Pan-tropical modelling of land cover and land-use change trajectories for newly deforested areas

Coca Castro, Alejandro

Awarding institution:
King's College London

Pan-tropical modelling of land cover and land-use change trajectories for newly deforested areas

Alejandro Coca Castro

Thesis submitted to King's College London
for the degree of Doctor of Philosophy

Department of Geography
King's College London

March 2020

There is always hope...

Abstract

Despite extensive efforts in modelling and monitoring forest cover and forest change, few advances in the study of land cover (LC) and land use (LU) dynamics following deforestation (so-called post-loss LC/LU change) exist. The current plethora of multi-source datasets with considerable high-frequency records and emerging technologies (e.g. cloud-computing) to process them, provide new possibilities in the study of long-term (>2 years) LC/LU dynamics over deforested areas. Building on these advances, this research aims to develop models for mapping and analysing post-loss LC/LU dynamics pantropically, based on time series analysis of earth observation and contextual data, for use in environmental and agricultural studies. The first part of this research conducts a state-of-art review which explores a rich body of theory, definitions and global policies relevant to the study of post-loss LC/LU change. Furthermore, a description of existing studies showing methodological advances at a large scale, from national to global, provide guidance for potential directions of work. As the majority of these advances rely on georeferenced data, the next step of this research investigates multiple data sources, from field records to coarse satellite imagery of over 120 sites detected as disturbed by Terra-i (a near real-time monitoring system for LC conversion). This local-scale but data-rich investigation helps to highlight opportunities and challenges in the study of long-term post-loss LC/LU change.

By integrating the lessons learned above, the second part of the research deploys an end-to-end supervised deep learning architecture. This particular architecture is capable of ingesting and processing large volumes of MODIS satellite image time series in the generation of 19-years of LC predictions suited to the study of post-loss LC change trajectories. The scalability of the proposed model is evaluated through its implementation at macro-regional scale, i.e. the Amazon region, using big data principles and cloud-computing technologies. The implementation has brought some challenges, such as assessing the effectiveness of the model for different LC classification schemes derived from global LC maps. The results show the calibrated model outperforms conventional machine learning techniques used by other on-going initiatives mapping long-term post-loss LC/LU change. In addition,

the feasibility of transferring the calibrated model to predict LC in other tropical areas in Latin America, Asia and Africa is successfully demonstrated.

The final part of this thesis describes how post-loss LC change data generated for the Amazon can be characterised and grouped into different typologies. In doing so, this research introduces the minimum number of years for identifying these typologies and how they are linked to potential land-use types and impacts on carbon dynamics. The findings corroborate the need to go beyond simplistic assumptions, such as a full recovery or complete conversion to non-forest after disturbance, to enhance estimates of the impacts of deforestation. In addition, this research offers illustrative examples of the influence of 12 spatial layers representing inaccessibility, biophysical and policy factors on the observed post-loss LC change trajectories. It has been found some of these variables, such as proximity to farmland, proximity to urban/built-up, proximity to protected areas with sustainable use, and elevation, can explain human-related post-loss LC change.

This PhD project is innovative in its use and deployment of novel methods (e.g. deep learning and sequence analysis) and technologies (e.g. cloud-computing) to analyse large volumes of high-frequency earth observation data and chronological post-loss LC change records. Furthermore, this research supports modellers of ecosystem services, biodiversity, and other deforestation-relevant topics by going beyond the immediate state of deforestation to understanding the typologies and dynamics of long-term post-loss LC/LU change trajectories, which will have impacts on these areas of research.

Acknowledgements

As we humans find ourselves in strange times with COVID19, more than ever before, we need support and hope. I thank God for guiding me in coping with such stressful times and in delivering my PhD project on schedule. It has been an exciting but challenging four-year journey where I have established a new family with my wife, built discipline, hard skills, awesome friendships, and a valuable network. I could not have done this journey without the help of many people who I want to mention here.

Foremost, I would like to acknowledge my primary supervisor, Professor Mark Mulligan, for providing his input, support and invaluable feedback, and for his patience and willingness to share his wide knowledge on modelling from the very beginning through the completion of this PhD. Mark first encouraged me to contribute to the exciting field of land-use change research, since my enrolment at King's as a master student in 2014. I am extremely grateful to have had such an excellent supervisor and mentor who has allowed me to be innovative, exploring novel data-driven approaches that I have found beneficial to the field and my career.

I would like to extend my gratitude to the International Center for Tropical Agriculture - CIAT, in particular the Terra-i team led by Dr Louis Reymodin, for supplying data (field campaign records and very high-resolution satellite data) and resources (workstations and storage) which were crucial to the successful completion of this PhD research. I am personally very thankful to Louis who has trusted in my ability in conducting research since my internship at CIAT in May 2011. The experiences at the Terra-i initiative exposed me to the challenges associated with creating user-driven geospatial information, but they also persuaded me of the incredible opportunities this information provides for solving real-world problems.

I am deeply grateful to the Alcaldía Mayor de Bogotá and Fundación Ceiba for awarding me the scholarship Beca Rodolfo Llinás. Without their funding this research would not have been possible. I would like also to thank the Department of Geography and the EOES hub at King's College London for their financial aid in supporting my attendance at international conferences which were key to engaging with experts in the field and enhancing my PhD project. I would like to acknowledge

Google for a generous grant in Google Cloud Platform credits which were essential to the running of large-scale experiments in this research.

This research would not have been possible without the assistance of multiple academics and professionals who provided their valuable advice at different stages of this research. I am particularly indebted to my second supervisor, Dr Thomas Smith from the London School of Economics, and external academics including Dr Andrés Pérez-Uribe and Dr Hector Fabio Satizábal from the School of Business and Engineering Vaud, Marc Rußwurm from the Technical University of Munich, and Eric Rahn from CIAT. Additionally, special thanks to Aloïs Tilloy, Claudia Gutiérrez-Arellano, Daniel Fisher, Kris Chan, Laura Canevari and Samantha Day from the Department of Geography at King's, for reading through my draught chapters and offering their helpful suggestions.

To all my friends, I do not have enough words to those who supported me on this journey. I would like to express my sincere 'Muchas Gracias' to my PhD 'Familia' at King's for making my life truly special with their friendship. Thank you all for many memorable moments at lunchtime, coffee breaks, dinners, GIFs, birthday celebrations, night-outs and hikes. Equally significant, my most warm thanks also go to Ana and Jose; Astrid, Dan and Flora; my cousins Freddy and Lili; Julie and Familia; and Emma and Christina; thanks for your unconditional support and making London feel like home. I also like to thank my friends back home in Bogotá and Cali ("The Branch of Heaven") who have not forgotten about me since moving to the UK.

I would not be the person I am now without the help of my family. I am eternally grateful to my parents, Luis Guillermo and Alicia for their continuous support, love, trust and guidance in all steps of my life. I am also very thankful to my brothers, Andrés and Camilo, and my extended family, Julita, Nair, Rosa, Doña Nelcy and Luis, with whom I have celebrated common achievements from our early days until now.

Finally, I would like to thank my best friend, love and wife Alejandra. Thank you for providing inspiration and endless encouragement during the completion of my PhD. Your presence in both the rewarding and challenging moments that a PhD leads us through was fundamental to my successfully making it to the finish line. I look forward to sharing many more projects, adventures and years together with you!

Abbreviations

AI	- Artificial Intelligence
ANNs	- Artificial Neural Networks
CNN	- Convolutional Neural Network
CGIAR	- Consultative Group on International Agricultural Research
CIAT	- International Center for Tropical Agriculture
DL	- Deep Learning
ESRI	- Environmental Systems Research Institute
EO	- Earth Observation
ESA	- European Space Agency
FAO	- Food and Agriculture Organization
FRA	- Forest Resources Assessment
FT	- Forest transition
FIRMS	- Fire Information for Resource Management System
GBDX	- Geospatial Big Data platform
GEE	- Google Earth Engine
GIS	- Geographical Information System
GFC	- Global Forest Change
GRU	- Gated Recurrent Unit
HANTS	- Harmonic ANalysis of Time Series
IGBP	- International Geosphere-Biosphere Programme
ETM+	- Enhanced Thematic Mapper Plus
LC	- Land Cover
LCCS	- Land Cover Classification System

LSTM - Long Short-Term Memory

LU - Land Use

ML - Machine Learning

MLR - Multinomial Logistic Regression

MODIS - MODerate Resolution Imaging Spectroradiometer

MTLCC - Multitemporal Land Cover Classification Network

NASA - National Aeronautics and Space Administration

NDVI - Normalized Difference Vegetation Index

OA - Overall Accurary

OM - Optimal Matching

QA - Quality

RAISG - Amazon Geo-Referenced Socio-Environmental Information Network

REDD+ - Reducing emissions from deforestation and forest degradation

RF - Random Forest

RFE - Recursive Feature Elimination

RNN - Recurrent Neural Network

SA - Sequence Analysis

SEI - Shannon Evenness Index

SDG - Sustainable Development Goals

SITS - Satellite Image Time Series

SRTM - Shuttle Radio Topography Mission

SVM - Support Vector Machine

TL - Transfer Learning

TR - Transition Rates

VHR - Very High Resolution

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

1 Introduction

1.1 Context

Forests play a substantial role in biodiversity protection, climate regulation and human livelihoods (Chazdon *et al.*, 2016). The partial success of countries in reducing forest cover loss in the last decade shows that the commitments of zero net deforestation by 2030 are plausible (MacDicken, 2015). However, halting deforestation globally will require both qualitative and quantitative estimates to address and identify unsustainable commodity chains (Meyfroidt *et al.*, 2014) and other emerging drivers related with forest loss including large infrastructure (Laurance *et al.*, 2014) and mining projects (Asner *et al.*, 2013; Edwards *et al.*, 2014; Abood *et al.*, 2015; De Sy *et al.*, 2019).

Novel data analytics methods and big data technologies have been proposed as a solution from simply mapping forests to monitoring the complex dynamics of forest ecosystems structure, function and composition (Reiche *et al.*, 2016). For instance, operational and robust near-real time forest monitoring systems such as the FORest Monitoring for Action (Hammer *et al.*, 2009), Terra-i (Reymondin *et al.*, 2012), Global Forest Cover (Hansen *et al.*, 2013) and Global Land Analysis and Discovery (Hansen *et al.*, 2016) have largely contributed to estimate and identify hotspots of newly (>2000) forest loss areas pantropically (Tang *et al.*, 2019).

In an era of climate change, land cover (LC) and land use (LU) change after forest loss, also referred in the literature as post-loss LC/LU change (Pendrill and Persson, 2017), is one of the current uncertainties and challenges in the calculation of global/regional carbon fluxes (Morton *et al.*, 2006; Prestele *et al.*, 2017; Pendrill and Persson, 2017; Ramankutty and Graesser, 2017; De Sy *et al.*, 2019; Tang *et al.*, 2020) and impacts on ecosystem services and biodiversity (Wilson *et al.*, 2017). This is particularly true in developing countries in the tropics where data on forest inventories and related forest change dynamics are generally of insufficient quality and geographically limited (Romijn *et al.*, 2013). Consequently, in these countries there are few notions about historical reference

levels and LC/LU associated emissions after deforestation, these remaining relevant to financial ecosystem service payment schemes such as REDD+ (Ryan *et al*., 2014; Sloan and Sayer, 2015; De Sy *et al*., 2019).

1.2 Research problem

Despite extensive efforts in modelling and monitoring forest cover and forest change, few advances in the study post-loss LC/LU change (Pendrill and Persson, 2017; De Sy *et al*., 2019; Pendrill, Persson, Godar, and Kastner, 2019) exist. This aspect is more urgent for areas in the tropics where different expansion into native vegetation occurs (Curtis *et al*., 2018), therefore where impacts on biodiversity and carbon stocks are expected to be the greatest (Myers *et al*., 2000; Saatchi *et al*., 2011).

Multiple uncertainties might be reduced with further developments measuring and modelling LC/LU change over newly deforested areas. For example, the multiple pathways of crop expansion (Meyfroidt *et al*., 2014), in particular at the cost of forest, and related with certain practices (i.e. fire, clear cutting), can be unmasked and monitored. Attempts to update these pathways are key due to the appearance (and dominance) of emerging drivers of forest loss such as mining and commercial cropping shaped by complex interactions of underlying forces of social, political, economic, technological and cultural aspects (Hosonuma *et al*., 2012).

Thanks to advances in data availability and emerging technologies e.g. cloud computing new opportunities have emerged in LC/LU change research (Bey, Sánchez-Paus Díaz, *et al*., 2016; Soliman *et al*., 2017; Saah *et al*., 2019). For instance, the increasing availability of free of charge remote sensing and low-cost environmental/social sensing data (e.g. through crowdsourcing) is expected to enhance existing LC/LU change models (Joshi *et al*., 2016; Fritz *et al*., 2017). In order to support an efficient use and extraction of information from these data-rich resources, the development of robust data analytics and big data technologies are playing a pivotal role for handling, extracting, analysing and producing novel knowledge for LC/LU change science from large volumes of data (Camara *et al*., 2016; Saah *et al*., 2019).

Despite the above advances, the study of post-loss LC/LU change at large scales, from national to global, has mainly focused on the immediate (1 or 2 years) change with limited temporal scope using either sampling-based methods (De Sy *et al*., 2015; Tyukavina *et al*., 2017; De Sy *et al*., 2019), GIS-based operations (Pendrill and Persson, 2017) or visual interpretation of satellite data at very coarse scale (Curtis *et al*., 2018). Only a few groups of scholars and practitioners have proposed methodologies to map long-term (>2 years) trajectories of post-loss LC/LU change at annual (MapBiomas Project, 2019; Fagua and Ramsey, 2019) or biennial (Almeida *et al*., 2016; Arévalo *et al*., 2019) time steps. Tang *et al*. (2020) recently pointed out the benefits of analysing LC dynamics over post-disturbed areas through developing a spatially explicit carbon bookkeeping model. The authors claim the awareness of these dynamics by the model provides better estimates of carbon emissions and uptake than traditional assumptions of a full recovery or complete conversion to non-forest after disturbance.

Henceforth, based on the above premises and previous work related to the study of dynamics in deforested areas (Coca-Castro *et al*., 2014; Coca-Castro, 2015), this research aims to contribute in the development of models for mapping and analysing long-term trajectories of post-loss LC/LU change pantropically. The derived information and outcomes are targeted to scholars and policy makers consuming and/or producing research in LC/LU change science and those addressing impacts of LC/LU change on ecosystem services.

1.3 Aim

To develop models for mapping and analysing long-term trajectories of post-loss land cover and land-use change pantropically, based on time series analysis of earth observation (EO) and contextual data, for use in environmental and agricultural studies.

To achieve this aim, a set of steps will be required. These steps can be conveyed in the following two research questions:

*Can **land cover** change trajectories following deforestation be monitored at the large scale, based on pre-existing contextual data and time series analysis of freely available multispectral EO data?*

*To what extent can mapped land cover change trajectories be characterised and explained by socio-environmental conditions and indicate which **land uses** are the proximate causes of deforestation?*

1.4 Objectives

The following five objectives are set to fulfil the aim and research questions:

Objective 1. To explore existing knowledge and identify gaps regarding post-loss LC/LU change modelling and monitoring for tropical areas

To identify the advances and gaps in the study of post-loss LC/LU change, it is fundamental to examine the main principles of forest cover and forest cover change from both theoretical, technical and policy perspectives. Therefore, the narrative behind forest research such as forest transitions and drivers of forest change is revisited through a literature review. For the technical advances, a brief overview is made regarding the data availability and methodological advances in LC/LU change, particularly focused on post-loss LC/LU. Finally, the relevance and implications of the study of post-loss LC/LU change are highlighted according to the current policy frameworks.

Objective 2. To conduct an exploratory analysis over deforested sites with high levels of LC/LU information from ground reference observations to multi-temporal satellite data

Complementary to the state-of-the-art review, two data-rich study areas in Peru are analysed to a better understanding of main opportunities and challenges for the study of long-term trajectories of post-loss LC/LU change. Both areas consist of a high quantity of information from *in-situ* records in the field to free-of-charge repeated observations from satellite data. Inspired by existing tools for visual interpretation of LC/LU sample areas such as Geo-wiki (Fritz *et al*., 2009), Collect Earth (Bey, Sánchez-Paus Díaz, *et al*., 2016) and LACO-Wiki (See *et al*., 2017), a customised workflow is presented for interpreting and analysing post-loss LC/LU dynamics at annual time steps per study area.

Objective 3. To assess pre-existing global LC classifications and high frequency multispectral time series suited to large-area study of post-loss LC change trajectories

Satellite imagery and labelled data (reference labels) are generally the main sources for training and validating predictive models for LC/LU mapping. This objective is focused on introducing the proposed method for generating large-area LC/LU predictions using satellite time series (SITS) suited to the study of post-loss LC/LU change. In this regard, the feasibility of multiple global LC maps to extract labelled data is assessed. Due to none of these existing maps effectively represent LU types, only LC is considered. The selection of the best suited satellite and labelled data to train the proposed model is conducted by a visual inspection and validation protocol designed to measure its feasibility to the study of post-loss LC change.

Objective 4. To calibrate and implement the proposed data analytics technique for generating 19-years of LC information using the best suited satellite and labelled data across a large area in the tropics

The proposed method for generating large-area LC predictions for the study of post-loss LC change is calibrated and assessed against other conventional algorithms. The assessment includes mainly two aspects; 1) model evaluation using performance metrics i.e. overall accuracy, between predicted and observed values; and 2) spatial and temporal coherence, i.e. visual inspection of the model predictions.

The scalability of the proposed model is evaluated through its implementation at macro-regional scale using big data principles and cloud-computing technologies. The selection of a large area helps verifying the generalisation capabilities of the proposed model. This verification is performed in a data-rich region in the pantropics like the Amazon where regional-tuned LC/LU data exist. In addition to this verification, the feasibility of transferring the trained model using satellite and labelled data across the Amazon to predict LC in other pantropical areas in Latin America, Asia and Africa is demonstrated.

Objective 5. To understand the spatial and temporal patterns of observed post-loss LC change trajectories

Multi-temporal LC predictions generated over the Amazon are then used in combination with a deforestation dataset to produce long-term post-loss LC change data. This objective aims to characterise spatial and temporal patterns of observed post-loss LC dynamics and how they are linked to potential land-use types and impacts on carbon dynamics. Additionally, factors associated with the causal chain leading to forest change as reported in the literature e.g. environmental and socio-economic conditions are assessed to assess their influence in the observed post-loss LC dynamics.

1.5 Overview of the thesis

The content of this thesis comprises seven remaining chapters:

Chapter 2. A review of the study of pantropical post-loss land cover and land use change trajectories

Chapter 3. Using ground reference and very high-resolution imagery to better understand the challenges to mapping post-loss LC/LU change trajectories

Chapter 4. Assessment of pre-existing global LC classifications and high frequency multispectral time series suited to the study of post-loss LC change

Chapter 5. Calibration and comparison of the proposed model for the study of post-loss LC change

Chapter 6. Understanding the patterns of post-loss LC change trajectories over newly deforested areas across the Amazon region through data-mining

Chapter 7. Conclusions and future work

2 A review of the study of pantropical post-loss land cover and land use change trajectories

2.1 Introduction

Land system science can be defined as an interdisciplinary field that acts as a platform to integrate different perspectives and dimensions of global change research (Verburg *et al.*, 2015). As part of this field, a group of scholars and practitioners have focused on observing and modelling land cover (LC) and land use (LU) change from pantropical forest. In this regard, initiatives such as the latest Global Forest Resources Assessment (FRA) of the Food and Agricultural Organization of the United Nations (FAO) (MacDicken, 2015) shows a notable trend of forest loss occurring in the tropics, with either stable or expanding forest in other domains (Sloan and Sayer, 2015). Whilst FAO-FRAs's findings are periodically reported worldwide, land system scholars have highlighted the essential role of modelling and monitoring LC/LU change following deforestation (also called post-loss LC/LU change) (Hosonuma *et al.*, 2012; Salvini *et al.*, 2014; Pendrill and Persson, 2017; De Sy *et al.*, 2019; Tang *et al.*, 2020)

Figure 2.1 frames three key components related to the study of pantropical forest change. Theoretical, methodological and global policy considerations intersecting these components are briefly described in the following sections with a particular focus on immediate i.e. direct conversion (1-2 years) and long-term i.e. trajectories (>2 years) post-loss LC/LU change over large areas, from national to global scale.

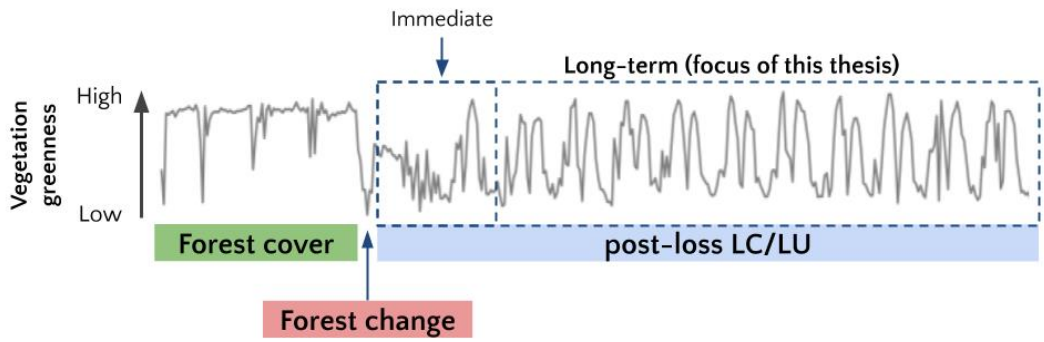


Figure 2.1 Diagram showing three main components in the study of forest change. Y-axis indicate the level of vegetation greenness as measured by earth observation data. Studies in post-loss LC/LU can be split in short-term i.e. immediate and long-term i.e. trajectories.

2.2 Theoretical considerations

2.2.1 Forest and land system change concepts

It is estimated that more than 800 official definitions of forest exist to date and consequently this proliferation of forest terms exacerbates conflicts on the meaning and interpretation of forest change (Sexton *et al*., 2016; Chazdon *et al*., 2016). Existing forest definitions mainly vary according to five attributes: tree cover, tree height, plot size, and the actual and potential height of vegetation (Sexton *et al*., 2016). The most accepted definition worldwide of forest stems from FAO and the International Geosphere-Biosphere Programme (IGBP). FAO defines forests mainly by the criterion of tree cover greater than 10% with a height or expected height of 5-m or more (FAO, 2000). IGBP gives a more differentiated definition using a criterion of greater than 60% tree cover for forest, specifying savannah as 10–30% and woody savannah as 30–60% cover (Loveland and Belward, 1997). An expanded definition of FAO's and IGBP's definitions is proposed by the United Nations Framework Convention on Climate Change (UNFCCC). UNFCCC's definition is broadly used by the participating countries due to its flexibility of selecting thresholds between 10 and 30% of tree cover (UNFCCC, 2001). A further discussion of forest definitions and their influence on forest cover and forest change reporting can be found in Sexton *et al*. (2015) and Chazdon *et al*. (2016).

In addition to forest definition, it is important to differentiate the meaning of land cover and land use (Joshi *et al*., 2016). The term land cover is represented by the features characterising the earth's land surface and immediate subsurface, including biota, soil, topography, surface and groundwater, as well as human structures (Lambin *et al*., 2000). LC is determined by direct observation of the earth's surface and it is crucial for the development of physical environmental models (Comber, 2008). Land use refers to the purposes for which humans exploit the land cover (Lambin *et al*., 2000) and it is mainly required for policy and planning purposes (Comber, 2008). Although studying and reporting changes in LC are still important for land system research, there has been increased focus on integrating LC and LU information (Zhang *et al*., 2019). This integration is beneficial in discriminating LC classes with different LU properties, as different LU, even with the same LC, have differential environmental and social impacts

(Joshi *et al*., 2016). One example is in Latin America where natural grasslands are barely discriminated from planted pastures in terms of cover, yet they are fundamentally different land uses (Verburg *et al*., 2015).

The distinction between LC and LU is also fundamentally important when discussing other terminologies in land system science such as deforestation. Hence, it is important to determine whether to assess deforestation as a change in LC or LU. These two approaches can be distinguished as forest change (for a LC change) and reforestation/deforestation (for a LU change), respectively. The latter is clearly of human agency, whereas the former also occurs through natural processes such as river flooding (Pendrill and Persson, 2017).

2.2.2 The forest transition and Kuznets curve theories

The concepts of forest transition and Kuznets curve might be good starting points to illustrate the theory behind LC/LU change in forested areas. The forest transition (FT) theory aims to explain the spatial-temporal dynamics, causes, and implications of an observed shift from net forest loss to net forest gain (Mather, 1992). While this transition was initially observed in developed countries (Drummond and Loveland, 2010), some examples have been reported in several developing countries too (Pendrill, Persson, Godar, and Kastner, 2019). The Kuznets curve proposes an inverted relationship between environmental and socioeconomic conditions, stressing that a society's environment deteriorates as its economy develops, but is followed by improved environmental quality as *per-capita* income levels rise (Drummond and Loveland, 2010).

The original FT concept, proposed by Mather (1992), suggests a 'U-shaped curve' for forest land with respect to two time phases: 1) a declining phase in the early stages of economic development (i.e. mainly land requirement for agriculture) followed by 2) a recovery phase through conserving remaining primary forest, plantations, reforestation and/or land abandonment. Since its inception, the FT theory has co-evolved with land system science (Rudel *et al*., 2020). For instance, new transition phases have been added following an inverse 'J-shaped curve' over time. Furthermore, each transition phase has been attributed to certain LU types (Hosonuma *et al*., 2012) and was recently mapped for all countries worldwide (Pendrill, Persson, Godar, and Kastner, 2019) (see **Figure 2.2**).

The environmental Kuznets curve serves as a general framework for explaining two non-exclusive pathways, the economic development path and the forest scarcity path, usually related with the FT theory (Rudel *et al*., 2005). While both paths have been successfully characterised in certain regions (Meyfroidt and Lambin, 2011), some studies show that the Kuznets curve can partially explain forest gain. For instance, Crespo Cuaresma *et al*. (2017) show the hypothesised link between economic development and deforestation exists for developing countries however it does not apply for developed countries.

