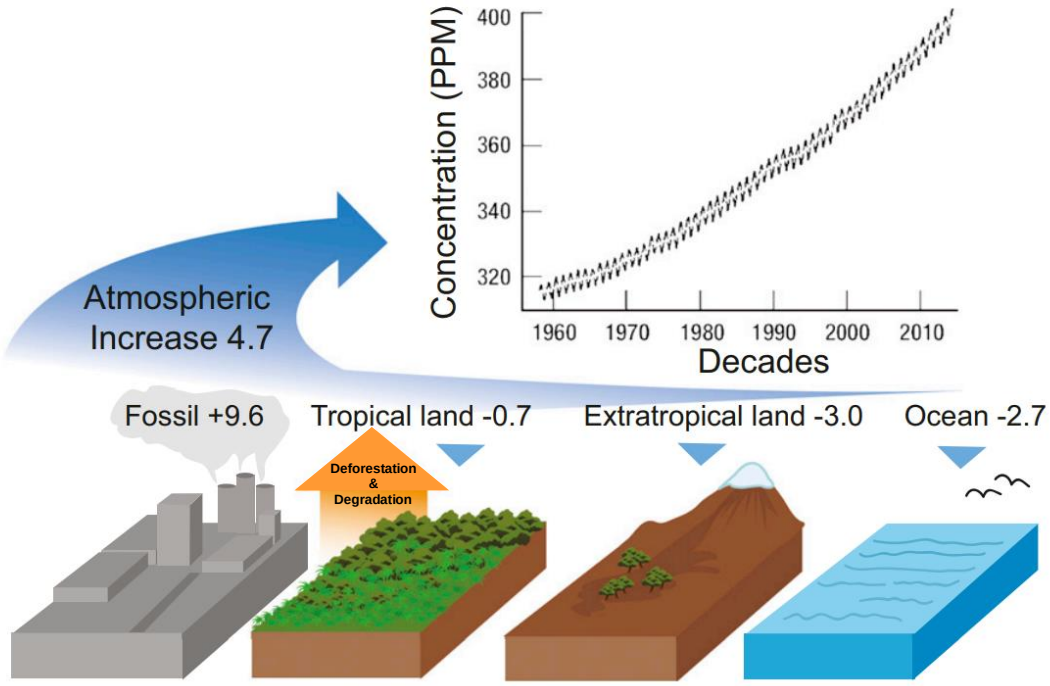


Figure 1.1 Global carbon cycle representation, showing the emissions from fossil fuels (and concrete), the two components of the land sink, and the oceanic sink. Image modified from Sellers et al. 2018, with units in $GtCyr^{-1}$.

1.3 Selective logging

Selective logging in the tropics generally occurs in waves, following a well-documented cycle. First logging roads are built to enter forested tracts and facilitate a managed harvest. The largest and most economically valuable trees are usually harvested first. After logging has finished many of the smaller, internal logging roads are decommissioned and the forest is left to regenerate before the next harvest cycle in 20-60 years (Pinard and Putz 1996; Putz et al. 2001, 2012; Blaser et al. 2011). There is growing evidence that in order to increase the sustainability of forest management practices in the tropics harvest cycles should be nearer 100 years, but the trend appears to be in the opposite direction with premature reentry being increasingly common (Blaser et al. 2011; Putz et al. 2012; Richardson and Peres 2016). After successive rounds of logging many forested tracts are cleared for agriculture or settlements as the logging frontier moves elsewhere.

Two main types of forest management plans occur in the tropics and are referred to as Conventional Logging (CL) and Reduced Impact Logging (RIL). RIL differs from CL in that there is careful road and skid trail planning, directional felling of trees to minimize collateral damage to adjacent trees, and pre-harvest liana cutting where possible to limit additional canopy damage (Putz and Pinard 1993). For decades RIL saw little uptake and application because of

entrenched ideology, assumed cost, and lack of information (Putz et al. 2000; Asner et al. 2009). While use of RIL practices remains generally low, growing concerns about the impacts of poor forest stewardship on biodiversity and carbon has made RIL an appealing win-win strategy for balancing development and ecological goals (Gibson et al. 2011; Ellis et al. 2019; Maxwell et al. 2019). Specifically, after accounting for the amount of wood volume removed, RIL activities do better at maintaining biodiversity than CL practices (Bicknell et al. 2014) while simultaneously sequestering more carbon during regrowth (Putz et al. 2008; Martin et al. 2015). However, if the current trends in loss and degradation of tropical forests continue, there is reason to believe tropical forests may transition to being a net carbon source (Mitchard 2018). The development and championing of policy actions consistent with RIL practices are needed and ought to be included in national climate mitigation strategies.

Globally, the intensity of selective logging operations vary by an order of magnitude (<10 m^3 ha^{-1} to >150 m^3 ha^{-1}); however intensities are generally <50 m^3 ha^{-1} outside of Southeast Asia and very low (<10 m^3 ha^{-1}) in Africa (Sist 2000; Putz et al. 2001). Historically Brazil encouraged logging and land clearance as part of its settlement and development activity between 1970 and 1990 (Asner et al. 2009). Widespread, unmanaged logging ravaged large portions of Mato Grosso, Pará, and Rondônia, accounting for more than 90% of production from the Brazilian Amazon (Asner et al. 2009). In an effort to address some of the impacts rampant deforestation and logging had caused, Brazil adopted the CONAMA resolution (CONAMA 2009), which imposed a number of restrictions on logging operations, including (among other things) limiting logging intensities to 30 m^3 ha^{-1}.

The ecological impacts of selective logging on tropical forests are well studied. Selectively logged forests have been shown to have increased microclimatic variability (Stratford and Robinson 2005), increased soil erosion (Douglas 1999; Hartanto et al. 2003), reduced tree diversity (Berry et al. 2008; Martin et al. 2015), altered forest phenology (Koltunov et al. 2009), and lowered levels of biodiversity (Burivalova et al. 2014). In addition, logging road networks have big implications for primary tropical forests (Kleinschroth et al. 2015, 2016; Kleinschroth and Healey 2017), becoming pipelines of human access into previously inaccessible forested areas. Roads also create forest edges that can alter abiotic processes like microclimate (Williams-Linera et al. 1998), change plant and animal species composition (Tabarelli et al. 2012), increase fire susceptibility (Armenteras et al. 2013), and ultimately weaken forest resilience (Murcia 1995; Kleinschroth and Healey 2017). However, forests subjected to selective logging generally maintain higher levels of biodiversity than other anthropogenic land use types in the tropics, such as plantations or secondary forests (Gibson et al. 2011; Edwards et al. 2014). These findings have resulted in some authors advocating for the

conservation value of logged forests, arguing that preventing their clearance should be seen as the next-best alternative to protecting primary forest in the tropics (Edwards et al. 2011, 2014).

Large uncertainties remain in assessing the true impacts of selective logging because the technological advances in detecting and monitoring logging at large spatial scales are only just emerging (Hethcoat et al. 2019). Improvements in detection of selective logging in the tropics would enable the mapping of intact primary forest as well as identify regions previously logged that possess high conservation value. Equally, there is an ever-increasing need to detect and account for the estimated 50-90% of tropical timber on the international market harvested illegally at very low intensities (Kleinschmit et al. 2016; Brancalion et al. 2018). Reliable mapping of forest degradation from selective logging is a key piece in understanding the terrestrial portion of the carbon budget and the role of land-use in turning tropical forests into net carbon emitters (Baccini et al. 2017; Mitchard 2018). Moreover, verifiable forest monitoring systems are urgently needed for tropical nations and conservation groups seeking to report and/or mitigate carbon emissions through improved forest stewardship (GOFC-GOLD, 2016). Earth observation technologies will play a key role in the current and future development of forest monitoring systems to track forest degradation, as they have with deforestation in the past (GFOI 2016).

1.4 Remote sensing of forest disturbances

Remote sensing is the most accurate and cost effective way to monitor forests at large scales (GFOI 2016). The last decade has seen the realization of comprehensive deforestation monitoring globally (Hansen et al. 2013), enabling operationalized forest monitoring programs on the national and international level. These advances were made possible, in part, because of the opening of the Landsat archives in 2009 for free use. The 60-fold increase in data downloads and the rapid growth in the scientific, private, and civil sectors resulted in similar policies being adopted under the European Space Agency's (ESA) Copernicus Program (Zhu et al. 2019). Free Earth observation data is now available on a scale like never before, and with deforestation monitoring having achieved an operationalized quality, attention has now shifted to pursuing equivalent gains in monitoring degradation (Langner et al. 2018; Bullock et al. 2018).

1.4.1 Optical data approaches

The satellites in the Landsat program have provided an unprecedented view of global change over the last forty years (Figure 1.2). Landsat 5 was the longest running Earth observation satellite in history, spanning more than 29 years of active duty in space. Ironically, Landsat 5 did not have a systematic acquisition plan from the outset. However, the value of the data was soon

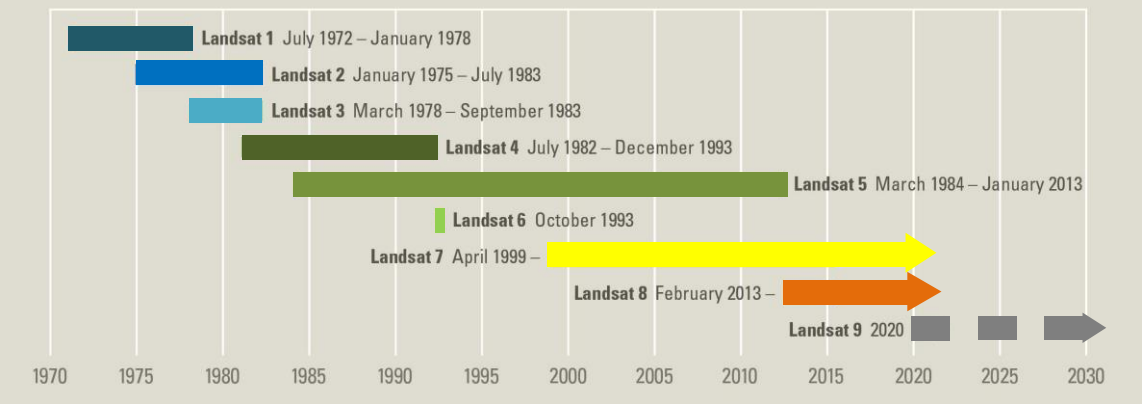


Figure 1.2 Timeline of the Landsat program missions (adapted from www.usgs.gov)

recognized and a consolidated global archive was created with the help of cooperating nations partnered with the Landsat program (Goward et al. 2006; Wulder et al. 2016). Over the years Landsat data has come to dominate the study of forested systems and has proven invaluable in improving our understanding of the Earth system (Wulder et al. 2016).

Several studies have sought to monitor selective logging in the tropics with Landsat data (Souza and Barreto 2000; Asner et al. 2002, 2004, 2005; Monteiro et al. 2003; Souza et al. 2005, 2013; Matricardi et al. 2007, 2010; Shimabukuro et al. 2014), with the vast majority of approaches utilizing spectral unmixing models. Briefly, spectral unmixing is a method that enables the fractions of various spectral features (often referred to as endmembers) within a single pixel to be estimated. Thus, the proportion of photosynthetic vegetation within a pixel, for example, can be estimated to look for changes in canopy cover through time. However all of the applications have involved moderately high logging intensities (>20 m^3 ha^{-1}) and their methods for identifying logging used simplistic decision trees with hardwired thresholds of change in endmember values to classify forest disturbances through time (e.g. Asner et al. 2005; Souza et al. 2013). The authors have generally acknowledged their methods are conservative and detect areas of selective logging at moderately high intensities that possess large canopy gaps and an abundance of spectrally distinct features, like log landing decks or large road networks. However, current forest monitoring methods (i.e. Hansen et al. 2013) now classify many of these areas as scattered deforestation detections (Figure 1.3). Consequently, Landsat data has been assumed to be too coarse to map and quantify selective logging at lower logging intensities and the amount of forest disturbance overlooked using these techniques is unknown.

More recently, time-series methods for detecting forest disturbances have been developed (Zhu et al. 2012; Zhu and Woodcock 2014; Bullock et al. 2018). These approaches have only just become practical with the development of Google Earth Engine (GEE) and its continued growth over the last two years (Gorelick et al. 2017). Previously, researchers needed

Figure 1.3 Deforestation detections from the Hansen et al. (2013) data (in red) in the Saracá-Taquera National Forests, Pará overlaid with selective logging tree locations (harvested at approximately 21 m^3 ha^{-1}) from the same year (grey circles). The deforestation detections here are selective logging activities. The map is centered at 56.17 W, 1.60 S.

to download every Landsat image in the time series in order to perform the analyses, which could consume large volumes of data storage. However, with the help of GEE, large-scale applications of these methods are now emerging in the literature and hold great promise at improving the detection of subtle forest disturbances (Zhu et al. 2016; Bullock et al. 2018).

1.4.2 Synthetic Aperture Radar data approaches

While the preceding decade has seen a number of improvements in detecting forest disturbances from space (Hansen et al. 2013; Tyukavina et al. 2017; Hethcoat et al. 2019), advances in forest monitoring have almost universally relied on optical satellite data from the Landsat program (Wulder et al. 2016). Yet, the effectiveness of optical data is limited in tropical

regions with frequent cloud cover. Synthetic Aperture Radar (SAR) data have a couple key of advantages over optical data that make it ideal for use in tropical systems. First, SAR data are an active signal, meaning they do not require solar illumination to acquire data and the transmitted signal can penetrate clouds and acquire data in regions obscured by cloud cover. In addition, the properties of the SAR signal (wavelength and polarization) influence its interaction with forest architecture and can provide detailed information about forest structure. In particular, SAR data in L- and P-band are known to exhibit a consistent relationship with aboveground biomass (Mitchard et al. 2009; Koch 2010; Saatchi et al. 2011), while X- and C-band SAR generally interact only with the forest canopy (Figure 1.4).

SAR data have been used in forestry applications since the early 1990s, with much of the initial development and early work utilizing aerial radar systems affixed to aircraft (Flores-Anderson et al. 2019). The launch of ERS-1 in 1991 was quickly follow by JERS-1 in 1992 and earmarked the start of continuous spaceborne SAR observation of Earth (Figure 1.5). However, the spaceborne SAR missions lack coordinated observation and scientific strategies (though we recognize comparing multiple missions across many national space agencies with the Landsat program is unfair). Consequently, the SAR data archives are spatially and temporally fragmented and in many cases the data products required commercial licences (i.e. a fee) to use. As a result, uptake by users has been more limited than optical data and the full potential of SAR has likely been under-utilized (Reiche et al. 2016).

Longer wavelength SAR (L- and P-band) primarily interacts with large branches and trunks, resulting in high backscatter from double-bounce. This enables accurate differentiation between forest and non-forest areas, because the loss of trees strongly decreases backscatter, and has been well studied over the years (Woodhouse 2005). In contrast shorter wavelength SAR, like X- and C-band, is less sensitive to forest change. Shorter wavelength SAR signals interact with an intact forest canopy in a similar manner as remnant understory vegetation (after deforestation). However, forest mapping has been demonstrated effectively in some applications (Saatchi et al. 1997; Antropov et al. 2016). More recently, polarimetric and interferometric methods have been developed that utilize phase information in the SAR signal

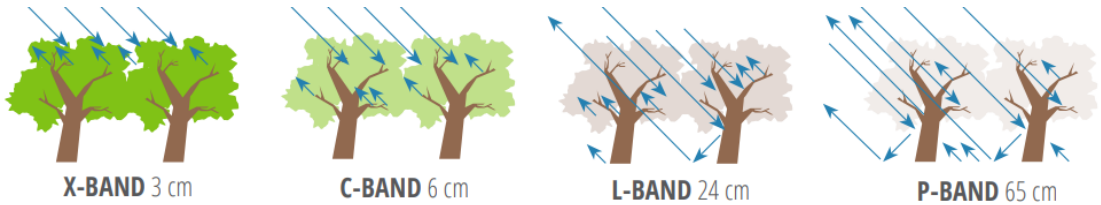


Figure 1.4 Differences in SAR signal interaction with forest canopy associated with wavelength (adapted from Flores-Anderson et al. 2019).

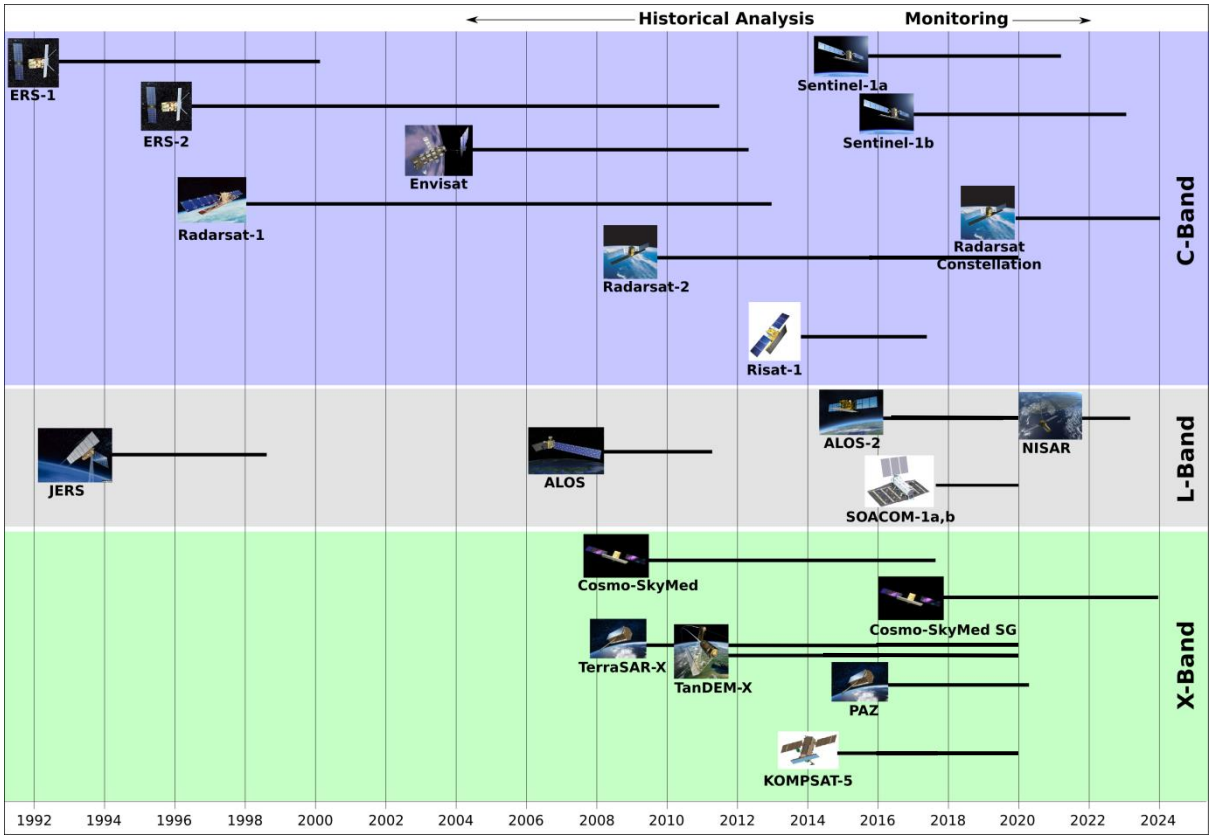


Figure 1.5 Timeline of spaceborne SAR missions (adapted from www.unavco.org).

to detect forest changes (Deutscher et al. 2013; Mathieu et al. 2013; Lei et al. 2018; Flores-Anderson et al. 2019). Yet, the limited temporal and spatial coverage of SAR data have hampered widespread application and use of these techniques to monitor forest disturbances (e.g. single-pass interferometric SAR is only available with TanDEM-X data).

The launch of Sentinel-1A in 2014 represented the first continuous global acquisition strategy for open SAR data. The rapidly expanding archives of Sentinel-1 have resulted in a number of time series methods for detecting forest change (Reiche et al. 2018a, b). With the successful launch of SAOCOM 1A in late 2018, the planned continuation of the Sentinel-1 missions (with C and D), and the anticipated launches of SAOCOM 1B in 2019 and NISAR in 2021, vast amounts of free C- and L-band SAR data will soon be available. Accordingly, methods are needed that utilize SAR data for large-scale forest monitoring, yet no study has used SAR for detection of selective logging activities with the aim of operationalized forest monitoring. Again, however, advancements in monitoring selective logging with SAR data are generally lacking, despite widespread recognition of both the need and the role it could play (Reiche et al. 2016; Mitchell et al. 2017).

1.5 Thesis aims and outline

The overarching objective of this thesis is to improve current abilities to monitor tropical selective logging with remote sensing data. We utilize both optical- and SAR-based methods for detecting logging and evaluate errors with field data from logging concessions across Brazil. Operationalized forest monitoring systems for detecting of selective logging at large spatial scales are urgently needed in all sectors of society and we sought to contribute to this body of work. The core of this thesis is composed of four chapters that generally build on or extend elements from earlier findings (Figure 1.6).

In Chapter 3, we develop an approach that uses detailed logging records, from a single logging concession in Rondônia, Brazil, to build machine learning algorithms for detecting selectively logged pixels in Landsat imagery. We demonstrate the feasibility of the approach and validate its effectiveness at a second logging concession, approximately 1500 km away, in Pará, Brazil.

In Chapter 4, we attempt to extend the approach that worked successfully in Chapter 3 with optical data to the detection of logging with synthetic aperture radar (SAR) data, with an expanded dataset on logging from across Brazil. Three different SAR datasets (two C-band and one L-band) are used to detect logging, but with poor performance. We go on to examine if a dense time series of C-band SAR from Sentinel-1 displayed a distinct breakpoint with the onset of logging.

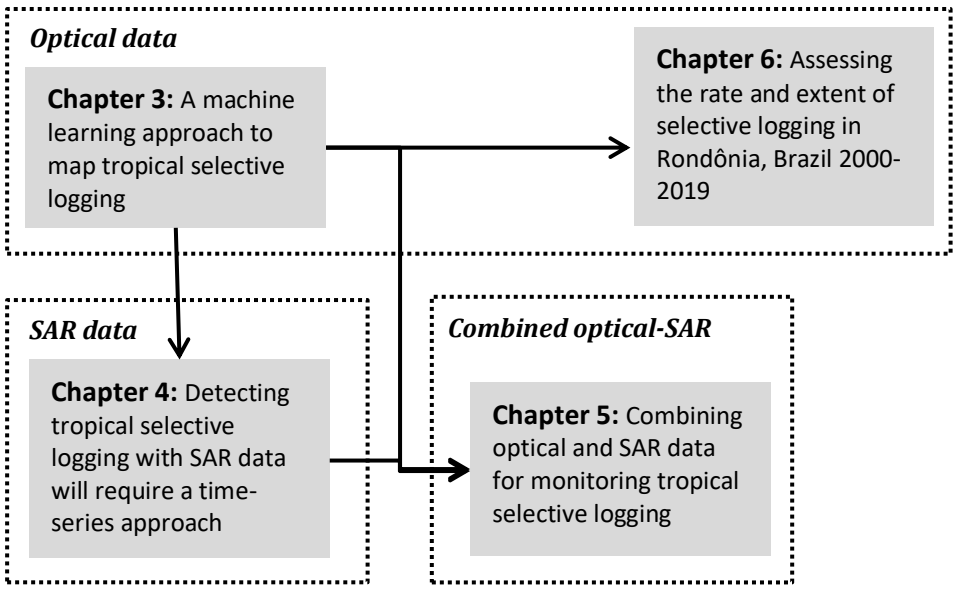


Figure 1.6 Overview of links between data chapters of this thesis.

In Chapter 5, we combine the approaches from Chapters 3 and 4 to understand how incorporating both optical and SAR data might improve detection of logging over the use of either on their own. We utilized a completely independent dataset from the earlier chapters; composed of selective logging records from across Brazil. We assess the results over a logging concession in Rondônia where we have knowledge of logging, generally, but no field data.

In Chapter 6, we further refine the approach developed in Chapter 3, generalizing the methodology to facilitate a large spatial and temporal scale assessment of selective logging. We create annual estimates of selective logging from 2000-2019 in Rondônia, Brazil and asses trends in logging through time.

Finally, in Chapter 7, I start by summarize the core thesis results. I then turn to discussing the implications of our findings for the development of regional- to pan-tropical scale logging maps and how this information can inform assessments of carbon losses and biodiversity impacts from selective logging at scale as well as enable the development of near real-time monitoring of selective logging expansion.

Chapter

2

Theory:

Detecting forest degradation with satellite data

2.1 Conceptualizing forest degradation

2.1.1 Definitions and drivers

There is no singular definition of degradation, as forests vary substantially around the world (in structure, diversity, functions, values, perceptions, etc.). Yet, degradation has generally come to be viewed, conceptually, as a loss of resilience; a reduction in the capacity to return to a pre-disturbance state (Ghazoul et al., 2015; Olsson et al., 2019; Simula, 2009; Thompson et al., 2013). Normally, degradation is reserved for anthropogenic disturbances, though the distinctions between natural and anthropogenic factors seem increasingly vague (e.g. global climate change affecting hurricane frequency in the Caribbean or shifting rainfall patterns in the Amazon basin). Nevertheless, throughout this thesis degradation is regarded as an anthropogenic phenomenon that is demonstrably linked (directly or indirectly) to human activities.

The global drivers of forest degradation include selective logging, anthropogenic fires, charcoal production, fuelwood collection, and livestock grazing within forests (Hosonuma et al., 2012). The prominence of particular drivers varies geographically, for example, selective logging constitutes >75% of all degradation within South America and Asia, but fuelwood collection and charcoal production dominate in tropical Africa (Hosonuma et al., 2012; Kissinger et al., 2012). While the global extent, severity, and expansion of various forest degradation activities are not well quantified, there is general agreement that degradation is an urgent and widespread problem.

The desire to monitor forest degradation (for the reasons outlined in Chapter 1) imposes the problematic task of first identifying the best metric(s) to track (i.e. define degradation), then determining appropriate spatial and temporal scales that sufficiently capture the transition or state-change to a degraded forest. These are no small achievements and the multitude of working definitions of forest degradation reflects a bespoke approach to defining degradation that suits local, regional, or national interests (Ghazoul et al., 2015). Nearly every aspect of degradation is contextual, because it is so complex and value-laden (Warren, 2002). Nevertheless, the UNFCCC (under REDD+) have reduced the idea down to a single, quantifiable metric of reduction in forest carbon stocks for the purposes of monitoring, verifying, and reporting progress under REDD+ (UN-REDD, 2018). Moreover, remote sensing technologies, used in combination with other available data, will be the primary tool for generating globally consistent metrics that can track degradation over appropriate time scales.

2.1.2 Selective logging and degradation

Selective logging activities that do not exceed the capacity for regrowth (i.e. sustainable forest management; SFM) are not regarded as degradation under REDD+; the idea being that forests

will recover if responsibly managed. However there is strong evidence that this is a gross oversimplification. First, even when SFM is performed, a lack of governance, monitoring, and enforcement mechanisms can result in subsequent degradation by other actors. Second, early re-entry into forests is increasingly common and the harvest regimes should be closer to 100 years, not the 30-40 years typical of the tropics (Blaser et al., 2011; Putz et al., 2012; Richardson & Peres, 2016). Finally, post-logging species composition has been shown to never fully recover after the initial harvest, a consequence of over-reliance upon relatively few, high value species that become functionally extinct within former logging concessions (Richardson & Peres, 2016). For the purposes of this thesis, we regard selective logging as forest degradation and use the terms interchangeable at times. However, we recognize that SFM practices can be done in such a way that selective logging does not constitute degradation. Crucially, the moniker of SFM requires evidence that has not been sufficiently demonstrated. Regardless of whether or not selective logging is considered degradation, there is an undeniable need to quantify the extent, intensity, and expansion of logging activities globally in order to inform progress toward climate targets, but also to better understand the impacts to the suite of goods and services forests provide.

2.2 Remote sensing

2.2.1 Satellite sensors and land surface attributes

2.2.1.1 Passive sensor data

Optical satellite sensors rely on capturing solar radiation that has been reflected by the Earth. Essentially a sophisticated spaceborne digital camera, the passive sensors onboard are sensitive to a section of the electromagnetic spectrum (Figure 2.1). Satellites further subdivide the optical region of the electromagnetic spectrum into numerous bands that record incoming radiation at

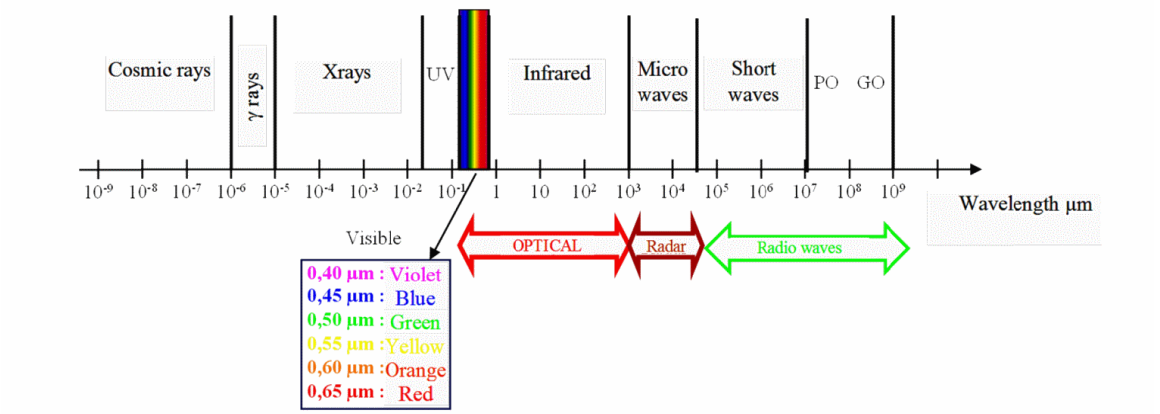


Figure 2.1 The electromagnetic spectrum with key wavelength ranges labelled (from Briottet, 2016).

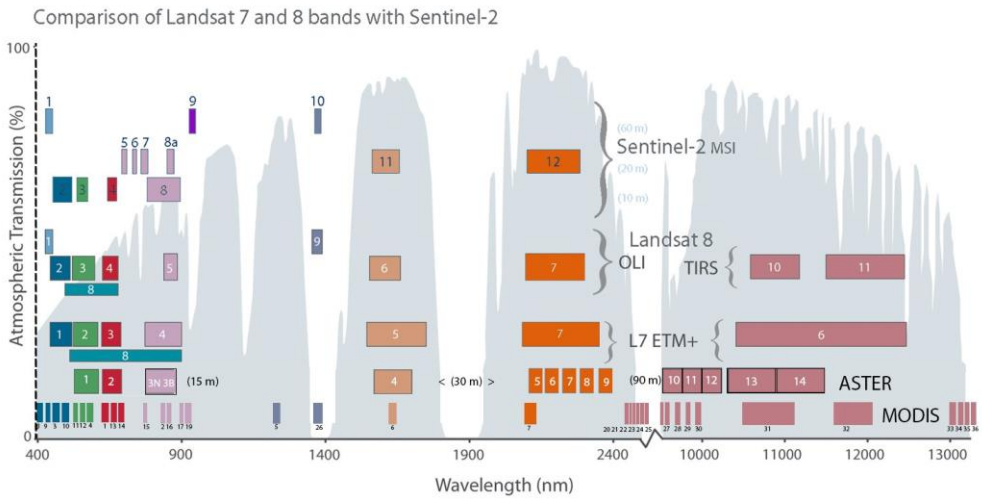


Figure 2.2 Band designation comparison between the MODIS, ASTER, Landsat, and Sentinel-2 sattelites. Note, the Y-axis relates to the grey spectral reflectance curve in the background (image from USGS Landsat Program).

specific wavelengths. The number of bands and the range of wavelengths each band covers differ slightly from instrument to instrument, but with broad overlap (Figure 2.2). For example, Landsat 8 data are divided into 11 sections (or bands), MODIS has 36 bands, and Hyperion (not shown in Figure 2.2) has 220 bands. The remainder of this section focuses on Landsat band designations, however, the general interpretations are the same for other satellite sensors with comparable wavelengths.

Table 2.1 provides detailed information about Landsat bands and includes some details on the particular features each band can discriminate. For example, the first band in Landsat 8 is regarded as an aerosol band - its sensitivity to deep blues and violets make it primarily used for mapping shallow water and detecting fine particles like dust or smoke in the atmosphere. Blue light tends to be easily scattered by the Earth's atmosphere and fine particles, consequently, the aerosol band is typically not used for terrestrial applications. Bands 2-7 are the workhorses for mapping land surface phenomena, correspond with the visible and infrared (IR) portions of the electromagnetic spectrum (Table 2.1). The visible bands (blue, green, red) are most similar to what the human eye would see; healthy vegetation appears green, unhealthy vegetation and soils are brown, urban features appear white and grey, and water is dark blue or black. The IR region is particularly useful in forest applications, as plants reflect, transmit, and absorb different portions of the near-IR spectrum. For example, healthy vegetation absorbs blue and red light during photosynthesis (and reflects green). In addition, plants with more chlorophyll tend to reflect more near-IR energy than unhealthy plants – a result of stronger chlorophyll reflectance. Thus, analyzing absorption and reflectance in visible and IR wavelengths can provide useful information about the productivity and health of vegetation as well as assist in identifying exposed soils. The optical signal associated with selective logging, depending on the

Table 2.1 Comparison of Landsat 7 (ETM+) and Landsat 8 (OLI) band designations, pixel size, wavelength coverages, and brief notes on applications (adapted from USGS Landsat Program).

Landsat 7 ETM+ (µm)		**Landsat 8 OLI (µm)**		**Applications**
		30m **Coastal/Aerosol** (0.435 - 0.451)	Band 1	Coastal and aerosol studies (fine dust and smoke).
Band 1	30m **Blue** (0.441 - 0 0 514)	30m **Blue** (0.452 - 0.512)	Band 2	Distinguishing soil from vegetation and deciduous from coniferous vegetation.
Band 2	30m **Green** (0.519- 0.601)	30m **Green** (0.533 - 0.590)	Band 3	Emphasizes peak vegetation, which is useful for assessing plant vigor. Barren lands, urban areas, and roads appear brighter.
Band 3	30m **Red** (0.631 - 0.692)	30m **Red** (0.636- 0.673)	Band 4	Discriminates vegetation and soil - strong chlorophyll absorption and strong reflectance for most soils.
Band 4	30m **NIR** (0.772- 0.898)	30m **NIR** (0.851 - 0.879)	Band 5	Emphasizes chlorophyll content and shorelines (water absorbs NIR). Best spectral region to distinguish vegetation varieties and conditions.
Band 5	30m **SWIR-l** (1.547 - 1.749)	30m **SWIR-l** (1.566 - 1.651)	Band 6	Discriminates moisture content of soil and vegetation; penetrates thin clouds.
Band 6	60m **TIR** (10.31 - 12.36)	100m **TIR-1** (10.60-11.19)	Band 10	Thermal mapping and estimated soil moisture.
		100m **TIR-2** (11.50-12.51)	Band 11	Improved thermal mapping and estimated soil moisture.
Band 7	30m **SWIR-2** (2.064- 2.345)	30m **SWIR-2** (2.107 - 2.294)	Band 7	Improved moisture content of soil and vegetation; penetrates thin clouds.
Band 8	15m **Pan** (0.515-0.896)	15m **Pan** (0.503 - 0.676)	Band 8	Sharper image definition.
		30m **Cirrus** (1.363 - 1.384)	Band 9	Improved detection of cirrus cloud contamination.

intensity, tends to relatively subtle and short lived in the Brazilian Amazon (Broadbent et al., 2006). Within the context of this thesis, however, the visible and IR bands associated with Landsat (Table 2.1) are predicted to show a discernable shift in the spectral response of pixels with the onset of selective logging (from exposed soils, loss of canopy cover, an increase in woody debris, etc.). Moreover,

2.2.1.2 Active sensor data

Synthetic Aperture Radar (SAR) data is an active signal, meaning it provides its own illumination in the form of microwave energy and does not, therefore, require reflected or radiated solar energy to detect features (Figure 2.1). Fundamentally, SAR works by measuring the Doppler shift of returned radar signals - resulting from the forward motion of the satellite. In doing so, SAR satellites also measure the portion of the transmitted signal that is returned; termed backscatter. Backscatter is influenced by a number of factors (see next sections), but this quantity forms the basis for all SAR imaging. Yet, because of the way microwaves interact with the atmosphere and ground, only a subset of the full radar frequencies are generally used in SAR satellites (Figure 2.3). The applications of SAR are numerous and include geology, crop monitoring, deforestation detection, sea ice measurement, disaster monitoring, and oceanic vessel monitoring, however, the remainder of this section focuses on terrestrial applications of SAR data, with a particular emphasis on forests.

SAR frequency plays an important role in determining the nature of the returned signal, as an object on the ground will tend to scatter longer and shorter wavelengths differently. Regardless of the wavelength, however, the nature of the returned signal depends on the 1) slope, 2) roughens, and 3) the dielectric properties (the ratio of the reflectivity of the signal relative to a vacuum) of scattering objects (Flores-Anderson et al., 2019; Woodhouse, 2005). In particular, rougher and wetter objects result in higher backscatter (appear brighter), while the effect of slope depends on the terrain aspect relative to the sensor. For example, higher backscatter results from slopes facing the sensor (foreshortening) and is seen as brighter facing

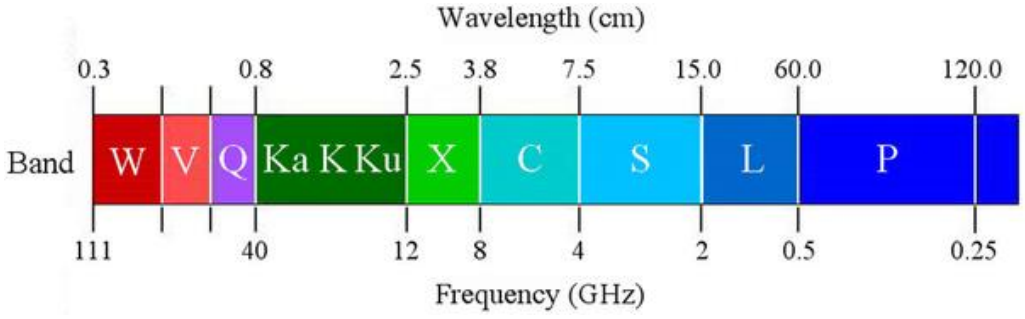


Figure 2.3 A portion of the radar region of the electromagnetic spectrum (from Ouchi, 2013).

slopes and dark back slopes stretching away from the sensor. For steep slopes, the position of the terrain relative to the sensor may cause these distortions to be greatly exaggerated, producing layover and shadowing. In layover the higher portions of the terrain are imaged first and makes high ridges or peaks appear to fall over toward the nadir. In shadowing the radar signal from the back slope is invisible to the sensor because the terrain blocks the signal from reaching the back slopes. Both result in a loss of data in the affected areas. The work presented in this thesis covers flat terrain of the Amazon basin and we were not impacted by these factors.

The properties of the SAR signal itself (wavelength and polarization) influence its interaction with forest architecture and can provide detailed information about forest structure. As a general rule, longer wavelengths penetrate further than shorter wavelength SAR data. For example, X- and C-band SAR data are generally scattered by the upper canopy of a forest whereas L- and P-band tend to interact only with large stems and trunks (Figure 2.4). This phenomenon in particular makes SAR data at longer wavelengths useful in assessing changes in forest biomass, however because of signal saturation, the relationship only applies to forests with biomass <150 Mg ha^{-1} (Koch, 2010; Mitchard et al., 2009; Saatchi et al., 2011).

In addition to wavelength, the polarization information associated with the SAR signal – that is the orientation of the plane of propagation (horizontal and vertical) – also impacts how the signal interacts with objects on the ground (and thus its backscatter). Briefly, there are three types of scattering: (1) rough surface, (2) double-bounce, and (3) volume (Figure 2.5). Each polarimetric channel tends to favour certain scattering mechanisms, such that the strength of the backscatter in specific channels can give an indication of the primary scattering mechanisms (Flores-Anderson et al., 2019). This information can greatly assist image classifications, as particular habitats, vegetation structures, land forms, soils types, etc. can produce unique combinations of backscatter across channels. Within the context of this thesis, we predicted discernible signals associated with SAR polarization, but also differences in logged and unlogged areas resulting from changes in backscatter after logging.

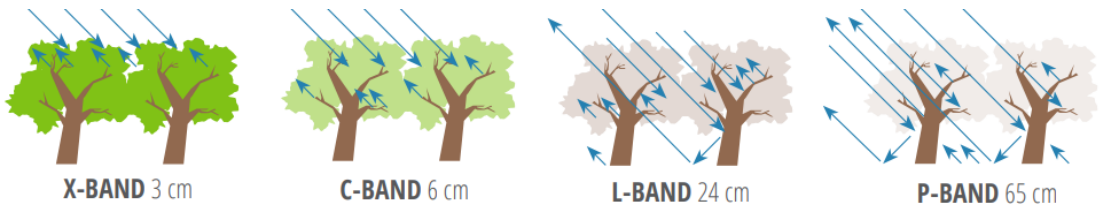


Figure 2.4 Differences in SAR signal interaction with forest structure associated with wavelength (adapted from Flores-Anderson et al. 2019).

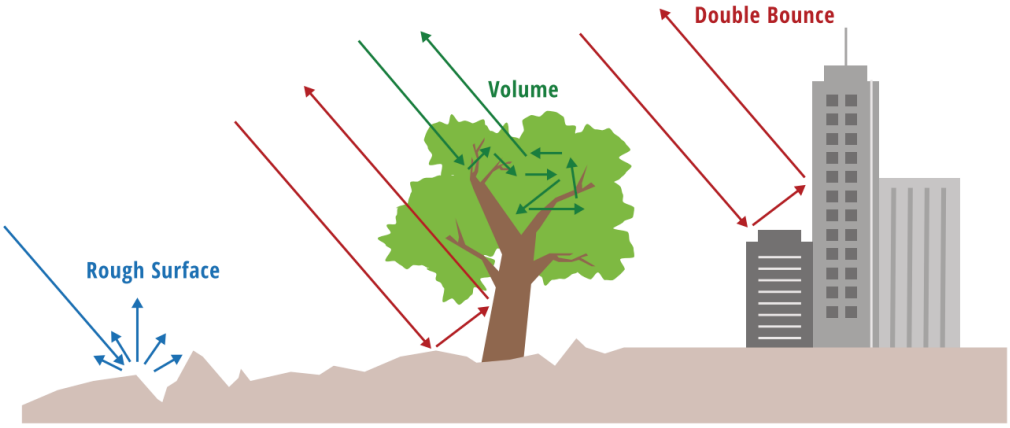


Figure 2.5 The three main scattering types for SAR data (from Flores-Anderson et al. 2019).

2.2.2 Spatial context

Mapping any land surface phenomena that operates across spatial scales, from a few pixels to globally abundant, can be aided by combining spectral and spatial information. That is, in addition to an individual pixel's spectral information, the spectral properties of adjacent or nearby pixels can help to further discriminate boundaries or changes in the arrangement of land surface features. A common way to include spatial-spectral information is by calculating image texture, essentially spatial variations in greyscale levels within a neighborhood. Such metrics can be grouped into what are known as first-order and second-order statistics, and are summarized for an individual pixel and thus preserve the spatial resolution of the original data. First-order statistics are the familiar summaries, like mean, variance, standard deviation, etc. within a particular window or neighborhood size. Second-order statistics summarize the grey-level differences between pairs of pixels within the neighborhood (Figure 2.6). A Grey Level Co-occurrence Matrix (GLCM) summarizes probabilities of co-occurring grey level pairs in various directions (Haralick et al., 1973). A suite of metrics can be calculates from the GLCM and their use has been widespread in remote sensing applications since their initial formulation, often improving per-pixel classifications (Hall-Beyer, 2017).

Second-order texture measure are used throughout this thesis to improve detection of selective logging activities. Logging affects patches of forest, not isolated pixels and context matters, since roads and skid trails accompany canopy gaps in logged forests. Moreover, the spatial resolution of Landsat and the freely available SAR imagery are such that a canopy gap may occupy more than a single pixel. There is, however, a need to better understand and integrate the particular forest attributes that are captured by the GLCM statistics. These metrics are often employed to improve classification, but their interpretation often remains an abstraction (Hall-Beyer, 2017). While we agree using GLCM metrics requires consideration, an

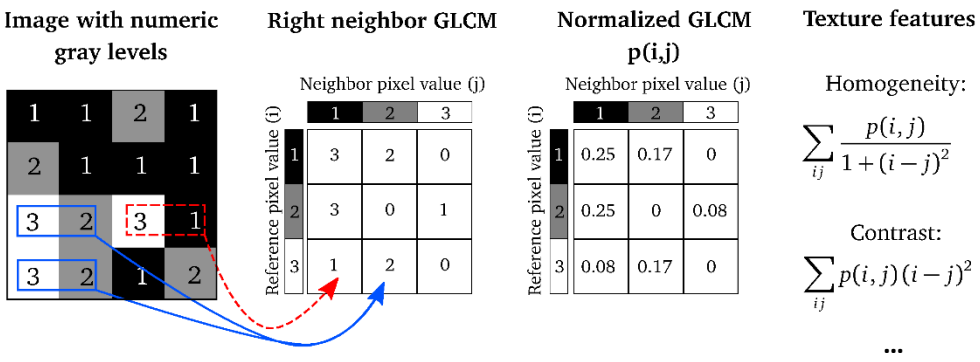


Figure 2.6 An illustration of how Haralick texture measures are computed. In a 4 × 4 image, three grey-levels are represented by numerical values from 1 to 3. The GLCM is constructed by considering the relation of each pixel to its neighborhood. In this simple example we only look at the neighbor to the right, however a rotationally invariant form is typically calculated. The GLCM acts like a counter for every combination of grey-level pairs in the image. For each pixel, its value and the neighboring pixel value are counted in a specific GLCM element. The normalized GLCM represents the estimated probability of each combination to occur in the image – since there are 12 right neighbor comparisons in this example. The Haralick texture measures are functions of the normalized GLCM, where different aspects of the grey-level distribution in the neighborhood are represented. For example, diagonal elements in the GLCM represent pixel pairs with equal grey-levels. The "Sum Average" texture measure is an average of the neighborhood, whereby individual pixel values are not weighed by its frequency of occurrence alone (as with the familiar mean), but by its frequency of occurrence in combination with certain neighbor pixel values. The "Contrast" equation results in elements with similar grey-level values having low weight, but elements with dissimilar grey-levels having high weight, this highlighting differences between combinations of shade in the neighborhood (figure adapted from Lofstedt et al., 2019; see Haralick et al., 1973 for notation details).

explicit understanding or detailed definition of the forest attributes captured by particular GLCM metrics may simply be beyond simple interpretation. Indeed, interpretation of tuning parameters associated with the hidden layers of neural networks are deeply abstract and often beyond simple interpretation. Nevertheless, we agree that a deeper understanding of GLCM interpretations are generally needed within the remote sensing community.

2.3 Machine learning theory

2.3.1 Overview

The objective of training a machine learning algorithm is to have it perform well, not just on the training data, but on new data it has never seen before. Thus, *generalizable* models are sought (i.e. neither under nor overfit). Machine learning methods could formerly be subdivided into two broad categories; supervised and unsupervised approaches, but the field has since widened to include varieties that do not fit neatly into either (e.g. semi-supervised, reinforced learning, self learning, meta-learning, etc.). We briefly touch on a few different approaches in the next section, however, this thesis exclusively utilizes supervised learning methods.

2.3.2 Machine learning approaches

Machine learning methods, including regression trees, support vector machines, and Random Forests, for classification of satellite imagery have been used for over two decades and can turn a suite of predictor variables weakly correlated with a response into a relatively strong classifier (Schulz et al., 2018). Throughout this thesis we utilized an ensemble supervised machine learning method known as Random Forests (Breiman, 2001). Over the last few years, owing to increased computational power, machine learning ensemble methods have become increasingly popular in remote sensing applications (Belgiu et al., 2016). Ensemble classifiers utilize two main approach, bagging (Breiman, 1996) and/or boosting (Schapire, 1990). Bagging involves training models on random subsets of the training data, in parallel, to make a conclusion about an observation that represents the average of all the decision trees built. In contrast, boosting is an iterative process whereby all the training data are used to the build a model and the misclassified observation are given higher emphasis in the next iteration (Schapire, 1990). The development of ensemble methods offers an opportunity to move beyond deterministic decision trees or simple rule-based thresholding for classifying changes (e.g. Asner et al., 2005; Souza et al., 2013) in remotely sensed imagery (Belgiu et al., 2016). However, these methods require adequate training data and sufficient datasets have previously not been available. The dataset we have compiled on selective logging, across a range of logging intensities, provides an opportunity to assess the suitability of ensemble methods for detecting subtle forest disturbances.

Most recently, a sub-genre of the broader family of machine learning methods, known as deep learning, has shown real promise for the future of remote sensing applications (Brodrick et al., 2019). In particular, convolutional neural networks (CNNs) have become one of the most popular methods for classifying remotely sensed imagery because of the way spatial context information is used to identify features (Brodrick et al., 2019; Nogueira et al., 2017). However, CNNs (and other deep learning methods) require large numbers of labelled images and are computationally demanding, making them impractical to design from scratch for remote sensing applications (Cheng et al., 2017; Nogueira et al., 2017). Indeed, the use of pre-trained CNNs with little to no modification have shown encouraging results recently (Nogueira et al., 2017), though the problem of detecting small objects amongst dense scenes (e.g. selective logging within continuous tropical forest) persists (Fu et al., 2019). Ultimately, however, CNNs (and other deep learning methods) are still so new that their development is somewhat limited to proof-of-concepts and they have yet to be made scalable within platforms like Google Earth Engine, unlike the ensemble methods mentioned previously (Gorelick et al., 2017).

2.4 Detection theory

Throughout this thesis we utilize a number of concepts from detection theory, a broad field with applications spanning signal processing, psychology, game theory, medicine, artificial intelligence, and others. More specifically, we rely on theory surrounding the sensitivity or discriminability of a signal. Conceptually, sensitivity refers to how easy it is to detect a particular signal within background noise. Different signals can be rare but strong, common and faint, or anything in between. The signal of selective logging within the Amazon basin is almost certainly faint, but the prevalence is not well known because of current inabilities to reliably detect it. Nevertheless, selective logging is assumed to be fairly uncommon, relative to the vastness of the Amazon; 10-20,000 km^2 of logging per year (Asner et al., 2005) is approximately 0.25% annually. While selective logging is almost certainly more common than this, we have regarded it as a rare event for the purposes of defining the framework by which we built our detection system.

Classifying a single pixel as logged or unlogged (i.e. making a prediction about its status) can be thought of like a hypothesis – one that can be tested using field data on each pixel's logging history. Thus, classifying a satellite image represents potentially millions of hypothesis tests. In a two-class problem such as this, millions of tests can result in potentially hundreds of thousands of false positive results (i.e. labelling a truly unlogged pixel as logged; Table 2.2). This is obviously unacceptable and is the primary reason we focused on methods to control the Type-I error rate (i.e. falsely rejecting the null hypothesis of unlogged forest). We rely on two concepts within detection theory that relate to hypothesis testing, defining acceptable rates of *false discovery* (Benjamini & Hochberg, 1995) and *false alarm* (Figure 2.7).

It may not be immediately noticeable, but choosing a false alarm rate (FAR) is the way null hypotheses are typically tested. An alpha is chosen, often 0.05, such that anything below this value is the rejection region (or a detection of logging in our case). Thus, a tolerable error rate is pre-determined and the detection rate (DR) is a direct a consequence of this value (i.e. one cannot control both the false alarm and the detection rate). Constant FARs have a long

Table 2.2 Confusion matrix outlining the definitions of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), detection rate (DR), false alarm rate (FAR), and false discovery rate (FDR) in the detection of selective logging used throughout this thesis.

DR = TP/(TP+FN) FAR = FP/(FP+TN) FDR = FP/(TP+FP)		**Reference**	
		Logged	Unlogged
Predicted	Logged	*TP*	*FP*
	Unlogged	*FN*	*TN*

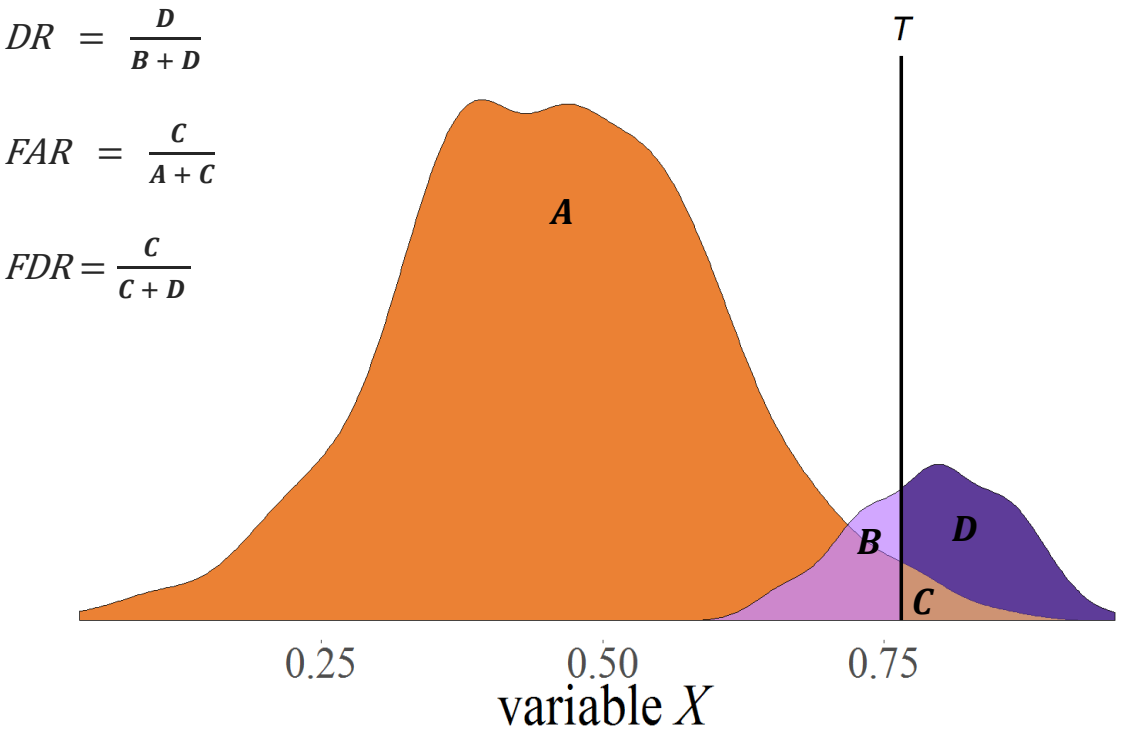


Figure 2.7 Diagram representing the trade-off between the detection rate (DR) and the false alarm rate (FAR) associated with using a threshold *T* (vertical black line) in the value of an arbitrary variable (*X*) to label pixels as logged and unlogged. The purple and orange colors correspond to density plots (scaled by the abundance of each class) for hypothetical logged and unlogged observations, respectively. The areas A and B are observations of unlogged and logged pixels, respectively, that will be labelled as unlogged. Similarly, C and D are observations of logged and unlogged pixels, respectively, that will be labelled as logged. Thus, region B is Type-II error and region C is Type-I error.

history in radar detection systems and are useful even when no targets exist (i.e. there is only background noise - or unlogged forest in our case). Linking it to an example of detecting selective logging, let us assume we want to define a threshold of canopy cover to map logging. We need to select a suitable canopy cover value that does not result in too many false detections of logging, so we might say we were willing to accept a 2% FAR. Thus 2% of all unlogged forest that was analyzed would be incorrectly labelled as logged. Crucially here, the FAR rate is the proportion of all unlogged pixels that are false positives (Figure 2.7, Table 2.2). This kind of selection process would be useful when one does not have field data on logging, yet a threshold of canopy cover needed to be identified. One could use a national park or some region known to have remained unlogged and identify a threshold value that produces the desired FAR. Thus, using a constant FAR, unlike FDRs, does not require both classes to be present and makes them extremely flexible.

In contrast to FARs, setting a FDR focusses on all the *detections* (i.e. everything above the threshold in Figure 2.7) and defining an acceptable proportion that are wrong. This difference is

subtle, but FDR controls the Type-I error rate and requires both classes to be present (i.e. any detection of logging in unlogged forest is wrong and thus a 100% FDR even with 1 false detection). Throughout this thesis we generally use FDRs, but discuss applications or scenarios where using a FAR would be more appropriate. Controlling Type-I error (i.e. using FDR) has an intuitive appeal over controlling the FAR in that the size or amount of unlogged forest is not known. Therefore defining an acceptable proportion of an unknown, presumable large quantity (unlogged forest in the Amazon) poses non-trivial uncertainty. In contrast, emphasizing detections, and the proportion of those detection that were wrong (i.e. FDR) presented a tractable solution to dealing with unknown quantities. The use of FDRs is not extremely common in remote sensing applications. However, given the imbalance in the abundance of the two land cover classes (logged and unlogged forest), the consequences of a particular fixed error rate can be dramatically different for the more abundant class (i.e. 10% of millions is many more than 10% of thousands). Thus, the use of FDRs and FARs

2.5 Conclusion

This thesis integrates the range of concepts outlined within this chapter. In particular, we explore changes in spatial-spectral features, aided by current machine learning models and key concepts from detection theory, to improve current abilities to detect and map tropical selective logging. We have tried to maintain a wide view of these concepts, however, specific details about each can be found within the text of each chapter as they relate to particular advancements.

Chapter

3

A machine learning approach to map tropical selective logging

MG Hethcoat, DP Edwards, JMB Carreiras, RG Bryant, FM França, and S Quegan

Published in *Remote Sensing of Environment*, 221 (2019), 569-582

DOI: 10.1016/j.rse.2018.11.044

Abstract

Hundreds of millions of hectares of tropical forest have been selectively logged, either legally or illegally. Methods for detecting and monitoring tropical selective logging using satellite data are at an early stage, with current methods only able to detect more intensive timber harvest (>20 m^3 ha^{-1}). The spatial resolution of widely available datasets, like Landsat, have previously been considered too coarse to measure the subtle changes in forests associated with less intensive selective logging, yet most present-day logging is at low intensity. We utilized a detailed selective logging dataset from over 11,000 ha of forest in Rondônia, southern Brazilian Amazon, to develop a Random Forest machine-learning algorithm for detecting low-intensity selective logging (< 15 m^3 ha^{-1}). We show that Landsat imagery acquired before the cessation of logging activities (i.e. the final cloud-free image of the dry season during logging) was better at detecting selective logging than imagery acquired at the start of the following dry season (i.e. the first cloud-free image of the next dry season). Within our study area the detection rate of logged pixels was approximately 90% (with roughly 20% commission and 8% omission error rates) and approximately 40% of the area inside low-intensity selective logging tracts were labelled as logged. Application of the algorithm to 6152 ha of selectively logged forest at a second site in Pará, northeast Brazilian Amazon, resulted in the detection of 2316 ha (38%) of selective logging (with 20% commission and 7% omission error rates). This suggests that our method can detect low-intensity selective logging across large areas of the Amazon. It is thus an important step forward in developing systems for detecting selective logging pan-tropically with freely available data sets, and has key implications for monitoring logging and implementing carbon-based payments for ecosystem service schemes.

3.1 Introduction

Earth's tropical forests are being rapidly lost and degraded by agricultural expansion and commercial logging operations, with population growth projected to further increase pressures on forests globally (Asner et al. 2005; DeFries et al. 2010). The ability to monitor forest disturbances is an important component in sustainable forest management, understanding the global carbon budget, and implementing climate policy initiatives, such as the United Nation's (UN) Reducing Emissions from Deforestation and Forest Degradation (REDD+) programme, which seeks to mitigate climate change and biodiversity losses through improved forest management practices (GOFC-GOLD, 2016). The UN anticipates that payments to nations under REDD+ initiatives, which compensate countries for conserving forests (and sequestering carbon), could reach $30 billion annually (Phelps et al., 2010, UN-REDD Programme, http://www.un-redd.org).

Remote sensing is considered the most accurate and cost-effective way to systematically monitor forests at broad spatial scales (Herold and Johns 2007; Achard et al. 2007; Shimabukuro et al. 2014). Large-scale monitoring of deforestation has significantly improved in recent years, and forest losses can be identified with accuracies greater than 90% using freely available satellite data (Hansen et al. 2013). In addition, near real-time deforestation tracking and alert systems are now possible with systems like DETER (Shimabukuro et al., 2012), FORMA (Hansen et al. 2013; Hammer et al. 2014), and Global Forest Watch (Hansen et al. 2016). In contrast, methods for detecting and monitoring forest degradation are less developed. Forest degradation is an ambiguous term, with over 50 different definitions and no internationally established description (Simula 2009; Ghazoul et al. 2015). This makes generalizing its impacts difficult, in part because degradation can include forests subject to varying intensities of selective logging, fire, artisanal gold mining, fuelwood extraction, etc., which has hampered the development of coordinated international forest policies to track and monitor forest degradation (Sasaki and Putz 2009; Ghazoul et al. 2015).

Here we focus on detecting a key driver of forest degradation globally, commercial logging operations. In contrast to forest clearance (i.e. deforestation), selective logging represents a more diffuse disturbance wherein only a subset of trees (typically the most economically valuable) are harvested (Putz et al. 2001; Fisher et al. 2014). The resulting forest maintains some degree of its original composition (e.g. canopy cover, biodiversity measures, carbon content, etc.) but is punctured by treefall gaps and logging roads and consequently lies on a continuum between primary forest and complete deforestation (Thompson et al. 2013; Ghazoul et al. 2015). The intensity of selective logging operations can vary in two main ways: (1) the volume of wood harvested typically ranges up about 50 m^3 ha^{-1}, as high as 150 m^3 ha^{-1} in Asia (Putz et al. 2001; Burivalova et al. 2014) and (2) the degree to which reduced-impact

logging is practiced, in which damage to the remaining forest is minimized by careful planning of road networks, skid trails, and directional felling of trees to limit additional tree or canopy damage (Putz and Pinard 1993). We acknowledge wood biomass can vary substantially across forest types globally and may not, by itself, be a perfect indicator of logging intensity. However, in this manuscript we define logging intensity in terms of wood volume extracted to be consistent with legal restrictions outlined in the Brazilian forest code.

Selective logging activities are often the first anthropogenic disturbance to affect primary tropical forests (Nepstad et al. 1999; Asner et al. 2009b) and are thought to be a major source of carbon emissions from degradation (Hosonuma et al. 2012; Pearson et al. 2017). Moreover, road networks associated with logging are often precursors to additional land-use changes (such as agricultural conversion or development of human settlements) and facilitate further degradation (e.g. increased susceptibility to fires or illegal logging) and forest losses (Matricardi et al. 2010; Kumar et al. 2014; Alamgir et al. 2017). Estimates suggest over 400 million ha of tropical forest, an area the size of the European Union, are earmarked in the tropical timber estate to be logged (Blaser et al. 2011). However, the extent of forest subjected to selective logging across the tropics has yet to be estimated (Asner et al. 2005).

Several authors have tried to address the challenges of using satellite data to estimate forest disturbances from selective logging in the tropics (Souza and Barreto 2000; Asner et al. 2002, 2004a; Souza et al. 2005; Asner et al. 2005; Matricardi et al. 2007, 2010; Shimabukuro et al. 2014). The majority of approaches employ classification of fractional images derived from spectral unmixing of Landsat scenes. Despite these advancements, Landsat imagery has been considered too coarse to monitor less intensive selective logging activities, with nearly all applications involving logging intensities > 20 m^3 ha^{-1} (Souza and Barreto 2000; Asner et al. 2002, 2004a; Souza et al. 2005; Asner et al. 2005; Matricardi et al. 2007, 2010; Shimabukuro et al. 2014). While most authors acknowledge their methods can detect areas of selective logging at moderately high intensities (> 20 m^3 ha^{-1}; 3-7 trees ha^{-1}), that possess large canopy gaps and an abundance of spectrally distinct features, like log landing decks or large road networks, their respective abilities to detect lower logging intensities are unknown. Therefore, using Landsat data to map and quantify selective logging at lower logging intensities (< 20 m^3 ha^{-1}) remains a major challenge, and the amount of forest disturbance overlooked using currently available techniques is unknown. Yet, growing concerns over the impacts of selective logging on carbon and biodiversity (Putz et al. 2008; Bicknell et al. 2014; Edwards et al. 2014; Martin et al. 2015; França et al. 2017) has led to increased use of improved forest management practices, such as reduced-impact logging (Putz and Pinard 1993). Consequently, the extent of tropical forests being logged at lower intensities and with reduced-impact is almost certainly expanding. In addition, there is an ever-increasing need to detect and account for the estimated 50-90% of

tropical timber on the international market harvested illegally at very low intensities (Kleinschmit et al. 2016; Brancalion et al. 2018). Therefore methods to detect subtle forest disturbances from satellite systems with regular global coverage are urgently needed, both to establish reference levels from historical data (e.g. the vast amount of freely available Landsat archives) and to obtain maximum benefit from current and future systems, such as Landsat 8, 9 and Sentinel-2 (Drusch et al. 2012; Roy et al. 2014).

The primary objective of this study was to develop a new method for detecting selective logging in moist tropical forest with Landsat data. It focuses on reduced-impact selective logging of intensity < 15 m^3 ha^{-1} (1-2 trees ha^{-1}), much lower than is typically reported in studies that use remote sensing data to estimate selective logging (Asner et al. 2004a; Souza and Roberts 2005; Asner et al. 2005), but still more than three times the background rate of natural mortality estimated for tropical forests (Clark et al. 2004; Brienen et al. 2015). We used detailed spatial and temporal logging records from Rondônia, Brazil, together with Landsat data, to build a machine learning algorithm for detecting selectively logged Landsat pixels. Machine learning (neural networks, decision trees, support vector machines, etc.) for classification of satellite imagery has been used with increasing success in recent years (Tuia et al. 2011) and can turn a suite of predictor variables weakly correlated with a response into a relatively strong classifier (Breiman 2001). The successful application of this algorithm to a test site in northern Pará, Brazil, approximately 1500 km from the location of algorithm development, demonstrates that this approach is transferable and can greatly improve existing methods of detecting subtle selective logging activities in the tropics.

3.2 Study sites and satellite imagery

Data from two test sites in the Brazilian Amazon were used in this study (Figure 3.1a). The Jamari site consists of *terra firme* tropical forest inside the Jamari National Forest, Rondônia, Brazil. The logging concession was subdivided into forest management units (FMUs) that were each approximately 2,000 ha (Figure 3.1b). Selective logging occurred within a single FMU in each year, at an intensity of approximately 10 m^3 ha^{-1} (1-2 trees ha^{-1}), beginning at the end of the wet season (roughly June) and continuing through the dry season (until November) from 2011 through 2015. Forest inventory measurements were recorded by trained foresters and included the spatial location of each marketable. At the Jamari site, heavy cloud cover typically occurs between October and May, but cloud-free images from Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper (ETM+), and Landsat 8 Operational Land Imager (OLI) were acquired approximately annually for 2008 to 2016 in the intervening dry season (Table 3.1). Note that the 2012 ETM+ images suffered from missing data as a result of the scanline corrector error and appear striped (Storey et al. 2005). For the analyses, we distinguished

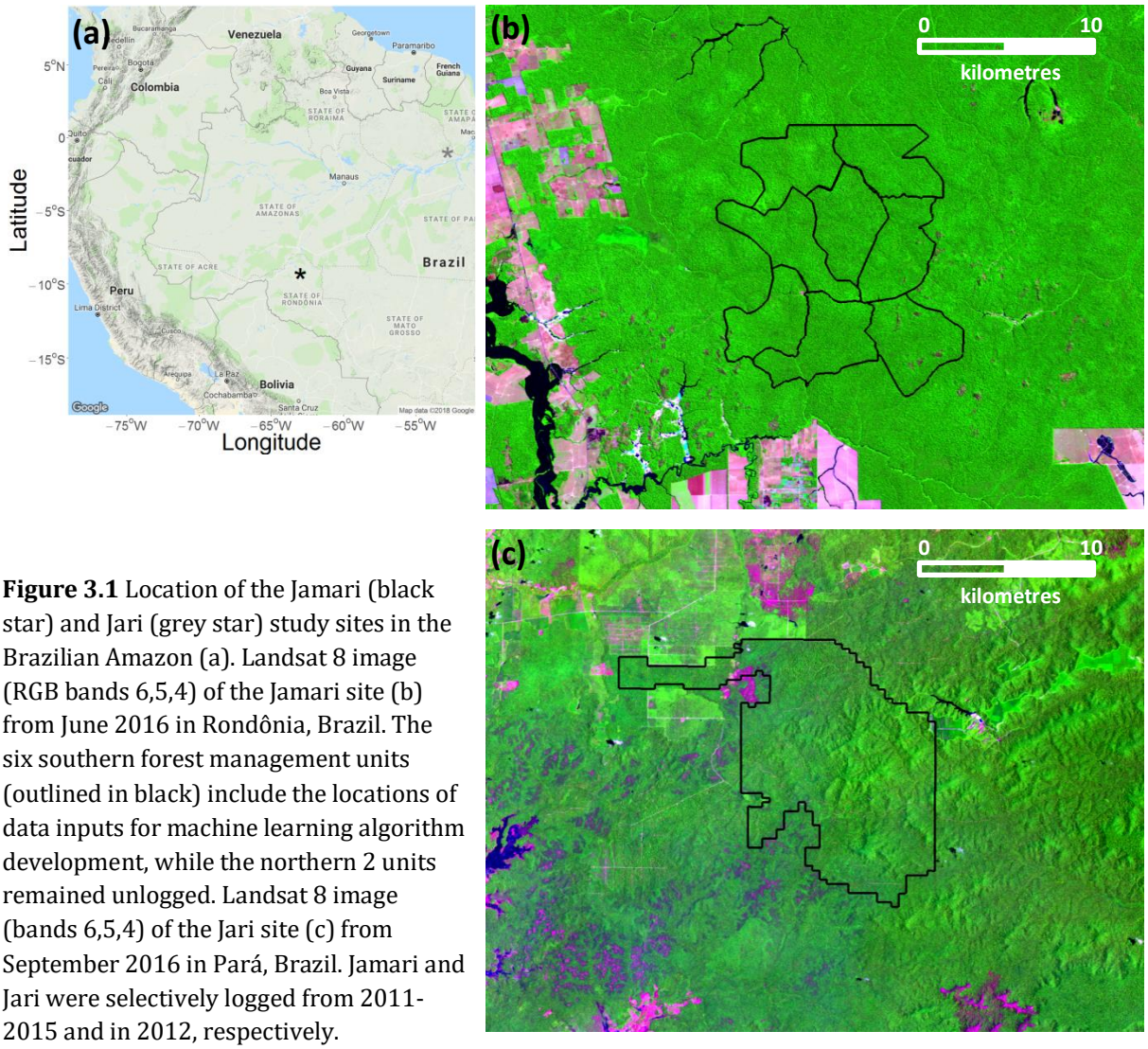


Figure 3.1 Location of the Jamari (black star) and Jari (grey star) study sites in the Brazilian Amazon (a). Landsat 8 image (RGB bands 6,5,4) of the Jamari site (b) from June 2016 in Rondônia, Brazil. The six southern forest management units (outlined in black) include the locations of data inputs for machine learning algorithm development, while the northern 2 units remained unlogged. Landsat 8 image (bands 6,5,4) of the Jari site (c) from September 2016 in Pará, Brazil. Jamari and Jari were selectively logged from 2011-2015 and in 2012, respectively.

