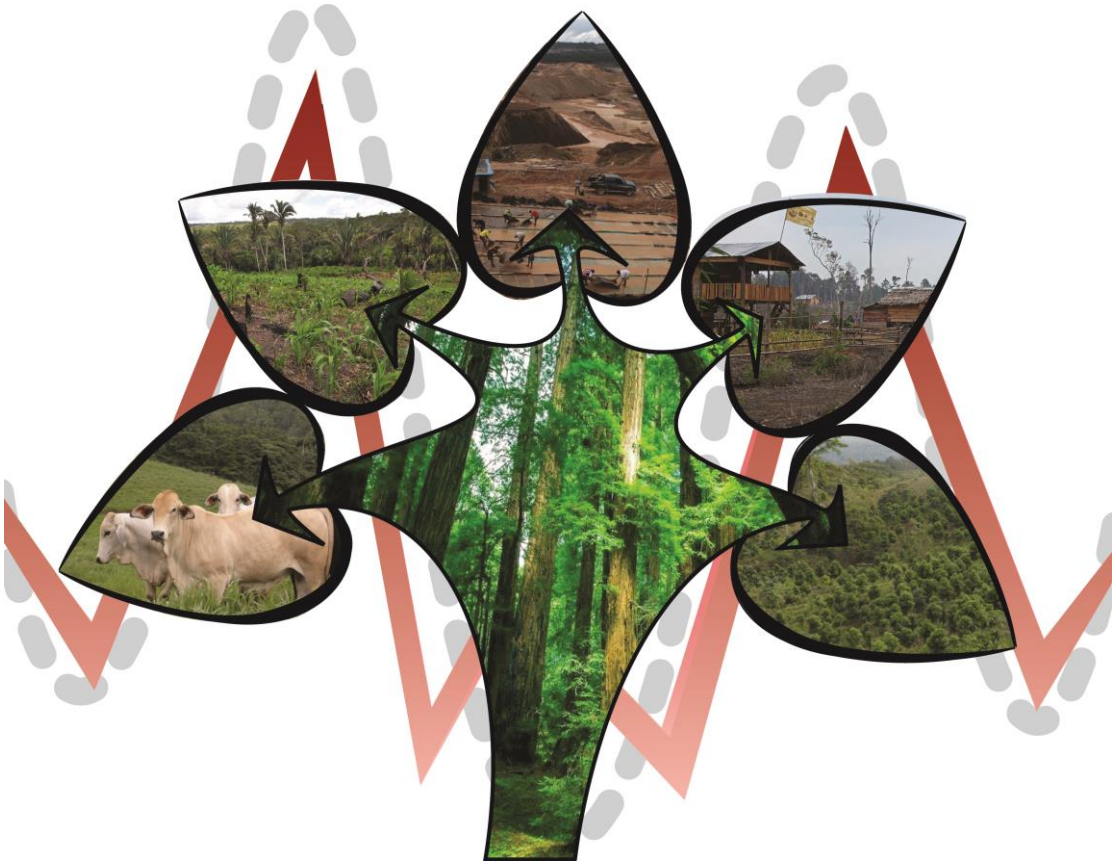
Centre for Geo-Information

Thesis Report GIRS-2015-40

Mapping Post deforestation Land Use in the Brazilian Amazon using Remote Sensing Time Series

Tom Bewernick



November 2015

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Mapping Post deforestation Land Use in the Brazilian Amazon using Remote Sensing Time Series

Tom Bewernick

Registration number 880312064060

Supervisors:

Jose Gonzalez de Tanago Menaca, Laboratory of Geo-information Science and Remote Sensing, Wageningen University

prof. dr. Martin Herold, Laboratory of Geo-information Science and Remote Sensing, Wageningen University

dr. ir. Jan Verbesselt, Laboratory of Geo-information Science and Remote Sensing, Wageningen University

A thesis submitted in partial fulfilment of the degree of Master of Science
at Wageningen University and Research Centre,
The Netherlands.

Date: 16.11.2015

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2015-40
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgements

In genuine love, I would like to express my gratitude to all the people and circumstances that brought me here to this moment of finishing my MSc thesis. I am especially thankful for all the wonderful insights beyond my thesis research in this demanding period.

Thank you Martin Herold and Jan Verbesselt for dedicating your precious time for my supervision. Your open and creative minds provided valuable guidance throughout this thesis. A special thanks also to Jose Gonzales de Tanago Menaca, who helped me most in the development of the topic and the method. Beyond that you supported me tremendously on a personal level, which I deem most valuable. I share your opinions and the motivation behind your actions. May happiness refund all your efforts.

Throughout this thesis I further gained valuable technical support from Ben DeVries, Mathieux De Cuyper, Loic Dutrieux, Michael Schulz, Johannes Eberenz, Benjamin Brede and Simon Besnard. Thank you all for sharing your collective knowledge with me. I hope your willingness for collaboration will accelerate the field of remote sensing as a whole. Thanks also to the R community for providing countless tutorials, hints and powerful software packages. Without you I could not develop the scripts to perform this research. This thesis exemplifies how open software initiatives contribute to knowledge dispersal in society.

Thank you, Efka, for bringing clarity and emotional shelter into my life. You remind me on what is important and help me understand the purpose of my doing. In the time in which we were writing our theses you made me understand that a sense of meaning can only be created when manifesting the inner purpose with action. Through you I can see beyond the work and live my dreams. What an angel you are.

I also want to express my gratitude to my mother, sister and father for their constant unconditional support. Without you I certainly wouldn't be here. I hope that the finalization of my study will bring us closer again.

Abstract

Government policies to populate the Brazilian Legal Amazon induced an ongoing process of forest conversion since the 1960's. This had big implications on climate change as enormous carbon stocks were released into the atmosphere. To understand the underlying forces driving deforestation and to assess anew carbon sequestration, knowledge of post-deforestation land use is crucial. The classification of land use requires multi-temporal remote sensing imagery to differentiate the seasonal characteristics of vegetation in target land use types. However, in the Amazon the availability of high resolution imagery is greatly constrained by persistent cloud cover, prohibiting accurate modelling of seasonal signal variations. To classify post-deforestation land use, we tested a novel approach aggregating Landsat time series stacks into temporally targeted image composites (whole agricultural year, rain season, dry season). Each composite was supplemented with 40 bands summarizing the signal variation of land use classes among a wide range of spectral indices. In order to determine the effect of different cloud masking procedures on the final classification, two independent sets of composites were produced: one processed with a single-date cloud-mask (Fmask), and the other using a multi-temporal cloud mask (Tmask). Target land use types cropland, pastures and secondary vegetation were classified employing a supervised machine learning approach (Random Forests) on temporal metrics of each composite. Better classification performance was achieved for image stacks that were produced with Tmask-derived cloud masks (overall accuracy = 88.8 %). While metrics derived from NBR time series were found to discriminate land use classes best, important predictors generally represented annual/seasonal mean and minimum values. Additionally, the classifier was tested among areas with different time lags since the last deforestation event. Although overall accuracy was lowest in areas with recent deforestation (84.1 %), results did not suggest a linear correlation between classification accuracy and the time lag to the last deforestation event. However, classification results may be strongly biased due to the strong prevalence of pastures among lag groups. Further, it was found that classification accuracy increased with additional observations, especially in the rain season. Accordingly, more accurate land use maps can be generated from data with a higher temporal frequency. The short revisit time and broad spectral coverage make Sentinel 2 a promising data source to embed the method into current forest monitoring systems.

Keywords: remote sensing, post-deforestation land use, image compositing, cloud masking, Random Forest, Landsat

Table of contents

List of figures

List of tables

List of annexes

Abbreviations

BLA	Brazilian Legal Amazon
ETM+	Enhanced thematic mapper (sensor on Landsat 7)
GCP	Ground control points
IBGE	Brazilian Institute of Geography and Statistics (IBGE)(IBGE)
INPE	Brazilian National Institute for Space Research
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
NIR	Near-infrared
LU/LC	Land use/land cover
MODIS	Moderate Resolution Imaging Spectroradiometer
OOB	Out-Of-Bag
REDD+	Reducing emissions from deforestation and forest degradation in developing countries; and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries
RF	Random Forest
RMSE	Root Mean Square Error
RS	Remote Sensing
SWIR	Short-wave-infrared
TCc	Tasseled Cap components
TM	Thematic Mapper (sensor on Landsat 5)
TOA	Top-of-atmosphere
TRMM	Tropical Rainfall Measuring Mission
TS	Time Series
TSS	Time Series Stacks
USGS	United States Geological Survey
VI	Vegetation index
WRS	Worldwide Reference System

1 Introduction

1.1 Background

In the 1960s and 1970s government policies and subsidies to develop, populate and integrate the Brazilian Legal Amazon (BLA) region into the rest of the country, induced extensive and predatory use of natural resources in that region (Araújo and Léna 2011). From that time onwards the rates of deforestation increased drastically, converting 17% of intact forests into other land uses (LU) (Azevedo-Ramos 2008).

Concerns about the implications of deforestation in the Amazon have largely focused on the effects on biological diversity (Dirzo and Raven 2003), changes in regional climate and precipitation (Baidya Roy and Avissar 2002), and global climate change (Houghton et al. 2000). Specifically the latter attracts strong international environmental interest as carbon emissions from deforestation and forest degradation are the second largest source of anthropogenic carbon emissions after the energy sector (IPCC 2007). About 60 to 80 billion tonnes of carbon are stored in the above-ground biomass of the Amazon forest, which exceeds global human-induced emissions in a decade (Azevedo-Ramos 2008). Due to deforestation about 200 million tonnes of these enormous carbon stocks are annually released to the atmosphere which accounts for 3% of global net carbon emissions (Numata et al. 2011).

In the case of Brazil, 70% of the converted forests of the last decades have been replaced by cattle ranching (Soares-Filho et al. 2006). That process was indirectly facilitated through subsidies for cattle ranching, infrastructure investments, land tenure issues and low law enforcement (Azevedo-Ramos 2008). Nowadays, Brazil is the largest beef exporter world-wide (FAO 2006). Due to the northward expansion of Brazilian cattle ranches, previously forested areas in the Amazon make up a big portion of the land (Morton et al. 2006). Despite the low productivity of the soil, clearing for livestock production is yet a profitable means to increase the value of the land (Davalos et al. 2014). Between 2005-2010, the deforestation rates rapidly decreased due to the expansion of protected forest areas and Brazilian land stewardship programmes (Nepstad et al. 2008). However, the rising international demand for biodiesel and soy beans recently boosted the expansion of large-scale agriculture in Brazil (Azevedo-Ramos 2008). This in turn caused a new rise of deforestation rates in the Amazon refuting the claims that cropland expansion does not go at the expense of tropical rainforest

(Morton et al. 2006). In order to adjust national policies to these developments, it is necessary to inform policy makers about the specific LU driving deforestation in the Amazon.

The UN-program to Reduce Emissions from Deforestation and forest Degradation and enhancing forest carbon stocks (REDD+), which was introduced in 2007, poses a potential incentive to enhance Brazils monitoring efforts. In part, the program proposes compensation for the conservation or reestablishment of carbon stocks through appropriate forest management. For the REDD+ implementation on national level, changes in carbon stocks have to be adequately monitored as required for national Measuring Reporting and Verification (MRV) systems. In fact, there is already an implemented forest monitoring system, namely the PRODES project, providing estimates on deforestation rates and annual forest change maps since 1988. PRODES data allows to assess the scale of the forest conversion, and therefore, can be used to derive rough estimates for carbon emissions (IPCC 2006). However, the product does not provide any information on the specific land use following a deforestation event, which is also required for national REDD+ strategies (UNFCCC 2011).

Data of the immediate post-deforestation land use is not only relevant for the assessment of deforestation drivers (but also to assess anew sequestration of atmospheric carbon in the biomass of regrowth vegetation or in soils (Morton et al. 2006). Furthermore, the type and duration of land use has significant implications on the recovery of forests in case of abandonment (Guariguata and Ostertag 2001).

1.2 Problem definition

The Brazilian Institute of Geography and Statistics (IBGE) produces repeated agricultural census statistics, but without enough detail to determine land use at the spatial scale of deforestation patches. The Global Land Cover maps (2000, 2009, 2012) produced by the FAO are spatially explicit but the temporal coverage of these datasets is too sparse to identify immediate post-deforestation LU. The Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover (2001-2011) maps are produced annually but the coarse spatial resolution of 500m does not allow to capture heterogeneous LU dynamics which occur particularly in temporary smallholder agriculture systems in the eastern Amazon (Aguiar, Câmara and Escada 2007). The TerraClass project of the Brazilian National Institute for Space Research (INPE) explicitly mapped the follow-up land use of cleared areas in the

Amazon at a spatial resolution of 30m. However, the maps are only available for the years 2008 and 2010 and the respective deforestation events date back to the year 2000. That implies that changes potentially occurred in the meantime and that TerraClass maps do not necessarily inform about immediate post-deforestation LU. Accordingly, there is currently a lack of repeated data demonstrating the follow-up LU after deforestation events in the Amazon region at adequate spatial detail.

Optical satellite imagery has been widely employed for the classification of land cover (LC). Most commonly, classification approaches match the spectral information of a mapping unit with the spectral signature of known LC (Lu and Weng 2007). However, these approaches are designed to produce single-date LC maps and require new training areas for repeated production. Additionally, LU is defined more broad than LC integrating aspects beyond the biophysical properties of the land surface. Some of these aspects become apparent though when looking at the temporal behaviour of LC, such as annual bare soil and vegetation-residue cycles in crop fields. In order to capture such generic cycles the seasonal signal variations can be modelled from multi-temporal Remote Sensing (RS) data with hyper-temporal resolution. Many approaches model the seasonal variation of the RS signal with sine functions or wavelet filters and then classify LU/LC based on curve typology and additional statistical parameters (Conrad et al. 2011, Petitjean, Inglada and Gancarski 2012, Viovy 2000). In order to distinguish inter-annual variations from noise, RS time series (TS) with a high temporal frequency are required. Recent changes in the data policy of the U.S. Geological Survey (USGS) and advancements in pre-processing algorithms allow to harness the full Landsat archive, which greatly enhances the availability of high-resolution imagery (30m) (Wulder et al. 2012).

(Zhu and Woodcock 2014b) developed an approach that combines both inter-annual Landsat observations for the continuous classification of LULC types, and intra-annual observations for the detection of LULC changes. In theory, such an approach may accomplish both detection of deforestation and classification of follow-up LULC. However, in the tropics the consistency of Landsat TS is greatly constrained by persistent cloud cover, reducing the number of valid observations to a fractional amount of the original data (Herold 2009). Therefore, the temporal density of Landsat observations is too scarce to model seasonal variation as sine curves in the tropics. Several studies integrated radar data as supplementary information which is generally not impaired by cloud cover (Kuplich 2006, Lu, Batistella and

Moran 2007). However the optical monitoring systems of the Brazilian Amazon could not be complemented by radar yet, and the acquisition of dense radar TS remains very costly. For the characterization of post-deforestation LULC in the tropics novel approaches are required to condense the information of the optical RS TS in such a way, that they capture the relevant temporal characteristics of individual LU/LC types. (Griffiths et al. 2013) developed a method for the classification of agricultural LU by compositing cloud-free images from all available observations of Landsat images at annual key stages of crop development and quantifying the temporal patters among reflectance bands and vegetation indices (VI). However, complete cloud-free composites still require a sufficient amount of cloud-free imagery, which is hardly available in tropical regions.

To gain enough reliable RS observations for the characterization of seasonal and annual patterns accurate cloud masking is required. Falsely unmasked clouds and cloud shadows influence optical sensors to a degree that the LU/LC of a pixel can be unidentifiable (Zhu and Woodcock 2012). Landsat land surface reflectance data provided by US Geological Survey (USGS) is already supplemented with a cloud and cloud shadow mask, namely the FMASK band. For every scene the FMASK band has been generated by a single-date algorithm developed by Zhu & Woodcock (2012). However, due to the limited spectral coverage of Landsat bands the threshold-based FMASK algorithm fails to detect identify all clouds, especially when different types of clouds prevail in one image (Zhu and Woodcock 2014a). Multi-temporal cloud detection algorithms often perform better because the magnitude of changes in successive images provides another indicator for the occurrence of clouds (Goodwin et al. 2013). On the other hand, masking pixels which underwent sudden changes in the spectral signature bears the possibility of masking out LC changes. Zhu & Woodcock (2014) developed a cloud detection methodology which excludes LC changes from cloud detection, namely the TMASK algorithm. The authors claim that the algorithm generates better cloud/cloud shadow mask than Fmask. However, Tmask was developed for Landsat TS with high data density. To our knowledge no research has been published where this algorithm has been tested in tropical regions with lower density of available imagery. Therefore, it is uncertain which cloud masking method is most beneficial for the spectro-temporal classification of post-deforestation LU.

2 Objectives and research questions

The main objective of this study was to develop an automatic method to classify LULC after deforestation events in the tropics using optical RS TS. In this context, it was aimed to assess different methodological choices (cloud masking approach and spectral indices) to describe the temporal patterns of multiple spectral indices for LULC classification. Further, we evaluated the performance of the developed method for classifying LU for pixels with different time lags to the last deforestation event. These objectives translate into the following research questions:

1. Does the TMASK cloud masking algorithm yield a better classification of low density TS than TS processed with FMASK?
2. Which spectral indices and temporal metrics bear the most relevant information for predicting LULC?
3. Is the developed classification approach affected by the time lag to deforestation events?

3 Study area

The LULC classification method was tested in the eastern part of Paragominas (Para, Brazil, Lat. 2.99° S, Long. 47.36° W, 200m above sea level) located in the south-eastern delta of the Amazon. The transitional climate of the region is characterized by periods of rainfall (December-May) and drought (July-January). The choice for this particular test site was legitimated by the higher availability of cloud-free RS data during the dry season compared to the central and western Amazon region which is denoted by higher persistency of clouds.