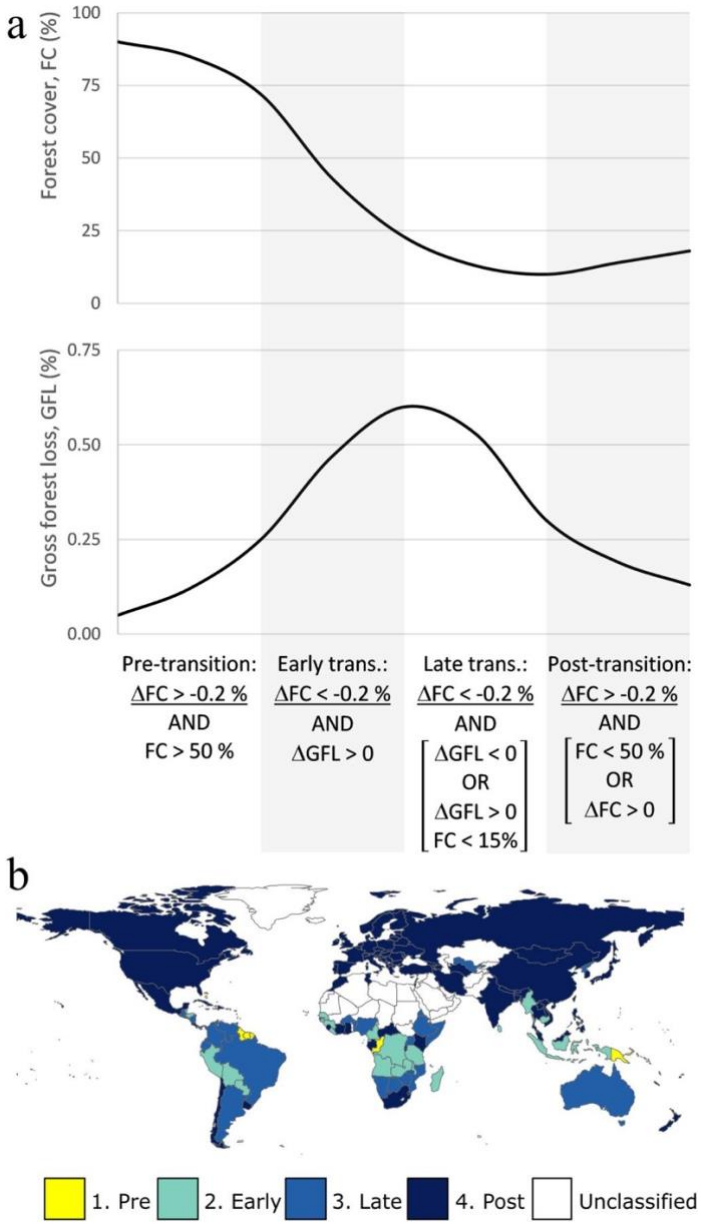


Figure 2.2 (a) Phases of the FT model according to the percentages of forest cover (FC), net forest cover change (ΔFC) and trend in gross forest loss (ΔGFL); and (b) distribution by country. Map shows most tropical countries are between *Early* to *Late* phases. Source: Pendrill, U. M. Persson, Godar, and Kastner (2019).

2.2.3 Approximations to the causes of pantropical forest change

There is an extensive narrative concerning the causes (and consequences) of pantropical forest change, see for example Meyfroidt *et al*. (2014). This section focuses on the essentials of how these causes are often categorised and their evolution since the inception of detailed spatial data have become available (i.e. earth observation data available from mid-1960's). Whilst these causes are usually associated with the land-use change outcome i.e. deforestation, they can also influence immediate and long-term post-loss LC/LU change trajectories as targeted in this thesis.

Multiple approaches and terminologies have been generated according to causal analysis in land system research (Meyfroidt, 2016) In particular, for a land use change event (outcome) such as deforestation, most existing literature typically distinguishes between two categories of causes; proximate and underlying (also called direct and indirect causes, respectively) (see **Figure 2.3**) (Lambin *et al*., 2001).

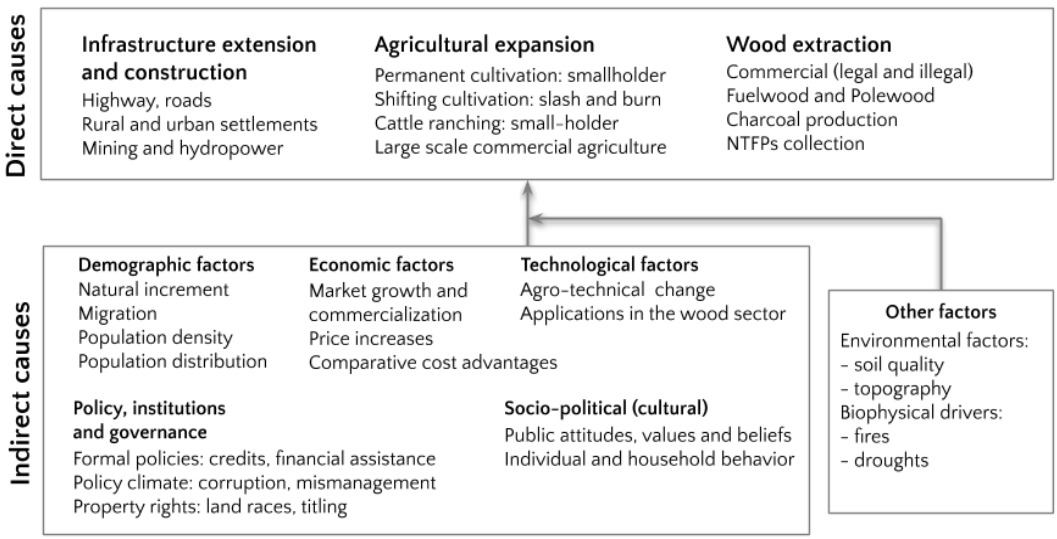


Figure 2.3 Direct and indirect causes of deforestation. Adapted from Geist and Lambin (2002).

On one hand, the proximate causes directly affect vegetation cover and thus it is typically land use or management causing the change (Meyfroidt, 2016) i.e. agriculture expansion, expansion of infrastructure and wood extraction (Lambin *et al*., 2001). On the other hand, the underlying causes work at different scales and are defined as factor(s) causing one or more proximate causes (Meyfroidt,

2016) i.e. demographic factors, economic factors, technological factors, policy and institutional factors, cultural factors and biophysical factors (Lambin *et al*., 2001). In terms of spatial scale, proximate causes generally operate at the local level (individual farms, households, or communities). In contrast, underlying causes can originate from the regional (districts, provinces, or country) or even global levels, with complex interactions between levels of organisation (Lambin *et al*., 2003). Chronologically, proximate causes are the factors involved in the last steps of the causal chain (or a continuous chain of processes between the cause and the outcome) and underlying causes are those involved in earlier steps in the chain. Further details about the dynamics between proximate and underlying causes of deforestation can be found in Meyfroidt (2016).

Rudel *et al*. (2009), Drummond and Loveland (2010), and Meyfroidt *et al*. (2013) claim that the causes of global deforestation can be divided into two time periods: state-enabled, smallholder deforestation (1965–1985) and enterprise-driven deforestation (1985–Present) (see **Figure 2.4**). The former was characterised by a high government intervention towards farmer-support, predominantly small-holders (i.e. agricultural modernisation or settlement schemes). For this period, poverty, poor agricultural technologies and population growth were viewed as the main causes of deforestation (Meyfroidt *et al*., 2013). In the latter period, the enterprise-driven, smallholder colonization schemes have declined as drivers of deforestation, in particular in Latin America and Southeast Asia (Asner *et al*., 2009; van Vliet *et al*., 2012). At the same time, fiscal austerity combined with the debt crises of the 1980s induced cost-cutting measures by central governments (Labán and Sturzenegger, 1994). Based on the crises of this period, governments have abandoned projects of new land settlement and consequently public resources for road infrastructure projects into forest-rich regions have decreased (Rudel *et al*., 2009). Meyfroidt *et al*. (2013) indicate that deforestation in the enterprise period has also been associated with international trade of agricultural products and remote urban demand, rather than rural population growth. For instance, certain countries, including Brazil (Arima *et al*., 2011; Macedo *et al*., 2012) and Indonesia (Carlson *et al*., 2012), have increasingly absorbed the rise of global demand for key agricultural products such as soybeans and palm oil, leading to rapid agricultural expansion (Meyfroidt *et al*., 2013).

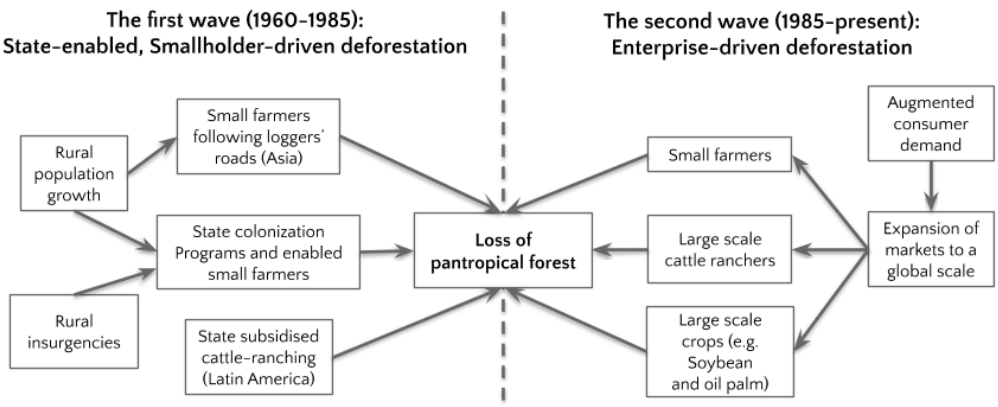


Figure 2.4 Main deforestation periods in the tropics as identified by Rudel *et al.* (2009). The scholars identified the dynamics in both periods by conducting a meta-analysis of 1975-2002 forest cover change between. The metanalysis included papers reporting the causes of tropical forest change, mostly in wet forests. Adapted from Rudel *et al.* (2009).

As part of the causal analysis, the study of post-loss LC/LU change emerges as a key component for a better understanding of the causes of current enterprise-driven deforestation pathways. van Vliet *et al.* (2012) suggest, in general, LC/LU conversions/transitions do not follow a fixed pattern. Moreover, the rate with which they can happen and the uncertain direction they may take can have different impacts on ecosystems (van Vliet *et al.*, 2012). According to Meyfroidt (2016), the combinations of causes, also called 'pathways', can lead to one or more outcomes. In other terms, this/these outcome(s) can be understood as the resulting LC/LU dynamics following forest loss. Because most land conversions are highly context specific, it is challenging to find studies synthesising post-loss LC/LU change. The few existing ones focus on immediate post-loss LC/LU change providing very broad categories at regional (De Sy *et al.*, 2015; Ramankutty and Graesser, 2017) and global level (Rudel *et al.*, 2009; Hosonuma *et al.*, 2012; De Sy *et al.*, 2019) (see **Figure 2.5**).

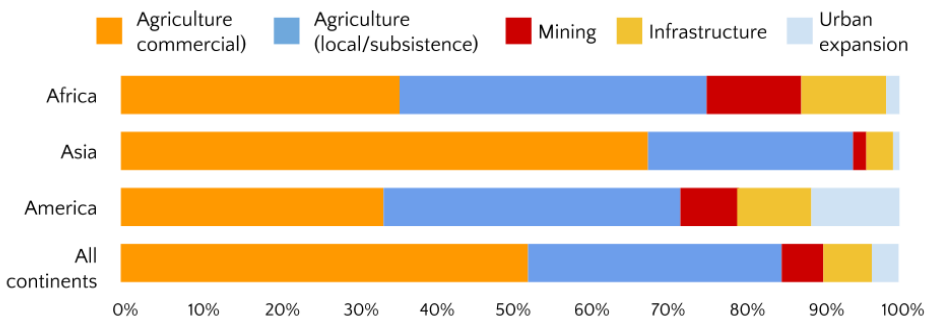


Figure 2.5 Continental-level estimations of the proportions of immediate post-loss LC/LU change, also known as direct drivers of deforestation, for the period 2000–2010, based on data from 46 tropical and subtropical countries. Adapted from Hosonuma *et al.* (2012).

2.2.4 Key determinant and location factors in pantropical forest change

Whilst multiple causes have been associated with forest change, only a set of them, termed as 'determinant' or 'spatial determinant', have been attributed as location factors or as a group of biophysical and socio-economic factors linked with spatial characteristics (spatial pattern or structure) of land systems (van Asselen and Verburg, 2012). According to Meyfroidt (2016), some of these factors directly affect the rate and quantity of LC/LU change, e.g. the area of forest cleared by new incoming migrants. Other factors influence the location of LC/LU change, e.g. the land suitability for cropland. The concept of 'determinant' might be best suited for relations that have a 'statistical explanation', but where causality is not necessarily asserted or assumed (Meyfroidt, 2016). A determinant may belong to either of the two categories of causes described in the previous section (proximate causes or underlying causes) as well as to the group of contextual factors i.e. factors which are supposed 'not to drive, but rather – to shape' forest change (Meyfroidt, 2016). The following subsections illustrate the main determinants used by spatially explicit models of frontier expansion causing deforestation in the tropics at regional (Aguiar *et al*., 2007; Scouvart *et al*., 2008; Redo *et al*., 2012) and global scale (Verburg *et al*., 2011).

2.2.4.1 Accessibility conditions and connections to markets

This group consists of accessibility conditions and connections to markets and ports as well as the proximity to roads, rivers and urban settlements (Aguiar *et al*., 2007). Accessibility to markets drives decisions by local agents by influencing the profitability of activities as a function of distance (transportation costs) to the central market/port (Scouvart *et al*., 2008). Moreover, some scholars (Pittman *et al*., 2010; Perz *et al*., 2007, 2015) suggest LC/LU and spatial pattern of land occupation is likely to vary by connectivity as well as by the conditions of the accessibility network (i.e. paved or unpaved roads) and market proximity. This joint combination affects the travel time. For example, Southworth *et al*. (2011) show LU types such as logging and pasture, which are associated with market-oriented activities, are sensitive to accessibility. In contrast, deforestation driven by illegal production systems such as coca crop plantations and illegal mining are mostly located in remote areas, in preference nearby to river tributaries (Asner *et al*., 2013; Chadid *et al*., 2015).

2.2.4.2 Occurrence of an extractive cycle

Contextual factors such as timber-production facilities and mineral deposits may have the capacity of stimulating the occupation of new areas (Aguiar *et al.*, 2007). The former factor promotes deforestation indirectly by opening access roads that attract spontaneous migrants into forest areas (Arima *et al.*, 2005). Moreover, logging also indirectly produces a temporal wealth in the area, which can then induce changes in the productive base (Scouvart *et al.*, 2008). Regarding the presence of mineral deposits, in particular gold mining, they can also attract new migrants by reorienting the already established colonists towards the extractive sector. Additionally, although a decrease in deforestation rates can be experienced while colonists leave their fields to engage in extractive activities, a sharp increase in deforestation can happen once extractive activities are abandoned (Scouvart *et al.*, 2008). Extractive cycles also lead to the development of settlements that under certain circumstances can evolve into towns and cities (Mertens, 2002).

2.2.4.3 Frontier age and type of agents

Scouvart *et al.* (2008) state that the age of a frontier is an important factor in explaining LC/LU dynamics in forests. In this way, households characterised by different levels of available labour force and production systems, determine the evolution of a frontier (H. Wang *et al.*, 2019). Over a certain period, land tenure generally changes according to the consolidation of land parcels (Scouvart *et al.*, 2008). The frontier age can be defined as the number of years (history) from the start of the colonisation (Scouvart *et al.*, 2008). For instance, in the Brazilian Amazon, an old frontier (20 years or more) is characterised by larger deforested areas but less rapid rates of change than a more recently opened area (less than 20 years), due mainly to the consolidation of land clearing and potential intensification of land use (Scouvart *et al.*, 2008).

In addition to the frontier age, the type of agents is a fundamental factor shaping deforestation processes and related LC/LU dynamics (Scouvart *et al.*, 2008). In terms of income, two groups of agents (or colonisers) can be distinguished: 1) those with reduced capital who mainly practice subsistence agriculture (mixed crops, agroforestry, swidden agriculture) and are often classified as first wave colonisers; 2) and those with higher capital practising models of more intensive

agriculture playing a major role as second wave colonisers (Scouvart *et al*., 2008). In other words, two groups of land-users are consistently identified in the literature: 1) smallholder cultivators, including settlers and traditional farmers, and 2) industrial farmers or plantation owners. These two groups differ in terms of their particular characteristics among regions, but their impacts might be similar (Ewers and Laurance, 2006). For instance, while most small-scale farmers in Central America have coffee plantations, in West Africa the major crop is cocoa. Similarly, commercial agriculture in the Brazilian Amazon consists mainly of soybean and cattle ranches, in comparison in Indonesia it is dominated by oil-palm plantations (Rudel, 2013).

2.2.4.4 Demographic

Some studies indicate an asymptotic relationship between population and deforestation, suggesting that population growth only plays an important role in the opening of new land areas (Pittman *et al*., 2010; Laurance *et al*., 2014). The urban and rural population can play different roles in forest change. While the former has a direct influence on forests through their conversion to other land uses, the latter has an indirect effect on deforestation through its demand for agricultural and forest products (Scouvart *et al*., 2008; DeFries *et al*., 2010).

2.2.4.5 Biophysical conditions and soil suitability

There are multiple biophysical factors that influence land-use decisions (soil quality, rainfall, terrain features and vegetation type), and some of them have major impact according to the particular characteristics of the study area (Scouvart *et al*., 2008). For the frontier expansion of cropland and pastoral land, the interannual variability in rainfall can in general play an essential role. Rainfall availability modifies land productivity and therefore produces a positive or negative feedback on the production of a given area under pastoral or cultivation use (Stéphenne and Lambin, 2001). In connection with soil quality, which mostly constrains tropical agriculture, it is related to chemical rather than physical soil properties i.e. soil nutrients (Lieth and Werger, 2012). These constraints also influence the type of production system. Moreover, certain combinations of both factors can lead to the development of agricultural activities and thus deforestation, e.g. high agricultural potential combined with a pronounced dry season, which facilitates the access and operation of agricultural machinery,

increases the probability of land clearing in the Brazilian Amazon (Scouvart *et al.*, 2008). In terms of clearing size, Ewers and Laurance (2006) suggest that the large-scale clearing activities might be less affected by small-scale variations in topography and soil conditions and thus more influenced by long, linear boundaries between adjacent, large-scale properties. The type of forest vegetation can also influence the size of clearings and land use activities after deforestation. For instance, changes in dry forests are more prone to large-scale activities than humid forests (Grau *et al.*, 2005).

Regarding the most relevant practices of forest clearing, fire has been associated as a tool for the conversion of land use from forest to agricultural or other purposes (Lima *et al.*, 2012; Armenteras *et al.*, 2017). Small- and large-scale farming operations can have different fire use preferences (i.e. large cattle ranches and industrial soybean farms using heavy machinery such as bulldozers and intensive use of water and fertilisers) which consequently produce different spatial patterns and impacts in the quality and quantity of other natural resources (i.e. water) (Ewers and Laurance, 2006).

2.2.4.6 Public policies

This group is mainly related to governmental actions, such as planned settlements and protection areas. In particular, the presence of protected areas can influence the type of LC/LU dynamics following deforestation (Jurjonas *et al.*, 2016). Additionally, another factor related to policies relies on the governance of tropical agricultural development i.e. commodity supply chain interventions such as the Roundtable on Sustainable Palm Oil in Indonesia and the Roundtable on Responsible Soy, the Soy Moratorium and the Cattle Agreement in Brazil (Meijer, 2015).

2.3 Methodological considerations

The study of forest change has been supported by a wide range of sources from local case studies using interviews and surveys to global scale studies based on models, remote sensing and/or other data e.g. crowdsourcing (Fritz *et al.*, 2017). This section is mainly focused on existing work in mapping forest change and post-loss LC/LU based on remote sensing. Satellite remote sensing data is by far

the main source in providing spatially-explicit information of LC/LU extent and dynamics at multiple scales (Verburg *et al.*, 2015).

2.3.1 Forest change

Different approaches have been proposed to monitor forest change using remote sensing data at multiple scales (Pendrill and Persson, 2017). Two main families of free-to-use products exist, those generated from (i) historical high-resolution imagery e.g. Landsat and (ii) composited high-temporal satellite data e.g. MODIS. While the former is useful for quantifying forest loss and gain for carbon accounting and long-term forest management, the latter is well-suited for capturing current forest change events in a timely manner (Tang *et al.*, 2019). **Table 2.1** describes the features and capacities of the existing deforestation datasets with pantropical coverage.

Table 2.1 Existing free-to-use deforestation datasets, ordered by spatial resolution, with full or partial pantropical coverage. Adapted from Mitchard (2016).

Dataset	Coverage	Spatial resolution	Input data	First year	Temporal resolution
Global Forest Change (GFC)[1]	Global	30-m	Landsat	2000	Annual
Global Land Analysis and Discover (GLAD)[2]	22 tropical countries	30-m	Landsat	2014	Weekly
FORest Monitoring for Action (FORMA)[3]	Pantropical	250-m	MODIS	2012	Biweekly
Terra-i[4]	Pantropical	250-m	MODIS	2004	Biweekly
JJ Fast[5]	77 tropical countries	5 ha	ALOS-2	2016	1.5 months

[1] Hansen et al. (2013) [2] Hansen et al. (2016) [3] Hammer et al. (2009) [4] Reymondin et al. (2012) [5] Watanabe et al. (2018).

Amongst the datasets introduced in the above table, it is evident that only two datasets, GFC and Terra-i, have relatively long track records of forest change covering the pantropics, since 2000 and 2004, respectively. For the study of post-loss LC/LU change trajectories, these records are well-suited to serve as target areas in studying chronological LC/LU dynamics. For instance, the GFC dataset was used to produce unbiased forest loss area estimates and the types

of forest disturbance across the Brazilian Amazon (Tyukavina *et al*., 2017). The same dataset has also been used to quantify the expansion of agriculture into forests (Pendrill and Persson, 2017; Pendrill, Persson, Godar, and Kastner, 2019) and determine associated carbon emissions globally (Pendrill, Persson, Godar, Kastner, *et al*., 2019).

2.3.2 Post-loss LC/LU change

In contrast to the increased capability to monitor and map forest change, most existing initiatives in mapping post-loss LC/LU at large scales, from national to global, has focused on the immediate (1 or 2 years) change with limited temporal scope. These initiatives can be grouped in three distinct groups using sampling-based methods (De Sy *et al*., 2015; Tyukavina *et al*., 2017; De Sy *et al*., 2019), GIS-based operations (Pendrill and Persson, 2017) or visual interpretation and processing of satellite data at very coarse scale (Curtis et al., 2018; Fagua and Ramsey, 2019). Only a few groups of scholars and practitioners, strongly supported by the capabilities of machine learning algorithms, have proposed methods to map long-term (>2 years) post-loss LC/LU change at annual (MapBiomas Project, 2019) or biennial (Almeida *et al*., 2016; Arévalo *et al*., 2019) time steps.

Table 2.2 summarises the main features of the above initiatives. In general, large blocks of 10-km x 10-km are used to inform post-loss LC/LU change at global level. Seven out of ten initiatives focus only on the immediate LC and/or LU transition after forest loss and are limited to representing a few periods of time, except Tyukavina *et al*. (2017)'s study. It is evident that the number of classes is generally reduced from regional to global extent, except for the GIS-based method. The particular example of GIS-based operations targeted proximate causes of forest loss in the form of agricultural expansion being non-agricultural classes discarded (Pendrill and Persson, 2017).

Table 2.2 Existing studies, ordered by geographical extent, supported by remote sensing data to map immediate or long-term post-loss LC/LU change from global to national level.

Reference, target and method	Extent	Number of classes	Description	Time and resolution
Curtis *et al.* (2018) Immediate Mix of machine learning and visual interpretation	Global	5 (commodity-driven deforestation, shifting cultivation, forestry, wildfire and urbanisation)	Visual interpretation of Google Earth data and prediction by a decision tree model	Single period: 2000-2015 Blocks of 10-km × 10-km
De Sy *et al.* (2019) Immediate Sample-based design	Global	9 (small-scale crop, large-scale crop, tree crop, pasture, mixed agriculture, infrastructure, other land use, water, unknown)	Visual interpretation of Landsat, Google Earth imagery and ESRI world imagery	Single period: 1990-2000 Blocks of 10-km × 10-km
De Sy *et al.* (2015) Immediate Sample-based design	Continental (South America)	15 (mixed agriculture, commercial crop, smallholder crop, tree crops, pasture, urban and settlements, roads and built-up, mining, other land, bare land, other wooded land, grass and herbaceous, wetlands, unknown)	Visual interpretation of Landsat, Google Earth imagery and ESRI world imagery over sampling units delimited by the systematic sampling design of FAO Global Forest Resources assessment	Two periods: 1990-2005 2000-2005 Blocks of 10-km x 10-km
Pendrill and Persson (2017) Immediate GIS based operations	Continental (Latin America) & Regional (Brazilian Amazon)	2 (pasture, cultivated)	Overlapping and lag time analysis from datasets of forest change, the Global Forest Change (GFC) and a land cover dataset, GlobeLand30	Single time: 2010 30-m
MapBiomas Project (2019) Long-term Machine learning	Regional (Pan-Amazon and Gran Chaco)	6 to 7 general classes (listed below). specific classes can be found in the project website	Landsat imagery using a combination of spectral mixture analysis (to map forests) and	Annual: 2001-2017 (Pan-Amazon & Gran Chaco)

	National (Brazil)	Pan-Amazon and Brazil: 6 (forest, non-forest natural formation, farming, non-vegetated area, water, non-observed) Gran Chaco: same as above plus spare vegetation	supervised Random Forest (to map other LC/LU classes)	& 1985-2017 (Brazil) 30-m
Almeida *et al.* (2016) Long-term Mix of machine learning and visual interpretation	Regional (Brazilian Amazon)	12 (deforestation, urban areas, mining, mosaic of uses, annual crops, herbaceous pasture, shrubby pasture, pasture with bare soil, regeneration with pasture, secondary vegetation, others, and non-observed areas)	Visual interpretation of Landsat satellite data in combination with MODIS phenology data and the PRODES deforestation mask	Biennial: 2008-2014 30-m
Tyukavina *et al.* (2017) Immediate Sample-based design	Regional (Brazilian Amazon)	13 (agroindustrial crops, agroindustrial trees, pasture, small-scale crops, roads, other due to human clearing, dam-related flooding, selective logging, fire, river meandering, windfalls, other due to natural disturbance)	Visual interpretation of annual Landsat composite images and Google Earth imagery via a sampling-based design (random sampling)	Annual: 2001-2013 10,000 samples (pixels of (30-m × 30-m)
Krylov *et al.* (2018) Immediate Sample-based design	Regional (Mexican Yucatán & Argentine Chaco)	10 in Yucatán (henequen, permanent corn, sugar cane, pasture, construction, temporary corn, burned, fallow, hurricane damaged, logging) 8 in Argentine Chaco (dry beans, Soy, Corn, Pasture, Construction, burned, artisanal logging, industrial logging)	Field surveys and visual interpretation of annual Landsat composite images and Google Earth imagery via sampling based design (two-stage sampling)	Period: 2000-2012 Blocks of 6 km x 6 km (11 pixels per block) in Yucatán Blocks of 20-km x 20-km (21 pixels per block) in Argentine Chaco

Fagua and Ramsey (2019) Immediate Mix of machine learning and visual interpretation	Regional (Chocó-Darien Global Ecoregion)	8 (woody vegetation, wetland, grassland, crops, palm plantations, settlements and infrastructure, continental water, bare areas)	Classification of time series of cloud-free MODIS vegetation index products using Random Forest	Annual: 2002-2015 250-m
(Arévalo *et al.*, 2019) Long-term Machine learning	Regional (Colombian Amazon)	7 (forest, secondary forest, grassland, urban, pastures/crops, water and other)	Zhu and Woodcock's (2014) algorithm for detecting change segments from Landsat images time series The segments are classified using a Random Forest model to assign LC labels	Biennial: 2001-2016 30-m

2.4 Global policy in forest cover and post-loss LC/LU change

Maximising forest area is not just a concern for conservationists and ecologists, it is widely recognised as a major issue in the climate change and development debate (Sloan and Sayer, 2015). New geospatial data and technologies are expected to be crucial for aiding a number of global policy initiatives. In particular, post-loss LC/LU change data are essential to the following mechanisms focused on the forestry-based sector and research: 1) the Intended Nationally Determined Contributions under the Paris Agreement, or INDCs state a long-term objective with voluntary mitigation actions (self-commitments) aimed at decarbonising economies (van der Gaast *et al.*, 2018); 2) the Sustainable Development Goals (SDGs) set out core intentions of a new global agenda on environment and development through a set of 17 goals (Hoen *et al.*, 2014); 3) the Reducing Emissions from Deforestation and Forest Degradation, known as REDD, and enhancing forest carbon stocks, known as REDD+, provide the incentive and framework for implementation of forest protection and conservation measures in developing countries (Angelsen and Rudel, 2013; Ryan *et al.*, 2014); and 4) the Bonn Challenge aims for the restoration of 150 million hectares of the world's deforested and degraded land by 2020, and 350 million hectares by 2030 (Lambin *et al.*, 2003). These mechanisms are strongly interlinked. For example,

REDD+ action might form part of INDCs as a potential mitigation option beside other forestry-related carbon trading schemes (van der Gaast *et al.*, 2018). Moreover, REDD+ can be a useful means to deliver on the SDGs in developing countries rather than exclusively in achieving sustainable forests (SDG 13 and 15), but also on other related goals such as the promotion of sustainable agriculture (SDG 2) and poverty eradication (SDG 1). The SDGs, in turn, act as an additional institutional incentive to pursue effective REDD+ actions (Lima *et al.*, 2015). The Bonn Challenge is a practical means of realizing some of these mechanisms like REDD+ and others the Convention on Biological Diversity (CBD) Aichi Target 15 and the Rio+20 land degradation neutrality goal (Lambin *et al.*, 2003).