"early" and "late" images for a given region. The early image was the last cloud-free image of the dry season in the *same year* the FMU was logged (typically in August, approximately 2-3 months before cessation of logging activities for the season). The late image was the first cloud-free image of the dry season in the *year after* cessation of logging activities (typically in June, approximately 8-12 months after the FMU was logged). We used early and late imagery to generate two separate datasets and build two separate algorithms in order to assess which time period provided better detection of selective logging. This is illustrated for a hypothetical logging season in Figure 3.2. The selection of two time periods reflects the fact that after 8-12 months, regrowth of foliage and other vegetation can reduce the spectral signatures required to identify canopy gaps and woody debris in tropical systems (Asner et al. 2004b, a; Broadbent et al. 2006).

The Jari site (Figure 3.1c) in Pará, Brazil, consists of *terra firme* tropical forest inside the 12,500 ha Jari concession that was selectively logged at an intensity of approximately 12 m^3 ha^{-1} (1-3 trees ha^{-1}) between July and December 2012. In contrast to Jamari, the Jari site lacked detailed information on where trees were removed, but the volume of wood (m^3) removed was recorded for 10 ha (400 m x 250 m) blocks in the concession. The Jari site allowed us to assess whether the algorithms developed using the Jamari dataset, located approximately 1500 km away, were transferable to this distant site. At Jari heavy cloud cover is common throughout the year, but we used the early and late time period imagery with the lowest cloud cover available to assess logging before and after logging activities occurred within the FMU (Table 3.1).

Table 3.1 Landsat 5 (TM), 7 (ETM+), and 8 (OLI) scenes used to build and assess Random Forest models developed to detect selective logging. The Jamari study site is path 232, row 066 and the Jari site is path 226, row 061.

Study Site	Acquisition Date	Scene Timing	Solar Zenith Angle	Landsat Sensor
Jamari	2008-07-28	Early	49.75	TM
	2009-07-31	Early	50.00	TM
	2010-07-18	Early	46.36	TM
	2011-08-06	Early	51.67	TM
	2012-08-16	Early	54.05	ETM+
	2013-08-27	Early	57.07	OLI
	2014-08-30	Early	58.84	OLI
	2015-09-02	Early	60.19	OLI
	2009-06-29	Late	43.79	TM
	2010-07-02	Late	43.63	TM
	2011-07-05	Late	44.30	TM
	2012-06-13	Late	41.64	ETM+
	2013-07-10	Late	42.26	OLI
	2014-06-11	Late	40.43	OLI
	2015-06-14	Late	40.47	OLI
	2016-06-16	Late	40.37	OLI
Jari	2011-11-08	Early	123.31	ETM+
	2012-11-10	Early	125.27	ETM+
	2011-07-03	Late	48.12	ETM+
	2013-08-17	Late	60.92	OLI

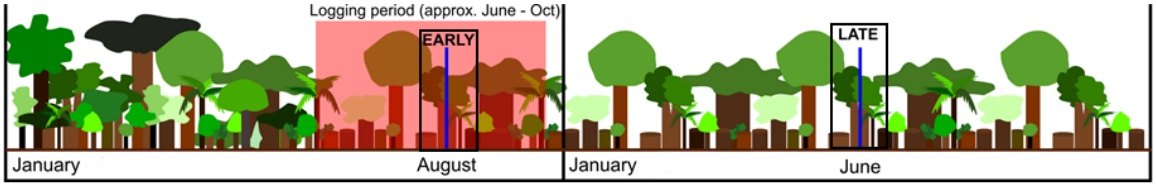


Figure 3.2 Timeline representation of a single forest management unit in the Jamari study site. Vertical blue lines indicate image acquisitions during the early and late time periods (black boxes) relative to when logging occurred (red box). In this example the early Landsat image was acquired part way through the logging season, so part of the management unit has yet to be cut. The late image is the first cloud-free image of the following dry season and is acquired approximately 8 months after the management unit was selectively logged.

3.3 Methods

3.3.1 Data inputs for detecting selective logging

For the Landsat scenes given in Table 3.1, the surface reflectance values for the Blue, Green, Red, Near Infrared, Shortwave Infrared 1 and Shortwave Infrared 2 bands were measured at each pixel where logging occurred (n = 13699) and 2000 randomly selected pixels in an adjacent FMU that remained unlogged. In addition, since logging activities tend to be accompanied by surrounding disturbances (residual damage to neighbouring unharvested trees and skid trails along which logs are extracted), seven texture measures were calculated for each band (mean, variance, homogeneity, contrast, dissimilarity, entropy, and second moment) to provide a local context for each pixel (Haralick et al. 1973; Castillo-Santiago et al. 2010; Beekhuizen and Clarke 2010; Rodriguez-Galiano et al. 2012). These were calculated within a 7x7 pixel window, chosen as a trade-off between minimizing window size while still capturing the disturbances in a selectively logged forest compared to an unlogged forest (see Section 3.4.1 for a brief comparison of larger and smaller window sizes). The various texture metrics were assigned to the centre pixel, thus maintaining pixel size (i.e. 30 m), and were added after preliminary modelling efforts with only the surface reflectance bands were found to perform inadequately (i.e. approximately double the rate of omission error of logged pixels; see Tables A1.1 - A1.4 for details). Because of possible Landsat inter-sensor differences, we added one final categorical variable that represented the sensor (TM, ETM+, or OLI) from which the image was acquired. The dataset thus comprised a 49-element vector (6 surface reflectance bands, 7 texture measures for each band, and a sensor-type indicator) for each pixel where logging occurred and an additional 2000 randomly selected pixels in an adjacent FMU that remained unlogged between 2008 and 2016.

The early and late datasets were reduced to exclude data from time periods close to when each FMU was logged. In the early dataset, for each FMU we excluded data from the year before logging because access roads were built and pixel values would therefore not represent

undisturbed forest. In addition, data from all years following logging were excluded (see Table A1.5 for details). For example, for an FMU logged in 2014 the early dataset comprised data from around August in 2008 through 2012 (representative of unlogged conditions) and August 2014 (representative of logged conditions), but excluded data from August 2013, 2015 and 2016. The same procedure was used for the late dataset. For example, for an FMU logged in 2014 the unlogged dataset included data acquired around June in 2008 through 2012, while the logged data was for June 2015. Data were excluded from June 2013 (roads being built in the FMU), 2014 (logging recently initiated), and June 2016 (2 years post-logging). In both the early and late datasets the data from 2000 randomly selected pixels in an adjacent FMU that remained unlogged were retained from all years because they were never logged. Note that for early data, the imagery was acquired before the final part of the FMU was logged; this introduced some errors into model training, because some pixels labelled as logged in the training data were still unlogged. Despite this, we demonstrate in Section 3.4.1 that detection of selective logging was better with early time period data.

3.3.2 Random Forest for detection of selective logging

We built Random Forest (RF) models using the *randomForest* package in program R version 3.3.1 (Liaw and Wiener, 2002; R Development Core Team, 2016). The RF algorithm (Breiman 2001) is a machine learning technique that uses an ensemble method to identify a response variable (here, whether a pixel was logged or unlogged) given a set of predictor variables (e.g. surface reflectance values). In contrast to a single decision tree, RF models employ multiple, independent decision trees (hence a forest). Random subsets of the training data are drawn, with replacement, to construct many trees in parallel, with each tree casting a vote on which class should be assigned to the input data. The withheld subset of the data, called the out-of-bag fraction, can be used for validation in the absence of independent validation data (Breiman 2001). To reduce generalization error and minimize correlations amongst predictors, RF uses a random subset of predictor variables in the decision at each node within a tree during construction. Prior to model training, feature selection was performed with the Boruta package (version 6.0.0) in Program R (version 3.3.1) and all variables were deemed significant (Figure A1.1).

We split the early and late datasets into 75% for training and 25% was withheld for validation. We used the out-of-bag data during model training to determine the threshold value for classification (i.e. model calibration, see Section 3.3.3.1). In order to ensure independence, the training and validation datasets were spatially filtered such that no observations in the training dataset were within 90 m of an observation in the validation dataset. RF models have only two tuning parameters: the number of classification trees to be produced (k), and the

number of predictor variables used at each node (m). We used 10-fold cross-validation to identify the number of trees (k = 1000) and the number of variables to use at each node (m = 5) that minimized the out-of-bag error rate on the training data.

3.3.3 Algorithm evaluation

3.3.3.1 Calibration: selecting the detection threshold

RF models typically use a simple majority vote to assign an observation to a particular class, for example, in binary decisions when more than 50% of the trees assign a pixel to a particular class (Breiman 2001). However, the proportion of votes cast for a particular class from the total set of trees can be obtained for each pixel and a classification threshold can be applied to this proportion (Liaw and Wiener 2002). We adopted this approach here, wherein the proportion of votes that predicted each observation to be logged, denoted as X and informally termed the *likelihood* a pixel was logged, was used to select the classification threshold. Model calibration (with the out-of-bag data) was then used to define a threshold, T, such that if $X > T$ the pixel was classified as logged (Figure 3.3).

Detection of logging involves only two classes, logged and unlogged forest, so the confusion matrix has the form:

		Reference	
		L	UL
Predicted	L	D_L	D_{UL}
	UL	$N_L - D_L$	$N_{UL} - D_{UL}$

where L and UL refer to logged and unlogged, N_L and N_{UL} are the numbers of logged and unlogged observations in the reference dataset, and D_L and D_{UL} are respectively the numbers of logged and unlogged pixels detected as logged. The total number of observations is $N = N_L + N_{UL}$. Since logging is a relatively rare event, both in our data and on the landscape (i.e. $N_L << N_{UL}$), it is appropriate to use the terminology of detection theory. Accordingly, we define the *detection probability* $P_d = D_L/N_L$ and *false detection probability* $P_{fd} = D_{UL}/N_{UL}$ as the probabilities that a logged or unlogged pixel is classified as logged, respectively. P_d is equivalent to 1 – the omission error of the logged class and P_{fd} is the omission error of the unlogged class.

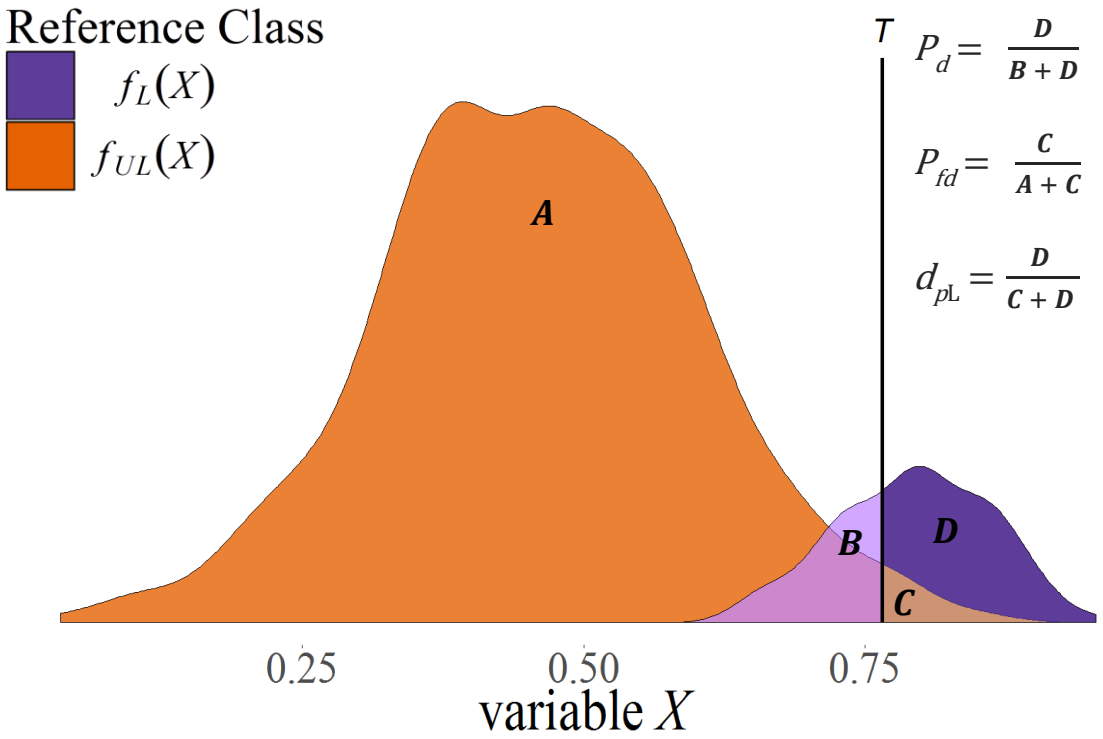


Figure 3.3 Diagram representing the trade-off between the probability of detection (P_d) and the probability of false detection (P_{fd}) associated with using a threshold T (vertical black line) on the variable X (the proportion of votes that predicted each observation to be logged) to label pixels as logged and unlogged. Here the purple and orange colors correspond to probability distribution functions of X for hypothetical logged, $f_L(X)$, and unlogged, $f_{UL}(X)$, observations, respectively (scaled by the sample size in each group). Thus, the areas A and B are the portions of the observations from unlogged and logged pixels, respectively, that will be labelled as unlogged. Similarly, C and D represent the portions of the observations from logged and unlogged pixels, respectively, that will be labelled as logged.

A pixel was classified as logged if X, the proportion of votes from RF that predict the pixel as logged, exceeds a given threshold T. Hence the detection and false detection probabilities depend on T and can be written

$$P_d(T) = \int_T^1 f_L(X)dX \quad (3.1a)$$

and

$$P_{fd}(T) = \int_T^1 f_{UL}(X)dX \quad (3.1b)$$

where $f_L(X)$ and $f_{UL}(X)$ are the probability distributions of X for the logged and unlogged classes, respectively (see Figure 3.3).

The selection of T involves a trade-off between increasing P_d and reducing P_{fd} (Figure 3.3). In making this choice, the overall accuracy, given by

$$A = \frac{D_L + (N_{UL} - D_{UL})}{N}, \tag{3.2}$$

is not a good guide, since it can be shown that A is maximal (equivalently, the overall probability of error is a minimum) when

$$\frac{f_L(X)}{f_{UL}(X)} = \frac{N_{UL}}{N_L}. \tag{3.3}$$

If N_L and N_{UL} were equal, the threshold would then be chosen at the intersection of $f_L(X)$ and $f_{UL}(X)$, but since $N_L << N_{UL}$ it has a much higher value (i.e. it moves to the right in Figure 3.3). This is because to increase overall accuracy it is more effective to reduce P_{fd} than to increase P_d, since there are so many more unlogged pixels (Schwartz 1984), and maximizing accuracy would lead to very few (or even no) detections. For example, if only 1% of an area was logged and all the pixels were classified as unlogged, the overall accuracy would be 99%. Thus, overall accuracy would not sufficiently balance the trade-off between true and false detections to meet our objectives.

Various criteria could be used to select a classification threshold, including maximizing Cohen's kappa (Cohen 1960) or defining an acceptable rate of omission error; ultimately however, there is no wrong threshold, since this depends on the objectives of prediction. The criterion used in this study to define T was to fix the proportion of detected pixels that were truly logged, defined here as d_{pL}:

$$d_{pL} = \frac{D_L}{D_L + D_{UL}} = \frac{1}{1 + \left(\frac{N_{UL}}{N_L}\right)\left(\frac{P_{fd}}{P_d}\right)}. \tag{3.4}$$

Adopting this criterion is equivalent to a Constant False Discovery Rate detector which is widely used in detection problems with rare events (Benjamini and Hochberg 1995; Neuvial and Roquain 2012). This fixes the rate of prediction error (i.e. type I) when labelling pixels as logged, because d_{pL} is equal to 1 minus the commission error of the logged class, thus limiting the rate of commission error. This approach enables the user to select the proportion of detections that will be false. It was chosen because in the detection of rare events (e.g. selective logging within the Amazon Basin, for example), the implications of a particular error rate when predicting over the majority class (i.e. unlogged forest) are greater than an equivalent error rate when predicting over the minority class (i.e. 10% of millions of unlogged pixels is far greater than 10% of thousands of selectively logged pixels). Thus, in order to avoid being swamped by false detections, we wanted to fix the proportion of all detected pixels that were incorrect and accept the level of accuracy associated with this criterion. The approach outlined here, therefore,

should be viewed from a detection theory perspective as opposed to simply being a classification problem.

Model calibration was used to calculate P_d, P_{fd}, and $d_{p\mathrm{L}}$ across the full range of threshold values. In practice this involved iterating through all values of T between 0 and 1 (in steps of 0.001), building each confusion matrix, and calculating the associated values of P_d, P_{fd}, and $d_{p\mathrm{L}}$. The threshold value was chosen such that $d_{p\mathrm{L}}$ = 0.85 in the training data (i.e. 15% of pixels classified as logged were actually unlogged). We initially set $d_{p\mathrm{L}}$ to 95% to strongly limit the rate of false detections, but this resulted in very high omission error of truly logged pixels (>75%). Consequently, $d_{p\mathrm{L}}$ was reduced to 0.85 by lowering the threshold, thus causing the detection and false detection rates to increase and causing more logged pixels to be detected. This value was then used to estimate P_d and P_{fd} during model assessment with the validation dataset.

3.3.3.2 Validation: assessing model accuracy

RF models were validated using a random, independent subset of the early and late datasets (described in Section 3.3.2). The threshold value of T, chosen during model calibration, was applied to the validation data and the associated error rates were calculated. The values of P_d, P_{fd}, and $d_{p\mathrm{L}}$ are presented across full range of threshold values to thoroughly illustrate model performance. Good practices outlined by Olofsson et al. (2014) were used to assess agreement and calculate unbiased error estimates when mapping selective logging detections. During mapping, non-forested areas were excluded using Brazil's national forest change product, PRODES (INPE 2015), and cloudy pixels were masked using the cloud mask provided with Landsat surface reflectance imagery. In addition, we provide the value of Cohen's kappa, κ, for comparison with other studies (Cohen 1960).

3.4 Results

3.4.1 Random Forest classification of selective logging at Jamari

The rates of true and false detection probabilities for the early and late validation data are shown in Figure 3.4 for the full range of T (black lines). These curves indicate how a given threshold value used for classification influenced the associated values of P_d, P_{fd}, κ, and $d_{p\mathrm{L}}$ in the validation data. For example, if a $d_{p\mathrm{L}}$ of 0.90 was used (indicating 10% of logging detections would be spurious) then the false detection rate (P_{fd}) would be < 1% for both datasets, but the detection rate (P_d) would be approximately 55% and 30% for the early and late datasets, respectively. These plots clearly demonstrate that there is no unambiguous way to choose an

optimal value for T, and the choice about its value is a trade-off between the number of true and false detections.

In general, these plots indicate that the early data provided a higher detection rate than the late data, for a given false detection rate. The early and late data had similar rates of commission error when labelling logged pixels, which is not surprising given we used this measure to constrain models during training. However, the late data had higher rates of omission error of logged pixels and detected less logging (Table 3.2). In addition, these plots demonstrate why using the threshold that maximized Cohen's κ would lead to higher false detection rates, as the threshold value is higher when $d_{p\mathrm{L}}$ = 0.85 than at maximum κ (i.e. pixels classified as logged must have a higher likelihood). Furthermore, because κ is high across a wide range of range of threshold values for both early and late data, slight differences in the likelihoods produced by the validation data could result in dramatic shifts in the value of T.

Although $d_{p\mathrm{L}}$ was fixed at 0.85 during model calibration (i.e. with the training data), the values calculated with the validation dataset were slightly lower (Table 3.2). Thus, the threshold value determined during model training did not produce the same values for $d_{p\mathrm{L}}$ when used against the validation dataset (i.e. some loss of performance). Slight differences in the proportion of logged observations (16.3% and 14.5% in training and validation, respectively) and minor differences in the ratio of P_{fd}:P_d between the training and validation datasets account for the disparity (see equation 3.4). In general model assessments seldom give identical performance across training and validation phases, and the difference here were marginal and yielded comparable model behavior.

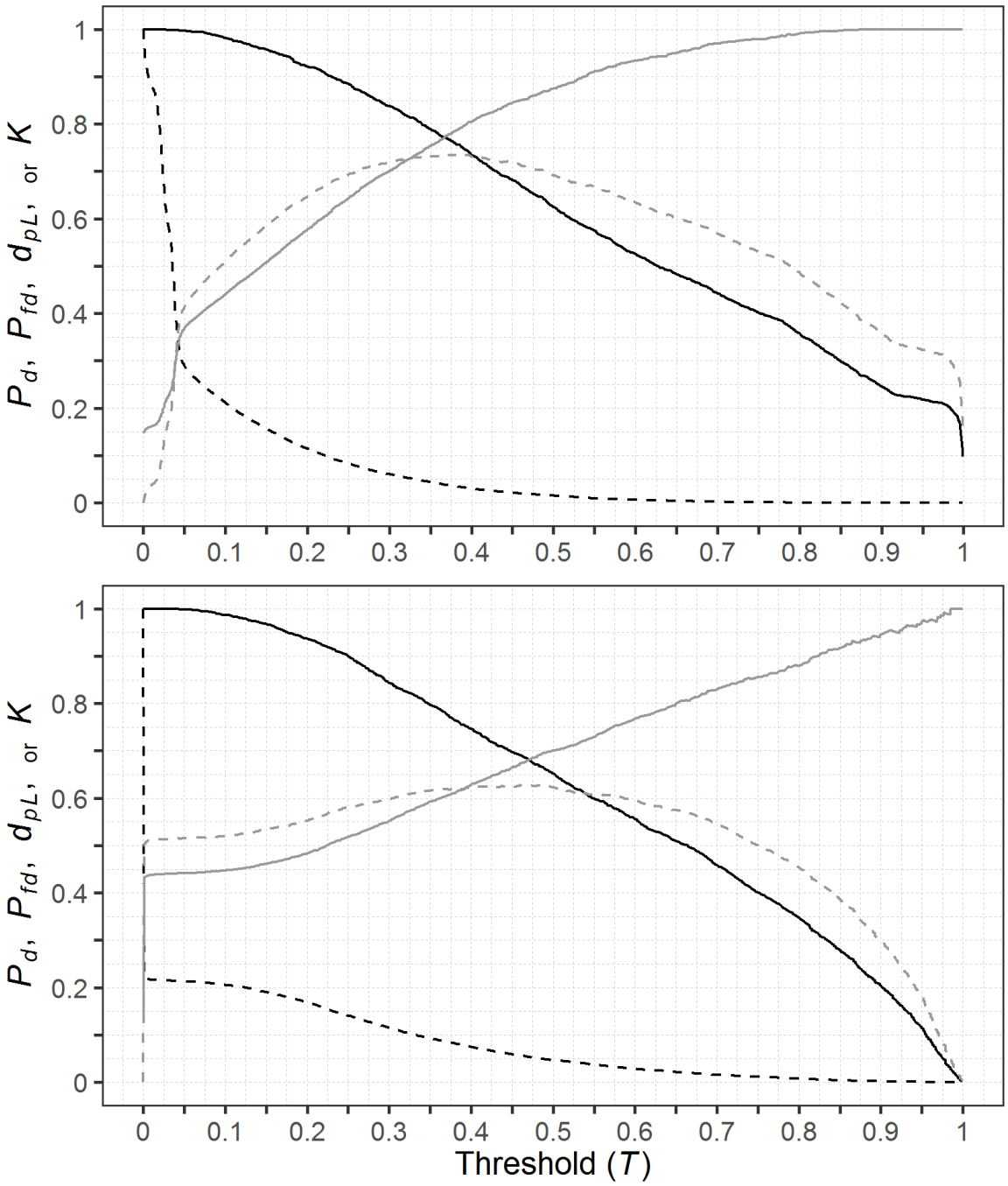


Figure 3.4 Trade-off curves between true (P_d) and false (P_{fd}) detection rates (solid and dashed black lines, respectively) for the early (top) and late (bottom) Random Forest models at the Jamari site as a result of varying the threshold value (T) for classification. Also shown are the corresponding values of d_{pL} (the proportion of detections that were truly logged) and Cohen's kappa (solid and dashed grey lines, respectively).

The early data displayed higher spatial correspondence between high likelihoods and the locations of logging in Jamari. This is illustrated in Figure 3.5, where the likelihood of logging provided by RF is shown on a colour scale and the individual locations of tree removal are indicated by black squares. The early model yields much higher likelihoods and these match well with reference logging data, whereas there is generally lower correspondence between reference logging locations and regions of highest likelihood in the late predictions. Note that we expect some logging locations to be omitted in the early data as the corresponding satellite data were acquired part way through the logging period and missed later logging. Evidence for this is provided by the inset regions expanded at the bottom of Figure 3.5 where the locations of the last 200 trees in the logging records for the season are displayed as crosses instead of

Table 3.2. Confusion matrix summarizing unbiased (Olofsson et al., 2014) results from Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging). The classification threshold (T) for RF models was set during model calibration such that the proportion of detections that were truly logged (d_{pL}) was fixed at 0.85, resulting in a T of 0.40 and 0.65 for the early and late datasets, respectively. The corresponding values for overall accuracy (OA), Cohen's kappa (κ), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

EARLY					**LATE**				
OA: 89.7%					OA: 91.7%				
κ: 0.78		Reference Class			κ: 0.40		Reference Class		
d_{pL}: 0.80		Logged	Unlogged	Commission	d_{pL}: 0.80		Logged	Unlogged	Commission
P_d: 0.92				Error (%)	P_d: 0.30				Error (%)
Predicted	Logged	0.313	0.076	19.5	Predicted	Logged	0.032	0.008	19.9
Class	Unlogged	0.027	0.584	4.4	Class	Unlogged	0.075	0.885	7.8
Omission Error (%)		8.0	11.5		Omission Error (%)		70.1	1.0	

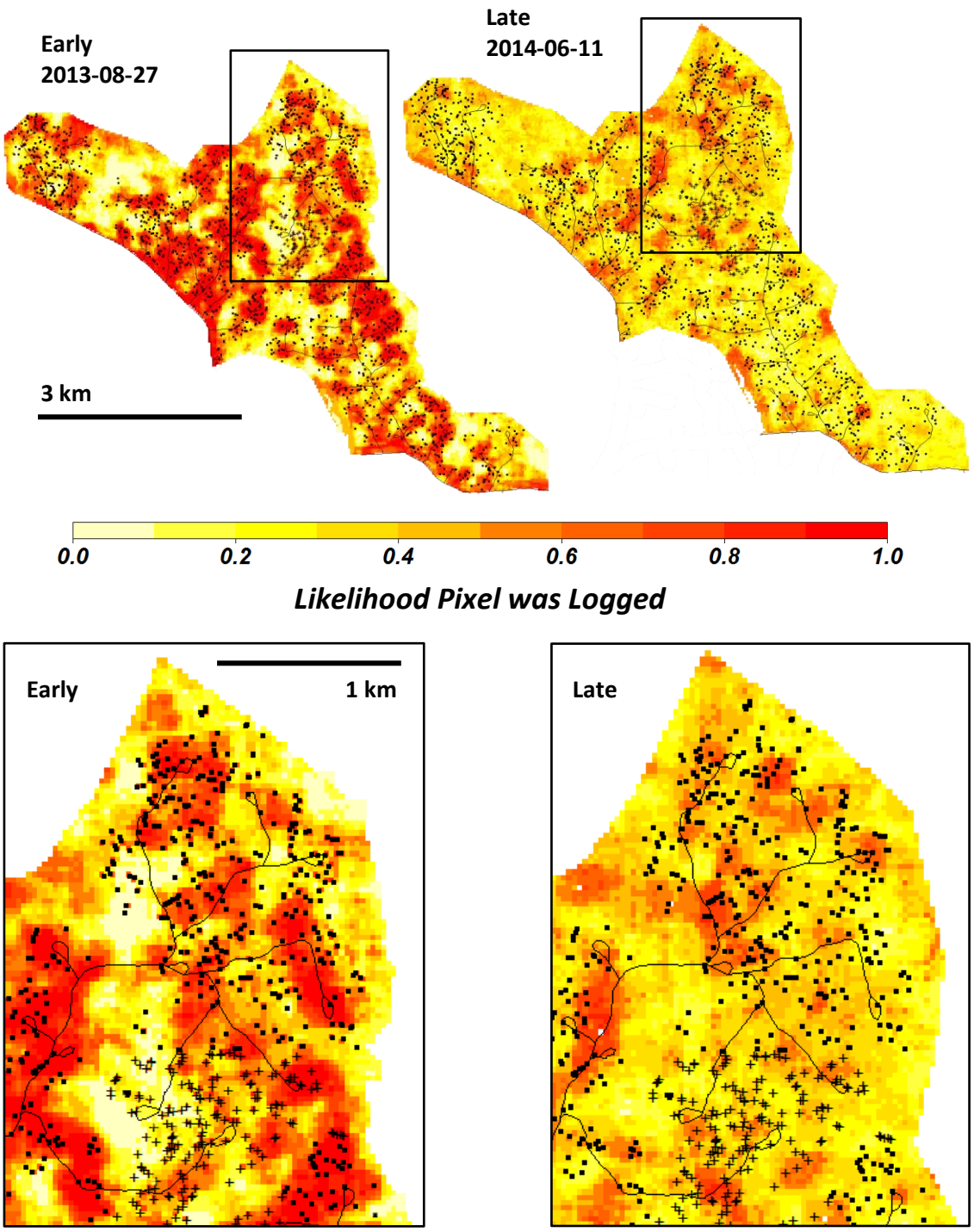


Figure 3.5 Example of a forest management unit in Jamari logged in 2013 showing the RF predicted likelihood that each pixel was logged (highest likelihoods in red) for the early and late data. Logging roads are thin black lines and tree removal locations are displayed as black squares and crosses. The black crosses (see insets for detail) coincide with the final 200 trees in the logging records for 2013.

squares. Many of these locations occur in low likelihood regions in the early data because these locations were probably unlogged at the time of the image acquisition (dates for specific tree removal were unavailable).

A further marked difference between the predictions is that, in general, far more pixels were labelled as logged in the early data than in the late, as can be seen by comparing the classifications in Figure 3.6, which shows the years between 2011 and 2015 for the early (top) and late (bottom) datasets, respectively. The FMU where logging occurred in each year is outlined in yellow and the 2015 image also shows the FMU to be logged in 2016 outlined in white. The early classifications appear to show some indication of a retained signal from the previously logged FMU (particularly 2012-08-16 and 2013-08-27 in Figure 3.6) that are less visible in the late classifications. In addition, the range of predicted logging likelihoods with late data was more variable from scene to scene, which resulted in some scenes having very few pixels of high likelihood of logging (see 2012-06-13 in Figure 3.6) and others with most of the study area predicted as logged (see 2016-06-16 in Figure 3.6). This suggests the threshold value from model calibration could not be used reliably for all late images and a scene-specific threshold value might need to be calculated for each image to provide better correspondence with logging activities.

The true proportion of logged pixels in each FMU (from the logging records) was roughly 12% in a given year (mean = 11.8%; standard deviation = 2.4%), but the early classifications consistently labelled a greater number of pixels as logged (Figure 3.7). For example, the proportion of pixels assigned in each FMU for early acquisitions was expected to be around 25% (10% truly logged and 15% false positives), but nearly twice as many were identified. However, forest disturbances from selective logging affect patches of forest and not just the pixels where trees were logged. Extra detections would be expected because of additional tree and canopy damage associated with tree removals, roads, and construction of skid trails. Note that the rate of false detections over unlogged FMUs (open diamonds in Figure 3.7) is roughly as expected for the early algorithm and most dates for the late algorithm, but is significantly different for the late algorithm for the FMU logged in 2015. The late scene for this FMU clearly shows anomalous behaviour and displays high likelihood of logging over most of the study area, including known unlogged regions (see Figure 3.6).

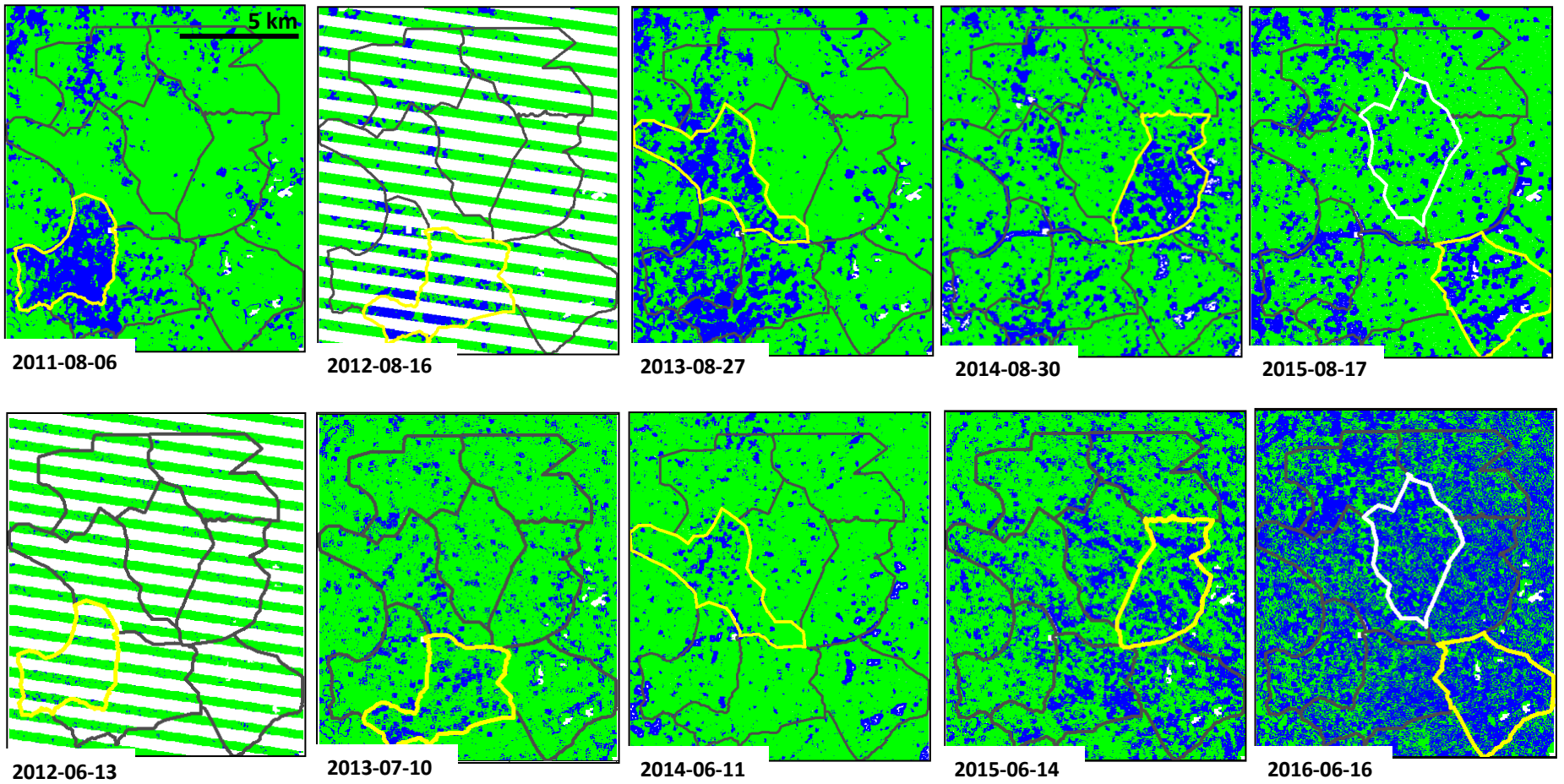


Figure 3.6 Classifications for Jamari between 2011 and 2016 with early (top) and late (bottom) Landsat data. The forest management units (FMUs) are outlined in black and the FMU logged in each year (where logging should be detected) is outlined in yellow. Blue and green represent classifications for logged and unlogged forest, respectively. White areas are no-date and correspond to the Landsat 7 scan-line corrector error (stripes) and pixels that were non-forest (irregular patches) in Brazil's Program to Calculate Deforestation in the Amazon (PRODES) database. The FMU logged in 2016 is outlined in white (far right) and the top two FMUs in each image remained unlogged.

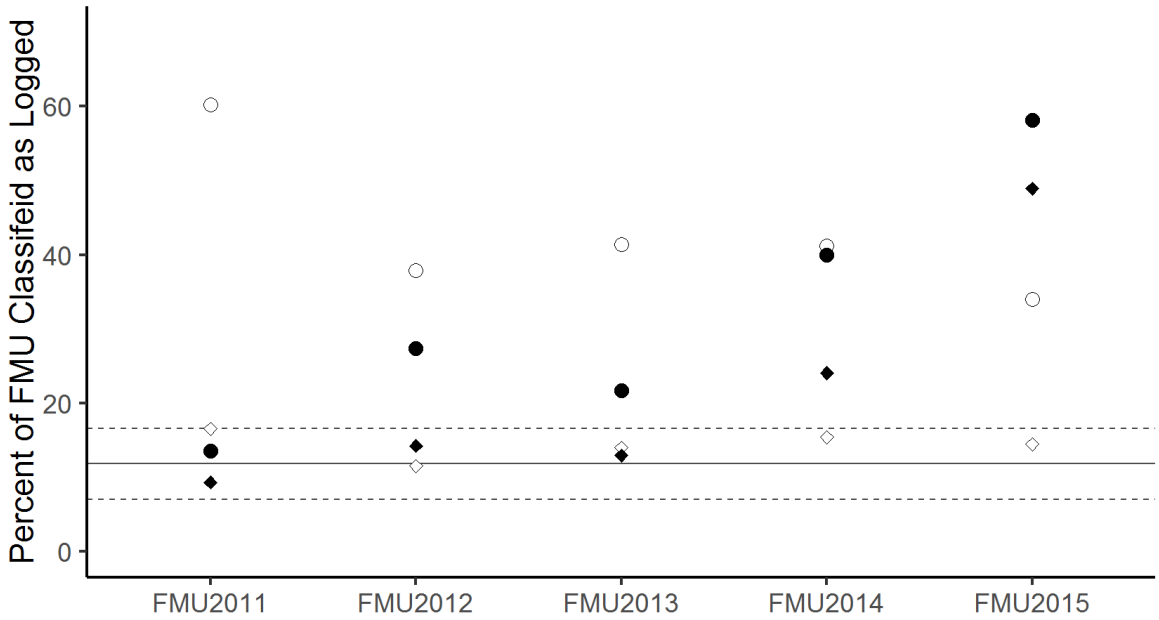


Figure 3.7 The proportion of pixels in each FMU that were classified as logged in Figure 3.6 for the early (open symbols) and late (closed symbols) algorithms. Circles are the logged FMUs in each year and diamonds are values from an FMU that remained unlogged. The black line represents the mean ± 1 standard deviation (dashed lines) of the true rate of logging across all FMUs. Values are unbiased (Olofsson et al., 2014) to account for possible sampling bias in the validation data.

We used the early algorithm to predict over the available Landsat time series in Jamari that coincided with logging in four FMUs (see Table A1.6 for image dates) and plotted the detections of logging through time (Figure 3.8). As expected, the proportion of detected pixels increased through the logging season during the year a given FMU was logged. There was also a drift upwards in the unlogged FMU, but the detections peaked just above the expected rate of 12% by late August (Figure 3.8). Importantly, known unlogged regions will not exhibit a $d_{p\mathrm{L}}$ of 0.85 (i.e. a false discovery rate of 15%), as any and all detections in known unlogged areas are wrong (i.e. a $d_{p\mathrm{L}}$ = 0). Consequently, the false alarm rate is the expected proportion of detections (i.e. P_{fd} = 11.5% in Table 3.2). This suggests that the algorithm performed as would be expected for tracking forest disturbances through time in both logged and unlogged FMUs. In particular, forest patches subjected to selective logging should display measurable increases in detections as the logging season progresses and known unlogged regions will exhibit the expected false alarm rate.

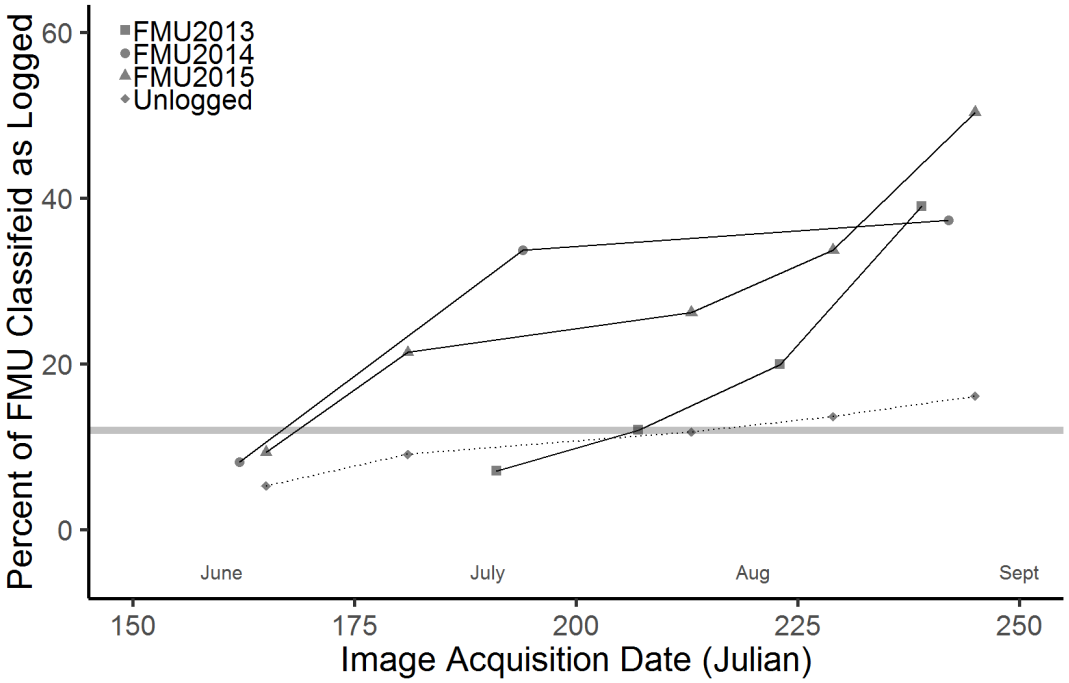


Figure 3.8. The proportion of pixels classified as logged through time in three logged and one unlogged FMU using the early RF model. Triangles, circles, and squares represent logged FMUs (solid lines) and diamonds are an unlogged FMU (dotted line). The grey horizontal line at 12% is the approximate detection rate expected for unlogged regions. Values are unbiased (Olofsson et al., 2014) to account for possible sampling bias in the validation data.

We assessed the impact of the window size used to calculate texture measures on the proportion of pixels labelled as logged FMUs for three logged and one unlogged FMU in the early data (Figure 3.9). Reducing the window size from 7x7 to 3x3 lowered the proportion labelled as logged by nearly 50% within each FMU, resulting in smaller clusters of pixels with high likelihoods (Figure 3.9). However, as noted above, forest disturbance from selective logging affects chunks of forest and not just the pixels where trees are cut. Thus, depending on the scale of interest, larger or smaller window sizes may be better for identifying patches of forest that have been selectively logged. In contrast, reducing the window size had little impact on the false detection rate over unlogged regions, remaining close to the 12% expected irrespective of window size (Figure 3.9). This suggests that the choice of window size is independent of the false positive rate over undisturbed forested areas and primarily affects likelihoods around pixels that the algorithm identifies as disturbed.

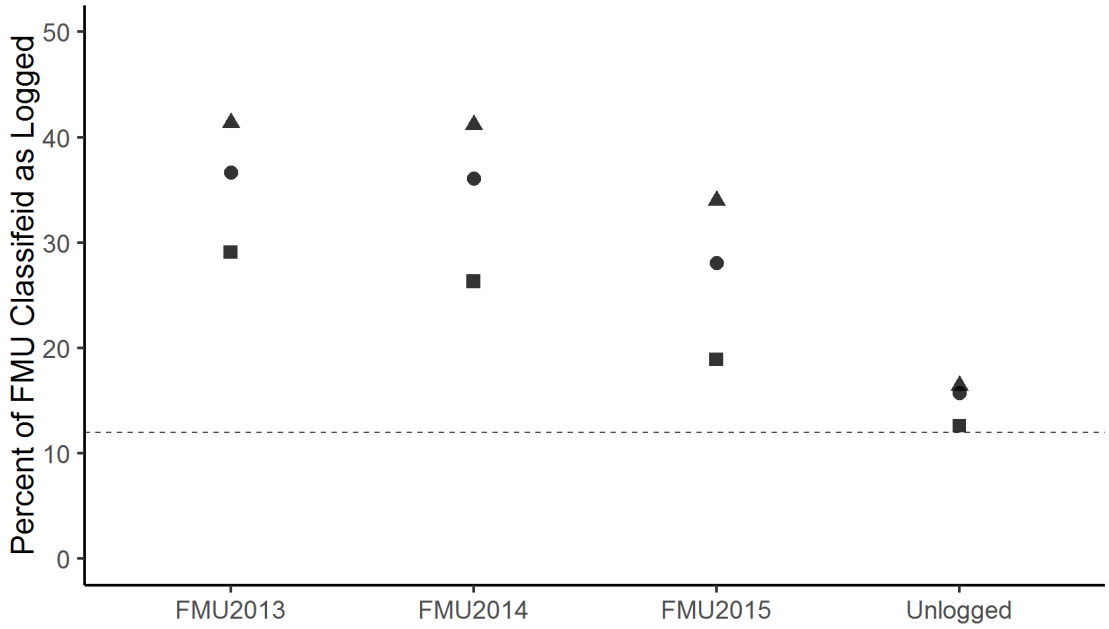


Figure 3.9 The proportion of pixels classified as logged in three logged FMUs and one unlogged FMU from RF models using texture measures with different window sizes. Triangles, circles, and squares represent windows used for texture calculation of 7x7, 5x5 and 3x3 pixels, respectively. The dashed line at 12% is the approximate detection rate expected for unlogged regions. Values are unbiased (Olofsson et al., 2014) to account for possible sampling bias in the validation data.

3.4.2 Random Forest predictions of logging at Jari

The majority of the best available (sufficiently cloud-free) Landsat scenes over Jari were from the ETM+ sensor, which suffered the scan-line corrector error, so approximately 22% of each image has missing data that appear as white stripes in Figures 3.10 and 3.11 (Storey et al. 2005). Nonetheless, this allowed us to see behaviour similar to Jamari, wherein predictions using early data clearly identified active logging (Figure 3.10) and predictions using late data detected very little logging (Figure 3.11). In particular, with late data most of the study area was classified as unlogged both before and after logging. Additionally, with early data the predictions of logged pixels in the year before logging were close to the expected rate of false positives over unlogged regions (approximately 12%). However, with late data the rate of false positives was not close to the expected rate over unlogged regions. Maps for the year before logging are displayed to demonstrate that the early dataset identified the correct year in which logging occurred and did not simply predict high amounts of logging for every year.

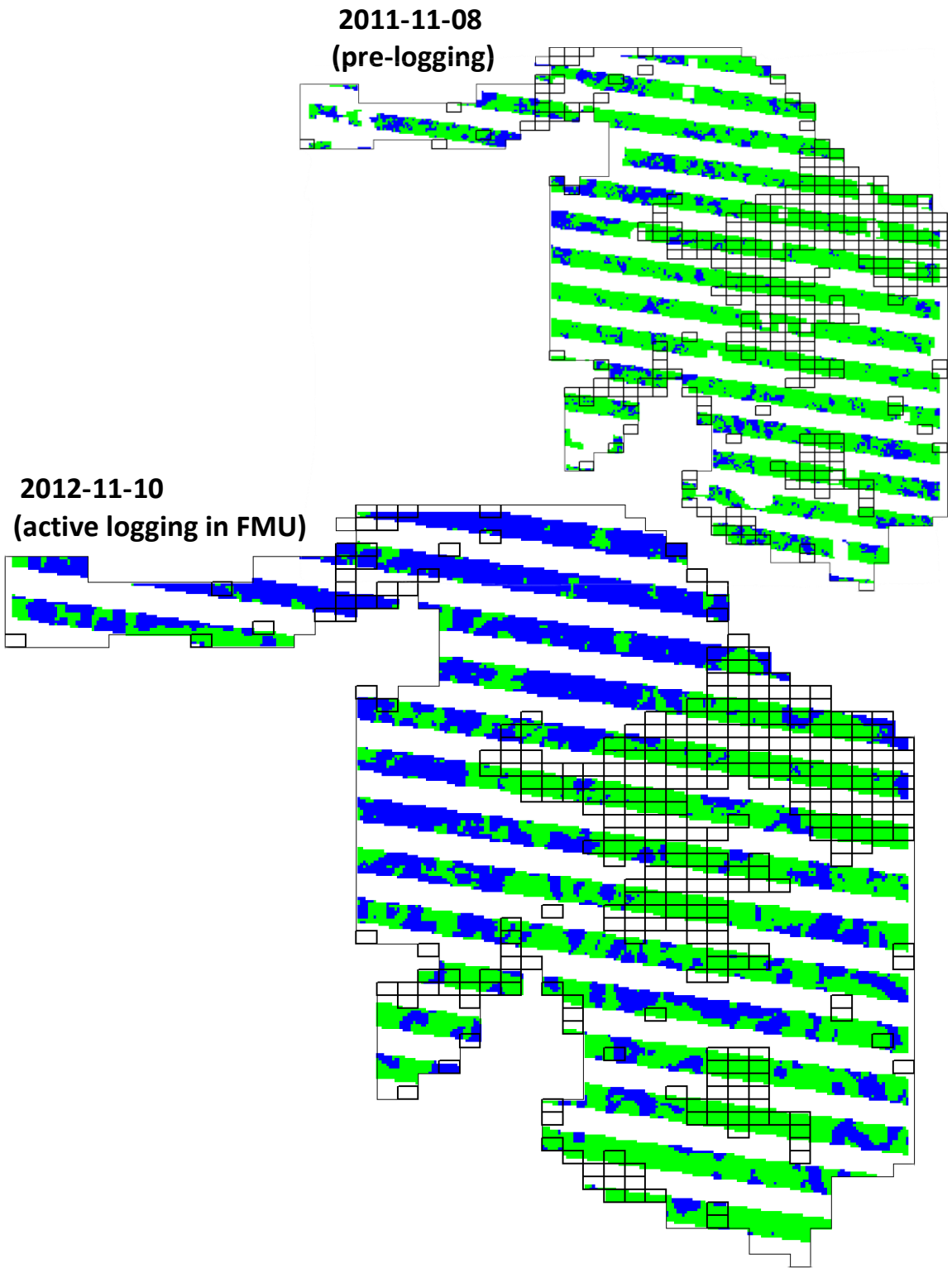


Figure 3.10 Logged (blue) and unlogged (green) predictions at the Jari study site using a Random Forest model trained from early Landsat inputs. Predictions from November 2011 (top) were before logging activities began and from November 2012 (bottom) while active logging was ongoing. Clouds were masked out and appear as irregular white patches (top). Missing data regions from the Landsat 7 scan-line corrector error appear as white stripes through the maps. Black boxes indicate the 10 ha blocks inside the Jari concession that were not logged.

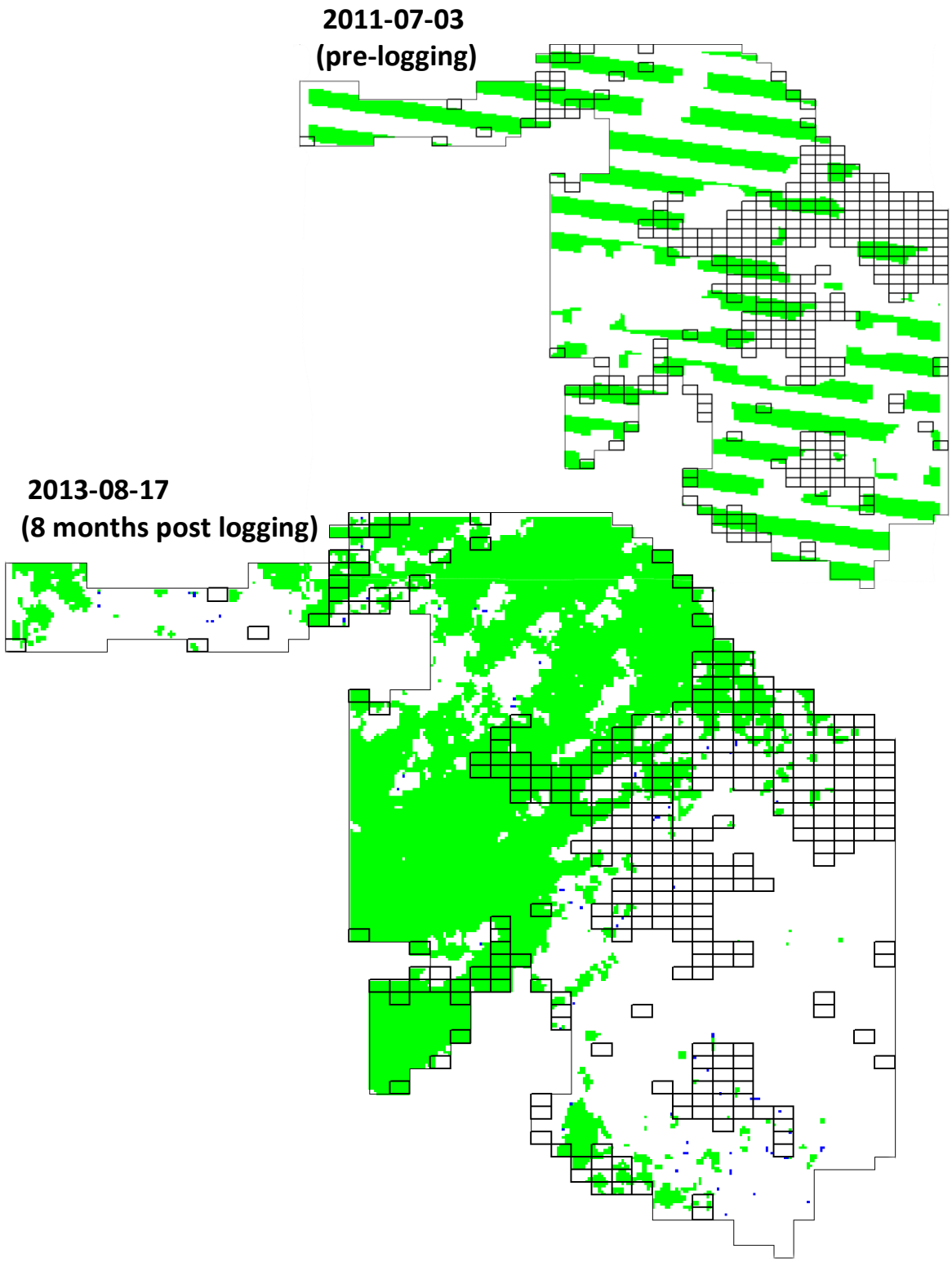


Figure 3.11. Logged (blue) and unlogged (green) predictions at the Jari study site using a Random Forest model trained from late Landsat inputs. Predictions from July 2011 (top) were before logging activities began and from August 2013 (bottom), approximately 8 months post-logging. Clouds were masked out and appear as irregular white patches. Missing data regions from the Landsat 7 scan-line corrector error appear as white stripes through the map (top). Black boxes indicate the 10 ha blocks inside the Jari concession that were not logged.

In total, an area of 6152 ha was visible in Jari after removing clouds and missing data gaps from the SLC error in the year of logging. Of this area, 1710 ha was not logged (black boxes in Figures 3.10 and 3.11). Since we lacked detailed logging records and only knew which 10 ha blocks were logged, a formal accuracy assessment of logging detections was not possible. However, when using the unbiased proportions and the threshold from Table 3.4 to classify predictions, the early algorithm labelled 2316 ha (38%) as logged (Figure 3.10). This value is consistent with predictions from Jamari where approximately 40% of logged FMUs were labelled with early data (see Figure 3.7). In addition, the rate of commission error when predicting logged pixels (i.e. 1 - d_{pL}) was 19.8%, which is also consistent with the rate of commission error between the validation data and prediction errors found for the Jamari site.

Table 3.3 Confusion matrix summarizing unbiased (Olofsson et al. 2014) results from Random Forest (RF) model classifications of logged and unlogged observations at Jari with Landsat data. The thresholds (T) developed at Jamari were used to classify predictions at Jari and were 0.40 and 0.65 for the early and late datasets, respectively (Table 3.2). The corresponding values for overall accuracy (OA), Cohen's kappa (κ), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

EARLY					**LATE**				
OA: 89.0%					OA: 92.2%				
κ: 0.77		Reference Class			κ: 0.05		Reference Class		
d_{pL}: 0.80		Logged	Unlogged	Commission	d_{pL}: 0.80		Logged	Unlogged	Commission
P_d: 0.93				Error (%)	P_d: 0.03				Error (%)
Predicted	Logged	0.351	0.085	19.5	Predicted	Logged	0.002	0.005	19.9
Class	Unlogged	0.025	0.538	4.4	Class	Unlogged	0.078	0.919	7.8
Omission Error (%)		6.7	13.7		Omission Error (%)		97.3	0.06	

3.5 Discussion

The spatial resolution of Landsat data has previously been considered too coarse to monitor selective logging activities (Asner et al. 2002), with most applications involving logging intensities >20 m^3 ha^{-1} at sites with an abundance of spectrally distinct features (Souza and Barreto 2000; Souza et al. 2005; Asner et al. 2005). However, we have demonstrated that Landsat surface reflectance data can be used effectively, in a supervised machine learning framework, to detect subtle spectral changes from selective logging at low intensities. Although a definitive estimate of the amount of logging activities that have previously gone undetected is difficult to determine, a dataset of 824 logging permits from the state of Pará, Brazil found 18% of permits authorized for logging were harvested at intensities < 20 m^3 ha^{-1} (Richardson and Peres 2016). Thus, our approach has the potential to significantly increase current abilities to detect and monitor selective logging activities that up to now have been, at best, marginally detectable (see Appendix A1, Section A1.5 for a comparison between our method and CLASlite, Asner et al., 2009a). In addition, the approach outlined here has the distinct advantage of being able to make predictions about forests on a single scene to map disturbances, instead of requiring successive cloud-free images like many approaches (Asner et al. 2009a). This is particularly important since a single, low-cloud scene may be all that is available for a given region (see Souza, Jr et al. 2013).

Only the algorithm developed with data close to the time of active logging (i.e. the early data) performed well at detecting selective logging. Many logged pixels were omitted when using data from the first cloud-free image of the next dry season (i.e. late). In addition, only the algorithm trained with imagery close in time to the logging events was transferable to new areas (Figures 3.10 and 3.11). Thus, our results suggest images acquired during, or very soon after, active logging are needed to map low intensity selective logging. This is partly because logging activities typically occur in the dry season when cloud-free imagery is more likely to be available, but also because the spectral changes associated with low-intensity selective logging practices are subtle and short-lived and rapidly become obscured under even limited regrowth (Broadbent et al. 2006).

The decision to fix the proportion of logging detections that were correct (i.e. limiting the commission error when predicting logged pixels) defined the classification threshold applied to the likelihoods produced by the RF models developed at Jamari. This threshold would likely give different values of d_{pL} in regions that contain different proportions of logged and unlogged observations (see equation 3.4). Indeed, the threshold value from model training produced a slightly higher d_{pL} when assessed against the validation dataset, yet these data were from the same study site. In addition, depending on the distribution of likelihoods produced by the RF models, different datasets might yield different threshold values, for example because of higher

selective logging intensities. However, assuming both classes are present, the proportion of detected pixels that are wrong (i.e. 1 - d_{pL}) would be expected to remain invariant. Hence if the same threshold were applied over the whole of the Amazon basin, we would expect approximately 20% of all detections to be wrong and 11.5% of truly intact forest pixels to be identified as logged. This could be used to refine the algorithm (in the absence of field data on logging locations) by examining the rate of false detections over known unlogged regions or protected areas to achieve a similar error rate. Adopting this threshold (i.e. P_{fd} = 11.5) would make the method equivalent to a Constant False Alarm Rate detector which is widely used in detection problems with rare events (Scharf, 1991). A d_{pL} of 85% was the value chosen here as a compromise that gives a high detection rate (0.92 for early data, see Table 3.2) while keeping the proportion of detections that are false to an acceptable level. However, other values of d_{pL} could be chosen, depending on the predictive objectives of the particular application. This is precisely why Figure 2.4 shows the full range of threshold values; to enable a detailed assessment of model performance with higher or lower values of T or d_{pL}.

An important issue when assessing detections of selective logging is that *patches* of forest are affected, not just the isolated pixels where trees are removed. The area around logged pixels is certain to be disturbed because of canopy damage associated with tree removals and the construction of roads and skid trails, but the precise amount is unknown. Consequently, taking as a reference purely the pixels where trees were known to be removed is inadequate for assessing the disturbance due to logging. Indeed, the true rate of logged pixels at Jamari was approximately 12% (mean = 11.8%; standard deviation = 2.4%), but this represents a minimum expected detection rate and the associated forest disturbances would result in more detections. The early algorithm labelled approximately 40% of the area inside FMUs in Jamari and Jari as logged. This may be a more realistic estimate and is likely close to the upper limit of what constitutes forest disturbance for this level of logging. However, because the choice of window size for texture measure calculation affected the proportion of pixels labelled as logged (Figure 3.9), the appropriate window size for a particular application needs to be considered. Smaller windows resulted in fewer detections, but use of too small a window risks being unable to adequately measure texture arising from forest disturbances from selective logging. Thus, the specific application would best dictate the optimum approach and the user should, if possible, use window sizes matched to the expected or known spatial spread of forest disturbance around tree removals.

Selective logging rates in the Brazilian Legal Amazon (BLA) are thought to have remained relatively stable since 2000, with Pará and Mato Grosso enduring the highest rates of selective logging (Souza et al. 2013; Betts et al. 2017). However, our findings suggest that their assessments of forest disturbance and the associated carbon emissions are likely

underestimated. Machine learning approaches (neural networks, decision trees, support vector machines, etc.) for classification of satellite imagery have been used with increasing frequency and success since their initial applications to remote sensing questions in the 1990's (Tuia et al. 2011), but their effectiveness relies heavily on adequate training data. Our results suggest that detailed logging records ought to be a reporting requirement for logging companies or for REDD+ projects related to logging. These datasets could be used for building, improving, and updating models similar to the one presented here, with the aim of facilitating the creation of pan-tropical estimates of (legal and illegal) selective logging activities.

From a conservation perspective, the ability to identify regions of forest that are selectively logged is useful for mapping primary forest, but also for delineating logged forests with conservation value. Forests subjected to selective logging generally maintain far higher levels of biodiversity than other modified habitats, such as plantations or secondary forests (Gibson et al. 2011; Edwards *et al.*, 2014). Moreover, even after accounting for the amount of wood removed, reduced impact logging activities (like those at our study site in Jamari) do better at maintaining biodiversity than conventional selective logging practices (Bicknell et al. 2014) while simultaneously sequestering more carbon during regrowth (Putz et al. 2008; Martin et al. 2015). Thus, in the context of REDD+ or alternative conservation initiatives, forests affected by low intensity selective logging offer high biodiversity value and carbon sequestration potential. Accordingly, our method could be used for identifying and prioritizing forest tracts suitable for such initiatives.

3.5.1 Study limitations

While the minimum mapping unit remained 30 m, the use of texture measures resulted in some spatial aggregation of logging predictions (see Figures 3.5 and 3.6). This was expected around logged pixels, as a result of canopy gaps, skid trails, and roads, but clustered detections were also present in unlogged FMUs (see Figure 3.6). Ideally, predictions of logging in unlogged FMUs would have shown a diffuse 12-15% of spurious detections. Attempts to refine the accuracy of a final predictive map, by performing a post-processing step in which either likelihoods or classified pixels are re-examined (e.g. using a window analysis to apply neighbourhood rules whereby likelihoods or counts of nearby pixels are re-evaluated against some criteria) to enhance the detection rate or limit the false detection rate further, would prove difficult (Huang et al. 2014). However, using a smaller window size for texture calculation, such as 5x5 pixels, would reduce this effect. Ultimately, the optimal window size for textures depends on the objectives of the application and understanding how different window sizes affect detection and false detection rates.

Landsat surface reflectance data is known to exhibit occasional strong scene-to-scene and within-scene variations because of discontinuities across focal plane modules (Morfitt et al. 2015) and seasonal changes in solar viewing angles (Roy et al. 2016), respectively. We did not take these effects into account and likely affected algorithm performance in some instances (e.g. 2016-06-16 in Figure 3.6). Thus, a large scale application of the approach outlined here should include a step to normalize surface reflectance data across scenes to facilitate detection of the subtle and short lived spectral changes associated with low-intensity selective logging practices (Broadbent et al. 2006).

Our analysis used a binary classification (logged and unlogged forest) yet tropical forest landscapes are a heterogeneous mixture of land uses (e.g. secondary forests, burned areas inside forests, agricultural fields). We avoided some of these complexities by using the PRODES forest designations to remove urban areas, agricultural fields, and deforested areas that had regenerated to secondary forest. However, our method cannot distinguish between disturbance types and is best suited for tracts of remaining forest that contain logging concessions. In addition, selective logging represents a range of forest disturbance intensities and we would have preferred to use the logging dataset in a regression framework (i.e. a continuous response, such as logging intensity). However, the range of logging intensities within our Jamari dataset was very limited, since it was such a low intensity concession. Consequently, a regression approach was not suitable for the Jamari dataset and we chose to use classification. Additional datasets could fold into the framework here and might facilitate a continuous response approach as those datasets become available.

Finally, our analyses used freely available optical datasets. However, the problems associated with using optical imagery in the tropics, including the limited availability of cloud-free images over many regions and the rapid regeneration of tropical forest vegetation, remain major obstacles to pan-tropical assessments of tropical selective logging rates. Methods that integrate optical and radar dataset into a single algorithm would likely further improve the detection of tropical selective logging activities (Joshi et al. 2016; Higginbottom et al. 2018; Reiche et al. 2018).

3.6 Conclusion

Loss and degradation of forests in the tropics has important implications for global climate change, local populations and biodiversity (Lewis et al. 2015). Methods to reliably map forest disturbances from selective logging would be a key contribution to quantifying the terrestrial portion of the carbon budget and the role of land-use change in tropical forests emissions (Baccini et al. 2017). In addition, reliable forest monitoring systems are actively sought after by tropical nations and conservation groups tasked with mitigating global climate change through

improved forest management practices (GOFC-GOLD, 2016). Our results should stimulate further assessments of regional rates of low-intensity selective logging in tropical forests.

Our analysis, based on training Random Forest models with detailed records of tree removals, has demonstrated that Landsat data can be effective at detecting selective logging at much lower intensities than has previously been reported. To be successful, the input satellite data needs to be acquired within a few months of the logging, as the subtle signal caused by logging (and the more extensive disturbance associated with logging) is rapidly lost. Although we had less complete knowledge of logging activities at the Jari site, the algorithm developed at Jamari appeared to transfer successfully to this site (despite being 1500 km away). Hence there is reason to expect that it could be applied at much wider scales.

Acknowledgements

We would like to thank *AMATA Brazil* for providing access to logging records for Jamari and *Jari Florestal* for logistical support. MGH was funded by the Grantham Centre for Sustainable Futures. JMBC was funded as part of NERC's support of the National Centre for Earth Observation. FMF was CNPQ and NERC-funded (NE/P004512/1; PELD-RAS 441659/2016-0, respectively) and was funded by CAPES (BEX5528/13-5) and CNPq-PELD site 23 (403811/2012-0) during the long-term monitoring in Jari. We thank three anonymous reviewers for their thoughtful comments and suggestions that greatly improved the manuscript.