In the past decades the Paragominas was marked by rapid LULC changes. During the 1960's the Brazilian government encouraged cattle production in Amazonia through low-interest loans and other financial incentives (Hecht, Norgaard and Possio 1988). In the following decades Paragominas became one of the centers of the emerging cattle and timber industry (Verissimo et al. 1992, Nepstad 1989). However, in the course of the years many pastures degraded leaving behind a complex mosaic of shrubland, secondary forests and extensive pastures (Serrao and Toledo 1989). Nowadays, favourable market conditions are driving large-scale landholders to shift towards intensive soy, corn and rice production (Balazs 2001, Morton et al. 2006). Simultaneously, due to competitive disadvantages and the increasing unavailability of profitable land small landowners sell or abandon their land (Balazs 2001). This socio-economic framework determines a highly dynamic LU mosaic, making this area a suitable pilot site to set up and test the proposed LU classification method.

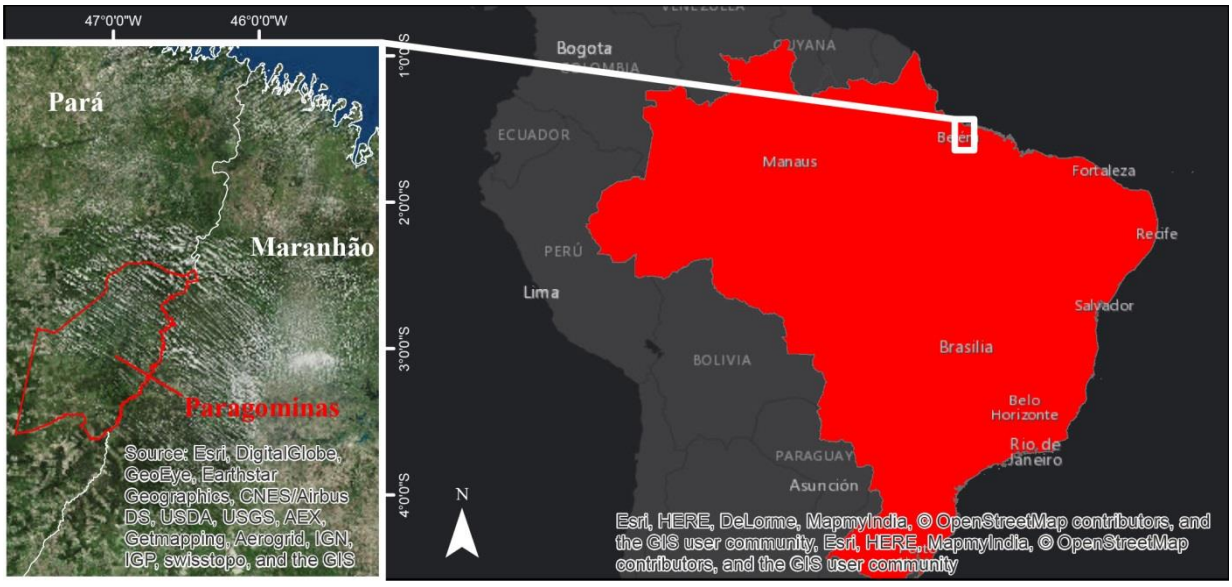


Figure 1: Location of study area, Paragominas (Para, Brazil)

4 Materials

4.1 Data

4.1.1 Land use reference data

As part of the TerraClass project, the Brazilian National Institute for Space Research (INPE) produced two LU maps for the Brazilian Legal Amazon (BLA) for the years 2008 and 2010. The target polygons in these maps were derived from PRODES data and indicate only the LU in previously deforested areas. The actual classification approach is based on a mix of interpreting spectral responses, texture and history of the area with the aid of grey image slicing. Landsat imagery was employed as input data. Consequently, the maps have a spatial resolution of 30m, which makes TerraClass a suitable reference for LULC analysis with Landsat. TerraClass maps were not generated automatically, employing high manual efforts and expert knowledge for visual interpretation of image composites. Assuming a high reliability of this approach, the accuracy of TerraClass maps was not directly assessed. Producers claim that the data is an accurate representation of the real LULC, which was verified with helicopter flights (Almeida et al. 2009).

TerraClass distinguishes 9 different LU classes which are listed in Table 1. Not all LU types are present in the study area. Water bodies were not taken into account as they were masked from Landsat images during the pre-processing procedure. Furthermore, reforestation was not considered, because the class was only incorporated in the TerraClass map 2010. The mapping requirements are further clarified in section 5.2. Urban LU was excluded from the analysis as well because too few pixels of this class were present in the study area for an elaborate investigation. Four pasture classes distinguished by TerraClass were aggregated into one pasture class due to the strong prevalence of "intensive pastures" among all pastures (98.5 %).

Table 1: Distribution of LU classes in the study area

LU type	Mean patch size (ha)	Total area (ha)	Share of total area (%)
Agriculture	**166.18**	**33236.06**	**14.8**
Pasture (aggregated)	**43.38**	**100556.21**	**44.4**
Intensive pasture	55.72	99069.34	43.9
Extensive pasture	1.73	410.45	0.1
Degraded pasture	3.55	1076.41	0.4
Urban	81.75	2289.13	1.0
Secondary vegetation	**8.06**	**89919.51**	**39.8**

The final LU classes which were examined in this study are described below.

4.1.1.1 Cropland

Biophysically, this LU type is generally characterized by bare soil in the beginning of the season, followed by a rapid increase in vegetation growth during the first half of the agricultural year. During or after harvesting the vegetation is suddenly removed, sometimes followed by a second cycle of cultivation within one year. In the case of double cropping cycles key stages in the crop development may be temporally shifted. TerraClass merges areas with single and double cropping cycles in one class.

From a RS perspective, cropland is characterized by high annual signal variation of vegetation indices. Furthermore, the presence of bare soil is readily identifiable in the TS of most spectral indices and marks a major distinctive feature to other LU types with constant vegetative cover.

Figure 2: Aerial photographs of cropland (source: Almeida, 2009)

4.1.1.2 Pastures

TerraClass distinguishes 3 different pasture classes in the study area: intensive, extensive, and degraded pasture. The types vary especially in terms of woody vegetation cover. Depending on the grazing intensity (i.e. stock density per area) vegetative ground cover ranges from short grass to early stages of scrub encroachment. Furthermore, some pasture lands are burned or fertilized for temporary new growth; others are abandoned and eventually turn into either forest regrowth or unproductive bush fallow (Da Veiga et al. 2003).Therefore, a broader range of spectral and temporal characteristics might be observed for pastures.

In contrast to cropland, annual vegetation changes of pastures come about gradually, implying a lower variability of the RS signal especially when looking at a smaller time frame (e.g. single season). Pastures might also display patches of bare soil. However, this feature is only present in the class "degraded pasture"(Almeida et al. 2009), which hardly exist in the study area (see Table 1).

Figure 3: Examples of pastures (source: Almeida, 2009)

4.1.1.3 Secondary vegetation

Secondary vegetation refers to spontaneous forest regrowth as a consequence of land abandonment or longer cycles of timber production (Almeida et al. 2009). This LU class features an ongoing process of biomass increase and ranges from early successional stages to secondary forest (ibid.).

Due to the relatively high above-ground biomass levels, secondary vegetation is characterized by consistently high values of vegetation indices. However, designed to respond to photosynthetic activity, most vegetation indices already saturate at low biomass levels. This

potentially hampers the differentiation to other LU types (Birdsey et al. 2013). RS literature suggests that spectral indices sensitive to moisture are more suitable to distinguish vegetation with high biomass levels (Xiao et al. 2002, Fiorella and Ripple 1993, Jin and Sader 2005). Therefore, it is assumed that secondary vegetation shows a more unique response in the shortwave-infrared (SWIR) band and related spectral indices (Frazier et al. 2014, He et al. 2011).

Figure 4: Aerial photograph of secondary vegetation (source: Almeida, 2009)

4.1.2 Data on deforestation years

As part of the Amazon Deforestation Monitoring Project (PRODES), digital deforestation maps were generated annually since 2000 at a resolution of 90m (Espaciais 2002). The PRODES methodology employs spectral mixing algorithms, image segmentation and unsupervised classification for the identification of soil, vegetation and shade fractions from which change maps are derived (Shimabukuro et al. 1998).

This data complements LU data derived from TerraClass with the information of when a particular LU patch was deforested. Due to persistent cloud cover PRODES does not identify the exact year of deforestation in all areas. In these cases the maps do not indicate a single year of deforestation but rather a range of possible years (Câmara, de Morisson Valeriano and Vianei Soares 2006). This was partly resolved by reclassifying PRODES deforestation years into two-year intervals incorporating only pixels with certain years of deforestation and with an uncertainty of one year. The latter were only included in case the range of possible deforestation years falls exactly in the pre-defined two-year intervals. All pixels with an uncertainty of more than a year were excluded from the analysis. In order to match the resolution and origin of Landsat and TerraClass data, PRODES maps (originally 90m resolution) were resampled to 30m.

4.1.3 Remote sensing data

Optical RS data for this study was comprised of Landsat images from the TM (Thematic Mapper) and ETM+ (Enhanced Thematic Mapper) sensors, provided by the USGS, for a period of observation between 01.07.2009 – 01.07.2010. The test site was demarked in such a way that it falls completely within the extent of one WRS (Worldwide Reference System) location (path=222; row=62). Although this area is covered completely by one WRS tile it is also covered partially by the neighboring tiles. Therefore, in order to maximize the amount of RS observations, all available images from three tiles were downloaded (path/row=222/62; path/row=222/63, and path/row=223/62). Landsat data was downloaded via the ESPA ordering interface (https://espa.cr.usgs.gov). An overview of all imagery used in this study is given in Annex 1.

Due to a slight shift of satellite orbits, the coverage of Landsat images representing the same WRS tile is slightly variable (Huang et al. 2009a). That means observations on the edge of an image do not necessarily lie within the extent of another image with the same WRS location. In order to preserve data for processing, images were expanded to one common area mask encompassing all available observations. Subsequently, the expanded images were clipped to the extent of the study area. This procedure was conducted separately for images within the same WRS location or Landsat tile, producing three sets of images with matching geographical extent.

At the moment of acquisition the images were already radiometrically and atmospherically corrected by the automated modules in the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006b). Thus the employed RS data contains scenes processed at the levels L1T (full Terrain Correction), or in case of high cloud cover or absence of a digital elevation model, L1G (Systematic Correction) (Roy et al. 2010). Further pre-processing steps implemented in this study are described in section 5.1.

The choice for Landsat imagery was mainly based on its free availability and the relatively high spatial resolution of 30m which allows for the identification of fine-grain LU mosaics such as in smallholder agriculture. Furthermore, Landsat acquisitions cover the whole globe and have the longest historical record of all RS systems (Harris et al. 2012). Also the spectral characteristics of Landsat allow to differentiate various LU types (Griffiths et al. 2013, Yang and Lo 2002).

4.2 Software

Most computations in this thesis were performed in the R programming environment (R Core Development Team 2015). For the processing of RS data mainly R-packages "BfastSpatial", "zoo" and "raster" were used. The Tmask algorithm was implemented using stand-alone software provided via https://code.google.com/p/tmask-algorithm/ (Zhu and Woodcock 2014a). LU classification and statistical analysis of results were performed using the R-packages "randomForest" and "party". Graphical outputs were mainly generated using the "ggplot2" packages and ArcMap software (v.10.2).

5 Methodology

This chapter describes the conducted processing and analysis steps to reach the defined thesis objective. The main procedure is divided into three steps which are illustrated in Figure 5. Pre-processing (section 5.1) addresses the preparation of Landsat data for the pixel-wise computation of annual and seasonal summary metrics. This step produces two sets of TSS representing different spectral indices, which were computed with different cloud masking procedures. Each set was then segmented into annual and seasonal image stacks, for which pixel-based summary statistics were computed (section 5.2). These metrics were used as input variables for a Random Forest (RF) classification model (section 5.3). This classifier was applied to generate two post-deforestation LU maps, one from data processed with Fmask, and one from Tmask data.

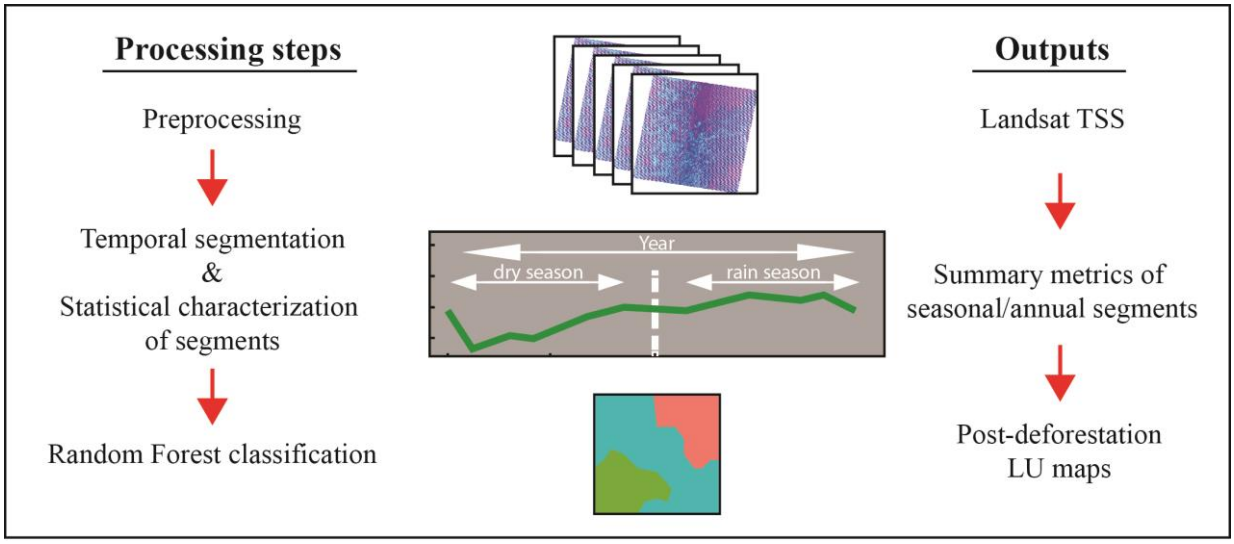


Figure 5: Main processing steps and resulting outputs

5.1 Landsat pre-processing

The general chain of implemented pre-processing steps is presented in Figure 6. Geometric correction, extraction of spectral bands and derivation of spectral indices was applied generically to all Landsat scenes. Subsequently, groups of partially overlapping images were stacked, and purged of clouds and cloud shadows. The procedures are explained in more detail in the following subsections.

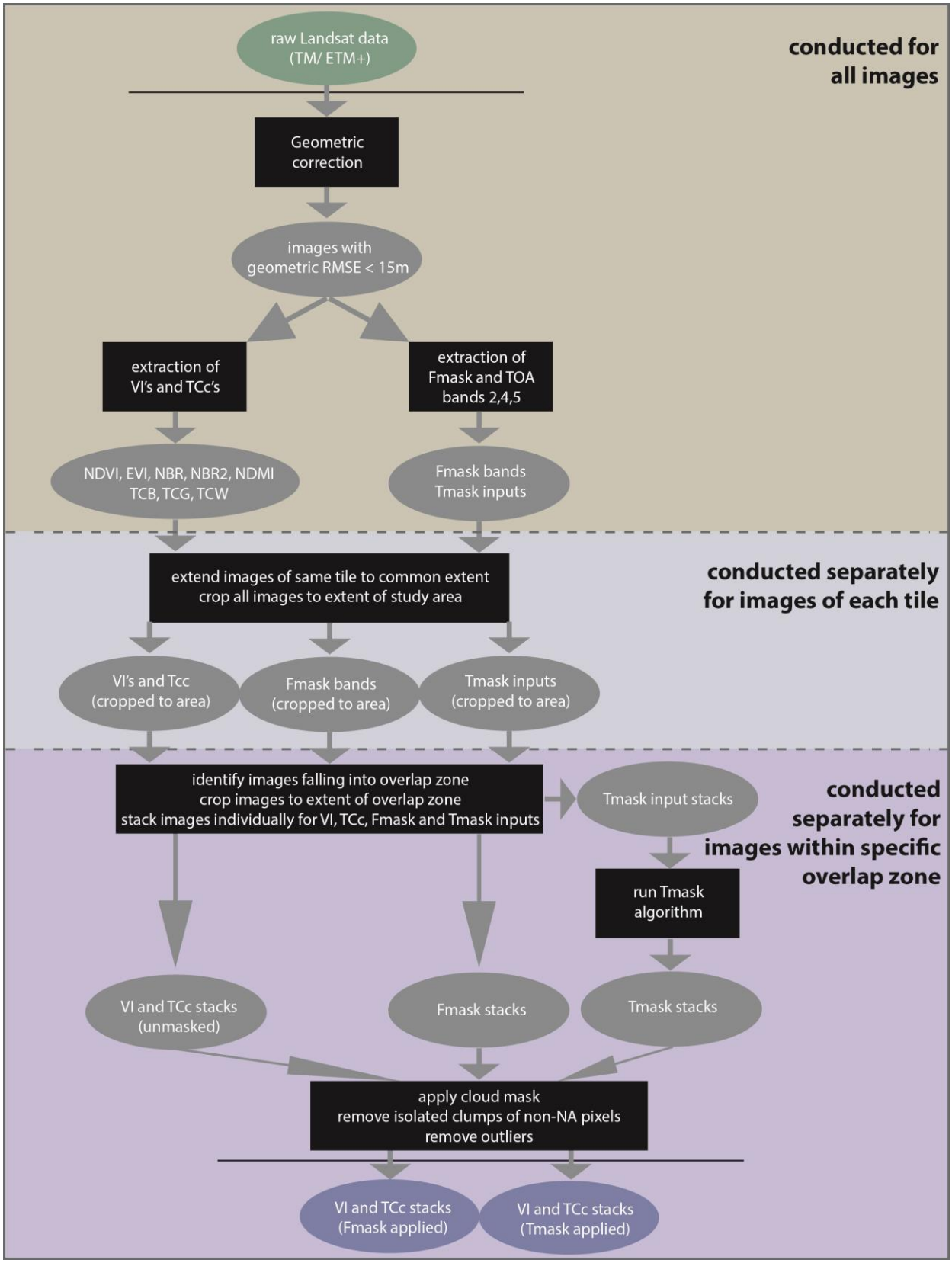


Figure 6: Workflow for Landsat pre-processing and creation of Time Series Stacks

5.1.1 Geometric correction

The geometric registration accuracy of all Landsat images was provided by the USGS with the metadata as the Root Mean Square Error (RMSE) of Ground Control Points (GCP) in x and y direction. High geometric accuracy is particularly important for the temporal analysis of RS imagery, as pixels representing exactly the same geolocation need to be stacked among different images. Already a shift of one pixel (~30m in x-y direction) could insert a non-representative observation into the TS of the pixel (Townshend et al. 1992). Therefore, all imagery with a geometric RMSE above a threshold of half a pixel (15m) was excluded from further the analysis. In particular, images at processing level L1G were excluded, as they do not guarantee sub-pixel accuracies (Lee et al. 2004).