Complementary to the above forested-related mechanisms, the generation and analysis of post-loss LC/LU change data is beneficial to current initiatives of deforestation-free supply chains. For instance, Costa Rica's Monitoring of Land Use Change within Production Landscapes (MOCUPP) initiative (MOCUPP, 2017) links land tenure information with forest coverage to monitor different commodities, mainly pineapple. Hargita *et al.* (2020) recently compares similarities and differences of deforestation-free supply chains initiatives against REDD+. The authors stress the potential of a common monitoring system. In this regard, new geospatial data such as post-loss LC/LU change trajectories can play a pivotal role to guide these initiatives and others like the Bonn Challenge to ensure a long-term effect on forest conservation.

2.5 Main gaps and opportunities in the study of post-loss LC/LU change

The above review allowed identifying three main knowledge gaps and opportunities related to characterising and modelling post-loss LC/LU change:

- The theory and definitions behind forest change are well established in the literature. Historical dynamics and determinants related to deforestation have been widely characterised in the pantropics. Further investigation of post-loss LC/LU change is needed to complement these theoretical assumptions. For instance, it would be interesting to study post-loss LC/LU change trajectories according to the state-enabled, smallholder deforestation (1965–1985) and enterprise-driven deforestation (1985–

Present) periods. Furthermore, it would be relevant to assess how determinants and location factors of deforestation described in **Section 2.2.4** evolve over long-term post-loss LC/LU change.

- According to the section of methodological considerations, it was found the pathways of expansion of the current agricultural and non-agricultural drivers of global forest change are mainly based on information aggregated in coarse spatial mapping units e.g. country-scale or blocks of 10km per 10km. In addition, most of the existing studies at large scales, from national to global, focus on the immediate conversion after forest loss with limited temporal extent. In order to approximate the complexity of LC/LU dynamics in deforested landscapes, hidden patterns such as forest converted temporarily to pasture and time after commercial agriculture might only be unmasked with a dedicated spatially explicit post-loss LC/LU change data generated over a considerable time period (>2 years).

- The initiatives mapping long-term post-loss LC/LU change rely on traditional machine learning algorithms e.g. Random Forest with low capability to fully exploit the spatial and temporal contextual information from earth observation data (Petitjean, Kurtz, *et al.*, 2012; Rußwurm and Körner, 2018b). In addition, the performance of these algorithms generally depends on time-consuming pre-processing e.g. cloud removal and human-designed rules to train them. In this regard, novel machine learning algorithms like Deep Neural Networks represent an opportunity of end-to-end mapping of large-area LC/LU change from dense satellite image time series with minimal pre-processing (Rußwurm and Körner, 2018b). As it is demonstrated in **Chapter 4 to 5**, the implementation of these novel techniques requires a rigorous assessment to ensure spatio-temporal dimension are captured consistently for the target application, in this case mapping long-term post-loss LC/LU change.

3 Using ground reference and very high-resolution imagery to better understand the challenges to mapping post-loss LC/LU change trajectories

3.1 Introduction

This chapter aims to describe the context, opportunities and limitations for mapping long-term post-loss LC/LU change. The analysis and results are derived from multiple sources of data, including *in-situ* records in the field and multi-temporal satellite data. The *in-situ* records analysed are primarily based on previous work by Coca-Castro *et al.* (2014). This particular work contains a protocol to validate deforestation areas located in the eastern Peruvian Amazon (elaborated on below). More recently, the protocol has been widely applied on additional field campaigns conducted in the northern Peruvian Amazon (Paz, Reymondin, *et al.*, 2018a, 2018b) and following these, in Honduras (Paz, Tello, *et al.*, 2018).

Inspired by existing tools for visual interpretation of LC/LU sample areas – e.g. Geo-wiki (Fritz *et al.*, 2009), Collect Earth (Bey, Sánchez-Paus Díaz, *et al.*, 2016) and LACO-Wiki (See *et al.*, 2017) – this work presents a customised workflow for integrating multi-temporal satellite data obtained from different sources and *in-situ* information. This workflow ensures the dynamics of subsequent LC/LU change over post-deforested areas (so called post-loss LC/LU change) are captured.

3.2 Background

The release of new LC/LU products is directly related to the availability of remotely-sensed imagery acquired by satellites (Bontemps *et al.*, 2012). This development has consequently led to the collection and sharing of reference databases by the Earth Observation (EO) user community. These databases are crucial to systematically assess and benchmark existing and future products, particularly given the uncertainty in land-related products (Waldner *et al.*, 2019).

The collection of LC/LU reference data is commonly derived from either: i) ground-based data in the field; ii) geo-referenced airborne data (Amaral and D'Alge, 2009), and/or; iii) interpretation of free satellite imagery (Lesiv *et al*., 2018) or geo-tagged photographs (Antoniou *et al*., 2016). Although the two first approaches provide a high level of detail, they remain challenging due to the cost associated with such collection efforts, particularly over large areas (Bastin *et al*., 2013; Amaral and D'Alge, 2009). The latter approach is therefore more accessible and includes the engagement of not only experts but also citizens in remote sensing or geospatial sciences (Waldner *et al*., 2019).

Thanks to the rise of new technologies and the freely available QuickBird, WorldView, GeoEye and other Very High Resolution (VHR) imagery hosted by Google Earth, ESRI and Bing Maps, initiatives for visual interpretation of imagery for understanding LC/LU in sampled areas such as Geo-wiki (Fritz *et al*., 2009), Collect Earth (Bey, Sánchez-Paus Díaz, *et al*., 2016) and LACO-Wiki (See *et al*., 2017) have emerged. Besides providing standards for collecting and analysing LC/LU reference data, these tools leverage the global remote sensing expert community for regional knowledge creation thereby contributing to the technical capacity of institutions responsible for LC/LU change reporting (Tyukavina *et al*., 2018; Schepaschenko *et al*., 2019).

Table 3.1 outlines the main features of the two main open tools, Geo-wiki and Collect Earth, which share some similarities but also differences according their system components and design i.e. architecture. For instance, whilst Collect Earth allows to customise the interface, this feature is not possible in Geo-wiki.

Whilst the databases collected by the aforementioned initiatives contribute to reducing the gap of reference data, the time-span analysed by most of these efforts is necessarily limited and specific to the target map validated (Tsendbazar *et al*., 2018). These limitations make it difficult to re-use this type of information for certain applications, in particular those requiring periodic observations such as long-term post-loss LC/LU change. Therefore, the analysis of *in-situ* data collected by Coca-Castro *et al*. (2014) and Paz, Reymondin (*et al*., 2018b) was extended using continuous satellite imagery obtained from multiple sources

towards a better understanding of the precedent and subsequent LC/LU change over deforested areas.

Table 3.1 A comparison of the two main open tools for visual interpretation of VHR imagery. Adapted from Schepaschenko *et al.* (2019)

Feature	Collect Earth	Geo-Wiki
High-resolution imagery	Direct access to current and historical imagery such as Landsat is available in Google Earth Engine (GEE)	Current and historical fine-resolution imagery is available through Sentinel Hub as a Web Map Service (WMS)
Imagery dates	The dates of the Google VHR images are recorded manually. Microsoft Bing Map dates are read automatically via an API.	
Ancillary layers	Only those layers that are available in GEE, but users can upload their own layers	Any ancillary layer can be displayed as a WMS.
Sample design	Users cannot generate a sample in Collect Earth, but scripts are provided on the Collect Earth website to generate a country-level sampling grid through the GEE code editor	A simple random sample can be created, but otherwise the sample is created using GIS and then hardcoded into the Geo-Wiki branch by developers
Level of customization	Users can customize the interface for any specific data collection needs, including different data types and levels of spatial detail	Customization is currently not possible. Branches are created ad hoc to meet user needs
Code availability	Open access on GitHub	Not available
Data access	The data collected using Google Earth belong to the group or persons who initiated the data collection campaign. Only they can decide if they share these data with others	The data can be downloaded from the Geo-Wiki website or from the PANGAEA repository

3.3 Study area and context

The *in-situ* observations and auxiliary datasets analysed through this work are located in the Peruvian Amazon, specifically in the regions of Ucayali and Loreto (see **Figure 3.1**). These observations were collected as part of a series of field campaigns led by the Terra-i initiative (described in **Section 3.4.1.2**) across the Peruvian Amazon which is the second-largest portion of Amazon forest after Brazil (RAISG, 2012). The records in Ucayali were collected in October 2013 focusing on areas surrounding the city of Pucallpa. The field campaign in Loreto

was conducted in June 2015 across locations near the city of Yurimaguas. Both Ucayali and Loreto have received considerable impact due to anthropogenic-driven disturbances (Oliveira *et al.*, 2007). The two regions alongside Madre de Dios include the majority of primary forest remaining in the country (Piu and Menton, 2014).

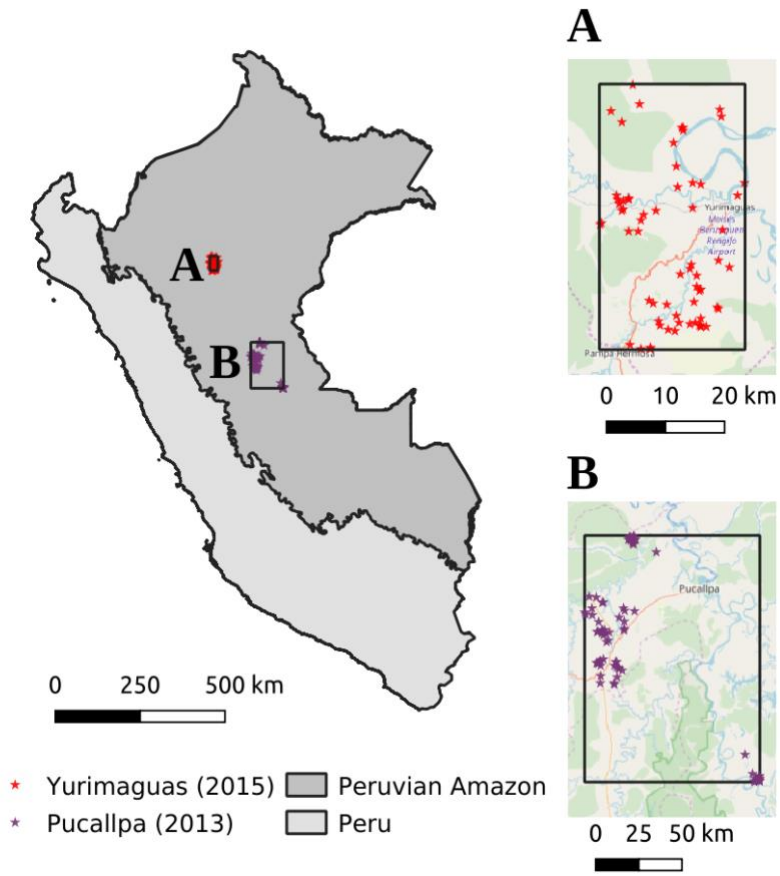


Figure 3.1 Location of the *in-situ* observations across the study areas of Ucayali (map inset A) and Loreto (map inset B) in the Peruvian Amazon. The average annual precipitation and temperature in both areas range between 1,800 mm to 2,000 mm per year and 25.2°C to 26°C, respectively with a dry season from June to September (Kobayashi *et al.*, 2014).

Ucayali is experiencing a dramatic increase in area of total land harvested for agriculture (Oliveira *et al.*, 2007). Previous studies in the same location define the landscape as a mosaic of forest (patches of old-growth and second-growth forest) surrounded by pastures, oil palm plantations, and smallholder farms (Gutiérrez-Vélez and DeFries, 2013; Schwartz *et al.*, 2017). This landscape has been shaped by a highway and networks of roads connecting Pucallpa to Lima and other urban centers for more than six decades. Thanks to the accessibility

afforded by both road and river, the area has attracted many migrants from other regions of Peru (Uriarte *et al*., 2012).

Loreto, where *in-situ* data are distributed principally in areas surrounding the city of Yurimaguas, has also experienced similar high rates of deforestation as Ucayali. As with the study area in Pucallpa, accessibility has been a key factor shaping deforestation patterns in the Yurimaguas area, including also industrial-scale oil palm plantations at the border region between Loreto and San Martin (Valqui *et al*., 2015; Arima, 2016). The main economic activities are cropland, livestock and forestry production. Although these activities are primarily for subsistence, some cash crops such as cocoa, oil palm and papaya are expanding in the area (Valqui *et al*., 2015). In contrast to the city of Pucallpa, population growth and in-migration has been slower and stable in Yurimaguas. While significant migration flows in the direction of the Lower Amazon i.e. Ucayali, Loreto has experienced a higher rate of out-migration than in-migration causing a net loss in population (Menton and Cronkleton, 2019).

3.4 Methodology

The workflow to analyse the *in-situ* observation related to deforested areas is depicted in **Figure 3.2**. Each area was analysed using a group of variables as described in **Table 3.2**. These variables were interpreted according to the grain of the reference deforestation dataset and including a square buffer of 500-m. This buffer allows the observation of variables analysed of neighbouring areas beyond the pixel scale, to capture neighbourhood effects for contextual variables. A two-stage analysis was conducted using the input data described in **Section 3.4.1**. Stage 1 referred to a characterisation of the target areas at the time of field collection. In stage 2, a multi-temporal analysis from 2004 to 2018 was carried out to capture the LC/LU change dynamics over time.

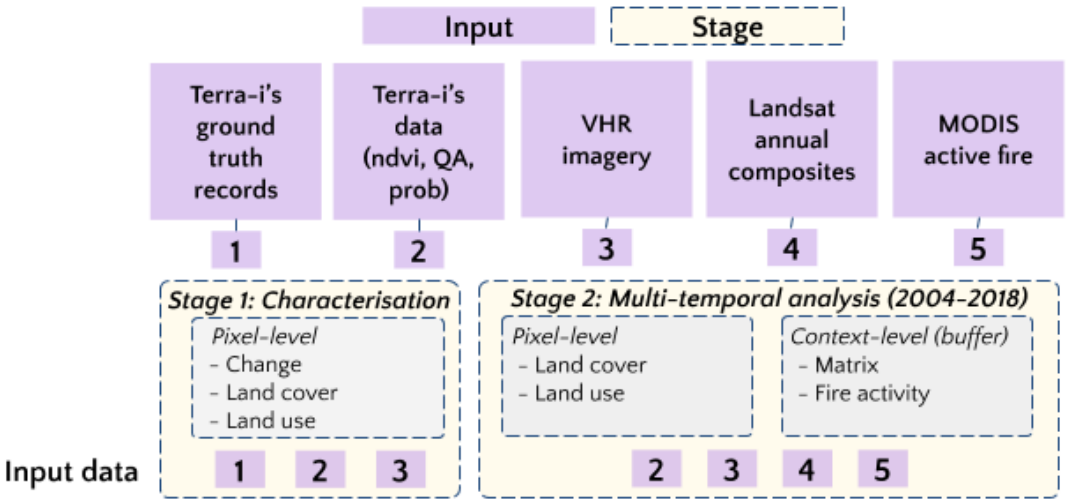


Figure 3.2 Input data and components analysed over the deforested areas.

Table 3.2 Variable and range of values interpreted by area analysed.

Variable	Classes	Values
Change	3	1) false positive, 2) true positive, 3) early/late detection
Land cover	10	1) tree cover, 2) shrub cover, 3) herbaceous vegetation/grassland, 4) cultivated and managed, 5) mosaic of cultivated and managed/natural vegetation, 6) flooded/wetland, 7) urban, 8) snow and ice, 9) barren and 10) open water
Land use	7	1) forest 2) cropland (tree crops, semi-permanent, seasonal, unknown) 3) grassland (unmanaged, managed) 4) wetland, 5) settlement 6) water, 7) Otherland (road, barren)
Matrix	4	1) >50% forest and natural, 2) mosaic intact forest-human disturbance, 3) >50% human disturbance, 4) other
Fire activity	2	1) null 2) active (more than a single fire activity)

3.4.1 Input data

3.4.1.1 Deforestation dataset

From the deforestation datasets introduced in **Chapter 2** (see **Section 2.3.1**), Terra-i was used as reference. Terra-i is based on MODIS MOD13Q1 satellite data product with a 250-m spatial resolution and a 16-day temporal resolution (Reymondin *et al.*, 2012). The current system is capable of monitoring forest loss at the pantropical scale. The tool's consistency has been tested by ground-truthing field campaigns in certain countries. These campaigns have also collected information about the approximate drivers of forest loss.

In comparison to the GFC dataset (Hansen *et al*., 2013), which contains a larger database of annual forest change events (since 2000), the Terra-i has a lower maximum lag between the real change and the system report. While the GFC dataset is updated every 2 years, the Terra-i system is updated within 2 months. Moreover, the Terra-i dataset retains a greater temporal resolution (with an image every 16-days) allowing a better approximation of the day of the year when a given disturbance occurs.

In addition to utilising Terra-i deforestation alerts, this work also incorporated some intermediate products of the Terra-i model (see **Table 3.3**). The analysis of these products enhanced the understanding of how this particular model serves in flagging disturbance events.

Table 3.3 Summary of the intermediate products analysed over the Terra-i alerts studied. The access to these products was granted by the Terra-i project.

Product	Description
Raw NDVI	Time series of NDVI extracted from MOD13Q1
Quality (QA)	Quality flag (bad or good observations) as defined by the Terra-i
Smoothed NDVI	Raw NDVI time series are reconstructed by using the Harmonic Analysis of NDVI Time Series (HANTS)[1] algorithm and QA observations
Probability	A threshold value is set to define the probability, from 0 to 1, of change

[1] *Roerink et al. (2000)*

3.4.1.2 Ground-truthing field campaigns

From the database of field campaigns accessed through the Terra-i project, two field campaigns conducted in the Peruvian Amazon (Coca-Castro *et al*., 2014; Paz, Reymondin, *et al*., 2018b), were analysed in this work. These study areas are the focus of the exploratory data analysis of this chapter as described in **Figure 3.2**.

In total, some 120 areas were used for the collection of *in-situ* data (photographs, annotation to determine land cover and land use) by the Terra-i team (see **Table 3.4**). The ground inspection consisted of the verification of 9 subsample areas distributed over each validated pixel (see **Figure 3.3**). This distribution was selected according to a technical guide for Land Use, Land-Use Change and Forestry studies issued by the Intergovernmental Panel on Climate Change

(IPCC) (Penman *et al*., 2003). The selection of the validated areas did not follow a strict statistical sampling design as proposed by Olofsson *et al*. (2013). Instead, a naive approach was followed according to the road conditions and logistical constraints as indicated by Hyman and Barona (2010).

Table 3.4 Description of the field campaigns conducted in the Peruvian Amazon.

Field campaign	Method	Number of pixels by year validated	Version of the Terra-i dataset validated
Ucayali, Peru	Ground inspection	22 (2011) 35 (2012)	2004_01_01_2013_04_07
Loreto, Peru	Ground inspection	32 (2013) 27 (2014) 4 (2015)	2004_01_01_2015_04_07

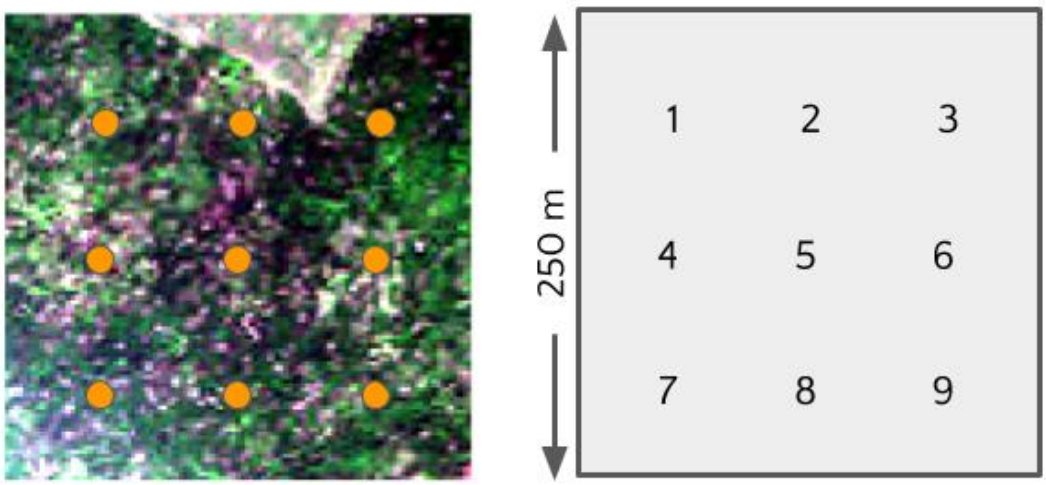


Figure 3.3 Sampling scheme conducted by field campaigns to inspect 250-m MODIS disturbed areas as detected by Terra-i across the study areas of Ucayali and Loreto in the Peruvian Amazon. Left, very high-resolution satellite image obtained from DigitalGlobe's GBDX. Right, scheme with the distribution of the sub-sample areas.

3.4.1.3 Auxiliary satellite datasets

Very High Resolution (≤ 1-m from DigitalGlobe), moderate (30-m from Landsat) and coarse (250-m from MODIS) resolution satellite images were inspected across the target deforested areas. While the first two provide sufficiently high resolution to visualise LC/LU change over time, the MODIS product used – active fire data from the MODIS's fire information for resource management system (FIRMS) – is sufficiently sensitive to determine if any fire activity might be associated with an observed forest disturbance.

For this study, Very-High-Resolution (VHR) data were provided by DigitalGlobe's GBDX platform through the International Center for Tropical Agriculture (CIAT). Landsat and MODIS data were accessed and processed via Google Earth Engine (GEE) (Gorelick *et al.*, 2017).

Very-High-Resolution Satellite Imagery

VHR data have been a fundamental source of calibration and validation for remote sensing studies and products. The spatial resolution of these images is high enough to clearly distinguish types and related spatial patterns of LC/LU. According to DigitalGlobe's GBDX catalogue, only multi-spectral images were considered for this work. The set of images explored includes Quickbird, Geoeye-1, Worldview-1, Worldview-2 and Worldview-3 (see **Table 3.5**). For those images with multispectral information, the Normalized Difference Vegetation Index (NDVI) was computed and visualised accordingly using the Visible (RGB) and Near Infrared (NIR) bands of the electromagnetic spectrum.

Table 3.5 Main features of the VHR images inspected. Source: GBDX (2019).

Satellite	Data Range	Spectral Bands
Quickbird (QB02)	2002-2014	Panchromatic Multispectral 4-bands
Geoeye-1 (GE01)	2008 - present	Panchromatic and Multispectral 4-bands
Worldview-1 (WV01)	Sept 2007 - present	Panchromatic
Worldview-2 (WV02)	Oct 2009 - present	Panchromatic Multispectral 8-bands
Worldview-3 (WV03)	Oct 2014 - present	Panchromatic Multispectral 8-bands SWIR 8-bands

The imagery available in Google Earth and Microsoft Bing Maps held mosaics of multiple images from different time periods, different spatial resolutions and various image providers. Whilst it is true certain areas around the globe have a good spatial and temporal coverage by both providers, the access to VHR via GBDX platform has guaranteed the inspection of more than a single image per year. Also, the catalogue of VHR facilitated the verification of LC/LU change from

2002 which closely coincides with Terra-i's first year of detection (2004 onwards).

Landsat

Multi-annual composites of Landsat 7 ETM+ and Landsat 8 observations were created from 2001 to 2018. Composites are derived using the best available pixel observation (from the target period e.g. year) for any given pixel location. These composites can be produced generally using predefined rules according to the information needed (White *et al*., 2014). To facilitate the characterisation of the deforested areas, the composites were pansharpened using the panchromatic band. Pansharpening was key for improving the visual clarity of the composites for interpretation. This process should be used cautiously when spectral analysis and derived analysis are conducted (Pradhan, 2005) however this was not the purpose of this work.