Chapter

4

Detecting tropical selective logging with SAR data requires a time series approach

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, and S Quegan

Under review in *Remote Sensing of Environment*

Abstract

Selective logging is the primary driver of forest degradation in the tropics and reduces the capacity of forests to harbour biodiversity, maintain key ecosystem processes, sequester carbon, and support human livelihoods. While the preceding decade has seen a tremendous improvement in the ability to monitor forest disturbances from space, advances in forest monitoring have almost universally relied on optical satellite data from the Landsat program, whose effectiveness is limited in tropical regions with frequent cloud cover. Synthetic aperture radar (SAR) data can penetrate clouds and have been utilized in forest mapping applications since the early 1990s, but no study has exclusively used SAR data to map tropical selective logging. A detailed selective logging dataset from three lowland tropical forest regions in the Brazilian Amazon was used to assess the effectiveness of SAR data from Sentinel-1, RADARSAT-2 and PALSAR-2 for monitoring tropical selective logging. We built Random Forest models in an effort to classify pixel-based differences in logged and unlogged areas. In addition, we used the BFAST algorithm to assess if a dense time series of Sentinel-1 imagery displayed recognizable shifts in pixel values after selective logging. Random Forest classification with SAR data (Sentinel-1, RADARSAT-2, and ALOS-2 PALSAR-2) performed poorly, having high commission and omission errors for logged observations. This suggests little to no difference in pixel-based metrics between logged and unlogged areas for these sensors. In contrast, the Sentinel-1 time series analyses indicated that areas under higher intensity selective logging (> 20 m^3 ha^{-1}) show a distinct spike in the number of pixels that included a breakpoint during the logging season. BFAST detected breakpoints in 50% of logged pixels and exhibited a false alarm rate of approximately 10% in unlogged forest. Overall our results suggest that SAR data can be used in time series analyses to detect tropical selective logging at high intensity logging locations within the Amazon (> 20 m^3 ha^{-1}). These results have important implications for current and future abilities to detect selective logging with freely available SAR data from SAOCOM 1A, the planned continuation missions of Sentinel-1 (C and D), ALOS PALSAR-1 archives (expected to be opened for free access in 2020), and the upcoming launch of NISAR.

4.1 Introduction

Selective logging is the primary driver of forest degradation in the tropics (Curtis et al. 2018; Hosonuma et al. 2012). Logging reduces the capacity of forests to harbour biodiversity, maintain key ecosystem processes, sequester carbon, and support human livelihoods (Baccini et al. 2017; Barlow et al. 2016; Lewis, Edwards, and Galbraith 2015). However, large uncertainties remain in assessing the true impact of selective logging because the technological advances in detecting and monitoring logging at large scales are only just emerging (Hethcoat et al. 2019). The ability to reliably map forest degradation from selective logging is a key element in understanding the terrestrial portion of the carbon budget and the role of land-use in turning tropical forests into net carbon emitters (Baccini et al. 2017). In addition, reliable forest monitoring systems are urgently needed for tropical nations and conservation groups seeking to report and/or mitigate carbon emissions through improved forest stewardship (GOFC-GOLD, 2016).

While the preceding decade has seen a tremendous improvement in the ability to detect forest disturbances from space (Hansen et al. 2013; Hethcoat et al. 2019; Tyukavina et al. 2017), advances in forest monitoring have almost universally relied on optical satellite data from the Landsat program. Yet, the effectiveness of optical data is limited in tropical regions with frequent cloud cover like the northwest Amazon and central Africa. Synthetic aperture radar (SAR) data can penetrate clouds and have been utilized in forest mapping applications since the early 1990s (reviewed in Koch, 2010). However, the SAR data archives are spatially and temporally fragmented, and in many cases the data products required commercial licences for their use. Consequently, uptake by users has been more limited than optical data and the full potential of SAR has likely been under-utilized (Reiche et al. 2016).

SAR backscatter, particularly at L- and P-band, is sensitive to changes in carbon stocks in forests with biomass < 300 Mg ha^{-1} (Koch 2010; Mitchard et al. 2009; Saatchi et al. 2011). This enables accurate differentiation between forested and non-forested areas and has been well studied (e.g. Shimada et al., 2014). More recently, polarimetric and interferometric methods have been developed that utilize information in the SAR signal to detect forest changes (Deutscher et al. 2013; Flores-Anderson et al. 2019; Lei et al. 2018; Mathieu et al. 2013). Yet, the limited temporal and spatial coverage of SAR data have hampered widespread application and use of these techniques to monitor forest disturbances (e.g. single-pass interferometric SAR is only available with TanDEM-X data). Moreover, advancements in monitoring selective logging with SAR data are generally lacking, despite widespread recognition of both the need and the role it could play (Mitchell, Rosenqvist, and Mora 2017; Reiche et al. 2016).

The launch of Sentinel-1A in mid-2014 represented the first continuous global acquisition strategy for open SAR data. Since that time two studies have exclusively used

Sentinel-1 to map deforestation (Antropov et al. 2016; Delgado-Aguilar et al. 2017), with others utilizing a fusion of optical and SAR data (Joshi et al. 2016; Reiche et al. 2015; Reiche, Hamunyela, et al. 2018; Reiche, Verhoeven, et al. 2018). While methods that fuse optical and Sentinel-1 have been successful, their continued dependence on optical imagery nevertheless limits their utility in regions with frequent cloud cover. With the successful launch of SAOCOM 1A in late 2018, the planned continuation of the Sentinel-1 missions (with C and D), and the anticipated launches of SAOCOM 1B in 2019 and NISAR in 2021, vast amounts of free C- and L-band SAR data will soon be available. Accordingly, methods are needed that utilize SAR data for large-scale forest monitoring, yet no study has used Sentinel-1 for detection of selective logging activities.

The primary objective of this paper was to assess the ability of Sentinel-1 to detect tropical selective logging. Detailed spatial and temporal logging records from three regions in Brazil were used to develop and test the effectiveness of two different detection techniques: (1) exploiting pixel-based differences between logged and unlogged locations in single images and (2) detecting change in a time series of pixels known to be logged.

Pixel-based methods for detecting changes in remotely sensed imagery often utilize differences between pixel values or other mathematically derived metrics in time or space, for example before and after some disturbance or in areas known to be disturbed and undisturbed within the same image (reviewed in Hussain et al., 2013). These differences can be used for classification, employed in machine learning, or analyzed temporally to map change. Recently, the detection of selectively logged regions in single images has been demonstrated successfully with optical data from Landsat (Hethcoat et al. 2019). Accordingly, we sought to evaluate whether similar methods could be transferred to SAR data. The selective logging records were used to build supervised machine learning models to detect selective logging. Machine learning methods have many applications in remote sensing and have been used with increasing frequency and success (Lary et al., 2018). We performed equivalent analyses with SAR data from the C-band RADARSAT-2 and L-band PALSAR-2 sensors to compare the performance of longer wavelength (i.e. L-band PALSAR-2) and higher resolution data (both RADARSAT-2 and PALSAR-2 have higher sensor resolution).

In addition, we used all the available Sentinel-1 archives in a time series analysis to monitor pixel values for breakpoints in the time series of locations that had been selectively logged. Time series methods have increasingly been used for monitoring changes in pixel values, in part because of the availability of vast archives of imagery on cloud computing platforms like Google Earth Engine (Gorelick et al. 2017), but also because of the recognition that seasonal or longer term trends in pixel values can be less susceptible to erroneously characterizing change (Bullock, Woodcock, and Olofsson 2018; Verbesselt, Zeileis, and Herold

2012; Zhu 2017). Given that forest disturbances from selective logging are often subtle and short-lived, detecting changes with SAR data over large regions will present technological and algorithmic challenges. However, a critical assessment of detection capabilities and a careful understanding of the performance of these data types is essential for advancing forest monitoring techniques in the tropics.

4.2 Study area and data

4.2.1 Study area and selective logging data

Selective logging data from three lowland tropical forest regions in the Brazilian Amazon were used in this study (Figure 4.1). The Jacunda and Jamari regions are inside the Jacundá and Jamari National Forests, Rondônia, while the Saraca region is inside the Saracá-Taquera National Forest, Pará. Forest inventory data from 14 forest management units (FMUs) selectively logged between 2012 and 2017 were used, comprising over 32,000 individual tree locations. Unlogged data from three additional locations, one inside each study region, comprised over 11,500 randomly selected point locations known to have remained unlogged during the study period (Table 4.1).

4.2.2 Satellite data and pre-processing

All available C-band Sentinel-1A Ground Range Detected scenes in descending orbit and Interferometric Wide mode (VV and VH) were utilized in Google Earth Engine (GEE) over the study regions through November 2018. These had incidence angles of 38.7°, and 38.7°, and 31.4° for Jacunda, Jamari and Saraca, respectively. GEE is a cloud computing platform hosting calibrated, ortho-corrected Sentinel-1 scenes that have been processed in the following steps using the Sentinel-1 Toolbox: (1) thermal noise removal; (2) radiometric calibration; and (3) terrain correction using the Shuttle Radar Topography Mission (SRTM) 30 m digital elevation model (DEM). The resulting images had a pixel size of 10 m.

Single Look Complex C-band RADARSAT-2 scenes in Fine mode (HH and HV) were obtained from the Canadian Space Agency. Twelve ascending scenes, with an incidence angle of 30.7°, coincided with selective logging records and were acquired between 2011 and 2012. Pre-processing of images was done with the Sentinel-1 Toolbox and included: (1) radiometric calibration; (2) multi-looking (by a factor of 2 in azimuth) to produce square pixels; and (3) terrain correction using the SRTM 30 m DEM. The resulting images had a pixel size of 10 m.

Level 2.1 L-band PALSAR-2 scenes (HH and HV) were obtained from the Japan Aerospace Exploration Agency (JAXA) with a pixel size of 6.25 m. Four geometrically corrected scenes coincided with selective logging records and were acquired between 2016 and 2017

with incidence angles of 28.5° in ascending orbit. Image digital number was converted to normalized backscatter using the calibration factors provided by JAXA.

4.2.3 Speckle filtering

SAR data are inherently speckled from interference between scattering objects on the ground (Woodhouse, 2017) and often require reduction of speckle prior to analyses. Many speckle-reduction methods involve spatial averaging, but the associated loss of spatial resolution was likely to hinder the detection of the subtle signal from selective logging activities. Thus, following the SAR pre-processing steps detailed above for each data type, the final step involved multi-temporal filtering to reduce speckle (Quegan and Yu 2001). Multi-temporal filtering reduces speckle by averaging a pixel's speckle through time (as opposed to a spatial average). A 7x7 pixel window was used. The equivalent number of looks after speckle filtering for Sentinel-1, RADARSAT-2 and PALSAR-2 was approximately 15, 5 and 5, respectively.

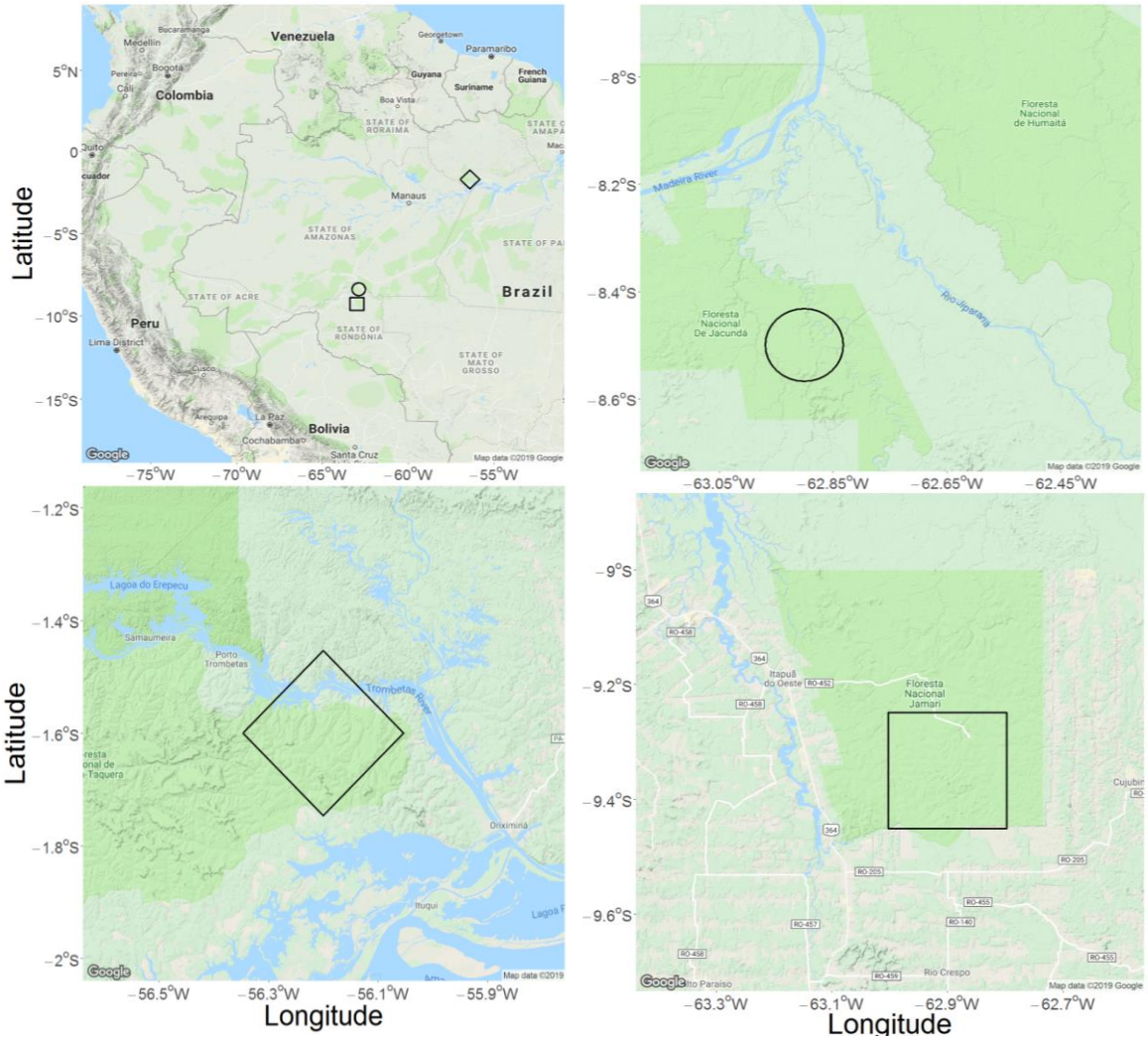


Figure 4.1 Location of the Jacunda (circle), Jamari (square), and Saraca (diamond) study regions in the Brazilian Amazon.

Table 4.1 Data used in the classification of selective logging from three study regions in the Brazilian Amazon. The forest management unit (FMU), logging intensity, sample size (pixels), and overlap with satellite data coverage are shown for Sentinel-1 (S), RADARSAT-2 (R), and PALSAR-2 (P).

FMU	Logging Intensity (m^3 ha^{-1})	N	Coverage
Jacunda_I_2016	6	2,290	S
Jacunda_I_2017	9	2,822	S
Jacunda_II_2015	15	2,613	S
Jacunda_II_2016	10	1,815	S
Jacunda_II_2017	7	1,310	S, R*
Jacunda_Reserve	0	3,000	S*, R*
Jamari_I_2015	22	1,094	S, R*
Jamari_I_2016	10	653	S, R*
Jamari_I_2017	12	911	S, R*
Jamari_III_2012	10	3,071	R
Jamari_III_2015	11	3,042	S, R*
Jamari_III_2016	9	2,058	S, R*, P
Jamari_III_2017	11	2,597	S, R*, P
Jamari_Reserve	0	5,912	S*, R*, P*
Saraca_Ia_2017	12	3,769	S
Saraca_II_2016	25	3,223	S
Saraca_II_2017	21	4,729	S
Saraca_Reserve	0	3,000	S*

* FMU was unlogged at time of acquisition and data represent unlogged observations

4.3 Methods

4.3.1 Supervised classification with Random Forest

4.3.1.1 Data inputs for classifying selective logging

For each satellite data type (Sentinel-1, RADARSAT-2, and PALSAR-2) data were extracted at each pixel where logging occurred and randomly selected pixels in nearby regions that remained unlogged. Thus, the data inputs for logged and unlogged observations came from a single scene for each study region (i.e. a space-for-time study design in contrast to images before and after logging from the same location). Selective logging at the study areas only occurred during the dry season, approximately June-October in a given year, and data were extracted from images acquired as late into the logging period as possible (Table A2.1) to

ensure the majority of pixels had been subjected to logging, but also before the onset of the rainy season (Hethcoat et al. 2019). In addition, logging activities tend to be accompanied by surrounding disturbances (canopy gaps, skid trails, patios, and logging roads) resulting in forest disturbances beyond just the pixels where a tree was removed. Accordingly seven texture measures were calculated for each polarization (sum average, sum variance, homogeneity, contrast, dissimilarity, entropy, and second moment) to provide a local context for each pixel (Haralick et al., 1973). These were calculated within a 7x7 pixel window, chosen as a trade-off between minimizing window size while still capturing the variability in selectively logged forests compared to unlogged forests. Finally, a composite band was calculated as the ratio of the co- polarized channel to the cross-polarized channel (i.e. HH/HV or VV/VH). Each dataset thus comprised a 17-element vector (2 polarization bands, their ratio composite band, and 7 texture measures for each polarization) for each pixel where logging occurred and randomly selected pixels that remained unlogged.

4.3.1.2 Random Forests for classification of selective logging

We built Random Forest (RF) models using the *randomForest* package in program R version 3.5.1 (Liaw and Wiener, 2002; R Development Core Team, 2018). The RF algorithm (Breiman 2001a) is an ensemble learning method for classification. Each dataset was split into 75% for training and 25% was withheld for validation. In order to further ensure the independence of training and validation datasets, the validation data were spatially filtered such that no observations in the training dataset were within 90 m of an observation in the validation dataset. RF models have two tuning parameters: the number of classification trees grown (k), and the number of predictor variables used to split a node into two sub-nodes (m). We used a cross-validation technique to identify the number of trees and the number of variables to use at each node that minimized the out-of-bag error rate on each training dataset (Table A2.2). Prior to model training, feature selection was performed with the Boruta package (version 6.0.0) in Program R (version 3.3.1) and all variables were deemed significant (Figures A2.1-3).The importance of each predictor variable was assessed during model training, using Mean Decrease in Accuracy, defined as the decrease in classification accuracy associated with not utilizing that particular input variable for classification (Breiman 2001b).

4.3.1.3 Model validation: assessing accuracy

RF models were validated using a random subset of the full dataset for each sensor (described in Section 4.3.1.2). By default, RF models assign an observation to the class indicated by the majority of decision trees (Breiman, 2001a). However, the proportion of trees that voted for a particular class from the total set of trees can be obtained for each observation and a

classification threshold can be applied to this proportion (Hethcoat et al. 2019; Liaw and Wiener 2002). We adopted such an approach, wherein the proportion of trees that predicted each observation to be logged, informally termed the *likelihood* a pixel was logged, was used to select the classification threshold. A threshold, T, was defined such that if *likelihood* > T the pixel was classified as logged (Figure 4.2).

The confusion matrix then has the form:

		Reference	
		L	UL
Predicted	L	D_L	D_{UL}
	UL	$N_L - D_L$	$N_{UL} - D_{UL}$

where L and UL refer to logged and unlogged classes, N_L and N_{UL} are the numbers of logged and unlogged observations in the reference dataset, and D_L and D_{UL} are the numbers of logged and unlogged pixels detected as logged, respectively. We defined the *detection rate* $DR = D_L/N_L$ and *false alarm rate* $FAR = D_{UL}/N_{UL}$ as the frequency that a logged or unlogged pixel was classified as logged, respectively. Thus, the DR is equivalent to 1 minus the omission error of the logged class and the FAR is the omission error of the unlogged class. In addition, we defined the *false discovery rate* (FDR):

$$\mathrm{FDR} = \frac{D_{UL}}{D_L + D_{UL}} = 1 - \frac{1}{1 + \left(\frac{N_{UL}}{N_L}\right)\left(\frac{\mathrm{FAR}}{\mathrm{DR}}\right)}. \quad (4.1)$$

The FDR is the proportion of all observations that were detected as logged that were actually unlogged, and is equivalent to the commission error of the logged class. The FDR is an assessment of the rate of prediction error (i.e. type I) when labelling pixels as logged and can be used in detection problems with rare events or unbalanced datasets, such as selectively logged pixels within the Amazon Basin (Benjamini and Hochberg 1995; Hethcoat et al. 2019; Neuvial and Roquain 2012). A high DR and low FDR is clearly desirable, but these cannot be fixed independently in two-class detection problems and both depend on the threshold value (Figure 4.2). For example, if achieving a 95% detection rate led to a FDR of 50%, then half of all predictions of logging would be incorrect. This level of performance would make estimates of selective logging extremely uncertain. The value of the classification threshold (T) therefore represents a trade-off between true and false detections. In practice, a viable detection method would expect to achieve a DR > 50% while limiting the FDR to 10-20% to have any value for widespread forest monitoring. The performance of each sensor was assessed by plotting the DR, FAR and FDR values as T varied from 0 to 1 to facilitate discussion of model performance.

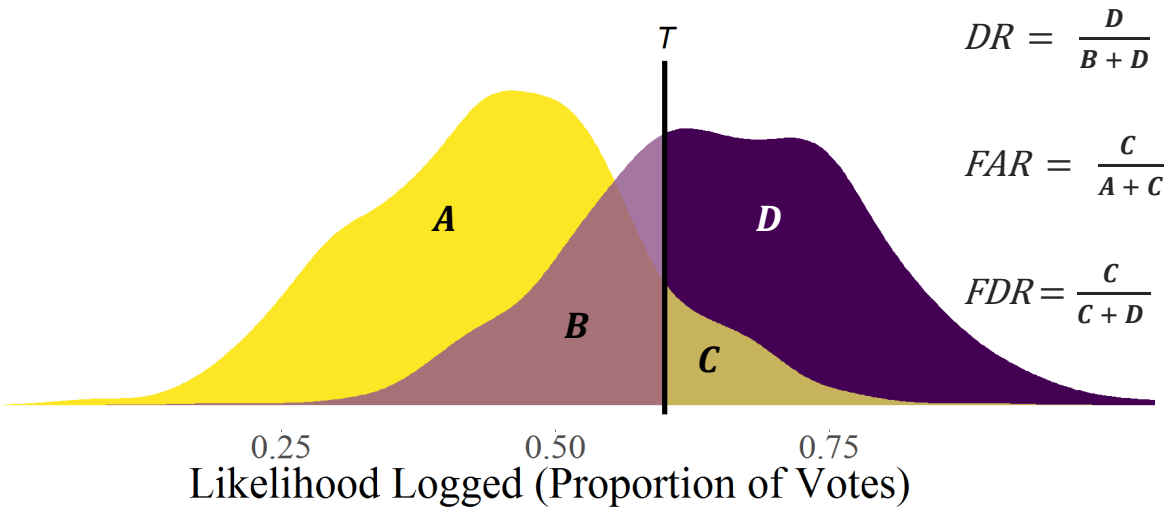


Figure 4.2 Diagram representing the trade-off between the detection rate (DR) and the false alarm rate (FAR) associated with using a threshold *T* (vertical black line) to label pixels as logged and unlogged based upon the proportion of votes that each observation was predicted to be logged. The purple and yellow colors correspond to density plots for hypothetical logged and unlogged observations, respectively. Thus, the areas A and B are the portions of the observations from unlogged and logged pixels, respectively, that will be labelled as unlogged. Similarly, C and D represent the portions of the observations from logged and unlogged pixels, respectively, that will be labelled as logged.

4.3.1.4 Sentinel-1 classification of high intensity logging

Most of the selective logging data in this study were low-intensity (<15 m^3 ha^{-1}) and we anticipated the logging signal to be weak and difficult to detect. Consequently, we also considered a reduced Sentinel-1 dataset that included only those FMUs with logging intensities above 20 m^3 ha^{-1} (n = 3 sites) and the unlogged data (n = 3 sites) to assess if Sentinel-1 could be used for detecting selective logging activities near the legal limit within the Brazilian Legal Amazon. Unfortunately RADARSAT-2 and PALSAR-2 imagery did not cover the highest intensity logging sites, so we could not perform equivalent analyses with these datasets. RF classification and validation was performed on this subset of the Sentinel-1 data in the manner detailed above for the full dataset.

4.3.2 Time series analyses

We tested whether a time series of Sentinel-1 data displayed discernible changes in pixel values after selective logging with the Breaks for Additive Seasonal and Trend (BFAST) algorithm (Verbesselt et al. 2010, 2012) in program R (R Core Team, 2018). BFAST estimates the timing of abrupt changes within a pixel-wise time series (breakpoint hereafter) and has been successfully utilized with a range of data types (e.g. Landsat, MODIS, SAR, etc.) to map phenology, deforestation, water inundation, and more. BFAST was used to assess if a suitable model with

one or no breakpoints was appropriate (i.e. the *bfast01* command) and included tests for coefficient and residual-based changes in the expected value (i.e. the conditional mean). BFAST works in 4 steps; first a linear model is fit to the pixel-wise time series (the model is parameterized by the data internally), next a model with 1 breakpoint is estimated (one that minimizes the segmented residual sum of squares), then a sequence of tests are performed (BIC, supLM, supF, OLS-MOSUM) assuming a null hypothesis of zero breaks, finally each test result is aggregated and a single test decision can be used or the combined assessment for any test statistic (Verbesselt et al. 2010; Zeileis 2005). BFAST can test for multiple breakpoints within a pixel's time series, however, given that our dataset only spanned a one-year time period we considered a single disturbance model to be most appropriate.

The metrics used in searching for breakpoints in the full Sentinel-1 time series (approximately 55 scenes from October 2016 – August 2018) were the two most important predictor variables identified from RF models. The limited temporal coverage of RADARSAT-2 and PALSAR-2 at our study sites precluded time series analyses with these datasets. Where breakpoints were identified, we determined if they coincided with the timing of selective logging activities (June – October) and regarded these as true detections. Breakpoints in unlogged areas and breakpoints outside the timing of logging activities were considered false detections. In addition, the relationship between the frequency of breakpoints within an FMU and its logging intensity was examined to understand potential thresholds in logging intensity above which variables could be used to monitor selective logging activities through time series analyses.

Finally, we examined if the relationship between logging intensity and the rate of detections and false alarms was consistent between logging locations (i.e. a scattered subset of pixels in an area) and an entire region (i.e. all pixels within a bounding box). The timing of breakpoints was mapped for two 500 m X 500 m test regions within the Saraca study area (one logged and one unlogged). A limited number of small test regions were chosen because of the computationally expensive nature of the pull request in Earth Engine (e.g. two 1 km regions query > 1 million records for export). Only breakpoints during the time period associated with logging were mapped (June – October).

4.4 Results

4.4.1 Random Forest classification of selective logging

The single-image detection results for all sensors revealed that in order to get false discovery rate (FDR) values sufficiently low (e.g. 10-20%), the corresponding detection rates (DR) of selective logging were of almost no value (< 5%) for reliably forest monitoring. In general, the following results suggest that regions that have experienced selective logging do not show

consistent differences from unlogged areas in the metrics we used for classification. The second analysis (section 4.4.2) therefore deals with detection of selective logging with time series data and provides better results.

4.4.1.1 Sentinel-1

Random Forest detection performance for Sentinel-1 is shown in Figure 4.3 (top). Both the detection and false alarm rates were close to 1 until the threshold exceeds ~0.4, meaning almost every pixel in an image would be detected as logged. This suggests difficulty distinguishing logged and unlogged observations, and many unlogged observations were being misclassified as logged (Figure A2.4). In general, the detection, false alarm, and false discovery rates (across the range of threshold values) were insufficient for reliable classification of selective logging with Sentinel-1 data at the intensities within our study areas (6-25 m^3 ha^{-1}). For example, even if a FDR of 30% were acceptable, this would yield a detection rate < 20%, which would be of little practical value. Thus, attempts to strongly limit the false discovery rate (commission error of logged observations) would require a high threshold value and result in very few detections. Overall, this suggests that using single images from Sentinel-1on their own to detect and map selective logging activities would be fraught with error with the classification approach used here.

4.4.1.2 RADARSAT-2

Random Forest performance for RADARSAT-2 is shown in Figure 4.3 (middle). Both the false alarm rate and the detection rate rapidly declined as the threshold value was initially increased, again suggesting difficulty in distinguishing logged and unlogged observations. In contrast to Sentinel-1, RADARSAT-2 was less likely to label an observation as logged and very few observations had likelihood values above 0.5 (Figure A2.5). It should be noted that the logging records that coincided with RADASAT-2 data were from a single FMU that was relatively low intensity (10 m^3 ha^{-1}). Consequently, the performance displayed here may not be a full appraisal of RADARSAT-2 capabilities. Given how poorly the model performed, however, it is uncertain that a vast improvement would occur with better training datasets. Overall, our results suggest that RADARSAT-2 data cannot be used to effectively monitor low-intensity selective logging activities using pixel-based differences between logged and unlogged areas. However, additional tests with data at higher logging intensities should be pursued.

4.4.1.3 PALSAR-2

Random Forest classification performance for PALSAR-2 is shown in Figure 4.3 (bottom). In general, the performance of PALSAR-2 was equally poor at distinguishing logged and unlogged

observations as RADARSAT-2 and Sentinel-1 (Figure A2.6). The final rise in the false discovery rate in Figure 4.3, before it drops to zero, is the result of calculating proportions from very small sample sizes (e.g. 5 of 10 observations predicted logged were actually unlogged). Similar to RADARSAT-2, the selective logging data that coincided with PALSAR-2 imagery was from two relatively low-intensity FMUs (9 - 11 m^3 ha^{-1}). Again, however, more data at higher logging intensities seems unlikely to improve classification performance to the desired level. For example when the data from Sentinel-1 was restricted to just the low intensity sites used in the PALSAR-2 analyses, there was effectively no change in the rates of detection and false discovery compared to the results from all logging intensities with Sentinel-1 (Figure A2.7 and Table A2.7). Thus, the lack of higher intensity logging data probably had little impact on the results for PALSAR-2. In general, this suggests that the limitations in distinguishing logged and unlogged pixels are inherent in the data and metrics we used for classification (for all three data sets).

4.4.1.4 Sentinel-1 classification of high intensity logging

Detection performance of Sentinel-1 data for the highest intensity FMUs is shown in Figure 4.4. Despite limiting the detection task to the most intensively logged FMUs (as well as unlogged observations), the detection rate and false discovery rate values were comparable to the results that used the full range of logging intensities. Instead, improvement in model performance was associated with better discrimination of unlogged observations (i.e. compare the commission and omission errors for the unlogged class between Tables 4.2 and 4.5). Essentially, the model was able to better identify unlogged forest, presumably because the more "confusing" observations (i.e. the low intensity FMUs) were absent and could not muddle the distinction between logged and unlogged observations (Figures A2.8). Overall, our results suggest Sentinel-1 data cannot be used in the classification of pixel-based differences to monitor selective logging activities with reasonable precision, even at the most intensively logged regions within the Amazon.

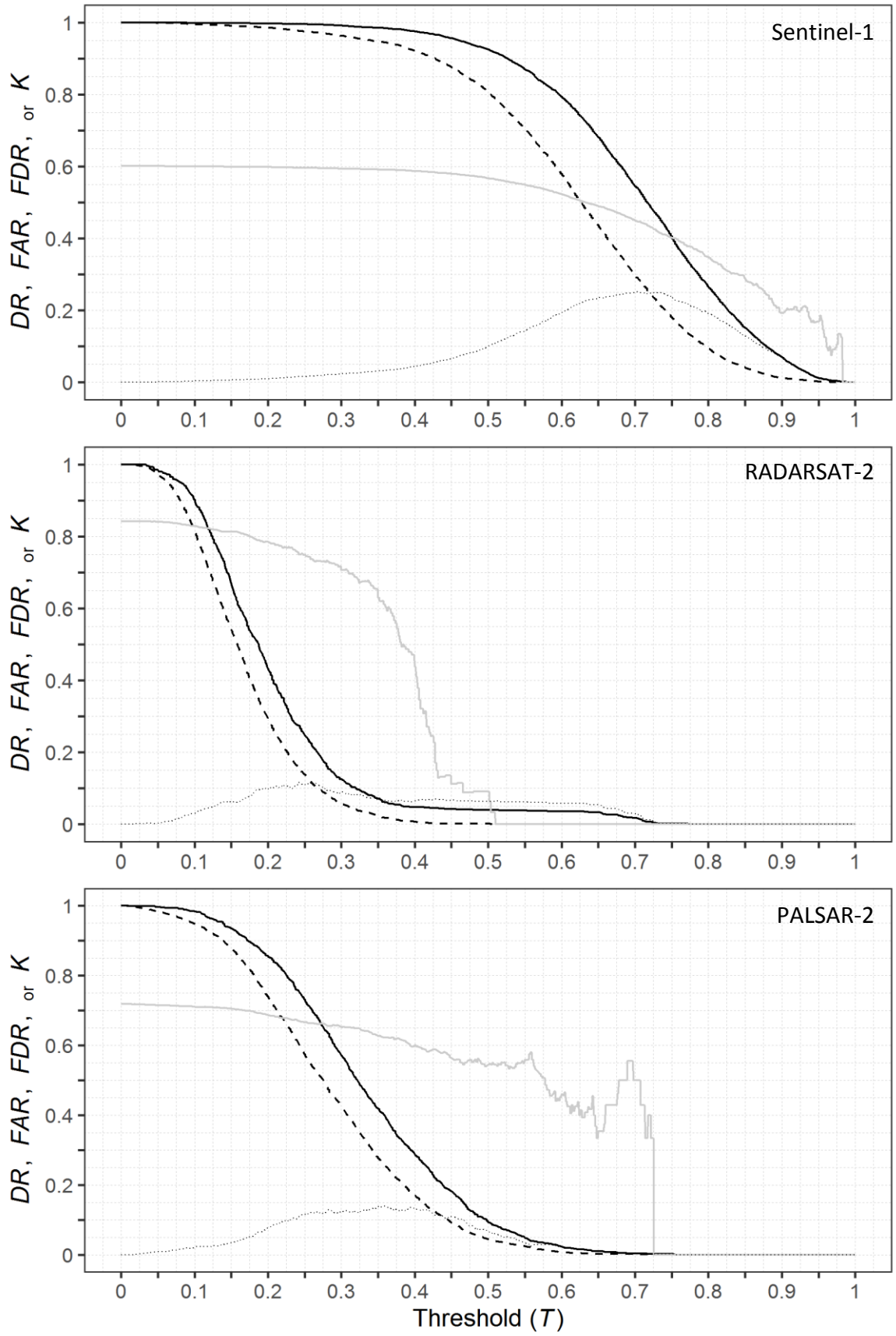


Figure 4.3 Random Forest model performance across the range of threshold values (*T*) for classification with SAR data. The Detection Rate (DR) and False Alarm Rate (FAR) are the solid and dashed black lines, respectively. Also shown are the corresponding values of the False Discovery Rate (FDR) and Cohen's kappa (*k*) as solid and dotted grey lines, respectively.

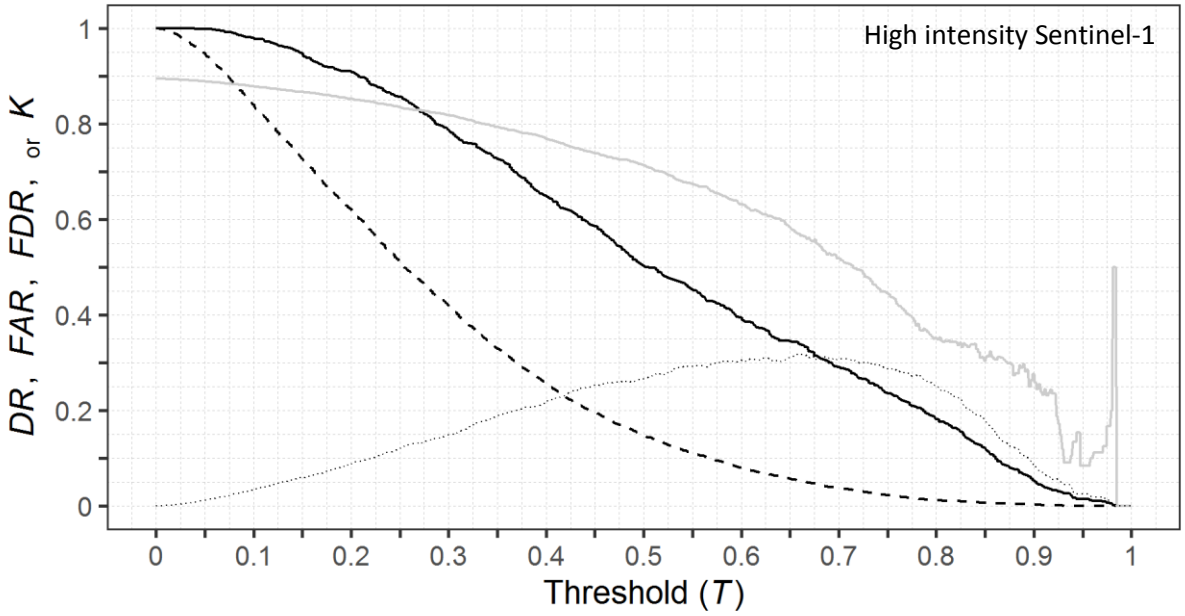


Figure 4.4 Random Forest model performance across the range of threshold values (*T*) for classification of Sentinel-1 data with a subset of the most intensively logged sites. The Detection Rate (DR) and False Alarm Rate (FAR) are the solid and dashed black lines, respectively. Also shown are the corresponding values of the False Discovery Rate (FDR) and Cohen's kappa (solid and dashed grey lines, respectively).

4.2 Sentinel-1 time series analyses

The two most important predictor variables from the Sentinel-1 RF model were the Sum Average metric (Haralick 1973) on the VV and VH bands (Figure A2.9, Equation A2.1). A plot of VV sum average values through time for six randomly selected tree harvest locations at the Saraca site is shown in Figure 4.5 and suggests selective logging decreased the value of this metric. In addition, histograms of the timings associated with all breakpoints at three FMUs are shown in Figure 4.6 and indicates the time frame of the breakpoints mainly occurred within the logging season for those FMUs logged above 20 m^3 ha^{-1}. In contrast, the time periods associated with breakpoints at lower logging intensities were shifted toward the onset of the rainy season in late 2017 – early 2018, however, all FMUs showed an uptick in breakpoints associated with the rainy season (Figure 4.6). This suggests that Sentinel-1 time series data could be used to detect and monitor selective logging activities from areas that have experienced logging close to the legal limit in Brazil (30 m^3 ha^{-1}), particularly if the detection time-frame is narrowed to within the known logging season.

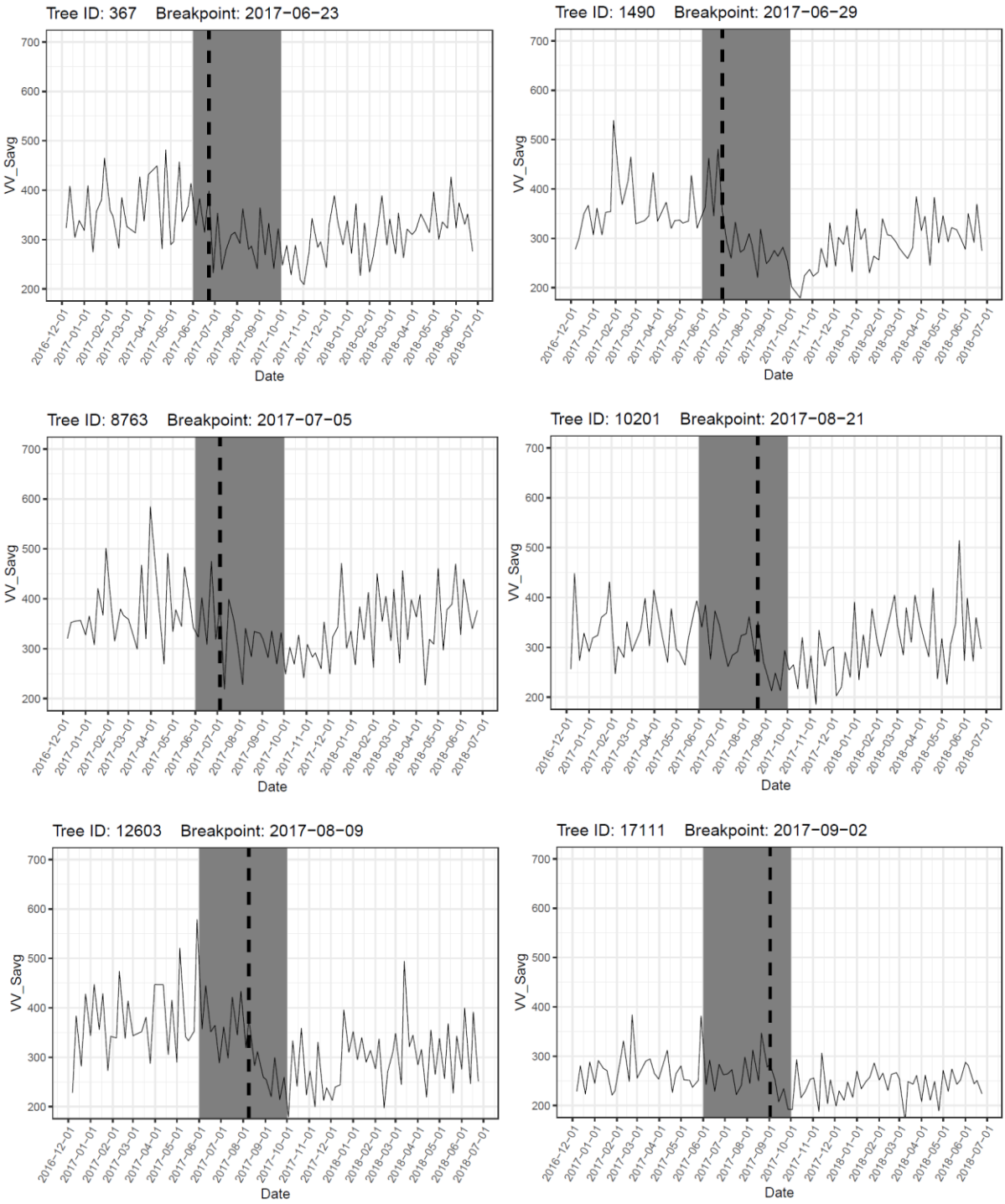


Figure 4.5 Breakpoint dates identified by the BFAST algorithm from six randomly selected points within the Saraca study region. The time series of the VV sum average texture measure is plotted in black, the selective logging period is shaded in grey, and the identified breakpoint date is labelled with a vertical dashed line.

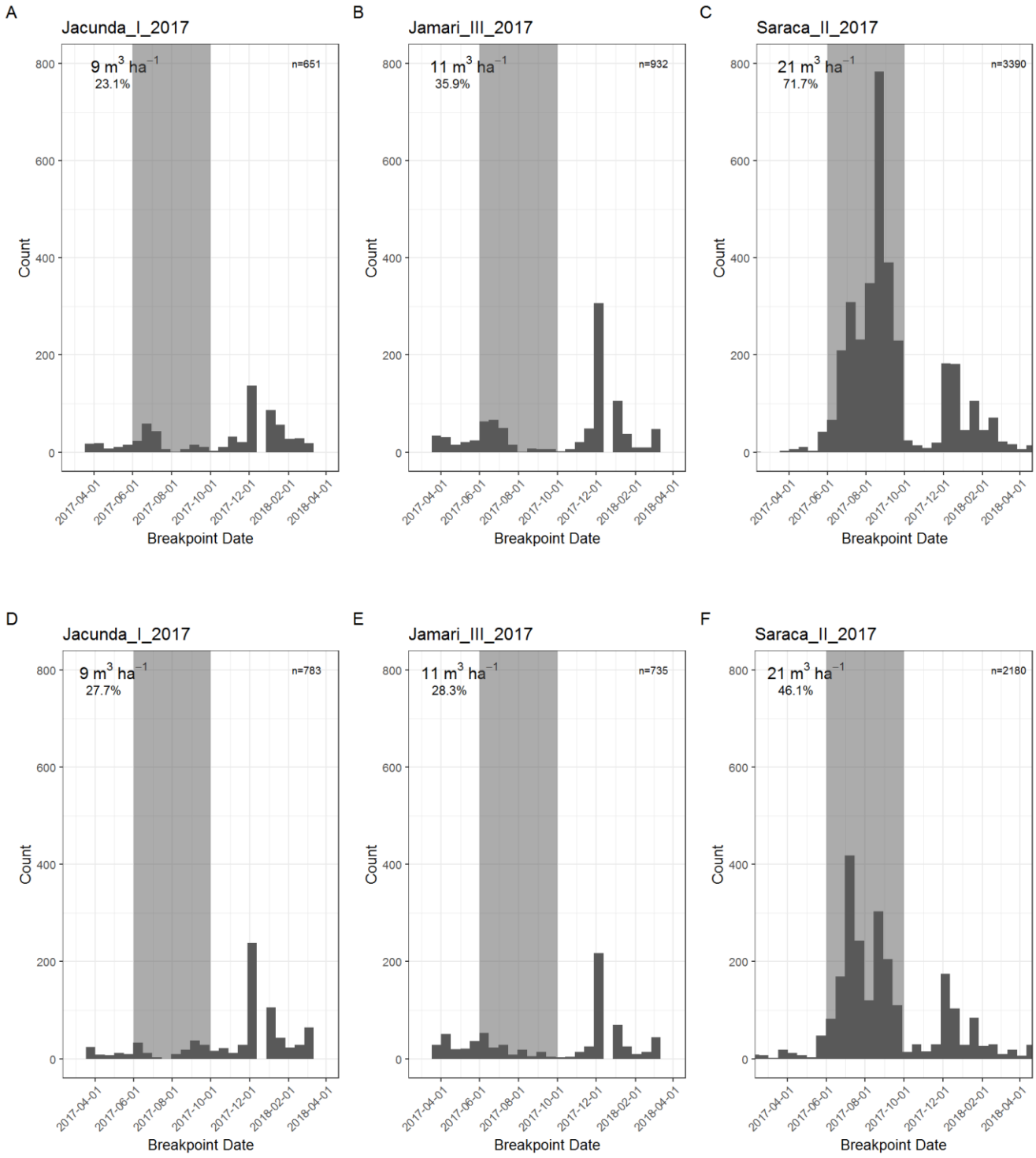


Figure 4.6 Histograms of breakpoint dates associated with time series analyses of the Sentinel-1 sum average texture measure for three study regions in the Brazilian Amazon for the VV (top row) and VH (bottom row) bands. The logging intensity and the proportion of observations with breakpoints in the data are in the upper left of each panel. The time period coinciding with logging activities is shaded in grey.

When the value of the VV sum average metric was monitored through time in pixels known to be logged and unlogged, the proportion of pixels with a significant breakpoint in their time series increased as the logging intensity of the FMU increased (Figure 4.7A). Approximately 70% of logged pixels in high logging intensity FMUs had a breakpoint, however, nearly 25% of unlogged pixels showed a breakpoint in their time series (i.e. 25% false alarm rate). This false alarm rate was generally consistent through logging intensities approaching 15 m^3 ha^{-1} and suggests no signal in pixels logged at low to moderate intensities (Figure 4.7A). When the breakpoints were assessed only over the time period associated with logging (to remove the false peak associated with the rainy season), the relationship showed a similar pattern whereby the FMUs logged at the highest intensities showed a large rise in breakpoints above a background false alarm rate that was relatively constant up through moderate logging intensities (Figure 4.7B). At the highest intensities, the detection rate was > 50% and the false alarm rate was approximately 10%. These results further support the idea that FMUs logged at low to moderate intensities do not show a distinct time series signal whereas FMUs logged at higher intensities do. Overall, this suggests that FMUs logged at intensities closer to the legal limit within the Brazilian Legal Amazon (30 m^3 ha^{-1}) should show a noticeable spike in the number of breakpoints within its time series above a background false alarm rate and could be used to detect logging activities in the dry season.

Approximately 55% and 20% of pixels in the logged and unlogged test regions had a breakpoint during the logging season (Figure 4.8A and B). These values are generally in agreement with our prior results from the subset of pixels where trees were removed (see Figure 4.7B). While 55% of the pixels in the logged test region did not have a tree removed, selective logging is associated with forest disturbances that go beyond the individually logged pixels (e.g. canopy gaps, skid trails, logging roads, etc.) and additional detections are expected. Only about 5% of the pixels in the logged test region were actually logged, however, it is clear from the Planet imagery (Figure 4.8C and D; Planet Team 2017) that more than 5% of the forest patch was disturbed by logging activities. Given the false alarm rate was around 20%, the difference between detections and false alarms might represent a value comparable with the amount of forest disturbance expected at this intensity (i.e. about 30%).

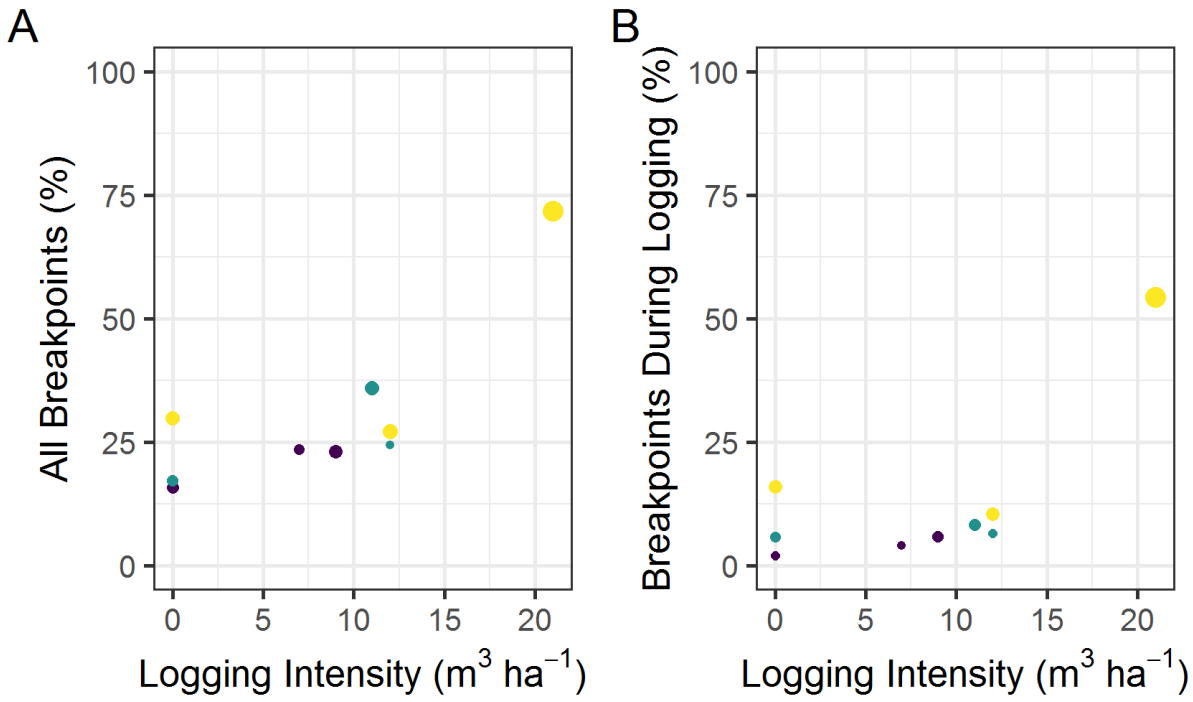


Figure 4.7 The relationship between the proportion of observation within a Forest Management Unit (FMU) that had a breakpoint identified within its Sentinel-1 VV sum average texture measure time series and the logging intensity of the FMU. The proportion of all observations (A) and the proportion that had a breakpoint that coincided with the logging season (B) are shown separately. The circle size corresponds to number of observations at each FMU and yellow, green, and purple colors represent the Saraca, Jamari, and Jacunda sites, respectively. See the supplementary material for the same analyses with the second and third best metric from Random Forest (Figure A2.10).

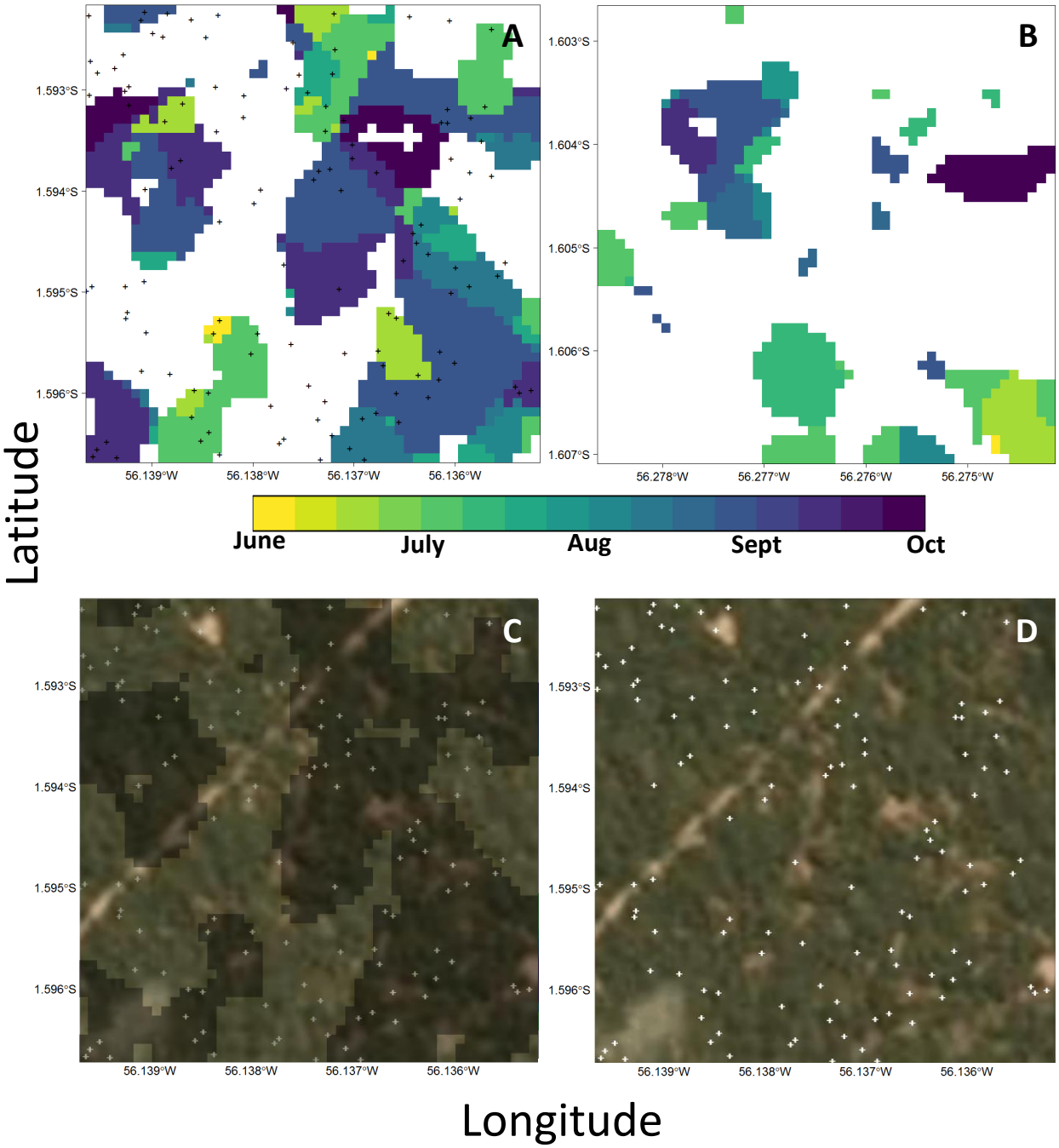


Figure 4.8 Map of predicted breakpoint dates for two 500m X 500m test regions, one logged (A) and one unlogged (B), in the Saraca National Forest, Para, Brazil. Logged tree locations are black crosses and the date of the breakpoint for each pixel is color coded by week, with white representing no breakpoint. Planet imagery (3 m) from 28 August 2017 overlaid with and without breakpoint locations (C and D) for the logged area (trees in white). Approximately 54% and 21% of the pixels in the logged and unlogged regions had breakpoints, respectively.

4.5 Discussion

We present the first multi-sensor comparison of SAR data for monitoring a range of selective logging intensities in the tropics. We demonstrated that L-band PALSAR-2, C-band RADARSAT-2, and C-band Sentinel-1 data performed inadequately at detecting tropical selective logging when using pixel-based attributes for classification. However, when analysing a time series of Seninel-1 texture measures, logged pixels displayed a strong tendency for a breakpoint in their time series as the logging intensity of the FMU increased. Moreover, the timing associated with the identified breakpoint generally coincided with active logging at the highest logging intensities. Overall, our results suggest that Sentinel-1 data could be used to monitor the most intensive selective logging, but a time series approach would be required to detect change. A number of studies have used Sentinel-1 time series data to monitor deforestation (Bouvet et al. 2018; Reiche, Hamunyela, et al. 2018; Reiche, Verhoeven, et al. 2018), often in combination with optical data, however our study is the first to show it has the potential to be used exclusively to monitor selective logging.

4.5.1 Variable importance

In a number of cases the most important predictor variables from RF models involved the co-polarized channel (Figure A2.9), despite the generally accepted view that the cross polarized channel is best for detecting changes in forest cover (Joshi et al. 2016; Reiche, Hamunyela, et al. 2018; Ryan et al. 2012; Shimada et al. 2014). The HH polarization of PALSAR-2 data has previously been shown to be sensitive to the early stages of deforestation, resulting from single-bounce scattering from felled trees (Watanabe et al. 2018). Our results support the idea that the co-polarized channel (for L- and C- band SAR) is useful and should not be ignored in forest disturbance detection analyses (e.g. Reiche et al., 2018a). While shorter wavelength SAR data, like C- and X-band, are known to be less sensitive to forest structure, because the radar signal mainly interacts with the forest canopy (Woodhouse, 2017; Flores-Anderson et al., 2019), the higher backscatter values in the co-polarized channel for all three sensors suggests predominantly rough surface backscattering from the forest canopy (as volume scattering generally results in roughly equal backscatter between co- and cross-polarized channels). This suggests that forest tracts subjected to more intensive selective logging than we studied (conventional logging permits with larger canopy gaps, large road networks, and many log landing areas) should possess a signal in the co-polarized channel that could be used to detect changes in canopy cover and should not be discarded (e.g. Reiche et al., 2018a).

Random Forest models offer an objective approach to selecting important variables for use in time series analyses. The Mean Decrease in Accuracy rankings were used to select the sum average texture measure in the time series results, corroborate their rankings (see Figures

4.7 and A2.9). The detection rate was highest with the best, lower with the second best, and lower still with the third. SAR data often has fewer bands than optical data, for example, so the choice of which metric to use in time series analyses may be more straightforward. However, many studies do not compare the results among metrics to select an optimal, relying instead on supposition (e.g. Reiche et al., 2018a). Our findings suggest Mean Decrease in Accuracy is useful for variable selection, even if the Random Forest models themselves are of little practical use (e.g. Figure 4.3).

4.5.2 Texture measures and detecting selective logging

In all cases the texture measures had the highest variable importance rankings (Figure A2.9). This corresponds with previous results with optical data, where detection of selective logging relied on the contextual information embodied within their calculation (Hethcoat et al. 2019). Similar to their results, the predictions of logging in our test areas were spatially correlated, presumably a consequence of the spatial window used in the calculation. Again, however, extra detections are expected from the accompanying forest disturbances associated with logging. Yet, in the context of accuracy assessment, an issue that has not received much attention within the remote sensing literature is how to report selective logging detections in the absence of robust field data on canopy gaps, roads networks, skid trails, log landing decks, etc. Others have shown that selective logging can be associated with 30-50% forest disturbance (Asner et al. 2002, 2004; Putz et al. 2019), depending on the intensity and logging practices (reduced impact versus conventional). Clearly Figure 4.8A has false discoveries associated with the breakpoint detections, but some of the detections that do not occur at a tree location undoubtedly correspond with canopy gaps seen in the Planet imagery.

While the texture information clearly helped with detection of selective logging, a sensible understanding of what the sum average metric means, in terms of characterizing forest disturbances from selective logging or understanding the structural changes to forests associated with increasing and decreasing values, remains unknown. Attempts to generalize and interpret the meaning of textures have proven difficult over the years. However, some have suggested that high values in measures like variance, dissimilarity, entropy, and contrast were associated with visual edges whereas average, homogeneity, correlation, and angular second moment were associated with subtle irregular variations from continuous regions like forests or water (Hall-Beyer 2017). More work is needed to understand the interpretation of textures measures that are so often employed in remote sensing classifications.

3.5.3 Combining sensors for classification

We chose not to combine any of the data types used here, partly because the inconsistent spatial and temporal coverage precluded such an analysis, but also because we wanted to assess the detection capabilities of each sensor on its own. Methods that combine data from multiple sensors (both other SAR platforms and/or optical data from Landsat or Sentinel-2) would likely perform better, corresponding with results for monitoring deforestation (Mercier et al. 2019; Reiche et al. 2015, 2016; Reiche, Verhoeven, et al. 2018). Indeed, prior work with Landsat data has shown strong detection of selective logging at similar intensities (Hethcoat et al. 2019), yet this work sought to establish a baseline with the SAR sensors available. The general direction and momentum for the advancement of detecting subtle forest disturbances from spaceborne SAR will likely require time series, polarimetric, and data fusion approaches, particularly in light of our findings that pixel-based differences between logged and unlogged areas with SAR backscatter alone cannot do the job effectively.

4.5.4 Longer time series in the tropics

Sentinel-1A began acquiring imagery regularly (approximately every 12 days) in late 2016 for most of Brazil, with Sentinel-1B following in late 2018. Consequently, a time series assessment was only possible for a single calendar year (roughly 2017) with the logging data sets we had access to. The BFAST algorithm is generally flexible and can be tuned with a baseline period if sufficient data are available, enabling assessments of longer and more variable time series (Verbesselt et al. 2010). The limited time series available is likely the reason many breakpoints for the less intensively logged sites occurred in December, presumably with the onset of the rainy season in earnest and an uptick in backscatter associated with moisture. Our analysis, however, was limited to a simpler test of one or no breakpoints – future work should explore how longer time series might improve detection of lower intensity logging, where seasonal patterns in backscatter can be established as a baseline to help reduce false alarms.

4.6 Conclusion

Tropical selective logging is fundamentally connected to global climate, biodiversity conservation, and human wellbeing (Lewis et al. 2015). Selective logging is often the first disturbance to affect primary forest (Asner et al. 2009), with road networks and ease of access facilitating further disturbances (e.g. increased fires, hunting or illegal logging). Efforts to detect and map selective logging with Sentinel-1, because of its global coverage and anticipated continuation missions (i.e. Sentinel-1C and D), are urgently needed to understand the capabilities this data stream might offer at advancing detection of tropical selective logging activities. With the successful launch of SAOCOM 1A in late 2018, the planned continuation of Sentinel-1 (with C and D), the opening of the ALOS PALSAR-1 archives, and the anticipated

launches of SAOCOM 1B in 2019 and NISAR in 2021, an immense volume of freely available C- and L-band SAR data will, hopefully, usher in a new era of forest monitoring from space with SAR data. Our findings suggest that time series methods should be effective at detecting the most intensive selective logging in the Amazon with these data sets. Moreover, if a distinct dry season is characteristic of the study region, focusing detecting during this time frame can further bolster detection by removing false positive detections associated with seasonal rainfall.

Acknowledgements

MGH was funded by the Grantham Centre for Sustainable Futures. JMBC was funded by the Natural Environment Research Council (Agreement PR140015 between NERC and the National Centre for Earth Observation). RADARSAT-2 imagery was provided by MDA through an ESA agreement under proposal 126091 and PALSAR-2 imagery was provided by JAXA. We would like to thank Planet Labs for access to imagery through the education and outreach program.

Chapter

5

Combining optical and SAR data for monitoring tropical selective logging

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, and S Quegan

Abstract

Earth's tropical forests play a key role in the global carbon and hydrological cycles, maintaining biological diversity, and supporting the global economy and human livelihoods. Yet, continued loss and degradation of tropical forests, coupled with swelling population and energy demands, are putting increasing pressure on forests globally. In recognizing some of these challenges, the United Nation's (UN) has developed the Reducing Emissions from Deforestation and Forest Degradation (REDD+) programme, which seeks to mitigate climate impacts and biodiversity losses through improved forest management. However, consistent and reliable forest monitoring systems are still needed to monitor tropical forests at large scales and REDD+ projects have seen little progress in reporting and monitoring impact. Recent advances in combining optical data and Synthetic Aperture Radar (SAR) data have shown promise for improved ability to monitor forest losses, particularly in cloudy regions. However, to date, no study has examined combining optical and SAR data from selective logging monitoring. We used detailed selective logging records from three lowland tropical forest regions in the Brazilian Amazon to test the effectiveness of combining Landsat and Sentinel-1 for selective logging detection. We built Random Forest models to classify pixel-based differences in logged and unlogged regions to understand if combining optical and SAR improved the detection capabilities over optical data alone. We found that the classification accuracy of models with optical data from Landsat 8 alone was slightly higher than models that combined SAR and Landsat. In general, detection of selective logging was high in both models (Landsat only and Landsat-SAR combined) with the validation dataset, but performance was lower over new regions. Overall our results show that adding SAR data did not improve the detection of selective logging and the optical data was dominating the importance and performance of models. The results have important implications for current and future abilities to detect selective logging with freely available satellite data. While we have shown limited capabilities with C-band here, the anticipated opening of the ALOS PALSAR-1 archives should stimulate research investigating similar methods to understand if longer wavelength SAR might help with classification.

5.1 Introduction

Earth's tropical forests play an important role in the global carbon and hydrological cycles, maintaining biological diversity, and supporting the global economy and human livelihoods (Pan et al. 2011; Lewis et al. 2015; Barlow et al. 2016; Baccini et al. 2017). However continued loss and degradation of tropical forests, coupled with swelling population and energy demands, are putting tremendous pressure on forests globally (Edwards et al. 2019). In recognizing some of these challenges, the United Nation's (UN) has developed the Reducing Emissions from Deforestation and Forest Degradation (REDD+) programme, which seeks to mitigate climate impacts and biodiversity losses through improved forest management practices (GOFC-GOLD, 2016). In order to be eligible for REDD+ funding, however, developing countries must show progress toward reducing degradation and deforestation emission. Yet, consistent and reliable forest monitoring systems are still needed to monitor tropical forests at large scales and most REDD+ projects have seen little progress in reporting and monitoring impact (Milbank et al. 2018).

Satellites offer the most accurate and cost effective way to monitor forests for country level reporting under REDD+. The technological capabilities to monitor tropical forests with satellite data have greatly improved over the last 10-15 years. Reliable deforestation alerts are available in near real-time from a number of organization, like Global Forest Watch and the Brazilian National Institute for Space Research (Hansen et al. 2013; Diniz et al. 2015). In contrast, detection and monitoring of forest degradation has lagged behind because of the complex and subtle disturbances associated with the range degradation activities (Ghazoul et al. 2015). Recent advances in monitoring selective logging with optical data have showed promise in monitoring forest degradation (Bullock et al. 2018; Hethcoat et al. 2019), however, optical data are limited in some regions with frequent cloud cover.

Synthetic Aperture Radar (SAR) data offers potential to advance detection of forest disturbances in regions with frequent cloud cover. SAR satellites transmit radio waves and do not require solar illumination for data acquisition. Thus SAR sensors can penetrate clouds and operate at night, and have been used in forest mapping since the early 1990s (reviewed in Koch, 2010). Historically, the SAR data archives have been spatially and temporally scattered, with few programs operating systematically to acquire global data. The Japanese Space Agency (JAXA) are an exception with the ALOS missions, however, those data products are under commercial licenses and imagery costs over £1500. The launch of Sentinel-1 in late 2014 has provided free C-band SAR data with global coverage every 5-12 days. This has spurred the development of using dense time series of Sentiel-1 for detecting deforestation (Reiche et al. 2015, 2018b). In addition, methods that combine optical and SAR data have been developed to improve detection of deforestation (Vaglio Laurin et al. 2013; Reiche et al. 2015; Joshi et al.

2016; Mercier et al. 2019). However it remains unclear if the recent advances in degradation detection and monitoring (Bullock et al. 2018; Hethcoat et al. 2019) could be improved with the combination of optical and SAR data.

We have chosen to focus on a key driver of forest degradation globally, selective logging operations. While selectively logged forests have been shown to have increased microclimatic variability (Stratford & Robinson, 2005), increased soil erosion (Douglas, 1999; Hartanto et al., 2003), reduced tree diversity (Berry et al., 2008; Martin et al., 2015), altered forest phenology (Koltunov et al., 2009), and lowered levels of biodiversity (reviewed in Burivalova et al., 2014), forests subjected to selective logging generally maintain higher levels of biodiversity than other anthropogenic land use types, such as plantations or secondary forests (reviewed in Edwards et al., 2014). Moreover, recent works have shown that even after accounting for the amount of wood removed, RIL has a greater effect on maintaining biodiversity than conventional selective logging (CL) practices (Bicknell et al., 2014) while simultaneously sequestrating more carbon during regrowth (Putz et al., 2008b). Thus, in the context of REDD+ or alternative conservation initiatives, forests impacted by RIL offer high biodiversity value and carbon sequestration potential, making them ideal for carbon and biodiversity co-benefits. However, commercial logging is often the first anthropogenic disturbance event to affect primary forests and is an agent for additional changes, facilitating more forest losses and other forms of degradation (Nepstad et al., 1999; Asner et al., 2005, 2006, 2009). In general, improved methods are needed to detect and monitor tropical selective logging activities, whether for identifying areas for inclusion in REDD+ type programs or enabling national monitoring efforts to qualify for REDD+ funding.

Recently, Hethcoat et al. (2019) have demonstrated success in detecting and mapping selective logging with optical data from the Landsat program. In addition, they also found that SAR data, on its own, was insufficient for accurately detecting selective logging across a range of logging intensities (Hethcoat et al. in prepaparation). The primary objective of this work was to extend those analyses (i.e. combine methods from Chapters 3 and 4) to understand how combining optical data from Landsat and SAR data from Sentinel-1 might improve detection capabilities of tropical selective logging. Specifically, while we showed promising detection capabilities with Landsat data on its own (Chapter 3) and weak detection with SAR data on its own (Chapter 4), we anticipated further increases in performance over those seen in Chapter 3. We utilized generally similar methods, but generated a completely independent data set from those used in Chapters 3 and 4.