For many other images no geometric accuracy information of the GCP model was available. These scenes suffer excessive cloud cover which precluded the possibility to detect the land GCP by the geometric correction algorithm used by the USGS (Gonzalez de Tanago Menaca 2012). Also these scenes were excluded.

5.1.2 Extraction of spectral indices

From the raw data a multitude of vegetation indices (VI), wetness indices (WI), burn indices (BI) and Tasselled Cap components (TCc) were extracted. The respective indices including the equations from which they were derived are listed in Table 2.

NDVI, EVI and NDMI were already provided as ready-to-use imagery with the downloaded data (Masek et al. 2006a). Exploiting the difference between the red and NIR band, NDVI and EVI are both sensitive to photosynthetic activity of vegetation (Huete et al. 1997, Xu et al. 2011). For that reason they are frequently used for the classification of LULC (Hansen et al. 2000). However, NDVI and EVI both tend to saturate at moderate biomass levels (Huete et al. 2002). Therefore, they are rather suitable for the distinction of vegetation types with low-biomass and for the detection of bare soil, such as in agricultural LU (Wardlow, Egbert and Kastens 2007). In contrast, NDMI incorporates the difference between Landsat SWIR and NIR bands, and is therefore more sensitive to moisture content rather than photosynthetic activity. That bears the advantage that LU classes with high biomass levels, such as different stages of forest succession (in this thesis defined as secondary vegetation), can be better discriminated (Fiorella and Ripple 1993).

To support the differentiation of secondary vegetation, a number of alternative spectral indices, commonly used in forest monitoring, were included. Normalized Burn Ratios (NBR and NBR2) were also provided with the downloaded data. Additionally, we performed a Tasseled Cap Transformations to merge information of all Landsat bands into three indices: Tasseled Cap Greenness (TCG), Tasseled Cap Brightness (TCB), and Tasseled Cap Wetness (TCW). For the transformation the coefficients defined for surface reflectance data (Crist 1985) were used.

Table 2: Vegetation indices and Tasseled Cap components

Index	Equation	Source
Vegetation indices		
NDVI	$= \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red}}$	(Tucker 1979)
EVI	$= \text{G} * \frac{\text{NIR} - \text{red}}{\text{NIR} + C_1 * \text{red} - C_2 * \text{blue} + \text{L}}$	(Huete, Justice and Liu 1994)
Wetness indices		
NDMI	$= \frac{\text{NIR} - \text{SWIR1}}{\text{NIR} + \text{SWIR1}}$	(Horler and Ahern 1986)
Burn indices		
NBR	$= \frac{\text{NIR} - \text{SWIR2}}{\text{NIR} + \text{SWIR2}}$	(Key and Benson 1999)
NBR2	$= \frac{\text{SWIR1} - \text{SWIR2}}{\text{SWIR1} + \text{SWIR2}}$	unknown
Tasselled Cap components		
TCG	$= -0.1603 * blue - 0.2819 * green - 0.4934 * red + 0.7940 * NIR - 0.0002 * SWIR1 - 0.1446 * SWIR2$	(Crist 1985)
TCB	$= 0.2043 * blue + 0.4158 * green + 0.5524 * red + 0.5741 * NIR + 0.3124 * SWIR1 + 0.2303 * SWIR2$	(Crist 1985)
TCW	$= 0.0315 * blue + 0.2021 * green + 0.3102 * red + 0.1594 * NIR - 0.6806 * SWIR1 - 0.6109 * SWIR2$	(Crist 1985)

5.1.3 Cloud masking

Downloaded Landsat data was already provided with the Fmask band, which can be used to mask out clouds and cloud shadows imagery (Zhu and Woodcock 2012). The Fmask band originates from a single-date algorithm which employs a probability mask and a minimum threshold for cloud detection. Furthermore, the outside edges of identified clouds are dilated by 3 pixels. Therefore, the Fmask product is provided at a resolution of 90m.

Additionally, a second cloud mask was produced, using the Tmask algorithm. Tmask benefits from the initial cloud masking provided by Fmask and additional multi-temporal information of the top-of-atmosphere (TOA) reflectance bands 2,4 and 5. Tmask employs a pixel-based TS model consisting of sines and cosines to estimate TOA reflectance in the respective bands. If an observation, initially identified as "clear" by Fmask, deviates more than 0.04 (or 400, if DN values are used) from the predicted/modelled TOA reflectance values the pixel is masked as cloud or cloud shadow. The TS model also anticipates LC change on a pixel, preventing it to be falsely masked. Tmask is capable of detecting thin cloud edges and their shadows, and therefore, was implemented without dilation. Accordingly, Tmask produced a pixel level cloud mask with a spatial resolution of 30m (Zhu and Woodcock). The algorithm was implemented separately for each overlap zone, following the instructions provided via https://code.google.com/p/tmask-algorithm/ (Zhu and Woodcock 2014a).

Both downloaded Fmask and computed Tmask layers were used to mask out clouds and cloud shadows from raw Landsat images, producing two image stacks for each overlap zone. Additionally, the cloud masks for each image were further refined by filtering out small clumps of unmasked observations that were surrounded by masked pixels. The maximum area threshold for these unmasked "islands" to be deleted was 0.45 ha, equivalent to five Landsat pixels (see Figure 7).

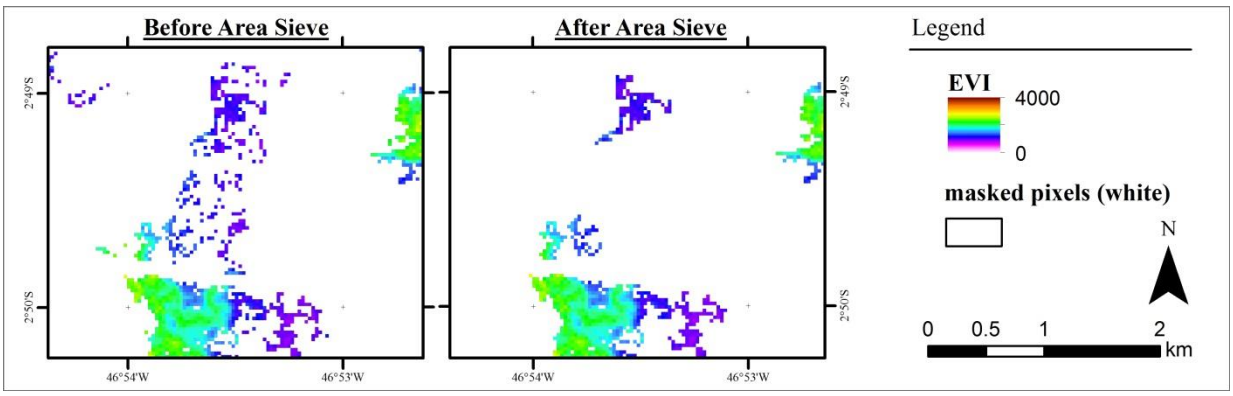


Figure 7: Implementation of area sieve

5.1.4 Outlier removal

A number of factors introduce noise into TS derived from Landsat imagery, such as sun angle, sensor drift, atmospheric condition, and geometric misregistration (Kennedy, Yang and Cohen 2010) Depending on the nature of noise and statistical descriptors of the TS, outliers may have significant influence on the LU classification result. Due to the low density of RS observations in the study area multi-temporal outlier detection or smoothing algorithms (e.g. sine or wavelet filtering) were not employed, in order to avoid that temporally isolated observations would be falsely identified as invalid. Therefore, an alternative approach was implemented in this study.

Histograms were computed individually for each spectral index from a single non-cloudy Landsat scene. From these histograms the probability distribution was estimated using non-parametric kernel densities (see Annex 2). Essentially, kernel density plots resemble a smoothed histogram from which global minimum and maximum thresholds for acceptable observations were determined. These thresholds were set visually, identifying only observations with extreme deviation from all other values, regardless of the apparent LC. Admittedly, this method only addresses few observations, allowing a high level of noise to remain in individual pixel TS. However, that enables us to compare the performance of Fmask and Tmask, because a number of outliers (e.g. due to cloud shadows) originate from deficient cloud masking.

5.1.5 Construction of Time Series Stacks (TSS)

Sequences of Landsat images addressing spectral bands or indices can be bundled into ready-to-use Time Series Stacks (TSS). TSS are acquired at nominal temporal intervals for a specific WRS location and enable the pixel-wise computation of any type of TS analysis (Huang et al. 2009b). With the implementation of LEDAPS processing prior to the publication of downloadable data, the production of TSS including data of Landsat TM and ETM+ sensors became a fairly simple procedure.

However, one important drawback of using Landsat data is the low temporal resolution. Despite the revisit time of 16 days, persistent cloud coverage limits the availability of valid surface reflectance data to a fractional amount. Therefore, only few, often temporally isolated, observations are recorded for the rain season. Moreover, the malfunction of the Scan Line

Corrector (SLC) in 2003 causes systematic data gaps in all Landsat 7 images which were used in this study.

In order to enhance the data availability for the reference period, Landsat imagery of neighbouring WRS tiles overlapping with the study area was included. Four overlap zones were defined, each with a unique set of overlapping images (see Figure 8). The extent of each zone was outlined manually. Minima/maxima of x and y coordinates of zone extents were determined visually and are documented in Annex 3. Overlapping images were then cropped (and/or extended) to the extent of each overlap zone and stacked producing 32 TSS (8 spectral indices * 4 zones). Through this procedure many additional observations were gained (see Figure 9). The zone defined as "NE" displays the amount of data that would be available without the inclusion of neighbouring tiles.

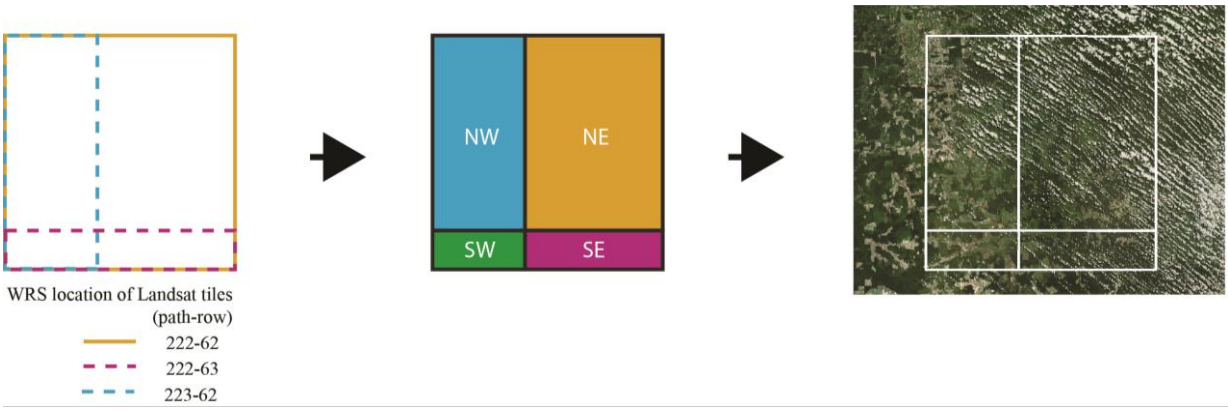


Figure 8: Overlap zones of Landsat images from different WRS tiles

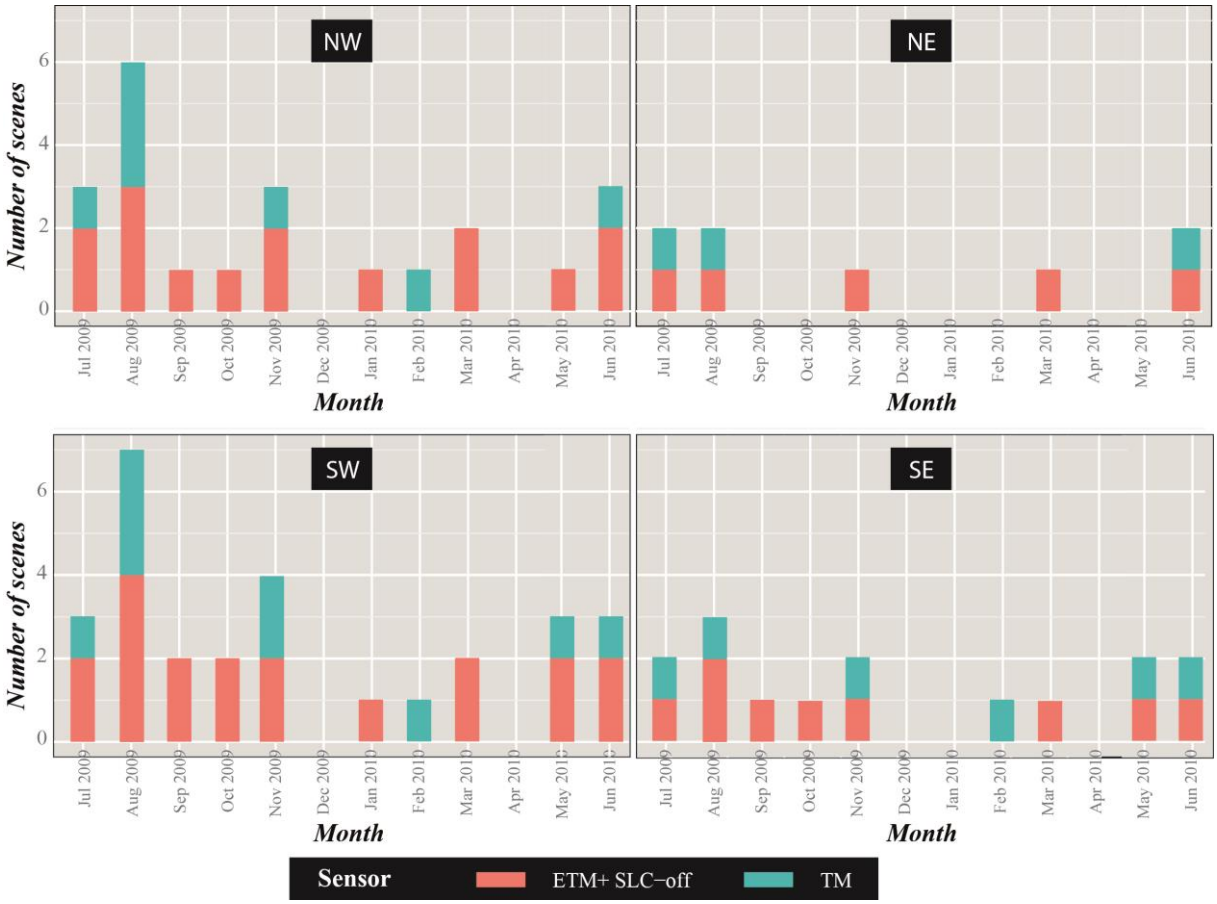


Figure 9: Available Landsat images for zone-specific TSS

5.2 Temporal segmentation and statistical summary metrics

In this study, TSS of spectral indices were temporally segmented, producing three TSS representing the whole agricultural year, rain and dry season (see Figure 10). The agricultural year was temporally bounded by the beginning and the end of the cropping cycle, derived from annual minima of MODIS EVI2 TS. The onset of rain- and dry season was identified from TRMM (Tropical Rainfall Measuring Mission) TS. Both EVI2 and TRMM data were obtained using the TS visualization tool provided by INPE (http://www.dsr.inpe.br/laf/series/en/map.php). The segmented TSS can be conceptualized as extended snapshots which complement each other by specifying particular seasonal or annual characteristics of spectral indices.

For the purpose of LU classification simple statistical metrics were used to characterize the spectral–temporal variability of target LU within each segment (see Figure 10). Altogether 120 parameters were extracted (8 indices * 5 variables * 3 segments). The selection criteria for individual metrics are based on their representativeness of seasonal and annual LC characteristics. Seasonal minimum and maximum values were used mainly to identify bare

soil (e.g. in cropland) and high biomass levels (secondary vegetation) (Hansen et al. 2011). Additionally, standard deviation and amplitude were extracted to characterize the variation of the vegetative cover due to phenology or disturbances (pasture). Annual and seasonal mean values were used to distinguish LU's with generally high (secondary vegetation) and low (pasture) biomass levels.

Due to the use of statistical summary metrics the dimension of time was largely eliminated and merely retained in the target period of each segment. This improves computation efficiency and potentially minimizes the influence of outliers on classification. Another advantage of the method is that specific seasonal characteristics (i.e. crop emergence, ploughing) can be identified and significantly enhance the differentiation of specifically cropland (Prishchepov et al. 2012).

Conceptually, the method is closely related to the work of (Griffiths et al. 2013). For the purpose of agricultural LU classification they produced seasonal image composites, supplemented with many bands summarizing the spectral characteristics within spring, summer and fall. Our approach differs from the latter as it segments the year only in rain and dry season. Segmenting more than two seasons was not possible due to limited data availability. Further, we included the whole agricultural year to gain more reliable mean values of spectral indices.