The generation of aforementioned composites were carried out by using a GEE built-algorithm, *ee.Algorithms.Landsat.simpleComposite*. This method is divided into three steps (i) computation of a Top of the Atmosphere (TOA) composite from the input collection and applies standard TOA calibration; (ii) assignation of a cloud score by pixel using the *SimpleLandsatCloudScore* algorithm. The lowest possible range of cloud scores at each pixel is selected and then per-band percentile values for the accepted pixels are computed; (iii) use of the *LandsatPathRowLimit* algorithm to select only scenes with minimal cloud cover in regions (Gorelick *et al*., 2017).

MODIS Active Fire Data

The MODIS active fire product, FIRMS, is generated using the standard MODIS MOD14/MYD14 Fire and Thermal Anomalies product. Each active fire location represents the centroid of a 1-km pixel flagged as containing one or more active fires at the time of the satellite overpass. A confidence value, ranging from 0% (lowest confidence) to 100% (highest confidence), can be accessed by active fire location (C.O. Justice *et al*., 2002).

In GEE, the FIRMS dataset, which is originally provided as vector file, is rasterized. This procedure is conducted as follows: (i) for each fire point, a 1-km bounding

box (BB) is defined, (ii) pixels in the MODIS sinusoidal projection that intersect the FIRMS BB are identified, (iii) if multiple FIRMS BBs intersect the same pixel, the one with higher confidence is retained; in the case of equivalence, the brighter pixel is retained as most likely representing a genuine and also the largest fire event. For this work, all available active fires were included without filtering by value of confidence. Other scholars have included MODIS pixels irrespective of the level of confidence, in particular for landscapes where small size fires occur (Tansey *et al.*, 2008; Fornacca *et al.*, 2017) as is common in the study areas (Schwartz *et al.*, 2015). The count of daily active fires was summed and retained to a 16-day period according to MOD13Q1 dataset. This pre-processing ensured the analysis of each area according to the timestep used by the Terra-i model.

3.4.2 Variables

3.4.2.1 Change

For this work, the term *change*, particularly *forest loss*, represents the replacement of pristine and/or secondary tree cover by anthropogenic land uses (e.g. farmlands, mining and/or human settlements) or by natural events (e.g. flooding, wildfires). This clarification is relevant as some of Terra-i's detections also include abrupt disturbances in other natural/anthropogenic land surfaces (e.g. burning pasture, changes in tree plantations, among others).

According to the terminology aforementioned, three types of *change* were assigned to each of Terra-i's *in-situ* reference sites:

- *False positive*: tree cover remained unchanged or a disturbance existed but over a non-forest/woody cover (e.g. pasture burning).
- *True positive*: a change occurred and matched with the detection date.
- *Early/late detection*: change started before/after the detection date. A tolerance of 2 years before/ahead the detection date was considered to flag a given area under this category.

3.4.2.2 LC/LU classification

Predefined LC/LU classification schemes were used to characterise both field and satellite data at the date of collection and remaining years (from 2004 to 2018). These schemes were chosen according to existing standards by web-

based crowdsourced data collection platforms such as Geo-wiki (Fritz *et al*., 2009) and Collect Earth (Bey, Sánchez-Paus Díaz, *et al*., 2016).

The reference LC scheme includes a set of 10 simple classes 1) *tree cover*, 2) *shrub cover*, 3) *herbaceous vegetation/grassland*, 4) *cultivated and managed*, 5) *mosaic of cultivated and managed/natural vegetation*, 6) *flooded/wetland*, 7) *urban*, 8) *snow and ice*, 9) *barren* and 10) *open water*. These LC types, used as standard by the Geo-wiki initiative (Fritz *et al*., 2013), are a simplified version of IGBP land cover classification (Loveland *et al*., 2000). The translation between IGBP's classification scheme and Geo-Wiki's classification is depicted in **Table 3.6**.

Table 3.6 Relationships between the classification scheme of the IGBP and Geo-wiki's generalised land cover classification. Source: Fritz *et al*. (2013).

IGBP class(es)	Geo-wiki scheme
Evergreen needleleaf forest Evergreen broadleaf forest Deciduous needleleaf forest Deciduous broadleaf forest Mixed forest Woody savannas Savannas	Tree cover
Closed shrublands Open shrublands	Shrub cover
Grasslands	Herbaceous vegetation / Grassland
Croplands	Cultivated and managed (CM)
Cropland/natural vegetation mosaic	Mosaic of CM / natural vegetation
Permanent wetlands	Flooded / wetland
Urban and built-up	Urban
Snow and ice	Snow and ice
Barren or sparsely vegetated	Barren
Water bodies	Open water

To aid the visual interpretation of LC classes across the target deforested areas, the global LC crowdsourced dataset authored by Fritz *et al*. (2017) was inspected in GEE. A subset of this dataset, which uses Geo-wiki's LC scheme, was created according to the extent of the Amazon as defined by the Amazon Geo-Referenced Socio-Environmental Information Network (RAISG) corresponding to socio-ecological criteria including watershed, political and ecological boundaries (see **Figure 3.4**).

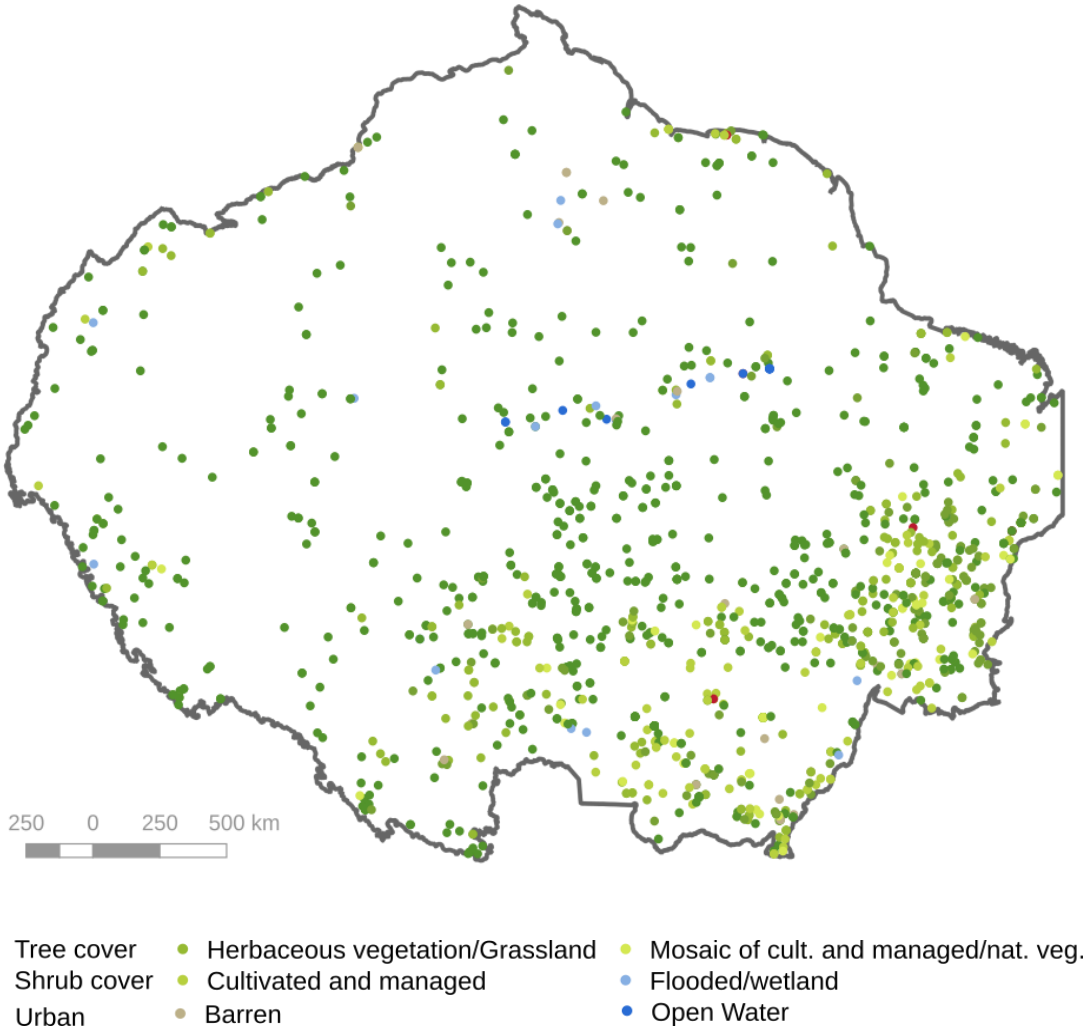
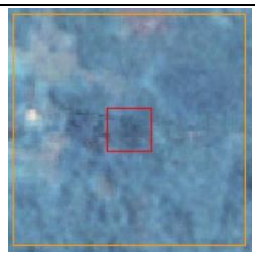
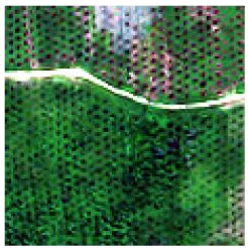
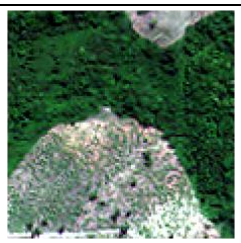
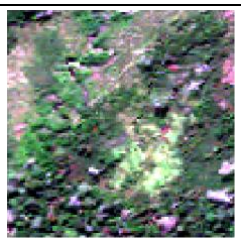


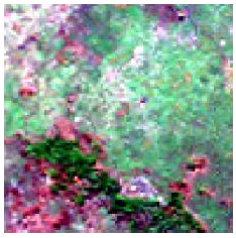
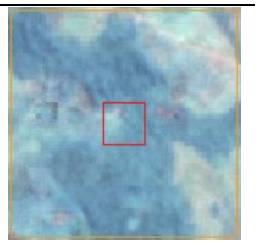
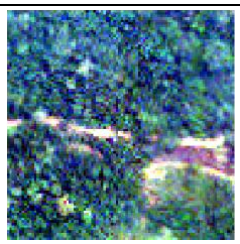
Figure 3.4 Distribution of locations inspected from the global dataset of crowdsourced land cover provided by Fritz *et al.* (2017). The extent of the Amazon is delineated by the black line as defined by the Amazon Geo-Referenced Socio-Environmental Information Network (RAISG, 2012).

Whilst LC describes the features characterising land surface analysed, LU refers to the purposes for which humans exploit the land cover (Lambin *et al.*, 2000). For the aims of this work, the IPCC land-use classification was also extended and adapted. This classification allowed for the description of more detailed information of land activities over the target deforested areas, in particular those related with cropland and grassland categories (see

Table 3.7). This LU classification was relevant due to Geo-wiki's LC scheme containing classes where natural and anthropogenic land surfaces overlap. For instance, the *tree cover* class can include natural forest and tree cover plantations (woody plantations, oil palm, fruit trees, among others) (Fritz *et al.*, 2013).

Table 3.7 Major land use types at different sources. A representation of the size of the MODIS 250-m in Landsat imagery is provided by a red square box. * *taken by Terra-i team*.

Land use	Field photos*	VHR image	Landsat
1 Forest			
2A Cropland Tree crops (oil palm, cocoa, woody plantations)			
2B Cropland Semi-permanent crops (rice)			
2C Cropland Seasonal crops (papaya, plantain, cassava, fishponds)			
2D Cropland Unknown (soil preparation, fallow)			
3A Pasture Unmanaged			

3B Pasture Managed (active land management i.e. burning)			
4 Wetland			
5 Settlement			
6 Water			
7 Otherland (road, barren)			

3.4.2.3 Intact/disturbed forest matrix

For the multi-temporal analysis, a description of the intact/disturbed forest matrix using a 500-m buffer area by year was conducted. Five classes were established accordingly; 1) *>50% forest and natural*, 2) *mosaic intact forest-human disturbance*, 3) *>50% human disturbance*, 4) *Other* and 5) *NoData*.

3.4.2.4 Fire activity

In addition to the analysis of the landscape matrix, the fire activity per year was also estimated. The daily fire activity data were aggregated by year to produce annual estimates of fire active pixels.

3.4.3 Software and implementation

The variables mentioned in **Section 3.4.2** were recorded by field campaign separately into Google spreadsheets. These spreadsheets were created using a customised Google form to facilitate the selection of a single option/response by variable. The outputs were then plotted in bar charts using *seaborn*, a python-based library.

R software v.3.4.3. was used for identifying trajectories of LU change. The observed annual LU over the target sites were analysed using the *lucCalculus* v.1.0 R package (Maciel *et al*., 2019). Based on Allen's interval logic (Allen, 1984), this package provides a set of functions to express cases of recurrence, conversion and evolution in LU change, e.g. forest regrowth, land abandonment, agriculture intensification. For the aim of this chapter, the tool was used to identify the most common land conversions from forest to other LU types. To introduce the logic behind the *lucCalculus* tool, **Table 3.8** shows an example of an expression accepted to find the locations i.e. pixels that until 2003 were forest areas that did not suffer any type of degradation, but from 2004 have been turned into other land uses, such as pasture or single cropping. The functions HOLDS and FOLLOWS allow the tool to select the target LC classes and indicate the transition of land change to pasture or single cropping can occur any time within the interval 2004 to 2016. **Figure 3.5** shows the locations where the expression in **Table 3.8** is evaluated as true.

Table 3.8 Example of an expression computed in *lucCalculus* to select locations i.e. pixels with a specific land use transition, in this case conversion of forest after 2003. Source: Maciel (2017).

Search for all forest areas that have been replaced by pasture or single cropping after the year 2003.
$\forall l \in \mathcal{L}, \forall t_{i,} t_{j,} \in T, HOLDS(l, "Forest", t_{i,}) \wedge (HOLDS(l, "Pasture", t_{j,}) \vee (HOLDS(l, \text{Single\_cropping}, t_{j,}) \wedge FOLLOWS(t_{i,} t_{j,}), where\ t_{i,} = [2001, 2003], t_j = [2004, 2016]$

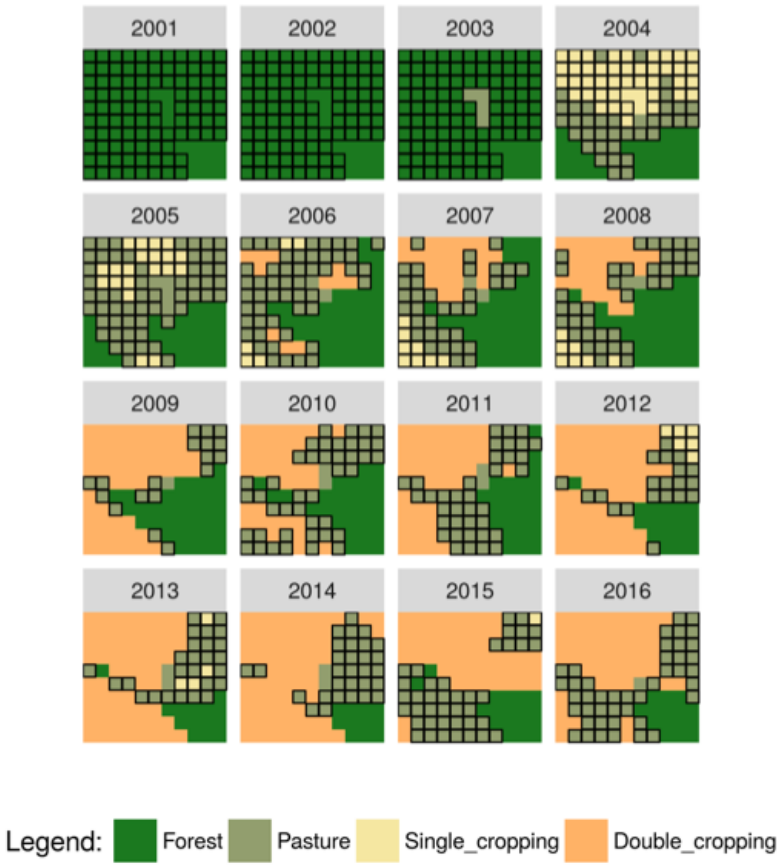


Figure 3.5 Spatial distribution of 2001-2016 land use types across a small area tested by Maciel (2017). Locations i.e. pixels where forest turned into pasture or single cropping and queried using the expression of **Table 3.8** are highlighted in black. Source: Maciel (2017).

The above target locations i.e. pixels, highlighted in black in **Figure 3.5**, can be further visualised as a sequence plot (see **Figure 3.6**). This particular plot shows transitions are not sequential through time as most of locations lack of continuous coloured lines. Moreover, the plot highlights different LU pathways across the area of interest. For instance, some locations used for pasture are turned into single cropping and later become pasture again. In addition to the sequence plot, the *lucCalculus* tool also offers the generation of bar plots to facilitate the identification of the direct conversions from the target LC cover over time. For the purposes of this chapter, the logic of the tool and corresponding plots as described in the previous examples were used to analyse the observed annual LU data over the target sites derived from **Section 3.4.2.2**.

Finally, QGIS version 3 was primarily used for vector and raster operations, i.e. preparing records in a readable format for the tools used such as GEE, GBDX tool and lucCalculus as well as producing cartographical outputs.

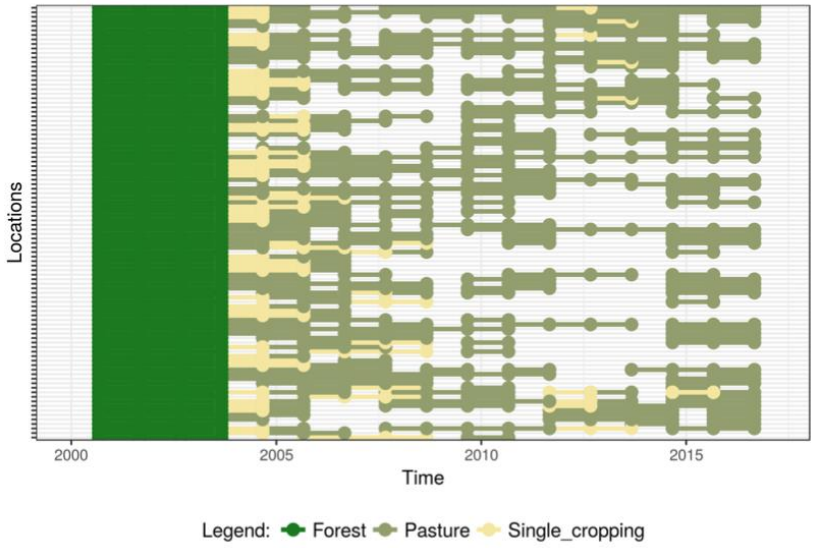


Figure 3.6 Sequence of land use transitions along the years. Source: Maciel (2017).

3.5 Results

3.5.1 Characterisation of the field records

Table 3.9 shows the results of verifying the change occurrence across the sites inspected. From the total areas analysed (N=120), 63% of the field records presented disturbance within the period during which Terra-i detected disturbance (true positives), with a further 25% early or late detections. A low proportion (12%) includes false positives or areas which remained unchanged.

Table 3.9 Proportions of the categories of change and non-forest estimation by field campaign and total. The standard deviation in proportion of non-forest is denoted by ±.

Field campaign	Proportion (%) by change type			Proportion (%) of non-forest
	False positives	True positives	Early/late detection	
Ucayali, PER (N=57)	16	63	25	77 ± 22
Loreto, PER (N=63)	8	56	28	66 ± 33
Total (N=120)	12	63	25	66 ± 33

For the characterisation of LC (see **Figure 3.7**), the areas inspected were mostly related with the class Cultivated and managed vegetation followed by Mosaic of cultivated and managed/natural vegetation. While the proportion of both classes was higher for the sites in Loreto, some sites in Ucayali presented a considerable proportion of Flooded/wetland and Open water surface.

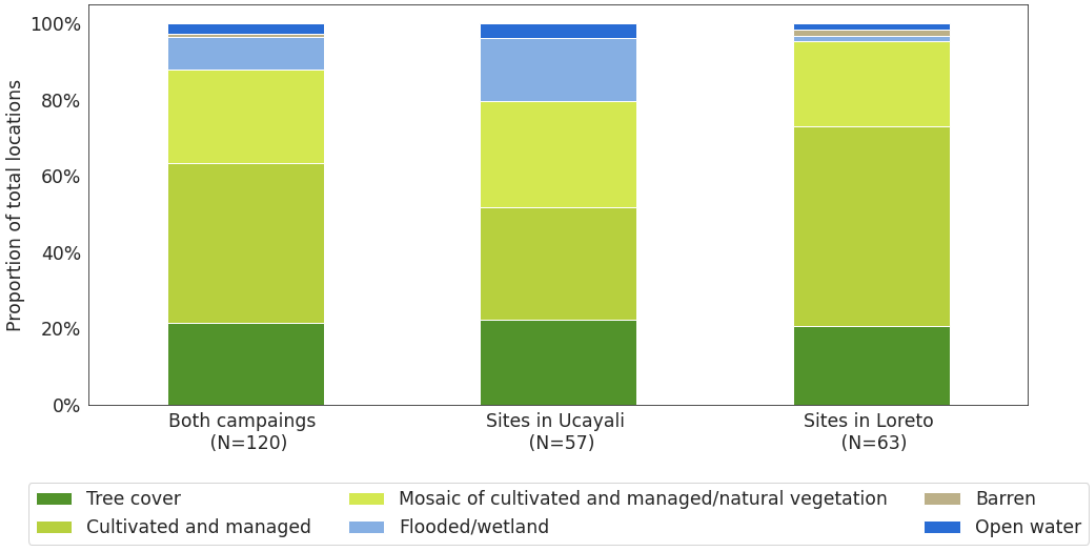


Figure 3.7 Distribution of land cover classes classified at the time of field collection in both campaigns and by single campaign.

The interpretation of LU types revealed some overlapping of disturbed and natural coverages within certain observed LC types (see **Figure 3.8**). For both campaigns, 80% of the total samples classified as *Tree cover* surface were overlapped by *Cropland* land use. While the *Grassland* and *Cropland* land use categories are equally distributed within the *Mosaic of cultivated and managed/natural vegetation* land cover class, *Cropland* was considerably more dominant within the *Cultivated and managed*. For the *Flooded/wetland* and *water* classes, they were overlapped by the *Wetland* land use.

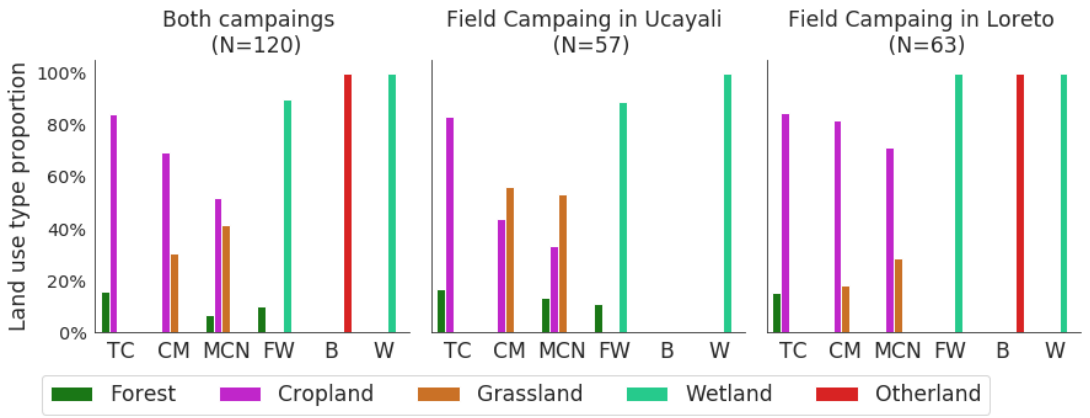


Figure 3.8 Distribution of land use types by observed land cover at the time of field collection in both campaigns and by single campaign. X-axis labels represent land cover classes Tree cover (TC), Cultivated and managed (CM), Mosaic of cultivated and managed/natural vegetation (MCN), Flooded/wetland (FW), Barren (B), Open Water (W).

3.5.2 Land cover change trajectories

The changes in LC over time varied by field campaign. For the sites in Ucayali, there was a slight increase of the *Cultivated and managed*, particularly at the cost of the *Tree cover* and *Mosaic of cultivated and managed/natural vegetation* (see **Figure 3.9**). The areas with *Flooded/wetland* remained constant over time and some percentage was likely replaced by *Open water*. For the sites in Loreto, the expansion of the *Cultivated and managed* was more pronounced than Ucayali. This trend was remarkably evident from the years of Terra-i alerts inspected (2013 to 2015), when this class represented over 60% of the samples.

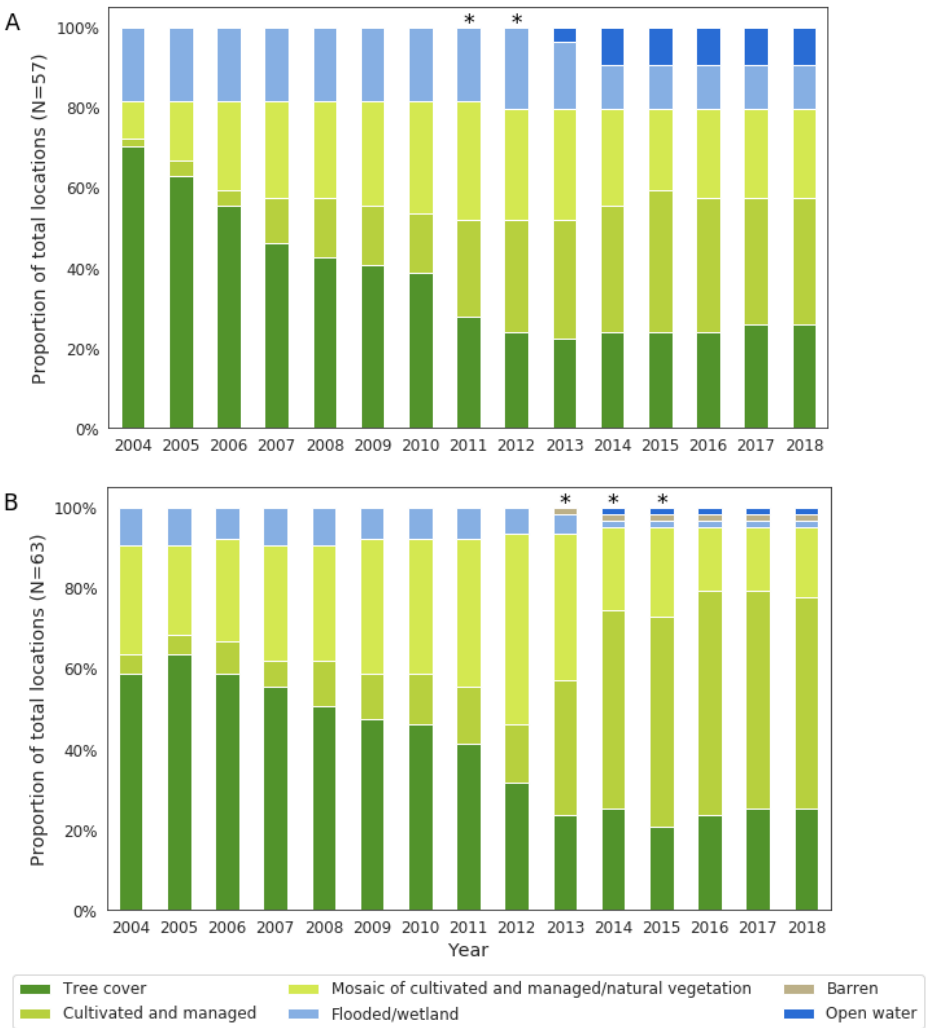


Figure 3.9 2004-2018 annual land cover change characterised across the sites of the field campaign in Ucayali (A) and Loreto (B), Peru. For the sites in Ucayali, there was a slight increase of the *Cultivated and managed* class, particularly at the cost of the *Tree cover* and *Mosaic of cultivated and managed/natural vegetation*. For the sites in Loreto, the expansion of the *Cultivated and managed* class was more pronounced than Ucayali. Asterisks (*) located above bars denote years of Terra-i's detection over the samples analysed by study area.