5.2 Study area and data

5.2.1 Study are and selective logging data

Selective logging data from three lowland, *terra firma* tropical forest regions in the Brazilian Amazon were used in this study (Figure 5.1). The Jacunda and Jamari regions, inside the Jacundá and Jamari National Forests, Rondônia, and the Saraca region, inside the Saracá-Taquera National Forest, Pará. Forest inventory data from 11 forest management units (FMUs) selectively logged between 2016 and 2017 were used, comprising over 25,000 individual tree locations. Unlogged data from three additional locations, one inside each national forest (Jacunda, Jamari, and Saraca), comprised approximately 8,000 randomly selected point locations known to have remained unlogged during the study period (Table 5.1).

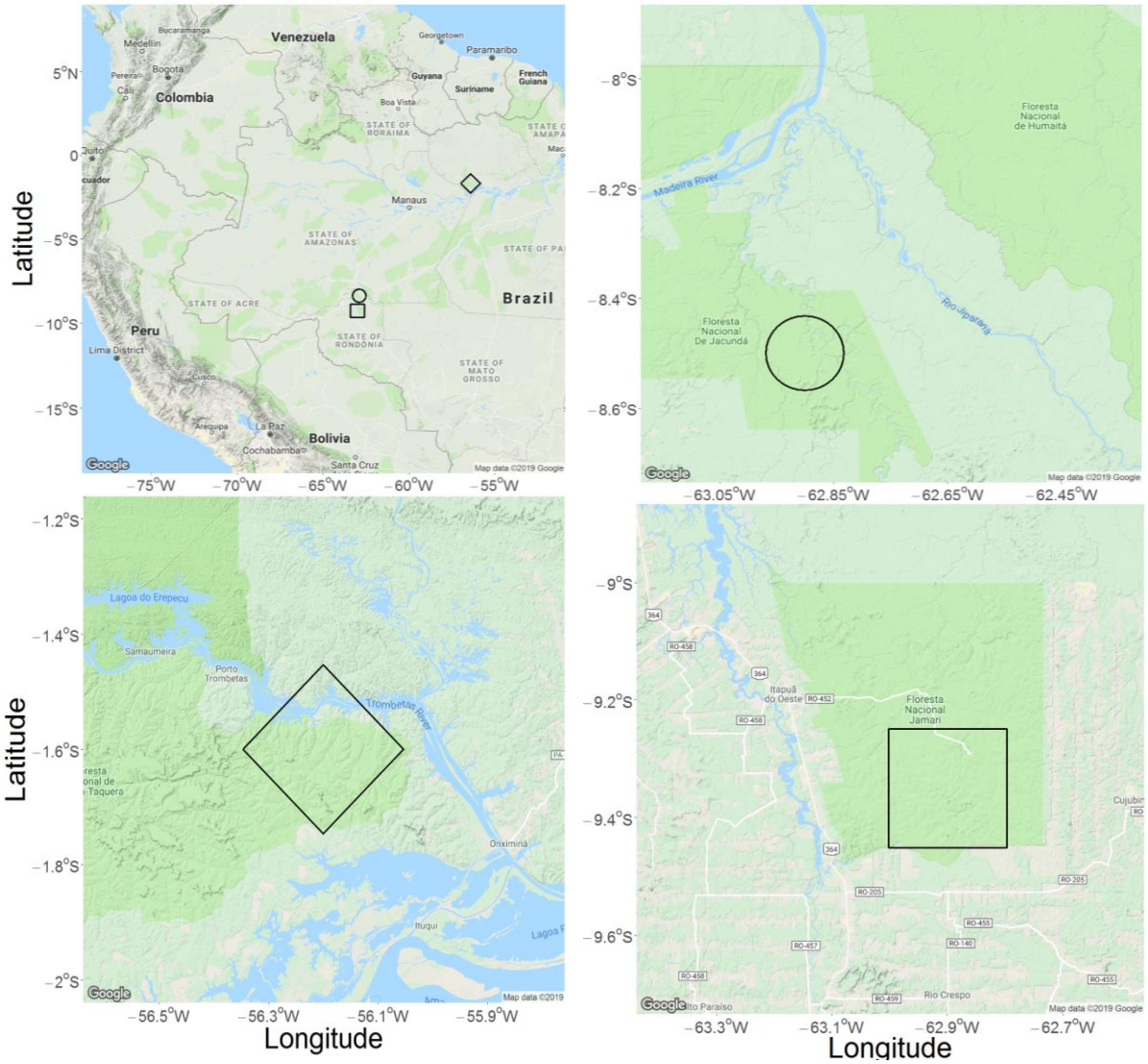


Figure 5.1 Location of the Jacunda (circle), Jamari (square), and Saraca (diamond) study regions in the Brazilian Amazon.

Table 5.1 Data used in the classification of selective logging from eleven forest management units (FMU) at three study regions in the Brazilian Amazon.

FMU	Logging Intensity (m^3 ha^{-1})	N
Jacunda_I_2016	6	2,290
Jacunda_I_2017	9	2,822
Jacunda_II_2016	10	1,815
Jacunda_II_2017	7	1,310
Jacunda_Reserve	0	3,000
Jamari_I_2016	10	653
Jamari_I_2017	12	911
Jamari_III_2016	9	2,058
Jamari_III_2017	11	2,597
Jamari_Reserve	0	1,912
Saraca_Ia_2017	12	3,769
Saraca_II_2016	25	3,223
Saraca_II_2017	21	4,729
Saraca_Reserve	0	3,000

5.2.2 Satellite data and processing

The Landsat 8 and Sentinel-1 data archives were queried in Google Earth Engine (GEE) to obtain a single image over each FMU that was late into the dry season logging period, but before the onset of the rainy season each year (Hethcoat et al. 2019). This was to ensure that as many of the logging locations had been logged, but a cloud free Landsat 8 image was still available. While Sentinel-1 can penetrate clouds, backscatter can be affected by rainfall, surface water, and soil moisture (Flores-Anderson et al. 2019). Consequently, Sentinel-1 imagery was acquired within a similar time frame and the Landsat imagery at each site. A summary of image path, row, and acquisition dates can be seen in Table A3.1.

GEE is a cloud computing platform hosting satellite imagery that has been processed to varying levels. We used the Landsat 8 Surface Reflectance collection and the Sentinel-1 Ground Range Detected, Interferometric Wide mode (VV and VH) collection. GEE calibrates and ortho-corrects Sentinel-1 imagery in the following steps using the Sentinel-1 Toolbox: (1) thermal noise removal; (2) radiometric calibration; and (3) terrain correction using the Shuttle Radar Topography Mission (SRTM) 30 m digital elevation model (DEM). The resulting images have a pixel size of 10 m. We further processed the Sentinel-1 imagery to remove inherent noise (i.e.

speckle) in the SAR signal (Quegan and Yu 2001) and finally we reduced the pixel resolution (via a mean) to produce 30 m pixels that corresponded with Landsat 8 pixels at each location.

Given that forest disturbances from selective logging affect patches of forest and have associated canopy gaps, skid trails, etc. we calculated four Grey Level Co-occurrence Matrix (GLCM) metrics for each band. A 5x5 window was used to calculate the Sum Average, Entropy, Contrast, and Angular Second Moment metrics (Haralick et al. 1973). A 5x5 window was used to further reduce the correlations amongst predictor variables and model predictions of logging (see Figure 3.9 and discussions therein). The full dataset thus comprised a 45-element vector (6 Landsat surface reflectance bands, 24 Landsat texture measures, 3 SAR bands, 12 SAR texture measures) for each pixel where logging occurred and an additional 2000 randomly selected pixels in an adjacent FMU that remained unlogged between. The data were exported from GEE and collated in R version 3.5.1 (R Core Team 2018) for analyses.

5.3 Methods

5.3.1 Supervised classification with Random Forest

We built Random Forest (RF) models using the *randomForest* package in program R version 3.5.1 (Liaw and Wiener, 2002; R Development Core Team, 2018). The RF algorithm (Breiman 2001) is a machine learning technique that uses an ensemble method to identify a response variable (here, whether a pixel was logged or unlogged) given a set of predictor variables (e.g. surface reflectance values). In contrast to a single decision tree, RF models employ multiple, independent decision trees (hence a forest). Random subsets of the training data are drawn, with replacement, to construct many trees in parallel, with each tree casting a vote on which class should be assigned to the input data. The withheld subset of the data, called the out-of-bag fraction, can be used for validation in the absence of independent validation data (Breiman 2001). To reduce generalization error, RF also uses a random subset of predictor variables in the decision at each node within a tree during construction.

We split the early and late datasets into 90% for training and 10% was withheld for validation. We spatially filtered the training and validation datasets such that no observation from training was within 90 m of an observation within the validation dataset. RF models have only two tuning parameters: the number of classification trees to be produced (k), and the number of predictor variables used at each node (m). We used 10-fold cross-validation to identify the number of trees (k = 500) and the number of variables to use at each node (m = 7) that minimized the out-of-bag error rate on the training data.

5.3.2 Model validation

RF models were validated using a random subset of the full dataset for each sensor (described in Section 5.3.1). By default, RF models assign an observation to the class indicated by the majority of decision trees (Breiman, 2001a). However, the proportion of trees that voted for a particular class from the total set of trees can be obtained for each observation and a classification threshold (T) can be applied to this proportion (Liaw and Wiener 2002; Hethcoat et al. 2019). We used similar methods from Hethcoat et al. (2019) to selection the detection threshold, however for context, we provide model performance across all values of T.

The confusion matrix then has the form:

		Reference	
		L	UL
Predicted	L	D_L	D_{UL}
	UL	$N_L - D_L$	$N_{UL} - D_{UL}$

where L and UL refer to logged and unlogged classes, N_L and N_{UL} are the numbers of logged and unlogged observations in the reference dataset, and D_L and D_{UL} are the numbers of logged and unlogged pixels detected as logged, respectively. We defined the *detection rate* $DR = D_L/N_L$ and *false alarm rate* $FAR = D_{UL}/N_{UL}$ as the frequency that a logged or unlogged pixel was classified as logged, respectively. Thus, the DR is equivalent to 1 minus the omission error of the logged class and the FAR is the omission error of the unlogged class. In addition, we defined the *false discovery rate* (FDR):

$$\text{FDR} = \frac{D_{UL}}{D_L + D_{UL}} = 1 - \frac{1}{1 + \left(\frac{N_{UL}}{N_L}\right)\left(\frac{\text{FAR}}{\text{DR}}\right)}. \quad (5.1)$$

The FDR is the proportion of all observations that were detected as logged that were actually unlogged, and is equivalent to the commission error of the logged class. See Hethcoat et al. (2019) for further explanation.

5.4 Results

5.4.1 Landsat 8 only

We present the results from a model that only used Landsat 8 data initially to form the baseline upon which the combined results will be compared. Random Forest performance for the model that used only Landsat 8 is shown in Figure 5.2. As the threshold value increased, the false alarm

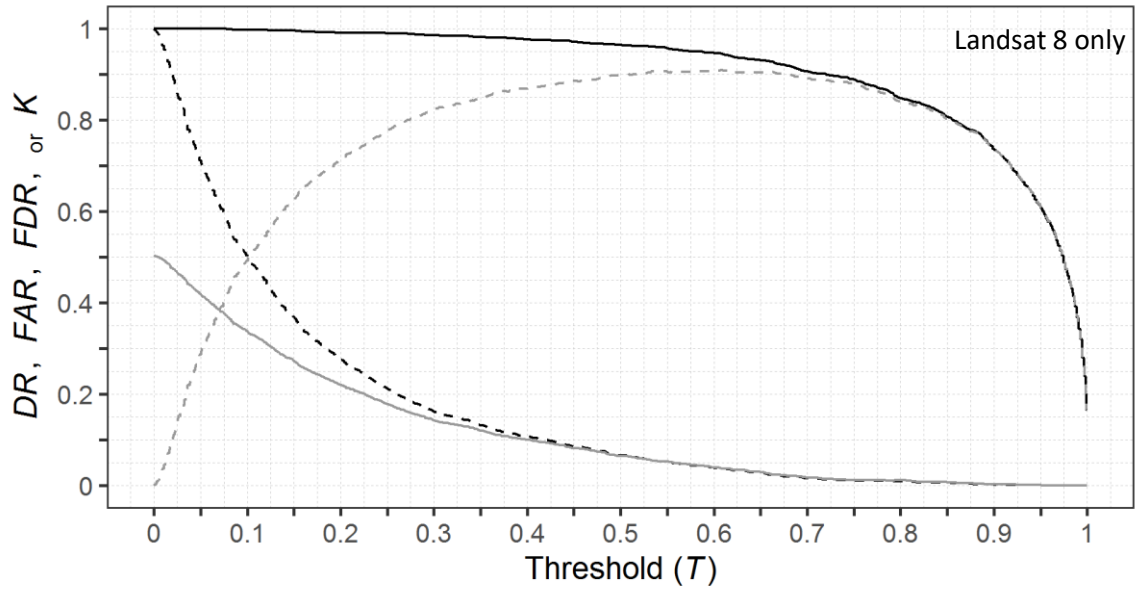


Figure 5.2 Random Forest model performance using Landsat 8, bands 2-7 and four GLCM textures measure for each band. The detection rate and the false alarm rate are the solid and dashed black lines, respectively. The false discovery rate and Cohen's kappa are the solid and dashed grey lines, respectively.

rate declined gradually and the detection rate stayed very high, displaying an impressive ability to distinguish logged observations (Table 5.2). This detection rate was generally higher than the results from Chapter 3 (though the false alarm rate is initially higher here). We chose a very high threshold value in this case to strongly limit the FDR, causing higher omission of logging than Chapter 3. This was done because despite limiting FDR in model training it was generally slightly higher when predicting to new areas in Chapter 3 (see Figure 3.6). Thus, we hoped to pre-empt this when producing maps with future models. When this model was applied to

Table 5.2 Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at three study areas in the Brazilian Amazon, derived from Landsat 8 data. Data were split into 90% training and 10% validation. Matrix numbers are pixel counts with the validation data (n = 4068). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Landsat 8 only				***T* = 0.9**
OA: 87.3%				
κ: 0.75		Reference Class		
FDR: 0.7%		Logged	Unlogged	Commission
DR: 75.0%				Error (%)
Predicted Class	Logged	1514	11	0.7
	Unlogged	506	2037	19.9
	Omission Error (%)	25.0	0.5	

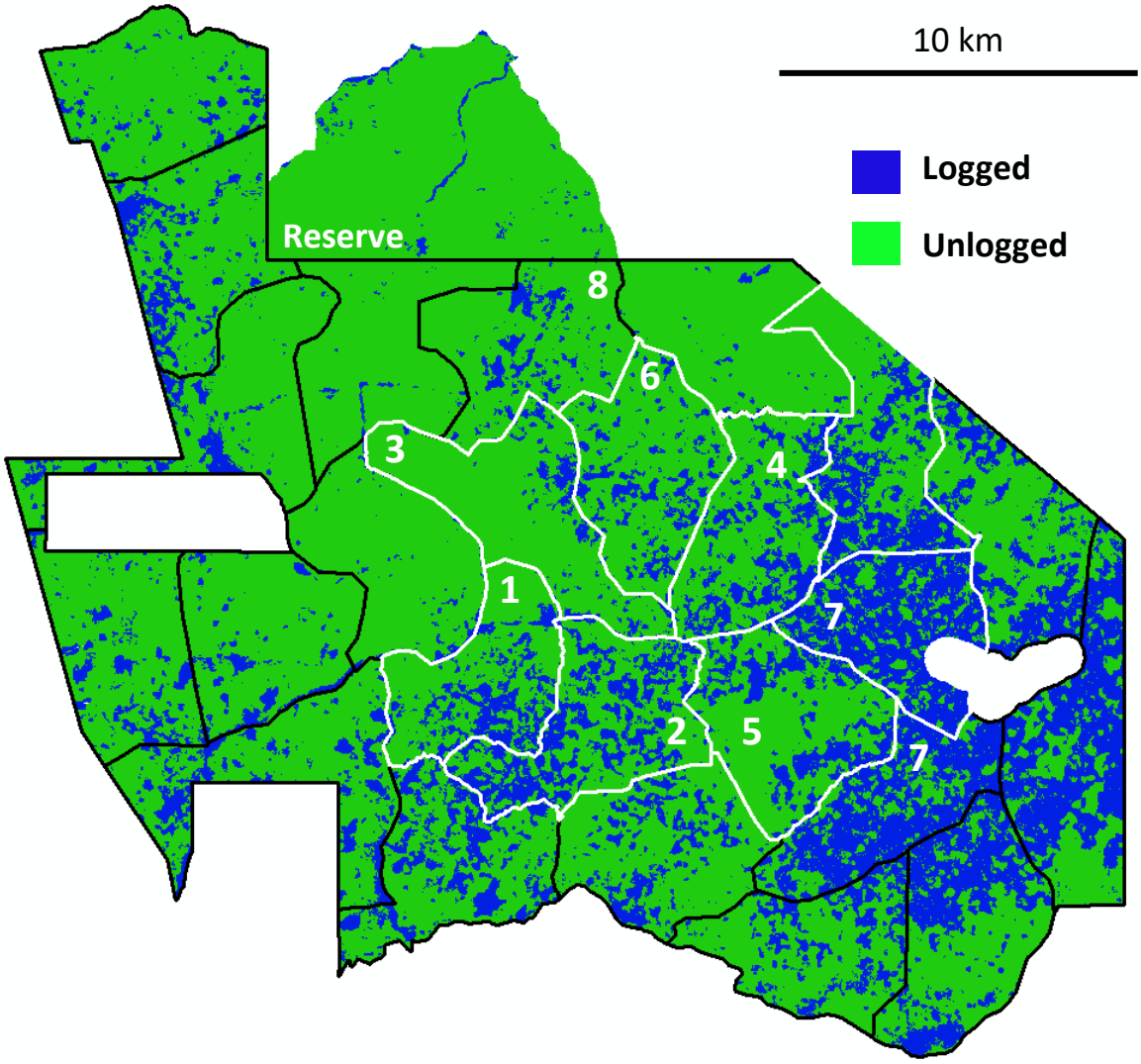


Figure 5.3 Classified map (using a 0.9 threshold) of selective logging detections for the Jamari region with the Landsat 8 model. The forest management units that have been logged are bounded in white, those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black). Numbers correspond to the year of logging, with year 2011 being 1 through year 2018 being 8. Note two FMUs were logging in 2017.

Landsat 8 imagery over an entire logging concession, it was clear that the model did not perform quite as well as the validation suggested (Figure 5.3). The unlogged regions south and east of the units cut in 2017 showed more than half of the area predicted to have been logged. In addition, regions that had been logged previously showed a retained signal beyond a year or two that was observed in Chapter 3 (see Figure 3.6). Further refinements to get the highest possible accuracy were beyond the scope of our objectives here, as these results are for comparison to the combined Landsat Sentinel-1 model (next). In general, however, the results here are in line with previous findings in Chapter 3.

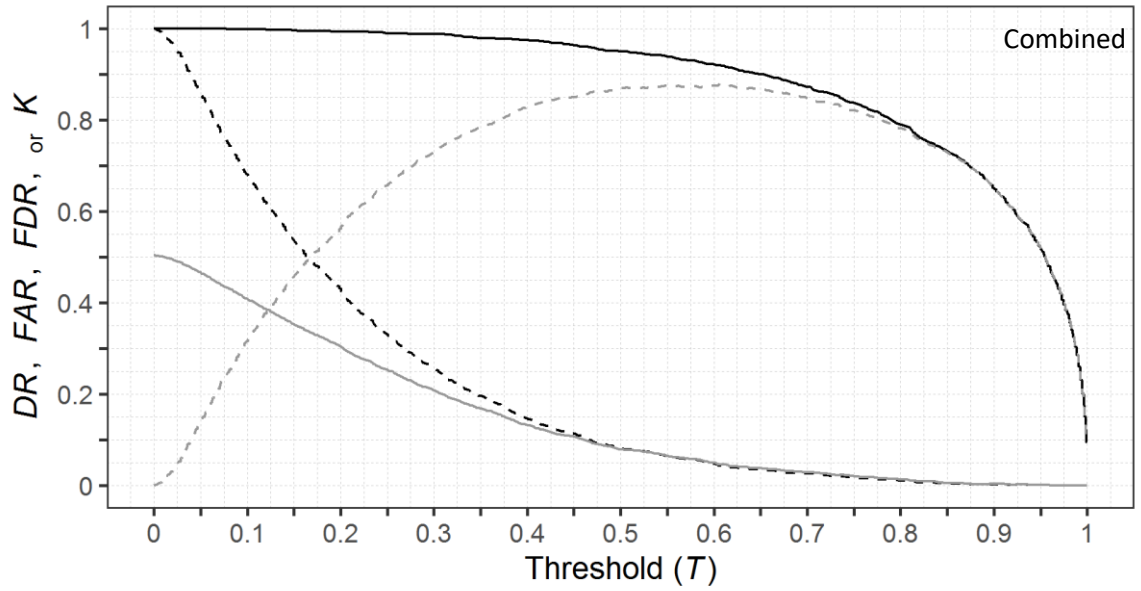


Figure 5.4 Random Forest model performance using combined Landsat 8 and Sentinel 1 data. The detection rate and the false alarm rate are the solid and dashed black lines, respectively. The false discovery rate and Cohen's kappa are the solid and dashed grey lines, respectively.

5.4.2 Optical and SAR combined

Random Forest performance for the model that used both Landsat 8 and Sentinel-1 is shown in Figure 5.4. Again, the false alarm rate gradually declined and the detection rate stayed very high as the threshold value increased. The detection rate declined slightly faster and the false alarm rate declined slightly slower in this combined model when compared to the Landsat only model, though the differences were negligible (Table 5.3). In general, the results with the combined model were very similar to those with the Landsat only model, suggesting very little performance boost. When this model was applied to Landsat 8 and Sentinel-1 imagery over an

Table 5.3 Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at three study areas in the Brazilian Amazon, derived from Landsat 8 and Sentinel-1 data. Data were split into 90% training and 10% validation. Matrix numbers are pixel counts with the validation data (n = 4068). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Combined				***T* = 0.9**
OA: 82.7%				
κ: 0.65		Reference Class		
FDR: 1.0%		Logged	Unlogged	Commission
DR: 65.7%				Error (%)
Predicted Class	Logged	1328	13	1.0
	Unlogged	692	2035	25.4
	Omission Error (%)	34.3	0.6	

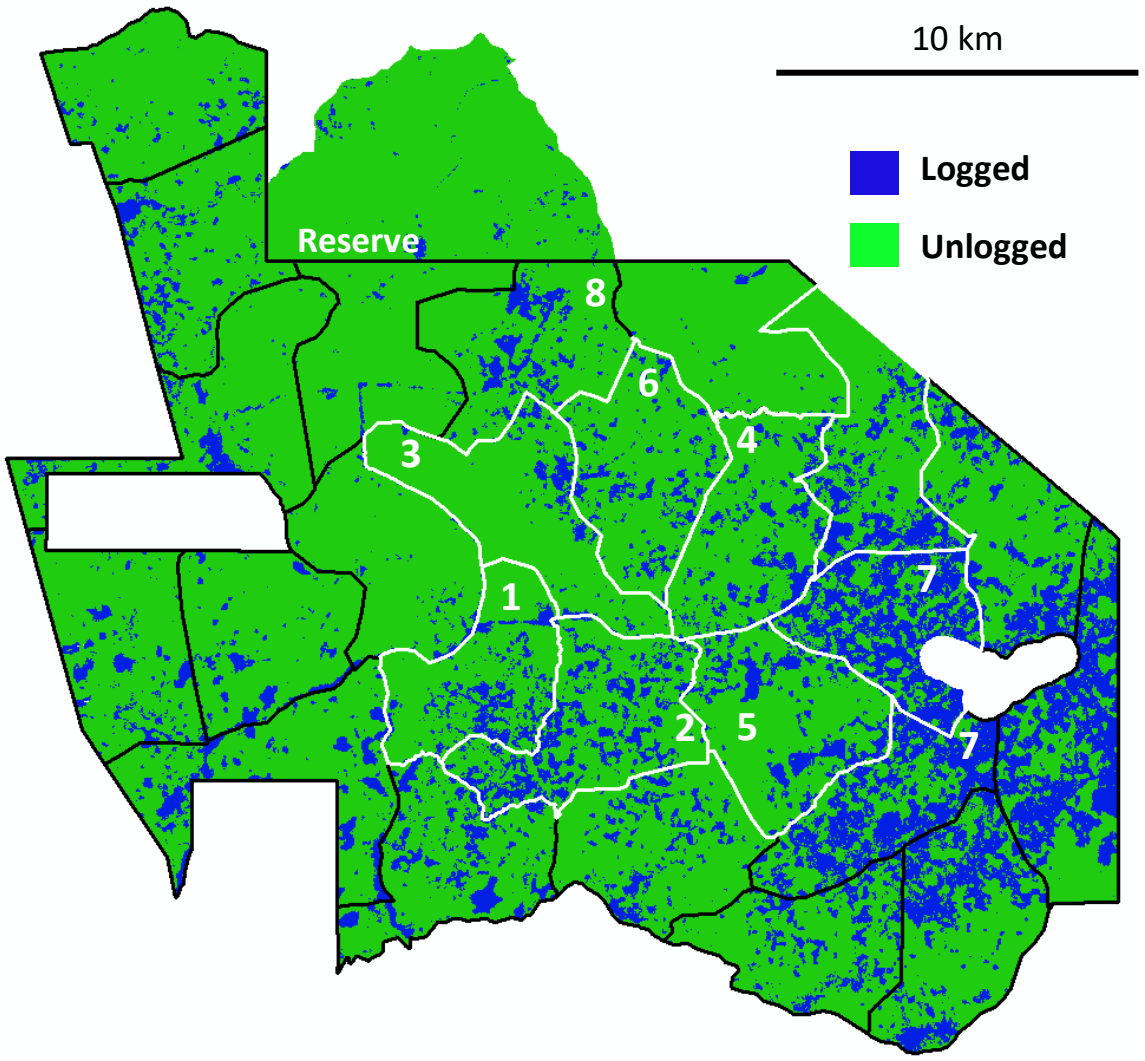


Figure 5.5 Classified map (using a 0.9 threshold value) of selective logging for the Jamari region using the combined Landsat 8 and Sentinel-1 model. Landsat data are from 2017-09-07 and the Sentinel-1 data are from 2017-09-25. The forest management units (FMU) that have been logged are bounded in white and those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black). Numbers correspond to the year of logging, with year 2011 being 1 through year 2018 being 8. Note two FMUs were logging in 2017.

entire logging concession (Figure 5.5), it was clear that the predictions were being driven by the Landsat data, as Figures 5.3 and 5.5 are almost identical. In trying to understand the model's performance, we examined the predictions from a separate model that only had Landsat 8 data and no GLCM texture measures (in an effort to further limit potential spatial autocorrelation in predictions and understand what could be causing large regions to be labelled logged). The results from this model further support the idea that the surface reflectance data were driving the classification in Figure 5.5 (Figure 5.6). We observed this phenomenon previously in our mapping efforts in Chapter 3, where subtle variations in surface reflectance, from varying solar angles, resulted in occasionally erratic predictive behaviour (see Figure 3.6, panel 2016-06-16).

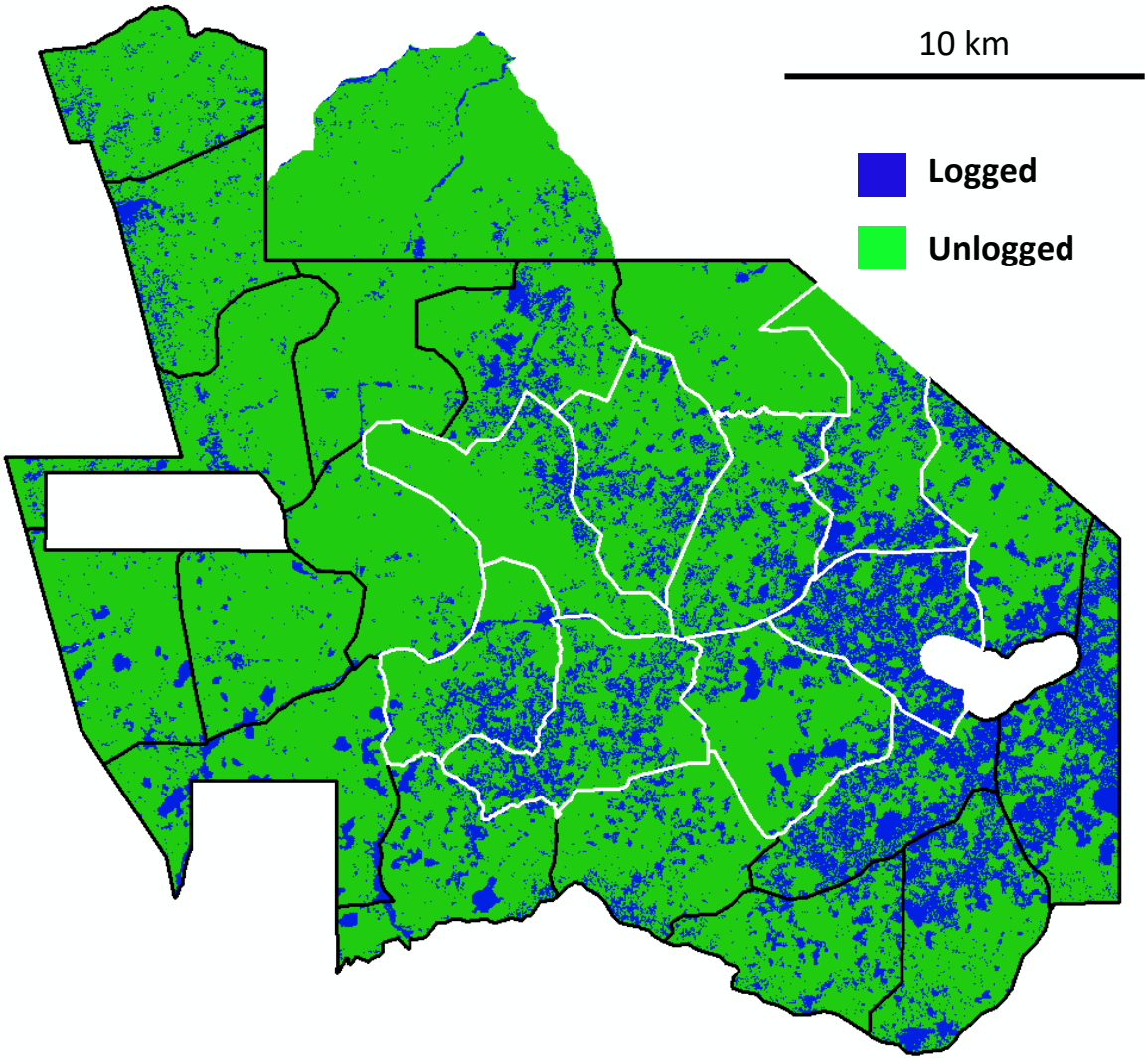


Figure 5.6 Classified map (using a 0.9 threshold) of selective logging detections for the Jamari region with the Landsat 8 model without GLCM textures. The forest management units that have been logged are bounded in white, those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black).

Known as the bidirectional reflectance distribution function (BRDF), once characterized Landsat data can be adjusted to harmonize reflectance data and help reduce this effect (Roy et al. 2016). Ultimately, however, this would not change the fact that the adding SAR data did not improve detection of logging. Overall, these findings suggest that Sentinel-1 offers little extra information content within a supervised classification scheme and the Landsat data are driving the classification results. This extends our previous findings in Chapter 4, that SAR data was too noisy for supervised classification of selective logging on their own (see Figure 4.3), to include their use in combination with optical data (for supervised classification).

5.4.3 Sentinel-1 only

Random Forest performance for the model that only used Sentinel-1 data is shown in Figure 5.7. The false alarm and detection rates gradually declined as the threshold value increased, with the rates declining roughly in parallel, suggesting difficulty in distinguishing logged and unlogged observations (Table 5.4). These findings generally corroborate the results from Chapter 4, wherein classification performance with Sentinel-1 was poor. Those data were at full resolution (i.e. not spatially averaged to align with Landsat 8 pixels) and included many more unlogged observations, yet the results are generally similar. We have included the Sentinel-1 only results here for completeness.

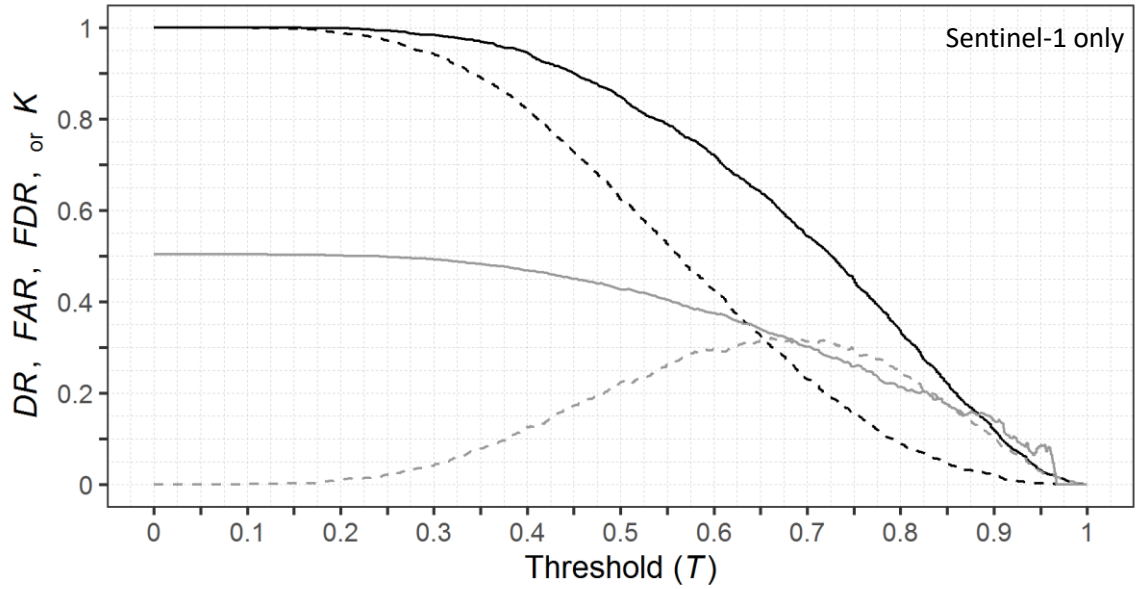


Figure 5.7 Random Forest model performance using only Sentinel 1 data. The detection rate and the false alarm rate are the solid and dashed black lines, respectively. The false discovery rate and Cohen's kappa are the solid and dashed grey lines, respectively.

Table 5.4 Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at three study areas in the Brazilian Amazon, derived from Sentinel-1 data. Data were split into 90% training and 10% validation. Matrix numbers are pixel counts with the validation data (n = 4068). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Sentinel-1 only				***T* = 0.9**
OA: 52.9%				
κ: 0.05		Reference Class		
FDR: 24.1%		Logged	Unlogged	Commission
DR: 7.5%				Error (%)
Predicted Class	Logged	151	48	24.1
	Unlogged	1869	2000	48.3
	Omission Error (%)	92.5	2.3	

5.5 Discussion

We have shown that combining C-band SAR from Sentinel-1 with optical data from Landsat does not offer a performance advantage over simply using optical data to detect tropical selective logging. Prior studies have shown an improvement in deforestation monitoring by combining optical and SAR (Erasmi and Twele 2009; Vaglio Laurin et al. 2013; Reiche et al. 2015, 2018a). Unfortunately a similar improvement did not occur when detecting selective logging. It should be noted that most of the logging records were relatively low intensity (Table 5.1) and only 2 sites were close to the legal limit within the Amazon ($30 m^3$ ha^{-1}). However, we also show in Chapter 4 that classification with Sentinel-1 still performed relatively poorly even when restricted to the most intensively logged data, so this is unlikely to be the primary cause. Thus, it seems that other methods should be explored to detect selective logging with optical and SAR data (e.g. deep learning or time series methods).

The detection rates displayed in the results with the validation data were generally higher (and lower commission error) than the results from Chapter 3. A few factors are likely causing this. First, the dataset used here include sites logged at much higher logging intensities than in Chapter 3; in some cases >3 times the intensity (compare Table 5.1 and Appendix 1, Table A1.7). Areas logged at higher intensity are generally easier to distinguish from unlogged locations (Appendix 2, Figure A2.4). Second, GEE does not allow the user to specify two important parameters GLCM textures require for calculation (and the default settings are not listed within the documentation): 1) the number of grey levels to categorize the image into (i.e. the size of the GLCM matrix) and 2) the minimum and maximum value ranges within the data to quantize into the grey levels. This latter parameter is key, because if the range of minimum and maximum values in one particular region is greatly different from another (for example in SAR imagery occasionally high backscatter values can occur that are greatly outside the range expected from natural scattering objects) then the calculations for the texture measures are shifted between the regions and RF models can quickly pick up on this difference. We had this problem in the original analyses of Chapter 4 (where one of the SAR data types initially showed nearly perfect classification performance, but it was later discovered to be being driven by this phenomenon). When we manually controlled the range of values to reflect possible values of backscatter (power from 0-1 for example) the very high classification performance disappeared. In examining the raw values of Landsat and Sentinel-1 going into the GLCM calculations there don't seem to be any particular sites that could be causing this (Figures A4.1 and A4.2). We removed Landsat 8 Band 3 to see if some of the variation seen in that variable was causing it, but the results did not change. The second possibility is that the training and validation data were very similar, despite the spatial filtering we performed. This was happening to some extent, as there was clearly more than 1% commission error of logged observations in Figure

5.5 (see Table 5.3). Finally, the difference between the proportion of the logged observations between training and testing datasets was different in Chapter 3. Here, the training was 65% logged and the validation was 50%, but in Chapter 3 the values were both approximately 15% (see the starting d_{pL} values in Figure 3.4 and FDR values here). Random Forest classification performance, particularly in two-class problems, has previously been shown to be impacted by imbalanced datasets (Chen et al. 2004). This problem general affects extremely imbalanced datasets and we avoided this by defining our own classification threshold (as opposed to using the default settings).

Recent analyses have shown strong detection of forest degradation with a time series analysis of Landsat data (Bullock et al. 2018). In addition, our prior work has shown a time series approach is likely needed for utilizing SAR in logging detection (Chapter 4). Our results here should stimulate research looking to combine optical and SAR data in a joint time-series approach. While the availability of dense time series of historical SAR data are simply not there, this approach could be used in forward looking analyses or the development of alert systems that combine the frequency of data acquisitions of a combined Landsat, Sentinel-2 and Sentiel-1 workflow.

Acknowledgements

MGH was funded by the Grantham Centre for Sustainable Futures. JMBC was funded by the Natural Environment Research Council (Agreement PR140015 between NERC and the National Centre for Earth Observation). We would like to thank the Google Earth Engine developers and community for access to the datasets that made this work possible.

Chapter 6

Mapping the rapid expansion of selective logging in the south-west Brazilian Amazon

MG Hethcoat, JMB Carreiras, DP Edwards, RG Bryant, CA Peres, and S Quegan

In preparation for submission to *Environmental Research Letters*

Abstract

Tropical forests harbour the highest biodiversity on the planet and are essential to human livelihoods and the global economy. However, continued loss and fragmentation of forested landscapes, coupled with a rapidly rising global population is placing incredible pressure on forests globally. The United Nations has developed the Reducing Emissions from Deforestation and forest Degradation (REDD+) programme in response to the challenges facing tropical forests and in recognition of the role they can play in climate mitigation. However, REDD+ requires consistent and reliable forest monitoring of forest disturbances and currently does not include for forest degradation because of an inability to track it effectively. In this paper we extend a recent analysis enabling the detection of selective logging at the scale of a logging concession to a regional-scale assessment of selective logging activities. We utilized logging records from across Brazil to train a supervised classification algorithm for detecting logged pixels in Landsat imagery then predicted the extent of logging over a 20 year period throughout Rondônia, Brazil with the help of Google Earth Engine. We estimate that 41.0% of the State of Rondônia remained undisturbed forest through 2019, with 3.4% having undergone selective logging and 25.7% being deforested between 2000 and 2019. Selective logging was mapped with 13% Commission Error and 45% Omission Error over the twenty year period. In general, rates of selective logging were twice as high in the first decade relative to the last decade of the period. Our approach is step in this direction of an operationalized selective logging monitoring system capable of detecting subtle forest disturbances over large spatial scale.

6.1 Introduction

The ten countries reporting the highest forest losses over the last fifteen years are all within the tropics (FAO, 2016). Tropical forests are among the most biodiverse ecosystems on the planet while simultaneously playing a crucial role in the global carbon and hydrological cycles and supporting human livelihoods and the global economy (Pan et al. 2011; Lewis and Maslin 2015; Edwards et al. 2019). Moreover, there is increasing recognition that tropical forests will play a vital role in nature-based solutions to mitigating climate impacts and reaching targets outlined in the Paris Climate Agreement (Houghton et al. 2015; Griscom et al. 2017). However continued loss and degradation of tropical forests, coupled with a rising global population and growing energy demands, are putting enormous pressure on forests globally (Edwards et al. 2019).

In response to the both challenges and opportunities tropical forests present, the United Nations (UN) has developed the Reducing Emissions from Deforestation and forest Degradation (REDD+) programme. REDD+ aims to simultaneously mitigate climate impacts and maintain the myriad of services forests provide (e.g. flood prevention, control soil erosion, maintain biodiversity, cultural traditions, etc.) through sustainable forest management (UN-REDD 2018). An essential component in REDD+, however, is consistent and reliable monitoring systems for national-level reporting of greenhouse gas emissions associated with anthropogenic activities affecting forests. Methodological guidelines in monitoring and reporting emissions from degradation have been broadly linked to those used by the Intergovernmental Panel on Climate Change (IPCC) to facilitate a consistent framework for estimating reference levels and emissions from various REDD+ activities (GFOI, 2016). Yet the IPCC and REDD+ still lack specific methodological details on quantifying emissions from forest degradation (IPCC 2006; Pearson et al. 2014). This is because degradation is notoriously difficult to quantify, as it includes a variety of forest disturbances (e.g. fire, selective logging, mining, hunting, invasive species, etc.). In addition, forest degradation can often operate on a spatial and temporal scale incompatible (i.e. relatively small scale and short-lived) with reporting at the national level (Hosonuma et al. 2012; Pearson et al. 2014; Ghazoul et al. 2015). Consequently, REDD+ initiatives do not currently report emission associated with degradation (Hosonuma et al. 2012).

Advancements in remote sensing have made satellite data the most practical and cost-effective way to monitoring forests at large spatial scales. The preceding decade has witnessed rapid improvement in the spatial and temporal accuracy of deforestation monitoring (Hansen et al. 2013, 2016; Reiche et al. 2018). Simultaneously, abilities to map both the spatial extent and severity of fires has improved their detection and assessment of impacts (Peres et al. 2006; Matricardi et al. 2010). Yet, widespread detection and monitoring of selective logging activities has lagged behind, despite recognition that selective logging is a key driver of both deforestation and forest degradation (Hosonuma et al. 2012; Pearson et al. 2017). While

sustainable management of forest that result in no longer-term loss of carbon stocks (through less destructive harvest techniques like reduced-impact logging) are not considered degradation in the context of REDD+, there is clearly a need to take account of and adequately monitor extractive practices that could result in reduced carbon stocks. Selective logging is often the initial anthropogenic disturbance event to impact primary forests, with the opening of road networks and improved access to forested lands facilitating further degradation (e.g. fuel wood removal, spread of invasive species, illegal logging, mining, and fires) or forest clearance for pastures, agriculture, or human settlements.

Efforts to improve detection of selective logging have arisen periodically in the literature (e.g. Asner et al., 2005; Broadbent et al., 2008; Matricardi et al., 2010; Souza, Jr et al., 2013; Souza et al., 2005). In all cases the approach was either a proof-of-concept and not repeated at scale or the canopy damage associated with the intensity of selective logging was so high that many of the detections are later mapped as forest loss in the Hansen et al. (2013) data (e.g. Asner et al., 2006; see Results Section 6.3.2, Figure 6.9 and Appendix 4, Figure A4.4 in this study for more). Simultaneously, because of the role tropical forests are poised to play in tackling climate targets and growing concerns about the impacts to other services (biodiversity, water provisioning, cultural, etc.), the amount of tropical forests logged at lower intensity and with better management practices is likely to grow. In addition, there is an ever increasing need to detect and account for the estimated 50-90% of tropical timber on the international market harvested illegally at very low intensities (Kleinschmit et al. 2016; Brancalion et al. 2018).

The majority of the work focused on detecting logging has utilized spectral unmixing of before-after images to estimate forest disturbances between time steps (e.g. Souza, Jr et al., 2013). This approach has been criticised, as a single image analyses can omit forest disturbances occurring later and/or cloudy regions not visible during scene acquisitions. More recently, advancements in data access and handling (e.g. Google Earth Engine) have enabled time series methods to be developed that track pixel values over a long period to monitor forest disturbances (Bullock et al. 2018). Yet, these same advancements have allowed for more complex image mosaics to be produced, where a single image can now be composed of individual pixels spanning any time period, minimizing scene loss from clouds (Gorelick et al. 2017).

Recently, Hethcoat et al. (2019) developed a method that used logging records to train supervised classification algorithms for detecting logging activities. However, their methods have only been applied at the scale of the logging concession and have not been demonstrated operationally. The primary objective of this work was to extend the methodology proposed by Hethcoat et al. (2019), moving beyond the scale of a forest management plan or logging concession, to a regional-scale assessment of selective logging activities. We utilized detailed

logging records to train a supervised classification algorithm for detecting selectively logged pixels then predicted the extent of logging over a 20 year period throughout Rondônia, Brazil.

6.2 Study area and data

6.2.1 Study area

The Brazilian state of Rondônia is located along the western edge of the country, bordering Bolivia. Comprising 237,576 km^2, the state is one of the most deforested regions in the Amazon (Pedlowski et al., 2005; Tyukavina et al., 2017). While Pará and Mato Grasso have endured the highest rates of selective logging (Tyukavina et al. 2017), the smaller size of Rondônia and distinct dry season make it an ideal candidate for a preliminary upscaling of the methodology proposed by Hethcoat et al. (2019).

6.2.2 Selective logging data

Selective logging data from four lowland tropical forest regions in the Brazilian Amazon were used to build the detection algorithm (described in Section 6.3.1). The Jacunda and Jamari regions were inside the Jacundá and Jamari National Forests, in Rondônia, while the Saraca and Cikel regions were in the Saracá-Taquera National Forests and Paragominas municipality, Pará, respectively (Figure 6.1). Forest inventory data from 19 forest management units (FMUs) selectively logged between 2010 and 2017 were used, comprising over 55,000 individual tree locations. Unlogged data from three additional locations, one inside each national forest (Jacunda, Jamari, and Saraca), comprised over 11,500 randomly selected point locations known to have remained unlogged during the study period (Table A4.1).

6.2.3 Satellite data and processing

6.2.3.1 Generating training data for logging detection algorithm

All available Landsat 5, 7, and 8 surface reflectance data that coincided with logging were utilized in Google Earth Engine (GEE). At each FMU the Landsat archives were queried to find a single scene with the lowest cloud cover that was late into the dry season, but before the onset of the rainy season, to ensure the majority of logging was completed (Hethcoat et al. 2019). A linear spectral unmixing model, developed and validated over a range of forest disturbance types within the Amazon (Souza et al. 2005; Bullock et al. 2018), was used to convert surface reflectance into proportions of Bare Ground (BG), Photosynthetic Vegetation (PV), and Non-Photosynthetic Vegetation (NPV) in each pixel (Table 6.1). In addition, the normalized burn ratio (NBR) was calculated (Equation 6.1), because it has been shown to highlight changes in BG and NPV relative to PV and has demonstrated strong change detection capabilities in evergreen tropical forests (Grogan et al. 2015; Shimizu et al. 2017; Langner et al. 2018).

$$NBR = \frac{NIR - SWIR2}{NIR + SWIR2} \tag{6.1}$$

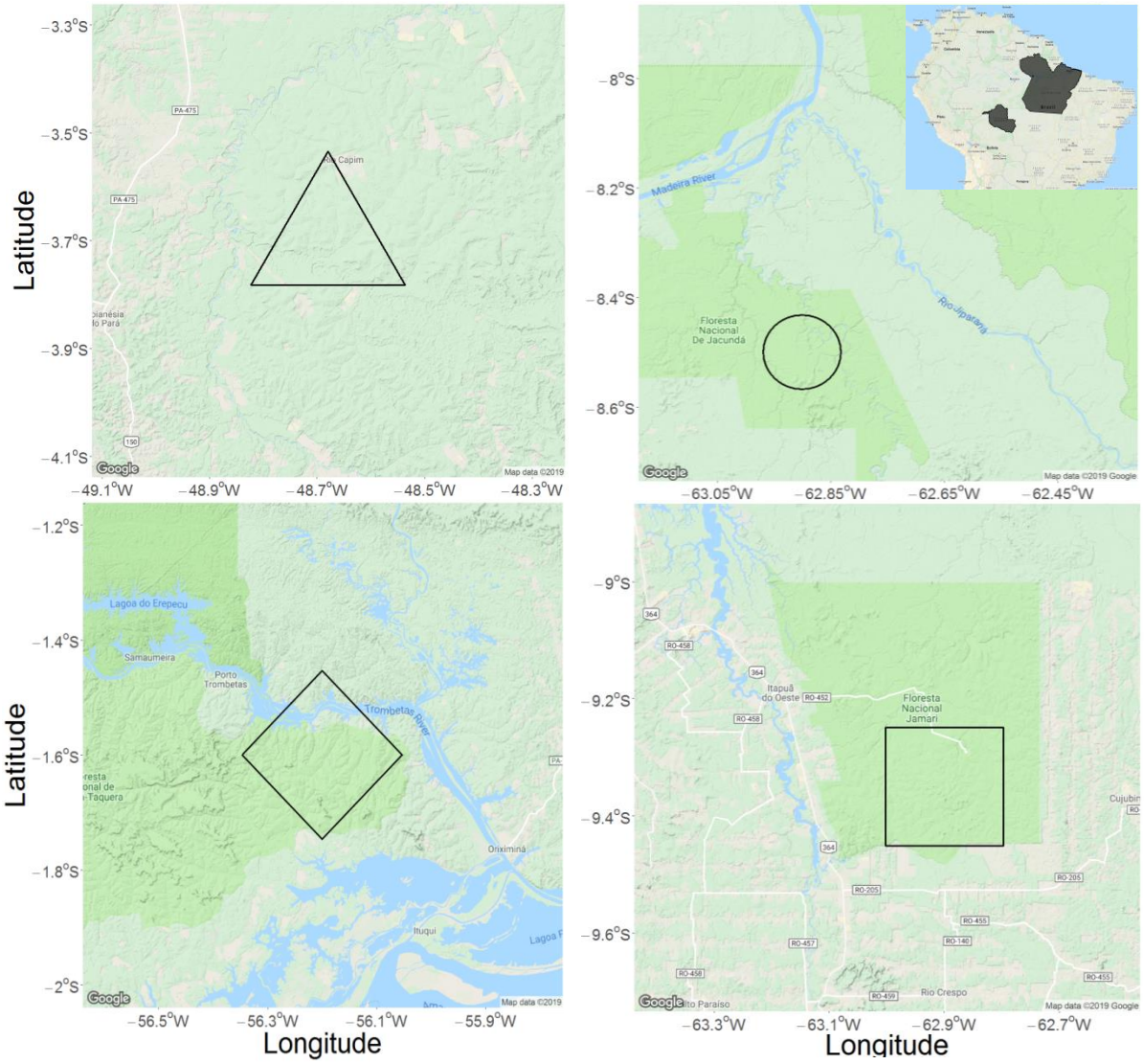


Figure 6.1 Location of the Cikel (triangle), Jacunda (circle), Jamari (square), and Saraca (diamond) study regions in the Brazilian Amazon. Cikel and Saraca are in Pará and Jacunda and Jamari are in Rondônia.

Table 6.1 Spectral endmembers used for unmixing analysis, developed from (Souza et al. 2005; Bullock et al. 2018), to calculate proportions of bare ground(BG), Photosynthetic Vegetation (PV), and Non-Photosynthetic Vegetation (NPV).

Endmember	Blue	Green	Red	NIR	SWIR1	SWIR2
BG	0.20	0.30	0.34	0.58	0.60	0.58
PV	0.05	0.09	0.04	0.61	0.30	0.01
NPV	0.14	0.17	0.22	0.30	0.55	0.30

The PV, NPV, and NBR values were spatially normalized in a self-referencing step to reduce subtle variation in their values through time, as a result of differing atmospheric conditions and solar illumination (Equation 6.2).

$$PV_n = PV_{median} - PV \quad (6.2a)$$

and

$$NPV_n = NPV_{median} - NPV \quad (6.2b)$$

and

$$NBR_n = NBR_{median} - NBR \quad (6.2c)$$

The median of a 150 m radius pixel window was calculated and the centre pixel value was subtracted from that median (Langner et al. 2018). Thus, normalized PV, NPV, and NBR values ranged between -1 and 1. An early version of our detection algorithm did not perform this step and suffered from extremely erratic predictive behaviour in adjacent Landsat paths from different dates (Figure A4.1). The values for the spatially normalized spectral unmixing and normalized burn ratio for the logged and unlogged observations were exported from GEE and compiled into a single dataset for Random Forest model training (Section 6.3.1).

6.2.3.2 Generating annual mosaics for Rondônia

All available Landsat 5, 7, and 8 data over Rondônia were utilized in Google Earth Engine (GEE). A cloud-free mosaic was made from the latest cloud-free pixel within the dry season (i.e. a quality mosaic using Julian day; see Table A4.2 for date ranges in each year). Clouds were masked using the QA band and an additional 300 m radius buffer was applied to cloudy pixels to minimize cloud shadows not identified by the QA mask. For the first year of analysis (2000) we only included pixels with forest cover > 90% (Hansen et al., 2013; Hansen data hereafter) in an effort to exclude open canopy forests, regenerating secondary forests, and areas generally not suitable for selective logging concessions that might result in false positives.

At each time step, the pixels that had been identified by the Hansen data as being deforested in that year were removed (i.e. since those areas had been identified as deforestation in that year we did not want to predict logging there). In addition, deforested pixels in the preceding year had one pixel buffer removed from its edges to reduce logging detections associated with deforestation. Finally, regions identified by the Moderate Resolution Imaging Spectrometer (MODIS) monthly burned area product (MCD64A1.006) were removed. Thus, the resulting "valid" pixels that the model could assess if they had been logged were regions that had greater than 90% tree cover in the year 2000, had not been deforested that year (or years prior), and had not burned. These annual mosaics were exported from GEE to be used for predicting the occurrence of selective logging in each year (Section 6.3.1).

6.3 Methods

A diagrammatic overview of the approach from Sections 6.2 and 6.3 can be seen in Figure 6.2.

5.3.1 Building a detection algorithm

We built Random Forest (RF) models using the randomForest package (version 4.6) in program R version 3.5.1 (Liaw and Wiener, 2002; R Development Core Team, 2018). The RF algorithm (Breiman, 2001) is a machine learning technique that uses an ensemble method to identify a response variable (here, whether a pixel was logged or unlogged) given a set of predictor variables. In contrast to a single decision tree, RF models employ multiple, independent decision trees where random subsets of the training data are drawn to construct many trees in parallel. Each tree casts vote on which class should be assigned to the input data. The withheld subset of the data, called the out-of-bag fraction, can be used for validation in the absence of independent validation data (Breiman, 2001).

We randomly allocated 90% of the data for training and withheld 10% for validation. In addition, the training and validation datasets were spatially filtered such that no observations in the training dataset were within 90 m of an observation in the validation dataset. RF models have only two tuning parameters: the number of classification trees to be produced (k), and the number of predictor variables used at each node (m). We used 10-fold cross-validation to identify the number of trees (k = 700) and the number of variables to use at each node (m = 2) that minimized the out-of-bag error rate on the training data.

6.3.2 Algorithm evaluation

6.3.3.1 Selecting the detection threshold

RF models typically use a majority vote to assign an observation to a particular class, with the class that received the most votes being assigned (Breiman 2001). However, the proportion of

votes cast for a particular class from the total set of trees can be obtained for each pixel and a classification threshold can be applied to this proportion (Liaw and Wiener 2002; Hethcoat et al. 2019). We adopted this approach here, wherein the proportion of votes that predicted each observation to be logged, informally term the *probability* a pixel was logged, was used to select the classification threshold. A threshold, T, was defined such that if $probability > T$ the pixel was classified as logged.

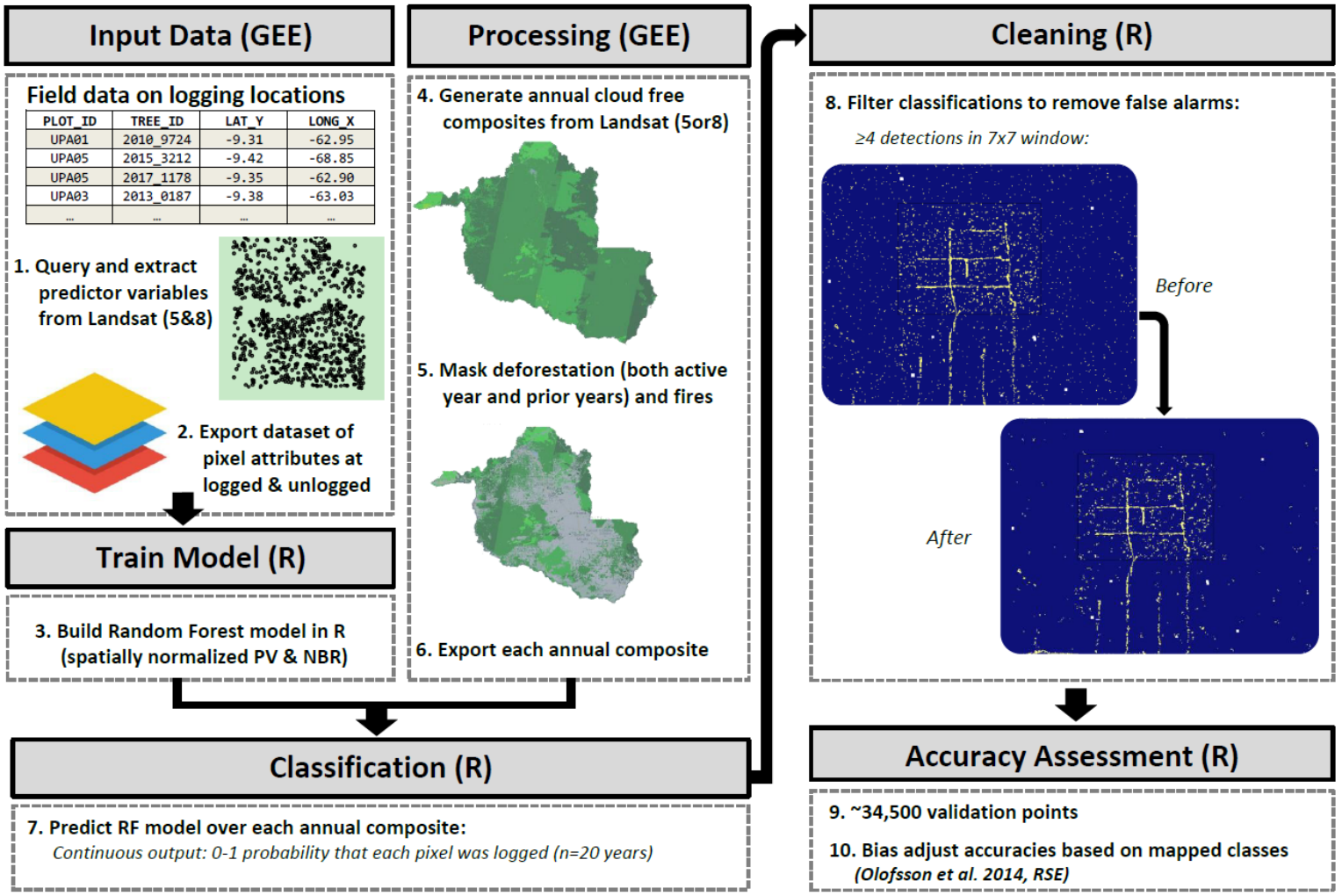


Figure 6.2 Workflow summarizing methods. The platform utilized for each step is in parentheses, with GEE being Google Earth Engine and R being the statistical software developed by R Core Team.

In this case detection of logging involved only two classes, logged and unlogged forest, so the confusion matrix has the form:

		Reference	
		L	UL
Predicted	L	D_L	D_{UL}
	UL	$N_L - D_L$	$N_{UL} - D_{UL}$

where L and UL refer to logged and unlogged classes, N_L and N_{UL} are the numbers of logged and unlogged observations in the reference dataset, and D_L and D_{UL} are the numbers of logged and unlogged pixels detected as logged, respectively. We defined the *detection rate* $DR = D_L/N_L$ and *false alarm rate* $FAR = D_{UL}/N_{UL}$ as the frequency that a logged or unlogged pixel was classified as logged, respectively. Thus, the DR is equivalent to 1 minus the omission error of the logged class and the FAR is the omission error of the unlogged class. In addition, we defined the *false discovery rate* (FDR):

$$\text{FDR} = \frac{D_{UL}}{D_L + D_{UL}} = 1 - \frac{1}{1 + \left(\frac{N_{UL}}{N_L}\right)\left(\frac{\text{FAR}}{\text{DR}}\right)}. \quad (6.3)$$

The FDR is the proportion of all observations that were detected as logged that were actually unlogged, and is equivalent to the commission error of the logged class. The FDR is an assessment of the rate of prediction error (i.e. commission error or type I) when labelling pixels as logged and can be used in detection problems with rare events or unbalanced datasets, such as selectively logged pixels within the Amazon Basin (Benjamini and Hochberg 1995; Neuvial and Roquain 2012; Hethcoat et al. 2019). A high DR and low FDR is clearly desirable, but these cannot be fixed independently in two-class detection problems and both depend on the threshold value. For example, if achieving a 95% detection rate led to a FDR of 50%, then half of all predictions of logging would be incorrect. This level of performance would make estimates of selective logging extremely uncertain. The value of the classification threshold (T) therefore represents a trade-off between true and false detections. In practice, a viable detection method would expect to achieve a DR > 50% while limiting the FDR to 10-20% to have any value for widespread forest monitoring. Model performance was assessed by plotting the DR, FAR and FDR values as T varied from 0 to 1 to facilitate selection of an appropriate threshold for classification.

6.3.3 Predicting selective logging through time

6.3.3.1 Classification and post-processing

After the classification threshold was applied (Section 6.4.1) and a subset of pixels were assigned as logged we applied a post-processing step to remove isolated detections speckled amongst undisturbed forest. We applied a focal window and removed any detection with less than 3 other detections in a 7x7 pixel window. Thus, only detections with at least 3 others in the neighbourhood were considered valid and remained. The window size and number of other detections were chosen through an iterative process of testing different values over the Jamari test region.

6.3.3.2 Assessing map accuracy

Good practices outlined by Olofsson et al. (2014) were used to assess agreement and calculate unbiased error estimates when mapping selective logging detections. We only assessed the accuracies of selective logging and stable forest (i.e. logged and unlogged pixels) and did not consider deforestation and fires, as these have been done elsewhere (Hansen et al. 2013; Turubanova et al. 2018; Giglio et al. 2018).

6.4 Results

6.4.1 Detection algorithm performance against validation data

The performance of the detection algorithm for classifying the validation dataset is shown for all threshold values in Figure 6.3. The false alarm rate rapidly declined as the threshold initially increased and was very low above thresholds of 0.5, suggesting an excellent ability to identify unlogged observations. In addition, the detection rate remained around 50% (after initially decreasing) through a threshold of about 0.6, then declined more rapidly. This suggests about half of the logged observations had very low probabilities and few logged observations had probabilities between 25-50% (Figures A4.2 and A4.3). Thus, the best detection we could obtain, while reducing the false alarm rate to a tolerable level, was approximately 50%. We chose a threshold value of 0.6 for labelling an observation as logged for two reasons. First, the detection rate was close to 50% and the FDR was around 15%; roughly aligning with the minimum level of performance we sought. Second, above thresholds of 0.6 the detection rate dropped quickly with little change in the FDR (Figure 6.3). In general, this suggests our model was relatively conservative at labelling an observation as logged, as we were willing to forego detecting more than half of the selective logging actually present in order to reduce false alarms in unlogged regions. A confusion matrix summarizing model performance at this threshold value can be seen in Table 6.2.

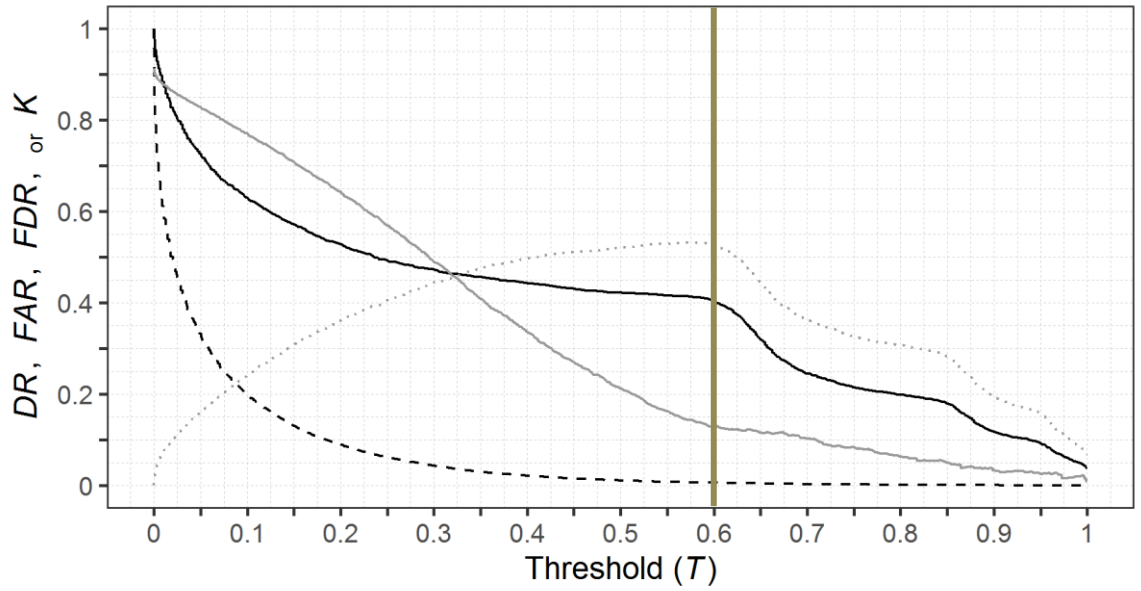


Figure 6.3 Random Forest model performance across the range of threshold values (*T*) for detecting selectively logged observations in the validation dataset. The Detection Rate (DR) and False Alarm Rate (FAR) are the solid and dashed black lines, respectively. Also shown are the corresponding values of the False Discovery Rate (FDR) and Cohen's kappa (*k*) as solid and dotted grey lines, respectively. The vertical line at 0.6 represents the chosen detection threshold value used for classification.

Table 6.2 Confusion matrix summarizing results from Random Forest (RF) model of logged and unlogged observations at four study regions in the Brazilian Amazon. The classification threshold (*T*) for RF models was set to limit FDR to approximately 15%. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

OA: 94.3%				***T* = 0.6**
κ: 0.53		Reference Class		
FDR: 12.9%		Logged	Unlogged	Commission
DR: 40.4%				Error (%)
Predicted Class	Logged	1212	179	12.9
	Unlogged	1785	31450	5.4
	Omission Error (%)	59.6	0.6	

6.4.2 Mapping selective logging through time

We estimate that 41.0% of the State of Rondônia remained undisturbed forest through 2019, with 3.4% having undergone selective logging and 25.7% being deforested between 2000 and 2019 (Figure 6.4). In general, there was little spatial bias associated with logging activities and detections were distributed evenly throughout the state (Figure 6.5). The areas in Figure 6.4 are the sum of all deforestation and selective logging events in a given year. Moreover, once

deforested or identified as having been logged, the pixel retained the first disturbance value (i.e. pixels could not undergo multiple disturbances events and thus be double counted). The amount of selective logging was about twice as high in the first ten years of the period relative to the last ten years. In addition, selective logging rates were generally not well correlated with the rates of deforestation. Thus, knowing the amount of forest loss in a given year would not necessarily inform the amount of selective logging expected (Figure 6.4). Interestingly, the rates of selective logging were lowest in the final couple years of the period, with the exception of 2010 which we believe is slightly underestimated (see Discussion Section 6.5). In addition, the deforestation data from 2019 were unavailable at the time of analyses and so some of the selective logging detections in that year likely occurred in deforested areas that would have been removed (i.e. the rates for 2019 are probably a little inflated).

Many of the detections were obviously logging road networks (both main access roads and smaller internal roads) that generally go undetected by the Hansen dataset (Figure 6.6). In addition, there are countless instances where detections preceded deforestation by a year or two (Figure 6.7), demonstrated by logging detections occurring inside areas later identified as deforested (i.e. subtle forest disturbances preceding total clearance was detected). These instances are the only cases when pixels could get included in calculations of logging and deforestation (i.e. double counted in Figure 6.4), but only if the logging detections were in years preceding deforestation. This is because once a pixel was identified as having been deforested it was removed from consideration for logging.

Finally, we explore the results the scale of a forest management unit at a location where we have general knowledge of logging, but limited field data. First, a forest management unit that was selectively logged in 2018, but where we did not have any data on logging locations, shows some false detections (in the colors preceding 2018), but both the year of logging are correctly identifiable and the internal logging road construction in 2015 was accurately detected (Figure 6.8). Next, the number of false alarms over an area known to have remained unlogged (a forest reserve area associated with the logging concession) a false alarm rate of approximately 2% (Figure 6.9). These results bolster confidence in the estimates of selective logging and demonstrate their effectiveness.