One limitation of our approach is that temporal segments are bounded by the agricultural year instead of the calendar year. Thus, unlike TerraClass maps the reference period for the proposed classification product spreads over two years: 2009 and 2010. TerraClass maps refer only to the years 2008 and 2010, which means that LU changes during the second half of 2009 are not visible. As the proposed classification approach aims to identify a single LU type for the period between 01.07.2009 – 01.07.2010, only pixels with known stable LU in that time could be used to apply the method. Unchanged pixels were located by comparing TerraClass maps for 2008 and 2010. Incorporated only in TerraClass maps 2010, the LU type reforestation had to be completely eliminated from the analysis to avoid the introduction of uncertainty.

Additionally, most of the statistical metrics require at least 2 observations in the respective temporal segment. Therefore, we further limited the classification approach to pixels with at least 2 observations in rain- and dry season.

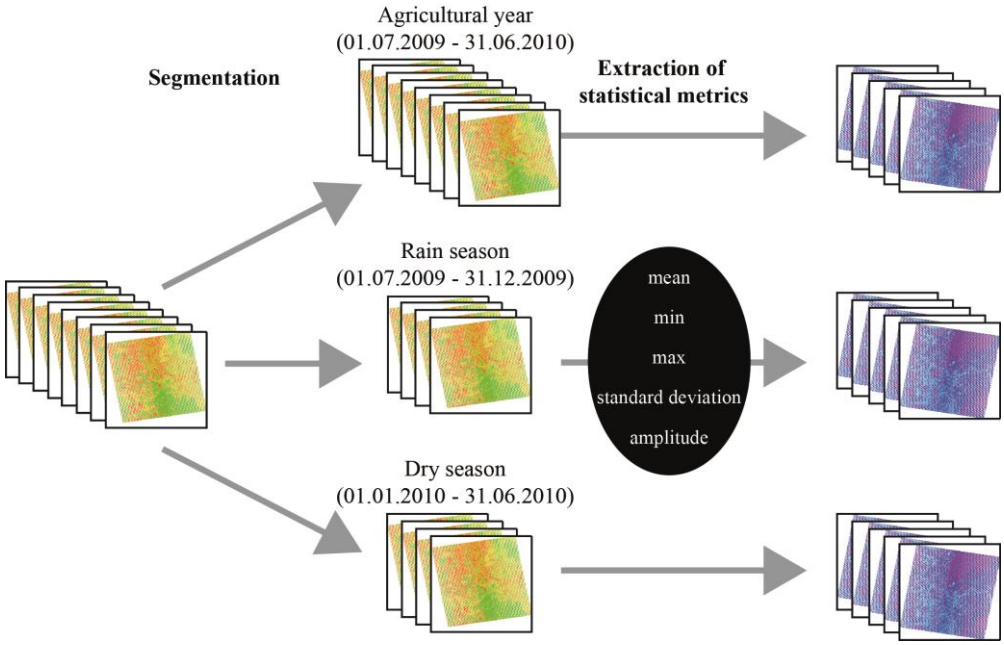


Figure 10: Concept of temporally targeted Time Series Stacks

5.3 Random forest classification

The statistical descriptors of all three TSS were used as input variables for a Random Forest classification model. Among many multivariate machine learning techniques, the RF algorithm is one of the most robust non-parametric methods (Breiman 2001).

RF is an ensemble method originating from Bagging or Bootstrap aggregation techniques. RFs fit a committee of decision trees to a training dataset, each casting a vote for the predicted class. Compared to other bagging methods, individual trees in the classifier don not only use a random subset of training samples (in-bag), but also a random subset of input variables. These subsets are compiled with the aim to create a diverse set of de-correlated trees. Essentially, each decision tree represents a noisy but unbiased model. By averaging them, variance is greatly reduced without a substantial increase in model bias.

The trees consist of splitting nodes which employ different sets of predictors. In the tree growing procedure the most informative variables are weighted at each node and iteratively tuned to define the optimal path for the final prediction. Once all decision trees are grown, each tree is weighted and the splitting parameters among them are averaged. In this way, variables are selected to optimally fit the desired class. Additionally, unused training samples in each tree (out-of-bag) are used for internal cross-validation throughout the building of the classifier. Given a high level of randomness (few trees) and enough training samples this so

called Out-Of-Bag (OOB) error represents an unbiased estimate of classification accuracy (Hastie, Tibshirani and Friedman 2009, Breiman 2001).

Unlike many other classification algorithms, RFs can effectively deal with high levels of variable interaction and collinearity. Given the big amount of independent variables in this study, it was further expected that variables demonstrate different types of distribution. Being a non-parametric decision tree classifier, Random Forests do not require standardization or any assumptions about the probability distribution of the assessed variables. (ibid.)

Another advantage of RFs is that they correct for decision trees' habit of overfitting by building a large set of trees, each using different subsets of original training data (Breiman 2001, Hastie et al. 2009). The claim that this holds generally true was contested by the argument that overfitting is prevented only if trees in the ensemble are limited and enough training samples are provided (Segal 2004, Statnikov, Wang and Aliferis 2008). More information is provided in section 5.3.2.

5.3.1 Training

Due to the availability of reference data the RF algorithm could be implemented in supervised mode, without time-consuming delineation of training areas. The main advantage of this approach is that classification results correspond to pre-defined classes and do not have to be interpreted (Gonzalez de Tanago Menaca 2012, Gonzales and Woods 2007). Training areas were derived from LU polygons in the TerraClass dataset. In order to prevent classification bias due to high prevalence of single LU classes an even amount of training pixels per class was used as input to the classifier. The size of training samples was determined by the data availability in different implementation scenarios (see 5.3.3).

5.3.2 Tuning

Random Forests are generally considered a black box. However, a few parameters of the classifier can be tuned in order to improve model performance and computation efficiency (Srivastava 2015), or to avoid overfitting (Segal 2004, Statnikov et al. 2008). The implemented tuning procedure is conceptualized in Figure 11 and described in detail below. The definite tuning parameters used in the final implementation of the RF classifier are shown in Annex 5.

.

Mtry-parameter:

The Mtry-parameter represents the number of variables randomly sampled as candidates at each split. Throughout the forest growing procedure this value is kept constant. Representing the degree of randomization in the model this parameter is the most important tuning parameter (Geurts, Ernst and Wehenkel 2006).

The standard R implementation of the Random Forest package provides functionalities to tune the Mtry-parameter automatically. First, a RF model with an initial Mtry=10 was computed. The model was repeated, inflating and deflating the initial Mtry-parameter by a factor of 1.9. In each iteration, classification accuracy (expressed as OOB-error) was reported. The iteration stopped when the relative accuracy improvement was lower than 0.02.

Number of trees

Generally, more trees increase the model accuracy because this parameter reduces the variance of the model. However, more trees require deeper forests to balance model bias. Thus, while accuracy improvements gradually diminish, computation time increases dramatically. Consequently, an efficient implementation requires careful monitoring of the number of trees in the model. Furthermore, given very noisy data the number of trees needs to be limited to avoid overfitting (Segal 2004, Statnikov et al. 2008).

Random forests implicitly record the OOB-error each time a new tree is added to the model. A model using 501 trees was built and the accuracy for any given number of trees plotted. The plot was used to determine the optimal number of trees by visually locating the point of error convergence. This criterion was chosen to balance performance and overfitting. Attention was paid to tune the number of trees to an odd number so that voting-ties can be broken.

Input variables

With the randomization in both sample and feature selection, the trees often use uninformative features for node splitting. Thus, limiting the input features to only good predictors prevents the introduction of noise and generates more accurate trees. This is especially valid when working with high-dimensionality data (Nguyen, Huang and Nguyen 2015).

To identify (un-)important variables an alternative procedure was implemented using "conditional variable importances". The method is explained in more detail in section 5.3.5. Conditional variable importances are reported as the mean decrease in model accuracy after the permutation of the respective variables. To compute this metric, a premature classifier was built using the previously tuned Mtry and Ntree parameters. Predictors with a mean accuracy decrease below 0 (variables that impair classification performance) were removed from the dataset.

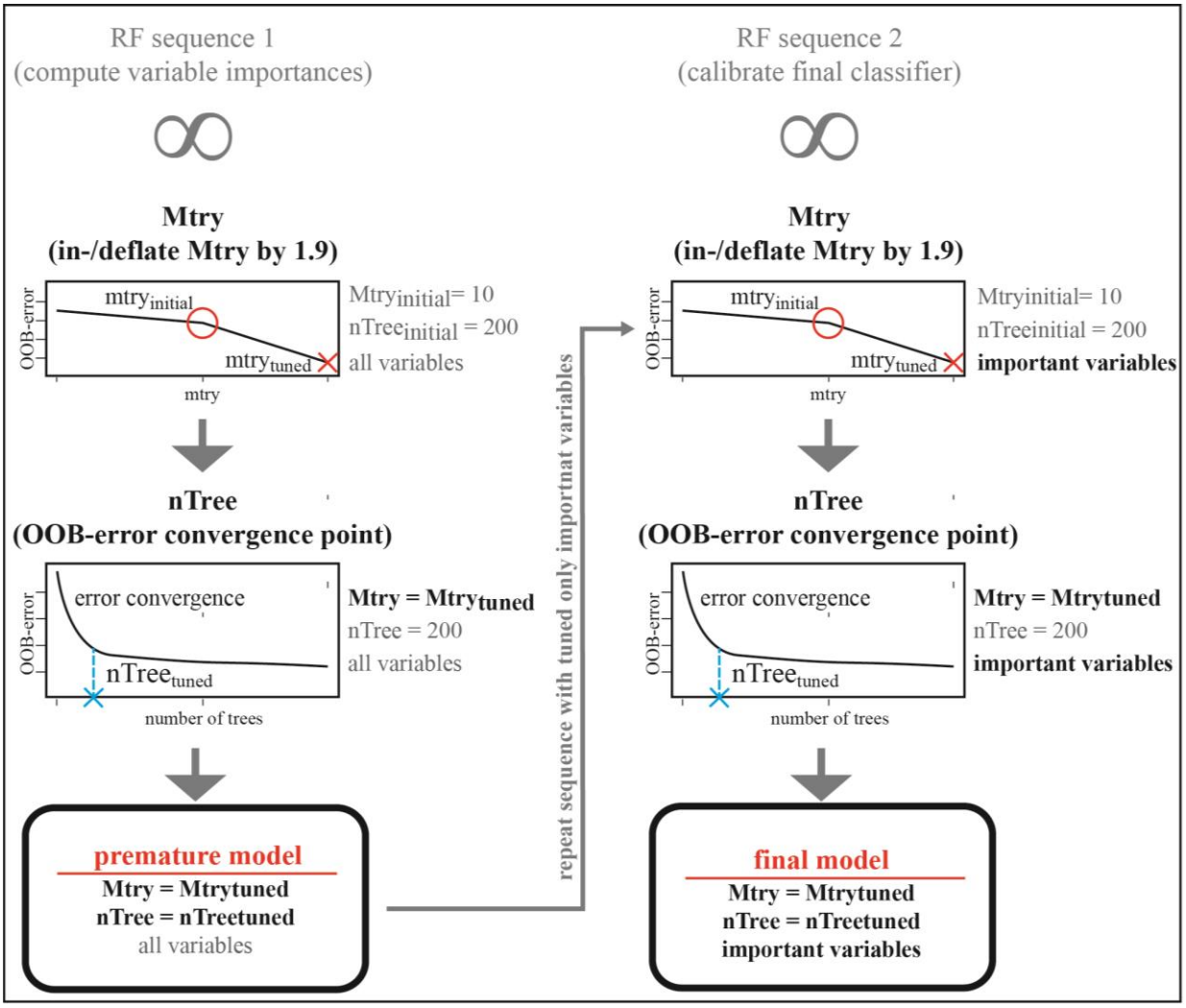


Figure 11: Implemented tuning procedure for Random Forest models

5.3.3 Analysis scenarios

To answer the respective research questions RF models were calibrated and tested in different scenarios. All models were constructed following the same approaches for training, tuning and elimination of unimportant variables.

A) General models to compare classification results for Fmask and Tmask data:

For data processed with both Fmask and Tmask two independent classification models were calibrated. Class-specific training pixels (n=10000) were sampled randomly within the study area. The calibrated RF classifiers were used to create two LU maps, using data processed with Fmask and with Tmask.

B) Stratified models based on data availability:

Additionally, it was tested how well the classifiers perform under different scenarios of data availability. Separate training and validation datasets for Fmask and Tmask data were stratified based on the number of available observations in the rain and dry season. For each count of observations, 5000 samples were used to train a unique RF model.

C) Isolated models with metrics of a single spectral index:

Variable importances from the "full model" provide valuable insights into the usability of individual variables. However, dealing with high-dimensional data metrics from different spectral indices always interact or correlate to some degree. In order to test the informative value of individual spectral indices, RF classifiers were built using 10000 training samples and the respective metrics of a single index in isolation.

D) Testing on groups with different deforestation lag:

Additionally, the robustness of the classification model was tested with respect to different time lags dating back to the last deforestation event (0-4; 4-8; 8-12; 12-16 years). In this scenario no new classifier was calibrated, mainly because of the imbalanced prevalence of individual LU types (see Table 3). Moreover, too few pixels were available to train and validate reliable classifiers in each group. Instead, the best performing model from scenario A was used to predict the LU in all groups. Depending on the outcome of scenario A, test data was chosen to match the better classifier (either Fmask or Tmask). Subsequently, LU was predicted for all pixels in each group.

Table 3: LU-specific pixel availability among groups of deforestation lag

Deforestation lag	Prevalence of LU types (number of Landsat pixels)		
	cropland	pasture	Secondary vegetation
4	1351	5093	2241
8	10315	29798	17146
12	66630	351107	154930
16	184	11638	1588

5.3.4 Validation and accuracy assessment

The accuracy of classification models in scenarios A, B and C was validated using an independent test set, 40% of the size of training data. For these scenarios, an even number of validation pixels per LU class were sampled randomly in the study area. Due to the low prevalence of some LU types in deforestation lag groups all available pixels were validated in scenario D. Predicted LU from the testsets was then compared with reference data, derived from TerraClass maps. Validation results were compiled in a confusion matrix from which classical accuracy measures were derived: overall accuracy, as well as class-specific user and producer accuracies (Congalton 1991). Using the internal OOB error of RF models for accuracy assessment was dismissed because it is biased in favor of training subsets (Mitchell 2011).

5.3.5 Variable importances

RFs implicitly evaluate the importance of individual variables for the final classification. This functionality was harnessed to assess the usability of spectral indices and statistical TS descriptors for the proposed method.

In the standard implementation of RFs (R-package: “randomForest”) the importance of individual variables is either reported as 1) Mean Gini Decrease or 2) Mean Accuracy Decrease after permuting individual variables (Breiman and Cutler 2008). Generally, the higher these values the more important an individual variable can be considered. However, these metrics are highly unreliable when using unscaled (Strobl, Boulesteix and Augustin 2007) or correlated variables (Archer and Kirnes 2008). For a robust identification of (un-)important variables we built an alternative RF model using all variables, conditional inference trees as base learners and subsampling without replacement (function cforest() in R-package: “party”). Variable importance was then measured as Mean Accuracy Decrease but

with conditional permutation of predictors. Just like in the original RF implementation this metric quantifies the decrease in model accuracy when permuting individual variables. However, conditional permutation first subsets variables into groups of un-correlated predictors, then permutes them one-by-one within these groups. Subsequently, groups with a new combination of uncorrelated variables are built and the permutation procedure repeated. The main advantage of this method is that the loss in explained model variance can be related to the permuted variable and is not disguised by other correlated variables. More information on this alternative implementation is given in (Strobl, Hothorn and Zeileis 2009).

6 Results

6.1 Effect of cloud masking on classification results

Our results indicate that Fmask and Tmask produced different cloud masks generating pixel TS with different data density. Generally, more pixels were masked when using the Tmask band (see Figure 12). That difference became especially apparent at the edge of WRS tiles. Accordingly, the required minimum number of 2 observations in rain and dry season for classification was met less frequently using Tmask (76.6 % of pixels) compared to Fmask (99%)

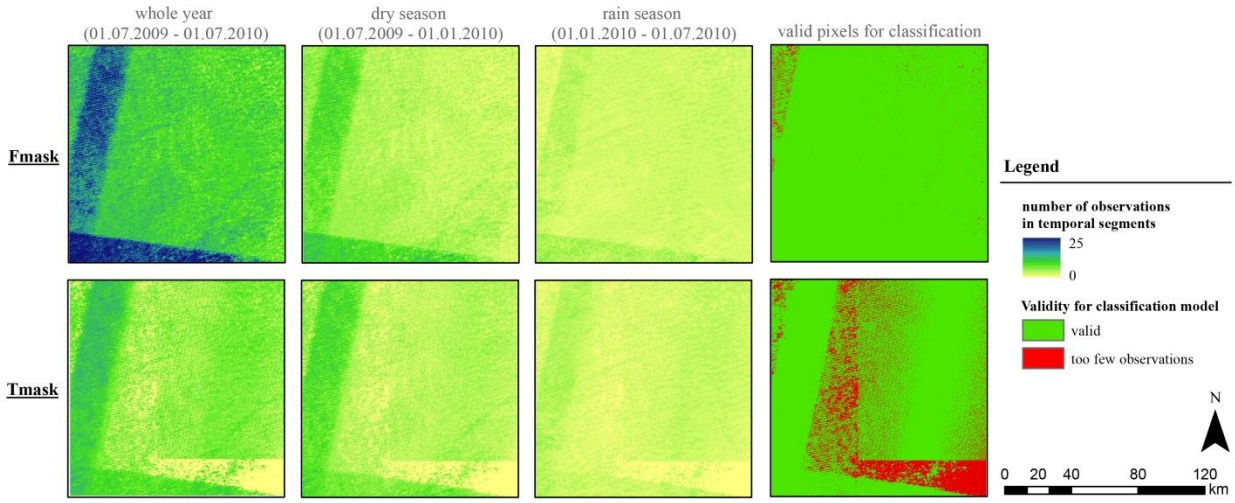


Figure 12: Available observations in pixel TS pre-processed with Fmask and Tmask

The total accuracy of the masks produced by Fmask and Tmask was not quantified directly. However, visual inspection of cloud masks suggested that Tmask tended to detect more cloud shadows than Fmask (see Figure 13). Unmasked cloud shadows had major effects on the TS of individual pixels. This was especially the case for spectral indices that were derived from the NIR band.