3.5.3 Intact/disturbed forest matrix and fire activity

The analysis of the intact/disturbed forest matrix showed forest-intact landscapes are reduced over time (see **Figure 3.10**). This reduction is evident until the years reported as disturbed by Terra-i, 2011-2012 and 2013-2015 for Ucayali and Loreto, respectively. After these dates, the landscapes had a gradual conversion to areas of greater than 50% disturbed.

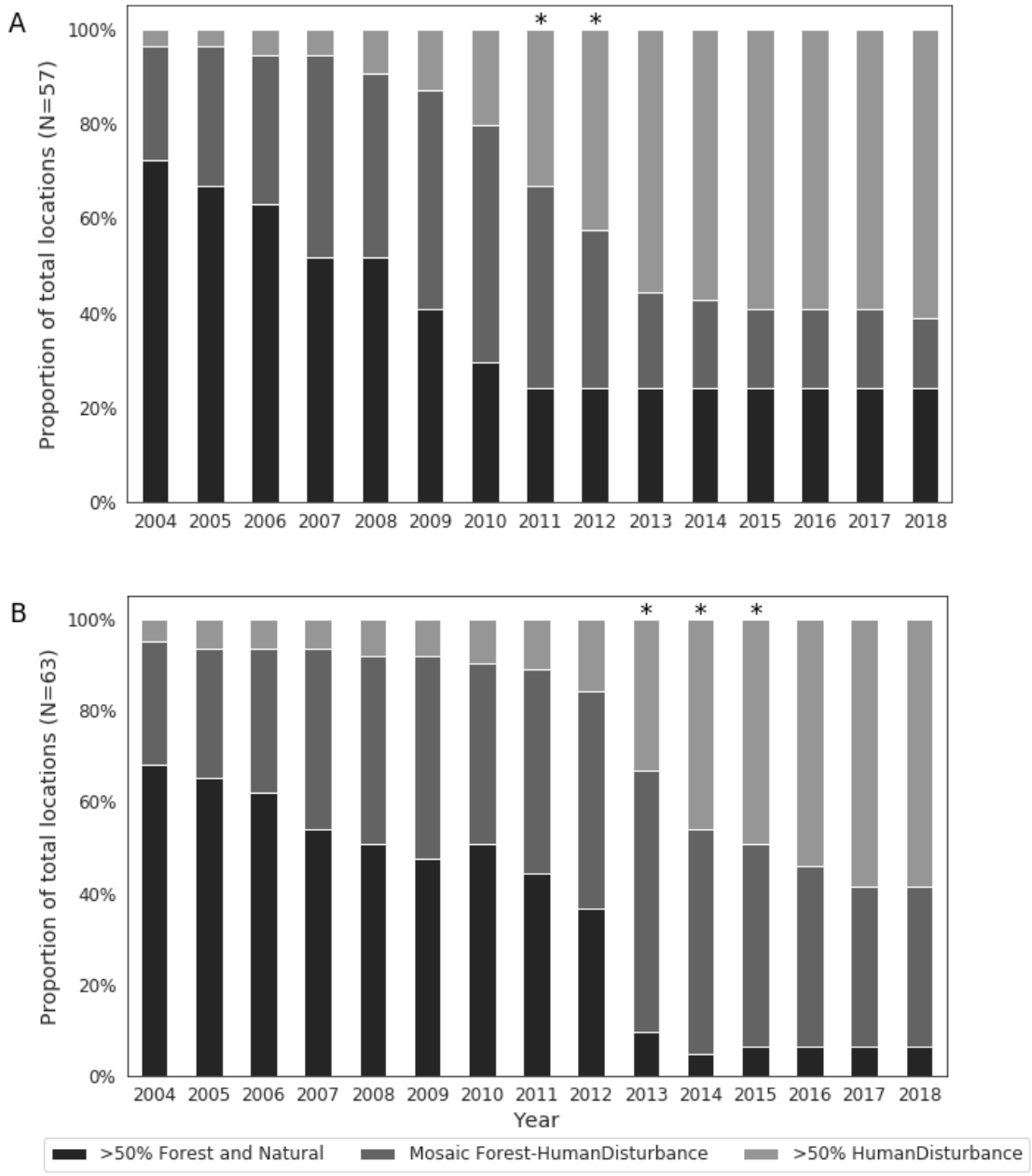


Figure 3.10 2004-2018 forest-intact/human-disturbed matrix characterised across the sites of the field campaign in Ucayali (A) and Loreto (B). In both areas, forest-intact landscapes *(>50% Forest and Natural*) are reduced over time. Asterisks (*) located above bars denote years of Terra-i's detection over the samples analysed by study area.

The relationship between the afore-mentioned landscape types and fire activity over time is denoted in **Figure 3.11**. In both areas, the rate of fire activity was higher across the samples of *50% Forest and Natural* landscape, particularly before Terra-i's disturbance dates. After these dates, the landscape types of *Mosaic* and *greater than 50% disturbance* experienced a higher fire activity rate.

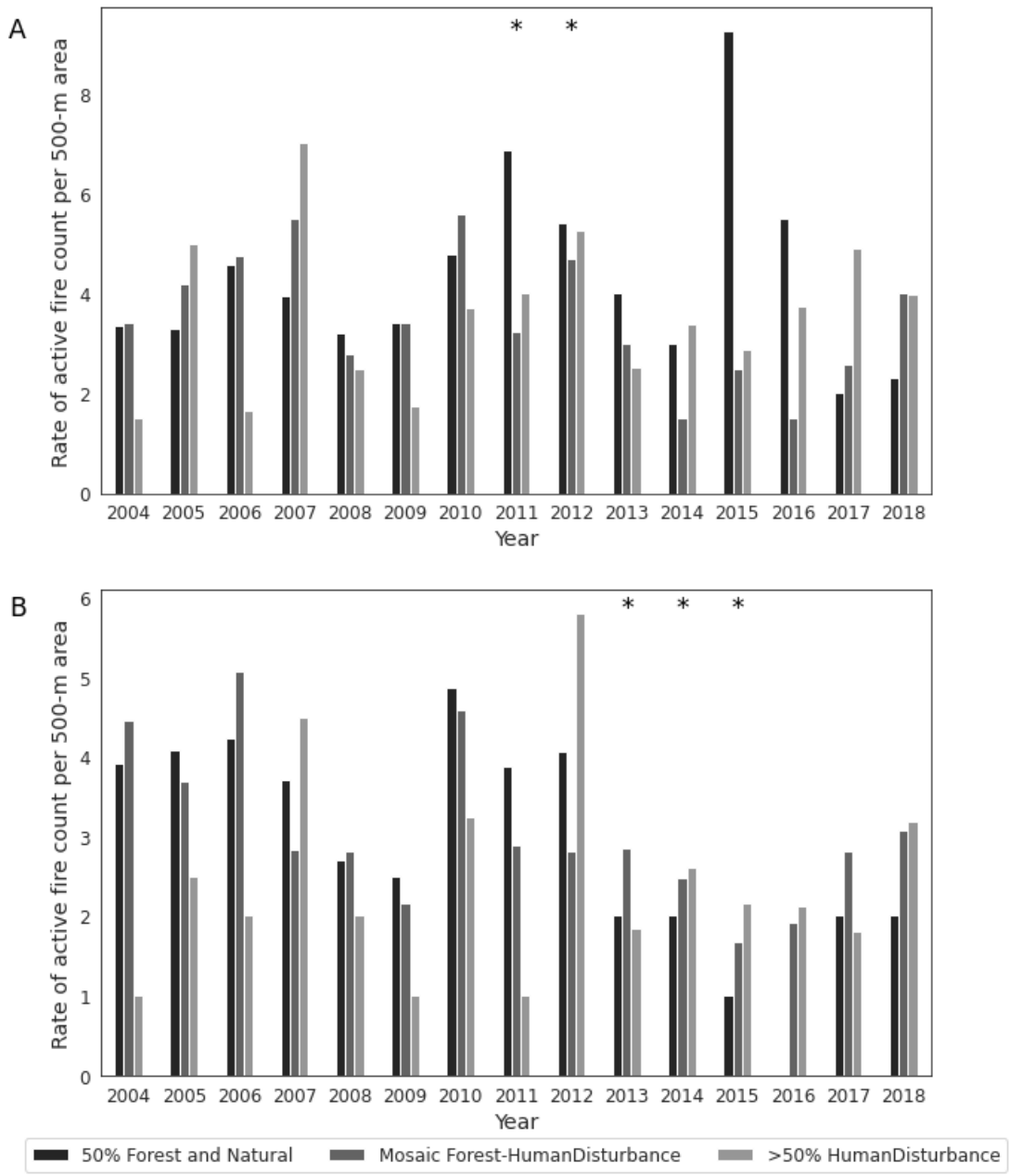


Figure 3.11 2004-2018 rate of active fire by landscape matrix type characterised across the sites of the field campaign in Ucayali (A) and Loreto (B). In both areas, the rate of fire activity was higher across the samples of *50% Forest and Natural* landscape, particularly before Terra-i's dates of disturbance. After these dates, the landscape types of *mosaic* and *greater than 50% disturbance* experienced a higher rate of fire per year. Asterisks (*) located above bars denote years of Terra-i's detection over the samples analysed by study area.

3.5.4 Land use change trajectories

The *lucCalculus* tool suited for reasoning LU change trajectories using spatio-temporal information such as depicted in **Figure 3.12**. This information was complementary to the multi-temporal analysis of land cover (see **Section 3.5.2**) and landscape (see **Section 3.5.3**).

Figure 3.12 Spatial distribution of 2004-2018 land use types across the areas (250-m MODIS pixels) inspected in Loreto, Peru. For this area, most of the pixels changed from Forest to Cropland, particularly after the Terra-i's disturbance dates, 2013 to 2015.

With the aforementioned spatio-temporal information the most common transitions from forest to other LU types were studied per field campaign. As it was introduced in **Section 3.4.3**, the *lucCalculus* tool facilitates the analysis of LU transitions (Maciel *et al.*, 2019). For instance, **Figure 3.13** illustrates a sequence plot of deforested sites in Ucayali and Loreto representing the transition *Forest* (green) to *Cropland* (purple). This transition was evident in 17 out of 57 and 33 out 63 locations, respectively. To be grouped under this category, the location, indicated by each row, must be labelled with the *Forest* class (green line) for at least 5 or 7 consecutive years according to end year of the Terra-i's validation, 2012 and 2015 in Ucayali and Loreto, respectively.

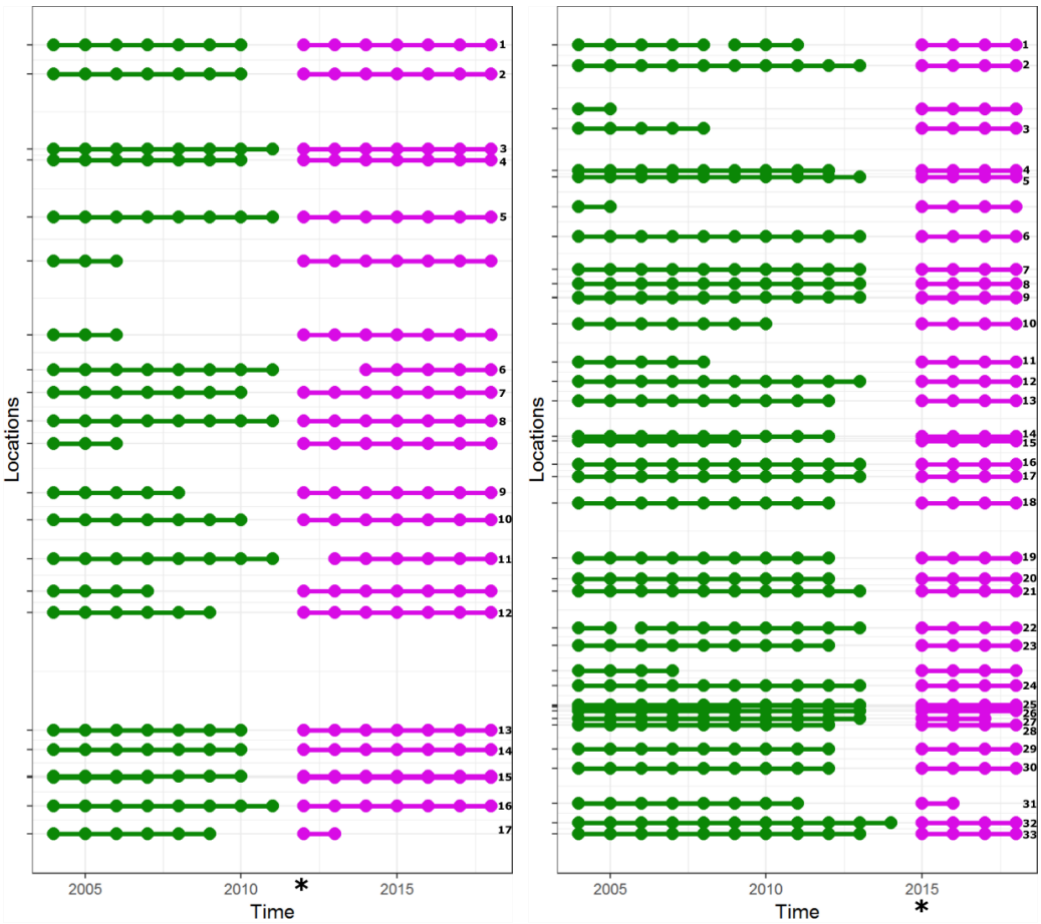


Figure 3.13 Example of pixels i.e. locations with *Forest* (green) to *Cropland* (purple) conversion across the deforested sites inspected from Ucayali (left) and Loreto (right) records. The temporal gaps, represented by the absence of continuous coloured lines, might be associated with intermediate LU types such as grassland which were not included in the query. The forest to cropland conversion was expressed in 17 out of 57 and 29 out 63 locations in Ucayali and Loreto, respectively. Locations (y-axis) are ordered by longitude, from west (top) to east (bottom). Only those locations with a given number of consecutive years, at least 5 (Ucayali) or 7 (Loreto), were considered as part of the defined trajectory and are numbered accordingly at the end of the sequence. Asterisk (*) located in the x-axis denotes Terra-i's end year of validation by study area.

The second most common trajectory represented changes from *Forest* to *Grassland*. According to the expression used in the *lucCalculus* tool to reflect this particular conversion, there are fewer deforested sites, 6 out of 57 and 4 out 63 locations in Ucayali and Loreto, respectively (see **Figure 3.14**), than the *Forest* to *Cropland* trajectory. The temporal gaps, represented by the absence of continuous coloured lines between green (forest) and orange (grassland), might be associated with intermediate LU types such as cropland. For instance, the first row at the left panel in **Figure 3.14**, which is the most western deforested site in Ucayali, shows that forest (green) goes from 2004 to 2005, then a gap exists until 2012 when grassland (orange line) is observed for 7 consecutive years.

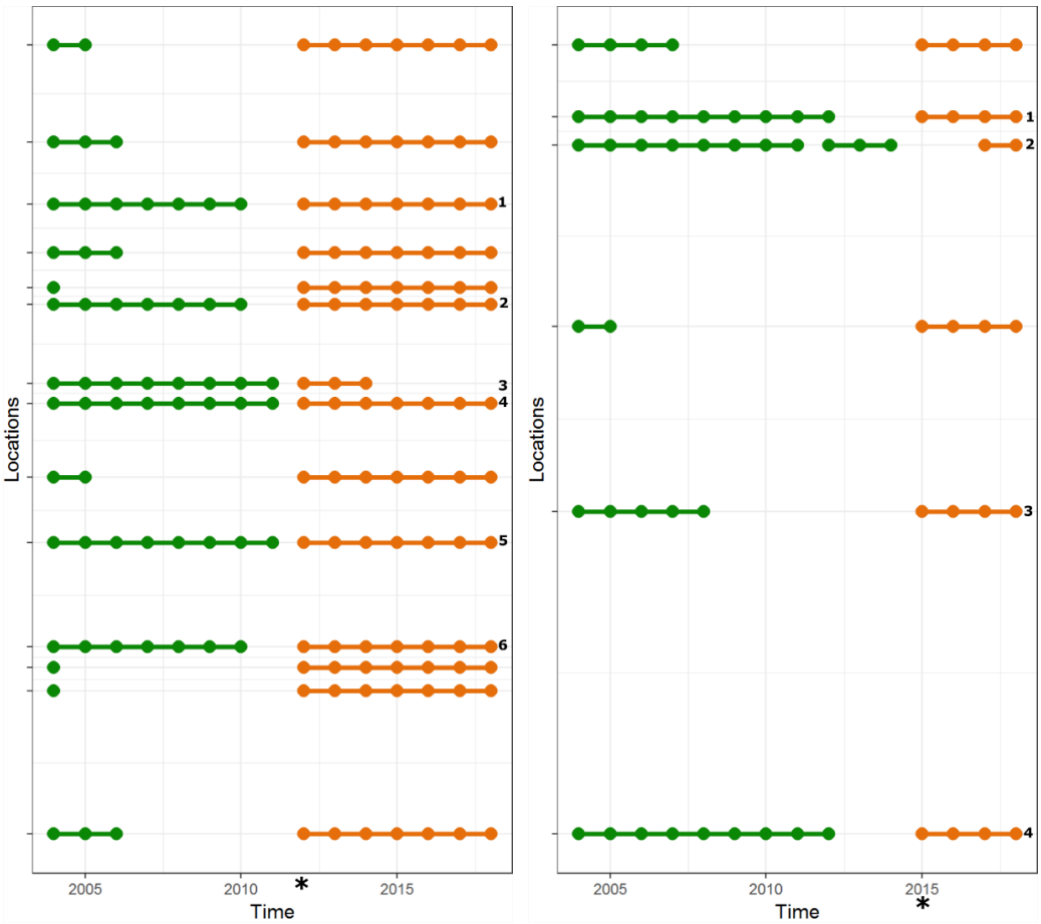


Figure 3.14 Example of pixels i.e. locations with *Forest* (green) to *Grassland* (orange) conversion across the deforested sites inspected from Ucayali (left) and Loreto (right) records. The temporal gaps, represented by the absence of continuous coloured lines, might be associated with intermediate LU types such as cropland which were not included in the query. The forest to grassland conversion was expressed in 6 out of 57 and 3 out 63 locations in Ucayali and Loreto, respectively. Locations are ordered by longitude, from west (top) to east (bottom). Only those locations with a given number of consecutive years, at least 5 (Ucayali) or 7 (Loreto), were considered as part of the defined trajectory and are numbered accordingly at the end of the sequence. Asterisk (*) located in the x-axis denotes Terra-i's end year of validation by study area.

The analysis of direct forest conversion over time using the *lucCalculus* tool reveals most forest changes occurred during the Terra-i's detection years (see **Figure 3.15**). The majority of locations converted within these years were *Forest* to *Cropland* followed by *Forest* to *Grassland.* The remaining years experiencing any type of conversion might be associated with the early/late detections identified in **Table 3.9**.

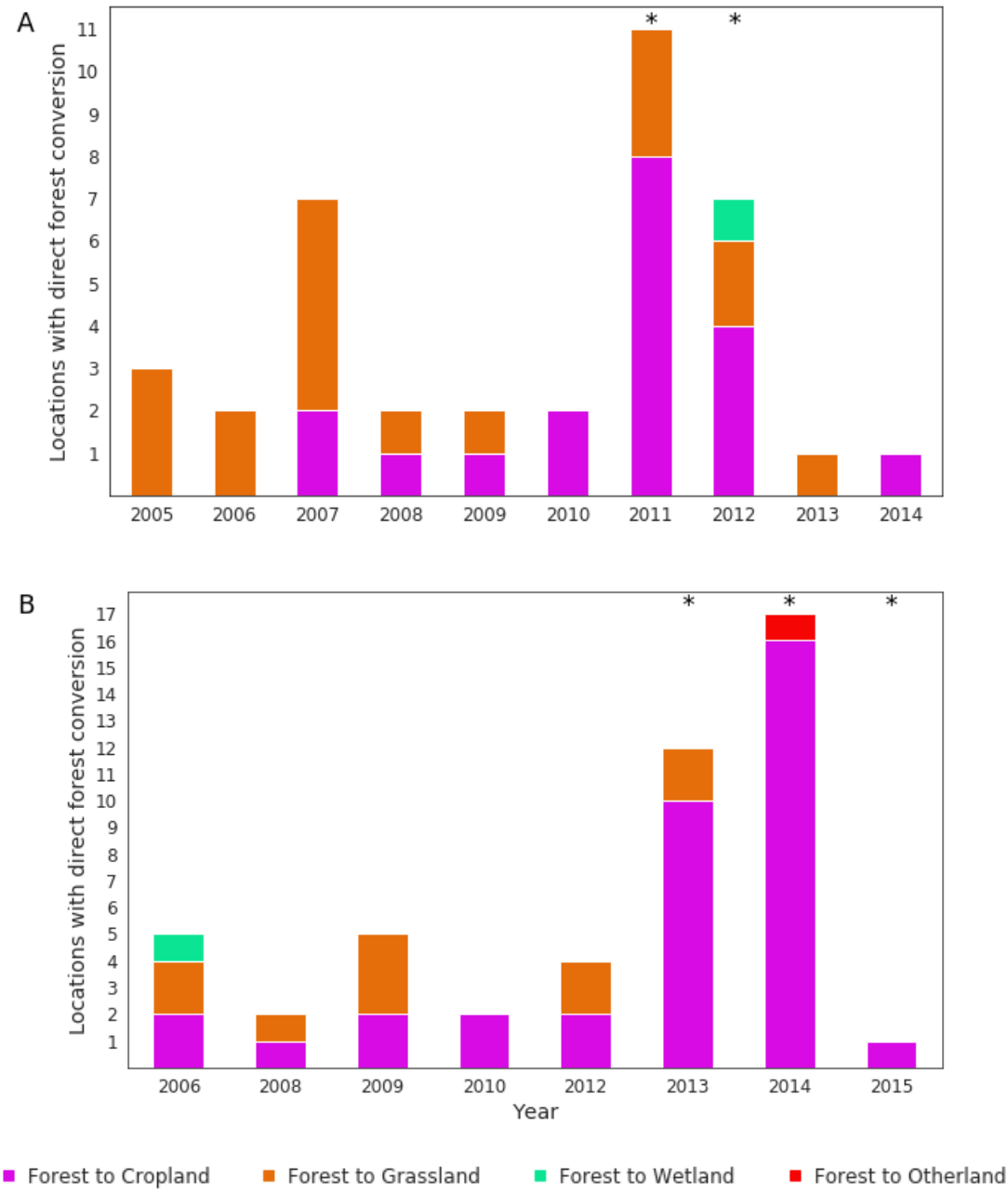


Figure 3.15 Direct conversion from forest to other land use types estimated by *lucCalculus* according to the 2004-2018 land use annotations across the sites in Ucayali (A) and Loreto (B). Asterisks (*) located above bars denoted years of Terra-i's disturbance detection over the samples analysed by study area.

3.6 Discussion and conclusions

The use of different multiple sources of geo-referenced data, including satellite data and field records, provided different insights into the dynamics of post-loss LC/LU. The *in-situ* field records as registered by a group of experts (Terra-i team and collaborators) provided the highest level of detail in this work. Similar types of databases are indeed very scarce in the tropics. Where they exist e.g. Geo-wiki's global dataset of LC/LU (Fritz *et al*., 2017) some limitations remain in terms of geographical coverage and resolution i.e. very coarse resolution 1-km. VHR and Landsat images were useful in understanding how spatially LC/LU change trajectories are distributed at pixel and landscape level across the study sites. Although MODIS data were less able to provide this level of detail, the high-temporal information was important to capture human-driven practices associated with deforestation such as fires.

For early warning forest monitoring systems such as Terra-i, there is an increasing number of frameworks assessing their performance focused on timing and minimum detectable size of forest disturbance (Tang *et al*., 2019). Access to VHR satellite images were useful in assessing the former aspect according to percentage of non-forest area at the time of field collection. The analysis of this particular variable in **Section 3.5.1** which was assessed only over *true positives* samples, indicated a minimum threshold of 33% to define a MODIS pixel as deforested. This result is in agreement with previous studies which claim 30% is a robust threshold below which change is hardly detectable in MODIS data (Sulla-Menashe *et al*., 2014; Huang and Friedl, 2014). For the performance in terms of timing, it was found that almost 25% of the detections across Ucayali and Loreto sites were early/late detections. Tang *et al*. (2019)'s study indicated that detection capability of an early warning system might vary according to the size and rates of deforestation activity within the target area. The sites studied in this work share similar socio-environmental conditions, which was reflected in very small differences within each proportion of type of change (false positives, true positives and lagged detections).

According to the characterisation stage, *Cultivated and managed* and Mosaic *of cultivated and managed/natural vegetation* are the most dominant LC classes (>60%) over the above deforested areas. While the former indicates that most of

the surface of the pixels under this class was almost completely disturbed (greater than 60%), the latter represents sites with the minimum threshold of 30% afore described where a considerable proportion of natural vegetation (e.g. forest fragments) remains. This remaining proportion is defined according to Geo-wiki's global LC validation sites (see **Table 3.6**) which used IGBP classification (Loveland and Belward, 1997) as reference nomenclature to define this particular class. For both study areas, *Cropland* followed by *Grassland* were the most dominant LU types. The agriculture expansion, including a wide variety of cropland types and intensification levels as observed in the study sites such as large-scale tree crops (e.g. oil palm, cocoa) and small-scale agriculture (e.g. rice, papaya) has been reported as recent drivers of deforestation in the Peruvian Amazon (Vijay *et al*., 2018; Marquardt *et al*., 2019; Ravikumar *et al*., 2017).