The bias adjusted confusion matrix, summarizing errors for the proportions of mapped classes (Olofsson et al. 2014), is shown in Table 6.3 and is generally consistent with the results from the validation data. These findings reiterate a 55% omission of logging detections and a 13% FDR (i.e. 13% commission error when predicting logged observations). Consequently, our estimates of selective logging should be viewed as conservative, particularly in light of the fact that we excluded areas considered deforested by the Hansen data. Indeed, in some cases regions known to be selectively logged were identified as deforestation by the Hansen data (Figures

6.10 and A4.4), limiting our ability to detect them that year. Thus, the annual amounts of permitted, legal selective logging may be closer to double what is reported here.

Consistently, about 55% (±8% SD) of selective logging detections were within 1km of deforestation activities occurring in the same year (Figure 6.11). Thus, the majority of selective logging activities in Rondônia occurred in close proximity to deforestation presently detectable through the weekly Global Land Analysis & Discovery alerts system (Hansen et al. 2016). This result is in line with the well documented cycle involving selective logging as a driver of and precursor to land clearance (Curtis et al. 2018).

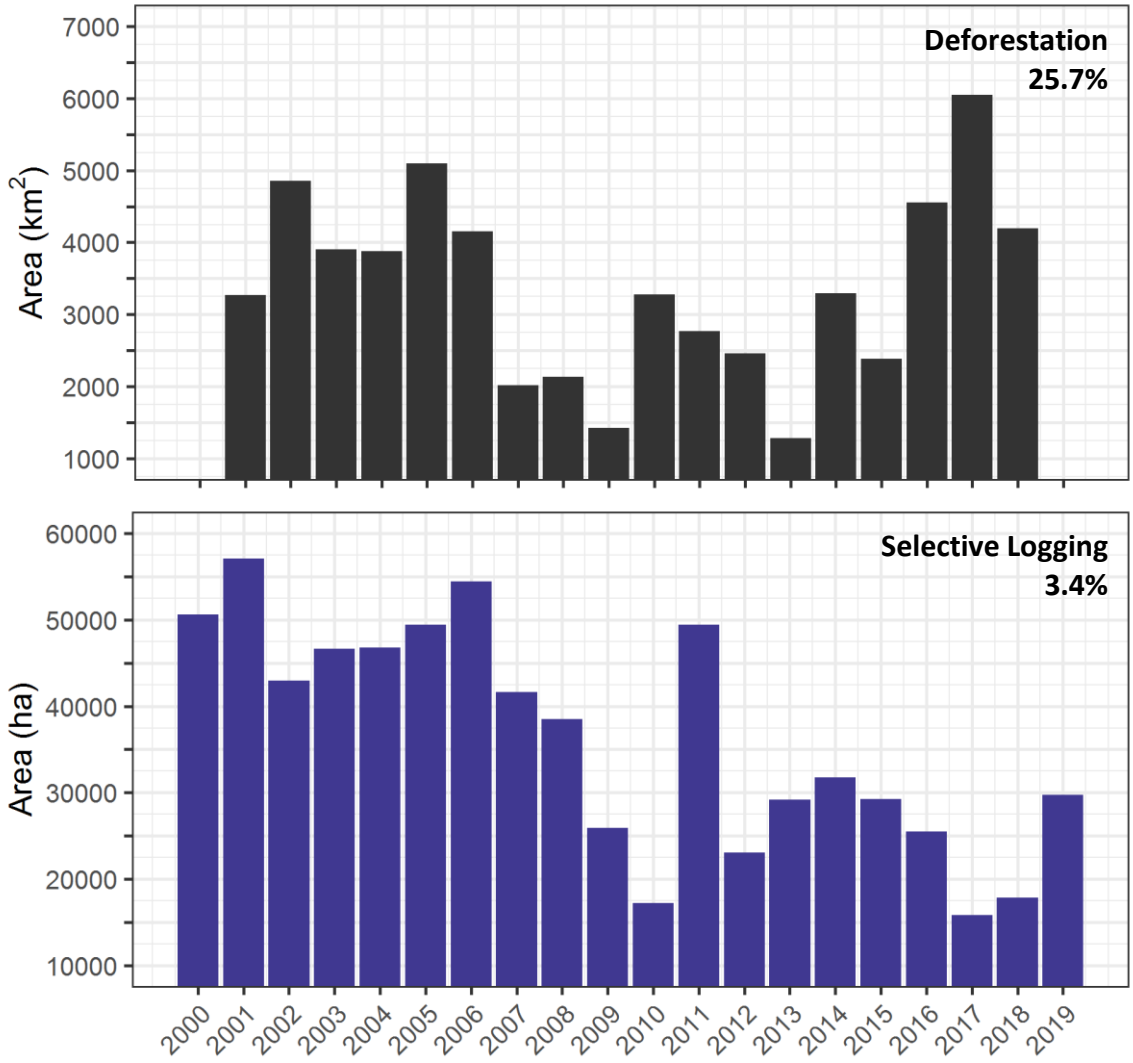


Figure 6.4 Annual amount of Rondônian forest affected by deforestation (from Hansen et al. 2013) and selectively logging (this study). Note the deforestation data from 2019 were unavailable at the time of analyses.

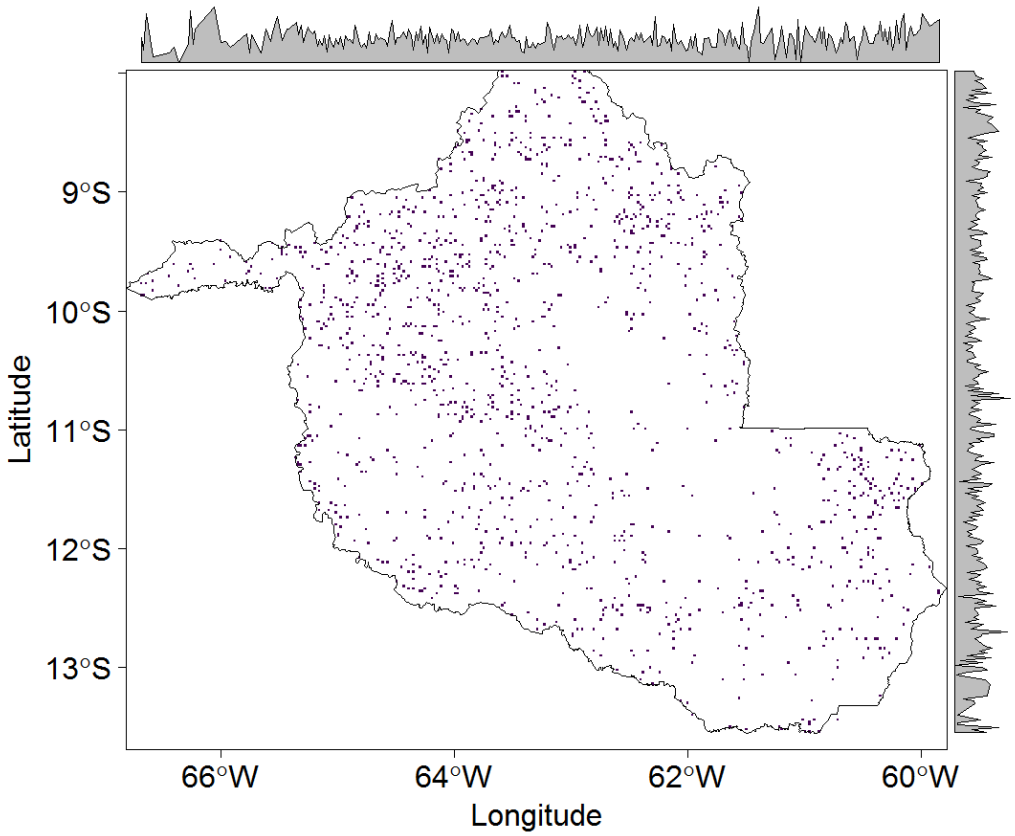


Figure 6.5 Spatial distribution of selective logging detections throughout Rondônia (density plots along outside axes). Note, the 30m scale of the detections (in purple) was distorted in the process of plot rendering and the pixels shown here should not be interpreted as either comprehensive or precise. They were left in only to give the impression of very general locations of detections (see Figures 6.6-6.9 for a detailed view of detections).

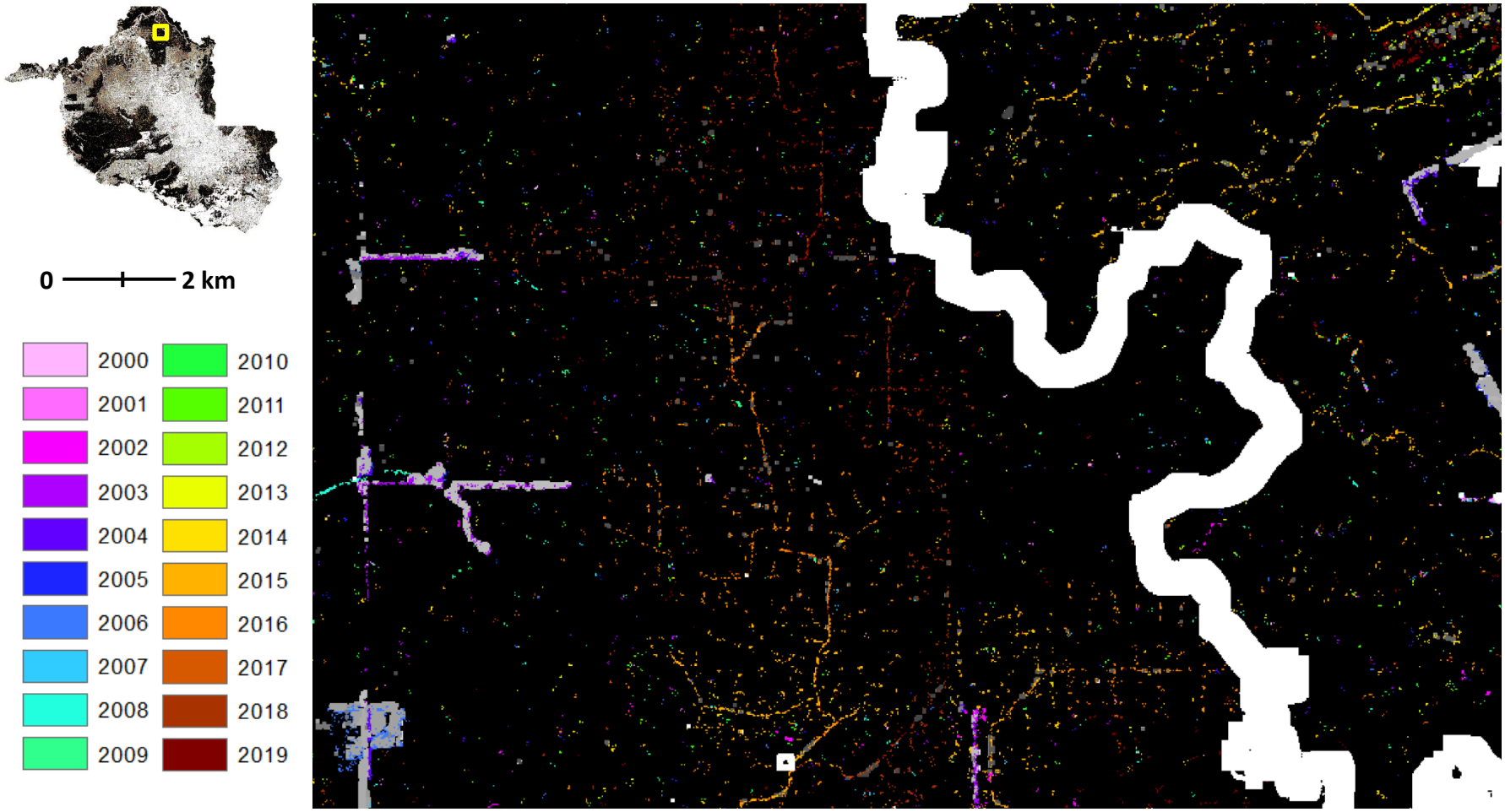


Figure 6.6 Example region showing selective logging road networks, with stable forest in black, Hansen forest losses in grey shades, and the Preto River in white. The map is centred on 62.875 W, 8.478 S.

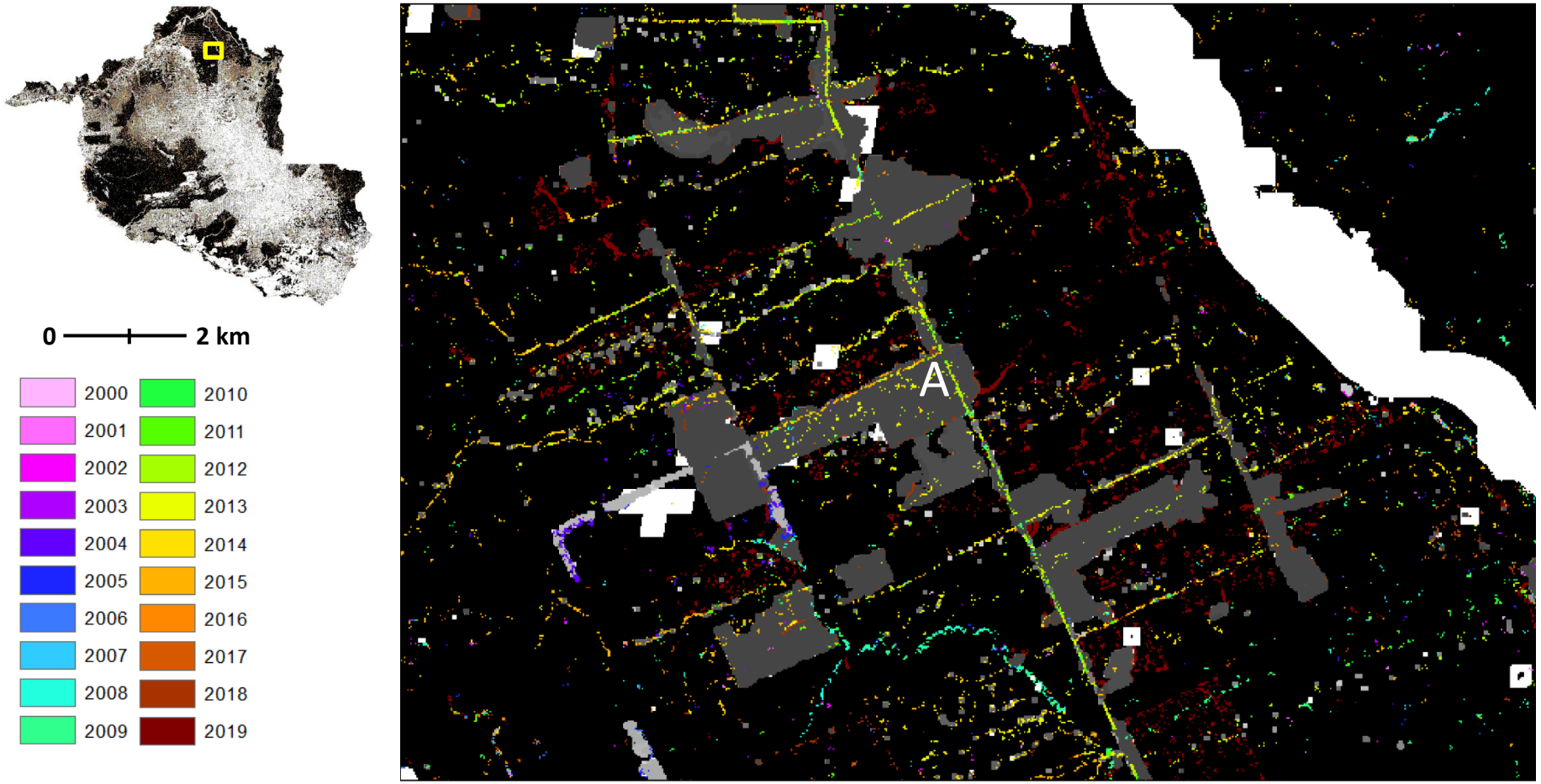


Figure 6.7 Example region showing early detection of deforestation. The expansion of roads and early forest disturbances (A, in green-yellow-orange colors) were detected before the deforestation events occurred, hence they are on top of the forest loss layer from Hansen (in grey shades). Stable forest is in black, burned areas are white squares, and the Jiparaná River is the in top right in white. The map is centred on 62.722 W, 8.410 S.

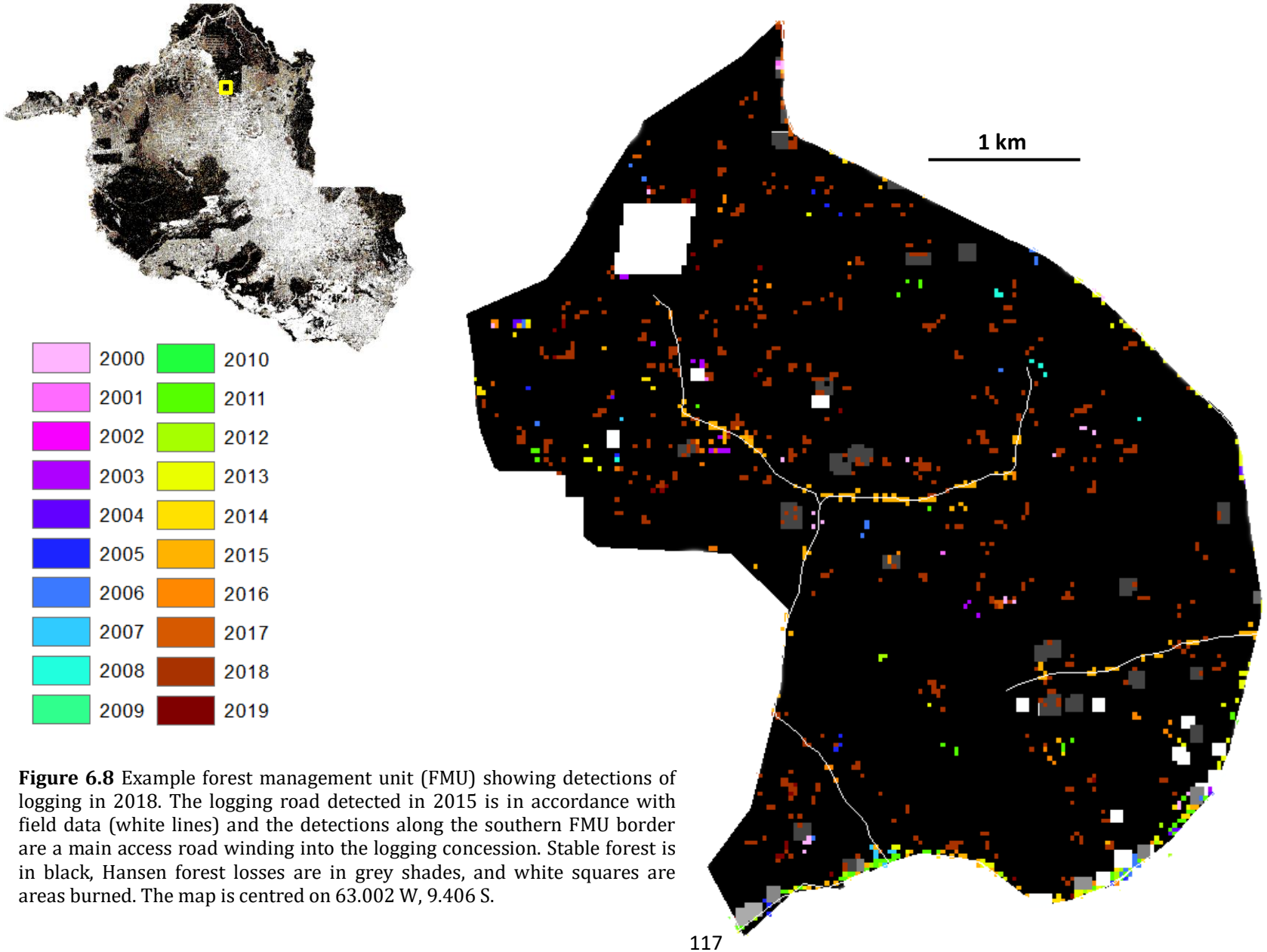


Figure 6.8 Example forest management unit (FMU) showing detections of logging in 2018. The logging road detected in 2015 is in accordance with field data (white lines) and the detections along the southern FMU border are a main access road winding into the logging concession. Stable forest is in black, Hansen forest losses are in grey shades, and white squares are areas burned. The map is centred on 63.002 W, 9.406 S.

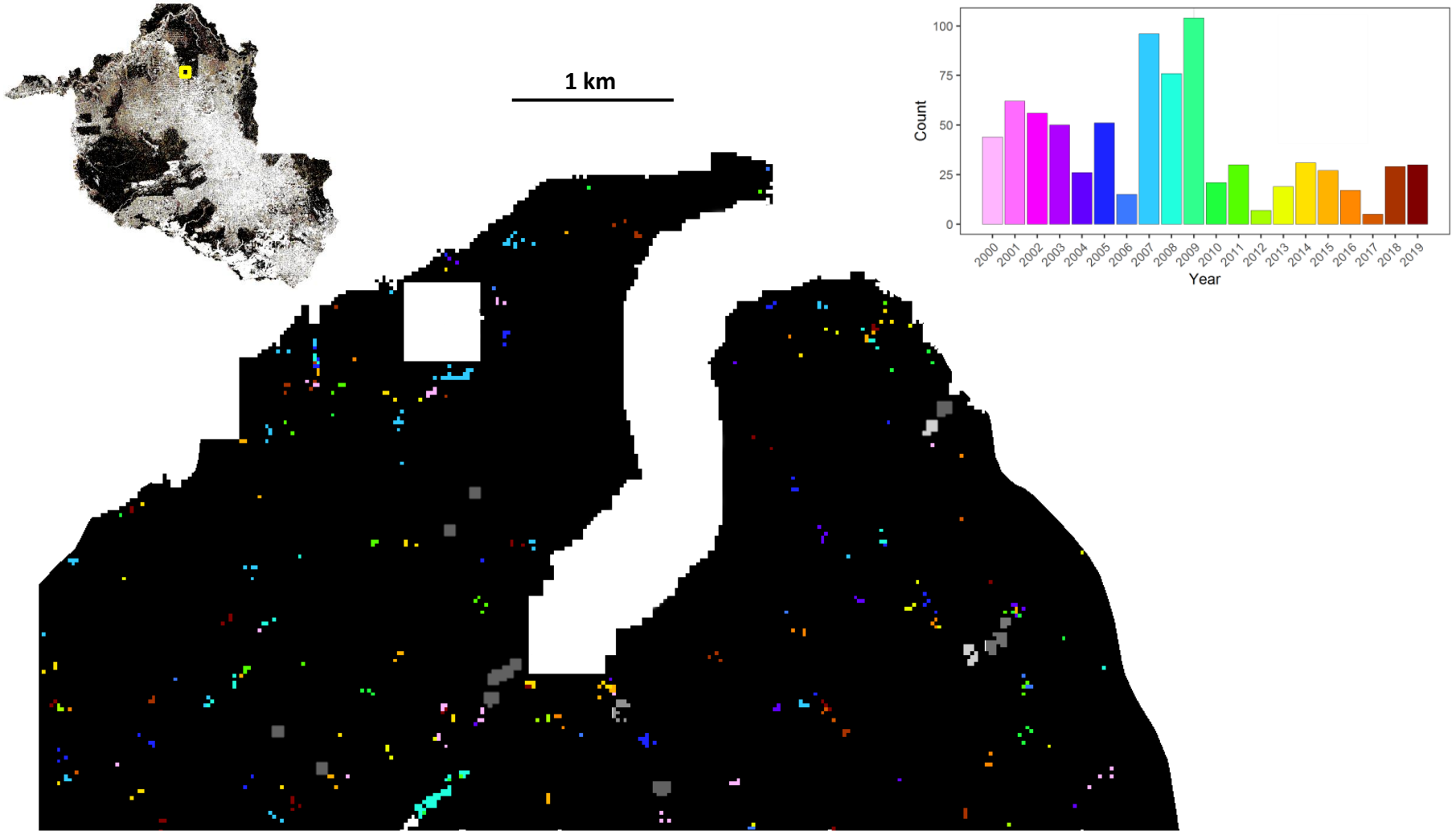


Figure 6.9 Example forest reserve area (i.e. unlogged forest) inside a logging concession in the Jamari National Forest showing false detections. Stable forest is in black, Hansen forest losses are in grey shades, and white areas are burned forest and water. Only 2.3% of pixels (n=796) are false alarms within the reserve over the 20 year period. The map is centred on 63.022 W, 9.266 S.

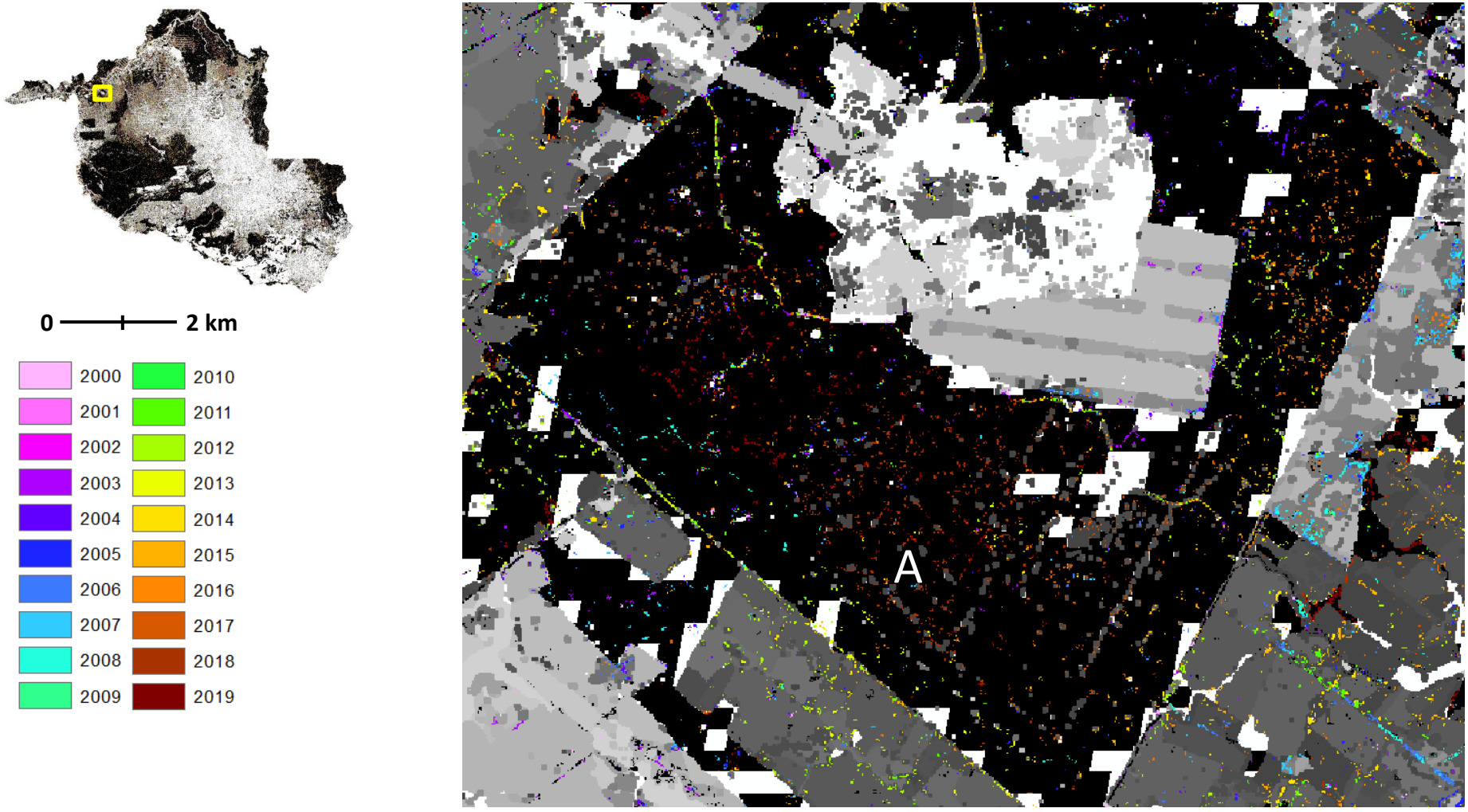


Figure 6.10 Example region showing deforestation detections by the Hansen dataset (in grey shades) that resulted from selective logging activities (A). These areas were excluded prior to prediction and thus limited our ability to fully map selective logging. Stable forest is in black and burned areas are in white. The map is centred on 64.825W, 9.665 S.

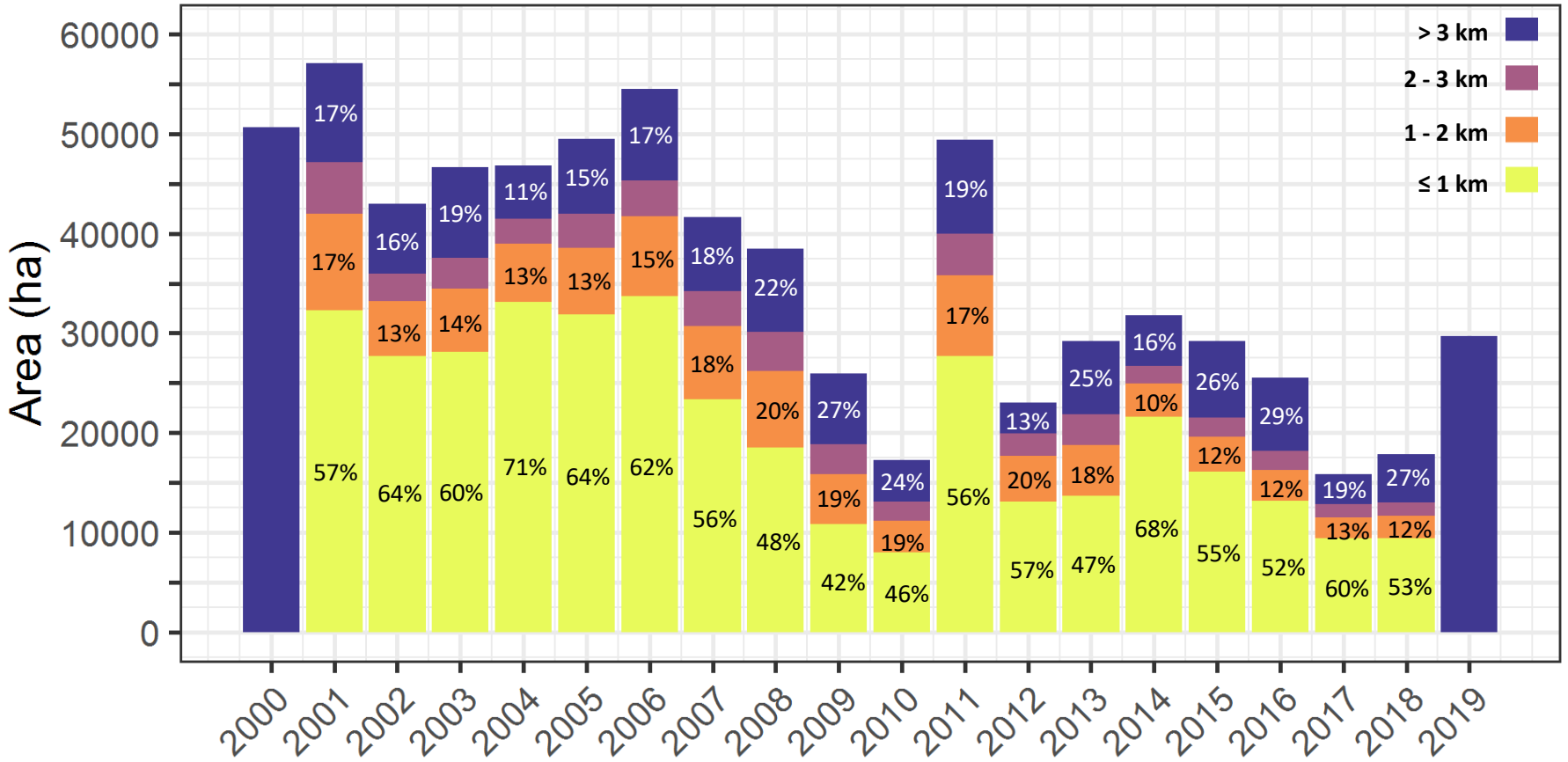


Figure 6.11 Selectively logging detections over four distance categories from deforestation activities in the same year.

Table 6.3 Confusion matrix summarizing unbiased (Olofsson et al., 2014) area estimates from mapping logged and unlogged pixels in Rondônia, Brazil. The classification threshold (*T*) for RF models was set during model training to limit FDR to approximately 15%. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

OA: 94.0%				***T* = 0.6**
κ: 0.64		Reference Class		
FDR: 12.9%		Logged	Unlogged	Commission
DR: 44.5%				Error (%)
Predicted Class	Logged	0.06	0.01	12.9
	Unlogged	0.05	0.88	5.4
	Omission Error (%)	45.5	1.0	

6.5 Discussion

We have demonstrated the scalability of the approach outlined by Hethcoat et al. (2019) to map tropical selective logging with Landsat data, advancing beyond the scale of a logging concession or forest management plan to regional-scale assessments of logging activities with historical data. Achieving this required changes to the original methodology detailed therein, moving away from surface reflectance values and utilizing a spatial normalization step to ameliorate abrupt shifts in pixel values resulting from varying solar illumination and atmospheric conditions in image mosaics. In doing so, however, we have further demonstrated how invaluable Landsat data are, and will continue to be, for monitoring Earth's forests. Advancements in data, tools, and techniques continue to enable greater abilities to monitor forests from space and offer great hope for conservation, yet they increasingly inform us of the extent of the damages already realized.

Two decisions fundamentally affected estimates of selective logging; the value of the classification threshold (*T*) and the values associated with the window size and the number of additional detections needed in the post-processing routine. At the time of writing the sensitivity of the results to changes in these values is being explored. In particular, we seek to decrease the omission of logging through lowering the threshold and/or altering the window size and detection requirements in the post-processing step. However a detailed understanding of how these changes would impact FDR is needed and these analyses are currently ongoing. In addition to the methodological decisions mentioned, an additional choice made from the outset was to exclude forests with canopy cover <90%, as defined within the Hansen data. Brazil defines a forest as having >10% canopy cover and >5 m height (GFOI 2016). However, we sought to restrict our analyses to continuous tropical forests (i.e. not secondary forest, cerrado, gallery forests, or otherwise modified forests) as best as we could. The Amazon Deforestation

Satellite Monitoring Project (PRODES) dataset begins in 2001 and could not be used for removing secondary forests, plantations, or areas previously deforested and predicting over older secondary forest (e.g. >15 years) in some cases could not be completely avoided. However, for financial viability reasons, most commercial logging leases occur in relatively intact tropical forest this likely did not affect our estimates greatly.

In general, our numbers on annual logging rates are almost certainly underestimates for a few key reasons. First, as stated in the omission error rates, about half of the logging was actually detected in a given year (Table 6.3). In addition, while our method makes predictions on the scale of the individual pixel, forest disturbances from selective logging affect patches of forest, not isolated pixels. Thus, the amount of intact forest within a selectively logged FMU can vary substantially from 50-75% (Putz et al. 2019), despite the proportion of pixels having had a tree removed being closer to 10%. Some have utilized a buffer (often 180 m) around logging road networks or landing decks (Souza and Barreto 2000; Monteiro et al. 2003; Matricardi et al. 2010), yet the authors have acknowledge high commission and omission errors associated with this approach. We welcome a renewed discussion from the community on best practices for estimating area affected by selective logging in the absence of extensive field data.

Two particular years from our analyses need further discussion. The selective logging detections from 2010 are almost certainly an underestimate and the detections from 2011 are likely an overestimate (relative to other years and not in the manner previously discussed above). Two factors working concurrently are believed to be impacting the predictions from those years. First, the cloud free window was more limited in 2010 and the time frame associated with the dry season mosaic was earlier and narrower than most other years (Table A4.2). The cloudiness of 2010 has been documented in other forest mapping exercises in the Brazilian Amazon recently (Qin et al. 2019). On its own this would have resulted in fewer detections, because the dry season mosaic was about three weeks narrower than average and fewer pixels would have been logged at that point in the season. Second, 2010 was a particularly high fire year within the Amazon (Aragão et al. 2018), consequently large regions were excluded from our analyses that probably coincided with some logging detections (Figure 6.12). Very much related, logging detections increased dramatically in 2011 (Figure 6.4), likely the result of delayed detection of logging activities missed in 2010 (i.e. showing up a year later), combined with additional detections from the fire scars from 2010 that were insufficiently mapped by the MODIS burned area product. While these idiosyncrasies might affect an annual estimate of logging, these kinds of anomalies would be dampened in an operationalized product that utilized the 5-year rolling average under reference level reporting for REDD+ (GFOI 2016).

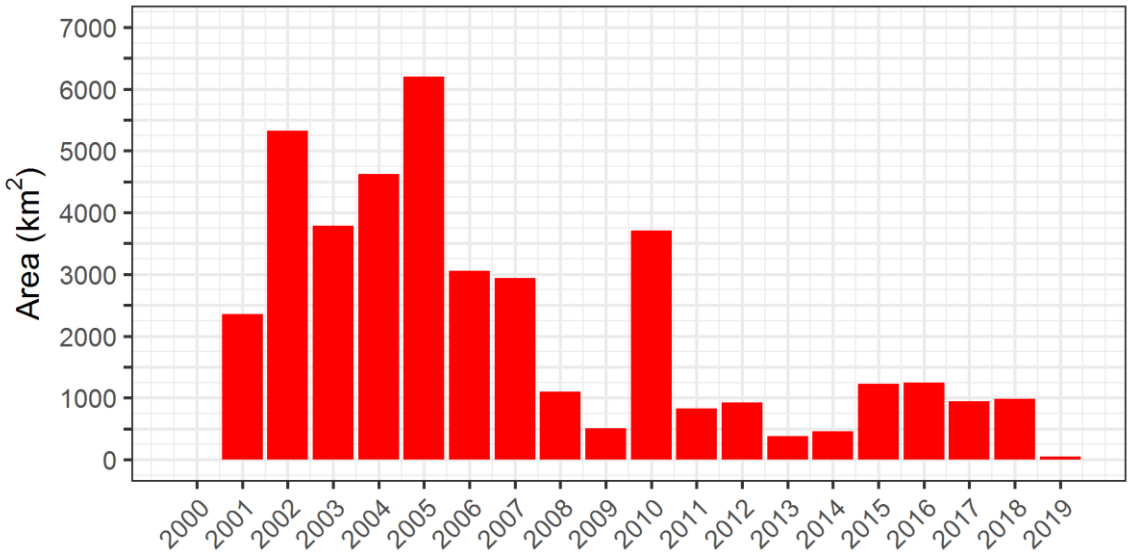


Figure 6.12 Annual amount of Rondônia burned, calculated by the MODIS MCD64A1.006 product. Note the MCD64A1 data are incomplete for 2000 and 2019, as the product became operational in late 2000 and the archives were only queried through September 2019. These summaries include all fires (i.e. in non-forest and prior deforested lands). In addition, fires in previously burned areas were included in tallies and thus represent an annual amount of total burned area.

A comparison between our results and others is difficult, as detection of selective logging exclusively has not been done before. However, our estimates are generally higher than other work quantifying degradation within Rondônia. For example, the only other studies to assess degradation over a similar time period combined all forms of degradation (Souza et al. 2013; Bullock et al. 2018). Souza et al. (2013) estimated about 500,000 ha of annual degradation within the whole of the Amazon from 2001- 2010 (but twice that amount in 2008), with roughly ~7% occurring in Rondônia (35,000 ha annually). Bullock et al. (2019) estimated roughly 50,000 ha annually from 2000-2005 and >75,000 ha annually from 2006-2013 within Rondônia. Our estimates are closer to those from Bullock et al. (2019) and the total area selectively logged over the period (3.4%) is about half of the 6% they found for degradation (logging and fire). Yet, our estimate of about 40,000 ha of logging annually from 2000-2010 occupies a large fraction of both their estimates of degradation, particularly given what we know about fire history in Rondônia over the period (Figure 6.12). However neither study had access to selective logging data specifically for training and validation of their results. In addition, if we breakdown the 1% omission error of unlogged forest (Table 6.3), this suggests about 97,000 ha of unlogged forest were identified as logged over the 20 year period (i.e. <2,000 ha per year). Thus, our estimates likely reflect a demonstration of the advancement in detection our method has enabled and do not suggest they are erroneously inflated.

One of the immediately noticeable aspects in the detections of selective logging was the abundance of linear features within forest tracks (i.e. logging roads). Road building has big implications for primary tropical forests (Kleinschroth et al. 2015, 2016; Kleinschroth and Healey 2017) and improving their detection is critical to our understanding of their tenure and the continued loss of intact forest landscapes (Potapov et al. 2008, 2017). Roads create forest edges that can alter abiotic processes like microclimate (Williams-Linera et al. 1998), change plant and animal species composition (Tabarelli et al. 2012), increase fire susceptibility (Armenteras et al. 2013), and ultimately weaken forest resilience (Murcia 1995; Kleinschroth and Healey 2017). Moreover recent work has shown that tropical forests globally may be nearing a tipping point where fragmentation will begin to dramatically increase (Taubert et al. 2018). The tropics are thought to have around 50 million forest fragments, encompassing nearly 50 million km of edge (Brinck et al. 2017). Monitoring the emergence and spread of roads is therefore critical to understanding the disturbance frontiers of intact forests globally.

It is important to highlight some caveats regarding our approach and the results. First, like all studies in the tropics that use exclusively optical data, some areas were excluded from analyses each year because of clouds. Despite creating a mosaic of all available pixels in each year, approximately 260,000 ha annually (± 240,000 ha SD; min: 5,600 ha; max: 893,000 ha; ~1% of Rondônia) was impacted by clouds in a given year and were re-included in the subsequent year assuming no disturbance had occurred. Second, each mosaic was made up of only a single pixel per location and was tantamount to a single image analysis. While the use of spatial normalization enabled us to overcome the weaknesses associated with vary solar illumination and atmospheric conditions, any selective logging that might have occurred after the date of the cloud-free pixel in the mosaic was excluded, further limiting our ability to detect logging. Third, there was generally an inability for our approach to distinguish between logging and fire. We limited this by removing burned areas annually, in the MCD64 burn product, yet the difference between the scale of those datasets (500 m) and Landsat (30 m) is certain to result in commission and omission of burned area removal. Collectively, most of the caveats further suggest we underestimated the amount of selective logging annually.

6.6 Conclusion

Globally, only 25% of forests are considered undisturbed (i.e. primary forest), with the remaining 75% being described as “other naturally regenerated forest” (FAO, 2016). Large-scale monitoring of forest degradation remains an elusive goal for supporting REDD+ initiatives and reporting country-level contributions toward emissions reductions (GFOI 2016). While sustainable management of forest that result in no longer-term loss of carbon stocks (through less destructive harvest techniques like reduced-impact logging) are not considered

degradation in the context of REDD+, there is clearly a need to take account of and adequately monitor extractive practices that could result in reduced carbon stocks. Selective logging is often the initial anthropogenic disturbance event to impact primary forests. The opening of road networks and improved access to regions once heavily forested can facilitate additional activities associated with degradation (e.g. fuel wood removal, spread of invasive species, illegal logging, mining, and fires) or forest clearance for pastures, agriculture, or human settlements. Our approach is step in this direction, towards an operationalized selective logging monitoring system capable of detecting subtle forest disturbances over large spatial scale.

Acknowledgements

MGH was funded by the Grantham Centre for Sustainable Futures. JMBC was funded by the Natural Environment Research Council (Agreement PR140015 between NERC and the National Centre for Earth Observation). We would like to thank the Google Earth Engine developers and community for access to the datasets that made this work possible.

Chapter 7

Synthesis and Discussion

7.1 Primary findings

This thesis has advanced our understanding of the capabilities and limitations optical and SAR data possess for large-scale monitoring of selective logging from space. We have demonstrated that Landsat data can be used to detect forest degradation and will continue to be an invaluable resource for forest monitoring applications globally. In contrast, SAR data (C-band Sentinel-1, RADARSAT-2 and L-band PALSAR-2) was shown to be ineffective for monitoring selective logging within a classification framework. However, time series analyses with Sentinel-1 displayed a change in the SAR signal with the onset of logging and likely offer a way of detecting high intensity commercial logging with the Amazon and potentially beyond.

Improving current abilities to detect and map tropical selective logging is essential for understanding the impacts on global biodiversity and tracking and mitigating the climate implications of forest degradation. Yet, large uncertainties remain in understanding the true impacts of selective logging because the advances in detection and monitoring at large spatial scales are only just emerging (Hethcoat et al. 2019). Progress in detection of selective logging in the tropics would enable the mapping of primary forest as well as identify logged regions that possess high conservation value. In addition, quantifying the extent of forest degradation from selective logging is a key step in refining our understanding of the terrestrial portion of the carbon budget (Baccini et al. 2017; Mitchard 2018; Le Quéré et al. 2018). Below, I summarize and discuss the key findings of this thesis within the context of the development of regional- to pan-tropical scale logging maps. Specifically, I discuss how this information can inform assessments of carbon losses and biodiversity impacts from selective logging at scale. I go on to outline the application of these results and the further work needed to develop an operationalized selective logging product and consider the potential for the development of a near real-time monitoring system.

Landsat data has previously been assumed to be too coarse to monitor selective logging in the absence of distinct spectral features like networks of logging roads and landing decks. We have shown that surface reflectance data can be used in a supervised machine learning framework to detect subtle spectral changes to forests from selective logging at low intensities. In Chapter 3, we begin on the scale of a logging concession and develop an algorithm that demonstrates high detection of logging (around 90%), with about 40% of the area inside logging concessions being identified as logged. We show that spatial-contextual information (from Grey Level Co-occurrence Matrix (GLCM) texture measures) was crucial in improving detection of logging. Unlike previously developed methods, our approach was able to make predictions about logging on a single image of a region, instead of requiring successive cloud-free acquisitions to compare changes in pixel values (Asner et al. 2005, 2009; Souza et al. 2013). This is particularly important in the tropics, because a single, low-cloud acquisition is

sometimes the best you get in a region (see Souza, Jr et al. 2013). Thus, our approach has the potential to significantly increase current abilities to detect and monitor selective logging activities that up to now have been, at best, marginally detectable.

In Chapters 4 and 5 we aim to extend this methodology to include SAR data, but with limited success. In Chapter 4 we present the first multi-sensor comparison of SAR data for monitoring a range of selective logging intensities in the tropics. We demonstrated that L-band PALSAR-2, C-band RADARSAT-2, and C-band Sentinel-1 data performed inadequately at detecting tropical selective logging when using pixel-based attributes for classification, even when using only the most intensely logged data. Indeed, when SAR data was used in combination with optical data, classification performance was lower than results with optical data alone (Chapter 5). However, we found that about 50% of the pixels in forest management units logged at the upper end of intensities with the Legal Amazon (20-30 m^3 ha^{-1}) showed a clear breakpoint in their time series of Sentinel-1 data (Chapter 4). Moreover, the breakpoints generally coincided with the timing of active logging. While these results are not a detection algorithm as such, since the accuracy of this approach was crude and relied on identifying regions of accumulated detections, they do represent a first demonstration of the potential. The timing of our logging data and the time series available precluded a longer time series analysis, but these results should stimulate more research into tracking Sentinel-1 data for moniroting intensive commercial harvest. Another important takeaway from our results in Chapters 4 and 5 was the finding that the co-polarized channel (i.e. VV or HH) was an important variable in classification. This is in contrast with the generally accepted view that the cross polarized channel is best for detecting changes in forest cover (Joshi et al., 2016; Reiche et al., 2018a; Ryan et al., 2012; Shimada et al., 2014). Our results support the idea that the co-polarized channel (for L- and C- band SAR) is useful and should not be ignored in forest disturbance detection analyses (e.g. Reiche et al., 2018a).

On the scale of a Brazilian state, or indeed the Amazon basin, the methodology developed in Chapter 3 was intractable. Image mosaics, composed of many satellite acquisitions merged together, are required to monitor forests at regional scales. Varying solar illumination and atmospheric effects impact the quality of each scene and, when combined, can result in abrupt changes in pixel values that can confuse an algorithm resulting in a cloud shadow being labelled as forest disturbance. Informed with what worked in Chapter 3 we amended the methodology to reduce atmospheric aberrations (by using spectral unmixing) and modified the way spatial context each pixel was included (i.e. the spatial referencing rather than GLCMs). In general estimates of logging were more conservative in Chapter 6 compared to Chapter 3. The sacrifice for working at scale was some loss of power in detection; though we could have detected more but would have had more false alarms and sought to avoid this. Critically, we

have produced a general methodology that is capable of detecting selective logging over large spatial and temporal scales and demonstrated its effectiveness.

7.2 Implications

Tropical forests store billions of tons of carbon. While the emissions estimates from selective logging are much lower than those from deforestation (Asner et al. 2010), recent work has shown that taking full accounting of degradation activities suggests much higher emissions than previously thought (Maxwell et al. 2019). However, Maxwell et al. (2019) simulated selective logging in proximity to road networks, because we just do not have the ability to estimate the full extent of selective logging globally. Maps of forest degradation from selective logging would enable the identification of regions suitable for inclusion in REDD+ initiatives and better accounting of the climate implications. In addition, RIL techniques that emphasize reduce carbon emissions (RIL-C) have been suggested as a tool to incentivize voluntary carbon markets (Ellis et al. 2019). However a lack of emissions verification systems, because of the difficulty in monitoring logging, has limited the adoption of these practices (Ellis et al. 2019). While the UN anticipates that payments to nations under REDD+ initiatives, could reach $30 billion annually (Phelps et al., 2010), large-scale monitoring of forest degradation remains an elusive goal for supporting REDD+ initiatives and reporting country-level contributions toward emissions reductions (GFOI 2016).

Logged tropical forests have been shown to restore much of their aboveground carbon in under 20 years (West et al. 2014). While forest management practices that result in no longer-term loss of carbon stocks are not considered degradation in the context of REDD+, others have suggested that current logging practices touted as sustainable are driving economically viable species to extinction and never fully recover (Richardson and Peres 2016). In addition, it has been shown that carbon-centered conservation strategies are likely to leave behind the most diverse tropical forests (Ferreira et al. 2018). More specifically, Ferreira et al. (2018) found that while biodiversity and carbon were positively associated in highly degraded and secondary forests, intact forests showed no relationship between carbon and biodiversity. Reliable logging maps would enable a better accounting of the relationships between timber harvest and the full suite of goods and services tropical forests provide.

The impacts of selective logging on biodiversity have been the focus of countless studies over the years, with the vast majority showing negative impacts for most taxa (Gibson et al. 2011; Burivalova et al. 2014). Improved abilities to map selective logging are critically important for projecting biodiversity losses from forest disturbances across scales. In addition, the identification of selectively logged forests with high conservation value falls under the umbrella of "+" activities that address biodiversity and development goals within REDD+ (Goetz

et al. 2015). Moreover, recent work has shown that when deforestation and forest degradation are studied together, their effects have mainly been examined in isolation of one another over the years, biodiversity losses are much higher (Barlow et al. 2016).

Selective logging is not a binary treatment. Logging represents a gradient of disturbance most closely linked to the intensity of harvest (Putz et al. 2019). Indeed, a global meta-analysis of logging impacts on biodiversity found intensity was the greatest predictor of species richness, after accounting for taxa, with a halving of species richness (or more) at logging intensities around 50 m^3 ha^{-1} (Burivalova et al. 2014). However, they also found that for most taxa, species richness was relatively stable at logging intensities <10 m^3 ha^{-1} (Burivalova et al. 2014). The methodologies developed in this thesis have shown the capability of detecting logging at these lower intensities (Chapters 3 and 6). Estimates of the amount of logging activities globally at these intensities are currently unavailable, however, intensities are generally lower in Africa and Latin America than Asia (Sist 2000; Putz et al. 2001). Yet, a dataset of over 800 logging permits from the state of Pará, Brazil found only 15 were for harvest intensities < 10 m^3 ha^{-1} (Richardson and Peres 2016). While low intensity permits probably represent a minority of legitimate, sustainable harvest practices, it has been estimated that some 50-90% of tropical timber on the international market is harvested illegally and at very low intensities (Kleinschmit et al. 2016; Brancalion et al. 2018). The extent to which our methodologies are capable of detecting illegal logging remains to be seen, but they will almost certainly depend upon the destructive quality and spatial extent of the operation (Brancalion et al. 2018). In general, however, the lower limit of logging intensities detectable with the methodology we present is not well understood and deserves further exploration.

7.3 Limitations

Our method was not able to take account of logging intensity. We examined and discussed the relationships between logging intensity and the likelihood a pixel was labelled as logged in a number of places throughout the thesis, but we did not estimate a logging intensity for a given pixel explicitly. Random Forest models can be used in a regression framework (as opposed to classification) and, technically, pixel attributes could be used to develop a model relating specific pixel values to logging intensities. We chose not to take this approach, because initially we only had access to a limited dataset on logging that was primarily composed of lower intensities (Chapter 3). As a consequence, relating logging intensity to pixel attributes gets reduced to whether large trees can be detected being removed, since a large proportion of the volume comes from a smaller number of big trees. In addition, the spatial resolution of Landsat data is likely too coarse at those lower logging intensities and sufficient variation in intensity just does not exist within a management unit. With the expanded datasets we utilized in later

chapters, however, such an approach might be possible and we are keen to explore this potential in future work.

Throughout the thesis I discuss the tradeoffs associated with detection and false alarms in identifying selective logging. Conceptually this is identical to the specificity-sensitivity tradeoff often exhibited in medical diagnostic tests and forms the basis of the receiver-operator characteristic (ROC) analysis. However, the detection, false alarm, specificity, and sensitivity measures are independent of changes in the prevalence of particular classes (e.g. the number of logged and unlogged pixels within a landscape) and they receive equal weight in balancing the trade-off in ROC analyses (i.e. assumed to be of equal importance for classification). Ignoring class imbalances inherent within data on selective logging would not offer the best approach, as the amount of unlogged forest in the Amazon likely dwarfs the locations selectively logged; certainly within many of our datasets this was true. Consequently, we utilized an approach meant to limit the rate of type-I error, because we cared more about erroneously labelling truly unlogged pixels as having been logged (i.e. d_{pL} from Chapter 3 and FDR in Chapters 4-6). Fixing this measure, the rate of commission error when predicting data as logged, enables the user to specific a level of prediction error deemed acceptable for a given application, while balancing the trade-offs in detection and false alarm. However, the use of row statistics (i.e. commission errors) within a confusion matrix depends on the prevalence of particular classes on the landscape. Thus fixing an FDR for one particular dataset and application does not mean the same FDR on another dataset will yield comparable rates of detection.

We did not utilize optical data from the Sentinel-2 program, despite some overlap in the logging datasets (Sentinel-2 starts in mid-2015). The increased spatial resolution of the Blue, Green, and Red channels (10 m relative to 30 m Landsat data) in the Sentinel-2 program likely offer increased detection abilities. However, we were interested in initially taking an historical perspective on logging, with the knowledge of the planned continuation mission and commitments to harmonizing data products across future Landsat missions (Loveland and Dwyer 2012; Wulder et al. 2019). Moreover, increased spatial resolution is not necessarily the silver bullet approach to building a large-scale detection system. First, geolocation data under dense canopy often have a 3-5m error associated with their position. Second, the computational burden is 9-fold for 10 m over 30 m spatial resolution, however, increased access to high performance systems like Google Earth Engine have helped. The Sentinel-2 program is a constellation of 2 satellites separated by 180°, enabling image acquisition every 5 days. This feature will greatly improve global forest monitoring and help with advancing better near-real-time systems. Future work on near-real-time logging detection systems should incorporate Sentienl-2 for individual, given Landsat's revisit period is about 16 days in the tropics, but also for use in combination with other optical datasets. For example, incorporating imagery from

Landsat 9, Sentinel-2, and Landsat 9 (2020) will provide a re-v-sit period of around 2-days (Wulder et al. 2019).

7.4 Conclusion

Tropical selective logging is fundamentally connected to global climate, biodiversity conservation, and human wellbeing. Improving land use maps of logging will enable a better understanding of biodiversity impacts, help refine the role tropical forest play in the carbon budget, and assist in prioritizing conservation efforts aimed at achieving biodiversity and carbon co-benefits. We have shown that logging records can be used to build a detection algorithm for monitoring selective logging activities at regional scales. In order to be effective our results show images acquired during, or very soon after, active logging are needed to map selective logging. This is partly because logging activities typically occur in the dry season when cloud-free imagery is more likely to be available, but also because the spectral changes associated with selective logging practices are subtle and short-lived and rapidly become obscured under even limited regrowth (Broadbent et al. 2006). Access to similar datasets could immediately fold into the analyses presented here and, potentially, further improve detection. Selective logging within national forests, or otherwise government controlled lands, represents a use of collectively owned resources and those datasets are part of a national heritage and should be available to promote the benefit of the national interest. The extent of logged forest in the tropic is vast, yet they represent the next best alternative to the protection of primary forest. Given that financially viable pathways for global action on forest degradation will be linked to climate mitigation potential, with the aim of achieving secondary benefits for biodiversity and human livelihoods, an improved understanding of selective logging's role the nexus of REDD+ is essential.

APPENDIX

A1

Supplementary Material to Chapter 3

“A machine learning approach to map tropical selective logging “

A1.1 Model performance without GLCM texture measures

Table A1.1 Confusion matrix summarizing EARLY Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging). Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 17,809). The classification threshold (T) for RF models was set during model calibration such that the proportion of detections that were truly logged (d_{pL}) was fixed at 0.85, resulting in a T of 0.69. The corresponding values for overall accuracy (OA), Cohen's kappa (k), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

EARLY - without GLCM Overall: 90.0% Kappa: 0.50 d_{pL}: 0.77 P_d: 0.43		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	1116	328	22.7
	Unlogged	1470	14900	9.0
	Omission Error (%)	56.8	2.2	

Table A1.2 Confusion matrix summarizing EARLY Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging) and GLCM texture measures. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 17,809). The classification threshold (T) for RF models was set during model calibration such that the proportion of detections that were truly logged (d_{pL}) was fixed at 0.85, resulting in a T of 0.69. The corresponding values for overall accuracy (OA), Cohen's kappa (k), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

EARLY – with GLCM Overall: 93.5% Kappa: 0.73 d_{pL}: 0.81 P_d: 0.73		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	1898	460	19.5
	Unlogged	685	14766	4.4
	Omission Error (%)	26.5	3.0	

Table A1.3 Confusion matrix summarizing LATE Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging). Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 17,847). The classification threshold (T) for RF models was set during model calibration such that the proportion of detections that were truly logged (d_{pL}) was fixed at 0.85, resulting in a T of 0.66. The corresponding values for overall accuracy (OA), Cohen's kappa (k), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

LATE - without GLCM Overall: 91.0% Kappa: 0.60 d_{pL}: 0.81 P_d: 0.54		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	1390	331	19.2
	Unlogged	1187	14943	7.4
	Omission Error (%)	46.1	2.2	

Table A1.4 Confusion matrix summarizing LATE Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging) and GLCM texture measures. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 17,847). The classification threshold (T) for RF models was set during model calibration such that the proportion of detections that were truly logged (d_{pL}) was fixed at 0.85, resulting in a T of 0.66. The corresponding values for overall accuracy (OA), Cohen's kappa (k), the proportion of detected pixels that were truly logged (d_{pL}), and the detection probability (P_d) are provided.

LATE - with GLCM Overall: 91.0% Kappa: 0.58 d_{pL}: 0.80 P_d: 0.51		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	1310	325	19.9
	Unlogged	1261	14943	7.8
	Omission Error (%)	49.0	2.1	

A1.2 Landsat pixels in RF model inputs

Table A1.5 Number of pixels from each forest management unit (FMU) used to build Random Forest models to detect selective logging with early and late Landsat scenes. Numbers of pixels from imagery acquired in 2012 are lower because of missing data regions resulting from the scan-line corrector error aboard Landsat 7.

Time period	Cut status	FMU	Image year								
EARLY			2008	2009	2010	2011	2012	2013	2014	2015	2016
	Unlogged										
		FMU 1	2554	2554							
		FMU 2	2728	2728	2728						
		FMU 3	1631	1631	1631	1631					
		FMU 4	2128	2128	2128	2128	1541				
		FMU 5	2658	2658	2658	2658	1889	2658			
		FMU 6	2000	2000	2000	2000	1468	2000	2000		
	Logged										
		FMU 1				2554					
		FMU 2					2035				
		FMU 3						1631			
		FMU 4							2128		
		FMU 5								2658	
LATE											
	Unlogged										
		FMU 1	2554	2554							
		FMU 2	2728	2728	2728						
		FMU 3	1631	1631	1631	1631					
		FMU 4	2128	2128	2128	2128	1531				
		FMU 5	2658	2658	2658	2658	1870	2658			
		FMU 6	2000	2000	2000	2000	1332	2000	2000		
	Logged										
		FMU 1					1894				
		FMU 2						2728			
		FMU 3							1631		
		FMU 4								2128	
		FMU 5									2658

A1.3 Feature selection

The Boruta package, version 6.0.0, was used to assess feature importance and all were deemed to contribute significantly to classification.

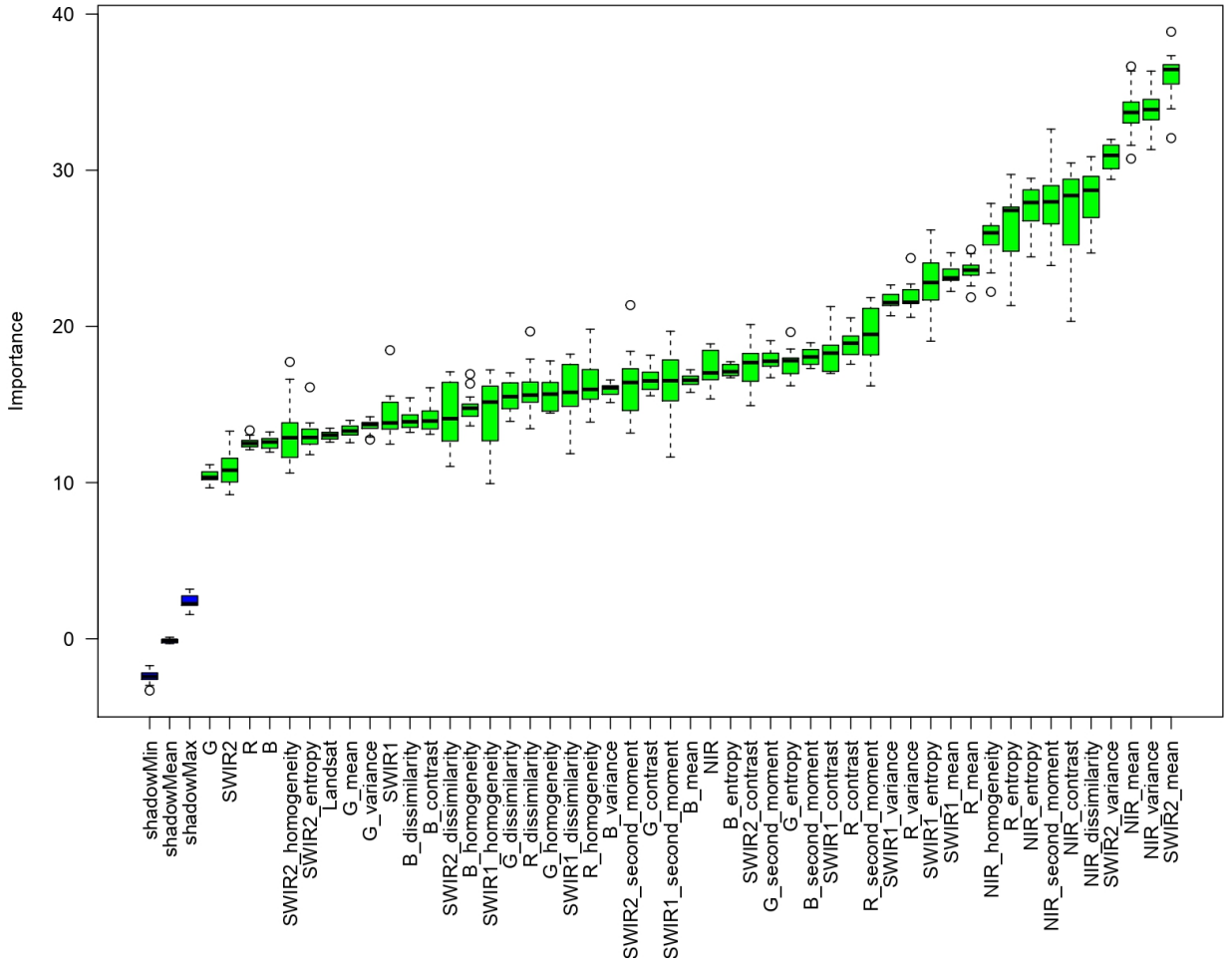


Figure A1.1 Variable importance measures from the Boruta package. Bars with green fill indicate significant contribution to classification, while yellow and red (not shown) indicate marginal and non-significant contributions, respectively. Thus, all variables were deemed important.

A1.4 Detection of selective logging through time at Jamari

Table A1.6 Landsat 8 (OLI) scene dates used from path 232, row 066 used to create Figure 3.8 in Chapter 3.

UPA2013	UPA2014	UPA2015	Unlogged
2013-07-10	2014-06-11	2015-06-14	2015-06-14
2013-07-26	2014-07-13	2015-06-30	2015-06-30
2013-08-11	2014-08-30	2015-08-01	2015-08-01
2013-08-27		2015-08-17	2015-08-17
		2015-09-02	2015-09-02

A1.5 Comparisons with CLASlite

A1.5.1 The CLASlite software

CLASlite is an unsupervised, pixel-based classification program developed specifically for mapping tropical deforestation and forest degradation (Asner et al., 2009a). It employs a spectral unmixing model that utilizes a vast spectral library (>250,000 observations) to distinguish the proportion of three endmembers within each image pixel: Bare Ground (BG), Photosynthetic Vegetation (PV), and Non-Photosynthetic Vegetation (NPV) that collectively sum to 100%. Changes in endmember values between time steps (i.e. image pairs) are used to identify forest disturbances. Specifically, CLASlite looks for a sharp rise in NPV (i.e. dead and dying vegetation from felled trees) and a simultaneous drop in PV (i.e. canopy cover) or an increase in BG (from roads, skid trails, or log decks used to stack cut tree before transport) to identify forest disturbance associated with loss of tree cover. Detection of forest degradation by CLASlite is automated and changes in pixel values between time steps are labelled as degraded using an internal decision tree (Asner et al., 2009a).

A1.5.2 Methods

CLASlite version 3.3 was used to process Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI scenes spaced approximately annually over the Jamari site between 2008 and 2016. CLASlite is actively maintained and updated (e.g. Sentinel-2 capabilities have been recently added) and the developers have continued to add features that enable the user to adjust settings, given their knowledge of the study site. One such feature attempts to filter out pixels erroneously labelled

as degraded by insisting that a pixel can only be classified as degraded if 5 pixels within its surrounding 7x7 window are also classified as degraded. This is because forest disturbances do not occur in isolation and other pixels in the neighbourhood would be expected to experience changes in endmember values related to degradation. This feature was turned off since the spectral signature of logging at Jamari was likely to be very subtle and we wished to give CLASlite the best opportunity to identify real changes. Hence all pixels that met CLASlite's criteria for changes in endmember values between time steps were labelled as degraded.

A1.5.3 Results

A1.5.3.1 CLASlite detection of degradation at Jamari

CLASlite did not label a single pixel as logged when using post-disturbance Landsat scenes (late), and, out of the 184 pixels labelled as logged in the full set of early time period images, only 34 (out of 11,006) had actually been logged. In addition, approximately 40% of CLASlite's detections were from the expansion of logging roads (Figure A1.2 and Table A1.7). Hence almost no degradation was detected by CLASlite in either the early or late time period Landsat scenes. Similarly, road networks were largely invisible, with only 74 pixels detected from more than 100 km of roads digitized in the Jamari site.

A1.5.3.2 CLASlite detections at Jari

CLASlite only labelled 71 ha of forest as degraded (Figure A1.3). In addition, CLASlite does not correct for the Landsat 7 scan-line corrector error and missing data regions are not ignored. In contrast, RF models with early data labelled 2316 ha as logged (Figure A1.3). Importantly, our method has the advantage of being able to make predictions about forest disturbances on a single scene to map degradation, as opposed to requiring successive cloud-free images, like CLASlite.

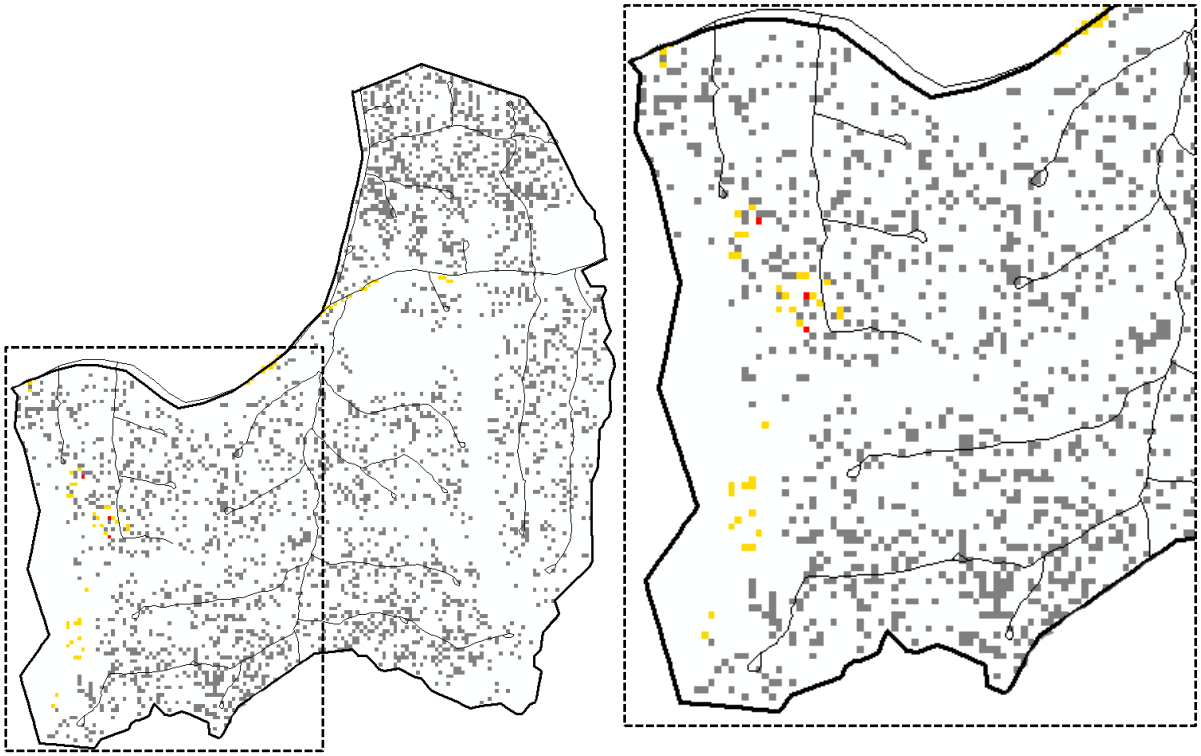


Figure A1.2 Example of a forest management unit in Jamari (logged in 2011) showing the locations of pixels that were logged in grey. CLASlite true and false detections using early Landsat scenes are shown as red and yellow respectively. GPS digitized logging roads are displayed as thin black lines. Pixels that coincided with mapped logging roads were removed from tallies of false detections.