In some cases, Tmask did not detect the very edges of cloud shadows, while Fmask, using a 3-pixel dilation, captured them. Furthermore, Tmask identified spurious cloud shadows, which limited the number of valid observations for classification.

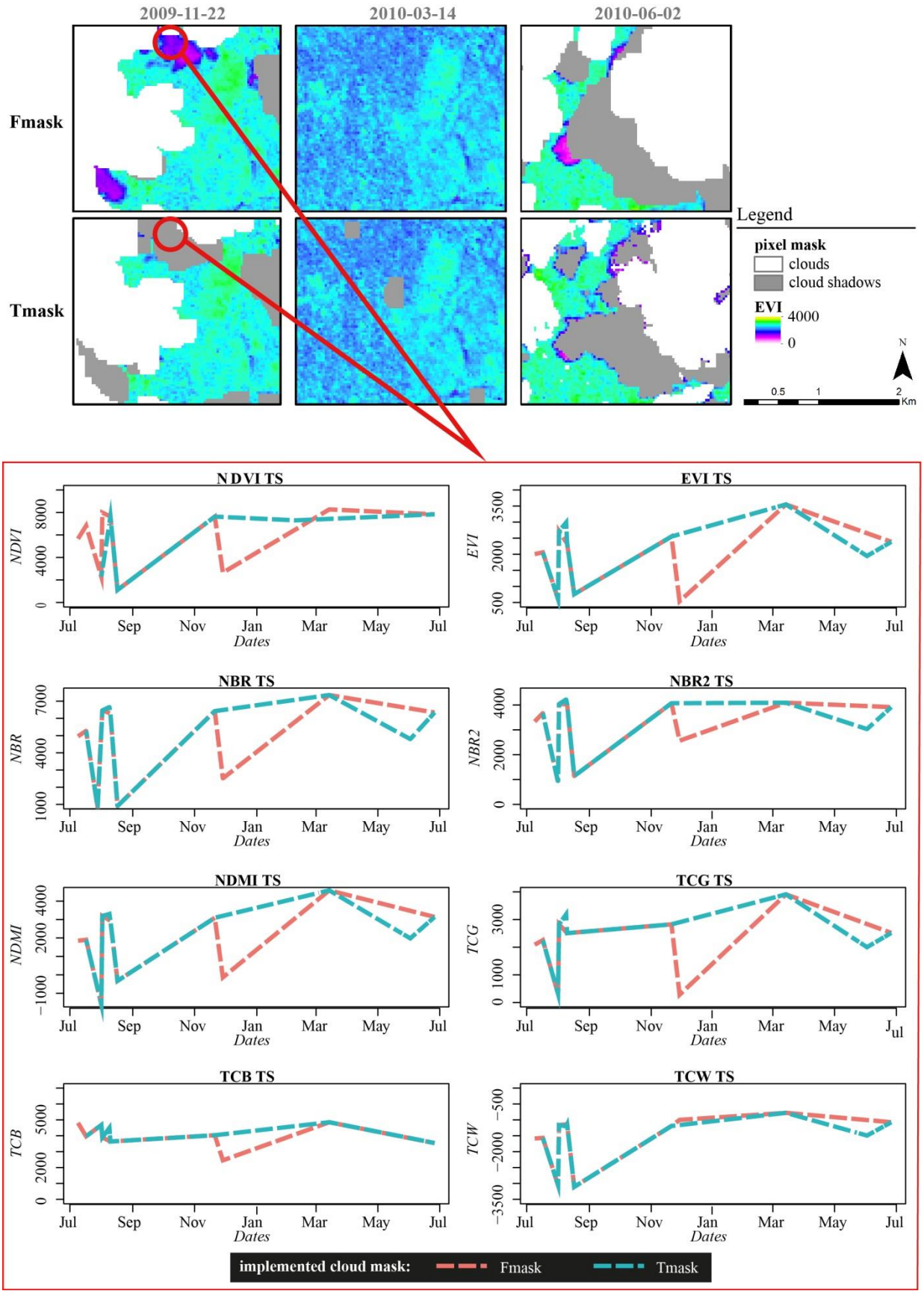


Figure 13: Detail of cloud masks in three successive images and resulting pixel TS of spectral indices

The two LU maps, generated from data pre-processed with Fmask and Tmask, had an overall accuracy well above 85% (see Table 4). While both maps can be considered very accurate, classifiers constructed with Tmask data perform slightly better than with Fmask data.

Table 4: Overall accuracy of Random Forest classifier for Fmask and Tmask dataset

Cloud masking	Accuracy (%)	Kappa (%)	Accuracy Lower (%)	Accuracy Upper (%)
Fmask	85.2	77.9	84.6	85.7
Tmask	88.8	83.1	88.2	89.3

The classification model tended to produce more accurate results with more available observations in pixel TS (see Figure 14). Particularly, in the rain season every additional observation significantly improved classification results. This trend was also visible, but less pronounced for pixels with more than 4 observations in the dry season. If less than 4 observations were available in the dry season added observations decreased the overall accuracy of the classifier. These findings held for both models generated from Fmask and Tmask data. Differences between Fmask and Tmask became especially pronounced when many data points were available in the rain-/dry season. While the model constructed with Tmask data demonstrated ongoing accuracy improvements with added observations, classification accuracy seemed to saturate around 90 % for Fmask.

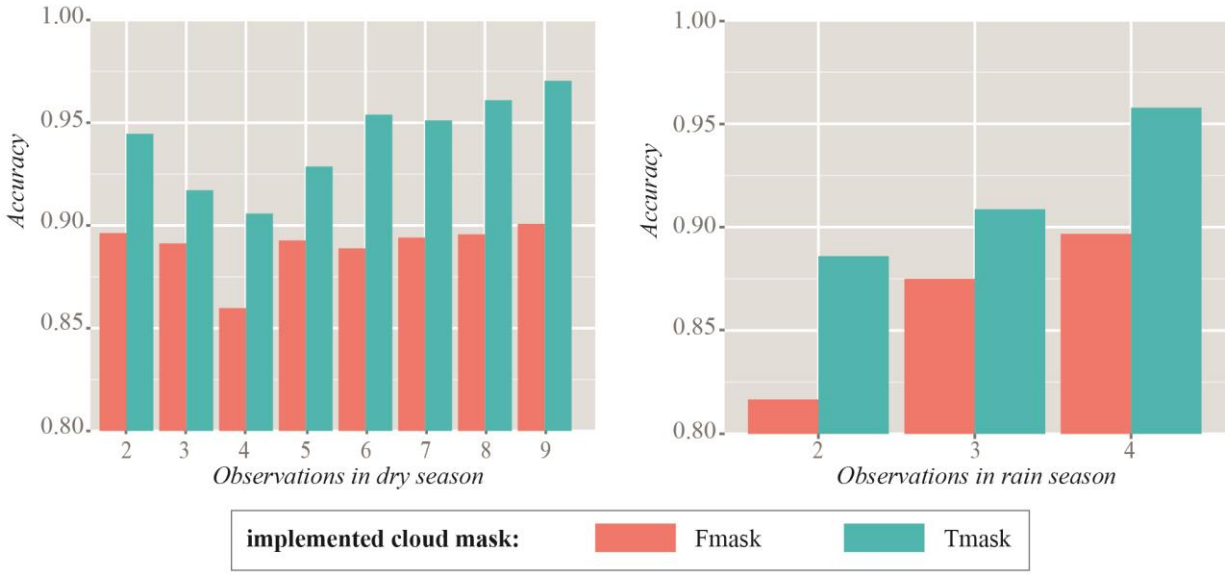


Figure 14: Overall accuracy of RF classifier in different scenarios of data availability

Figure 15 exemplifies the influence of different cloud masking procedures on the classification result. Note that the developed classification algorithm requires at least two observations in rain- and dry season. While Tmask masked more observations preventing the classification of many areas, Fmask frequently produced wrong classification results for the same pixels. Differences between both maps are especially pronounced along stripes originating from Landsat 7 images without functioning SLC. This feature is particularly visible full maps, attached in Annex 6. Full maps also reveal that LU maps generated from Fmask data provide a much higher coverage, especially along the eastern edge of the north-western overlap zone.

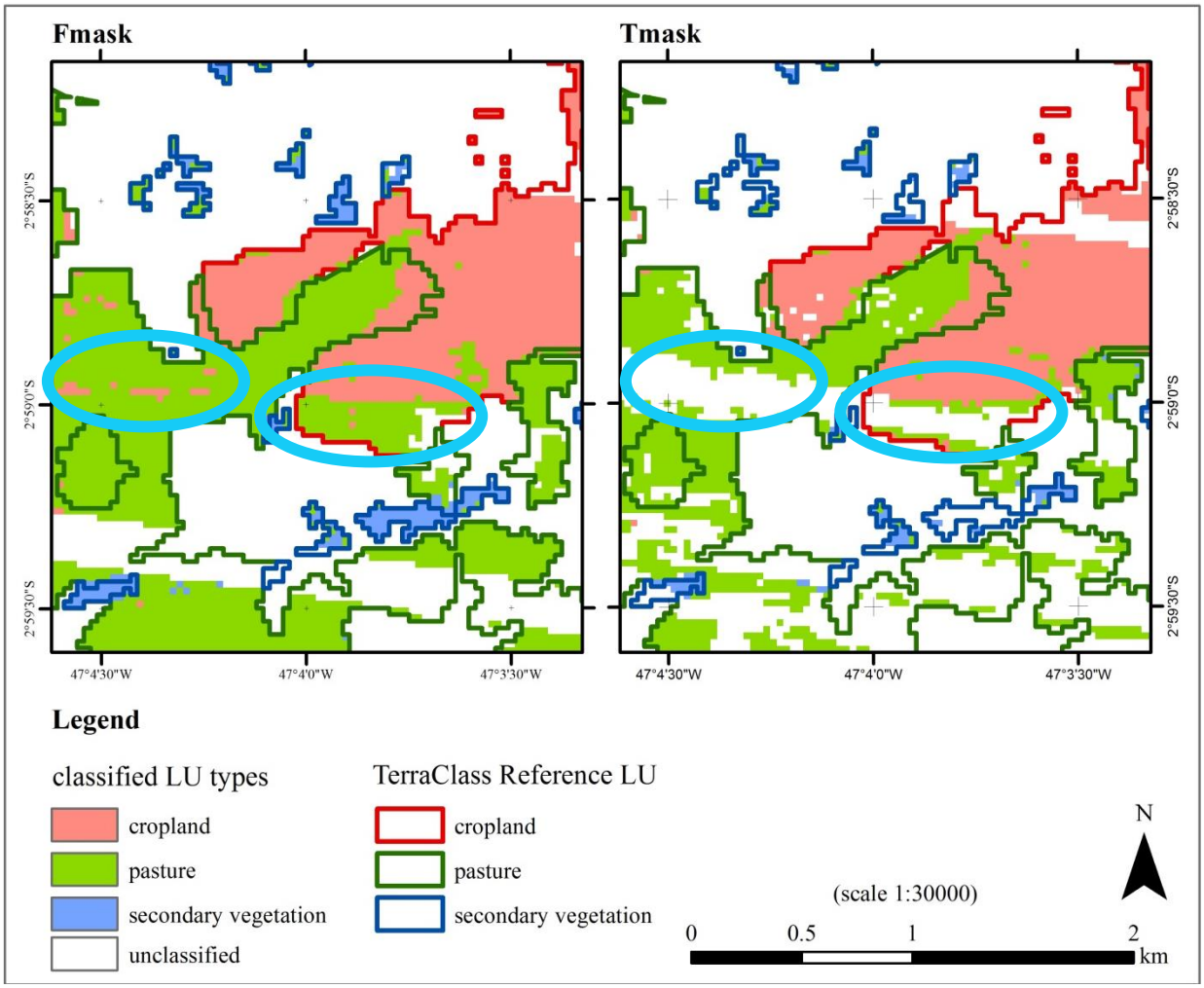


Figure 15: Detail of LU maps generated from Fmask and Tmask

Classification performance among individual classes is shown in Table 5. While secondary vegetation demonstrated highest user accuracy using Fmask data (88.2 %), cropland shows the highest user accuracy for Tmask data (91.8 %). Highest producer accuracies were achieved for cropland in both Fmask (90.7 %) and Tmask (93.1) classifiers. The classifiers

differed mainly in their ability to distinguish pastures from cropland which explains the differences in producer accuracy (Fmask = 76.6%, Tmask = 83.7%).

Table 5:Confusion matrix of predicted pixels and reference classes

Fmask				
Reference / **Predicted**	Cropland	Pasture	Secondary vegetation	Producer Accuracy (%)
Cropland	3272	306	31	90.7
Pasture	608	3427	440	76.6
Secondary vegetation	120	267	3529	90.1
User Accuracy (%)	81.8	85.7	88.2	
Tmask				
Reference / **Predicted**	Cropland	Pasture	Secondary vegetation	Producer Accuracy (%)
Cropland	3670	241	30	93.1
Pasture	219	3467	455	83.7
Secondary vegetation	111	292	3515	89.7
User Accuracy (%)	91.8	86.7	87.9	

6.2 Importance of temporal metrics and spectral indices

Among the computed temporal metrics of pixel TS, the mean value of NBR in the dry season was identified as the most important predictor for classification. Generally, variables expressing annual or seasonal mean and minimum values attained a higher importance than other variables. Annual and seasonal metrics seem equally important. Among the important predictors, summary metrics computed from NDMI TS were most prominent. Variables computed from NDVI, EVI and TCB TS did generally not achieve a high degree of importance. The majority of the least important metrics express annual and seasonal signal variability (standard deviation and amplitude) of spectral indices. The use of these metrics partly introduced noise into the classification model, which was illustrated by the higher model accuracy when omitting the respective variables. This was particularly the case for

metrics expressing NDVI variability. The importances of best and worst predictors are displayed in Figure 16.

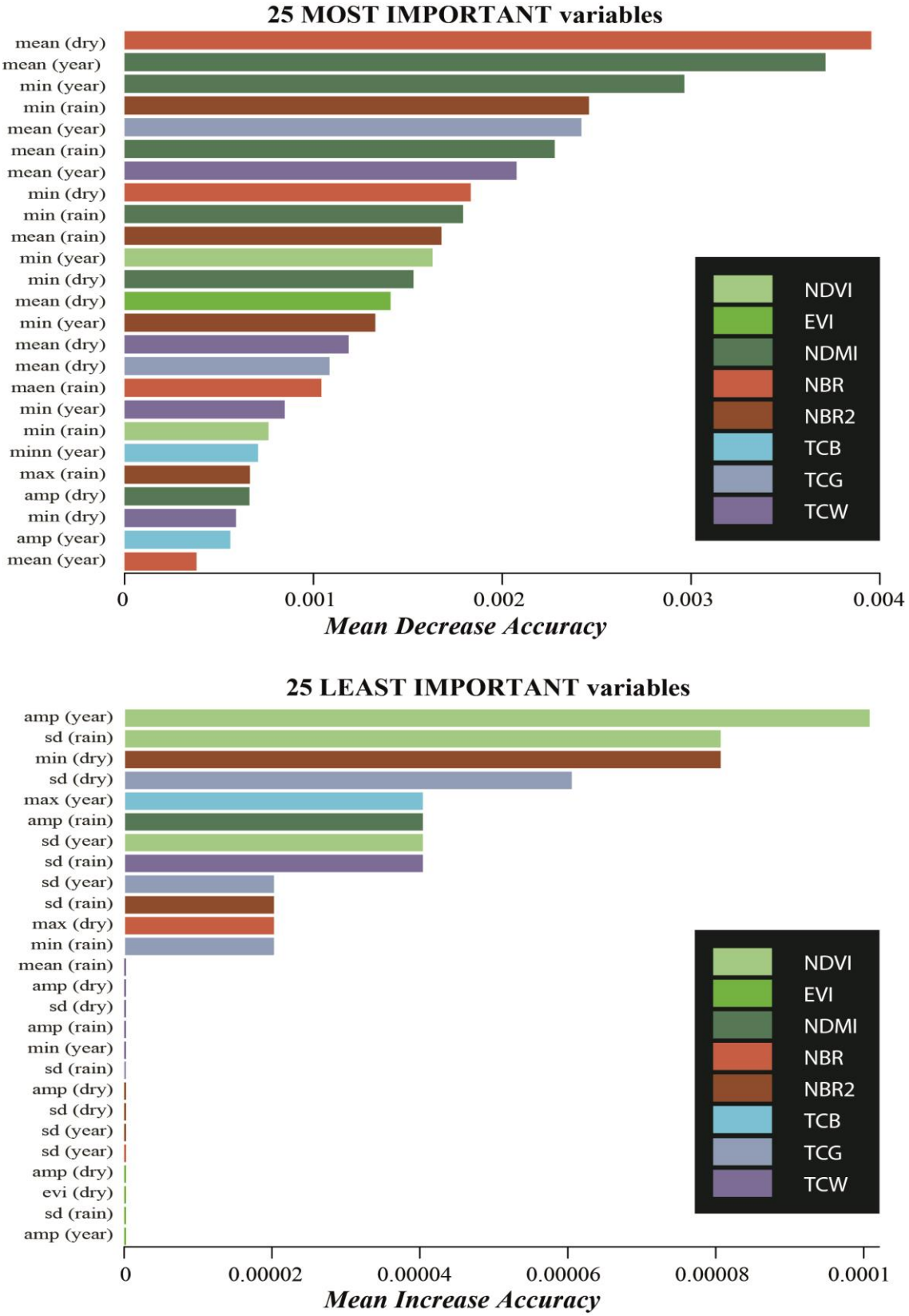


Figure 16: Variable importances in RF classifier

The accuracy of isolated models, using only metrics from one single spectral index, confirmed the variable importances reported from the general model (using all the different spectral indices tested). Highest overall classification accuracies were achieved when using metrics from NBR (83.7%) and NBR2 TS (82.9%). Also EVI, NDMI and NDVI achieved overall accuracies well above 80%. Among spectral indices TCB and TCG performed worst, demonstrating overall accuracies of 77.5% and 74.7% respectively (see Figure 17). Metrics derived from TCW had more predictive power (79.8%) than the other analysed TCc.