The multitemporal analysis of 15-years of satellite data, from 2004 to 2018, indicated a gradual expansion of both *Cultivated and managed* and *Mosaic of cultivated and managed/natural vegetation* at the cost of *Tree cover*. The trend of *Tree cover* reduction is likely to fit within the concept of forest transition curve (Mather, 1992; Hosonuma *et al*., 2012; Rudel *et al*., 2020). While this concept, which was previously introduced in **Section 2.2.2**, suggests the shift from net deforestation to net increase in forest cover, this was not completely evident during the post-deforested period of the sites analysed. To fully capture the forest curve is then necessary to either increase the number of years observed or study other geographical areas where the forest curve is expressed in a shorter period. It is worth to mention the task of comparing different areas with the proposed methodology might result challenging as it is time-consuming. Therefore, it is expected existing collaborative tools such as Geo-wiki or Collect Earth as described in **Table 3.1** improve their customisation capabilities for this purpose.

Complementary to the analysis of LC, the study of LU using the *lucCalculus* tool in **Section 3.5.4** provided a closer approximation regarding the evolution of certain conversions from forest to other LU types. For instance, while approximately half of the samples in Loreto followed the trajectory from *Forest to Cropland*, this split was lower for the sites across Ucayali. This might be associated with local communities occupying the sites studied in the western Amazon. Indeed, settler practicers can be regionally different, translating to

heterogeneous LU patterns (Fujisaka and White, 1998). This reasoning was particularly noticeable amongst the results which identify direct conversions over time. These results showed a larger number of pixels deforested from *Forest* to *Grassland* in Ucayali in comparison with Loreto. Suggesting there are areas where intermediate conversion also exists as a pathway e.g. *Forest* > *Grassland* > *Cropland*. A drawback of the *lucCalculus* tool is that the logic or reasoning of LU is only feasible to analyse with a limited number of sites. For a large amount of multi-temporal LC or LU records, the visualisation functions of this particular tool are not suitable as they are hardly interpretable. In this regard, some authors have developed more suitable tools for analysing and interpreting large amount of chronological LC/LU records (Gabadinho *et al*., 2011). The application of this sort of tools are further described and implemented in **Chapter 4** and **Chapter 6**.

Regarding the dynamics of change observed within the 500-m buffer areas (here denoted as landscapes) (see **Section 3.5.3**), it was evident that surrounding areas of the deforested sites were also changing and transforming over time. This particular aspect remains key to be addressed when deploying algorithms for large area monitoring of LC/LU change over deforested sites. Such algorithms should be capable to work at pixel level but considering the context information. For the observed deforested sites, the consolidation of disturbed landscapes (*>50% human disturbance*) was most prominent across the sites in Ucayali than Loreto. After Terra-i's deforestation dates, a lower percentage of *>50% forest and natural* landscape remained in the areas of Loreto than Ucayali. This might indicate that the size of forest conversion, by either clear-cutting or slash and burn, would be larger in the former region. These larger disturbances in Loreto might be associated with agriculture-driven deforestation in particular rice fields which were found to be spatially clustered in the landscape more so than other observed cropland types.

Coupled to the landscape matrix changes, the exploration of fire activity provided are in agreement with previous research about burning periods within intact and disturbed forest/natural vegetation matrices (Lima *et al*., 2012). In regions with high recurrence of land-use change such as developing countries, fire is commonly used for forest clearing, and as a land management practice for establishing agriculture and cattle ranching (Eva and Lambin, 2000; van der Werf

et al., 2010). The presence of a higher number of active fires over the matrices of >50% forest/natural vegetation and mosaics might be related to the transformation of most of these areas to agriculture. According to Morton *et al.*'s (2006) study in the Amazon, deforestation for cropland may involve burning for a longer period of time (days) than forest conversion for pasture.

The approach of visual inspection and respective classification by deforested site over time is not free of limitations. Due to its time-consuming nature and small-area coverage, it is less feasible for large-area analysis of post-loss LC/LU change. Instead, methods capable of covering large areas at low-cost such as image processing of remote sensing data are more desirable. The next chapters aim to introduce and evaluate the performance of a proposed method fulfilling this requirement. This method exploits pre-existing global LC maps and high frequency multispectral time series for generating LC data suited to the study of post-loss LC change trajectories.

Acknowledgements

The author acknowledges the International Center for Tropical Agriculture (CIAT) for granting access to the field records database accessed within this work. Field records were collected with the help of different sources of funding and are detailed as follows.

Field records in Ucayali were developed in the framework of the ASSETS (Attaining Sustainable Services from Ecosystems through Trade-off Scenarios Project) and also with funds of the CGIAR, Water, Land and Ecosystems (WLE) and Forest, Trees and Agroforestry (FTA) programmes. ASSETS project is part of the Ecosystem Services for Poverty Alleviation (ESPA), a research programme co-funded by UK Aid, the Natural Environment Research Council (NERC) and The Economic and Social Research Council (ESRC).

Field records in Loreto were developed in the framework of the Sustainable Amazonian Landscape Project. This project is part of the International Climate Initiative (IKI). The Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (BMU) supports this initiative on the basis of a decision adopted by the German Bundestag.

4 Assessment of pre-existing global LC classifications and high frequency multispectral time series suited to the study of post-loss LC change

4.1 Introduction

This chapter introduces the proposed method for land cover (LC) prediction suited to the analysis of post-loss LC change trajectories. The method is primarily based on novel methods in earth observation data analytics such as deep learning (DL) models and cloud-based platforms. Both resources allow the ingestion, analysis and processing of large volumes of satellite time series (SITS) and geo-referenced data (Parente *et al.*, 2019). The particular state-of-art DL model selected, which is based on Recurrent Neural Networks (RNNs), provides an end-to-end solution for large-area LC classification using SITS with minimal pre-processing (Rußwurm and Körner, 2018b). Hence, noise in SITS i.e. cloud obscured pixels or sensor artifacts is mitigated by internal structures of the model as is detailed in **Section 4.2.4.3.** Furthermore, the model extracts spatial and temporal features simultaneously with greater efficiency than traditional supervised algorithms in remote sensing (e.g. Random Forest) (Rußwurm and Körner, 2018b).

Though prior research has demonstrated that the proposed architecture provides consistent LC predictions using MODIS archives at 250 m for a single year across the Brazilian Amazon (Coca-Castro *et al.*, 2019), the implications of varying LC classification schemes i.e. number and type of LC classes have not yet been assessed. The scope in this chapter is to study the effectiveness of the proposed DL model according to multiple LC classification schemes (hereafter referred to as labelled datasets) derived from pre-existing global LC products. These products represent a wide range of spatial resolutions (from 100-m to 500-m) and LC classification schemes. Moreover, they are relatively coarser as the resolution of the target deforested areas as detected by the Terra-i system (250-m). The

selection of the labelled dataset for the generation of LC data best suited to the analysis of post-loss LC change considered several aspects. These aspects include an assessment of the traditional machine learning metrics (i.e. overall accuracy) as well as the cross-year consistency of the spatio-temporal predictions using high-resolution imagery (30-m Landsat). After defining which labelled dataset is best suited to the study of post-loss LC change, model calibration and comparison against other traditional models using SITS for LC classification are conducted as further detailed in **Chapter 5**.

4.2 Background

4.2.1 Time series data and analysis

Most real-world phenomena exhibit changes over time. Time series (TS) data are derived from measurements of an underlying phenomena and are represented as sequential instances. TS analysis aims to capture meaningful information and knowledge from those measurements (O'Reilly *et al.*, 2017). Research on TS analysis is conducted along one or more of the following dimensions: linear vs non-linear; parametric vs non-parametric; univariate vs. multivariate and evenly vs unevenly (Eckner, 2012).

Whilst the first two dimensions represent a set of methods to analyse and/or model TS data, the remaining two dimensions refer to implicit characteristics in which TS are observed. For univariate time series, a single variable or measurement (x) is recorded over time, $T = (x_1, \cdots, x_n)$ where $x_i \in \mathbb{R}$ for $1 \leq i \leq n$. On the other hand, instead of observing a single variable, multivariate time series represent two or more variables (x, y, r) observed subsequently over time, $T = \big((x_1, y_1, r_1), \cdots, (x_n, y_n, r_n)\big)$ where $x_i, y_i, r_i \in \mathbb{R}$ for $1 \leq i \leq n$. According to the sampling frequency, TS can be categorised as evenly or unevenly (also called regularly and irregularly) sampled. Both types of TS can be analysed by a wide spectrum of methods being more popular methods for regularly spaced TS (Eckner, 2012).

A diverse range of tasks are involved in TS characterization and analysis. These tasks are supported by large and interdisciplinary toolsets developed for problems ranging from forecasting, clustering, classification, anomalies and/or outlier detection, among others (Fulcher, 2018).

4.2.2 Satellite image time series (SITS)

The field of earth observation (EO) integrates the understanding, interpretation and establishment of relationships between Earth's surface measurements of electromagnetic energy and natural phenomena. Different spectral signatures by physical objects can be captured from satellites, drones, or even from a camera with a sensor capable of capturing wavelength information (Tatem *et al.*, 2008).

From the aforementioned sources, satellite images are by far the predominant source of EO data. Their success relies on large global and temporal availability and easy accessibility. In particular, free-of-charge satellite imagery from satellite missions such as MODIS and Landsat launched by the National Aeronautics Space Administration (NASA), or Sentinel by the European Space Agency (ESA) provide a wide variety of opportunities for EO applications (Bontemps *et al.*, 2012).

The collection of satellite images for a given geographical area at different times are denoted as Satellite Image Time Series (SITS). SITS can be described as a stack of images (a single layer per date) (Jönsson and Eklundh, 2004). The most discriminative characteristic of SITS data is the dimensionality or depth of its feature space. If the SITS capture a single feature, it is denoted as univariate SITS. In contrast, SITS data collected by multispectral or hyperspectral sensors, with a depth of two or more channels are categorised as multivariate SITS. **Figure 4.1** shows an example of a univariate SITS which can be represented in a 3D matrix (x, y, t) where x, y denotes a location (lat, lon) in a given time t.

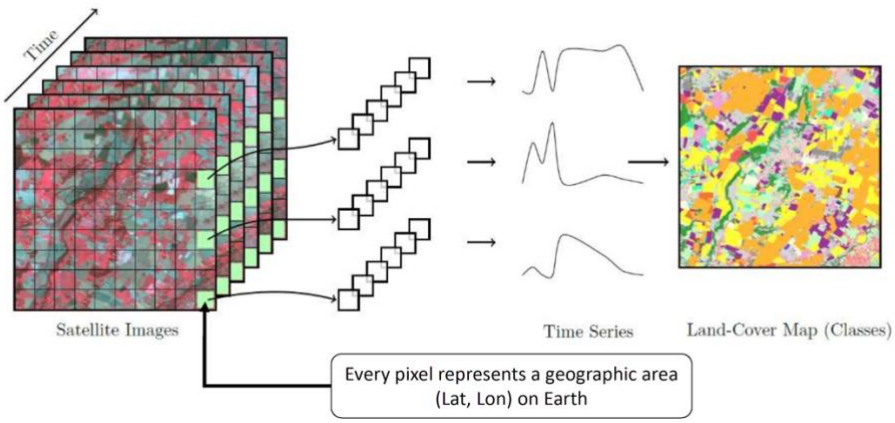


Figure 4.1 Representation of a SITS which pixel-level information allows for discriminating land cover types over a given geographical area. Adapted from Petitjean et al. (2012).

SITS present a set of properties and approaches for their analysis and modelling according to the target application (Petitjean, Kurtz, *et al*., 2012). For LC classification, a wide variety of traditional and emerging methods, mostly non-parametric supervised classifiers, exist (Wulder *et al*., 2018). Irrespective of the type of classifier, the suitable solution is guided by the application and end-user needs. Moreover, the selection of a given algorithm depends on the amount of data compositing SITS and the ability of the method to analyse them (Pelletier *et al*., 2016). The main aspects of both traditional and emerging approaches for SITS classification are covered in the following subsections.

4.2.3 Traditional approaches for SITS classification

Over the past decades, traditional classifiers, mostly supervised such as artificial neural networks (ANNs), support vector machines (SVM), decision tree and ensemble (e.g. Random Forest), have been shown to be efficient and accurate for producing LC maps at pixel-level and object-level from SITS (Li *et al*., 2014). These algorithms require a training input dataset to learn a decision rule, which is then used to predict the labels from an unlabelled dataset (Wulder *et al*., 2018).

When all observations of SITS are provided as input data to these classifiers, a major drawback remains in their low capabilities to identify specific temporal behaviours (Petitjean, Kurtz, *et al*., 2012). As a result, there is no influence on these classifiers' learning and outcomes from the temporal order in which satellite data are presented. For instance, shuffling the satellite images would return models with the same accuracy performance. This behaviour might induce a loss of temporal behaviour for classes with a temporal pattern e.g. seasonal land cover types (wetlands) (Pelletier *et al*., 2019).

The training set for these classifiers is commonly composed of phenological and/or statistical metrics extracted from SITS which allow discriminating targeted LC classes (Pittman *et al*., 2010; Jia *et al*., 2014; Azzari and Lobell, 2017). While some of these metrics aim to capture a temporal pattern i.e. the maximum vegetation index or the time of peak vegetation index, they have shown minimal effect on classification performance (Pelletier *et al*., 2016). More sophisticated approaches have tried to approximate to the temporal domain by combining temporal similarity measures e.g. Dynamic Time Warping and Euclidean distance,

with algorithms able to analyse them (Petitjean, Inglada, *et al*., 2012). These measures compare how two temporal sequences or time series are similar according to a set of rules and optimisations. Although these techniques have been applied widely in remote sensing (Xue *et al*., 2014; Baumann *et al*., 2017; Maus *et al*., 2016), they are often computationally expensive, a complete scan of the training set is needed to classify each test instance, and their complexity increases for large datasets (Xue *et al*., 2014).

4.2.4 Deep Learning in Remote Sensing and LC prediction

Deep learning (DL) has been considered as one of the breakout technologies in the artificial intelligence computation field during the last decade. It constitutes a group of modern techniques to discover knowledge and patterns from raw data that traditional machine learning algorithms have not been capable of analysing before. For decades, these traditional algorithms have used feature extraction routines and a considerable domain expertise to transform the raw data into a suitable internal representation from which the classifiers distinguish patterns (LeCun *et al*., 2015).

Rooted in the principles of Artificial Neural Networks (ANNs), DL is about 'deeper' ANNs that build a hierarchical representation from raw input data (Schmidhuber, 2015). These multiple levels of representations allow the learning of key aspects which are important to the learning task i.e. classification or regression. Thanks to the characteristic of automatic feature extraction, DL can solve more complex problems than procedures relying on time-consuming human-crafted features e.g. phenological and/or statistical metrics extracted from SITS (Pelletier *et al*., 2019). Deep ANNs did not receive much attention for a fairly long period until some advances such as the development of efficient multiprocessor graphics cards or graphics processing units (GPU), innovations in certain components (e.g. activation, normalisation and loss functions), and large training datasets became available (Schmidhuber, 2015).

A recent meta-analysis review conducted by Ma *et al*. (2019) shows DL started its momentum in remote sensing in 2014 when the number of works published in conferences and journals grew considerably (see **Figure 4.2a**). This trend could be explained by the capabilities of DL to handle some challenges of earth

observation data such as large volumes, multiple-scale information (different spatial and temporal resolutions), high-dimensionality (number of bands) and unstructured format (Zhu *et al*., 2017). Thanks to these capabilities, DL is currently applied to almost every step of remote-sensing image processing and analysis. Most of the successful applications of DL in remote sensing are focused on classification tasks (i.e., land cover and land use classification, object detection, and scene classification) (see **Figure 4.2b**). Furthermore, other non-conventional tasks have been outperformed by using DL-based techniques such as image registration (Merkle *et al*., 2018; Hughes *et al*., 2018), image fusion (Huang *et al*., 2015; Palsson *et al*., 2017), earth observation product validation (Xing *et al*., 2018) and acquisition of classification data (Chi *et al*., 2017).

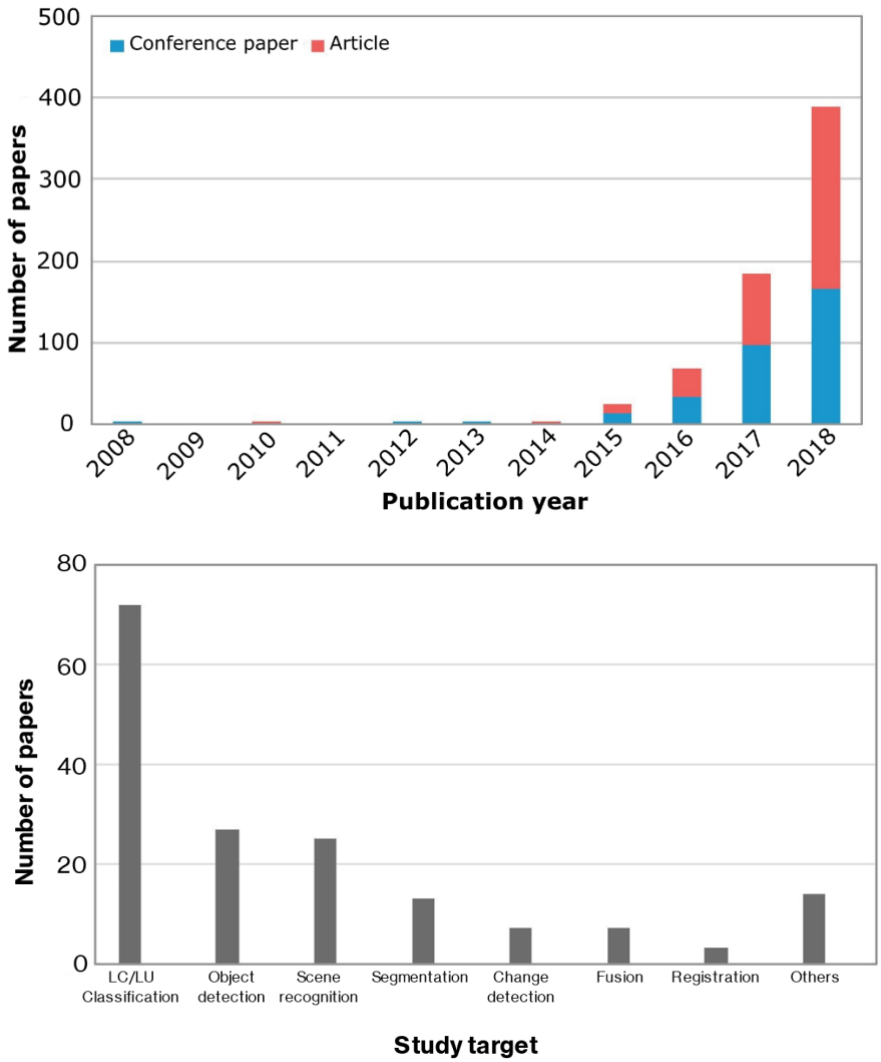


Figure 4.2 (a) Number of conference papers and articles in the Scopus database for a general search on ["deep learning" AND "remote sensing"] and (b) distribution of those publications for different study targets (b). Adapted from Ma *et al*. (2019).

It is important to mention the majority of studies compiled by Ma *et al.* (2019) were conducted using hyperspectral data or high (<10-m) to medium spatial (10-m to 30-m) resolution images (see **Figure 4.3a**). In addition, most of them were focused on the urban domain (see **Figure 4.3b**) followed by published articles in vegetation (including forest and cropland) and water areas. Other types of study areas including roads, snow, soil, and wetland were very few in number, and the authors did not list them individually but grouped with others.

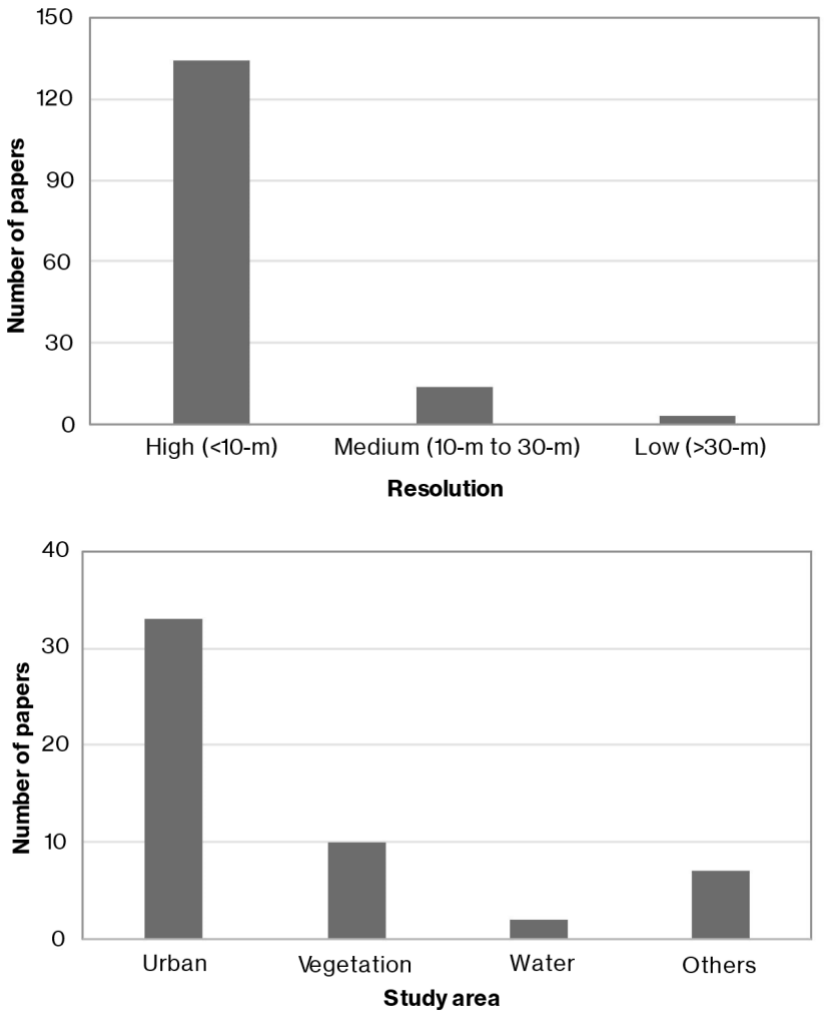


Figure 4.3 (a) Distribution of image spatial resolutions and (b) types of study areas in published articles concerning DL for remote sensing. Adapted from Ma *et al.* (2019).

For LC mapping, DL has widely tested on high resolution data due to their fine structural information (i.e. spatial details) (C. Zhang *et al.*, 2017; Marcos *et al.*, 2018; Li *et al.*, 2019; Zhang *et al.*, 2019). Conventional DL algorithms have also been applied to free-of-charge medium-resolution (10-m to 30-m) satellite images such as Landsat (Lyu *et al.*, 2018) and Sentinel data (Minh *et al.*, 2018; Cai *et al.*, 2018;

Hu *et al*., 2018). Although reasonable results have been obtained with these types of images, their performance is highly compromised by a lower level of structural information than with higher resolution images, see for example Helber *et al*. (2018). The reduction of spatial detail in medium and low-resolution satellite images is compensated by higher spectro-temporal information. Additionally, some scholars suggest adding the spatial information of nearby pixels (e.g. patch-based classification) provides better results than pixel-wise classification for medium resolution images (Sharma *et al*., 2018; Rußwurm and Körner, 2018b).

Whilst most of the DL models for LC classification use a cloud-free single image scene, raw or pre-processed (metrics i.e. mean, median derived from a stack of images), there is a growing interest by scholars to import high-temporal information of satellite image time series to DL models (Ji *et al*., 2018; Kamilaris and Prenafeta-Boldú, 2018; Ma *et al*., 2019). Amongst the supervised DL-based models, convolutional neural network (CNN) and recurrent neural network (RNN) models are commonly applied to SITS classification (Pelletier *et al*., 2019). For the purpose of this work, the general principles of ANNs as well as notions of CNN and RNN models are presented in the following sections, particularly focusing on existing work in LC classification using SITS.

4.2.4.1 General Principles

ANNs are computational models inspired by the knowledge of biological nervous systems. In comparison with standard machine learning approaches, ANNs are trained in an end-to-end manner, solely based on input data and output labels. The fundamental building blocks of these types of models are the neurons which could be considered as functions that take a given number of parameters and output a result (Bishop, 1995; Ripley, 2007).

The simplest possible neural network is a single neuron, which takes N inputs $x \in \mathbb{R}^D$ (plus a bias term) and outputs $h_{w,b}(x) = f(w^T x + b)$, where f is the activation function, w is the weight vector, and b is the bias term (see **Figure 4.4a**). For classification tasks, a loss function, which focuses on quantifying the divergence between predicted and actual output probabilities, is minimised at each training step by back-propagating residuals and adjusting the network weights and biases. ANNs are commonly arranged in multiple stacked layers with the hidden

output of one layer forming the input of the consecutive layer (see **Figure 4.4b**). The number of neurons, expressed as dimensions of hidden vectors m and number of layers l, are the main hyper-parameters to be set according to the application (Bishop, 1995; Ripley, 2007).

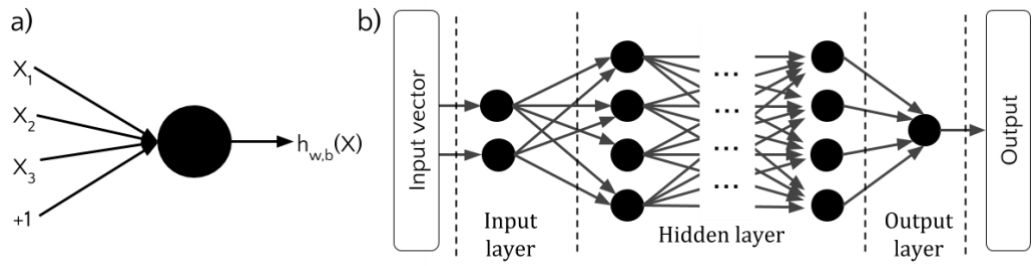


Figure 4.4 (a) Basic neural network unit (neuron, node) and (b) a typical multilayer neural network architecture. Adapted from Krenker *et al*. (2011).