Table A1.7 Summary of degradation detections by CLASlite using early Landsat scenes from the Jamari site between 2010 and 2015, with N_L and N_{UL} being the number of logged and unlogged observations, respectively. The time periods correspond to pre-logging and the year of logging for each forest management unit (FMU). From 2010 to 2014 logging road pixels were excluded from the number of false detections, however, only a partial road layer (a single, main access road) was available for the final FMU logged in 2015. Consequently road detections are underestimated and false detections are likely overestimated for FMU 5. The corresponding values for true (P_d) and false (P_{fd}) detection probabilities, the proportion of detections that were truly logged (d_{pL}), and Cohen's kappa (κ) are also given.

FMU logged	Time period (years)	Approximate logging intensity (m^3 ha^{-1})	N_L (pixels)	N_{UL} (pixels)	CLASlite detections (pixels)	Road detections (pixels)	True detections (pixels)	False detections (pixels)	P_d (%)	P_{fd} (%)	d_{pL} (%)	Cohen's kappa[1] (κ)
1	2010-2011	8.5	2554	15109	70	32	3	35	0.12	0.23	7.89	-0.002
2	2011-2012	9.9	2724	13647	12	3	2	7	0.07	0.05	22.22	0.001
3	2012-2013	8.2	1633	19959	12	4	5	3	0.31	0.02	62.50	0.005
4	2013-2014	8.0	2128	17648	21	17	3	1	0.14	0.01	75.00	0.002
5	2014-2015	11.3	2658	18755	69	18	21	30	0.79	0.16	41.17	0.011

[1] negative values for Cohen's Kappa occur when agreement is smaller than would be expected by chance

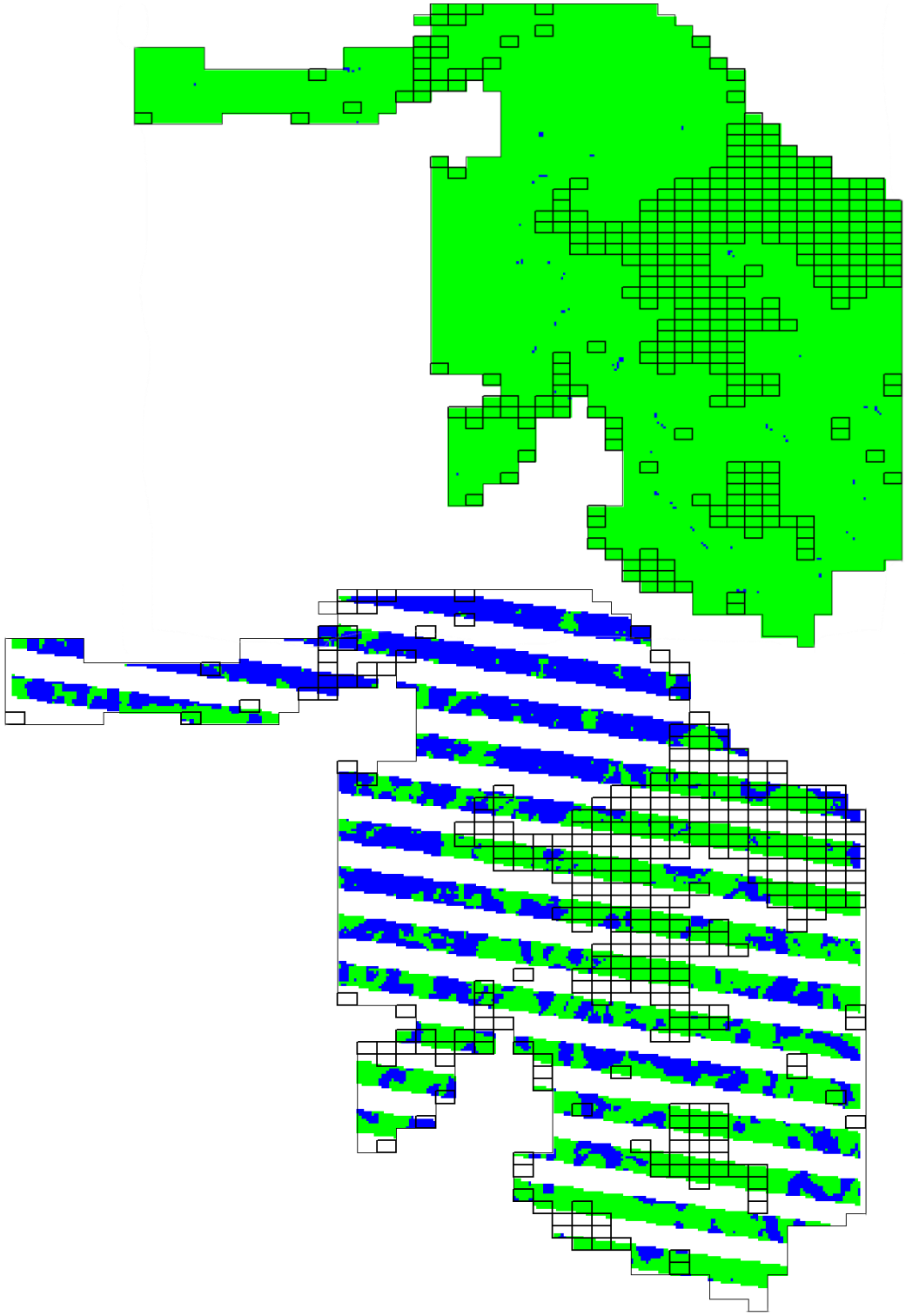


Figure A1.3 Classified maps of the Jari study site using CLASlite (top) and Random Forest (RF) models trained with early Landsat data (bottom). The image pairs used to create the classifications were from 8 November 2011 and 10 November 2012. Note that CLASlite does not correct for the Landsat 7 scan-line corrector error and missing data regions do not appear as white stripes. A threshold value of 0.65 was used for labelling logged pixels in RF models (see Table 2 in manuscript). Logged and unlogged forest pixels are displayed in blue and green, respectively. Black boxes are the 10 hectare blocks inside the Jari concession that were not logged.

A1.6 Random Forests and Gradient Boosted Model comparison

We compared the performance of gradient boosted models (Friedman et al., 2000; Friedman, 2002; Lawrence et al., 2004) and Random Forest models (Breiman, 2001) for classifying logged and unlogged pixels from the Jamari data set. We used the gbm package (Ridgeway, 2015) and the randomForest package (Liaw & Wiener, 2002) in program R (version 3.3.1). Gradient boosted models (GBMs) are sequentially built decision trees, as opposed to being built in parallel like Random Forest (RF) models. Decision trees (i.e. models) are fit iteratively to the training data and at each step the variation in the response that was not explained by the model is evaluated. We used the data inputs described in section 3.3.1 (Landsat Surface Reflectance, texture measures, etc.) and executed 10 runs for each model type. In each run a random subset of 70% of the data was used to train the model and the remaining 30% was used to evaluate the model built in that iteration. We present a confusion matrix with commission and omission errors associated with classification from a single model run and provide means and standard deviations of accuracies from the 10 model runs (Tables A1.8 and A1.9).

In general accuracies were very similar across modelling approaches (i.e. RF versus GBM models) and changes to the threshold criteria (i.e. cutoff values for labelling a pixel as logged or unlogged) would have had little impact on accuracies. The main advantages of using RF models were the fewer tuning parameters and shorter execution time in R (i.e. 4 minutes for each RF model to complete being trained and ready for use in predictions versus nearly 20 minutes for each GBM).

Table A1.8 Confusion matrix summarizing results from Random Forest classification models that utilized 56073 labelled point locations of known fate (before and after selective logging). Data were split into 70% training and 30% validation. Matrix numbers are representative outputs from a single model run with the test data. Error percentages are means from 10 model runs and standard deviations are given in parentheses.

Overall: 92.8% (0.2) Kappa: 0.70 (0.01)		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	2122	295	12.5 (0.5)
	Unlogged	1152	17257	6.5 (0.2)
	Omission Error (%)	36.3 (0.6)	1.7 (0.1)	

Table A1.9 Confusion matrix summarizing results from gradient boosted classification models that utilized 56073 labelled point locations of known fate (before and after selective logging). Data were split into 70% training and 30% validation. Matrix numbers are representative outputs from a single model run with the test data. Error percentages are means from 10 model runs and standard deviations are given in parentheses.

Overall: 92.4% (0.1) Kappa: 0.69 (0.2)		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	2157	442	17.2 (0.6)
	Unlogged	1151	17079	6.2 (0.2)
	Omission Error (%)	33.6 (0.8)	2.6 (0.1)	

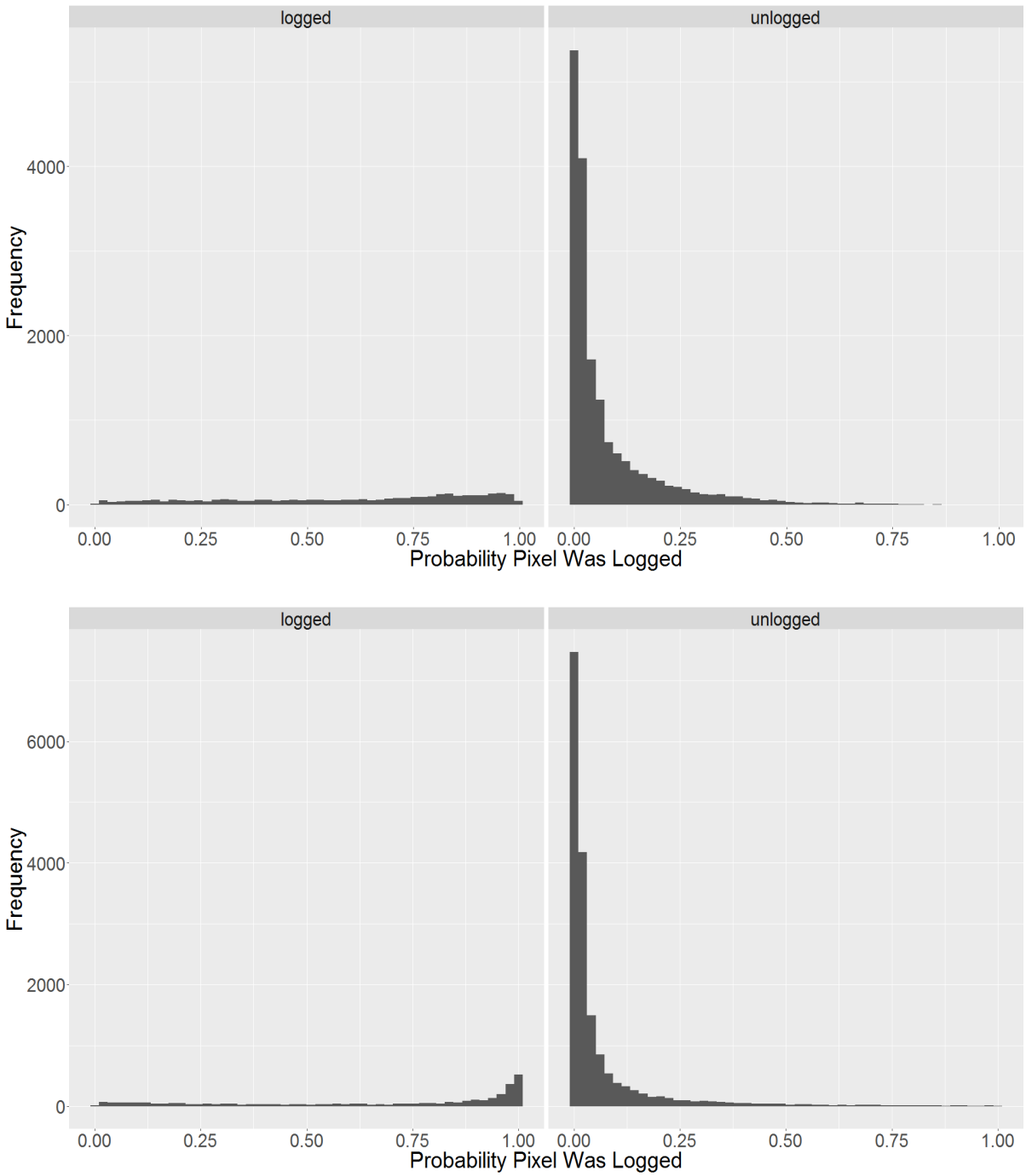


Figure A1.4 Histograms of predicted probabilities from pixels in the test dataset that were left out of algorithm training from Random Forest (top) and Gradient Boosted (bottom) models.

A1.7 Removing NDVI and EVI from RF models

An early version of our algorithm included two vegetation indices (NDVI and EVI) and had modest classification accuracies (Table A1.10). However, when we applied this algorithm back through time to track changes in the probabilities that a pixel had been logged at our study site (as a very simple validation exercise), predicted probabilities were either 0 or 100% depending on the year (i.e. the entire scene was classified as logged or unlogged). We examined the distributions of NDVI and EVI values through time from a large region that remained unlogged within the concession and noticed strong annual variations in NDVI and EVI values (Figures A1.5 and A1.6). We suspected this variation in NDVI and EVI across years (potentially from differences in phenology or foliar moisture content) would influence our ability to forecast to new areas. If NDVI and EVI were poor predictors within the same site through time we had little confidence in their abilities to forecast over new locations with potentially different phenology patterns. A common criticism of machine learning algorithms centre on overfitting of training datasets, resulting in loss of predictive performance when applied to new data or new regions. Consequently, we removed NDVI and EVI from model development.

Interestingly, EVI showed clear differences across Landsat sensors, with Landsat 8 OLI values being consistently higher that Landsat 5 and 7 (Figure A1.6). This effect is known for NDVI and some solutions have been suggested to overcome spectral differences across sensors (Roy et al. 2016) to enable use of long time-series datasets. However, to date, we have not seen any work demonstrating these effects for EVI.

Table A1.10 Confusion matrix summarizing results from Random Forest model classification that included EVI and NDVI at 69,430 labelled point locations of known fate (before and after selective logging). Data were split into 70% training and 30% validation. Matrix numbers are representative outputs from a single model run with the test data (n=20,829 pixels). Errors are means from 10 model runs and standard deviations are given in parentheses.

Overall: 89.1% (1.1) Kappa: 0.59 (0.02)		Actual Class		
		Logged	Unlogged	Commission Error (%)
Predicted Class	Logged	1803	636	25.8 (2.0)
	Unlogged	1534	16850	8.5 (0.3)
	Omission Error (%)	45.8 (1.2)	3.5 (1.6)	

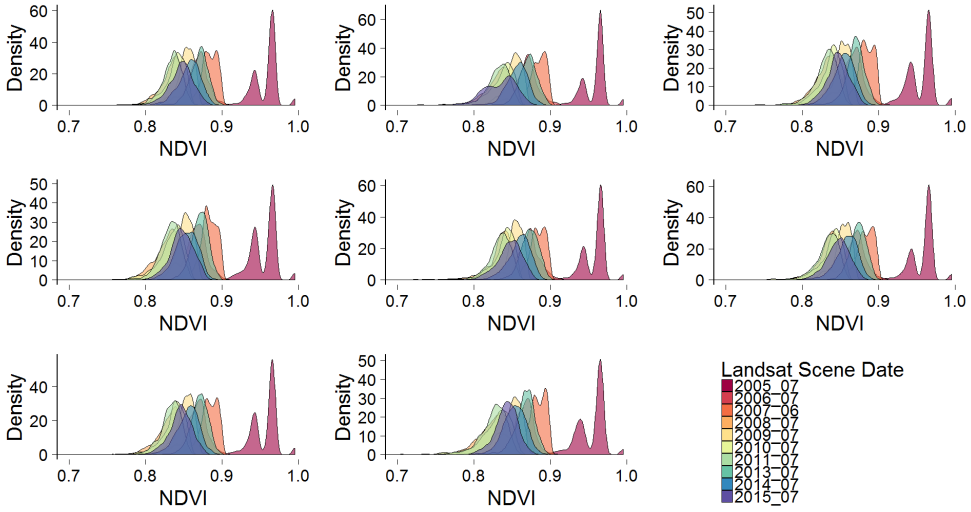


Figure A1.5 Density plots of the Normalized Difference Vegetation Index (NDVI) from 8 different forest management units (FMUs) within the Jamari site. All the FMUs remained unlogged over the 10 years plotted. Each panel represents a different FMU and the NDVI values are plotted for 2000 randomly selected pixels and colored by year. Regionally, FMUs displayed similar patterns in NDVI values (i.e. the FMU in upper left has similar NDVI values to the FMU in the lower left), but the same pixels showed strong differences across years (different colors within the same panel do overlap in some years).

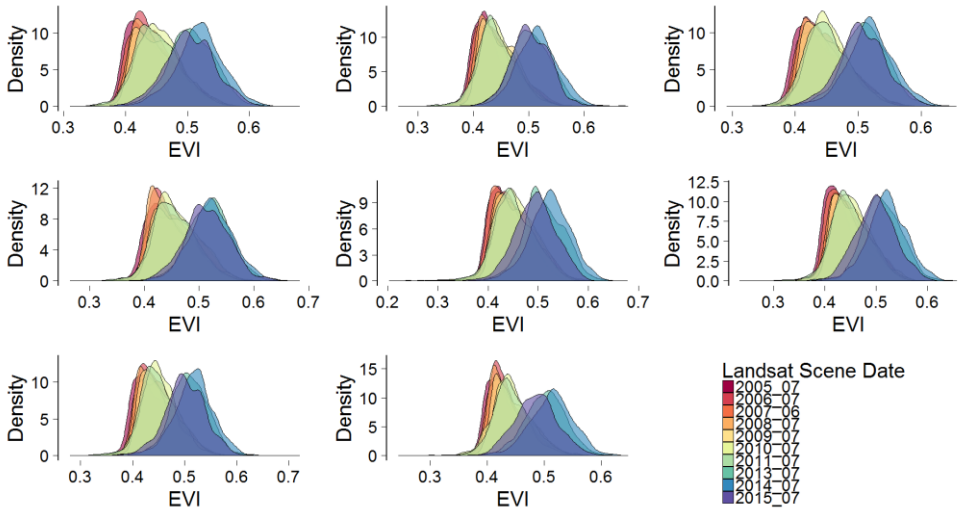


Figure A1.6 Density plots of the Enhanced Vegetation Index (EVI) from 8 different forest management units (FMU s) within the Jamari site. All the FMU s remained unlogged over the 10 years plotted. Each panel represents a single FMU and the EVI values are plotted for 2000 randomly selected points and colored by year. Regionally, UPAs displayed similar patterns in EVI values (i.e. the FMU in upper left has similar EVI values to the FMU in the lower left), but the same pixels showed strong differences across years (different colors within the same panel do overlap in some years).

A1.8 CLASlite: A review and validation with logging data

A1.8.1 Introduction

Earth's forests are being lost and degraded from, primarily, commercial logging operations (Eva et al. 2012; Hansen et al. 2013). Between 2000-2012, a combined area the size of Peru and Ecuador was cleared globally- a third of which was tropical forest (Hansen et al., 2013). The rapid depletion of forests in the tropics has severe implications for global climate change, local populations and biodiversity. Assessments of global carbon stocks suggest tropical forests store 40–60% of all carbon held in terrestrial vegetation (Saatchi et al. 2011; Pan et al. 2011; Baccini et al. 2012). In addition, tropical forests are known to harbour untold levels of biodiversity that are increasingly threatened by a myriad of anthropogenic disturbances (Gibson et al. 2011). Logging, whether selectively or clearcutting, is often a precursor to additional land-use changes (such as agricultural conversion or development of human settlements), with road networks facilitating further degradation and forest losses (Laurance et al. 2009, 2014).

The ability to map and quantify forest disturbances are an essential component of global conservation initiatives, such as the Reducing Emissions from Deforestation and Forest Degradation (REDD+) program. In addition, successful implementation of programs like REDD+ require the establishment of baseline data and reliable forest monitoring systems that enable accurate tracking of changes in carbon content or biodiversity measures. Remote sensing is considered the most accurate and cost-effective way to systematically monitor forests (Herold and Johns 2007; De Sy et al. 2012; Saatchi et al. 2015). Satellite monitoring of deforestation has come of age, with accuracies of 90 to 95% achievable with mid-resolution imagery to discriminate between forest and non-forest (Arino et al. 2015). In addition, near real-time tracking and alert systems are now possible with systems like DETER, FORMAS (Hammer et al. 2014), and others (Hansen et al. 2016). Present methods for detecting forest degradation, however, are poorly developed because the widely available data sets (e.g., Landsat or MODIS) are considered too coarse to measure the more subtle changes in forest structure associated with degradation (Herold et al. 2011).

Estimates vary widely, but forest degradation is thought to be a major contributor of greenhouse gas (GHG) emissions, comprising up to an additional 50% of emissions from forest loss alone (Asner et al. 2005, 2010b; Grace et al. 2014; Bustamante et al. 2016). However, forest degradation is an abused word, with over 50 different definitions and no internationally established description (Simula 2009; Ghazoul et al. 2015). While this makes generalizing the literature difficult, as degradation can include forests with varying intensities of selective logging, fire, or artisanal gold mining for example, a prescriptive application of a singular, rigid

characterization of forest degradation may not be appropriate either (Simula, 2009). In this review we have chosen instead to focus on a key driver of forest degradation, selective logging, as it is often the first anthropogenic disturbance event and is an agent for additional land-use changes (Nepstad et al., 1999; Asner et al., 2005, 2006, 2009).

In contrast to forest clearance (i.e., deforestation), forest degradation from selective logging represents a more diffuse disturbance wherein only a subset of trees (typically the most economically valuable) are harvested (Simula, 2009). The resulting forest maintains some semblance of its original structure (e.g., canopy cover, biodiversity measures, carbon content, etc.) and consequently falls along the continuum between primary forest and complete deforestation (Fearnside, 2000; Simula, 2009; Thompson et al., 2013; Ghazoul et al., 2015). The intensity of selective logging operations can vary in two main ways. First, and most straightforward, the number of trees removed (or volume of wood) can range from less than 1 tree ha-1 to greater than 20 (reviewed in Burivalova et al., 2014). Second, the degree to which damages to the remaining forest are minimized (e.g., careful planning of road networks or skid trails and directional felling of trees to minimize additional tree or canopy damage; termed reduced-impact logging). Studies of reduced-impact logging (RIL) practices have garnered increasing attention by researchers concerned with logging impacts on carbon stocks and biodiversity conservation (Putz & Pinard, 1993; Pereira et al., 2002; Asner et al., 2004a; Putz et al., 2008a; Edwards et al., 2012; Wilcove et al., 2013; Bicknell et al., 2014; Sist et al., 2014). While selectively logged forests have been shown to have increased microclimatic variability (Stratford & Robinson, 2005), increased soil erosion (Douglas, 1999; Hartanto et al., 2003), reduced tree diversity (Berry et al., 2008; Martin et al., 2015), altered forest phenology (Koltunov et al., 2009), and lowered levels of biodiversity (reviewed in Burivalova et al., 2014), forests subjected to selective logging generally maintain higher levels of biodiversity than other anthropogenic land use types, such as plantations or secondary forests (reviewed in Edwards et al., 2014). Moreover, recent works have shown that even after accounting for the amount of wood removed, RIL has a greater effect on maintaining biodiversity than conventional selective logging (CL) practices (Bicknell et al., 2014) while simultaneously sequestrating more carbon during regrowth (Putz et al., 2008b). Thus, in the context of REDD+ or alternative conservation initiatives, forests impacted by RIL offer high biodiversity value and carbon sequestration potential, making them ideal for carbon and biodiversity co-benefits.

Several authors have tried to address the challenges of using remotely sensed data to assess forest degradation from selective logging (Souza & Barreto, 2000; Asner et al., 2002, 2004a, 2005; Souza et al., 2005; Matricardi et al., 2007, 2010; Hirschmugl et al., 2014; Shimabukuro et al., 2014). Among these methods, the CLASlite software (Asner et al., 2009)

represents the state-of-the-art in automated optical imagery mapping to understand forest structure. In contrast to whole pixel classification, where a raster pixel is assigned to a single dominant land cover type, CLASlite uses spectral mixture analysis to determine the proportion of each land cover type inside each pixel. CLASlite's automated Monte Carlo unmixing (AutoMCU) algorithm utilizes a vast spectral library (over 250,000 observations) to distinguish 3 land cover types within each pixel; Bare Ground (BG), Photosynthetic Vegetation (PV), and Non-Photosynthetic Vegetation (NPV) whose proportions collectively sum to 100%. The Monte Carlo approach randomly selects spectra from the PV, NPV, and BG libraries and solves a set of linear equations to determine the fractional cover of each end member (Figure 1b). The selection process is repeated many times in each pixel to obtain the final fractions corresponding to the mean solution for BG, PV, and NPV (with a root mean squared error also calculated).

CLASlite has 4 main steps:

1- Convert the satellite image to surface reflectance (i.e., correct for atmospheric effects)
2- Estimate the proportion of BG, PV, and NPV within each pixel (AutoMCU described above)
3- Map forest cover within a single scene (i.e., generate a forest/non-forest map)
4- Map deforestation and forest degradation between image pairs (i.e., look for changes in BG, PV, and NPV through time steps)

In order to generate the forest cover map (step 3) CLASlite utilizes a simple decision tree to convert the AutoMCU output (proportions of BG, PV, and NPV from step 2) into a forest/non-forest map. The decision tree in step 3 defines forest as a pixel with PV ≥ 80% AND BG < 20%. As a consequence, non-forest is defined as a pixel with PV < 80% OR BG > 20%. Degradation is not assessed in step 3 and is only mapped through a time series analysis in step 4, where thresholds for changes in pixel values define the rules for classifying deforestation and degradation. The decision tree in step 4 is summarized below. The numbers following the CLASlite endmembers (BG, PB, and NPV) identify the time period associated each endmember value.

Deforestation:

(PV1 – PV2) ≥ 25%

OR

(BG1 ≤ 5%) AND (BG2 – BG1) ≥ 15%

OR

(PV2 < 80%) AND (NPV2 – NPV1) ≥ 20%

Degradation:

(NPV2-NPV1) ≥ 10 AND (PV1-PV2) > 10%

OR

(BG1 ≤ 5%) AND ((BG2-BG1) > 10%) AND (BG2 ≤ 15%)

CLASlite's internal decision tree provides some insight into the key spectral features used to identify areas experiencing forest degradation. Specifically, CLASlite looks for either a spike in NPV (i.e., dead and dying vegetation from felled tree, termed slash) and a drop in PV, or an uptick in BG (from roads, skid trails, or log decks used to stack cut tree before transport). Finally, in an effort to reduce the number of unlogged pixels that are classified as degraded, CLASlite uses a pixel filter at the end of step 4. To pass through the filter (and be classified as degraded) 5 pixels (non-contiguous) within a 7x7 moving window must also be classified as degraded. The idea here being that forest disturbance does not happen in isolation and other pixels in the neighbourhood should experience changes in end member fractions related to degradation.

Identifying and quantifying forest degradation from selective logging with remotely sensed data represents the vanguard of current technologies, and the literature suggests CLASlite is only capable of detecting selective logging areas of moderately high intensities (greater than 3 trees ha-1). Specifically, CLASlite has been validated over regions possessing spectrally distinct features like log landing decks, road networks, or significant amounts of slash (Asner et al., 2004a, 2004b, 2005, 2006, 2010; Broadbent et al., 2006; Oliveira et al., 2007; Carlson et al., 2013). However, with increasing prevalence of RIL and growing concerns over impacts on carbon and biodiversity, the extent of tropical forests that have been logged at lower intensities (less than 2 trees ha-1) is rapidly expanding.

Here, we review the existing literature on CLASlite to determine who, aside from Gregory Asner's group, has used the software, what intensities of degradation have been assessed, and in general, what key unknowns remain. In addition, we carried out two sets of analyses to better understand CLASlite. First, we explored correlations among the MODerate-resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) product and PV values from CLASlite. Lastly, we tested the ability of CLASlite to distinguish regions that had undergone low

levels of selective logging (approximately 1 tree ha-1) from areas that had not been logged (primary forest) in a test site located in the Brazilian Amazon (Figure A1.8.1a). Crucially, we wanted to understand the lower limits of CLASlite's abilities to detect selective logging.

A1.8.2 Review of CLASlite literature

We searched Google Scholar using the keywords "claslite" and "Carnegie Landsat Analysis System" to identify papers containing reference to Asner's (2009) software. In addition, we utilized the search tool available within Google Scholar to list the manuscripts that had specifically cited the CLASlite reference. We only considered published papers and excluded reports or book chapters. Moreover, we chose to include papers that contained results not relevant to logging or selective logging in order to understand the range of applications.

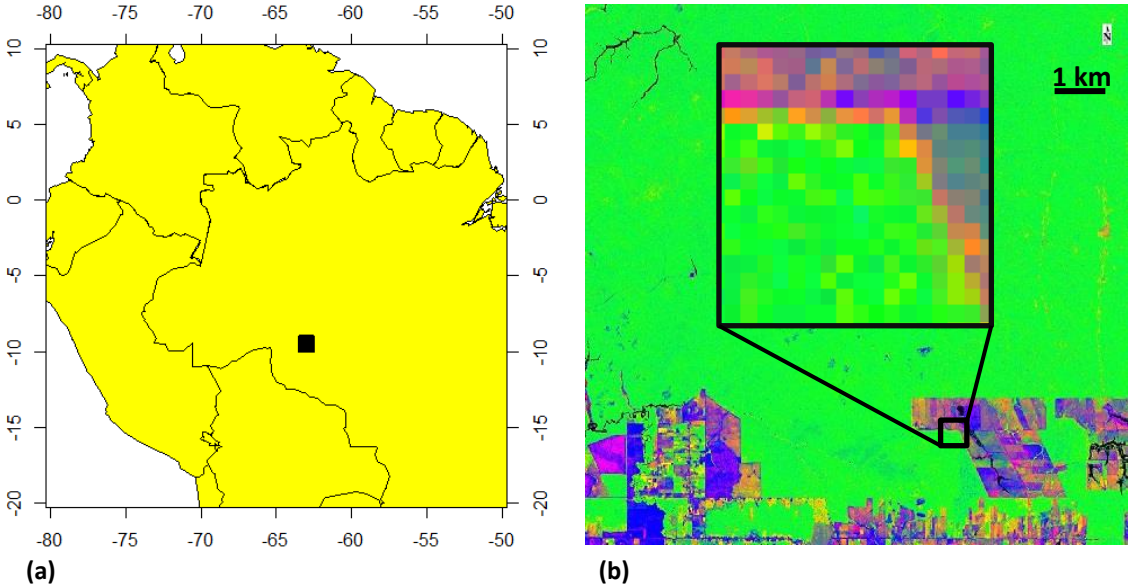


Figure A1.7 (a) Location of study site (black box) in the Jamari National Forest, Rondônia, Brazil and (b) CLASlite output from same location. Inset zoom (in b) illustrates the 30m x 30m pixels wherein unique colors indicate the combination of BG, PV, and NPV at each cell location. Bare ground (BG), photosynthetic vegetation (PV), and non-photosynthetic vegetation (NPV) are displayed in the colors red, green, and blue, respectively.

Literature searches yielded 197 papers under the keyword "claslite" and 99 papers under "Carnegie Landsat Analysis System"; note that these numbers included overlapping findings. An additional 115 papers were listed as having cited the CLASlite reference, again including overlapping results. From these lists we identified 36 papers that had used CLASlite, of which 12 did not include Gregory Asner as a co-author. Only 1 of the 12 papers assessed degradation; 8 calculated deforestation (forest/non-forest classifications) and simply referenced degradation within the introduction or discussion (i.e., no results), and 3 studies only used CLASlite to pre-process Landsat imagery to account for atmospheric effects (i.e., to convert from radiance to surface reflectance). The lone paper that included results on degradation assessed declines in forest canopy cover in the decades following the construction of a hydroelectric dam. The authors had no ground information regarding the intensity of disturbance and defined their own thresholds for deforestation and degradation. Deforestation was canopy cover (PV pixel values) less than 38% and degradation was PV values between 38% and 72% (Chen et al. 2015). However, Asner's papers indicate that, in densely forested tropical areas (like the Amazon basin where this study is from), PV values less than 67% represent total canopy opening- a consequence of light scattering effects and pixel adjacencies (Asner et al. 2006). In addition, CLASlite relies on changes in fractional endmembers between time steps, rather than threshold values within a single scene like the authors used, to classify degradation. Thus, most of the regions assessed as degraded by the above study were actually deforested

In general, there appears to be a lack of uptake and use of the CLASlite software outside of researchers linked to Asner's group. This may be, in part, a result of publication bias, since we chose to focus on manuscripts or book chapters and did not include reports from agencies or non-governmental organizations (NGOs). This latter user base is the target group Asner's CLASlite is largely intended for, in an effort to aid the accurate tracking of carbon stocks for global conservation initiatives like REDD+. That said, however, CLASlite does not overcome any of the standard limitations of using optical imagery in the tropics, including the limited availability of cloud-free images over many tropical regions, which can impede monitoring efforts (Asner 2001) and the rapid regeneration of tropical forest vegetation, which can hinder the detectability of canopy gaps or other important spectral features (Stone and Lefebvre 1998; Asner et al. 2004a). CLASlite's benefits, however, include the fact that the software is freely available (after completing and passing a series of online tutorials and examinations), its use of freely available datasets (Landsat archives and ASTER, now those data have been released for free use to the public), its simplicity in pre-processing satellite imagery (converting radiance to surface reflectance) and generating a deforestation layer, its future potential (now that Landsat 8 is operational and the problems associated with using Landsat 7 images that contain the scan

line corrector error will diminish), and finally the ability to integrate its outputs with other data types (e.g., LiDAR, synthetic aperture radar (SAR), vegetation indices like NDVI, or field data) into an analysis workflow.

Finally, we reviewed the CLASlite papers Asner's group has published, examining everything that used the CLASlite system, while paying particular attention to studies that dealt with forest degradation. Similar to the earlier studies we reviewed, the vast majority of Asner's papers utilized CLASlite to generate forest/non-forest maps (i.e., map deforestation) and only referenced degradation within the introduction and discussion. A lack of tree harvest (or volume) data from most manuscripts limits our ability to understand the lower limits of CLASlite's abilities to detect selective logging. Only their earlier studies (Asner et al. 2004b, a; Broadbent et al. 2006), indicate logging intensities (around 3-6 trees ha^{-1}) from field sampling locations, while their more recent papers treat airborne LiDAR data as "truth" and correlate remotely sensed PV with LiDAR based estimates of canopy coverage (Asner et al. 2010a; Carlson et al. 2012). The earlier studies that assessed degradation (Asner et al. 2004b, a; Broadbent et al. 2006), however, utilized imagery that was acquired within 6 months of forest disturbance events (to limit rapid regrowth of tropical foliage that might hide the signal). This final point is likely to be a key limit to CLASlite application. Researchers looking to identify new forest disturbances must deal with the frequency of cloud coverage over many tropical regions and, simultaneously, require imagery within a narrow window of time after disturbance (6 months).

Collectively, these works have highlighted the need to explore finer resolution optical imagery (e.g., WorldView, QuickBird, IKONOS), incorporate data from multiple sensors (e.g., SAR or hyperspectral imagery), and develop sophisticated image analysis techniques that integrate different data types (Joshi et al. 2016). Importantly, SAR data probably offers the best opportunities to overcome the limitations mentioned above. The absence of reliance upon cloud-free imagery and the regular return intervals SAR satellites provide make it an obvious choice. Finally, questions remain as to the lower limits of CLASlite's sensitivity to selective logging. The purpose of the remainder of this report is to address this question. We outline two sets of analyses carried out to better understand CLASlite and its limitations.

A1.8.3 Analyzing CLASlite's relation to MODIS VCF

In order to gain further information about the limits of CLASlite's abilities to detect selective logging, we compared the MODerate-resolution Imaging Spectroradiometer (MODIS) Vegetation Continuous Fields (VCF) product with the PV estimates from CLASlite over the same regions.

The VCF collection is derived from all seven bands of the MODIS sensor on-board NASA's Terra satellite and contains continuous estimates for percent vegetative cover globally at approximately 250m spatial resolution (product MOD44B.051, available at http://e4ftl01.cr.usgs.gov/MOLT/MOD44B.051/). We wanted to test how correlated CLASlite's PV values were with the VCF product.

A1.8.3.1 Methods- CLASlite's relation to MODIS VCF

CLASlite was used to process a single Landsat 8 scene (10 July 2013, path 232 row 66) over a test site in the Jamari National Forest, Rondônia, Brazil (Figure A1.8.1a) to generate the proportion of BG, PV, and NPV in each pixel (Figure 1.8.1b). The VCF product was then resized from 231m pixel resolution to 240m, using the nearest neighbour method, to allow a perfect alignment with the CLASlite output (i.e., 1 VCF pixel aligning with a set of 8x8 CLASlite pixels). The CLASlite output was then aggregated over an 8x8 pixel window, calculating the mean PV value for that area. Finally, values from the VCF product and the corresponding average PV values were extracted from 2000 randomly selected pixel locations. The relationship between VCFs and CLASlite PV values was tested using simple linear regression.

During this analysis (while re-reading the user guide for CLASlite) we discovered PV values from CLASlite are normalized based upon VCF values from the same locations (Asner et al. 2009). CLASlite looks at every pixel with PV values greater than 80% and compares it to the corresponding geographical location in the VCF product. CLASlite then adjusts the PV values such that the mean of the distributions of the coincident pixels from the VCF and the CLASlite image are equal. The NPV and bare substrate fractions are "adjusted accordingly" so that the sum of BG, PV, and NPV remains close to 100%. Consequently, in addition to examining the relationship between PV and VCFs across the entire range of values, we also pay particular attention to those data points within the upper range of PV values (1742 of the original 2000 points).

A1.8.3.2 Results and Discussion- CLASlite's relation to MODIS VCF

Across the full range of values, MODIS VCFs and CLASlite's PV values were highly correlated (Pearson's r = 0.899; Figure A1.8a). This result is not surprising, given what we learned about CLASlite's internal relation to the VCF product. However, the relationship between VCF and PV across the upper range of PV values (the region standardized by CLASlite) was not as highly correlated as Asner's notes suggest (Pearson's r = 0.528; Figure A1.8b). Because the CLASlite documentation indicates the mean PV values are adjusted to match the VCF values, a tighter relationship within this upper range of VCF values was expected. It is important to recall the PV

values presented here are a mean PV over an area of approximately 5.8 hectares. While the relationship is statistically significant, probably because the bulk of the data points fall within a narrow range of VCF values (between 88 - 96% tree cover), it suggests CLASlite's internal PV normalization methods are different from the approach carried out here.

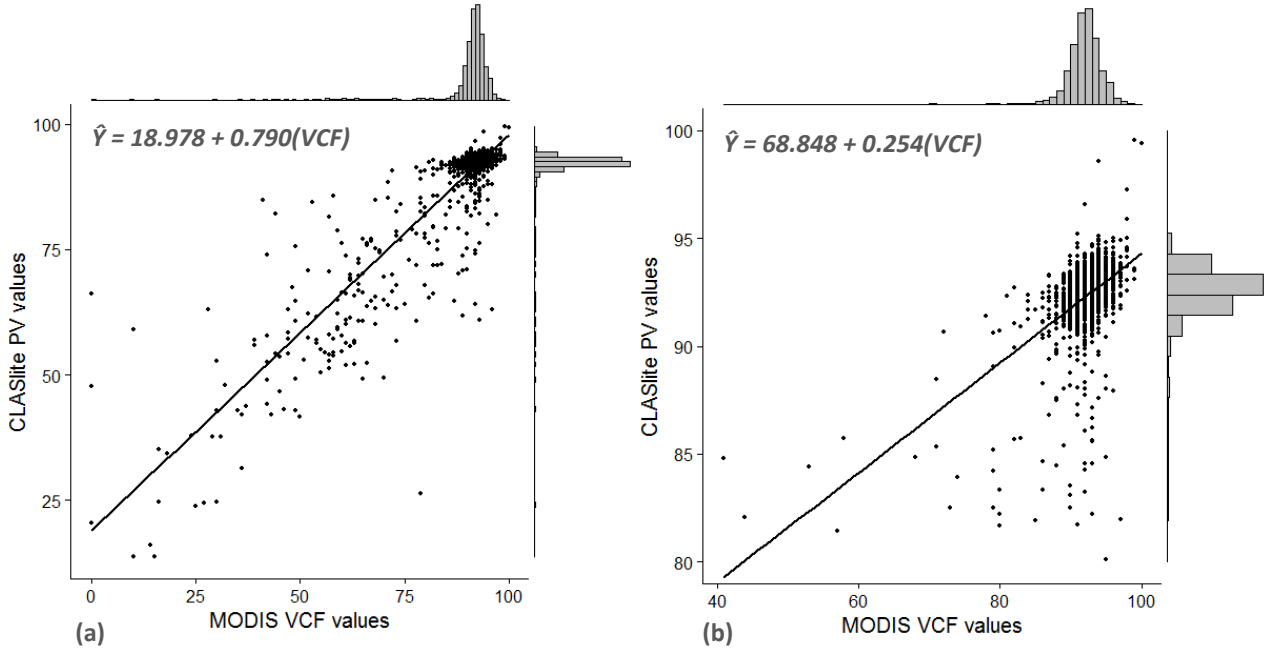


Figure A1.8 Scatterplot and linear regression for relationship between MODerate-resolution Imaging Spectro-radiometer (MODIS) Vegetation Continuous Fields (VCF) values and CLASlite photosynthetic vegetation (PV) values from (left) 2000 randomly selected point locations and (right) a subset of the same points (n=1742) that have PV values greater than 80. PV values were extracted from a July 2013 Landsat 8 scene that had been processed in CLASlite and MODIS values were obtained from an annual composite from 2013. Frequency histograms are overlaid on the opposite axes for each variable.

A number of older studies used VCF data (at 500m resolution) to generate estimates of tree cover over various regional scales (Hansen et al. 2000, 2002, 2008; Zhan et al. 2002). These works suggested a VCF value of around 25% signified deforested areas, whereas values in the more recent VCF product (250m resolution) approach zero over deforested areas (DiMiceli et al. 2011). Interestingly, as the CLASlite documentation suggests, PV values less than 67.5% signify total canopy openings (i.e., deforestation) in forested regions. This value corresponds quite well with the intercept (i.e., a VCF value of zero) from the equation describing the relationship between VCF and PV in figure 2b ($\hat{Y} = 68.848 + 0.254 \times VCF$). It is unclear if this is where Asner derived his estimates of minimum PV, if so, it suggests our approach was a close approximation and perhaps the internal CLASlite normalization requires a significant relationship (and not a near one-to-one relationship). Ultimately, however, this exercise did not give us any insight into the break point at which loss of forest is no longer detectable with CLASlite. We address this in the next section.

A1.8.4 Testing CLASlite with field data

We wanted to understand CLASlite's abilities to discriminate regions that had undergone low levels of selective logging (approximately 1 tree ha^{-1}) from areas that had not been logged (primary forest). CLASlite was used to process Landsat (5 and 8) scenes over a test site in Rondônia, Brazil (Figure A1.7a). The site encompasses tropical forests included in the 46,000 hectare Jamari Logging Concession, inside the Jamari National Forest. The concession is subdivided into management units that are approximately 1500– 2000 ha in area. Selective logging activities occur within a single management unit in each year, beginning at the end of the wet season (approximately April) and continuing through the dry season (until approximately November). A long-term study is in place within this concession with the objective of assessing the impacts of logging intensity and management techniques on tree biodiversity and carbon stocks. Forest inventory measurements have been recorded since operations began in 2011 and include detailed spatial information on each marketable tree species within the concession, its height, diameter, volume, and if it was logged in subsequent years. This dataset provided a means to assess the accuracy of CLASlite's ability to detect low selective logging (approximately 1 tree ha^{-1}) because there was a detailed account of where and when trees were selectively removed.

The objectives of this study were to answer 4 key questions:

1- Does CLASlite detect selective logging at the intensities present in the Jamari test site?
2- If not, can CLASlite's AutoMCU outputs (BG, PV, and NPV) be used to develop a novel decision tree classification system?
3- Again, if not, if we focus only on the most intense locations (pixels with most trees or volume removed) does this improve classification accuracy?
4- Finally, would adding additional metrics (like NDVI) improve the classification accuracy from our own decision tree?

A1.8.4.1 Methods Objective 1: Can CLASlite detect low intensity selective logging at the Jamari site?

CLASlite was used to process 2 Landsat scenes (before and after selective logging) over a single management unit that was selectively logged in 2012. The pre-logging image was from 6 August 2011 and the post-logging image was from 11 August 2013. Forest/non-forest maps at each time period (from step 3 in CLASlite) were used with all default settings to produce a

degradation map (step 4) between 2011 and 2013. The degradation output (from step 4) was visually inspected over the Jamari test site to determine if logging activities were detected. The number of pixels correctly classified as degraded is reported as a measure of accuracy.

A1.8.4.2 Results and Discussion Objective 1

With default settings, CLASlite did not identify a single pixel as degraded within the test site. Thus, on its own, CLASlite's internal decision tree system is not sufficient to identify the levels of selective logging (around 1 tree ha^{-1}) present in the Jamari test site. CLASlite does allow for some settings to be customized by the user, but these are limited to adjusting the thresholds for the deforestation map in step 3 to allow the users to exclude agricultural areas or regrowth that were incorrectly classified as forest (i.e., no settings can be adjusted for degradation in step 4). While this suggests CLASlite cannot be used to map degradation at our study site, it does not rule out using the outputs from the AutoMCU algorithm (BG, PV, and NPV fractional covers) to develop our own decision tree.

A1.8.4.3 Methods Objective 2: Can CLASlite outputs be used to make a novel decision tree?

We used CLASlite to process Landsat (5 and 8) scenes before and after selective logging activities at 4 management units that were cut sequentially between 2011 and 2015. Landsat scenes acquired after selectively logging were processed from 2 time periods (Figure A1.9); 1) the first cloud-free image of the dry season following the cessation of logging activities (approximately 8-12 months after logging, termed "late" hereafter), and 2) the last cloud-free image in the same year of logging (approximately 2-3 months before the cessation of logging activities for the season, termed "early" hereafter). The early time period was added after a review of the literature on CLASlite, since the regrowth of foliage (after 8-12 months) can reduce the spectral signatures required to identify canopy gaps (PV) and slash (NPV) in tropical systems (Asner et al. 2004b, a; Broadbent et al. 2006). Thus, we wanted to evaluate which time period after selective logging provided the best detection of logging activities at our site.

For the late time period BG, PV, and NPV values were extracted from 7192 point locations of known tree removals 2 years before logging (i.e., still primary forest) and at the same locations in the year following selective logging for a total of 14,384 points of known fate.

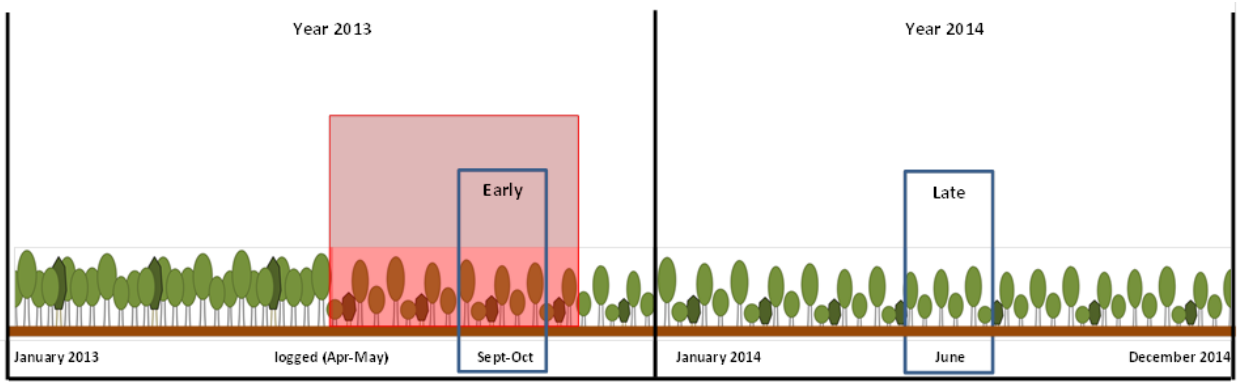


Figure A1.9 Timeline representation of a single forest management unit selectively logged in 2013. Blue boxes indicate when the Landsat imagery was acquired relative to when selective logging occurred (red box) for the early and late time periods. In this example, the selective logging occurred in year 2013 and the Landsat scene was acquired part way through the logging season for the early period (some of the management unit, to the right of the blue box, has yet to be cut). The late time period is the first cloud-free image of the dry season in the following calendar year and is approximately 8-12 months after selective logging occurred within the management unit.

For the early time period BG, PV, and NPV values were extracted from 7072 point locations of known tree removals 2 years before selective logging and at the same locations in the midst of the logging season for a total of 14,144 points of known fate. These numbers differ slightly (14,483 late points and 14,144 early points) because there is a gap in the available Landsat archive in the year 2012 and no imagery exists. Consequently, point locations that required a 2012 scene to be used to assess changes in end members after logging were excluded. Thus, for pixels logged in 2012, the early period would have required an image from approximately October 2012, but these data were not available and were excluded. Similarly, for the late time period, pixels logged in 2011 required imagery from approximately early June 2012 that was unavailable and therefore excluded. As a result, the early period analysed pixels from management units cut in 2011, 2013, and 2014, whereas the late period analysed pixels from management units cut in 2012, 2013, 2014.

The Random Forests decision tree algorithm (Breiman 2001; Liaw and Wiener 2002) was used to analyse the CLASlite outputs and build a new decision tree based on the attributes of known fate locations. Random Forests is an ensemble classifier that uses predictor variables (in this case our 3 constituent end-members from CLASlite; BG, PV, and NPV) to classify a response variable (whether the pixel was logged or unlogged forest). Bootstrapped samples are drawn to construct multiple decision trees (n = 500) and the predictors used to find the best split at each node are identified. Per convention, 70% of the data was assigned for classification development and the remaining 30% was used for validation. Here, accuracy was defined as the proportion of pixels correctly assigned by Random Forests as being selectively logged and unlogged, respectively.

A1.8.4.4 Results and Discussion Objective 2

Random Forest classification accuracy of selectively logged and unlogged pixels was around 60% using BG, PV, and NPV outputs from CLASLite, for both late and early time periods (Table A1.11). Pixels from unlogged forested areas were accurately classified only about 55-60% of the time in either scenario. In general, irrespective of time period, the Random Forest classification performed better than CLASlite's internal decision tree (zero pixels classified as degraded using CLASlite in objective 1 and ~60% from Random Forests here), but this value is still below an acceptable level of error. This suggests CLASlite's endmembers do not provide enough information to build a reliable classification system using just BG, PV, and NPV. Moreover it indicates a need to explore additional explanatory variables (such as vegetation indices or SAR backscatter) in order to help discriminate regions that have been selectively logged. In the next section, areas that had been selectively logged under the highest intensities within the concession were analysed in an attempt to improve classification accuracy and further clarify the criteria Random Forests used for classification.

Table A1.11 Confusion matrix for objective 2 summarizing results from Random Forests classification using only end-members derived from the CLASlite output at 14384 labelled point locations (late time period) and 14144 labelled points (early time period) of known fate (before and after selective logging). Data were split into 70% training points and 30% validation points in each case. Numbers are representative outputs from a single run of the Random Forest model with the test data. Accuracy percentages are means from 10 model runs and standard errors are given in parentheses. Overall accuracy was 63% (0.2) and 59% (0.1) for the late and early time periods, respectively.

Late		Predicted Class			Early		Predicted Class		
		Logged	Unlogged	User's Accuracy			Logged	Unlogged	User's Accuracy
Actual Class	Logged	1353	790	65.3% (0.7)	Actual Class	Logged	1262	912	57.6% (0.7)
	Unlogged	785	1388	61.0% (0.6)		Unlogged	791	1279	60.7% (0.7)
	Producer's Accuracy	62.4% (0.3)	63.9% (0.5)			Producer's Accuracy	59.7% (0.4)	58.6% (0.2)	

A1.8.4.5 Methods Objective 3: Does accuracy improve when only using pixels of highest logging intensity?

To address this objective we restricted the analysis to pixels from areas with the highest levels of logging intensity occurred. We focused on pixels where 3 or more trees had been removed and, separately, pixels where greater than 20m^3 of wood volume was removed. In all other respects, data analysis methods followed those of objective 2.

A1.8.4.6 Results and Discussion Objective 3

Overall Random Forest classification accuracy of logged and unlogged pixels was around 60 - 70%, irrespective of the time frame the imagery came from (late versus early), and metric of logging intensity (trees versus wood volume) used (Table A1.12). The dataset that only contained pixels where 3 or more trees had been removed had higher accuracies than the dataset that only contained pixels where more than 20m^3 of wood had been removed. We only display a confusion matrix for the trees removed dataset (and not wood volume) because results were similar. Wood volume accuracies were consistently lower by about 10% for each category.

The aim of focusing only on areas that were selectively logged under the highest intensities was based on the assumption that BG, PV, and NPV values would be sufficiently altered and this signal would assist Random Forests in classification. A comparison of accuracies between Table A1.11 and Table A1.12 shows only a modest improvement in classification, from ~60% with all the data (Table A1.11) to ~70% here (Table A1.12). While this result is an improvement, a target of 85% overall accuracy and no class less than 70% accurate is generally regarded a minimum criteria (Thomlinson et al. 1999). Thus, without any additional attributes from tree removal locations (such as vegetation indices or SAR backscatter) the Random Forest classification system falls short of acceptable with just BG, PV, and NPV. In the next section a vegetation index was calculated over the same locations and corresponding time periods to test if Random Forest classification was improved by the additional information.

Table A1.12 Confusion matrix for objective 3 summarizing results from Random Forests classification using only end-members derived from the CLASlite output at 390 labelled point locations (late time period) and 432 labelled points (early time period) of known fate (before and after logging). Data were split into 70% training points and 30% validation points in each case. Numbers are representative outputs from a single run of the Random Forest model with the test data. Accuracy percentages are means from 10 model runs and standard errors are given in parentheses. Overall accuracy was 69.0% (1.1) and 67.1% (1.1) for the late and early time periods, respectively.

Late		Predicted Class			Early		Predicted Class		
		Logged	Unlogged	User's Accuracy			Logged	Unlogged	User's Accuracy
Actual Class	Logged	1353	790	68.4% (2.3)	Actual Class	Logged	1262	912	65.5% (1.7)
	Unlogged	785	1388	69.8% (2.2)		Unlogged	791	1279	69.1% (2.0)
	Producer's Accuracy	68.0% (1.6)	70.6% (1.9)			Producer's Accuracy	67.6% (1.9)	67.1% (1.7)	

A1.8.4.7 Methods Objective 4: would adding NDVI improve the classification accuracy?

The same datasets from objectives 2 and 3 were used in objective 4, but in this step a Normalized Difference Vegetation Index (NDVI) product that had been calculated in ENVI was added. These data were extracted from the corresponding point locations at each of time periods mentioned above. Another run of the Random Forest decision tree classification was performed, this time adding the NDVI predictor to the datasets from objectives 2 and 3.

A1.8.4.8 Results and Discussion Objective 4

In every case the addition of NDVI improved classification accuracy of logged and unlogged points (in both time periods) for all tree removal pixels, only pixels where 3 or more trees were removed, and only pixels where greater than 20m^3 of wood was removed (Tables A1.13-15). The highest classification accuracies came when using all tree cut locations (objective 2 datasets) during the late period (Table A1.13) and the highest wood volume areas (objective 3 data) in the early period (Table A1.15). Finally, a summary table of classification accuracies with and without NDVI is provided (Table A1.16).

It is clear that NDVI can be used to improve the classification accuracy of selectively logged and unlogged pixels, which suggests that additional predictors (potentially from radar data or other vegetation metrics) might increase classification accuracy even further. In the future, radar datasets will be acquired over the Jamari test site in order to apply additional measures of forest structure with the goal of improving classification accuracy. Finally, the Random Forests output provides a scaled measure of variable importance, called the Gini index, which represents how much a model fit decreases with the exclusion of that particular variable. The Gini index for NDVI was roughly 4 times higher than NPV, 5 times higher than PV, and 10 times higher than BG. This result indicates a strong drop in model fit with the exclusion NDVI, an important next step would be an exploration of why NDVI improved classification so dramatically. What additional information was it providing and, in particular, how it might be correlated with measures of PV from CLASlite.

Table A1.13 Confusion matrix for objective 4 summarizing results from Random Forests classification using end-members derived from the CLASlite output and an NDVI product calculated in ENVI at 14384 labelled point locations (late time period) and 14144 labelled points (early time period) of known fate (before and after logging). Data were split into 70% training points and 30% validation points in each case. Numbers are representative outputs from a single run of the Random Forest model with the test data. Accuracy percentages are means from 10 model runs and standard errors are given in parentheses. Overall accuracy was 87% (0.2) and 80% (0.2) for the late and early time periods, respectively.

Late		Predicted Class			Early		Predicted Class		
		Logged	Unlogged	User's Accuracy			Logged	Unlogged	User's Accuracy
Actual Class	Logged	1760	424	80.2% (0.4)	Actual Class	Logged	1566	549	75.7% (0.3)
	Unlogged	129	2003	93.7% (0.2)		Unlogged	262	1867	85.2% (0.3)
Producer's Accuracy		92.8% (0.2)	82.4% (0.4)		Producer's Accuracy		83.5% (0.3)	78.0% (0.2)	

Table A1.14 Confusion matrix for objective 4 summarizing results from Random Forests classification using end-members derived from the CLASlite output and an NDVI product calculated in ENVI at 390 labelled point locations (late time period) and 432 labelled points (early time period) of known fate (before and after logging). Data were split into 70% training points and 30% validation points in each case. Numbers are representative outputs from a single run of the Random Forest model with the test data. Accuracy percentages are means from 10 model runs and standard errors are given in parentheses. Overall accuracy was approximately 86% (1.0) and 87% (1.0) for the late and early time periods, respectively.

Late		Predicted Class			Early		Predicted Class		
		Logged	Unlogged	User's Accuracy			Logged	Unlogged	User's Accuracy
Actual Class	Logged	51	5	86.3% (1.3)	Actual Class	Logged	48	10	85.0% (1.4)
	Unlogged	13	48	85.8% (1.7)		Unlogged	11	61	88.5% (1.1)
Producer's Accuracy		85.8% (1.9)	86.3% (1.5)		Producer's Accuracy		87.8% (1.1)	86.0% (1.3)	

Table A1.15 Confusion matrix for objective 4 summarizing results from Random Forests classification using end-members derived from the CLASlite output and an NDVI product calculated in ENVI at 650 labelled point locations (late time period) and 506 labelled points (early time period) of known fate (before and after logging). Data were split into 70% training points and 30% validation points in each case. Numbers are representative outputs from a single run of the Random Forest model with the test data. Accuracy percentages are means from 10 model runs and standard errors are given in parentheses. Overall accuracy was approximately 79% (1.2) and 90% (0.9) for the late and early time periods, respectively.

Late		Predicted Class			Early		Predicted Class		
		Logged	Unlogged	User's Accuracy			Logged	Unlogged	User's Accuracy
Actual Class	Logged	77	17	77.3% (2.1)	Actual Class	Logged	74	6	89.5% (1.0)
	Unlogged	15	86	81.3% (1.4)		Unlogged	2	70	91.9% (1.0)
	Producer's Accuracy	80.2% (1.5)	78.7% (1.8)			Producer's Accuracy	91.8% (0.9)	89.6% (1.2)	

A1.8.5 Conclusions

Taken together, our results show that CLASlite cannot detect selective logging at the levels that occurred at the Jamari test site (approximately 1 tree ha^{-1}). In particular, we showed that CLASlite did not identify a single pixel as degraded. This is certainly a drawback of the built-in decision tree CLASlite uses to label pixels as degraded, wherein a spike in NPV and a drop in PV are required to signal degradation. This is precisely why the early time period was added to our analyses, in an effort to detect a drop in canopy cover (PV) and an increase in slash (NPV) from the active logging. The fact that our classification from objective 2, where we used Random Forests to build a more complex decision tree than CLASlite, still underperformed (with accuracies around 60%) demonstrates the difficulties associated with detecting more subtle changes in forest attributes with Landsat-scale imagery. In most cases imagery from this early time period introduced more error (as some points were labelled logged before the selective logging had occurred in that year) and resulted in lower classification accuracies.

Another factor that probably impacted CLASlite's abilities to detecting degradation at Jamari was the final filtering process used (where CLASlite requires 5 other degraded pixels to be present in a 7x7 moving window in order to be classified as degraded). This is meant to minimize classifying unlogged pixels as degraded, since any anthropogenic disturbance event does not occur in isolation and should result in other associated disturbances. It may be that 5 pixels in a 7x7 moving window is too conservative to identify lower levels of selective logging. An option to adjust these settings within the software would prove useful, as this would enable users to fine-tune the output to correspond with their knowledge of the site.

Table A1.16 Summary table for accuracy results from objectives 2, 3, and 4 displaying the highest classification accuracies from either late or early time period. Underlined percentages were highest for the early time period.

With NDVI	User		Producer		Without NDVI	User		Producer	
	logged	unlogged	logged	unlogged		logged	unlogged	logged	unlogged
all pixels	80.2%	93.7%	92.8%	82.4%		65.3%	61.0%	62.4%	63.9%
≥ 3 trees	<u>85.0%</u>	<u>88.5%</u>	<u>87.8%</u>	<u>86.0%</u>		68.4%	69.8%	68.0%	70.6%
≥ 20m^3 volume	<u>89.5%</u>	<u>91.9%</u>	<u>91.8%</u>	<u>89.6%</u>		61.4%	59.4%	60.8%	60.7%

It is also important to recognize that the levels of selective logging generally reported equate to an average measure of intensity (i.e., number of trees removed over an entire management unit or concession). The arrangement of surface disturbances can greatly impact this measure of intensity. As an extreme example, consider rates of trees removal if ¼ of a management unit were deforested and the remaining ¾ left intact. Indeed, our test site was selectively logged at roughly 1 tree ha^{-1}, but there are dozens of regions where the intensity approaches 10-15 trees removed ha^{-1} and many other regions where no trees were removed within a hectare. CLASlite has been extensively validated, but with the lowest harvest intensities around 3-6 trees ha^{-1} (Asner et al. 2004b, a; Broadbent et al. 2006). Even though CLASlite could not detect degradation in Jamari, the fact that our results from objective 3, where we focused on areas of the highest selective logging intensities and used Random Forests to build a more complex decision tree than CLASlite, still underperformed (with accuracies around 70%) demonstrates how difficult detecting this type of disturbance can be.

Finally, our analyses used a before-after approach, where we analysed the same pixels 2 years before and in the year following logging. It would be useful to test other methods for assessing CLASlite's performance, such as cut locations after logging and randomly selected unlogged points from the same time period. This process would only require a single scene and might reduce the levels of uncertainty in the analysis. Moreover, now that a combined dataset of CLASlite outputs and NDVI increased classification above an acceptable level, an important next step will be to develop a predictive map over Jamari from Random Forests.

Overall, without being able to adjust CLASlite's internal decision tree system or modify the final degraded pixel filtering process, CLASlite should be thought of as an additional tool that researchers can use to derive forest attributes (BG, PV, and NPV). The need for a more sophisticated classification algorithm (like Random Forests) was required to leverage the information content, but still fell short of acceptable (around 60% accurate). Assuming we are able to identify additional metrics (beyond NDVI) that increase classification accuracy, this would produce a stronger paper overall- rather than simply criticizing the limitations of CLASlite (i.e., offer a method for improvement). In addition, with the ASTER datasets being made freely available (these are compatible with CLASlite), there may be additional improvement in performance with these data sets. In particular, the first 3 ASTER bands (near infra-red) are at 15m spatial resolution. We will be exploring these datasets in the weeks to come.

APPENDIX

A2

Supplementary Material to Chapter 4

“Detecting tropical selective logging with SAR data requires a time series approach “

Table A2.1 Acquisition dates for Sentinel-1, RDARSAT-2, and PALSAR-2 imagery used in classificication of selective logging.

Sensor	Dates
Sentinel-1	2016-02-15
	2016-09-27
	2016-09-30
	2017-08-20
	2017-08-22
	2017-08-29
	2018-08-15
	2018-08-29
RADARSAT-2	2012-08-19
PALSAR-2	2016-09-06
	2017-07-05

Table A2.2 Cross-validation results for the number of of trees (*k*) and the number of variables to use at each node (*m*) that minimized the out-of-bag error rate on each training dataset.

Sensor	nTree (*k*)	mTry (*m*)
Sentinel-1	600	1
RADARSAT-2	700	5
PALSAR-2	700	4
Sentinel-1 subset	800	1

A2.1 Feature selection

The Boruta package, version 6.0.0, was used to assess feature importance and all were deemed to contribute significantly to classification.

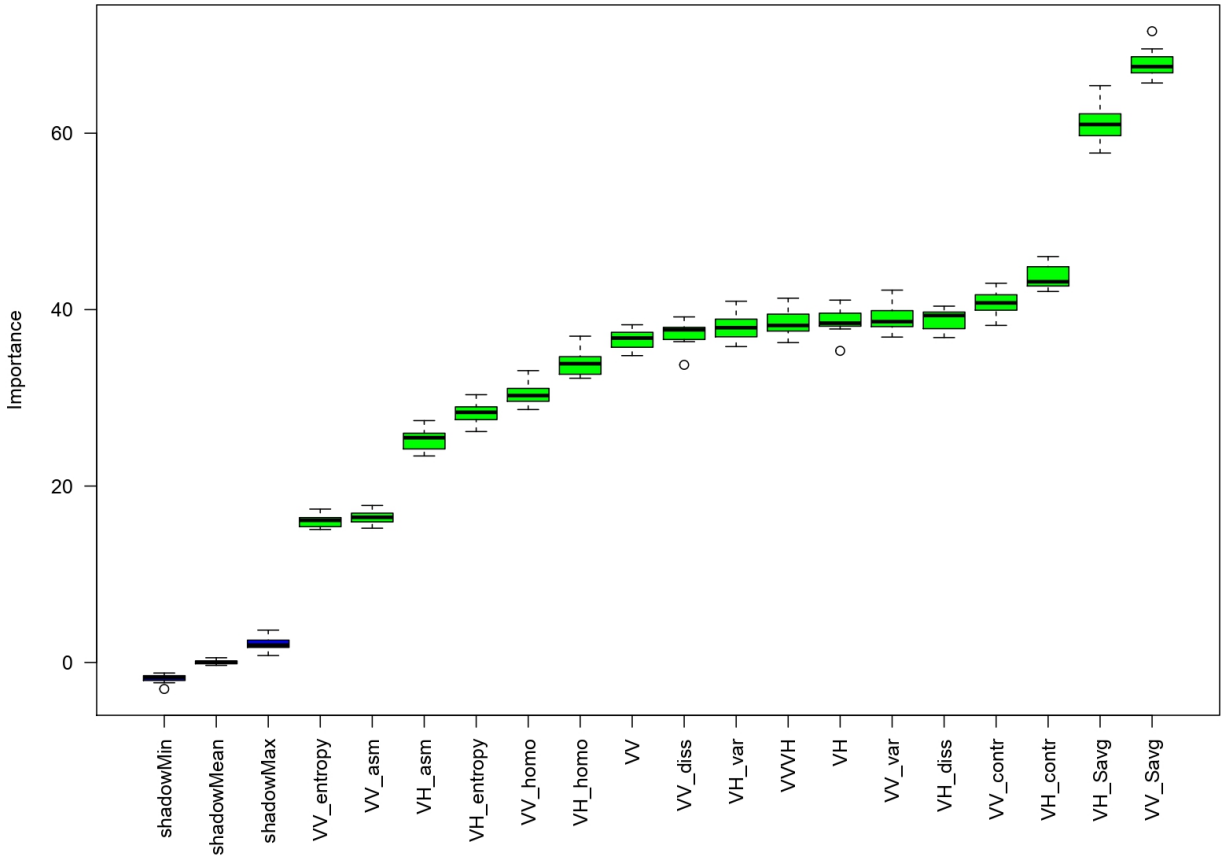


Figure A2.1 Sentinel-1 variable importance measures from the Boruta package. Bars with green fill indicate significant contribution to classification, while yellow and red (not shown) indicate marginal and non-significant contributions, respectively. Thus, all variables were deemed important.