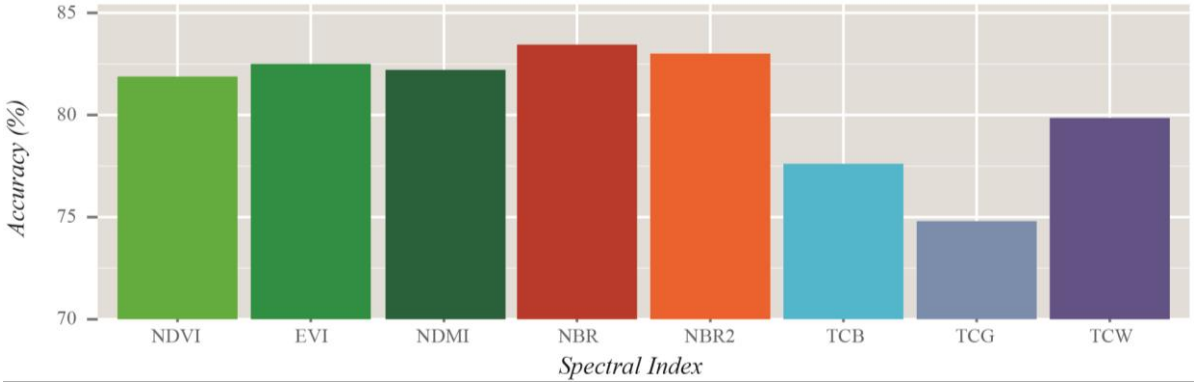


Figure 17: Overall accuracy of isolated models using metrics from only one spectral index

For all individual LU types, metrics from derived from NBR TS rank among the most explanatory variables (see Figure 18). Furthermore, TCW, NDMI, and NBR2 metrics were evenly good predictors for secondary vegetation. Using variables derived from NDVI TS produced maps with the lowest omission error for cropland. TCB and TCG had the least predictive power for all examined LU types.

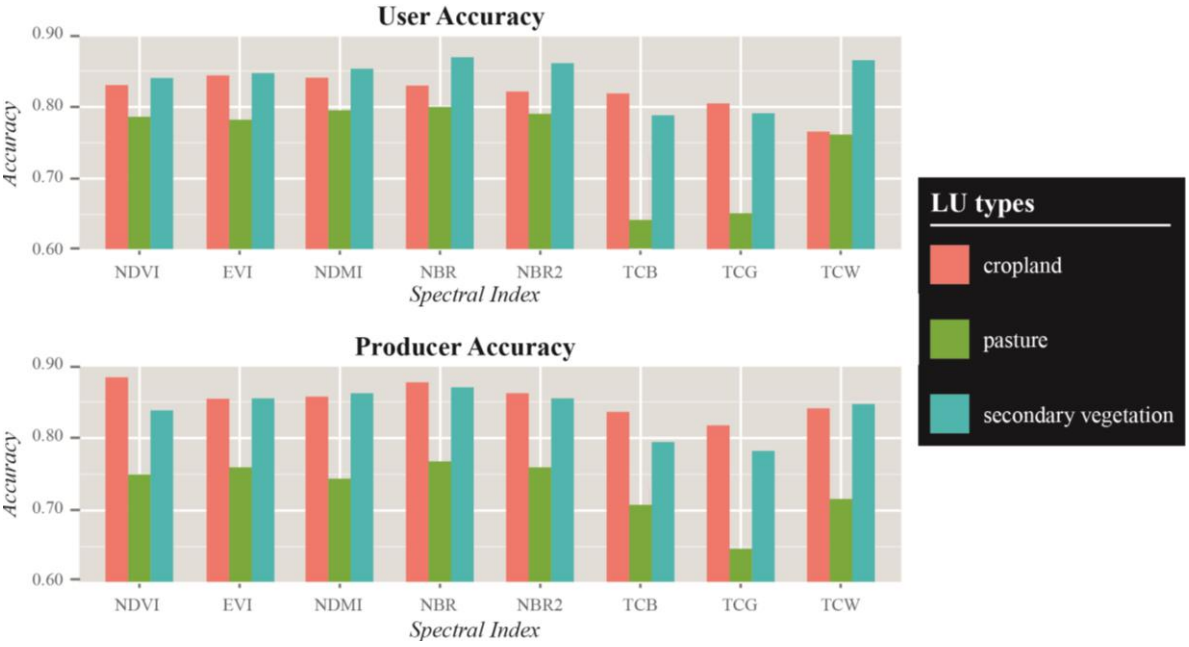


Figure 18: Classification accuracy of RF classifier with isolated spectral indices

6.3 LU classification with different post-deforestation lag

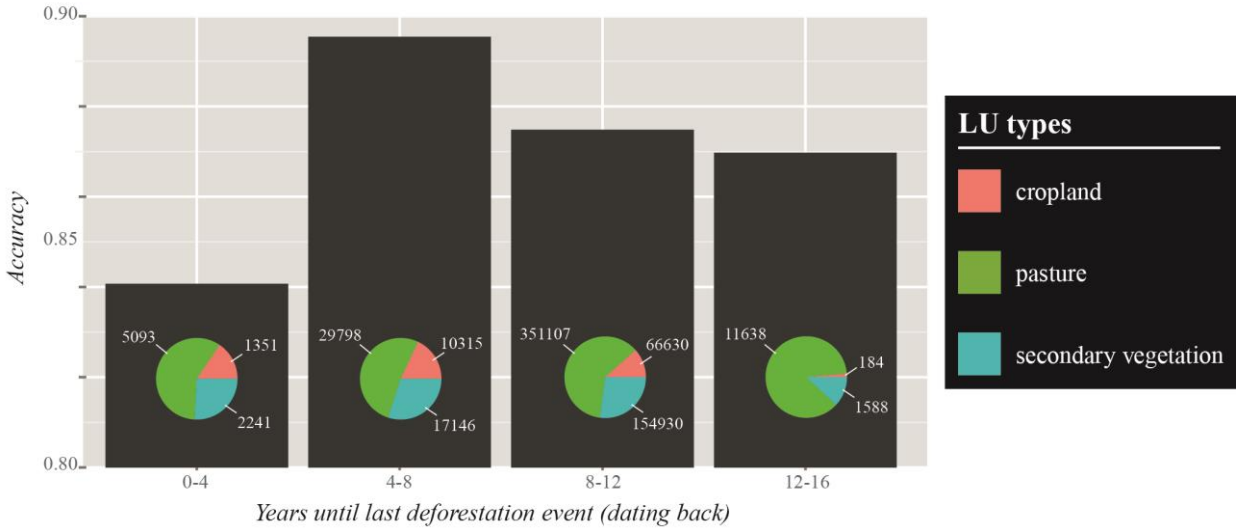


Figure 19: Classification accuracy and prevalence of LU classes for pixels grouped based on time lag to last deforestation event

For the classification of LU among different deforestation lag groups the calibrated RF model employing Tmask data was used.

The classifier performed worst for areas with recent deforestation (84.1%; 0-4 years) (see Figure 19). Pixels in the following group (4-8 years) demonstrated the highest overall classification accuracy (89.6%). Classification accuracy among groups with longer deforestation lags was slightly lower (87.5%; 8-12 years and 87%; 12-16 years).

Accuracy measures for individual LU types are displayed in Figure 20. Highest producer accuracies were achieved for pastures (93.4 – 99.3%), which corresponds with the strong prevalence of pastures in groups (see Table 3). Similarly, lowest producer accuracies for cropland and secondary vegetation were reported for the 12-16 year lag group, in which both classes were least abundant. Among all classes, best user accuracies were obtained for cropland, particularly evident in the group with the longest deforestation lag (97.8%). Similarly, secondary vegetation demonstrated the highest user accuracy in that group (95,2%). Pastures showed lowest user accuracies in the group with the most recent deforestation events (78.9%). Generally, the user accuracy of pastures in lag groups strongly resembled group-specific overall accuracies displayed in Figure 19.

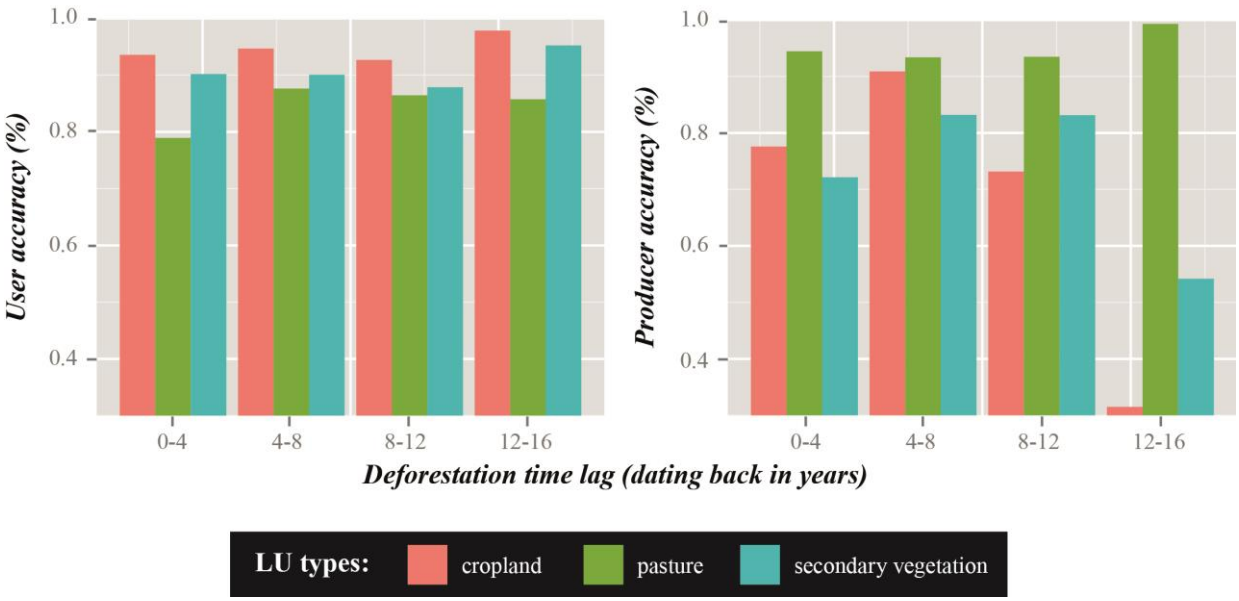


Figure 20: LU type-specific accuracies of classifiers built with data representing different deforestation time lags

7 Discussion and recommendations

7.1 Effect of cloud masking on classification results

In the framework of this study the accuracy of different cloud masking approaches was not quantified, as this would ideally require manual masks for every image (Zhu and Woodcock 2012). However, post-deforestation LU maps display that Tmask-derived cloud masks generated TS yielding slightly better classification results than using TS masked with Fmask. Since both maps were generated in an identical fashion, performance discrepancies must originate from different cloud masks produced.

Figure 21 displays the amount of available observations for pixels that were falsely classified with Fmask data and correctly classified with Tmask data. Falsely classified TS tend to feature more observations which suggests that less clouds or cloud shadows were masked. It is therefore assumed that the resulting outlying observations were propagated into computed temporal metrics. This hypothesis is partly supported by the findings displayed in Figure 14. The accuracy of Fmask classifiers saturated at 90 % regardless of the number of added observations. In contrast, Tmask classifiers steadily improved with every added observation to very high accuracies of 97 %. Following the same methodology for both classifiers, these findings can be explained by the comparatively high presence of unmasked outliers in Fmask.

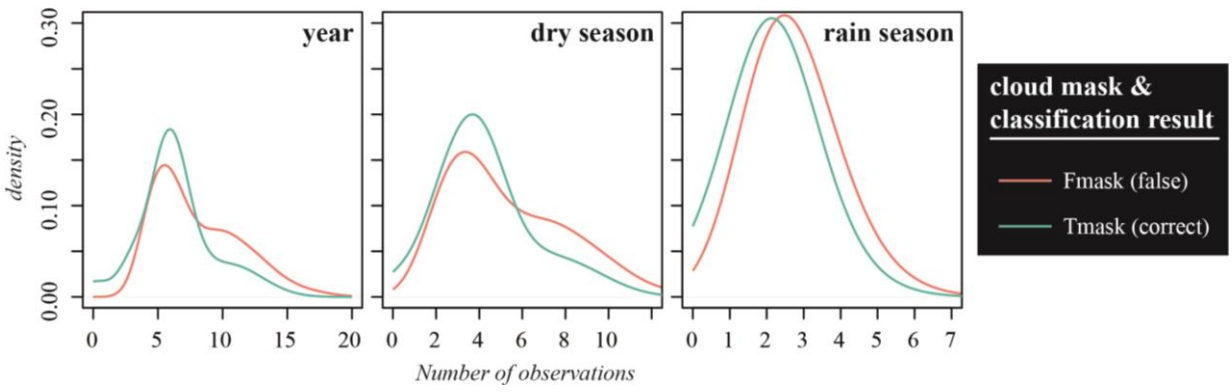


Figure 21: Probability distribution of unmasked observations in pixel TS processed with Fmask- and Tmask- derived cloud masks (n=24013; kernel bandwidth=1)

The classifiers mainly differed in their ability to classify pastures correctly. The Fmask classifier particularly confused pastures with cropland. Potentially, low minima in seasonal or annual composites representing unmasked cloud shadows were falsely interpreted as bare soil, one of the key distinctive features between pastures and cropland (Prishchepov et al. 2012).

Despite the better performance of Tmask, the overall accuracy of both LU maps is not conspicuously different. Originally developed for dense TS, Tmask requires at least 15 "clear" observations to estimate a robust TS model for cloud detection (Zhu and Woodcock 2014a). In the study area this was only the case for a minor amount of 0.3 % of all pixels. For all pixels with less than 15 observations Tmask employs a backup algorithm, which subtracts the TOA band 2 reflectance of a pixel from the median value of the TS at that location. Pixels are identified as "clear" if that difference is below a threshold of 0.04. Aiming to derive this median from TS that are uncontaminated by clouds and cloud shadows, Tmask utilizes the Fmask band for initial cloud screening. However, undetected clouds in the Fmask band artificially alter the median of TOA band 2 TS. This alteration is magnified in low-density TS. Therefore, additional unmasked clouds or cloud shadows in the TS may not deviate enough from the median and are thus propagated to Tmask outputs. Moreover, the Tmask backup algorithm does not distinguish between LC change and clouds. Potentially, the superior performance of Tmask becomes more pronounced if more cloud-free observations can be provided in pixel-TS.

From visual interpretation of individual images, we could identify some pixel values deviating strongly from spatially and temporally adjacent features (see Figure 13). These were mainly situated around the edges of clouds or cloud shadows. (Zhu and Woodcock 2014a) point out that Tmask especially outperforms Fmask regarding the detection of cloud shadows. While Fmask uses a complicated object-based geometry matching approach to detect shadows around clouds, Tmask uses location specific thresholds derived from pixel TS of TOA reflectance bands 4 and 5 (ibid.). Due to the darkening effect of cloud shadows on reflectance bands in the visible and NIR region, TS of most spectral indices were subject to the introduction of low outliers (Zhu and Woodcock 2012).

For future applications, we suggest to employ Tmask as it is location specific and detects more clouds/cloud shadows than Fmask. Tmask particularly benefits from the provision of an initial cloud mask and additional multi-temporal information. At present both can be readily employed using Landsat data. The implementation of Tmask can be further improved, especially in regard of providing a minimum of 15 clear observations. As clear observations do not have to be within a single year (Zhu and Woodcock 2014a), this requirement can easily be met by extending Tmask data inputs +/- one year in tropical areas with low data density. Furthermore, Tmask was implemented as a pixel-based approach. While the outputs look

generally accurate, the very edges of cloud shadows were not accurately masked. In order to avoid the introduction of these respective observations, one might consider dilation of cloud shadows by a fixed buffer. Using only three spectral bands (green, NIR, SWIR1), Tmask-like algorithms are applicable to various other optical RS systems. Given the comparatively shorter revisit time and more spectral bands for the detection of various cloud types (Gascon, Martimort and Spoto 2009), applying Tmask-like algorithms to multi-temporal Sentinel 2 data might produce more accurate cloud masks.

7.2 Importance of temporal metrics and spectral indices

Mean and minimum values of annual and seasonal TS segments, especially of NBR and NDMI, were identified as best predictors in the classification model. Harnessing the difference between NIR and SWIR reflectance, both indices are sensitive to the moisture content of vegetation (McDonald, Gemmell and Lewis 1998). Results from the isolated models confirm that indices sensitive to canopy moisture (NDMI, NBR and TCW) were particularly good predictors of secondary vegetation. Exemplary, these indices find frequent application in forest monitoring due to their usefulness to distinguish vegetation classes at higher biomass levels (Fiorella and Ripple 1993).

NBR stands out as it additionally characterized pastures best. Incorporating various vegetation types from short-grass to early stages of scrub encroachment pastures depict a high range of spectral-temporal characteristics. NBR was originally designed to map burned areas and fire severity (Key and Benson 2005). The index utilizes NIR and SWIR II bands. In terms of surface reflectance, vegetation and bare soil differ most in these bands (see Figure 22). Therefore, it is expected that NBR could specifically distinguish sparse vegetation in degraded or intensive pastures from bare soil, resulting in a lower confusion with croplands.