ANNs are characterised by their architecture (the network topology and pattern of the connections between the nodes), the method of determining connection weights, and activation function employed. Amongst the many types of ANNs that differ in connection pattern, feed-forward networks (FNNs) and recurrent neural networks (RNNs) are the most common with a broad number of applications. **Figure 4.5** illustrates the topologies of both types. In FNNs, the information flows in only one direction, from input to output. In contrast, RNNs process the information in both directions (feed-forward and feedback) generating a cyclic relation between network outputs and inputs (Bishop, 1995; Ripley, 2007).

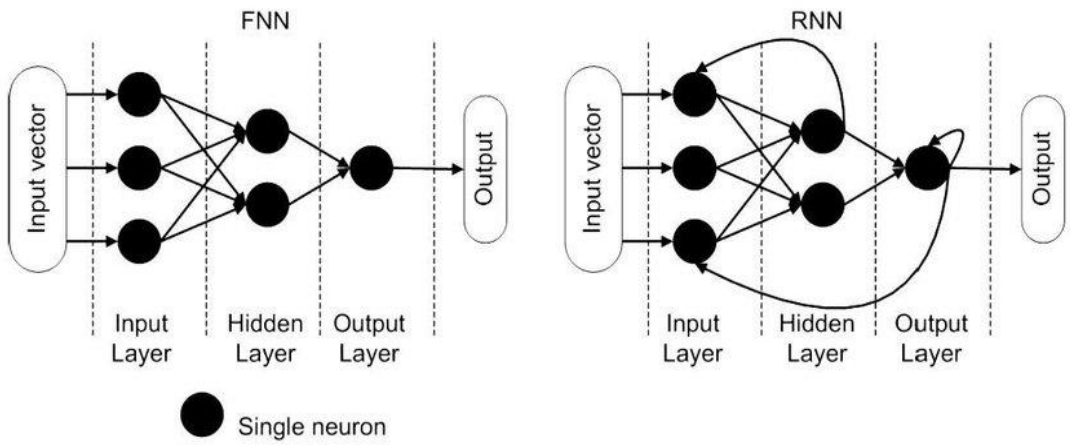


Figure 4.5 Topologies of feed-forward and recurrent neural networks (FNN and RNN in short). Different to FNN, RNNs process the information in both directions (feed-forward and feedback) generating a cyclic relation between network outputs and inputs. Adapted from Krenker *et al*. (2011).

Early studies concerning ANNs for remote sensing are summarised in Atkinson and Tatnall (1997). Some of these studies evaluated ANNs for LC classification using optical and radar imagery (Hepner *et al.*, 1990; Drummond and Loveland, 2010). Other initial applications include cloud identification and classification (Lee *et al.*, 1990) and the prediction of continuous variables e.g. biomass (Jin and Liu, 1997) and forest parameters (Wang and Dong, 1997).

4.2.4.2 Convolutional Neural Networks (CNN)

CNN is one of the most successful network architectures in DL-based applications for remote sensing, including LC classification (Ma *et al.*, 2019). CNN models, originally designed for computer vision tasks, use various convolutions in both x and y dimensions to identify the spatial structure of the data (Krizhevsky *et al.*, 2017). Three distinct types of hierarchical structures are present in CNNs: convolution layers, pooling layers, and fully connected layers. At each layer, the input image is convolved by a set of k kernels $W = \{W_1, W_2, \cdots, W_k\}$ and added biases $\gamma = \{b_1, b_2, \cdots, b_k\}$, each generating a new feature map X_k. These features are subjected to elementwise nonlinear transform $\sigma(\cdot)$, and the same process is repeated for every convolutional layer l: $X_k^l = \sigma\left(W_k^{l-1} * X^{l-1} + b_k^{l-1}\right)$ (Ma *et al.*, 2019).

The convolutional layers may be abstractly seen as a series of filters that transform an input image into another to highlight discriminative features. At the end of the network, these discriminative features are passed to a classifier e.g. fully connected layers which exploit the high-level features learned to classify input images into a set of targeted classes or make numeric predictions (Kamilaris and Prenafeta-Boldú, 2018) (see **Figure 4.6**). In contrast to traditional multilayer feed-forward networks, in CNNs the values of pixels within a neighbourhood of a certain size are aggregated using a permutation invariant function, typically the max or mean operation (also denoted as max-pooling or average pooling, respectively) (Ma *et al.*, 2019).

CNNs have been widely used for extracting patterns from remote sensing images, particularly high resolution optical and hyperspectral imagery (Ma *et al.*, 2019). The typical structures tested range from 1D to 3D CNNs targeting spatial and spectral dimensions separately (Liang and Li, 2016; Mou *et al.*, 2017; Cai *et al.*,

2018) or simultaneously (Y. Li *et al.*, 2017; Hamida *et al.*, 2018). Although less common than these dimensions, CNNs can also be used to analyse the temporal domain of remotely sensed time series. Inspired by previous work in CNNs for time series classification (Wang *et al.*, 2017) and video classification (Wu *et al.*, 2015), some scholars have proven 1D and 3D CNN-based architectures are also effective for handling SITS classification using the temporal (Pelletier *et al.*, 2019; Zhong *et al.*, 2019) and spatio-temporal (Hamida *et al.*, 2018; Ji *et al.*, 2018; Stoian *et al.*, 2019) dimensions. Although these architectures have contributed to LC classification using SITS, they require a certain level of pre-processing for example, resampling the temporal observations to a regular interval and relying on simplistic methods i.e. linear interpolation (Pelletier *et al.*, 2019; Zhong *et al.*, 2019) to handle gaps caused by no observation, bad data or cloud cover.

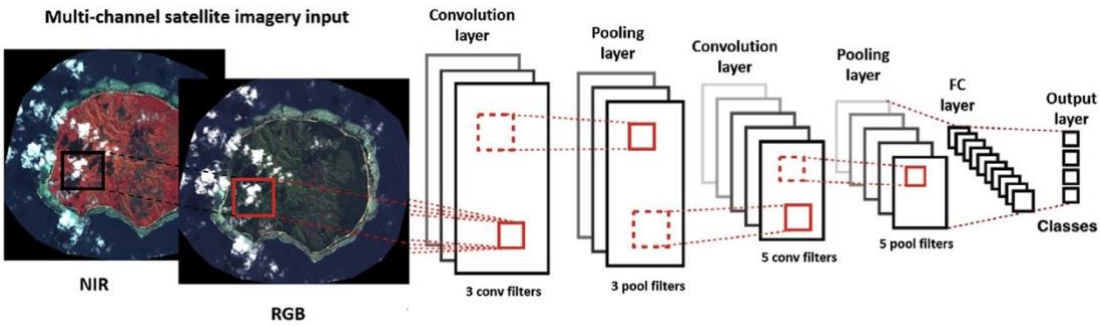


Figure 4.6 Example of a generic CNN pixel-wide classification for a single-time step satellite image classification as illustrated by Segal-Rozenhaimer *et al.* (2020). The network ingests both the pixel and spatial information from multi-channel satellite imagery (RGB + NIR) through a series of 2D convolutional layers, followed by pooling layers (mean/max operators for spatial dimensionality reduction), fully connected (FC) layer(s); and an output layer. The output layer can be a logistic regression (for binary classification), softmax (multi-class) or other classification output probability function.

4.2.4.3 Recurrent Neural Networks

In contrast with CNNs, which contain different types of layers that perform different functions (e.g., convolutional, pooling and nonlinear layers), the basic RNNs consist only of recurrent layers. RNNs are intrinsically suited for sequential data and have been successfully applied in machine translation (Bahdanau *et al.*, 2015), text summarization (Nallapati *et al.*, 2016) or speech recognition (Chorowski *et al.*, 2015). RNNs iteratively encode a sequence of observations $x = \{x_0, \cdots, x_t\}$. Since the data is processed sequentially, deeper representations are created through sequential updates $h_t \leftarrow x_t, h_{t-1}$ with context information from the previous representation h_{t-1}. Hence, a stack of recurrent layers return deep

high-level representations for either classification or regression tasks (Graves *et al*., 2013). **Figure 4.7** depicts a diagram of a typical neural RNN which unfolded produces an equivalent layered FNN with shared weights (Schmidhuber, 2015).

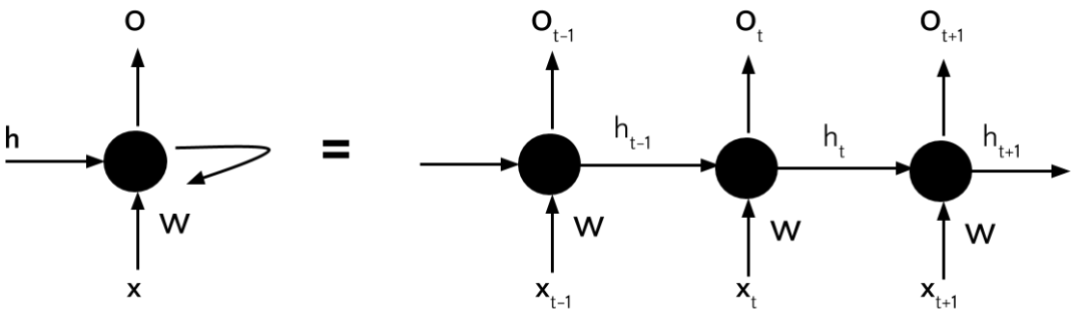


Figure 4.7 Representation of a folded (left) and unfolded (right) RNN unit. x (input); h (hidden state); W (weights); O (output). Adapted from Gakhov (2015).

The training in the basic RNN is carried out across multiple times using backpropagation, a method for calculating the gradient of network parameters. When applied to long sequences, the repeated use of the recurrent weight matrix during backpropagation makes the basic RNN susceptible to *vanishing and exploding gradient problems*. The vanishing problem is due to the repeated multiplication of small numbers together that results in smaller and smaller products, in particular affecting the earlier layers in the network. This problem prevents the model from learning short term aspects of the data. In contrast, exploding gradients are generated from repeatedly multiplying very large numbers resulting in larger and larger products (Bengio *et al*., 1994), meaning the model forgets the long term. To overcome both problems, specialized memory units using a gate-based system were proposed, initially Long Short-Term Memory (LSTM) memory cells (Hochreiter and Schmidhuber, 1997) and later, Gated Recurrent Unit (GRU) (Cho *et al*., 2014). As illustrated in **Figure 4.8**, a set of multiplicative units called *gates* in each of these memory units control the gradient flow through time and enable learning of long temporal relationships.

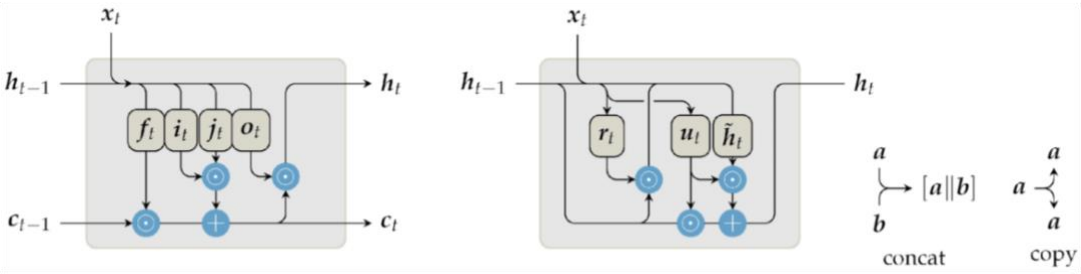


Figure 4.8 Schematic illustration of LSTM (left) and GRU (right) cells. The cell output h_t is calculated via internal gates (four and two in LSTM and GRU, respectively) and based on the current input x_t combined with prior context information h_{t-1}, c_{t-1}. LSTM cells are designed to separately accommodate long-term context in the internal cell state c_{t-1} from short-term context h_{t-1}. GRU cells combine all context information in a single, but sophisticated output h_t. $\odot$ and $\oplus$ represent element-wise multiplication and addition, respectively. Adapted from Rußwurm and Körner (2018b).

Under the assumption that observations of satellite images are time-dependent, RNNs have been extended for pixel-wise LC classification using either multitemporal optical (Jia *et al.*, 2017; Z. Sun *et al.*, 2019; Rußwurm and Körner, 2018b) or Synthetic Aperture Radar (SAR) data (Ienco *et al.*, 2017; Minh *et al.*, 2018; Ndikumana *et al.*, 2018; Y. Sun *et al.*, 2019). In addition, in order to process spatiotemporal data, RNNs have been combined with convolutions either by merging representations learned by the two types of networks (e.g. CNN-RNNs) (Ballas *et al.*, 2016) or by feeding a CNN model with the representation learning by a RNN model (e.g. ConvRNNs) (Shi *et al.*, 2015). Unlike CNN-RNNs, which interpret the output from CNN models, ConvRNNs use convolutions directly as part of reading input into the RNN units themselves (Pelletier *et al.*, 2019). For LC mapping, both types of spatiotemporal implementations of RNNs and others using different types of ANNs such as residual neural networks (ResNets) instead of CNNs have been explored (Benedetti *et al.*, 2018; Rußwurm and Körner, 2018b; Interdonato *et al.*, 2019) and compared (Garnot *et al.*, 2019b). In contrast to state-of-the-art DL-based methods for SITS such as CNN and ResNets, RNNs have greater generalization capabilities to be trained and predict multitemporal satellite data of different length and intervals (Rußwurm and Körner, 2018b). Furthermore, recurrence models and other emerging models based on attention mechanisms have achieved superior performance values on noisy raw SITS compared to convolutional models (Rußwurm and Körner, 2019).

4.3 Towards a spatially-explicit dataset of pantropical post-loss LC change trajectories using SITS and auxiliary data

According to Defries and Townshend (1999) and Hansen and Loveland (2012), experimental and operational approaches exist for large LC mapping and monitoring. This chapter and the subsequent chapter aim to contribute to the state-of-art of both approaches, in particular through the development of a novel and dedicated spatially explicit multi-temporal LC dataset suited to the study of post-loss LC change trajectories in the pantropics.

For the **experimental approach**, the following aspects have been identified as crucial to generating the proposed dataset by analysing SITS and contextual data:

- Exploitation of frequently available and free-of-charge remotely sensed data over the entire globe: instead of analysing specific dates, the model should be able to **analyse a broad spectrum of SITS** (Waske and Braun, 2009; Cai *et al.*, 2014; Camara *et al.*, 2016; Gómez *et al.*, 2016).

- **Adaptation to strong temporal variations by LC type:** the performance of traditional classification models for LC prediction is generally affected by temporal variations due to external conditions (precipitation, sunlight, temperature) (Jia *et al.*, 2017). Therefore, the patterns for a LC learned in a given period may fail to predict the same LC in subsequent periods due to differing external conditions resulting in misclassification (Jia *et al.*, 2017; Tardy *et al.*, 2019).

- Reliability and **logic in the transitions** by LC type: land covers have different temporal transition patterns. For instance, rules implemented into the model or post processing techniques should reduce illogical transitions. For instance, forest conversion to croplands/urban area is more likely to occur than urban area to forested area in a given time step (e.g. one year) (Gómez *et al.*, 2016).

Coupled with the elements required for an experimental approach, the main elements considered for an operational system which delivers reliable data products on a regular basis (Inglada *et al.*, 2017), are:

- Implementation of **automated** big data pipelines for efficient ingestion, storage, processing (parallel workflow) and analysis of earth observation data, including all available images, regardless of the extent of noise, and other georeferenced datasets. According to Tardy *et al.* (2019) for an operational LC mapping over large areas, automated processing is preferred.

- Usage of **existing databases** for training classifiers avoiding the dependence of manual reference sample collection. The traditional production of LC maps using remote sensing imagery is primarily governed by data availability. In addition to satellite images, labelled reference data are an essential element for classification. For large scale (i.e., over large areas) LC mapping at the annual frequency, the collection samples and labels either by field campaigns or by visual interpretation is prohibitive (Huang *et al.*, 2015). The exploitation of existing databases has been proposed as an alternative, to alleviate the bottleneck of delays in obtaining reference data to update LC maps (Tardy *et al.*, 2019).

- Reducing to a **minimum the operations** in the processing chain. Traditional pipelines for processing remote sensing data for LC mapping largely rely on particular pre-processing steps e.g. radiometric calibration, noise-filtering, feature extraction, among others. New pipelines with capabilities to ingest and analyse raw satellite data have emerged achieving state-of-art accuracies without dedicated pre-processing chains (Man *et al.*, 2018; Rußwurm and Körner, 2018b, 2019).

4.4 Study area and context

The study area corresponds to the Amazon region which represents the largest continuous region of tropical forest in the world. The region has special relevance for biodiversity (Dirzo and Raven, 2003; da Silva *et al.*, 2005), global nutrient cycles (Coe *et al.*, 2009), climate change (Cheng *et al.*, 2013; Achard *et al.*, 2014) and the support of numerous livelihoods (Perz *et al.*, 2013). Due to threats faced within this region such as forest loss and forest degradation, LC/LU monitoring has received great attention during the preceding decades, particularly in the Brazilian amazon (Tyukavina *et al.*, 2017).

Definitions of the boundaries of the Amazon vary widely between countries and context (Eva *et al*., 2005). For the purposes of this work, the Amazon as delineated by the Amazonian Network of Georeferenced Socio-Environmental Information (RAISG) was chosen (see **Figure 4.9**). This delineation encompasses approximately 7.8 million km2 using biogeographical criterion, with the exception of Ecuador and Brazil where legal-administrative criteria are used. A variety of studies and datasets using this delineation exist (Gullison and Hardner, 2018), including a recent launch of 2000-2017 annual maps of LC by the MapBiomas Amazonia initiative (Proyecto MapBiomas Amazonía, 2019).

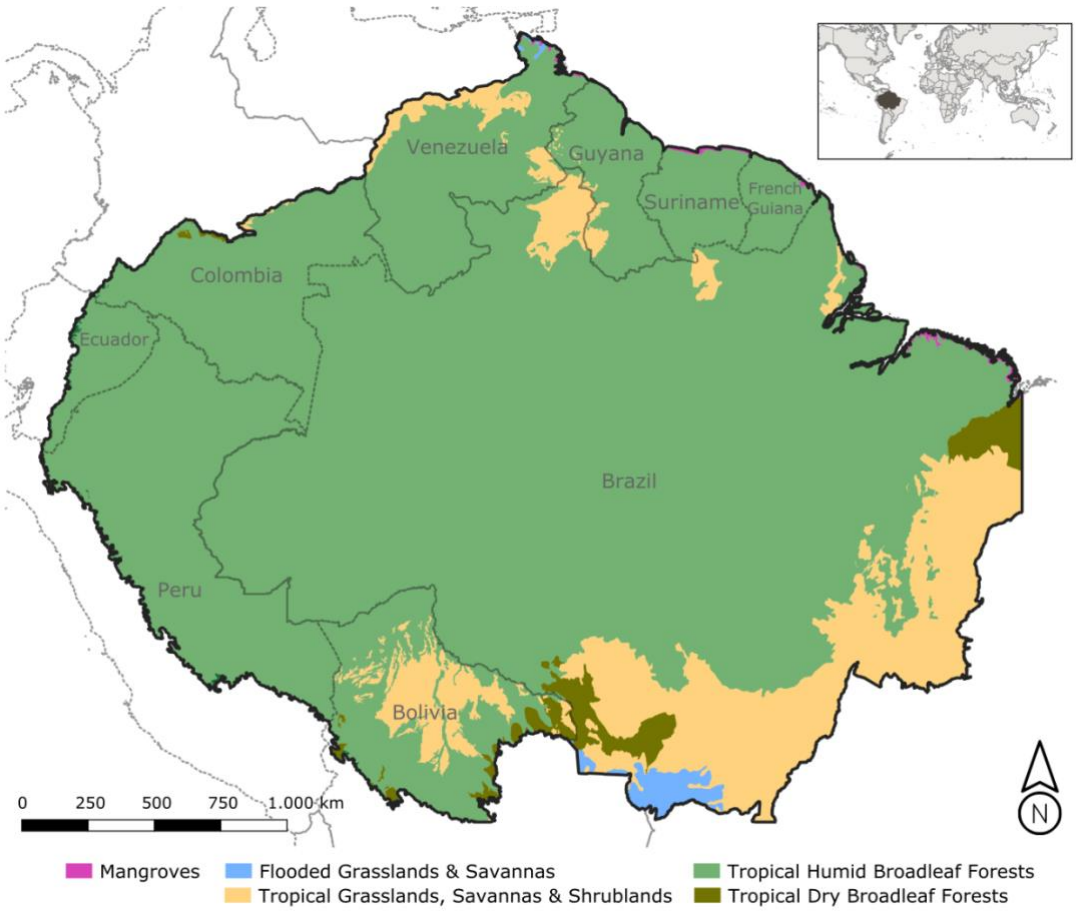


Figure 4.9 Amazon region delineated by RAISG (2012) including the WWF's major habitat types (Olson *et al*., 2001) and countries' boundaries.

4.5 Methodology

Figure 4.10 describes workflow with an aim to evaluate the performance of the proposed model according to calibration data derived from different pre-existing LC maps. This assessment is necessary to identify the optimal source of

reference data which produce satisfactory results to map LC change trajectories following deforestation.

The inputs of the workflow are:

1. Multi-temporal-spectral data (satellite data) (see **Section 4.5.2.1**).

2. A reference data with labelled samples (geographical positions for which LC is known) (see **Section 4.5.2.2**).

3. A reference dataset indicating the location of deforested areas (see **Section 4.5.5.3**).

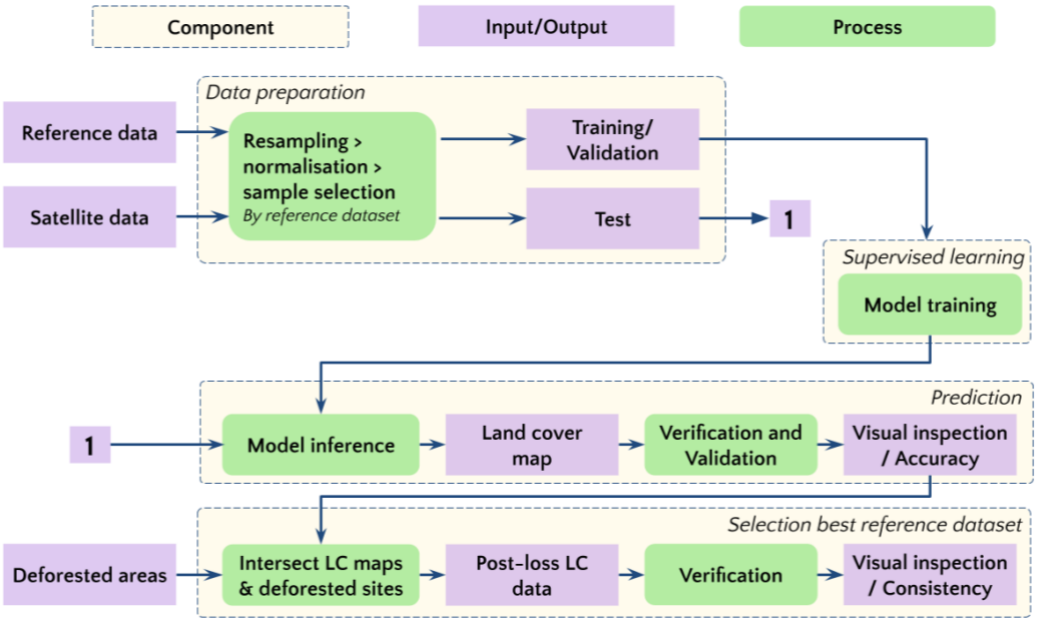


Figure 4.10 Diagram of the procedure to generate and validate LC maps and derived post-loss LC change trajectories.

Some pre-processing operations are conducted prior to model training including i) resampling reference and satellite data to the scale of interest, in this case at the spatial resolution of the reference deforestation dataset (see **Section 4.5.3.1**); ii) creation of new labelled datasets (also referred as hybrid datasets) derived from the group of pre-existing LC products (see **Section 4.5.3.2**) and iii) normalisation of the satellite data (see **Section 4.5.3.3**). A data partition is then performed according to a tiles-wise design in which tiles are distributed over specific regions of the study area (see **Section 4.5.4**). Following this, the

proposed supervised learning method is trained and evaluated for each labelled dataset according to a set of quantitative (i.e. overall accuracy) and qualitative aspects (i.e. visual inspection) (see **Section 4.5.5.1**). The model trained by the reference dataset is used to predict LC transitions over deforested sites (see **Section 4.5.5.2**). The visualisation of the observed trajectories is performed with a tool suited to analyse chronological LC data (see **Section 4.5.6**). The selection of the best labelled dataset is then made according to the outputs of the chain of processes in **Figure 4.10**.

4.5.1 Multitemporal Land Cover Classification Network

The Multitemporal Land Cover Classification (MTLCC) network, proposed by Rußwurm and Körner (2018b), employs a bidirectional sequential encoder using convolutional variants of RNNs for the task of multi-temporal land cover classification, as explained below. Although the network was initially designed for crop classification using Sentinel-2 data across a small geographical area in Germany, it has performed reasonably well in other locations with large and small training sets in Europe (Pelletier *et al*., 2019) and Africa (Rustowicz *et al*., 2019). For the purposes of this research, the MTLCC network covers key aspects related to operational aspects previously mentioned in **Section 4.3**. Firstly, it is an elegant framework to ingest and analyse large volumes of SITS by condensing their temporal dimension without further processing. Secondly, since the network assigns a given LC class by using the entire temporal series, atmospheric perturbations e.g. clouds can be seen and filtered as temporal noise according to the internal structures (so called gates) of the RNNs units (Rußwurm and Körner, 2018a, 2018b). This particular feature, which is further explored with the calibrated MTLCC network in **Chapter 5**, contributes in reducing some operations in the traditional processing chain of satellite images i.e. cloud filtering. Finally, the network and variants are flexible to multiple sources of satellite imagery including optical (Rußwurm and Körner, 2018b) and radar data (Rustowicz *et al*., 2019).

Figure 4.11 depicts the architecture of the MTLCC network. Inspired by sequence-to-sequence structures, the network encodes a sequence of satellite image observations $x = \{x_0, \cdots, x_T\}$ of $x \in \mathbb{R}^{h \times w \times d}$ into a fixed-length representation. The index T equals to the maximum length of the sequence and d

the input feature depth. As the number of satellite acquisitions can be variant and often shorter than T, particularly over large areas, the network handles missing sequence elements by padding them with a constant value being subsequently ignored at the iterative encoding steps. To eliminate bias towards the last observations in the sequence, the data are passed to the encoder in both sequential and reversed order. Network weights are shared between both passes.