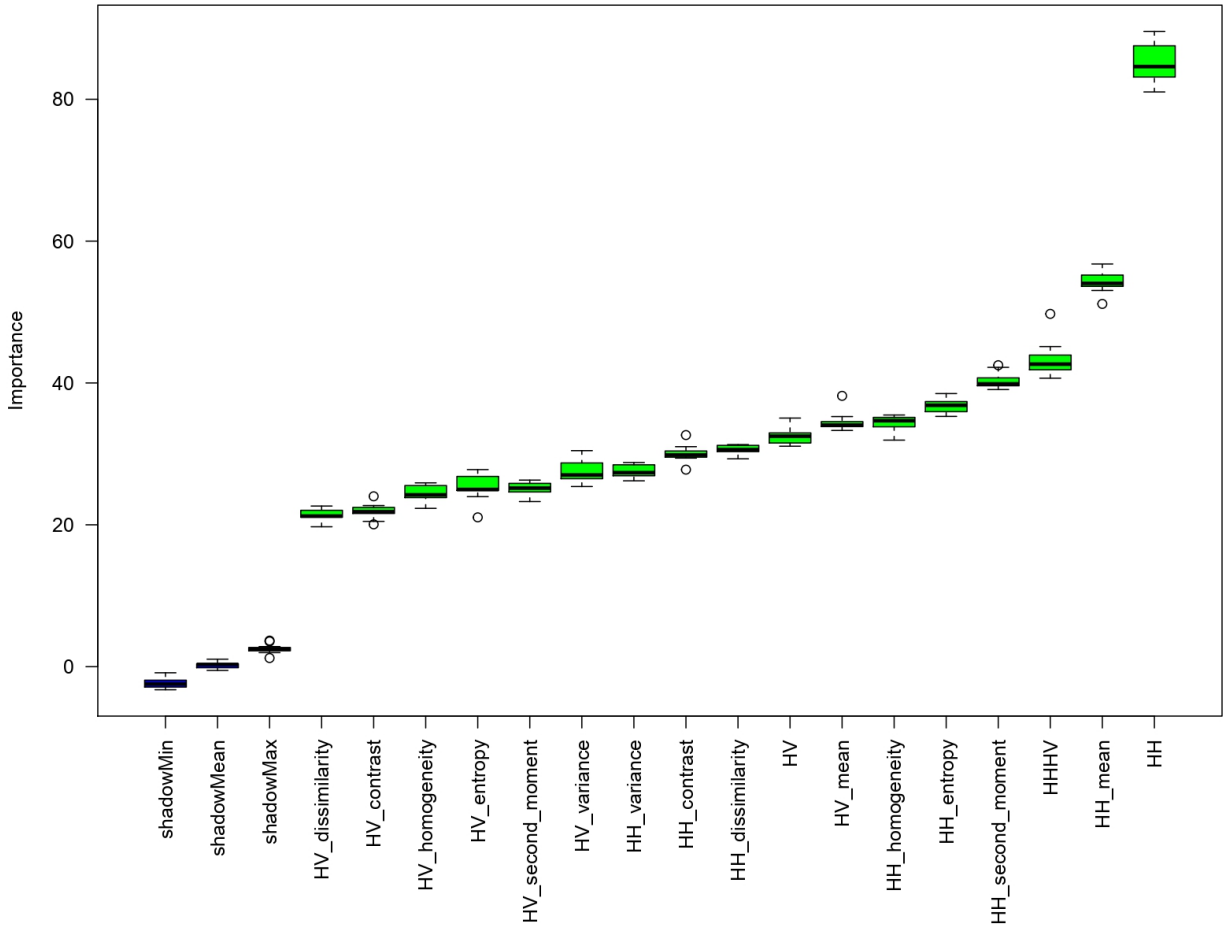


Figure A2.2 PALSAR-2 variable importance measures from the Boruta package. Bars with green fill indicate significant contribution to classification, while yellow and red (not shown) indicate marginal and non-significant contributions, respectively. Thus, all variables were deemed important.

.

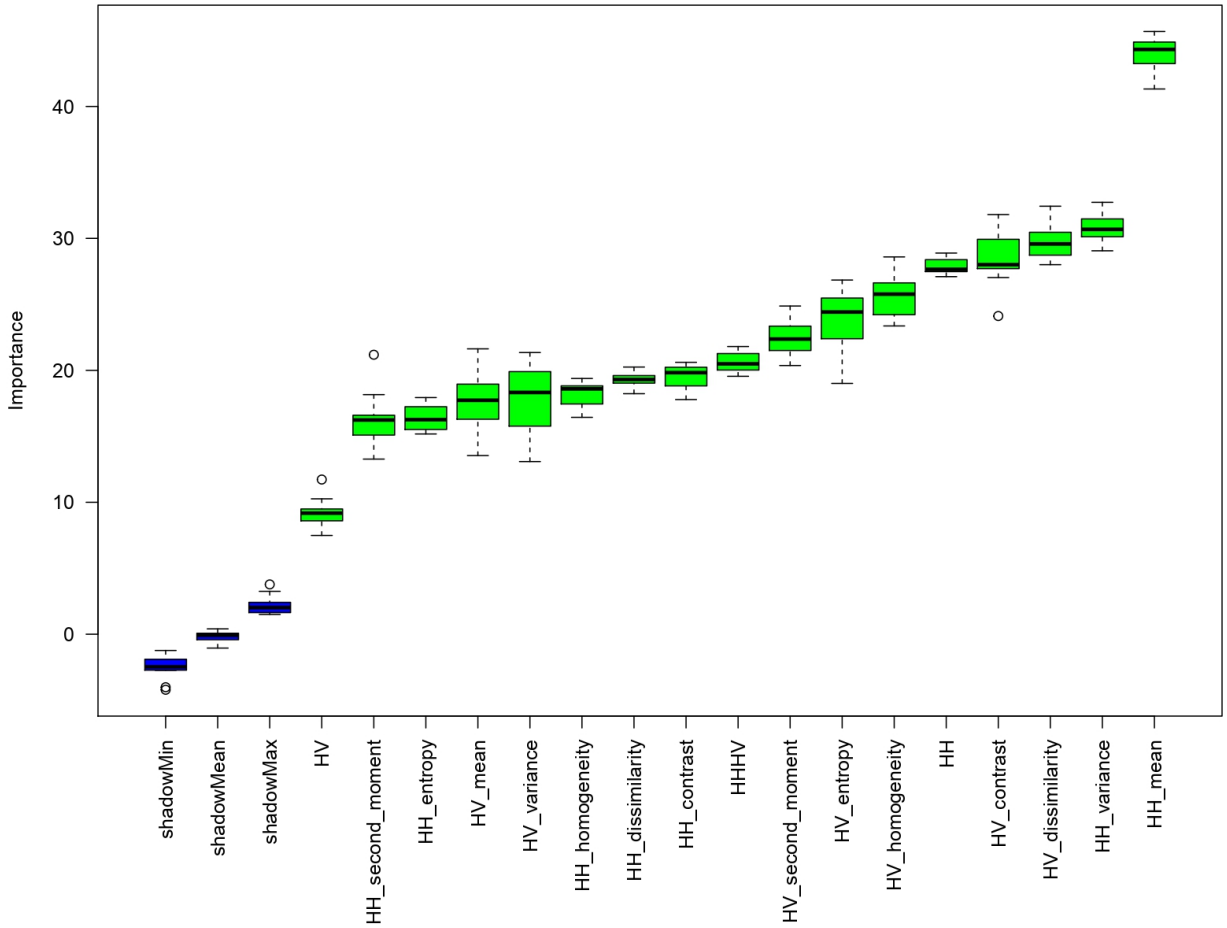


Figure A2.3 RADARSAT-2 variable importance measures from the Boruta package. Bars with green fill indicate significant contribution to classification, while yellow and red (not shown) indicate marginal and non-significant contributions, respectively. Thus, all variables were deemed important.

.

A2.2 Confusion Matrices from SAR classification

Table A2.3. Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at three study areas in the Brazilian Amazon, derived from Sentinel-1 data. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 13,401). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Sentinel -1				***T* = 0.71**
OA: 64.3%				
κ: 0.25		Reference Class		
FDR: 44.6%		Logged	Unlogged	Commission
DR: 53.5%				Error (%)
Predicted Class	Logged	2861	2299	44.6
	Unlogged	2489	5752	30.2
	Omission Error (%)	46.5	28.6	

Table A2.4 Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at two study areas in the Brazilian Amazon, derived from RADARSAT-2 data. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 4,903). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

RADARSAT-2				***T* = 0.24**
OA: 75.6%				
κ: 0.12		Reference Class		
FDR: 75.0%		Logged	Unlogged	Commission
DR: 27.4%				Error (%)
Predicted Class	Logged	211	643	75.0
	Unlogged	559	3490	13.8
	Omission Error (%)	72.6	15.4	

Table A2.5 Confusion matrix summarizing Random Forest (RF) model classifications of logged and unlogged observations at two study areas in the Brazilian Amazon, derived from PALSAR-2 data. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 4,122). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

PALSAR-2				***T* = 0.36**
OA: 64.2%				
κ: 0.14		Reference Class		
FDR: 62.4%		Logged	Unlogged	Commission
DR: 40.5%				Error (%)
Predicted Class	Logged	471	782	62.4
	Unlogged	692	2177	24.1
	Omission Error (%)	59.5	26.4	

Table A2.6 Confusion matrix summarizing Random Forest (RF) model classifications of the most intensively logged and unlogged observations at three study areas in the Brazilian Amazon, derived from Sentinel-1 data. Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 7,431). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Sentinel-1 High subset				***T* = 0.67**
OA: 88.5%				
κ: 0.32		Reference Class		
FDR: 55.9%		Logged	Unlogged	Commission
DR: 33.2%				Error (%)
Predicted Class	Logged	261	331	55.9
	Unlogged	524	6315	7.7
	Omission Error (%)	66.8	5.0	

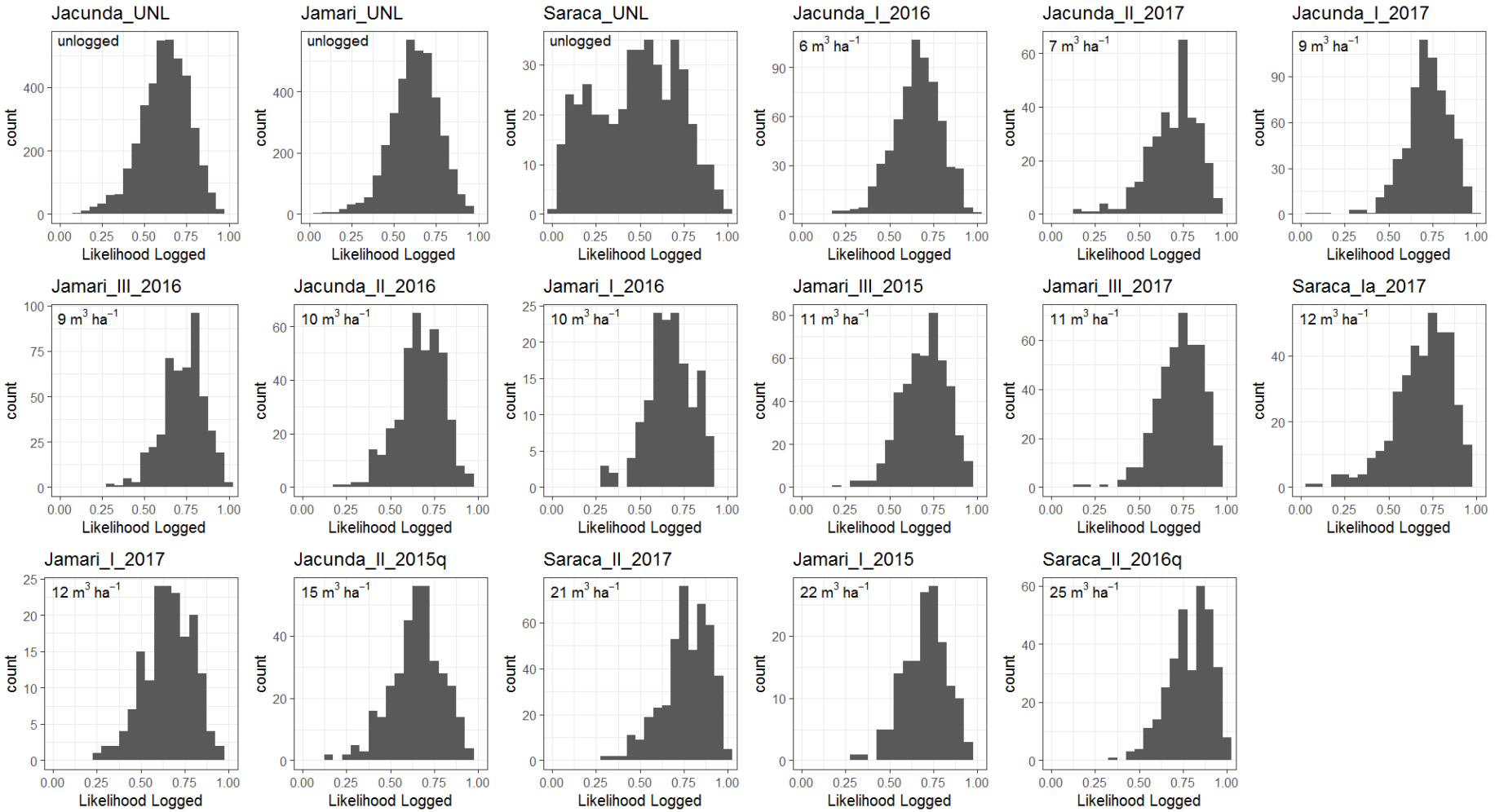


Figure A2.4 Histograms of the likelihoods (the proportion of votes for each class) for each observation with the full Sentinel-1 dataset (separated by FMU). The logging intensity is listed in the upper left of each panel.

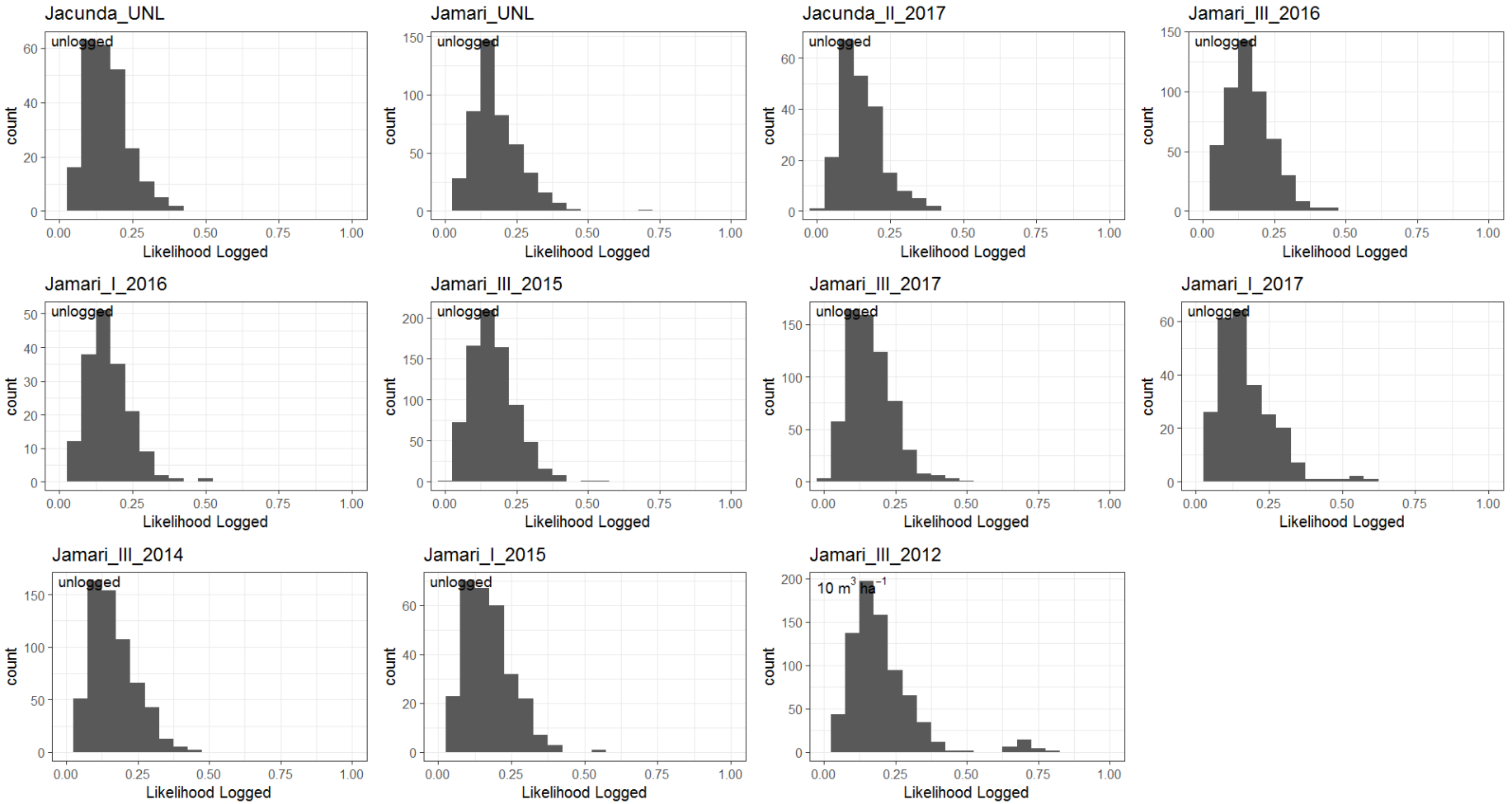


Figure A2.5 Histograms of the likelihoods (the proportion of votes for each class) for each observation with the RADARSAT-2 dataset (separated by FMU). The logging intensity is listed in the upper left of each panel.

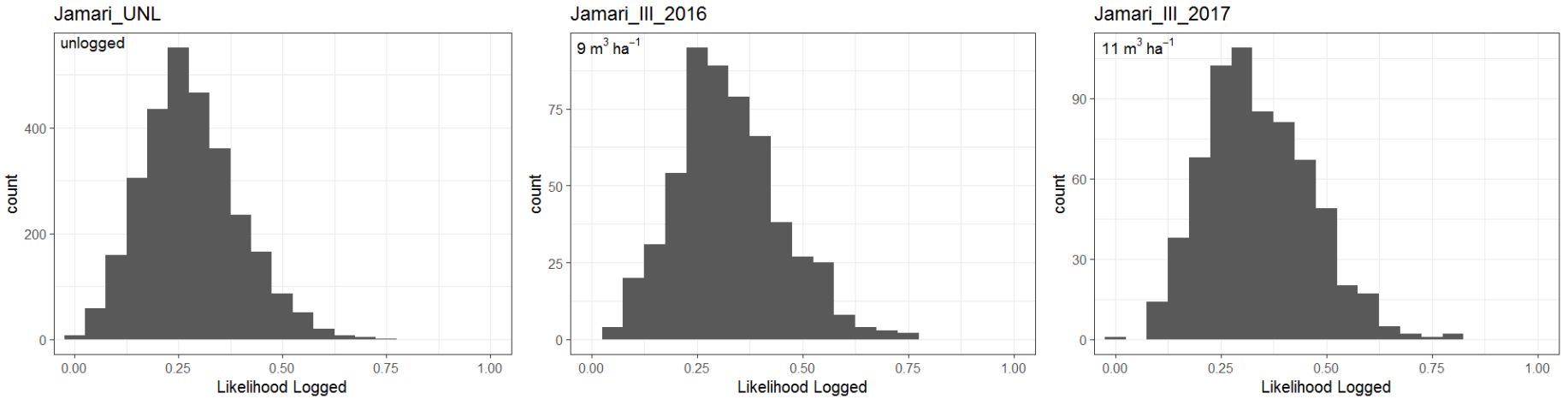


Figure A2.6 Histograms of the likelihoods (the proportion of votes for each class) for each observation with the PALSAR-2 dataset (separated by FMU). The logging intensity is listed in the upper left of each panel.

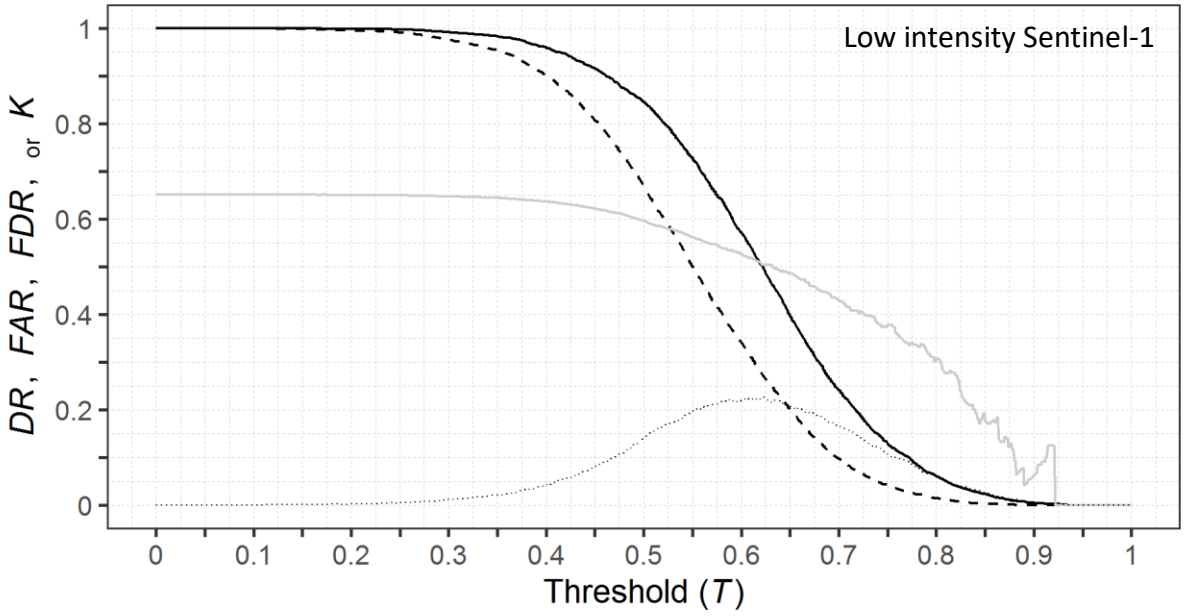


Figure A2.7 Random Forest model performance across all threshold values (*T*) for classification with Sentinel-1 (the same subset of low-intensity logging sites with that was used with the PALSAR-2). The Detection Rate (DR) and False Alarm Rate (FAR) are the solid and dashed black lines, respectively. Also shown are the corresponding values of the False Discovery Rate (FDR) and Cohen's kappa (*k*) as solid and dashed grey lines, respectively.

Table A2.7 Confusion matrix summarizing Random Forest (RF) model classifications of low-intensity logged and unlogged observations at two study areas in the Brazilian Amazon, derived from Sentinel-1 data (the same subset of sites used in the PALSAR-2 analyses). Data were split into 75% training and 25% validation. Matrix numbers are pixel counts with the validation data (n = 9,447). The classification threshold (*T*) for RF models was set to maximize Cohen's kappa. The corresponding values for overall accuracy (OA), the false discovery rate (FDR), and the detection rate (DR) are provided against the validation dataset.

Sentinel-1 Low subset				***T* = 0.62**
OA: 64.7%				
κ: 0.23		Reference Class		
FDR: 50.6%		Logged	Unlogged	Commission
DR: 50.3%				Error (%)
Predicted Class	Logged	1659	1700	50.6
	Unlogged	1638	4450	26.9
	Omission Error (%)	49.7	27.6	

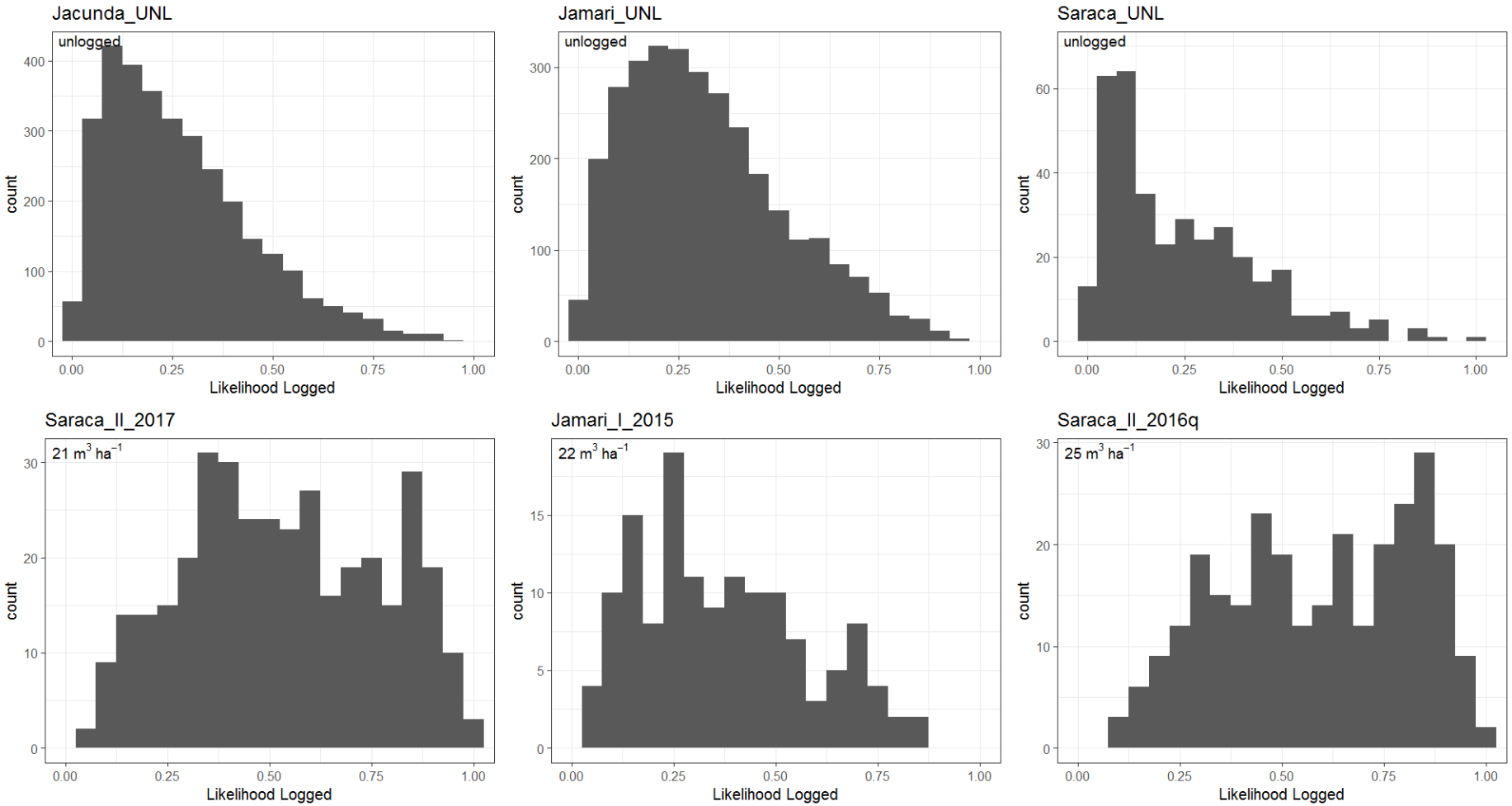


Figure A2.8 Histograms of the likelihoods (the proportion of votes for each class) for each observation with the subset Sentinel-1 dataset (separated by FMU). The logging intensity is listed in the upper left of each panel.

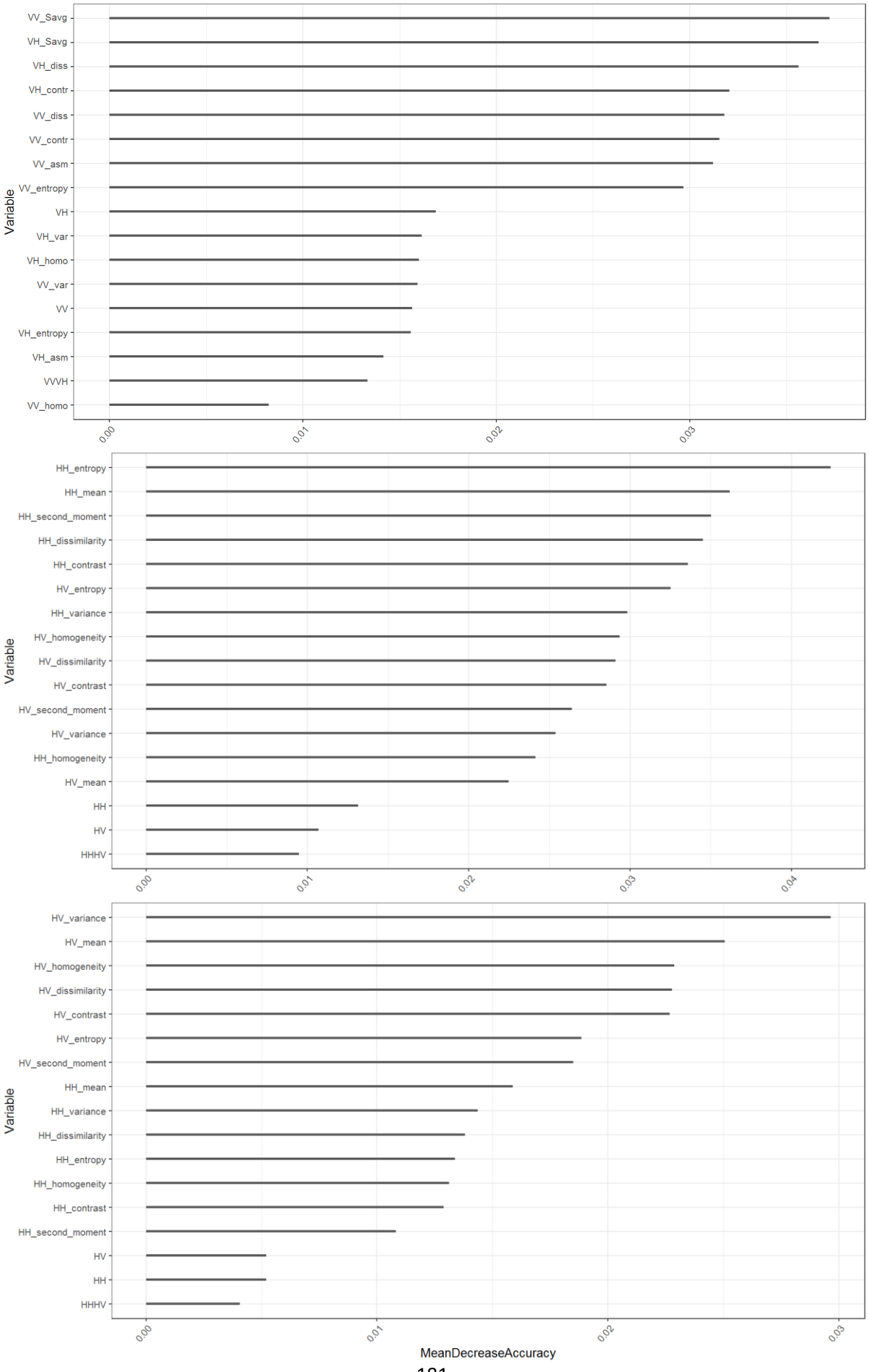


Figure A2.9 Random Forest model variable importance for Sentinel-1 (top), RADARSAT-2 (middle), and PALSAR-2 (bottom).

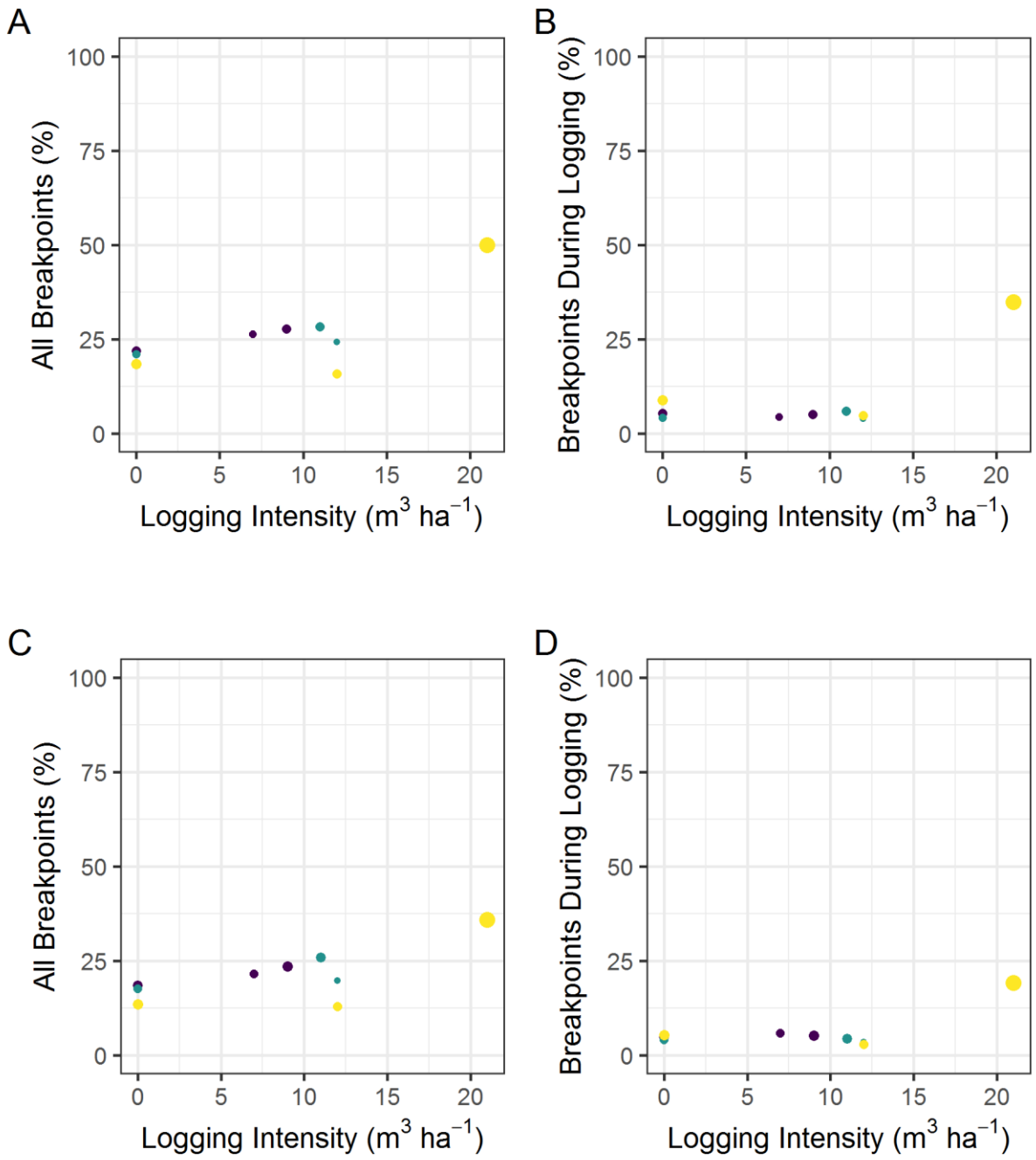


Figure A2.10 The relationship between the proportion of observation within a Forest Management Unit (FMU) that had a breakpoint identified within its Sentinel-1 sum average texture and dissimilarity measure time series and the logging intensity of the FMU for VH (bottom row). The proportion of all observations (A and C) and the proportion that had a breakpoint that coincided with the logging season (C and D) are shown separately. The circle size corresponds to number of observations at each FMU and yellow, green, and purple colors represent the Saraca, Jamari, and Jacunda sites, respectively.

Equation A2.1

$$Texture \sum Average = \sum_{i=2}^{2N_g} ip_{x+y}(i)$$

N_g: Number of distinct gray levels in quantized image

x and y are the coordinates (row and column) of an entry in the co-occurance matrix

$p_{x+y}(i)$ is the probability of co-occurances matrix coordinates summing to x+y

APPENDIX

A3

Supplementary Material to Chapter 5

“Combining optical and SAR data for monitoring tropical selective logging“

A3.1 Summary of satellite imagery

Table A3.1 Satellite imagery acquisition dates over the forest management units (FMU) in the Brazilian Amazon used in Random Forest classification of selective logging.

FMU	Landsat 8 PathRow_Date	Sentinel-1 Date
Jacunda_I_2016	232066_20160920	20161012
Jacunda_I_2017	232066_20170907	20170901
Jacunda_II_2016	232066_20160920	20161012
Jacunda_II_2017	232066_20170907	20170913
Jacunda_UNL	232066_20160803	20160930
	232066_20170806	20170901
Jamari_I_2016	232066_20160819	20160930
Jamari_I_2017	232066_20170923	20170925
Jamari_III_2016	232066_20160803	20161012
Jamari_III_2017	232066_20170907	20170913
Jamari_UNL	232066_20160803	20161012
	232066_20170907	20170925
Saraca_Ia_2017	229061_20171105	20170927
Saraca_II_2016	228061_20161111	20170822
Saraca_II_2017	228061_20170911	20170822
Saraca_UNL	228061_20161111	20161002
	228061_20170911	20170927

A3.2 Summary of raw inputs to RF models

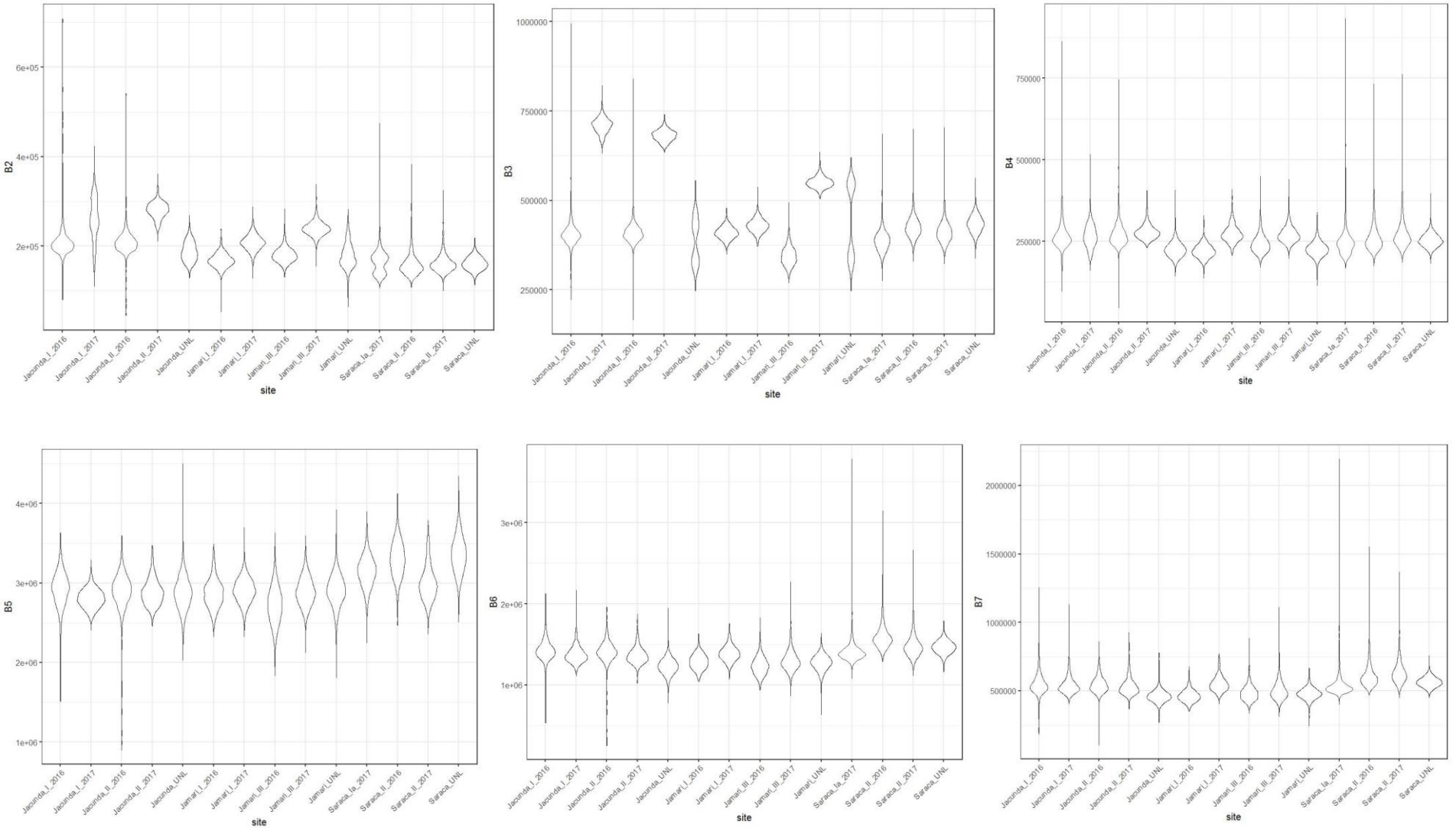


Figure A4.1 Violin plot of Landsat input variables used for GLCM calculations

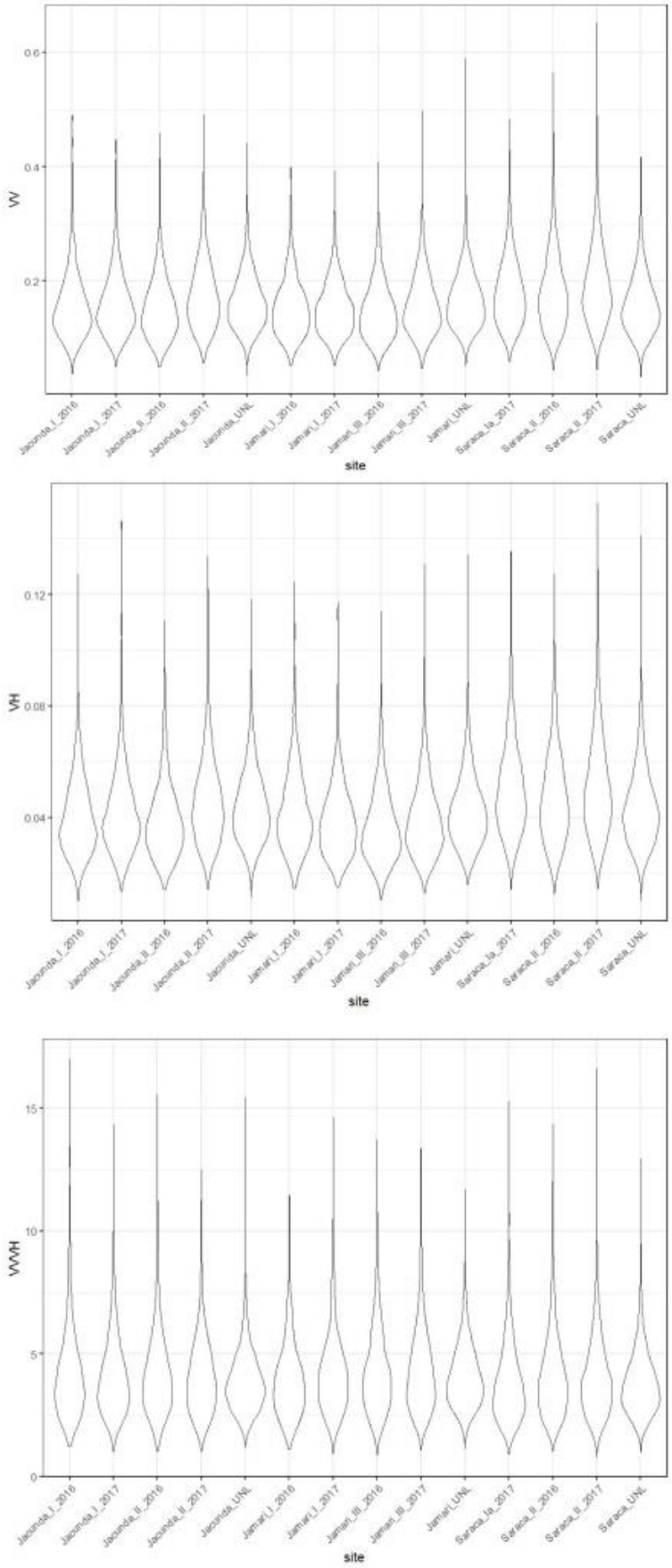


Figure A4.2 Violin plot of Sentinel-1 input variables used for GLCM calculations

A3.3 Additional figures

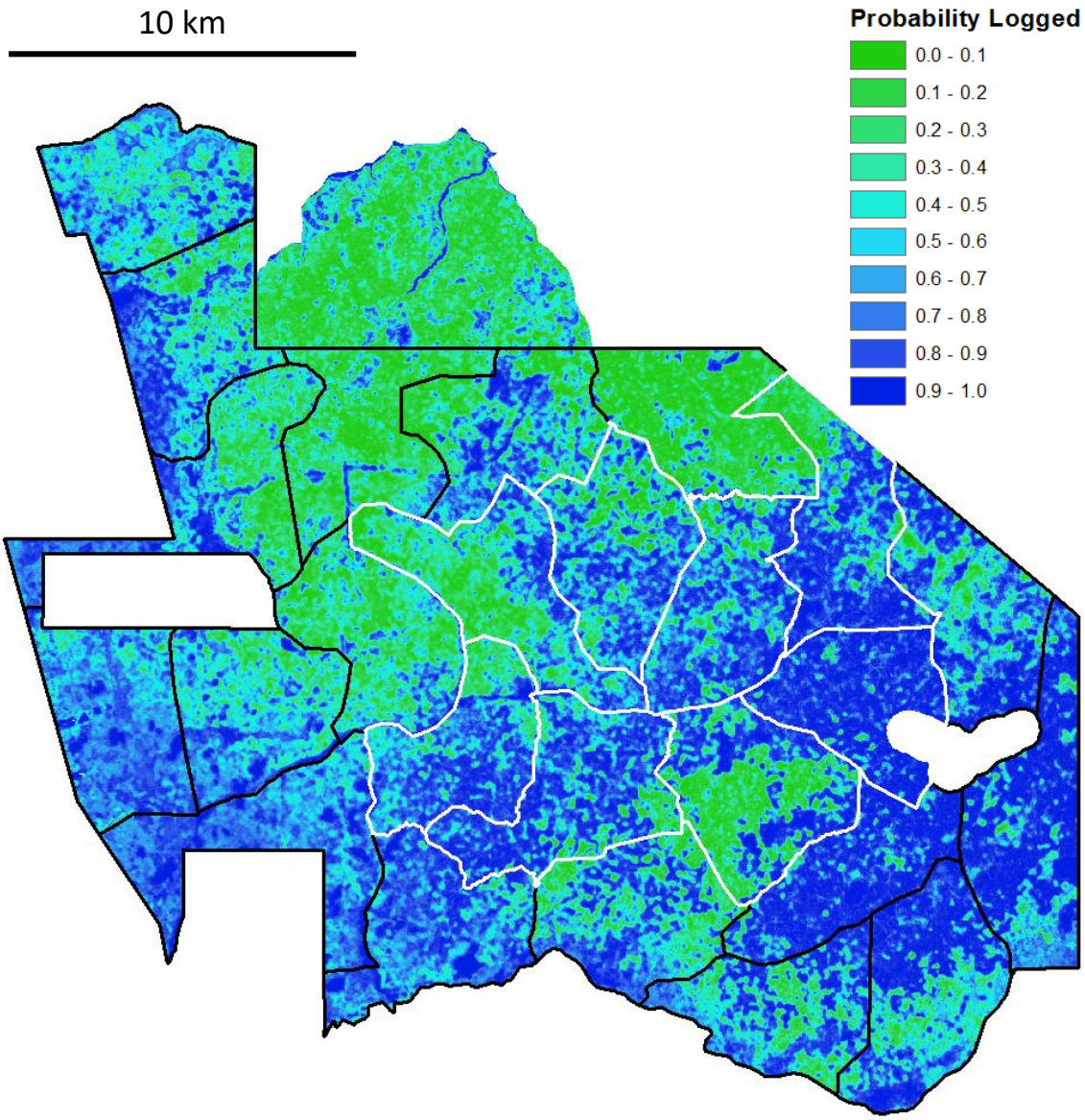


Figure A4.1 Map of selective logging probability for the Jamari region using the Landsat 8 model. Landsat data are from 2017-09-07. The forest management units that have been logged are bounded in white and those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black).

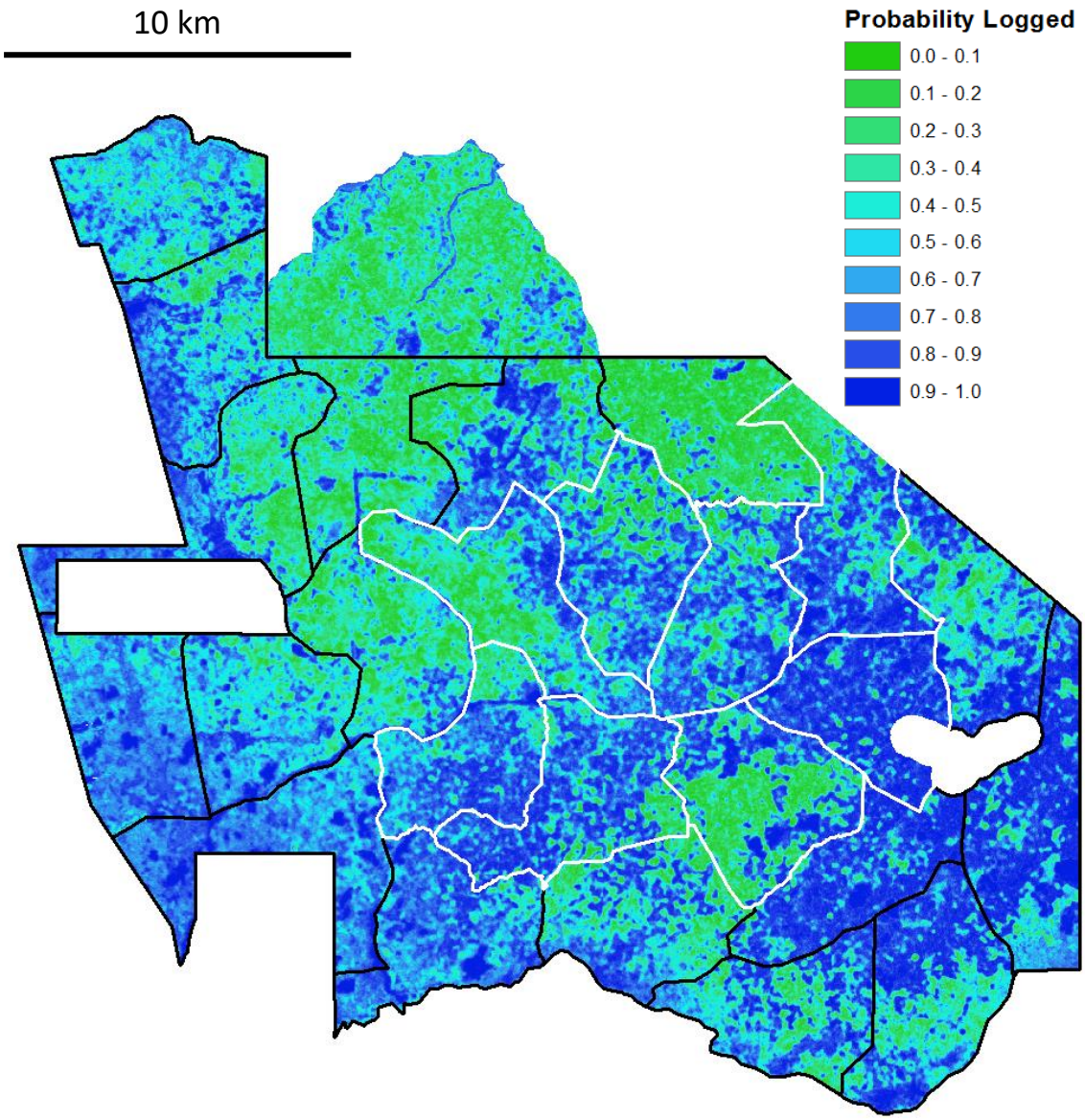


Figure A4.1 Map of selective logging probability for the Jamari region using the combined Landsat 8 and Sentinel-1 model. Landsat data are from 2017-09-07 and the Sentinel-1 data are from 2017-09-25. The forest management units that have been logged are bounded in white and those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black).

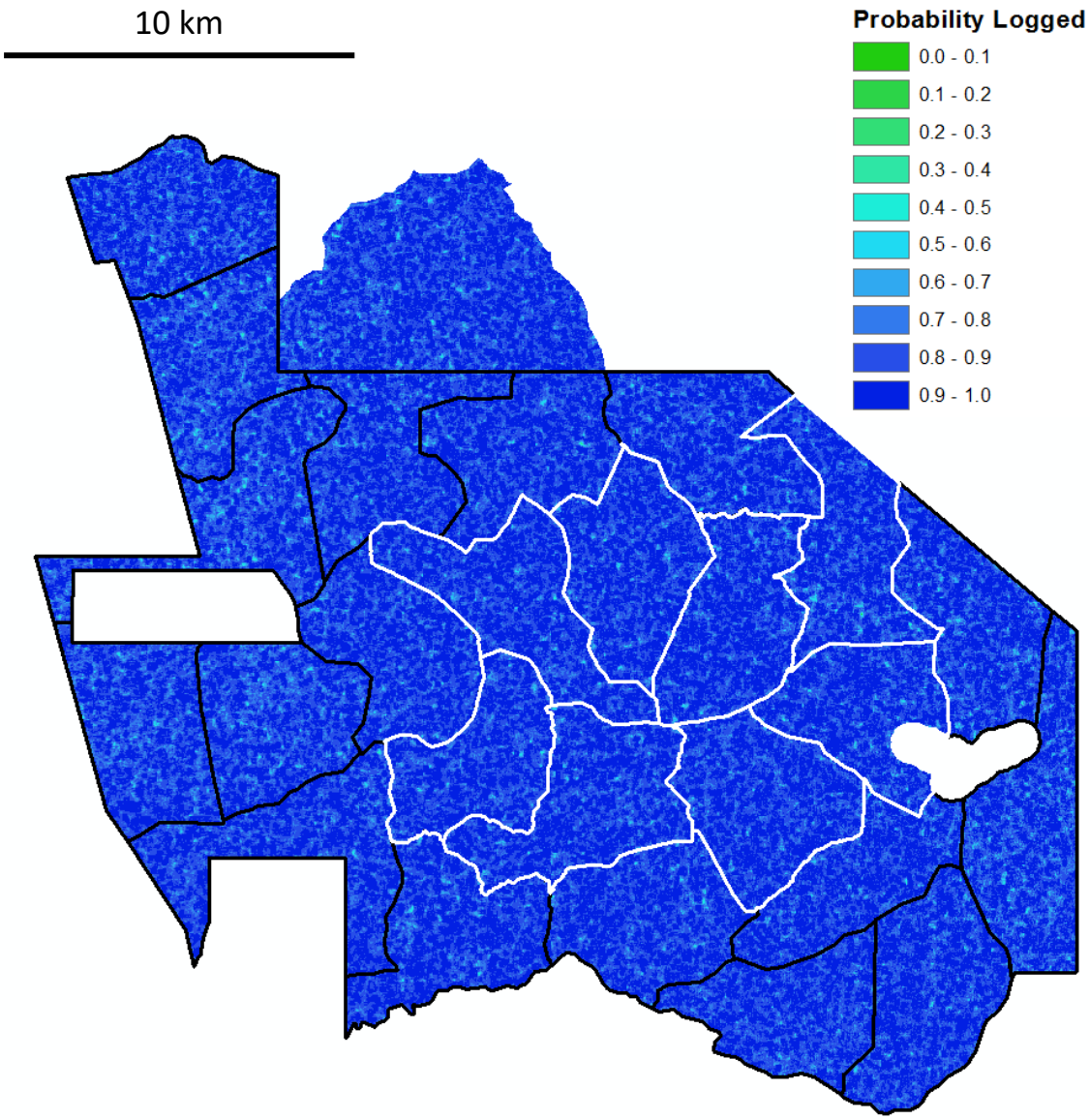
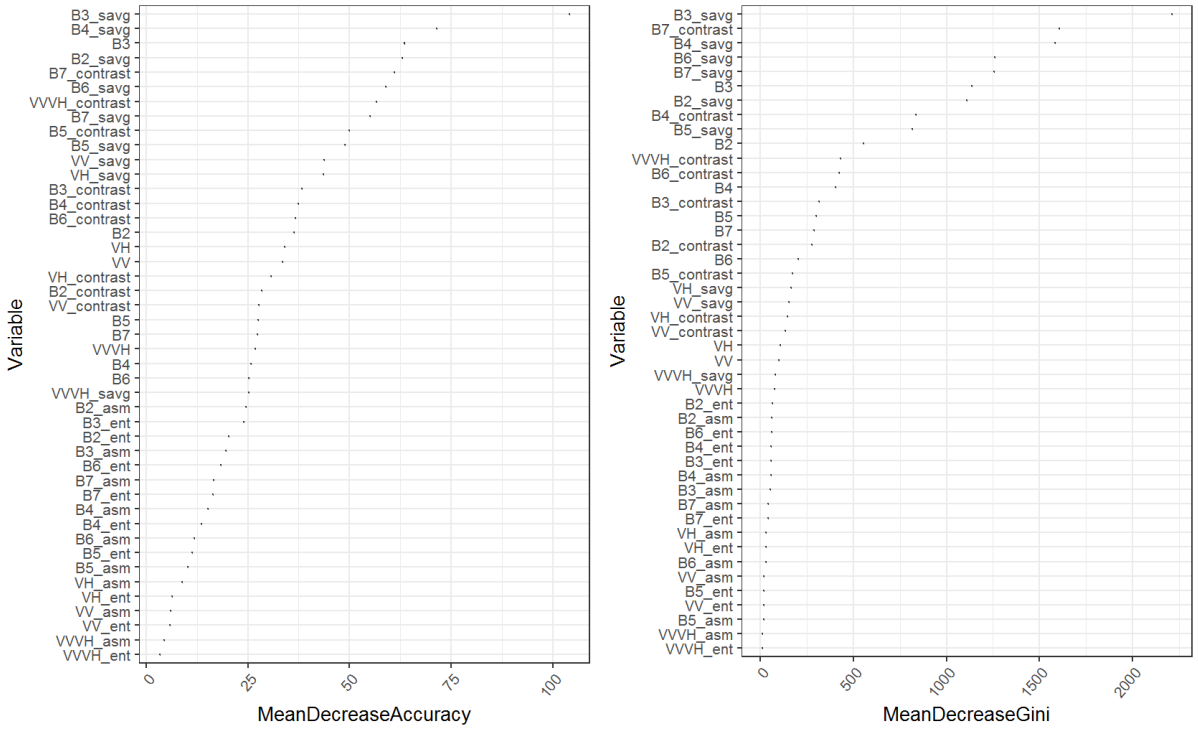


Figure A4.3 Map of selective logging probability for the Jamari region using the Sentinel-1 model from 2017-09-25. The forest management units that have been logged are bounded in white and those yet to be logged are bounded in black. The reserve area that remained unlogged is along the top and outside the boundary of the logging concession (i.e. not outlined in black).


Variable
B3_savg
B4_savg
B3
B2_savg
B7_contrast
B6_savg
VVVH_contrast
B7_savg
B5_contrast
B5_savg
VV_savg
VH_savg
B3_contrast
B4_contrast
B6_contrast
B2
VH
VV
VH_contrast
B2_contrast
VV_contrast
B5
B7
VVVH
B4
B6
VVVH_savg
B2_asm
B3_ent
B2_ent
B3_asm
B6_ent
B7_asm
B7_ent
B4_asm
B4_ent
B6_asm
B5_ent
B5_asm
VH_asm
VH_ent
VV_asm
VV_ent
VVVH_asm
VVVH_ent
0
25
50
75
100
MeanDecreaseAccuracy
Variable
B3_savg
B7_contrast
B4_savg
B6_savg
B7_savg
B3
B2_savg
B4_contrast
B5_savg
B2
VVVH_contrast
B6_contrast
B4
B3_contrast
B5
B7
B2_contrast
B6
B5_contrast
VH_savg
VV_savg
VH_contrast
VV_contrast
VH
VV
VVVH_savg
VVVH
B2_ent
B2_asm
B6_ent
B4_ent
B3_ent
B4_asm
B3_asm
B7_asm
B7_ent
VH_asm
VH_ent
B6_asm
VV_asm
B5_ent
VV_ent
B5_asm
VVVH_asm
VVVH_ent
0
500
1000
1500
2000
MeanDecreaseGini

APPENDIX

A4

Supplementary Material to Chapter 6

“Mapping the rapid expansion of selective logging in the south-west Brazilian Amazon“

A4.1 Summary of Landsat imagery

Table A4.1 Image acquisition dates from Landsat 5 (pre-2013) and Landsat 8 (post-2011) over the study regions

Site	PathRow_Date
Cikel_2011	223063_20060724
	223063_20070913
	223063_20081001
	223063_20090817
Jacunda_I_2016	232066_20050906
	232066_20060808
	232066_20070912
	232066_20080930
	232066_20091003
	232066_20100718
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160920
Jacunda_I_2017	232066_20050906
	232066_20060808
	232066_20070912
	232066_20080930
	232066_20091003
	232066_20100718
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160718
	232066_20170907
Jacunda_II_2015	232066_20050906
	232066_20060808
	232066_20070827
	232066_20080930
	232066_20090816
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918

Jacunda_II_2016	232066_20050906
	232066_20060808
	232066_20070912
	232066_20080930
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160920
Jacunda_II_2017	232066_20050906
	232066_20060824
	232066_20070912
	232066_20080930
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20170907
Jacunda_UNL	232066_20050906
	232066_20060808
	232066_20070912
	232066_20080728
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130811
	232066_20140814
	232066_20150918
	232066_20160803
	232066_20170806
Jamari_I_2015	232066_20060925
	232066_20070912
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830

	232066_20150918
Jamari_I_2016	232066_20060808
	232066_20070912
	232066_20080728
	232066_20090731
	232066_20100819
	232066_20110907
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160819
Jamari_I_2017	232066_20060808
	232066_20070912
	232066_20080728
	232066_20091003
	232066_20100819
	232066_20110907
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160819
	232066_20170923
Jamari_III_2011	232066_20060925
	232066_20070912
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20111009
Jamari_III_2012	232066_20060925
	232066_20070912
	232066_20080829
	232066_20090816
	232066_20100819
	232066_20111009
Jamari_III_2013	232066_20060925
	232066_20070912
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20111009
	232066_20130827

Jamari_III_2014	232066_20060925
	232066_20070912
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
Jamari_III_2015	232066_20060925
	232066_20070827
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
Jamari_III_2016	232066_20050720
	232066_20060925
	232066_20070912
	232066_20080829
	232066_20091003
	232066_20100819
	232066_20111009
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160803
Jamari_III_2017	232066_20060925
	232066_20070827
	232066_20081016
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160803
	232066_20170907
Jamari_UNL	232066_20060925
	232066_20070912

	232066_20080829
	232066_20091003
	232066_20100819
	232066_20110806
	232066_20130827
	232066_20140830
	232066_20150918
	232066_20160803
	232066_20170907
Saraca_Ia_2017	228061_20050910
	228061_20060913
	228061_20071002
	228061_20091023
	228061_20111029
	229061_20080909
	229061_20131212
	229061_20141215
	229061_20150929
	229061_20161017
	229061_20171105
Saraca_II_2016	228061_20051028
	228061_20060828
	228061_20071002
	228061_20080801
	228061_20091007
	228061_20111029
	228061_20130916
	228061_20140903
	228061_20151008
	228061_20161111
Saraca_II_2017	228061_20051028
	228061_20060828
	228061_20071002
	228061_20080801
	228061_20091007
	228061_20111029
	228061_20130916
	228061_20140903
	228061_20150906
	228061_20160924
	228061_20170911
Saraca_UNL	228061_20050926

228061_20071002
228061_20080801
228061_20091007
228061_20110826
228061_20140903
228061_20151008
228061_20161111
228061_20170911

A4.2 Spectral unmixing and spatial normalization

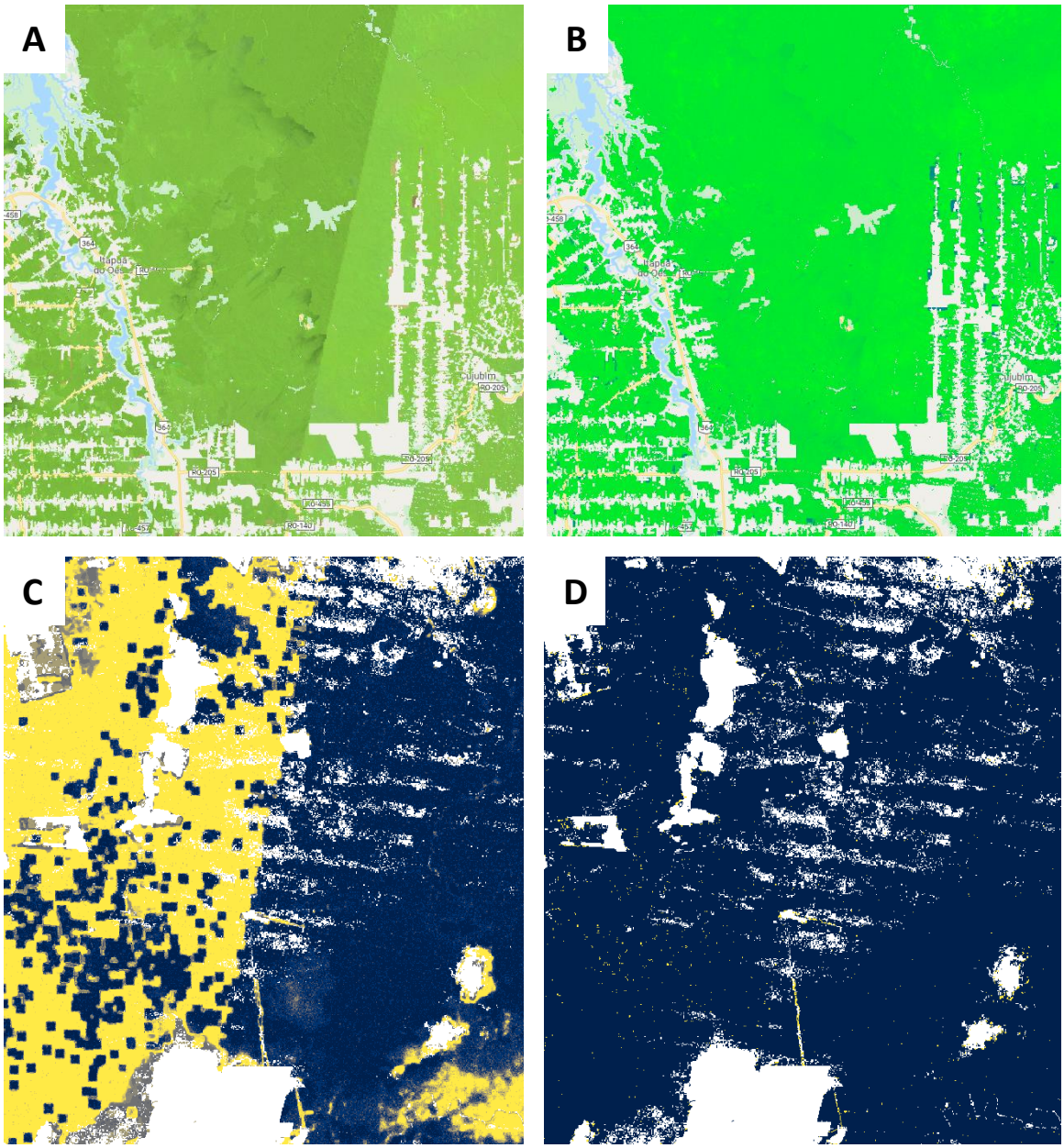


Figure A4.1 Examples of abrupt changes in Surface Reflectance (RGB of Landsat 5 bands 5,4,3) in adjacent Landsat paths (A) and the resulting impact on spectral unmixing (B). While more subtle than surface reflectance the resulting predictions of logging along abrupt changes were impacted, resulting in entire swaths of mosaics being predicted logged (C). The spatial normalization removed abrupt changes across adjacent paths and did not impact mode predictions (D).

A4.3 Summary of GEE mosaic dates

Table A4.2 Start and stop dates for the annual mosaic creation for Rondônia 2000-2019 and the number of days within the window (n Days). Landsat 5, 7 and 8 are TM, ETM, and OLI, respectively.

Year	startDate	stopDate	n Days	Landsat
2000	2000-07-01	2000-08-25	55	TM
2001	2001-07-01	2001-08-20	50	TM
2002	2002-06-15	2002-08-21	67	TM
2003	2003-06-15	2003-08-15	61	TM
2004	2004-06-20	2004-08-05	46	TM
2005	2005-07-01	2005-08-10	40	TM
2006	2006-07-01	2006-08-15	45	TM
2007	2007-07-01	2007-08-15	45	TM
2008	2008-07-01	2008-08-23	53	TM
2009	2009-07-01	2009-08-31	61	TM
2010	2010-07-01	2010-08-10	40	TM
2011	2011-07-01	2011-09-01	62	TM
2012	2012-07-01	2012-09-01	62	ETM
2013	2013-07-01	2013-09-12	73	OLI
2014	2014-07-01	2014-09-11	72	OLI
2015	2015-07-01	2015-09-20	81	OLI
2016	2016-07-01	2016-09-10	71	OLI
2017	2017-07-01	2017-09-06	67	OLI
2018	2018-07-01	2018-09-28	89	OLI
2019	2019-07-01	2019-09-30	91	OLI

A4.4 Additional figures

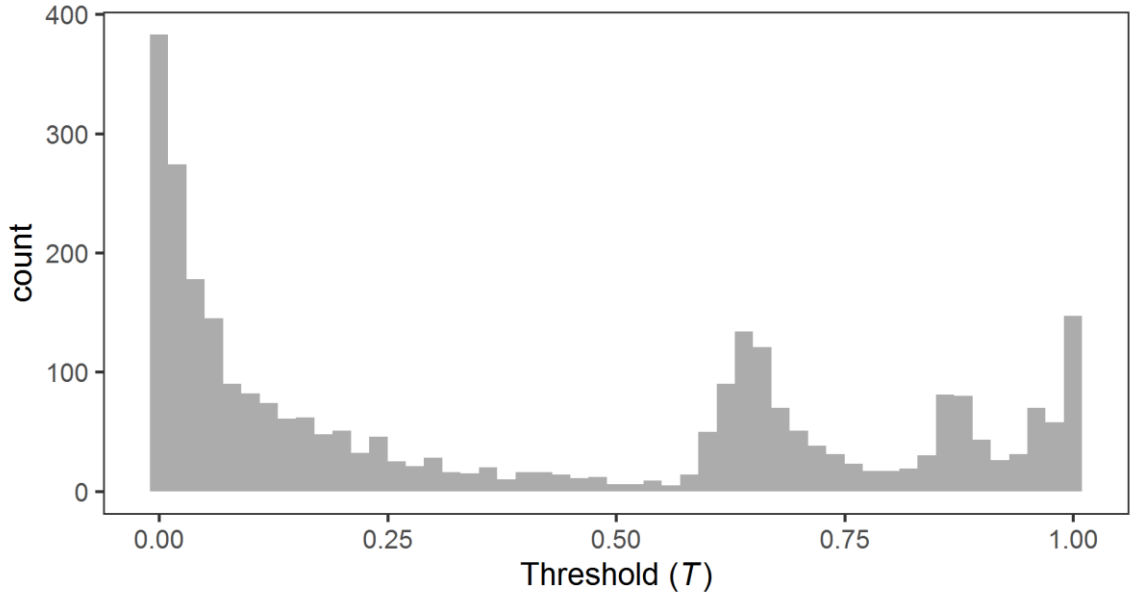


Figure A4.2 Histogram of predicted probabilities from the Random Forest model for logged observations in the validation dataset (all study sites pooled).