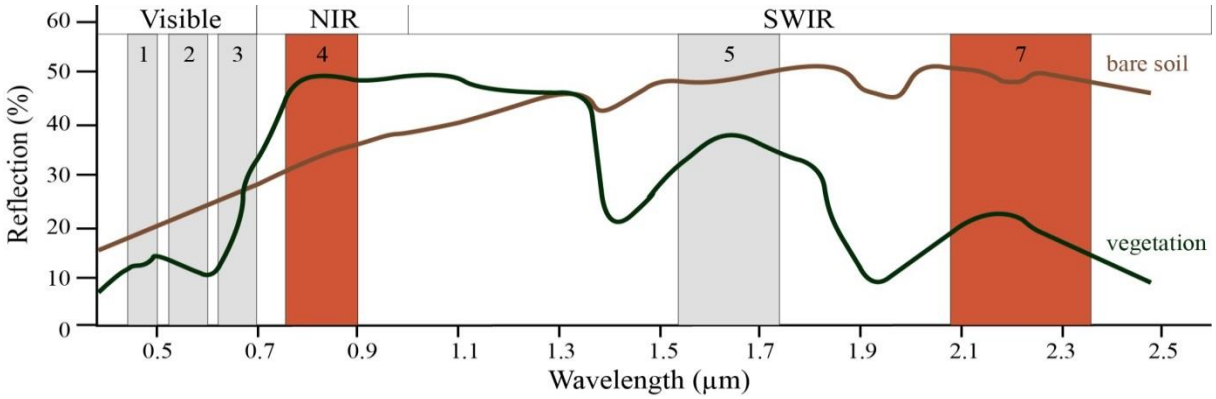


Figure 22: Spectral profiles of bare soil and vegetation (grey boxes = Landsat bands; red boxes = bands used by NBR) (source: Siegmund, 2005)

Conspicuously, statistical metrics expressing standard deviation and amplitude of temporal segments performed poor. This holds especially for NDVI, NBR2 and TCW. As minimum values of these spectral indices were found to be good predictors cloud contamination can be discarded as an explanation for these results. Similarly, maximum values of various spectral indices are not impaired by clouds, cloud shadows and other sources of noise (Holben 1986). A realistic expression of the real annual or seasonal signal variation relies on the detection of positive and negative RS signal peaks in the TS (Prishchepov et al. 2012). Employing low density TS, the probability of capturing these peaks is relatively small.

In this study, an extensive range of spectral indices was compared regarding their informative value for post-deforestation LU classification. A number of indices generated TS metrics which, even in isolation, produced classification results well above 80 %. Due to their different responses to biophysical characteristics of the earth surface, we suggest to combine a small set of spectral indices for LULC classification. The temporal profile of indices that are sensitive to photosynthetic activity (NDVI or EVI) provides an accurate representation of vegetation phenology (Huete et al. 2002). As these indices tend to saturate at higher biomass levels, they should be complemented by spectral indices responsive to canopy moisture (NBR, NDMI or TCW). The latter are particularly relevant for the detection of secondary vegetation or forest regrowth (DeVries et al. 2015). In scenarios of low data density we suggest to avoid the use temporal metrics expressing signal variability as TS are likely to obscure relevant signal peaks. Potentially, the data availability in tropical regions can be enhanced by using Sentinel 2 data. Sentinel 2 aims at the continuity of Landsat data which permits the use of comparable spectral indices with additional red channels for assessing vegetation characteristics. Potentially, the higher data density and more targeted bands of Sentinel 2 further improve the performance of the proposed classification methodology.

7.3 LU classification with different post-deforestation lag

When looking at individual LU types, both cropland and secondary vegetation seemed to show more unique spectral responses in the group with the longest deforestation lag, illustrated by the comparably low omission errors. Pastures were most frequently misclassified in the group with the shortest deforestation lag. That indicates a lower spectral separability of pastures to other LU types in immediate post-deforestation scenarios.

Despite these findings, results do not directly suggest a linear correlation between classification accuracy and the time since the last deforestation event. In fact, overall and class-specific producer accuracies are strongly biased by the dominating prevalence of pastures among all lag groups. Potentially, the reliability of class-specific user accuracies is also limited, as partly few test samples were available, especially for croplands in the 12-16 year lag group (see Table 3). By extending the research area, enough pixels could be gained to use sufficient and equal sample sizes among target classes. Another limitation of this analysis step was the high range of years within each group of deforestation lag. In the group with most recent deforestation events (0-4 years) it could not be distinguished between areas with very recent deforestation and others. In fact, all pixels with very recent deforestation (1-2 years) were excluded from the analysis as only pixels with stable LU between the years 2008-2010 were examined. By using a more targeted reference dataset, enough pixels experiencing very recent deforestation could be gained, so that samples can be grouped by single years.

Morton, 2006 #127 found that cropland and pastures can be detected well within one year after deforestation events using dense annual RS TS. Other research suggests that especially the detection of secondary vegetation may require longer monitoring windows after deforestation events (DeVries et al. 2015, Kennedy et al. 2010). That is in part due to varying characteristics of early succession stages and the high spatial heterogeneity in immediate post-deforestation landscapes (Moran and Brondizio 1998). Many examples in RS literature suggest data-driven approaches to capture slowly evolving and highly location-specific processes of anew forest succession (DeVries et al. 2015, Kennedy et al. 2010, Stueve et al. 2011). We propose to test the performance of the various TS methods for the detection of post-deforestation LU in scenarios with different data densities and time lags dating back to the last deforestation event.

7.4 Classification methodology

The developed classification model generated very accurate LU maps from data processed with both Fmask and Tmask. We demonstrated that seasonally targeted TS segments of various spectral indices produce metrics that are excellent predictors of post-deforestation LU, even in tropical areas suffering from very limited data availability.

In general, the comparability to other LU maps is restricted, as thematic legends and mapping standards of other studies generally differ (Herold et al. 2008). Using a very similar approach

to map agricultural land change (classes cropland, pastures and forest), (Griffiths et al. 2013) demonstrated the usability of the method for change detection in central Europe. They produced a tri-temporal change map at an accuracy of 90%, which suggests that the accuracy of individual maps must be at least as high as our product. The approach differs from ours as metrics were computed for 3 seasonal segments (spring, summer, fall). Therefore, the spectral response of LU types is temporally more specific. This however was not possible in our case as much less RS data is available for the BLA.

Other approaches circumvent constrained data availability by using lower resolution imagery. (Carreiras et al. 2006) used 1 km SPOT-4 VEGETATION data to map LU in Mato Grosso. Their LU maps display accuracies below our results (overall = 81-88%). However, more LU classes were incorporated in their product resulting in a higher probability of confusion. (Besnard 2014) contemplated the usability of different LC products for the analysis of deforestation drivers in South America, using 4 LC classes (see Table 6). Also compared to his findings our post-deforestation LU maps demonstrate higher accuracies. Performance differences may likely be pronounced in heterogeneous landscapes, which can be mapped more accurately with high-resolution imagery (Herold et al. 2008). Again, comparability is limited because these maps target global LU while our classification model was calibrated locally.

Table 6: Accuracy of Global LC products in South America (Besnard 2014)

LC product	Resolution (m)	Overall accuracy (%)	
		(2000)	(2005)
MODIS Land Cover	~ 463	72.9	67.1
GLC-SHARE	~ 1000	72.6	88.4
CCI-LC	300	90.2	88.4

The use of high-resolution post-deforestation LU data is especially important in the northern Amazon where swidden cultivation and agroforestry determine a highly complex land use mosaic (Moran et al. 2000). Potentially, the approach investigated in this thesis could be used to produce annual post-deforestation LU maps for the whole BLA. Certainly, TerraClass only provides training data for 2008 and 2010. Currently the Farmland Environmental Registry (CAR) is being developed in the area, which encourages land owners to report the use of their

property. RS data could be complemented with this information for annual calibration of the classification model.

Due to similar spectral coverage Sentinel 2 data is an interesting data source to improve the proposed method. The satellite outperforms Landsat regarding both temporal and spatial resolution. The shorter revisit time will provide more dense TS, which has several advantages:

- More accurate multi-temporal cloud masking procedure
- More accurate outlier detection using TS approaches
- Potentially enough data for three seasonal segments
- Less exclusion of pixels with too sparse data density in segments

Alternatively, Sentinel 2 data could be used to implement other classification approaches demanding high density TS. Recently, several methods combining continuous change detection and LULC classification were proposed (Maus 2014, Zhu and Woodcock 2014b). Given the lack of continuous LU data in the tropics, a comprehensive evaluation of different classifiers and data representations is highly recommended.

8 Conclusions

In this thesis, it was shown that post-deforestation LU can be classified very accurately using Landsat TS with low temporal density. Given the current unavailability of high-resolution LU maps in the Amazon, our research demonstrates that such products can be produced at low cost, both in terms of data acquisition and manual effort. If reproduced annually, these maps could provide valuable information for many applications and interest groups, among others:

- Assessment deforestation drivers **(policy making)**
- tracking post-deforestation LC trajectories and the associated fate of Carbon stocks (Ramankutty et al. 2007) **(climate studies)**
- Assessment of recovery potential of previous forests (Guariguata and Ostertag 2001) **(forestry and LU planning)**

Our research also provides valuable methodological insights regarding the mapping of post-deforestation LU in the tropics. Generating cloud masks with the Tmask algorithm enhanced classification results, compared with using the Fmask band provided with public Landsat data. For future applications we recommend to exploit a combination of spectral indices that are either sensitive to canopy moisture (NBR, NBR2, NDMI) or photosynthetic potential (EVI, NDVI) of vegetation. Given a low data density, temporal metrics expressing annual or seasonal mean and minimum (especially NBR) should be favoured over metrics expressing the variability of the RS signal (especially NDVI). In the framework of this study it could not be shown that classification accuracy generally corresponds with the time lag to the last deforestation event. Although the separability of pastures was limited in immediate post-deforestation scenarios, more targeted sampling approaches are needed to support these findings.

The classification approach in this study was mainly motivated by the limited availability of Landsat data. Alternatively, a number of interesting methods were developed for the classification of dense TS (e.g. (Maus 2014, Zhu and Woodcock 2014b). Sentinel 2 is a promising data source, both in terms of temporal data density and spatial resolution. Using Sentinel 2, potentially more advanced approaches can be implemented for cloud masking, outlier detection and LU classification to further enhance the accuracy and resolution of the produced maps.

9 References

Aguiar, A. P. D., G. Câmara & M. I. S. Escada (2007) Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity (Ecological Modelling). 209**,** 169-188.

Almeida, C. A., P. T.F., A. M. Barbosa, M. de Abreu, F. de Lucia Lobo, M. Silva, A. R. Gomes, L. Sadeck, L. de Medeiros, M. F. Neves, L. da Silva & P. Tamasauskas. 2009. METODOLOGIA PARA MAPEAMENTO DE

VEGETAC~AO SECUNDARIA NA AMAZONIA LEGAL. ed. G. J. F. Banon, 32. Sao Jose dos Campos, Brazil: INPE.

Araújo, R. & P. Léna (2011) Desenvolvimento Sustentável e Sociedade na Amazônia.

Archer, K. J. & R. V. Kirnes (2008) Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis,* 52**,** 2249-2260.

Azevedo-Ramos, C. (2008) Sustainable development and challenging deforestation in the Brazilian Amazon: the good, the bad and the ugly. *Unasylva 230,* 59**,** 12-16.

Baidya Roy, S. & R. Avissar (2002) Impact of land use/land cover change on regional hydrometeorology in Amazonia (Journal of Geophysical Research: Atmospheres). 107**,** 4-1-4-12.

Balazs, C. 2001. Deforestation and Human Land-Use Dynamics in the Brazilian Amazon: A Case Study Using Remote Sensing and GIS. In *Center for Environmental Studies*, 74. Berkeley: Brown University.

Besnard, S. 2014. Pan-tropical analysis of deforestation drivers using global land cover products and the role of deforestation drivers in global gross carbon emissions for 1990-2000-2005 time period. In *Laboratory of Geo-information Science and Remote Sensing*, 61. Wageningen University.

Birdsey, R., G. Angeles-Perez, W. A. Kurz, A. Lister, M. Olguin, Y. D. Pan, C. Wayson, B. Wilson & K. Johnson (2013) Approaches to monitoring changes in carbon stocks for REDD. *Carbon Management,* 4**,** 519-537.

Breiman, L. (2001) Random forests. *Machine Learning,* 45**,** 5-32.

Breiman, L. & A. Cutler. 2008. Random forests – Classification manual.

Câmara, G., D. de Morisson Valeriano & J. Vianei Soares. 2006. Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal INPE.

Carreiras, J., J. Cardoso Pereira, M. Lameiras Campagnolo & Y. E. Shimabukuro (2006) A land cover map for the Brazilian Legal Amazon using SPOT-4 VEGETATION data and machine learning algorithms. *Photogrammetric Engineering & Remote Sensing,* 72**,** 97–910.

Congalton, R. G. (1991) A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment,* 37**,** 35-46.

Conrad, C., R. R. Colditz, S. Dech, D. Klein & P. L. G. Vlek (2011) Temporal segmentation of MODIS time series for improving crop classification in Central Asian irrigation systems. *International Journal of Remote Sensing,* 32**,** 8763-8778.

Crist, E. P. (1985) A Tm Tasseled Cap Equivalent Transformation for Reflectance Factor Data. *Remote Sensing of Environment,* 17**,** 301-306.

Da Veiga, J. B., J.-F. Tourrand, R. Poccard Chapuis & M.-G. Piketty. 2003. Cattle ranching in the Amazon rainforest. In *Congrès forestier mondial*, 93-93. CAN: 12.

Davalos, L. M., J. S. Holmes, N. Rodriguez & D. Armenteras (2014) Demand for beef is unrelated to pasture expansion in northwestern Amazonia. *Biological Conservation,* 170**,** 64-73.

DeVries, B., M. Decuyper, J. Verbesselt, A. Zeileis, M. Herold & S. Joseph. 2015. *Tracking disturbance-regrowth dynamics in tropical forests using structural change detection and Landsat time series*.

Dirzo, R. & P. H. Raven. 2003. Global state of biodiversity and loss. In *Annual Review of Environment and Resources*, 137-167.

Espaciais, I. N. d. P. 2002. Deforestation estimates in the Brazilian Amazon. São José dos Campos: INPE.

FAO (2006) FAOSTAT [electronic resource].

Fiorella, M. & W. J. Ripple (1993) Analysis of Conifer Forest Regeneration Using Landsat Thematic Mapper Data. *Photogrammetric Engineering and Remote Sensing,* 59**,** 1383-1388.

Frazier, R. J., N. C. Coops, M. A. Wulder & R. Kennedy (2014) Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. *ISPRS Journal of Photogrammetry and Remote Sensing,* 92**,** 137-146.

Gascon, F., P. Martimort & F. Spoto. 2009. Sentinel-2 optical high resolution mission for GMES land operational services. 747404-747404-5.

Geurts, P., D. Ernst & L. Wehenkel (2006) Extremely randomized trees. *Machine Learning,* 63**,** 3-42.

Gonzales, R. C. & R. E. Woods. 2007. *Digital Image Processing*. New Jersey: Prentice Hall.

Gonzalez de Tanago Menaca, J. 2012. Towards an innovative remote sensing tool for tropical change monitoring by integrating spatial segments and time series analysis. In *Laboratory of Geo-information Science and Remote Sensing*, 104. Wageningen: Wageningen University.

Goodwin, N. R., L. J. Collett, R. J. Denham, N. Flood & D. Tindall (2013) Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM plus time series. *Remote Sensing of Environment,* 134**,** 50-65.

Griffiths, P., D. Muller, T. Kuemmerle & P. Hostert (2013) Agricultural land change in the Carpathian ecoregion after the breakdown of socialism and expansion of the European Union. *Environmental Research Letters,* 8.

Guariguata, M. R. & R. Ostertag (2001) Neotropical secondary forest succession: Changes in structural and functional characteristics (Forest Ecology and Management). 148**,** 185-206.

Hansen, M. C., R. S. Defries, J. R. G. Townshend & R. Sohlberg (2000) Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing,* 21**,** 1331-1364.

Hansen, M. C., A. Egorov, D. P. Roy, P. Potapov, J. C. Ju, S. Turubanova, I. Kommareddy & T. R. Loveland (2011) Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters,* 2**,** 279-288.

Harris, N. L., S. Brown, S. C. Hagen, S. S. Saatchi, S. Petrova, W. Salas, M. C. Hansen, P. V. Potapov & A. Lotsch (2012) Baseline Map of Carbon Emissions from Deforestation in Tropical Regions. *Science,* 336**,** 1573-1576.

Hastie, T., R. Tibshirani & J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

He, L. M., J. M. Chen, S. L. Zhang, G. Gomez, Y. D. Pan, K. McCullough, R. Birdsey & J. G. Masek (2011) Normalized algorithm for mapping and dating forest disturbances and regrowth for the United States. *International Journal of Applied Earth Observation and Geoinformation,* 13**,** 236-245.

Hecht, S. B., R. B. Norgaard & G. Possio (1988) The Economics of Cattle Ranching in Eastern Amazonia. *Interciencia,* 13**,** 233-240.

Herold, M. (2009) An assessment of national forest monitoring capabilities in tropical non-Annex I countries: .

Herold, M., P. Mayaux, C. E. Woodcock, A. Baccini & C. Schmullius (2008) Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment,* 112**,** 2538-2556.

Holben, B. N. (1986) Characteristics of Maximum-Value Composite Images from Temporal Avhrr Data. *International Journal of Remote Sensing,* 7**,** 1417-1434.

Horler, D. N. H. & F. J. Ahern (1986) Forestry Information-Content of Thematic Mapper Data. *International Journal of Remote Sensing,* 7**,** 405-428.

Houghton, R. A., D. L. Skole, C. A. Nobre, J. L. Hackler, K. T. Lawrence & W. H. Chomentowski (2000) Annual fluxes or carbon from deforestation and regrowth in the Brazilian Amazon. *Nature,* 403**,** 301-304.