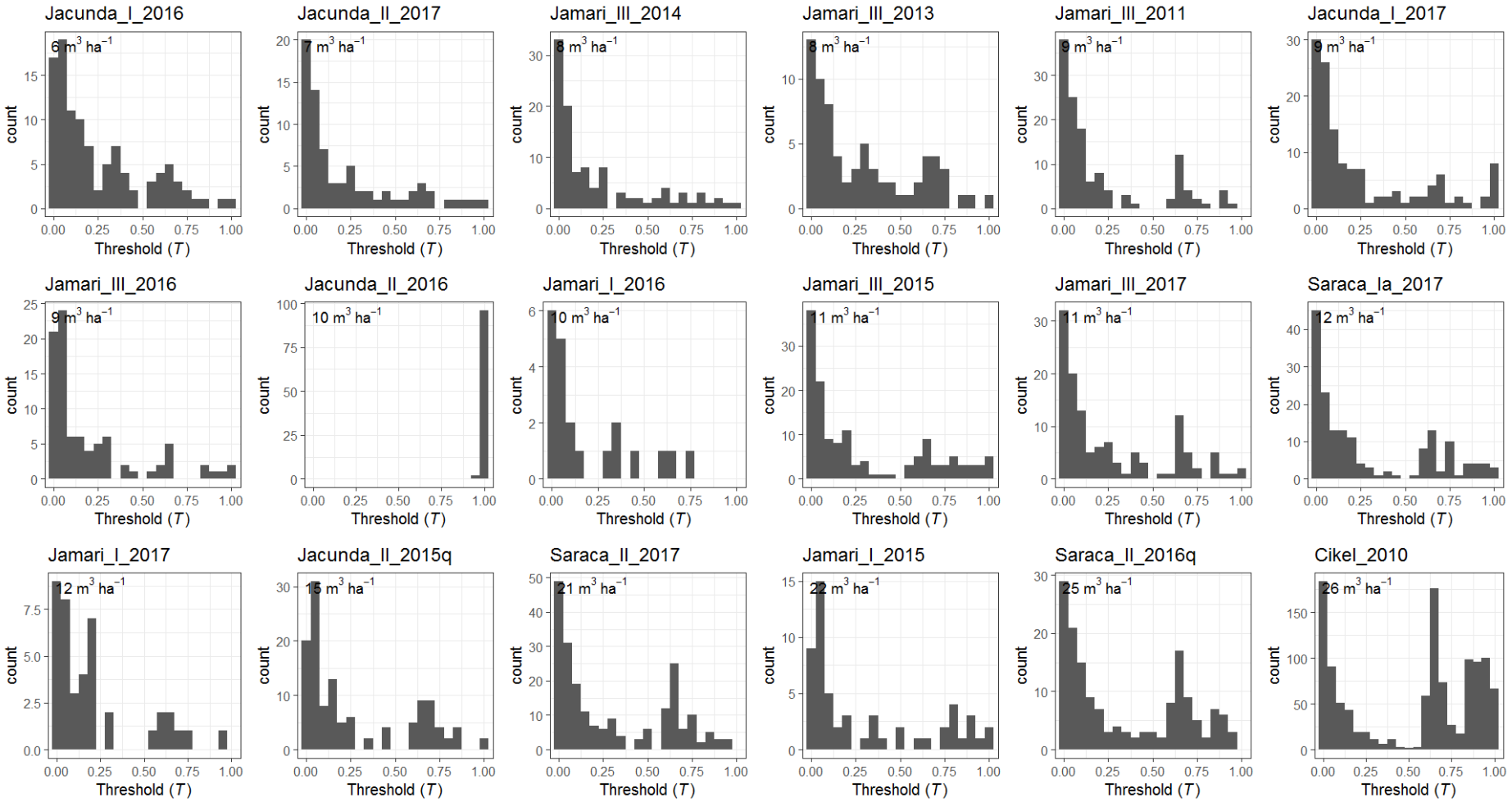


Figure A4.3 Histogram of predicted probabilities from the Random Forest model for logged observations in the validation dataset (separated by study site).

Figure A4.4 Deforestation detections from the Hansen data (in red) in the Saracá-Taquera National Forests, Pará overlaid with selective logging tree locations from the same year (grey circles). The deforestation detections here are selective logging and thus lowered our detection rate in some cases, as we excluded deforested areas before predicting logging. The map is centered at 56.17 W, 1.60 S.

References

Achard, F., R. S. DeFries, H. Eva, M. Hansen, P. Mayaux, and HJ Stibig. 2007. "Pan-Tropical Monitoring of Deforestation." *Environmental Research Letters* 2(4):045022.

Alamgir, Mohammed, Mason J. Campbell, Sean Sloan, Miriam Goosem, Gopalasamy Reuben Clements, Mahmoud I. Mahmoud, and William F. Laurance. 2017. "Economic, Socio-Political and Environmental Risks of Road Development in the Tropics." *Current Biology* 27(20):R1130–40.

de Andrade, Rafael B., Jennifer K. Balch, Amoreena L. Parsons, Dolors Armenteras, Rosa Maria Roman-Cuesta, and Janette Bulkan. 2017. "Scenarios in Tropical Forest Degradation: Carbon Stock Trajectories for REDD+." *Carbon Balance and Management* 12(1):6.

Antropov, Oleg, Yrjo Rauste, Anne Vaananen, Teemu Mutanen, and Tuomas Hame. 2016. "Mapping Forest Disturbance Using Long Time Series of Sentinel-1 Data: Case Studies over Boreal and Tropical Forests." Pp. 3906–9 in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Vols. 2016-Novem. IEEE.

Aragão, L. E. O. C., Liana O. Anderson, Marisa G. Fonseca, Thais M. Rosan, Laura B. Vedovato, Fabien H. Wagner, Camila V. J. J. Silva, Celso H. L. L. Silva Junior, Egidio Arai, Ana P. Aguiar, Jos Barlow, Erika Berenguer, Merritt N. Deeter, Lucas G. Domingues, Luciana Gatti, Manuel Gloor, Yadvinder Malhi, Jose A. Marengo, John B. Miller, Oliver L. Phillips, and Sassan Saatchi. 2018. "21st Century Drought-Related Fires Counteract the Decline of Amazon Deforestation Carbon Emissions." *Nature Communications* 9(1):1–12.

Arino, Olivier, John Armston, Gregory P. Asner, Luigi Boschetti, Barbara Braatz, Kim Calders, Emilio Chiuvieco, Ivan Csiszar, Mark Cutler, Mathias Disney, Takahiro Endo, Sandra Englhart, Michael Falkowski, Sandro Federici, Jonas Franke, Scott Goetz, Nancy Harris, Yasumasa Hirata, Dirk Hoekman, Anja A. Hoffman, Bernardus de Jong, Hans Joosten, Chris Justice, Josef Kellndorfer, Stephen Kull, Werner Kurz, Patrick van Laake, Eric Lambin, Richard Lucas, Mike McCall, Ronald McRoberts, Suvi Monni, Rebecca Moore, Erik Næsset, Ross Nelson, Glenn Newnham, Yosio Edemir Shimabukuro, Marc Paganini, Ian Paynter, T. R. H. Pearson, Jim Penman, Gary Richards, Ake Rosenqvist, David Roy, Jeremy Russell-Smith, Crystal Schaaf, Florian Siegert, David Shoch, Margaret Skutsch, Svein Solberg, Allan Spessa, and Michael Wulder. 2015. *A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Associated with Deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation*. The Netherlands.

Armenteras, Dolors, Tania Marisol González, and Javier Retana. 2013. "Forest Fragmentation and

Edge Influence on Fire Occurrence and Intensity under Different Management Types in Amazon Forests." *Biological Conservation* 159:73–79.

Asner, Gregory P. 2001. "Cloud Cover in Landsat Observations of the Brazilian Amazon." *International Journal of Remote Sensing* 22(18):3855–62.

Asner, Gregory P., Eben N. Broadbent, Paulo J. C. Oliveira, Michael Keller, David E. Knapp, and José N. M. Silva. 2006. "Condition and Fate of Logged Forests in the Brazilian Amazon." *Proceedings of the National Academy of Sciences of the United States of America* 103(34):12947–50.

Asner, Gregory P., Michael Keller, Marco Lentini, Frank Merry, and Carlos Souza. 2009. "Selective Logging and Its Relation to Deforestation." Pp. 25–42 in *Amazonia and Global Change*, edited by M. Keller, M. Bustamante, J. Gash, and P. S. Dias. Washington DC: American Geophysical Union.

Asner, Gregory P., Michael Keller, Rodrigo Pereira, Jr, Johan C. Zweede, and Jose N. M. Silva. 2004. "Canopy Damage and Recovery after Selective Logging in Amazonia: Field and Satellite Studies." *Ecological Applications* 14(sp4):280–98.

Asner, Gregory P., Michael Keller, Rodrigo Pereira, and Johan C. Zweede. 2002. "Remote Sensing of Selective Logging in Amazonia." *Remote Sensing of Environment* 80(3):483–96.

Asner, Gregory P., Michael Keller, and Jose N. M. Silva. 2004. "Spatial and Temporal Dynamics of Forest Canopy Gaps Following Selective Logging in the Eastern Amazon." *Global Change Biology* 10(5):765–83.

Asner, Gregory P., David E. Knapp, Aravindh Balaji, and Guayana Páez-Acosta. 2009. "Automated Mapping of Tropical Deforestation and Forest Degradation: CLASlite." *Journal of Applied Remote Sensing* 3(1):033543.

Asner, Gregory P., David E. Knapp, Eben N. Broadbent, Paulo J. C. Oliveira, Michael Keller, and Jose N. Silva. 2005. "Selective Logging in the Brazilian Amazon." *Science* 310(5747):480–82.

Asner, Gregory P., G. V. N. Powell, J. Mascaro, D. E. Knapp, J. K. Clark, J. Jacobson, T. Kennedy-Bowdoin, A. Balaji, G. Paez-Acosta, E. Victoria, L. Secada, M. Valqui, and R. F. Hughes. 2010. "High-Resolution Forest Carbon Stocks and Emissions in the Amazon." *Proceedings of the National Academy of Sciences* 107(38):16738–42.

Asner, Gregory P., George V N Powell, Joseph Mascaro, David E. Knapp, John K. Clark, James Jacobson, Ty Kennedy-Bowdoin, Aravindh Balaji, Guayana Paez-Acosta, Eloy Victoria, Laura

Secada, Michael Valqui, and R. Flint Hughes. 2010. “High-Resolution Forest Carbon Stocks and Emissions in the Amazon.” *Proceedings of the National Academy of Sciences of the United States of America* 107(38):16738–42.

Asner, Gregory P., Thomas K. Rudel, T. Mitchell Aide, R. S. DeFries, and Ruth Emerson. 2009. “A Contemporary Assessment of Change in Humid Tropical Forests.” *Conservation Biology* 23(6):1386–95.

Baccini, A., S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menashe, J. Hackler, P. S. A. Beck, R. Dubayah, M. A. Friedl, S. Samanta, and R. A. Houghton. 2012. “Estimated Carbon Dioxide Emissions from Tropical Deforestation Improved by Carbon-Density Maps.” *Nature Climate Change* 2(3):182–85.

Baccini, A., W. Walker, L. Carvalho, M. Farina, D. Sulla-Menashe, and R. A. Houghton. 2017. “Tropical Forests Are a Net Carbon Source Based on Aboveground Measurements of Gain and Loss.” *Science* 358(6360):230–34.

Barlow, Jos, Gareth D. Lennox, Joice Ferreira, Erika Berenguer, Alexander C. Lees, Ralph Mac Nally, James R. Thomson, Silvio Frosini De Barros Ferraz, Julio Louzada, Victor Hugo Fonseca Oliveira, Luke Parry, Ricardo Ribeiro de Castro Solar, Ima C. G. Vieira, Luiz E. O. C. Aragão, Rodrigo Anzolin Begotti, Rodrigo F. Braga, Thiago Moreira Cardoso, Raimundo Cosme De Oliveira Jr, C. M. Souza, Nárgila G. Moura, Sâmia Serra Nunes, João Victor Siqueira, Renata Pardini, Juliana M. Silveira, Fernando Z. Vaz-de-Mello, Ruan Carlo Stulpen Veiga, Adriano Venturieri, and Toby A. Gardner. 2016. “Anthropogenic Disturbance in Tropical Forests Can Double Biodiversity Loss from Deforestation.” *Nature* 535(7610):144–47.

Beekhuizen, Johan, and Keith C. Clarke. 2010. “Toward Accountable Land Use Mapping: Using Geocomputation to Improve Classification Accuracy and Reveal Uncertainty.” *International Journal of Applied Earth Observation and Geoinformation* 12(3):127–37.

Belgiu, Mariana, Lucian Drăgu, and Lucian Drăguţ. 2016. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” *ISPRS Journal of Photogrammetry and Remote Sensing* 114:24–31.

Benjamini, Y., and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B* 57(1):289–300.

Berry, Nicholas J., Oliver L. Phillips, Robert C. Ong, and Keith C. Hamer. 2008. “Impacts of Selective Logging on Tree Diversity across a Rainforest Landscape: The Importance of Spatial

Scale." *Landscape Ecology* 23:915–29.

Betts, Matthew G., Christopher Wolf, William J. Ripple, Ben Phalan, Kimberley A. Millers, Adam Duarte, Stuart H. M. Butchart, and Taal Levi. 2017. "Global Forest Loss Disproportionately Erodes Biodiversity in Intact Landscapes." *Nature* 547(7664):441–44.

Bicknell, Jake E., Matthew J. Struebig, D. P. Edwards, and Zoe G. Davies. 2014. "Improved Timber Harvest Techniques Maintain Biodiversity in Tropical Forests." *Current Biology* 24(23):R1119–20.

Blaser, Juergen, Alastair Sarre, Duncan Poore, and Steven Johnson. 2011. *Status of Tropical Forest Management 2011*. Vol. 38.

Bouvet, Alexandre, Stéphane Mermoz, Marie Ballère, Thierry Koleck, and Thuy Le Toan. 2018. "Use of the SAR Shadowing Effect for Deforestation Detection with Sentinel-1 Time Series." *Remote Sensing* 10(8):1250.

Brancalion, Pedro H. S., Danilo R. A. de Almeida, Edson Vidal, Paulo G. Molin, Vanessa E. Sontag, Saulo E. X. F. Souza, and Mark D. Schulze. 2018. "Fake Legal Logging in the Brazilian Amazon." *Science Advances* 4(8):eaat1192.

Breiman, Leo. 1996. "Bagging Predictors." *Machine Learning* 24(2):123–40.

Breiman, Leo. 2001a. "Random Forests." *Machine Learning* 45(1):5–32.

Breiman, Leo. 2001b. "Statistical Modeling: The Two Cultures." *Statistical Science* 16(3):199–231.

Brienen, R. J. W., O. L. Phillips, T. R. Feldpausch, E. Gloor, T. R. Baker, J. Lloyd, G. Lopez-Gonzalez, A. Monteagudo-Mendoza, Y. Malhi, S. L. Lewis, R. Vásquez Martinez, M. N. Alexiades, E. Álvarez Dávila, P. Alvarez-Loayza, A. Andrade, L. E. O. C. Aragão, Alejandro Araujo-Murakami, E. J. M. M. Arets, L. Arroyo, Gerardo A. Aymard C., O. S. Bánki, C. Baraloto, J. Barroso, D. Bonal, R. G. A. Boot, J. L. C. Camargo, C. V. Castilho, V. Chama, K. J. Chao, J. Chave, J. A. Comiskey, F. Cornejo Valverde, L. da Costa, E. A. de Oliveira, A. Di Fiore, T. L. Erwin, S. Fauset, M. Forsthofer, D. R. Galbraith, E. S. Grahame, N. Groot, B. Hérault, N. Higuchi, E. N. Honorio Coronado, H. Keeling, T. J. Killeen, W. F. Laurance, S. Laurance, J. Licona, W. E. Magnussen, B. S. Marimon, B. H. Marimon-Junior, C. Mendoza, D. A. Neill, E. M. Nogueira, P. Núñez, N. C. Pallqui Camacho, A. Parada, G. Pardo-Molina, J. Peacock, M. Peña-Claros, G. C. Pickavance, N. C. A. Pitman, L. Poorter, A. Prieto, C. A. Quesada, F. Ramírez, H. Ramírez-Angulo, Z. Restrepo, A. Roopsind, A. Rudas, R. P. Salomão, M. Schwarz, N. Silva, J. E. Silva-Espejo, M. Silveira, J. Stropp, J. Talbot, H. ter Steege, J. Teran-Aguilar, J. Terborgh, R. Thomas-Caesar, M. Toledo, M. Torello-Raventos, R. K. Umetsu,

G. M. F. van der Heijden, P. van der Hout, I. C. Guimarães Vieira, S. A. Vieira, E. Vilanova, V. A. Vos, and R. J. Zagt. 2015. “Long-Term Decline of the Amazon Carbon Sink.” *Nature* 519(7543):344–48.

Brinck, Katharina, Rico Fischer, Jürgen Groeneveld, Sebastian Lehmann, Mateus Dantas De Paula, Sandro Pütz, Joseph O. Sexton, Danxia Song, and Andreas Huth. 2017. “High Resolution Analysis of Tropical Forest Fragmentation and Its Impact on the Global Carbon Cycle.” *Nature Communications* 8:14855.

Broadbent, E., Gregory P. Asner, M. Keller, D. Knapp, P. Oliveira, and J. Silva. 2008. “Forest Fragmentation and Edge Effects from Deforestation and Selective Logging in the Brazilian Amazon.” *Biological Conservation* 141(7):1745–57.

Broadbent, Eben N., Daniel J. Zarin, Gregory P. Asner, Marielos Peña-Claros, Amanda Cooper, and Ramon Littell. 2006. “Recovery Of Forest Structure And Spectral Properties After Selective Logging In Lowland Bolivia.” *Ecological Applications* 16(3):1148–63.

Brodrick, Philip G., Andrew B. Davies, and Gregory P. Asner. 2019. “Uncovering Ecological Patterns with Convolutional Neural Networks.” *Trends in Ecology & Evolution* 34(8):734–45.

Bullock, Eric L., Curtis E. Woodcock, and Pontus Olofsson. 2018. “Monitoring Tropical Forest Degradation Using Spectral Unmixing and Landsat Time Series Analysis.” *Remote Sensing of Environment* 238(November):110968.

Burivalova, Zuzana, Çağan Hakkı Şekercioğlu, and Lian Pin Koh. 2014. “Thresholds of Logging Intensity to Maintain Tropical Forest Biodiversity.” *Current Biology* 24(16):1893–98.

Bustamante, Mercedes M. C., Iris Roitman, T. Mitchell Aide, Ane Alencar, Liana O. Anderson, L. E. O. C. Aragão, Gregory P. Asner, Jos Barlow, Erika Berenguer, Jeffrey Chambers, Marcos H. Costa, Thierry Fanin, Laerte G. Ferreira, Joice Ferreira, Michael Keller, William E. Magnusson, Lucia Morales-Barquero, Douglas Morton, Jean P. H. B. Ometto, Michael Palace, Carlos A. Peres, Divino Silvério, Susan Trumbore, and Ima C. G. Vieira. 2016. “Toward an Integrated Monitoring Framework to Assess the Effects of Tropical Forest Degradation and Recovery on Carbon Stocks and Biodiversity.” *Global Change Biology* 22(1):92–109.

Carlson, Kimberly M., L. M. Curran, D. Ratnasari, a. M. Pittman, B. S. Soares-Filho, Gregory P. Asner, S. N. Trigg, D. a. Gaveau, D. Lawrence, and H. O. Rodrigues. 2012. “Committed Carbon Emissions, Deforestation, and Community Land Conversion from Oil Palm Plantation Expansion in West Kalimantan, Indonesia.” *Proceedings of the National Academy of Sciences* 109(19):7559–64.

Castillo-Santiago, Miguel Angel, Martin Ricker, and Bernardus H. J. de Jong. 2010. “Estimation of Tropical Forest Structure from SPOT-5 Satellite Images.” *International Journal of Remote Sensing* 31(10):2767–82.

Chen, Chao, Andy Liaw, and Leo Breiman. 2004. “Using Random Forest to Learn Imbalanced Data.” *University of California, Berkeley* 110:1–12.

Chen, Gang, Ryan P. Powers, Luis M. T. de Carvalho, and Brice Mora. 2015. “Spatiotemporal Patterns of Tropical Deforestation and Forest Degradation in Response to the Operation of the Tucuruí Hydroelectric Dam in the Amazon Basin.” *Applied Geography* 63:1–8.

Cheng, Gong, Junwei Han, and Xiaoqiang Lu. 2017. “Remote Sensing Image Scene Classification: Benchmark and State of the Art.” *Proceedings of the IEEE* 105(10):1865–83.

Clark, David B., Carlomagno Soto Castro, Luis Diego Alfaro Alvarado, and Jane M. Read. 2004. “Quantifying Mortality of Tropical Rain Forest Trees Using High-Spatial-Resolution Satellite Data.” *Ecology Letters* 7(1):52–59.

Cohen, Jacob. 1960. “A Coefficient of Agreement for Nominal Scales.” *Educational and Psychological Measurement* 20(1):37–46.

CONAMA. 2009. *Brasil, Ministério Do Meio Ambiente. Resolução CONAMA Nº 406 de 02 de Fevereiro de 2009*.

Curtis, Philip G., Christy M. Slay, Nancy L. Harris, Alexandra Tyukavina, and Matthew C. Hansen. 2018. “Classifying Drivers of Global Forest Loss.” *Science* 361(6407):1108–11.

DeFries, R. S., Thomas Rudel, Maria Uriarte, and Matthew Hansen. 2010. “Deforestation Driven by Urban Population Growth and Agricultural Trade in the Twenty-First Century.” *Nature Geoscience* 3(3):178–81.

Delgado-Aguilar, M. J., F. E. Fassnacht, M. Peralvo, C. P. Gross, and C. B. Schmitt. 2017. “Potential of TerraSAR-X and Sentinel 1 Imagery to Map Deforested Areas and Derive Degradation Status in Complex Rain Forests of Ecuador.” *International Forestry Review* 19(1):102–18.

Deutscher, Janik, Roland Perko, Karlheinz Gutjahr, Manuela Hirschmugl, and Mathias Schardt. 2013. “Mapping Tropical Rainforest Canopy Disturbances in 3D by COSMO-SkyMed Spotlight InSAR-Stereo Data to Detect Areas of Forest Degradation.” *Remote Sensing* 5(2):648–63.

DiMiceli, C. M., M. L. Carroll, R. A. Sohlberg, C. Huang, M. C. Hansen, and J. R. G. Townshend. 2011. *User Guide for the MODIS Vegetation Continuous Fields Product Collection 5 Version 1*. College Park, Maryland.

Diniz, Cesar Guerreiro, Arleson Antonio De Almeida Souza, Diogo Correa Santos, Mirian Correa Dias, Nelton Cavalcante Da Luz, Douglas Rafael Vidal De Moraes, Janaina Sant Ana Maia, Alessandra Rodrigues Gomes, Igor Da Silva Narvaes, Dalton M. Valeriano, Luis Eduardo Pinheiro Maurano, and Marcos Adami. 2015. "DETER-B: The New Amazon Near Real-Time Deforestation Detection System." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8(7):3619–28.

Douglas, I. 1999. "Hydrological Investigations of Forest Disturbance and Land Cover Impacts in South-East Asia: A Review." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 354(1391):1725–38.

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

Edwards, D. P., T. H. Larsen, T. D. S. Docherty, F. A. Ansell, W. W. Hsu, M. a. Derhe, K. C. Hamer, and D. S. Wilcove. 2011. "Degraded Lands Worth Protecting: The Biological Importance of Southeast Asia's Repeatedly Logged Forests." *Proceedings of the Royal Society B: Biological Sciences* 278(1702):82–90.

Edwards, D. P., J. B. Socolar, S. C. Mills, Zuzana Burivalova, Lian Pin Koh, and David S. Wilcove. 2019. "Conservation of Tropical Forests in the Anthropocene." *Current Biology* 29(19):R1008–20.

Edwards, D. P., Joseph a. Tobias, Douglas Sheil, Erik Meijaard, and William F. Laurance. 2014. "Maintaining Ecosystem Function and Services in Logged Tropical Forests." *Trends in Ecology & Evolution* 29(9):511–20.

Ellis, Peter W., Trisha Gopalakrishna, Rosa C. Goodman, F. E. Putz, Anand Roopsind, Peter M. Umunay, Joey Zalman, Edward A. Ellis, Karen Mo, Timothy G. Gregoire, and Bronson W. Griscom. 2019. "Reduced-Impact Logging for Climate Change Mitigation (RIL-C) Can Halve Selective Logging Emissions from Tropical Forests." *Forest Ecology and Management* 438(February):255–66.

Erasmi, S., and A. Twele. 2009. "Regional Land Cover Mapping in the Humid Tropics Using Combined Optical and SAR Satellite Data—a Case Study from Central Sulawesi, Indonesia." *International Journal of Remote Sensing* 30(10):2465–78.

Eva, Hugh D., F. Achard, René Beuchle, Evaristo de Miranda, Silvia Carboni, Roman Seliger,

Michael Vollmar, Wilson A. Holler, Osvaldo T. Oshiro, Victor Barrena Arroyo, and Javier Gallego. 2012. “Forest Cover Changes in Tropical South and Central America from 1990 to 2005 and Related Carbon Emissions and Removals.” *Remote Sensing* 4(12):1369–91.

FAO. 2011. “Assessing Forest Degradation: Towards the Development of Globally Applicable Guidlines.” *Assessing Forest Degradation: Towards the Development of Globally Applicable Guidelines* 99.

FAO. 2015. *Global Forest Resources Assessment 2015*. Rome.

FAO. 2017. *Sustainable Woodfuel for Food Security*. Rome.

FAO. 2018. *The State of the World’s Forests 2018*. Rome.

Ferreira, Joice, Gareth D. Lennox, Toby A. Gardner, James R. Thomson, Erika Berenguer, Alexander C. Lees, Ralph Mac Nally, L. E. O. C. Aragão, Silvio F. B. B. Ferraz, Julio Louzada, Nárgila G. Moura, Victor H. F. F. Oliveira, Renata Pardini, Ricardo R. C. C. Solar, Ima C. G. G. Vieira, and Jos Barlow. 2018. “Carbon-Focused Conservation May Fail to Protect the Most Biodiverse Tropical Forests.” *Nature Climate Change* 8(8):744–49.

Fisher, Brendan, D. P. Edwards, and David S. Wilcove. 2014. “Logging and Conservation: Economic Impacts of the Stocking Rates and Prices of Commercial Timber Species.” *Forest Policy and Economics* 38:65–71.

Flores-Anderson, Africa Ixmucane, Kelsey E. Herndon, Rajesh Bahadur Thapa, and Emil Cherrington, eds. 2019. *The SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. NASA.

França, Filipe M., Fábio S. Frazão, Vanesca Korasaki, Júlio Louzada, and Jos Barlow. 2017. “Identifying Thresholds of Logging Intensity on Dung Beetle Communities to Improve the Sustainable Management of Amazonian Tropical Forests.” *Biological Conservation* 216(September):115–22.

Fu, Kun, Zhuo Chen, Yue Zhang, and Xian Sun. 2019. “Enhanced Feature Representation in Detection for Optical Remote Sensing Images.” *Remote Sensing* 11(18):1–17.

Gardner, Toby A., Jos Barlow, Robin Chazdon, Robert M. Ewers, Celia A. Harvey, Carlos A. Peres, and Navjot S. Sodhi. 2009. “Prospects for Tropical Forest Biodiversity in a Human-Modified World.” *Ecology Letters* 12(6):561–82.

GFOI. 2016. *Integrating Remote-Sensing and Ground-Based Observations for Estimation of Emissions and Removals of Greenhouse Gases in Forests*. Rome.

Ghazoul, Jaboury, Zuzana Burivalova, John Garcia-Ulloa, and Lisa a King. 2015. “Conceptualizing Forest Degradation.” *Trends in Ecology & Evolution* 30(10):622–32.

Gibbs, H. K., A. S. Ruesch, F. Achard, M. K. Clayton, P. Holmgren, N. Ramankutty, and J. A. Foley. 2010. “Tropical Forests Were the Primary Sources of New Agricultural Land in the 1980s and 1990s.” *Proceedings of the National Academy of Sciences* 107(38):16732–37.

Gibson, Luke, Tien Ming Lee, Lian Pin Koh, Barry W. Brook, Toby A. Gardner, Jos Barlow, Carlos A. Peres, Corey J. A. Bradshaw, William F. Laurance, Thomas E. Lovejoy, and Navjot S. Sodhi. 2011. “Primary Forests Are Irreplaceable for Sustaining Tropical Biodiversity.” *Nature* 478(7369):378–81.

Giglio, Louis, Luigi Boschetti, David P. Roy, Michael L. Humber, and Christopher O. Justice. 2018. “The Collection 6 MODIS Burned Area Mapping Algorithm and Product.” *Remote Sensing of Environment* 217(July):72–85.

Goetz, Scott J., Matthew Hansen, Richard A. Houghton, Wayne Walker, Nadine Laporte, and Jonah Busch. 2015. “Measurement and Monitoring Needs, Capabilities and Potential for Addressing Reduced Emissions from Deforestation and Forest Degradation under REDD+.” *Environmental Research Letters* 10(12):123001.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. “Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone.” *Remote Sensing of Environment* 202:18–27.

Goward, Samuel, Terry Arvidson, Darrel Williams, John Faundeen, James Irons, and Shannon Franks. 2006. “Historical Record of Landsat Global Coverage.” *Photogrammetric Engineering & Remote Sensing* 72(10):1155–69.

Grace, John, Edward T. A. Mitchard, and Emanuel Gloor. 2014. “Perturbations in the Carbon Budget of the Tropics.” *Global Change Biology* 20(10):3238–55.

Griscom, Bronson W., Justin Adams, Peter W. Ellis, Richard A. Houghton, Guy Lomax, Daniela A. Miteva, William H. Schlesinger, David Shoch, Juha V. Siikamäki, Pete Smith, Peter Woodbury, Chris Zganjar, Allen Blackman, João Campari, Richard T. Conant, Christopher Delgado, Patricia Elias, Trisha Gopalakrishna, Marisa R. Hamsik, Mario Herrero, Joseph Kiesecker, Emily Landis, Lars Laestadius, Sara M. Leavitt, Susan Minnemeyer, Stephen Polasky, P. V Potapov, F. E. Putz, Jonathan Sanderman, Marcel Silvius, Eva Wollenberg, and Joseph Fargione. 2017. “Natural Climate Solutions.” *Proceedings of the National Academy of Sciences* 114(44):11645–50.

Grogan, Kenneth, Dirk Pflugmacher, Patrick Hostert, Robert Kennedy, and Rasmus Fensholt. 2015. "Cross-Border Forest Disturbance and the Role of Natural Rubber in Mainland Southeast Asia Using Annual Landsat Time Series." *Remote Sensing of Environment* 169:438–53.

Hall-Beyer, Mryka. 2017. "Practical Guidelines for Choosing GLCM Textures to Use in Landscape Classification Tasks over a Range of Moderate Spatial Scales." *International Journal of Remote Sensing* 38(5):1312–38.

Hammer, Dan, Robin Kraft, and David Wheeler. 2014. "Alerts of Forest Disturbance from MODIS Imagery." *International Journal of Applied Earth Observation and Geoinformation* 33(1):1–9.

Hansen, M. .., R. S. DeFries, J. R. .. Townshend, R. Sohlberg, C. M. DiMiceli, and M. Carroll. 2002. "Towards an Operational MODIS Continuous Field of Percent Tree Cover Algorithm: Examples Using AVHRR and MODIS Data." *Remote Sensing of Environment* 83(1–2):303–19.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. Sohlberg. 2000. "Global Land Cover Classification at 1 Km Spatial Resolution Using a Classification Tree Approach." *International Journal of Remote Sensing* 21(6–7):1331–64.

Hansen, M. C., P. V Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342(6160):850–53.

Hansen, Matthew C., Alexander Krylov, Alexandra Tyukavina, P. V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. 2016. "Humid Tropical Forest Disturbance Alerts Using Landsat Data." *Environmental Research Letters* 11(3):034008.

Hansen, Matthew C., David P. Roy, Erik Lindquist, Bernard Adusei, Christopher O. Justice, and Alice Altstatt. 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(5):2495–2513.

Haralick, Robert M., K. Shanmugam, and Its'Hak Dinstein. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3(6):610–21.

Hartanto, Herlina, Ravi Prabhu, a. S. E. Widayat, and Chay Asdak. 2003. "Factors Affecting Runoff and Soil Erosion: Plot-Level Soil Loss Monitoring for Assessing Sustainability of Forest Management." *Forest Ecology and Management* 180(1–3):361–74.

Herold, Martin, and Tracy Johns. 2007. “Linking Requirements with Capabilities for Deforestation Monitoring in the Context of the UNFCCC-REDD Process.” *Environmental Research Letters* 2:045025.

Herold, Martin, Rosa Román-Cuesta, Danilo Mollicone, Yasumasa Hirata, Patrick Van Laake, Gregory P. Asner, C. M. Souza, Margaret Skutsch, Valerio Avitabile, and Ken MacDicken. 2011. “Options for Monitoring and Estimating Historical Carbon Emissions from Forest Degradation in the Context of REDD+.” *Carbon Balance and Management* 6(1):13.

Hethcoat, MG, D. P. Edwards, JMB Carreiras, R. G. Bryant, F. M. França, and S. Quegan. 2019. “A Machine Learning Approach to Map Tropical Selective Logging.” *Remote Sensing of Environment* 221(February 2019):569–82.

Higginbottom, Thomas P., Elias Symeonakis, Hanna Meyer, and Sebastian van der Linden. 2018. “Mapping Fractional Woody Cover in Semi-Arid Savannahs Using Multi-Seasonal Composites from Landsat Data.” *ISPRS Journal of Photogrammetry and Remote Sensing* 139:88–102.

Hosonuma, Noriko, Martin Herold, Veronique De Sy, Ruth S. De Fries, Maria Brockhaus, Louis Verchot, Arild Angelsen, and Erika Romijn. 2012. “An Assessment of Deforestation and Forest Degradation Drivers in Developing Countries.” *Environmental Research Letters* 7(4):044009.

Houghton, R. A., Brett Byers, and Alexander A. Nassikas. 2015. “A Role for Tropical Forests in Stabilizing Atmospheric CO2.” *Nature Climate Change* 5(12):1022–23.

Huang, Xin, Qikai Lu, Liangpei Zhang, and Antonio Plaza. 2014. “New Postprocessing Methods for Remote Sensing Image Classification: A Systematic Study.” *IEEE Transactions on Geoscience and Remote Sensing* 52(11):7140–59.

Hussain, Masroor, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. 2013. “Change Detection from Remotely Sensed Images: From Pixel-Based to Object-Based Approaches.” *ISPRS Journal of Photogrammetry and Remote Sensing* 80:91–106.

IPCC. 2006. *IPCC Guidelines for National Greenhouse Gas Inventories*.

Joshi, Neha, Matthias Baumann, Andrea Ehammer, Rasmus Fensholt, Kenneth Grogan, Patrick Hostert, Martin Jepsen, Tobias Kuemmerle, Patrick Meyfroidt, Edward T. A. Mitchard, Johannes Reiche, Casey Ryan, and Björn Waske. 2016. “A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring.” *Remote Sensing* 8(1):70.

Kissinger, Gabrielle, Martin Herold, and Veronique De Sy. 2012. *Drivers of Deforestation and Forest*

Degradation. or REDD+ Policymakers. Vancouver.

Kleinschmit, Daniela, Stephanie Mansourian, Christoph Wildburger, and Andre Purret. 2016. *Illegal Logging and Related Timber Trade – Dimensions, Drivers, Impacts and Responses*. Vol. 35.

Kleinschroth, Fritz, Sylvie Gourlet-Fleury, Plinio Sist, Fréderic Mortier, and John R. Healey. 2015. "Legacy of Logging Roads in the Congo Basin: How Persistent Are the Scars in Forest Cover?" *Ecosphere* 6(4):art64.

Kleinschroth, Fritz, and John R. Healey. 2017. "Impacts of Logging Roads on Tropical Forests." *Biotropica* 49(5):620–35.

Kleinschroth, Fritz, John R. Healey, Plinio Sist, Frédéric Mortier, and Sylvie Gourlet-Fleury. 2016. "How Persistent Are the Impacts of Logging Roads on Central African Forest Vegetation?" edited by L. Baeten. *Journal of Applied Ecology* 53(4):1127–37.

Koch, Barbara. 2010. "Status and Future of Laser Scanning, Synthetic Aperture Radar and Hyperspectral Remote Sensing Data for Forest Biomass Assessment." *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6):581–90.

Koltunov, Alexander, Susan L. Ustin, Gregory P. Asner, and Inez Fung. 2009. "Selective Logging Changes Forest Phenology in the Brazilian Amazon: Evidence from MODIS Image Time Series Analysis." *Remote Sensing of Environment* 113(11):2431–40.

Kumar, Sanath S., David P. Roy, Mark A. Cochrane, C. M. Souza, Chirstopher P. Barber, and L. Boschetti. 2014. "A Quantitative Study of the Proximity of Satellite Detected Active Fires to Roads and Rivers in the Brazilian Tropical Moist Forest Biome." *International Journal of Wildland Fire* 23(4):532.

Langner, Andreas, Jukka Miettinen, Markus Kukkonen, Christelle Vancutsem, Dario Simonetti, Ghislain Vieilledent, Astrid Verhegghen, Javier Gallego, and Hans-Jürgen Stibig. 2018. "Towards Operational Monitoring of Forest Canopy Disturbance in Evergreen Rain Forests: A Test Case in Continental Southeast Asia." *Remote Sensing* 10(4):544.

Laurance, William F., Gopalasamy Reuben Clements, Sean Sloan, Christine S. O'Connell, Nathan D. Mueller, Miriam Goosem, Oscar Venter, D. P. Edwards, Ben Phalan, Andrew Balmford, Rodney Van Der Ree, and Irene Burgues Arrea. 2014. "A Global Strategy for Road Building." *Nature* 513(7517):229–32.

Laurance, William F., Miriam Goosem, and Susan G. W. Laurance. 2009. "Impacts of Roads and Linear Clearings on Tropical Forests." *Trends in Ecology & Evolution* 24(12):659–69.

Lei, Yang, Robert Treuhaft, Michael Keller, Maiza Dos-Santos, Fabio Gonçalves, and Maxim Neumann. 2018. “Quantification of Selective Logging in Tropical Forest with Spaceborne SAR Interferometry.” *Remote Sensing of Environment* 211(March):167–83.

Lewis, Simon L., D. P. Edwards, and David Galbraith. 2015. “Increasing Human Dominance of Tropical Forests.” *Science* 349(6250):827–32.

Lewis, Simon L., and Mark A. Maslin. 2015. “Defining the Anthropocene.” *Nature* 519(7542):171–80.

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by RandomForest.” *R News* 2(3):18–22.

Loveland, Thomas R., and John L. Dwyer. 2012. “Landsat: Building a Strong Future.” *Remote Sensing of Environment* 122(October 2000):22–29.

Martin, Philip A., Adrian C. Newton, Marion Pfeifer, MinSheng Khoo, and James M. Bullock. 2015. “Impacts of Tropical Selective Logging on Carbon Storage and Tree Species Richness: A Meta-Analysis.” *Forest Ecology and Management* 356:224–33.

Mathieu, Renaud, Laven Naidoo, Moses A. Cho, Brigitte Leblon, Russell Main, Konrad Wessels, Gregory P. Asner, Joseph Buckley, Jan Van Aardt, Barend F. N. Erasmus, and Izak P. J. Smit. 2013. “Toward Structural Assessment of Semi-Arid African Savannahs and Woodlands: The Potential of Multitemporal Polarimetric RADARSAT-2 Fine Beam Images.” *Remote Sensing of Environment* 138:215–31.

Matricardi, EAT, D. L. Skole, M. A. Cochrane, M. A. Pedlowski, and W. Chomentowski. 2007. “Multi-Temporal Assessment of Selective Logging in the Brazilian Amazon Using Landsat Data.” *International Journal of Remote Sensing* 28(1):63–82.

Matricardi, EAT, David L. Skole, M. A. Pedlowski, W. Chomentowski, and Luis Claudio Fernandes. 2010. “Assessment of Tropical Forest Degradation by Selective Logging and Fire Using Landsat Imagery.” *Remote Sensing of Environment* 114(5):1117–29.

Maxwell, Sean L., Tom Evans, James E. M. Watson, Alexandra Morel, Hedley Grantham, Adam Duncan, Nancy Harris, P. V Potapov, Rebecca K. Runting, Oscar Venter, Stephanie Wang, and Yadvinder Malhi. 2019. “Degradation and Forgone Removals Increase the Carbon Impact of Intact Forest Loss by 626%.” *Science Advances* 5(10):eaax2546.

Mercier, Audrey, Julie Betbeder, Florent Rumiano, Jacques Baudry, Valéry Gond, Lilian Blanc, Clément Bourgoin, Guillaume Cornu, Carlos Ciudad, Miguel Marchamalo, René Poccard-

Chapuis, and Laurence Hubert-Moy. 2019. “Evaluation of Sentinel-1 and 2 Time Series for Land Cover Classification of Forest–Agriculture Mosaics in Temperate and Tropical Landscapes.” *Remote Sensing* 11(8):979.

Milbank, Charlotte, David Coomes, and Bhaskar Vira. 2018. “Assessing the Progress of REDD+ Projects towards the Sustainable Development Goals.” *Forests* 9(10):589.

Mitchard, E. T. A., S. S. Saatchi, I. H. Woodhouse, G. Nangendo, N. S. Ribeiro, M. Williams, C. M. Ryan, S. L. Lewis, T. R. Feldpausch, and P. Meir. 2009. “Using Satellite Radar Backscatter to Predict Above-Ground Woody Biomass: A Consistent Relationship across Four Different African Landscapes.” *Geophysical Research Letters* 36(23):L23401.

Mitchard, Edward T. A. 2018. “The Tropical Forest Carbon Cycle and Climate Change.” *Nature* 559(7715):527–34.

Mitchell, Anthea L., Ake Rosenqvist, and Brice Mora. 2017. “Current Remote Sensing Approaches to Monitoring Forest Degradation in Support of Countries Measurement, Reporting and Verification (MRV) Systems for REDD+.” *Carbon Balance and Management* 12(1):9.

Monteiro, A. L., C. M. Souza, and P. Barreto. 2003. “Detection of Logging in Amazonian Transition Forests Using Spectral Mixture Models.” *International Journal of Remote Sensing* 24(1):151–59.

Morfitt, Ron, Julia Barsi, Raviv Levy, Brian Markham, Esad Micijevic, Lawrence Ong, Pat Scaramuzza, and Kelly Vanderwerff. 2015. “Landsat-8 Operational Land Imager (OLI) Radiometric Performance On-Orbit.” *Remote Sensing* 7(2):2208–37.

Murcia, Carolina. 1995. “Edge Effects in Fragmented Forests: Implications for Conservation.” *Trends in Ecology & Evolution* 10(2):58–62.

Nepstad, Daniel C., Adalberto Verssimo, Ane Alencar, Carlos Nobre, Eirivelthon Lima, Paul Lefebvre, Peter Schlesinger, Christopher Potter, Paulo Moutinho, Elsa Mendoza, Mark Cochrane, and Vanessa Brooks. 1999. “Large-Scale Impoverishment of Amazonian Forests by Logging and Fire.” *Nature* 398(6727):505–8.

Neuvial, Pierre, and Etienne Roquain. 2012. “On False Discovery Rate Thresholding for Classification under Sparsity.” *The Annals of Statistics* 40(5):2572–2600.

Nogueira, Keiller, Otávio A. B. Penatti, and Jefersson A. dos Santos. 2017. “Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification.” *Pattern Recognition* 61:539–56.

Olofsson, Pontus, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. 2014. “Good Practices for Estimating Area and Assessing Accuracy of Land Change.” *Remote Sensing of Environment* 148:42–57.

Olsson, L., H.Barbosa, S. Bhadwal, A. Cowie, K. Delusca, D. Flores-Renteria, K. Hermans, E. Jobbagy, W. Kurz, D. Li, D. J. Sonwa, and L. Stringer. 2019. “Land Degradation.” Pp. 345–436 in *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*, edited by P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. P. Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, and J. Malley. Cambridge University Press.

Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes. 2011. “A Large and Persistent Carbon Sink in the World’s Forests.” *Science* 333(6045):988–93.

Pearson, T. R. H., Sandra Brown, and Felipe M. Casarim. 2014. “Carbon Emissions from Tropical Forest Degradation Caused by Logging.” *Environmental Research Letters* 9(3):034017.

Pearson, T. R. H., Sandra Brown, Lara Murray, and Gabriel Sidman. 2017. “Greenhouse Gas Emissions from Tropical Forest Degradation: An Underestimated Source.” *Carbon Balance and Management* 12(1):3.

Pedlowski, M. A., E. A. T. Matricardi, D. Skole, S. R. Cameron, W. Chomentowski, C. Fernandes, and A. Lisboa. 2005. “Conservation Units: A New Deforestation Frontier in the Amazonian State of Rondônia, Brazil.” *Environmental Conservation* 32(2):149–55.

Peres, Carlos A., Jos Barlow, and William F. Laurance. 2006. “Detecting Anthropogenic Disturbance in Tropical Forests.” *Trends in Ecology & Evolution* 21(5):227–29.

Phelps, J., E. L. Webb, and A. Agrawal. 2010. “Does REDD+ Threaten to Recentralize Forest Governance?” *Science* 328(5976):312–13.

Pinard, Michelle A., and F. E. Putz. 1996. “Retaining Forest Biomass by Reducing Logging Damage.” *Biotropica* 28(3):278–95.

Potapov, P. V, Matthew C. Hansen, Lars Laestadius, Svetlana Turubanova, Alexey Yaroshenko, Christoph Thies, Wynet Smith, Ilona Zhuravleva, Anna Komarova, Susan Minnemeyer, and Elena Esipova. 2017. “The Last Frontiers of Wilderness: Tracking Loss of Intact Forest

Landscapes from 2000 to 2013." *Science Advances* 3(1):1–14.

Potapov, P. V, Aleksey Yaroshenko, Svetlana Turubanova, Maxim Dubinin, Lars Laestadius, Christoph Thies, Dmitry Aksenov, Aleksey Egorov, Yelena Yesipova, Igor Glushkov, Mikhail Karpachevskiy, Anna Kostikova, Alexander Manisha, Ekaterina Tsybikova, and Ilona Zhuravleva. 2008. "Mapping the World's Intact Forest Landscapes by Remote Sensing." *Ecology and Society* 13(2):art51.

Putz, F. E., GM Blate, KH Redford, Robert Fimbel, and John Robinson. 2001. "Tropical Forest Management and Conservation of Biodiversity: An Overview." *Conservation Biology* 15(1):7–20.

Putz, F. E., Dennis P. Dykstra, and Rudolf Heinrich. 2000. "Why Poor Logging Practices Persist in the Tropics." *Conservation Biology* 14(4):951–56.

Putz, F. E., and Michelle A. Pinard. 1993. "Reduced-Impact Logging as a Carbon-Offset Method." *Conservation Biology* 7(4):755–57.

Putz, F. E., Pieter A. Zuidema, Michelle A. Pinard, Rene G. A. Boot, Jeffrey A. Sayer, Douglas Sheil, Plinio Sist, and Jerome K. Vanclay. 2008. "Improved Tropical Forest Management for Carbon Retention." *PLoS Biology* 6(7):e166.

Putz, F. E., Pieter A. Zuidema, Timothy Synnott, Marielos Peña-Claros, Michelle A. Pinard, Douglas Sheil, Jerome K. Vanclay, Plinio Sist, Sylvie Gourlet-Fleury, Bronson Griscom, John Palmer, and Roderick Zagt. 2012. "Sustaining Conservation Values in Selectively Logged Tropical Forests: The Attained and the Attainable." *Conservation Letters* 5(4):296–303.

Putz, Francis E., Tracy Baker, Bronson W. Griscom, Trisha Gopalakrishna, Anand Roopsind, Peter M. Umunay, Joey Zalman, Edward A. Ellis, Ruslandi, and Peter W. Ellis. 2019. "Intact Forest in Selective Logging Landscapes in the Tropics." *Frontiers in Forests and Global Change* 2(June):1–10.

Qin, Yuanwei, Xiangming Xiao, Jinwei Dong, Yao Zhang, Xiaocui Wu, Yosio Shimabukuro, Egidio Arai, Chandrashekhar Biradar, Jie Wang, Zhenhua Zou, Fang Liu, Zheng Shi, Russell Doughty, and Berrien Moore. 2019. "Improved Estimates of Forest Cover and Loss in the Brazilian Amazon in 2000–2017." *Nature Sustainability* 2(8):764–72.

Quegan, Shaun, and Jiong Jiong Yu. 2001. "Filtering of Multichannel SAR Images." *IEEE Transactions on Geoscience and Remote Sensing* 39(11):2373–79.

Le Quéré, Corinne, Robbie M. Andrew, Pierre Friedlingstein, Stephen Sitch, Judith Hauck, Julia

Pongratz, Penelope A. Pickers, Jan Ivar Korsbakken, Glen P. Peters, Josep G. Canadell, Almut Arneth, Vivek K. Arora, Leticia Barbero, Ana Bastos, Laurent Bopp, Frédéric Chevallier, Louise P. Chini, Philippe Ciais, Scott C. Doney, Thanos Gkritzalis, Daniel S. Goll, Ian Harris, Vanessa Haverd, Forrest M. Hoffman, Mario Hoppema, Richard A. Houghton, George Hurtt, Tatiana Ilyina, Atul K. Jain, Truls Johannessen, Chris D. Jones, Etsushi Kato, Ralph F. Keeling, Kees Klein Goldewijk, Peter Landschützer, Nathalie Lefèvre, Sebastian Lienert, Zhu Liu, Danica Lombardozzi, Nicolas Metzl, David R. Munro, Julia E. M. S. Nabel, Shin-ichiro Nakaoka, Craig Neill, Are Olsen, Tsueno Ono, Prabir Patra, Anna Peregon, Wouter Peters, Philippe Peylin, Benjamin Pfeil, Denis Pierrot, Benjamin Poulter, Gregor Rehder, Laure Resplandy, Eddy Robertson, Matthias Rocher, Christian Rödenbeck, Ute Schuster, Jörg Schwinger, Roland Séférian, Ingunn Skjelvan, Tobias Steinhoff, Adrienne Sutton, Pieter P. Tans, Hanqin Tian, Bronte Tilbrook, Francesco N. Tubiello, Ingrid T. van der Laan-Luijkx, Guido R. van der Werf, Nicolas Viovy, Anthony P. Walker, Andrew J. Wiltshire, Rebecca Wright, Sönke Zaehle, and Bo Zheng. 2018. “Global Carbon Budget 2018.” *Earth System Science Data* 10(4):2141–94.

Reiche, Johannes, Eliakim Hamunyela, Jan Verbesselt, Dirk Hoekman, and Martin Herold. 2018. “Improving Near-Real Time Deforestation Monitoring in Tropical Dry Forests by Combining Dense Sentinel-1 Time Series with Landsat and ALOS-2 PALSAR-2.” *Remote Sensing of Environment* 204(October 2017):147–61.

Reiche, Johannes, Richard Lucas, Anthea L. Mitchell, Jan Verbesselt, Dirk H. Hoekman, Jörg Haarpaintner, Josef M. Kellndorfer, Ake Rosenqvist, Eric A. Lehmann, Curtis E. Woodcock, Frank Martin Seifert, and Martin Herold. 2016. “Combining Satellite Data for Better Tropical Forest Monitoring.” *Nature Climate Change* 6(2):120–22.

Reiche, Johannes, Jan Verbesselt, Dirk Hoekman, and Martin Herold. 2015. “Fusing Landsat and SAR Time Series to Detect Deforestation in the Tropics.” *Remote Sensing of Environment* 156:276–93.

Reiche, Johannes, Rob Verhoeven, Jan Verbesselt, Eliakim Hamunyela, Niels Wielaard, and Martin Herold. 2018. “Characterizing Tropical Forest Cover Loss Using Dense Sentinel-1 Data and Active Fire Alerts.” *Remote Sensing* 10(5):777.

Richardson, Vanessa A., and Carlos A. Peres. 2016. “Temporal Decay in Timber Species Composition and Value in Amazonian Logging Concessions” edited by A. Zia. *PLOS ONE* 11(7):e0159035.

Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan.

2012. “Random Forest Classification of Mediterranean Land Cover Using Multi-Seasonal Imagery and Multi-Seasonal Texture.” *Remote Sensing of Environment* 121:93–107.

Roy, D. P., V. Kovalskyy, H. K. Zhang, E. F. Vermote, L. Yan, S. S. Kumar, and A. Egorov. 2016. “Characterization of Landsat-7 to Landsat-8 Reflective Wavelength and Normalized Difference Vegetation Index Continuity.” *Remote Sensing of Environment* 185:57–70.

Roy, D. P., M. A. Wulder, T. R. Loveland, Woodcock C.E., R. G. Allen, M. C. Anderson, D. Helder, J. R. Irons, D. M. Johnson, R. Kennedy, T. A. Scambos, C. B. Schaaf, J. R. Schott, Y. Sheng, E. F. Vermote, A. S. Belward, R. Bindschadler, W. B. Cohen, F. Gao, J. D. Hipple, P. Hostert, J. Huntington, C. O. Justice, A. Kilic, V. Kovalskyy, Z. P. Lee, L. Lymburner, J. G. Masek, J. McCorkel, Y. Shuai, R. Trezza, J. Vogelmann, R. H. Wynne, and Z. Zhu. 2014. “Landsat-8: Science and Product Vision for Terrestrial Global Change Research.” *Remote Sensing of Environment* 145:154–72.

Roy, D. P., H. K. Zhang, J. Ju, J. L. Gomez-Dans, P. E. Lewis, C. B. Schaaf, Q. Sun, J. Li, H. Huang, and V. Kovalskyy. 2016. “A General Method to Normalize Landsat Reflectance Data to Nadir BRDF Adjusted Reflectance.” *Remote Sensing of Environment* 176:255–71.

Rudel, Thomas K., R. S. DeFries, Gregory P. Asner, and William F. Laurance. 2009. “Changing Drivers of Deforestation and New Opportunities for Conservation.” *Conservation Biology* 23(6):1396–1405.

Ryan, Casey M., Timothy Hill, Emily Woollen, Claire Ghee, Edward T. A. Mitchard, Gemma Cassells, John Grace, Iain H. Woodhouse, and Mathew Williams. 2012. “Quantifying Small-Scale Deforestation and Forest Degradation in African Woodlands Using Radar Imagery.” *Global Change Biology* 18(1):243–57.

Saatchi, S., Miriam Marlier, Robin L. Chazdon, David B. Clark, and Ann E. Russell. 2011. “Impact of Spatial Variability of Tropical Forest Structure on Radar Estimation of Aboveground Biomass.” *Remote Sensing of Environment* 115(11):2836–49.

Saatchi, S., Joseph Mascaro, Liang Xu, Michael Keller, Yan Yang, Paul Duffy, Fernando Espírito-Santo, Alessandro Baccini, Jeffery Chambers, and David Schimel. 2015. “Seeing the Forest beyond the Trees.” *Global Ecology and Biogeography* 24(5):606–10.

Saatchi, Sasan S., João Vianei Soares, and Diogenes Salas Alves. 1997. “Mapping Deforestation and Land Use in Amazon Rainforest by Using SIR-C Imagery.” *Remote Sensing of Environment* 59(2):191–202.

Saatchi, Sassan S., Nancy L. Harris, Sandra Brown, Michael Lefsky, Edward T. A. Mitchard, William

Salas, Brian R. Zutta, Wolfgang Buermann, Simon L. Lewis, Stephen Hagen, Silvia Petrova, Lee White, Miles Silman, and Alexandra Morel. 2011. “Benchmark Map of Forest Carbon Stocks in Tropical Regions across Three Continents.” *Proceedings of the National Academy of Sciences* 108(24):9899–9904.

Salati, Eneas, Attilio Dall’Olio, Eiichi Matsui, and Joel R. Gat. 1979. “Recycling of Water in the Amazon Basin: An Isotopic Study.” *Water Resources Research* 15(5):1250–58.

Sasaki, Nophea, and F. E. Putz. 2009. “Critical Need for New Definitions of ‘Forest’ and ‘Forest Degradation’ in Global Climate Change Agreements.” *Conservation Letters* 2(5):226–32.

Schapire, Robert E. 1990. “The Strength of Weak Learnability.” *Machine Learning* 5(2):197–227.

Schulz, Karsten, Ronny Hänsch, Uwe Sörgel, Karsten Schulz, and Uwe Sörgel. 2018. “Machine Learning Methods for Remote Sensing Applications: An Overview.” P. 1 in *Earth Resources and Environmental Remote Sensing/GIS Applications IX*. Vol. 1079002, edited by U. Michel and K. Schulz. SPIE.

Schwartz, Mischa. 1984. *Information Transmission, Modulation and Noise: A Unified Approach to Communication Systems*. McGraw-Hill.

Sellers, Piers J., David S. Schimel, Berrien Moore, Junjie Liu, and Annmarie Eldering. 2018. “Observing Carbon Cycle–Climate Feedbacks from Space.” *Proceedings of the National Academy of Sciences* 115(31):7860–68.

Shimabukuro, Yosio Edemir, René Beuchle, Rosana Cristina Grecchi, and F. Achard. 2014. “Assessment of Forest Degradation in Brazilian Amazon Due to Selective Logging and Fires Using Time Series of Fraction Images Derived from Landsat ETM+ Images.” *Remote Sensing Letters* 5(9):773–82.

Shimada, Masanobu, Takuya Itoh, Takeshi Motooka, Manabu Watanabe, Tomohiro Shiraishi, Rajesh Thapa, and Richard Lucas. 2014. “New Global Forest/Non-Forest Maps from ALOS PALSAR Data (2007–2010).” *Remote Sensing of Environment* 155:13–31.

Shimizu, Katsuto, Raul Ponce-Hernandez, Oumer S. Ahmed, Tetsuji Ota, Zar Chi Win, Nobuya Mizoue, and Shigejiro Yoshida. 2017. “Using Landsat Time Series Imagery to Detect Forest Disturbance in Selectively Logged Tropical Forests in Myanmar.” *Canadian Journal of Forest Research* 47(3):289–96.

Simula, Markku. 2009. *Towards Defining Forest Degradation: Comparative Analysis of Existing Definitions*. Rome, Italy.

Sist, Plinio. 2000. “Reduced-Impact Logging in the Tropics: Objectives, Principles and Impacts.” *International Forestry Review* 2(1):3–10.

Souza, C. M., and P. Barreto. 2000. “An Alternative Approach for Detecting and Monitoring Selectively Logged Forests in the Amazon.” *International Journal of Remote Sensing* 21(1):173–79.

Souza, C. M., and D. Roberts. 2005. “Mapping Forest Degradation in the Amazon Region with Ikonos Images.” *International Journal of Remote Sensing* 26(3):425–29.

Souza, C. M., Dar a. Roberts, and Mark a. Cochrane. 2005. “Combining Spectral and Spatial Information to Map Canopy Damage from Selective Logging and Forest Fires.” *Remote Sensing of Environment* 98(2–3):329–43.

Souza, C. M., João Siqueira, Marcio Sales, Antônio Fonseca, Júlia Ribeiro, Izaya Numata, Mark Cochrane, Christopher Barber, Dar Roberts, and Jos Barlow. 2013. “Ten-Year Landsat Classification of Deforestation and Forest Degradation in the Brazilian Amazon.” *Remote Sensing* 5(11):5493–5513.

Stone, T. A., and P. Lefebvre. 1998. “Using Multi-Temporal Satellite Data to Evaluate Selective Logging in Para, Brazil.” *International Journal of Remote Sensing* 19(13):2517–26.

Storey, James, Pasquale Scaramuzza, Gail Schmidt, and Julia Barsi. 2005. “Landsat 7 Scan Line Corrector-Off Gap-Filled Product Development.” Pp. 23–27 in *PECORA 16 Conference Proceedings*.

Stratford, Jeffrey a., and W. Douglas Robinson. 2005. “Gulliver Travels to the Fragmented Tropics: Geographic Variation in Mechanisms of Avian Extinction.” *Frontiers in Ecology and the Environment* 3(2):85–92.

De Sy, Veronique, Martin Herold, F. Achard, Gregory P. Asner, Alex Held, Josef Kellndorfer, and Jan Verbesselt. 2012. “Synergies of Multiple Remote Sensing Data Sources for REDD+ Monitoring.” *Current Opinion in Environmental Sustainability* 4(6):696–706.

Tabarelli, Marcelo, Carlos A. Peres, and Felipe P. L. Melo. 2012. “The ‘Few Winners and Many Losers’ Paradigm Revisited: Emerging Prospects for Tropical Forest Biodiversity.” *Biological Conservation* 155:136–40.

Taubert, Franziska, Rico Fischer, Jürgen Groeneveld, Sebastian Lehmann, Michael S. Müller, Edna Rödig, Thorsten Wiegand, and Andreas Huth. 2018. “Global Patterns of Tropical Forest Fragmentation.” *Nature* 554(7693):519–22.

Thomlinson, John R., Paul V. Bolstad, and Warren B. Cohen. 1999. “Coordinating Methodologies for Scaling Landcover Classifications from Site-Specific to Global.” *Remote Sensing of Environment* 70(1):16–28.

Thompson, Ian D., Manuel R. Guariguata, Kimiko Okabe, Carlos Bahamondez, Robert Nasi, Victoria Heymell, and Cesar Sabogal. 2013. “An Operational Framework for Defining and Monitoring Forest Degradation.” *Ecology and Society* 18(2):art20.

Townsend, Alan R., Cory C. Cleveland, Benjamin Z. Houlton, Caroline B. Alden, and James W. C. White. 2011. “Multi-element Regulation of the Tropical Forest Carbon Cycle.” *Frontiers in Ecology and the Environment* 9(1):9–17.

Tuia, Devis, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Munoz-Mari. 2011. “A Survey of Active Learning Algorithms for Supervised Remote Sensing Image Classification.” *IEEE Journal of Selected Topics in Signal Processing* 5(3):606–17.

Turubanova, Svetlana, Peter V. Potapov, Alexandra Tyukavina, and Matthew C. Hansen. 2018. “Ongoing Primary Forest Loss in Brazil, Democratic Republic of the Congo, and Indonesia.” *Environmental Research Letters* 13(7):074028.

Tyukavina, Alexandra, Matthew C. Hansen, P. V Potapov, Stephen V. Stehman, Kevin Smith-Rodriguez, Chima Okpa, and Ricardo Aguilar. 2017. “Types and Rates of Forest Disturbance in Brazilian Legal Amazon, 2000–2013.” *Science Advances* 3(4):e1601047.

UN-REDD. 2018. *UN-REDD Consolidated 2018 Annual Report*.

Vaglio Laurin, Gaia, Veraldo Liesenberg, Qi Chen, Leila Guerriero, Fabio Del Frate, Antonio Bartolini, David Coomes, Beccy Wilebore, Jeremy Lindsell, and Riccardo Valentini. 2013. “Optical and SAR Sensor Synergies for Forest and Land Cover Mapping in a Tropical Site in West Africa.” *International Journal of Applied Earth Observation and Geoinformation* 21:7–16.

Verbesselt, Jan, Rob Hyndman, Glenn Newnham, and Darius Culvenor. 2010. “Detecting Trend and Seasonal Changes in Satellite Image Time Series.” *Remote Sensing of Environment* 114(1):106–15.

Verbesselt, Jan, Achim Zeileis, and Martin Herold. 2012. “Near Real-Time Disturbance Detection Using Satellite Image Time Series.” *Remote Sensing of Environment* 123:98–108.

Vitousek, Peter M. 1984. “Litterfall, Nutrient Cycling, and Nutrient Limitation in Tropical Forests.” *Ecology* 65(1):285–98.

Warren, A. 2002. “Land Degradation Is Contextual.” *Land Degradation & Development* 13(6):449–

59.

Watanabe, Manabu, Christian N. Koyama, Masato Hayashi, Izumi Nagatani, and Masanobu Shimada. 2018. “Early-Stage Deforestation Detection in the Tropics With L-Band SAR.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(6):2127–33.

West, Thales A. P., Edson Vidal, and F. E. Putz. 2014. “Forest Biomass Recovery after Conventional and Reduced-Impact Logging in Amazonian Brazil.” *Forest Ecology and Management* 314(2014):59–63.

Williams-Linera, G., V. Dominguez-Gastelu, and M. E. Garcia-Zurita. 1998. “Microenvironment and Floristics of Different Edges in a Fragmented Tropical Rainforest.” *Conservation Biology* 12(5):1091–1102.

Woodhouse, I. H. 2005. *Introduction to Microwave Remote Sensing*. CRC Press.

Wright, Jonathon S., Rong Fu, John R. Worden, Sudip Chakraborty, Nicholas E. Clinton, Camille Risi, Ying Sun, and Lei Yin. 2017. “Rainforest-Initiated Wet Season Onset over the Southern Amazon.” *Proceedings of the National Academy of Sciences* 114(32):8481–86.

Wulder, Michael A., Thomas R. Loveland, David P. Roy, Christopher J. Crawford, Jeffrey G. Masek, Curtis E. Woodcock, R. G. Allen, Martha C. Anderson, Alan S. Belward, Warren B. Cohen, John Dwyer, Angela Erb, Feng Gao, Patrick Griffiths, Dennis Helder, Txomin Hermosilla, James D. Hipple, Patrick Hostert, M. Joseph Hughes, Justin Huntington, David M. Johnson, Robert Kennedy, Ayse Kilic, Zhan Li, Leo Lymburner, Joel McCorkel, Nima Pahlevan, Theodore A. Scambos, Crystal Schaaf, John R. Schott, Yongwei Sheng, James Storey, Eric Vermote, James Vogelmann, Joanne C. White, Randolph H. Wynne, and Zhe Zhu. 2019. “Current Status of Landsat Program, Science, and Applications.” *Remote Sensing of Environment* 225(March):127–47.

Wulder, Michael A., Joanne C. White, Thomas R. Loveland, Curtis E. Woodcock, Alan S. Belward, Warren B. Cohen, Eugene A. Fosnight, Jerad Shaw, Jeffrey G. Masek, and David P. Roy. 2016. “The Global Landsat Archive: Status, Consolidation, and Direction.” *Remote Sensing of Environment* 185:271–83.

Zeileis, Achim. 2005. “A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals.” *Econometric Reviews* 24(4):445–66.

Zhan, X., R. a. Sohlberg, J. R. G. Townshend, C. M. DiMiceli, M. L. Carroll, J. C. Eastman, M. C. Hansen, and R. S. DeFries. 2002. “Detection of Land Cover Changes Using MODIS 250 m Data.” *Remote Sensing of Environment* 83(1–2):336–50.

Zhu, Zhe. 2017. “Change Detection Using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications.” *ISPRS Journal of Photogrammetry and Remote Sensing* 130:370–84.

Zhu, Zhe, Yingchun Fu, Curtis E. Woodcock, Pontus Olofsson, James E. Vogelmann, Christopher Holden, Min Wang, Shu Dai, and Yang Yu. 2016. “Including Land Cover Change in Analysis of Greenness Trends Using All Available Landsat 5, 7, and 8 Images: A Case Study from Guangzhou, China (2000–2014).” *Remote Sensing of Environment* 185:243–57.

Zhu, Zhe, and Curtis E. Woodcock. 2014. “Continuous Change Detection and Classification of Land Cover Using All Available Landsat Data.” *Remote Sensing of Environment* 144:152–71.

Zhu, Zhe, Curtis E. Woodcock, and Pontus Olofsson. 2012. “Continuous Monitoring of Forest Disturbance Using All Available Landsat Imagery.” *Remote Sensing of Environment* 122:75–91.

Zhu, Zhe, Michael A. Wulder, David P. Roy, Curtis E. Woodcock, Matthew C. Hansen, Volker C. Radeloff, Sean P. Healey, Crystal Schaaf, Patrick Hostert, Peter Strobl, Jean-Francois Francois Pekel, Leo Lymburner, Nima Pahlevan, and Ted A. Scambos. 2019. “Benefits of the Free and Open Landsat Data Policy.” *Remote Sensing of Environment* 224(February):382–85.