Huang, C., S. N. Goward, J. G. Masek, F. Gao, E. F. Vermote, N. Thomas, K. Schleeweis, R. E. Kennedy, Z. Zhu, J. C. Eidenshink & J. R. G. Townshend (2009a) Development of time series stacks of Landsat images for reconstructing forest disturbance history (International Journal of Digital Earth). *International Journal of Digital Earth,* 2**,** 195-218.

Huang, C. Q., S. N. Goward, J. G. Masek, F. Gao, E. F. Vermote, N. Thomas, K. Schleeweis, R. E. Kennedy, Z. L. Zhu, J. C. Eidenshink & J. R. G. Townshend (2009b) Development of time series stacks of Landsat images for reconstructing forest disturbance history. *International Journal of Digital Earth,* 2**,** 195-218.

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

Huete, A., C. Justice & H. Liu (1994) Development of Vegetation and Soil Indexes for Modis-Eos. *Remote Sensing of Environment,* 49**,** 224-234.

Huete, A. R., H. Q. Liu, K. Batchily & W. van Leeuwen (1997) A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment,* 59**,** 440-451.

IBGE. 2006. Censo Agropecuário 2006. Insituto Brasileiro de Geografia e Estatística.

IPCC. 2006. Volume 4 Agriculture, Forestry and Other Land Use. In *IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme,* ed. IPCC.

--- (2007) Climate Change 2007-The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report.

Jin, S. M. & S. A. Sader (2005) Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment,* 94**,** 364-372.

Kennedy, R. E., Z. Yang & W. B. Cohen (2010) Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - Temporal segmentation algorithms (Remote Sensing of Environment). 114**,** 2897-2910.

Key, C. H. & N. C. Benson. 1999. Measuring and remote sensing of burn severity: the CBI and NBR. In *Joint Fire Science Conference and Workshop,* eds. L. F. Neuenschwander & K. C. Ryan. Idaho: Boise.

Key, C. H. & N. C. Benson. 2005. Landscape assessment: ground measure of severity, the Composite Burn Index; and remote sensing of severity, the Normalized Burn Ratio. In *FIREMON: Fire Effects Monitoring and Inventory System,* eds. D. C. Lutes, R. E. Keane, J. F. Caratti, C. H. Key, N. C. Benson & L. J. Gangi. USDA Forest Service.

Kuplich, T. M. (2006) Classifying regenerating forest stages in Amazônia using remotely sensed images and a neural network (Forest Ecology and Management). 234**,** 1-9.

Lee, D. S., J. C. Storey, M. J. Choate & R. W. Hayes (2004) Four years of Landsat-7 on-orbit geometric calibration and performance. *Ieee Transactions on Geoscience and Remote Sensing,* 42**,** 2786-2795.

Lu, D., M. Batistella & E. Moran (2007) Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and Radarsat data (International Journal of Remote Sensing). 28**,** 5447-5459.

Lu, D. & Q. Weng (2007) A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing,* 28**,** 823-870.

Masek, J. G., E. F. Vermote, N. Saleous, R. Wolfe, F. G. Hall, F. Huemmrich, F. Gao, J. Kutler & T. K. Lim (2006a) A Landsat surface reflectance data set for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters,* 3**,** 68-72.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, F. Gao, J. Kutler & T. K. Lim (2006b) A Landsat surface reflectance dataset for North America, 1990-2000. *Ieee Geoscience and Remote Sensing Letters,* 3**,** 68-72.

Maus, V. 2014. Satellite time series analysis for land use/cover change detection, IIASA Interim Report IR-14-017. International Institute for Applied Systems Analysis.

McDonald, A. J., F. M. Gemmell & P. E. Lewis (1998) Investigation of the utility of spectral vegetation indices for determining information on coniferous forests. *Remote Sensing of Environment,* 66**,** 250-272.

Mitchell, M. W. (2011) Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. *Open Journal of Statistics,* 1**,** 205-211.

Moran, E. & E. Brondizio. 1998. Land-Use Change After Deforestation in Amazonia. In *People and Pixels: Linking Remote Sensing and Social Science,* eds. D. Liverman, E. Moran, R. Rindfuss & P. Stern, 94-120. Washington D.C.: National Academy Press.

Moran, E. F., E. S. Brondizio, J. M. Tucker, M. C. da Silva-Forsberg, S. McCracken & I. Falesi (2000) Effects of soil fertility and land-use on forest succession in Amazonia. *Forest Ecology and Management,* 139**,** 93-108.

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

Nepstad, D. (1989) Ranching, Logging, and the Transformation of an Amazonian Landscape. *TRI News,* 6**,** 3-6.

Nepstad, D. C., C. M. Stickler, B. Soares & F. Merry (2008) Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 363**,** 1737-1746.

Nguyen, T.-T., J. Z. Huang & T. T. Nguyen (2015) Unbiased Feature Selection in Learning Random Forests for High-Dimensional Data. *The Scientific World Journal,* 2015**,** 18.

Numata, I., M. A. Cochrane, C. M. Souza & M. H. Sales (2011) Carbon emissions from deforestation and forest fragmentation in the Brazilian Amazon. *Environmental Research Letters,* 6.

Petitjean, F., J. Inglada & P. Gancarski (2012) Satellite Image Time Series Analysis Under Time Warping. *Ieee Transactions on Geoscience and Remote Sensing,* 50**,** 3081-3095.

Prishchepov, A. V., V. C. Radeloff, M. Dubinin & C. Alcantara (2012) The effect of Landsat ETM/ETM plus image acquisition dates on the detection of agricultural land abandonment in Eastern Europe. *Remote Sensing of Environment,* 126**,** 195-209.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramankutty, N., H. K. Gibbs, F. Achard, R. Defriess, J. A. Foley & R. A. Houghton (2007) Challenges to estimating carbon emissions from tropical deforestation. *Global Change Biology,* 13**,** 51-66.

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

Segal, M. R. 2004. Machine Learning Benchmarks and Random Forest Regression.

Serrao, E. A. S. & J. M. Toledo. 1989. A procura de sustentabilidade em pastagens Amazonicas. In *Alternatives to Deforestation: Coexistence of humans and the Amazon forest,* ed. A. B. Anderson, 281. New York: Columbia University Press.

Shimabukuro, Y. E., G. T. Batista, E. M. K. Mello, J. C. Moreira & V. Duarte (1998) Using shade fraction image segmentation to evaluate deforestation in Landsat Thematic Mapper images of the Amazon Region. *International Journal of Remote Sensing,* 19**,** 535-541.

Soares-Filho, B. S., D. C. Nepstad, L. M. Curran, G. C. Cerqueira, R. A. Garcia, C. A. Ramos, E. Voll, A. McDonald, P. Lefebvre & P. Schlesinger (2006) Modelling conservation in the Amazon basin (Nature). 440**,** 520-523.

Srivastava, T. 2015. Tuning the parameters of your Random Forest model. In *Analytics Vidhya*.

Statnikov, A., L. Wang & C. F. Aliferis (2008) A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *Bmc Bioinformatics,* 9.

Strobl, C., A. L. Boulesteix & T. Augustin (2007) Unbiased split selection for classification trees based on the Gini Index. *Computational Statistics & Data Analysis,* 52**,** 483-501.

Strobl, C., T. Hothorn & A. Zeileis. 2009. Party on! A New, Conditional Variable Importance Measure for Random Forests Available in the party Package. 1-5. Munich: Department of Statistics, University of Munich.

Stueve, K. M., C. H. Hobie Perry, M. D. Nelson, S. P. Healey, A. D. Hill, G. G. Moisen, W. B. Cohen, D. D. Gormanson & C. Huang (2011) Ecological importance of intermediate windstorms rivals large, infrequent disturbances in the northern Great Lakes. *Ecosphere,* 2.

Townshend, J. R. G., C. O. Justice, C. Gurney & J. Mcmanus (1992) The Impact of Misregistration on Change Detection. *Ieee Transactions on Geoscience and Remote Sensing,* 30**,** 1054-1060.

Tucker, C. J. (1979) Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment,* 8**,** 127-150.

UNFCCC. 2011. Report of the Conference of the Parties on its sixteenth session, Decision 1/CP.16. paragraphs 70-72, Annex I

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky & J. W. Munger (2007) Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research-Biogeosciences,* 112.

Verissimo, A., P. Barreto, M. Mattos, R. Tarifa & C. Uhl (1992) Logging Impacts and Prospects for Sustainable Forest Management in an Old Amazonian Frontier - the Case of Paragominas. *Forest Ecology and Management,* 55**,** 169-199.

Viovy, N. (2000) Automatic Classification of Time Series (ACTS): a new clustering method for remote sensing time series. *International Journal of Remote Sensing,* 21**,** 1537-1560.

Wardlow, B. D., S. L. Egbert & J. H. Kastens (2007) Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment,* 108**,** 290-310.

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

Xiao, X. M., S. Boles, J. Y. Liu, D. F. Zhuang & M. L. Liu (2002) Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data. *Remote Sensing of Environment,* 82**,** 335-348.

Xu, L., A. Samanta, M. H. Costa, S. Ganguly, R. R. Nemani & R. B. C. L. Myneni (2011) Widespread decline in greenness of Amazonian vegetation due to the 2010 drought. *Geophysical Research Letters,* 38**,** n/a-n/a.

Yang, X. & C. P. Lo (2002) Using a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *International Journal of Remote Sensing,* 23**,** 1775-1798.

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

Zhu, Z. & C. E. Woodcock (2014a) Automated cloud, cloud shadow, and snow detection in multi-temporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment,* 152**,** 217-234.

Zhu, Z. & C. E. Woodcock (2014b) Continuous change detection and classification of land cover using all available Landsat data (Remote Sensing of Environment). 144**,** 152-171.

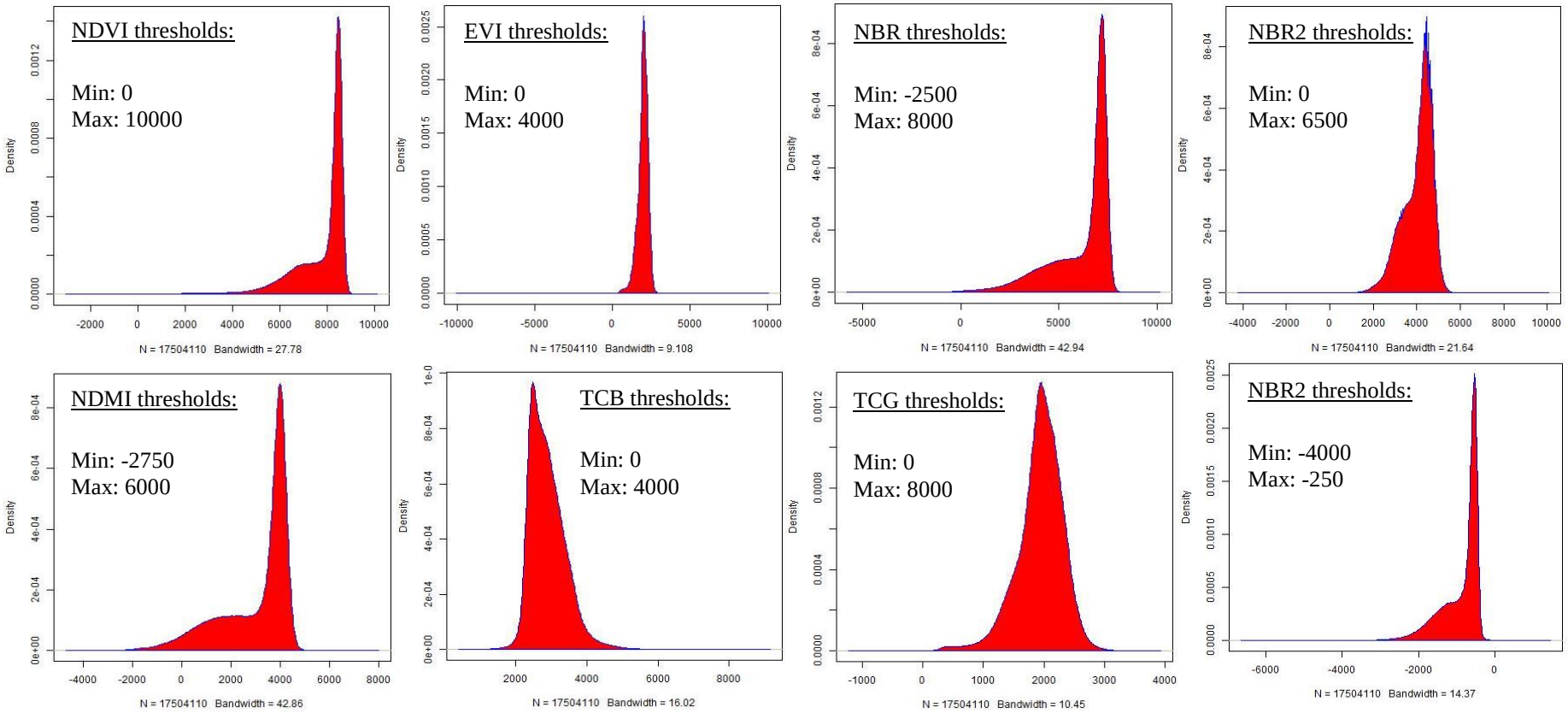
10 Annexes

Annex 1: Landsat imagery used in this study

Path-row	Acquisition date (day-month-year)	Cloud Cover * (%)	Data Type (sensor, level)	Geometric RMSE (m)	UTM (zone)
222-62	1-7-2009	38.7	ETM+ L1T	5.83	23
222-62	9-7-2009	18.08	TM L1T	4.122	23
222-62	17-7-2009	42.56	ETM+ L1T	7.755	23
222-62	2-8-2009	34.27	ETM+ L1T	5.921	23
222-62	10-8-2009	0.39	TM L1T	6.071	23
222-62	22-11-2009	40.83	ETM+ L1T	9.171	23
222-62	14-3-2010	35.38	ETM+ L1T	10.045	23
222-62	2-6-2010	5.34	ETM+ L1T	5.388	23
222-62	26-6-2010	0	TM L1T	5.32	23
222-63	1-7-2009	17.18	ETM+ L1T	4.096	23
222-63	9-7-2009	4.44	TM L1T	4.48	23
222-63	17-7-2009	5.02	ETM+ L1T	4.477	23
222-63	2-8-2009	8.69	ETM+ L1T	4.588	23
222-63	18-8-2009	29.61	ETM+ L1T	6.271	23
222-63	3-9-2009	57.44	ETM+ L1T	2.815	23
222-63	21-10-2009	40.72	ETM+ L1T	11.087	23
222-63	22-11-2009	2.74	ETM+ L1T	3.825	23
222-63	30-11-2009	32.53	TM L1T	6.721	23
222-63	9-5-2010	16.57	TM L1T	4.627	23
222-63	17-5-2010	48.8	ETM+ L1T	5.054	23
222-63	2-6-2010	28.44	ETM+ L1T	3.53	23
222-63	26-6-2010	0	TM L1T	4.763	23
223-62	8-7-2009	19.1	ETM+ L1T	4.712	22
223-62	1-8-2009	10.74	TM L1T	5.511	22
223-62	9-8-2009	46.72	ETM+ L1T	4.428	22
223-62	17-8-2009	0.08	TM L1T	4.598	22
223-62	25-8-2009	20.61	ETM+ L1T	4.511	22
223-62	26-9-2009	41.73	ETM+ L1T	4.416	22
223-62	28-10-2009	52.43	ETM+ L1T	3.82	22
223-62	21-11-2009	37.33	TM L1T	8.99	22
223-62	29-11-2009	45.08	ETM+ L1T	3.347	22
223-62	16-1-2010	54.22	ETM+ L1T	4.042	22
223-62	9-2-2010	50.19	TM L1T	10.083	22
223-62	5-3-2010	55.69	ETM+ L1G	3.478	22
223-62	8-5-2010	27.93	ETM+ L1T	10.864	22
223-62	25-6-2010	52.96	ETM+ L1T	3.859	22

* approximate cloud coverage given in metadata of downloaded scenes

Annex 2: Probability distribution of spectral index values in a non-cloudy scene



*Kernel densities were derived from Landsat scene "LT52220622009222" (sensor: TM, path-row: 222-62, date: 2009-222)

Annex 3: Geographic extents of overlap zones

Zone	Xmin	Xmax	Ymin	ymax	Overlap tiles
NW	Xmin (223-62)	Xmax (223-62)	Ymax (222-63)	Ymax (222-62)	222-62, 223-62
NE	Xmax (223-62)	Xmax (222-62)	Ymax (222-63)	Ymax (222-62)	222-62
SW	Xmin (223-62)	Xmax (223-62)	Ymin (222-63)	Ymax (222-63)	222-62, 222-63, 223-62
SE	Xmax (223-62)	Xmax (222-62)	Ymin (222-63)	Ymax (222-63)	222-62, 222-63

Annex 4: Usable pixels for LU classification after cloud masking with Fmask and Tmask

LU	Nr. of pixels with > 1 observation in rain and dry season		
	Total	>1 in both seasons (%)	>2 in both seasons (%)
Fmask			
Cropland	369141	68	22.9
Pasture	517072	74.2	18.3
Secondary Vegetation	80385	69.1	17
Tmask			
Cropland	369141	22.1	7
Pasture	517072	39.2	12.4
Secondary Vegetation	80385	35.9	11.8

Annex 5: Random forest tuning parameters

Model type	Cloud masking approach	mtry	nTree	Number of input variables	nodesize
Initial tuning		10	501	120	1
Full	Fmask	19	75	90	1
Full	Tmask	19	75	92	1
Isolated NDVI	Tmask	5	31	7	
Isolated EVI	Tmask	5	31	3	
Isolated NBR	Tmask	9	31	6	
Isolated NBR2	Tmask	3	31	3	
Isolated NDMI	Tmask	5	31	7	
Isolated TCG	Tmask	5	31	4	
Isolated TCB	Tmask	5	31	3	
Isolated TCW	Tmask	5	31	6	

Annex 6: Full post-deforestation LU maps generated from data processed with Fmask and Tmask-derived cloud masks