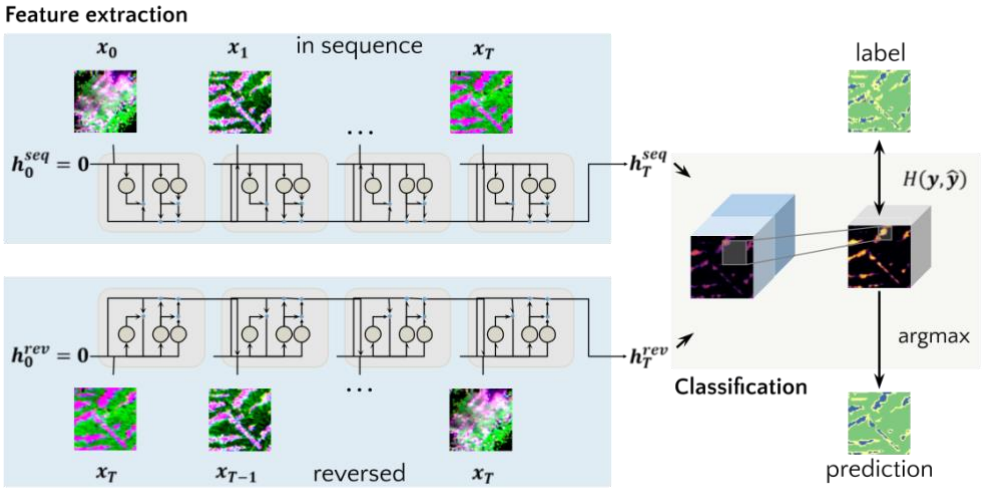


Figure 4.11 Scheme of the MTLCC model in this case using GRU cells to extract features in a bidirectional manner from an input sequence of satellite images $x = \{x_0, \cdots, x_T\}$ of $x \in \mathbb{R}^{h \times w \times d}$. This sequence is encoded to a representation $h_T = [\ h_T^{seq} \parallel h_T^{rev}]$. The observations are passed bidirectionally, in sequence (seq) and reversed (rev), to the encoder to eliminate bias towards recent observations. The concatenated representation h_T is then projected to softmax-normalized feature maps for each class using a convolutional layer. Adapted from Rußwurm and Körner (2018b).

Table 4.1 presents formulas of the types of convolutional variants of RNNs accepted by the MTLCC model. These variants are used by the network to account for neighbourhood relationships (spatial context) of satellite observations. For this work, the convolutional variant of GRU (ConvGRU) is preferred to the LSTM one (ConvLSTM) since it has yielded better and faster results in the field of LC classification using multi-temporal optical (Ienco *et al.*, 2017) and radar (Ndikumana *et al.*, 2018) remote sensing data. In place of separate vectors for long- and short-term memory, GRUs formulate a single, but more sophisticated, output vector. Furthermore, it uses only two gating units, namely the reset r_t and update u_t gates, to adaptively capture dependencies over multiple time scales and control how much memory and information is read and written to the output (Rußwurm and Körner, 2018b).

Table 4.1 Updated formulas of the convolutional variants of standard RNNs, long short-term memory (LSTM) cells and gated recurrent units (GRUs) accepted by the MTLCC model. A convolution between matrices a and b is denoted by $a * b$; $a \odot b$ is the element-wise multiplication (also known as the Hadamard product); and concatenation on the last dimension is marked by $[a \parallel b]$. The activation functions sigmoid $\sigma(x)$ and hyperbolic tangent $tanh(x)$ are used for non-linear scaling. Adapted from Rußwurm and Körner (2018b).

Gate	Variant	
	LSTM	GRU
Forget/Reset	$f_t \leftarrow \sigma([x_t \parallel h_{t-1}] * W_f + 1)$	$r_t \leftarrow \sigma([x_t \parallel h_{t-1}] * W_r)$
Insert/Update	$i_t \leftarrow \sigma(W[x_t \parallel h_{t-1}] * W_i)$ $j_t \leftarrow \sigma([x_t \parallel h_{t-1}] * W_j)$	$u_t \leftarrow \sigma([x_t \parallel h_{t-1}] * W_u)$
Output	$o_t \leftarrow \sigma([x_t \parallel h_{t-1}] * W_o)$ $c_t \leftarrow c_{t-1} \odot f_t + i_t \odot j_t)$ $h_t \leftarrow o_t \odot \tanh(c_t)$	$\hat{h} \leftarrow [x_t \parallel r_t \odot h_{t-1}] * W)$ $h_t \leftarrow u_t \odot h_{t-1} + (1 - u_t) \odot \tanh(h_t)$

When ConvGRU cells are used in the MTLCC network, the outputs $h_0^{seq}, h_0^{rev} \in \mathbb{R}^{h \times w \times d}$ are initialized with zeros. The outputs are derived from the latest-processed observation T in each reading, sequential h_T^{seq} or reverse h_T^{rev}. The concatenated final outputs $h_T = [h_T^{seq} \| h_T^{rev}]$ are the representation of the entire sequence and are passed to a convolutional layer for classification. A second convolutional classification layer projects the sequence representation h_T to softmax-normalized activation maps (so called confidence maps) $\hat{y}$ for n classes $h_T \in \mathbb{R}^{h \times w \times 2r} \rightarrow \hat{y}\ \mathbb{R}^{h \times w \times n}$. This layer is composed of a convolution with a kernel size of k_{class}, followed by batch normalization and a rectified linear unit (ReLU) or leaky ReLU non-linear activation function (Rußwurm and Körner, 2018b). The normalization is a standard technique for improving the speed, performance, and stability of artificial neural networks, particularly for RNNs (Laurent *et al.*, 2016). The model is trained with a cross entropy loss function, which is a measure for calculating the difference between two probability distributions, at each training step:

$$H(\hat{y}, y) = -\sum_i y_i \, log(\hat{y}_i) \qquad \text{Equation 4.1}$$

between the predicted activations $\hat{y}$ and reference labels y evaluates the prediction quality. To handle sparse labels, the loss is only calculated on image regions which have valid labels. The loss of all unknown pixels is set at these locations to zero.

The main tuneable hyper-parameters of the MTLCC network are the number of layers l, number of recurrent cells r and the sizes of the convolutional kernel k_{rnn} and the classification kernel k_{class}. Further exploration of these hyper-parameters is conducted in **Chapter 5**.

4.5.2 Datasets

In this section, the datasets used for the experiments are described. The datasets are composed of optical satellite images (input data) and reference LC data (so called labelled dataset). Observations from a single year, 2015, were used to train and evaluate the performance of the MTLCC model. This year was selected due to it being the only common year between all pre-existing LC products used as is detailed in **Section 4.5.2.2**.

4.5.2.1 Optical Satellite Data

The MODerate resolution Imaging Spectroradiometer (MODIS) data were the principal source of satellite imagery where surface reflectance time series were extracted. MODIS provides the electromagnetic radiation of the Earth in 36 spectral channels (C. O. Justice *et al.*, 2002). MODIS, onboard NASA's Terra (since 1999) and Aqua (since 2002) platforms, collect images at three different spatial resolutions: 250-m, 500-m and 1-km. The first seven bands are used to observe LC features in addition to cloud and aerosol properties.

For this work, MOD09Q1 and MOD09A1 products, at a spatial resolution of 250-m and 500-m were used. Both products are composites of 8-day periods which sum a total of 46 observations per year, except 2001 which has 45. Each pixel of the MOD09Q1 or MOD09A1 datasets contains the best possible Level 2 Gridded observation during an 8-day period and is usually selected based on high-observation coverage, low-view angle, absence of clouds or cloud shadow, and aerosol loading. MOD09Q1 provides surface reflectance in two bands (red = 620–670 nm; and Near-Infrared (NIR) = 841–876 nm). MOD09A1 also includes these bands plus five bands (blue = 459–479 nm; green = 545–565 nm; Short-Wave Infrared 1 (SWIR1) 86 = 1230–1250 nm; SWIR2 = 1628–1652 nm; SWIR3 = 2105–2155 nm) (C. O. Justice *et al.*, 2002).

Beside the surface reflectance bands, the day of the year was extracted and included as input band. This particular feature has been used by the original MTLCC network and other similar models (Rustowicz *et al*., 2019) to guide a better capture of the temporal aspects of satellite image observations.

4.5.2.2 Reference Data

Large extent LC mapping using satellite data is commonly challenged by its reliance on reference data. Some scholars have addressed this issue by rigorous *in-situ* field campaigns (Arvor *et al*., 2011) or by human interpretation, expert (C. Li *et al*., 2017; Tsendbazar *et al*., 2018) and non-expert (Fritz *et al*., 2009, 2017), of auxiliary satellite imagery (e.g. very high resolution images). In addition to these strategies, another emerging approach consists of automatically deriving calibration data using past or outdated LC datasets (Wessels *et al*., 2016; Inglada *et al*., 2017; Zhang and Roy, 2017; Cai *et al*., 2018; Kyere *et al*., 2019; Huang *et al*., 2015; Z. Sun *et al*., 2019; H. Wang *et al*., 2019).

Various pre-existing global LC products were explored to derive reference LC data. These products contain multiple classification schemes describing LC properties that are derived from observations spanning a year's input of satellite image data. For this study, five classification schemes extracted from three global LC datasets were assessed (see **Table 4.2**). These datasets represent reference data with a wide range of spatial resolutions (from 100-m to 500-m) which are relatively coarser than the resolution of deforested areas as detected by Terra-i (250-m). In addition to these datasets, the regional-tuned multi-temporal MapBiomas Amazonia product (Proyecto MapBiomas Amazonía, 2019) was used to compare the predictions generated by the MTLCC network trained with the coarse resolution reference LC data. Although this product does not have a global coverage as in the reference LC datasets, it contains annual per-pixel LC data from 2000 to 2017 with a high spatial resolution of 30-m. These characteristics make this particular dataset pertinent to the verification of modelled LC change trajectories across the study area. A detailed description of the methodology and reported accuracies of these datasets can be accessed in Appendix I.

Table 4.2 Characteristics of the different LC classification schemes and global products used in this research. The products are sorted according to their spatial resolution. Further info about each dataset/scheme is detailed in the sections below. The MapBiomas Amazonia maps [ID 6] are used only in the verification of the predictions of the MTLCC network trained with datasets derived from global LC products [IDs 1 to 5].

ID	Short name	Scheme	Number of classes (excl. no data)	Source (and version)	Resolution (m)	Temporal coverage (annual)
1	MODIS IGBP	IGBP	17	MCD12Q1 v6	500	2001-2017
2	MODIS FAO-LCCS1	FAO-LCCS	16	MCD12Q1 v6	500	2001-2017
3	MODIS FAO-LCCS2	FAO-LCCS	11	MCD12Q1 v6	500	2001-2017
4	ESA FAO-LCCS	FAO-LCCS	37	ESA-LC-CCI v2.0.7	300	2001-2015
5	CGLS-LC100 FAO-LCCS	FAO-LCCS	22	CGLS-LC100 v2	100	2015
6	MapBiomas Custom	Custom	9	MapBiomas Amazonia v1	30	2001-2017

4.5.3 Data preparation

4.5.3.1 Resampling

Scaling techniques were used prior to training to prepare the information required at the scale of interest, in this case at the spatial resolution of the target deforested dataset, Terra-i (250-m). Downscale/upscale operations were carried out according to the type of variable, continuous or categorical (Ge *et al*., 2019). For continuous values of reflectance data, all bands, except NIR and red which are provided by MOD09Q1 product at 250-m, were resampled using bilinear interpolation which is the same used by the original MTLCC model (Rußwurm and Körner, 2018b). This operation is made on the fly according to the library used to train the proposed model (see **Section 4.5.6**). For the labelled datasets, categorical downsampling was conducted to all pre-existing datasets, except for CGLS-LC100 FAO-LCCS which was upsampled due to its native resolution being finer (100-m) than the target resolution (250-m). The downscale consisted of assigning a given LC type to a pixel of 250-m according to the overlapping class of the coarser pre-existing LC map (MODIS and ESA). The upscaling procedure

used the dominant coverage in a pixel of 250-m from a group of overlapping pixels of 100-m, also known as majority resampling (Tao *et al*., 2015). Both operations were conducted through the cloud platform from which LC datasets were accessed (see **Section 4.5.6**).

4.5.3.2 Hybrid LC datasets

Inspired by previous work in LC mapping (Tsendbazar *et al*., 2015; Wessels *et al*., 2016), four additional labelled datasets (hereafter referred to as hybrid) were derived from the original LC datasets described in **Section 4.5.2.2** under the assumption they lead to greater accuracy in results. In this regard, four hybrid datasets were created. Two of them aimed to exploit the semantic and spatial aspects offered by the original LC data sources in a single year (2015). The remaining two exploit temporality features of two multitemporal land cover products, MODIS Land Cover Type product (2001-2018) and ESA-CCI (1995-2015). For the first group, the original classification schemes were modified to retain classes suited to the study of post-loss LC. For the second group, the hybrid datasets were generated by pixels with completely consistent LC types that had not changed over a considerable length of time. Only the stable pixels from 2001 to 2015, which is the period matched by MODIS Land Cover Type and ESA-CCI products, were retained for deriving the hybrid datasets per product. **Figure 4.12** illustrates a representation of the four hybrid datasets and main procedures for their generation.

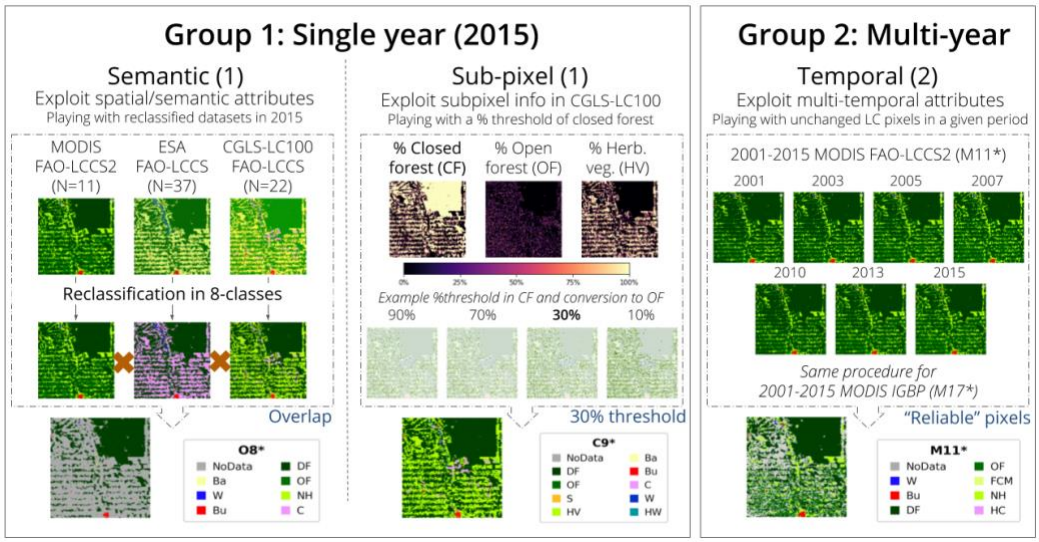


Figure 4.12 Illustration of the hybrid datasets derived from global LC products described in Section 4.5.2.2. The first group consists of maps generated

Single year derived hybrid datasets

For this group, the original classification schemes were modified to retain classes suited to the study of post-loss LC. Consequently, non-relevant classes in the original LC datasets such as *Snow and Ice* and *Moss and lichen* were removed. Furthermore, forest/tree like classes, which are disaggregated in most of the reference datasets by specific types, were merged into a single class i.e. closed like forest into closed forest and open like forest into open forest.

- MODIS/ESA/CGLS LC-Custom (O8*)

The O8* hybrid dataset aims to integrate some LC global products such as MODIS FAO-LCCS2 (N=11), ESA FAO-LCCS (N=37) and CGLS-LC100 FAO-LCCS (N=22) datasets. To reflect this condition, the map legends were harmonized to eight general LC classes (see **Table 4.3**). The harmonization followed the approach of previous work by Tsendbazar *et al*. (2017), which provides a guide for translating different legends of global LC maps into a common legend. After harmonizing the original datasets to the nine general LC classes, the reclassified maps were overlapped. Those pixels which do not overlap in all three datasets were classified as *NoData*.

Table 4.3 Relationships of the hybrid dataset eight general LC classes in different land classification system. The original LC class code by dataset was reclassified from 1 to the number of classes. *, **, *** stand for the associated LC class related to the codes of MODIS FAO-LCCS2, ESA FAO-LCCS and CGLS-LC100 FAO-LCCS datasets, respectively.

General LC class	MODIS FAO-LCCS2*	ESA FAO-LCCS**	CGLS-LC100 FAO-LCCS***
Dense Forests	5	7-9, 11, 12, 14, 15, 17	1-6
Open Forests	6	10, 13, 16, 18, 29, 30	7-12
Shrublands	11	20-22	13
Natural Herbaceous	8	19, 23, 31	14-15
Cropland	7, 9, 10	1-6	18
Urban and Built-up Lands	4	32	19
Barren	1	25-28, 33-35	17
Water Bodies	3	36	21, 22
No Data	0, 2	24, 37	16, 20

** 0: 'No data', 1: 'Barren', 2: 'Permanent Snow and Ice', 3: 'Water Bodies', 4: 'Urban and Built-up Lands', 5: 'Dense Forests', 6: 'Open Forests', 7: 'Forest/Cropland Mosaics', 8: 'Natural Herbaceous', 9: 'Natural Herbaceous-Croplands Mosaics', 10: 'Herbaceous Croplands', 11: 'Shrublands' ** 0: 'No data', 1: 'Cropland rainfed', 2: 'Cropland rainfed - Herbaceous cover', 3: 'Cropland rainfed - Tree or shrub cover', 4: 'Cropland irrigated or post-flooding', 5: 'Mosaic cropland (>50%) / natural vegetation*

*(tree/shrub/herbaceous cover) (<50%)', 6: 'Mosaic natural vegetation (tree/shrub/herbaceous cover) (>50%) / cropland (<50%)', 7: 'Tree cover broadleaved evergreen closed to open (>15%)', 8: 'Tree cover broadleaved deciduous closed to open (>15%)', 9: 'Tree cover broadleaved deciduous closed (>40%)', 10: 'Tree cover broadleaved deciduous open (15-40%)', 11: 'Tree cover needleleaved evergreen closed to open (>15%)', 12: 'Tree cover needleleaved evergreen closed (>40%)', 13: 'Tree cover needleleaved evergreen open (15-40%)', 14: 'Tree cover needleleaved deciduous closed to open (>15%)', 15: 'Tree cover needleleaved deciduous closed (>40%)', 16: 'Tree cover needleleaved deciduous open (15-40%)', 17: 'Tree cover mixed leaf type (broadleaved and needleleaved)', 18: 'Mosaic tree and shrub (>50%) / herbaceous cover (<50%)', 19: 'Mosaic herbaceous cover (>50%) / tree and shrub (<50%)', 20: 'Shrubland', 21: 'Shrubland evergreen', 22: 'Shrubland deciduous', 23: 'Grassland', 24: 'Lichens and mosses', 25: 'Sparse vegetation (tree/shrub/herbaceous cover) (<15%)', 26: 'Sparse tree (<15%)', 27: 'Sparse shrub (<15%)', 28: 'Sparse herbaceous cover (<15%)', 29: 'Tree cover flooded fresh or brakish water', 30: 'Tree cover flooded saline water', 31: 'Shrub or herbaceous cover flooded fresh/saline/brakish water', 32: 'Urban areas', 33: 'Bare areas', 34: 'Consolidated bare areas', 35: 'Unconsolidated bare areas', 36: 'Water bodies', 37: 'Permanent snow and ice' *** 0: 'Non-reliable', 1: 'Closed forest, evergreen needleleaf', 2: 'Closed forest, deciduous needleleaf', 3: 'Closed forest, evergreen broadleaf', 4: 'Closed forest, deciduous broadleaf', 5: 'Closed forest, mixed', 6: 'Closed forest, unknown', 7: 'Open forest, evergreen needleleaf', 8: 'Open forest, deciduous needleleaf', 9: 'Open forest, evergreen broadleaf', 10: 'Open forest, deciduous broadleaf', 11: 'Open forest, mixed', 12: 'Open forest, unknown', 13: 'Shrubs', 14: 'Herbaceous vegetation', 15: 'Herbaceous wetland', 16: 'Moss and lichen', 17: 'Bare / sparse vegetation', 18: 'Cultivated and managed vegetation/agriculture (cropland)', 19: 'Urban / built up', 20: 'Snow and Ice', 21: 'Permanent water bodies', 22: 'Open sea'*

- CGLS-LC100 LC-Custom (C9*)

In the original resampling of the CGLS-LC100 dataset, partially deforested areas at the extent of 250-m pixels are susceptible to get hidden. For instance, 100-m CGLS-LC100's pixels of the class *Closed Evergreen Broadleaved Forest (CEBF)* are resampled and assigned to the same category even other LC classes are present at the extent of the 250-m MODIS pixel. Consequently, there is a loss of information regarding the level of heterogeneity at sub-pixel level. This aspect gets crucial as the Terra-i system needs a minimum threshold of 33% of change to define a MODIS pixel as deforested as was previously demonstrated in **Chapter 3**.

To exploit the subpixel information offered by the CGLS-LC100 dataset, a hybrid LC dataset was derived to reflect small-scale disturbances which are misled in the resampled CGLS-LC100 dataset from 100-m to 250-m. The procedure mainly consisted in deriving maps based on the proportion of *Closed forest* pixels computed at the extent of MODIS 250-m. The mapped proportions, which go from 0 to 1, were grouped in ten ranges from 0.1 to 1 by 0.1. These ranges were used to replace at a given fraction of *Closed forest* with the dominant non-closed forest class per pixel, which in most cases was represented by *Open forest* pixels.

The approach above was tested over two contrasting tiles of 384 pixels x 384 pixels of MODIS 250-m. These tiles are located near the city of Pucallpa in the Ucayali, Peru and the municipality of Porto dos Gauchos in Mato Grosso, Brazil, both areas located inside the Amazon with different sizes and magnitude of forest loss. While the area in Peru consists of a large proportion of small- and medium-scale deforestation (Finer and Novoa, 2016), the area in Brazil is mostly dominated

by large-scale deforestation (Maus *et al.*, 2016). **Figure 4.13** displays spatial changes of aggregated open like forest pixels which replaced most of the closed forest according to increasing levels of fraction assessed in both areas. It is evident the relationship of increasing levels of fraction of *Closed forest* and the number of *Open forest* pixels being more pronounced in the area in Ucayali than Mato Grosso. In particular, the different levels of fraction show an exponential increase of pixels with *Open forest* class when fraction values are greater than 0.7. This value was therefore used as a threshold to create the new dataset. It is expected that the new dataset favours a better mapping of small-scale forest loss in areas with heterogeneous landscapes as was observed in the region in Peru.

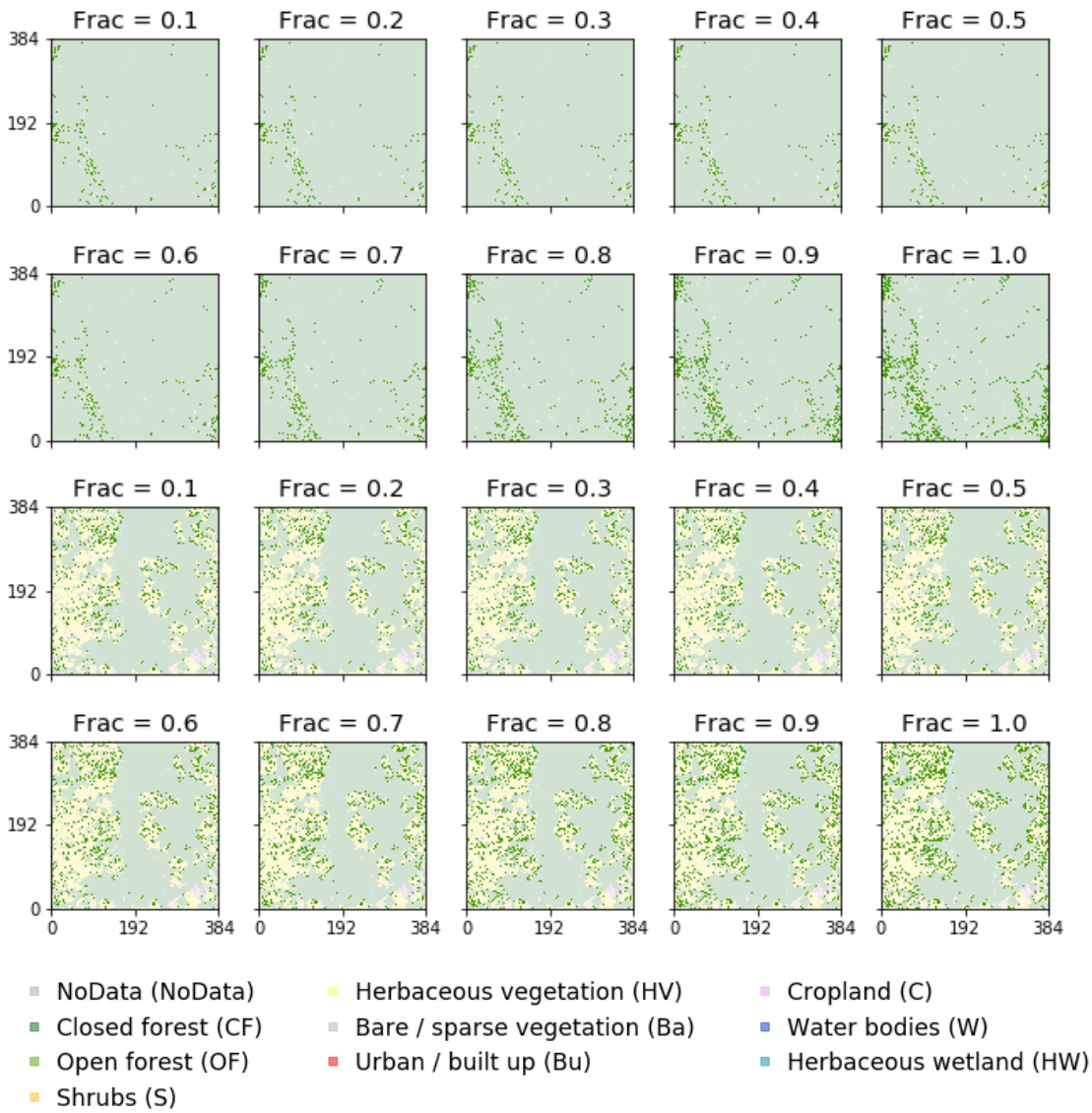


Figure 4.13 Spatial distribution of *Open forest* pixels (highlighted by a green colour) according to increasing levels of the fraction of closed forest determined at the spatial extent of MODIS 250-m.

Figure 4.14 shows the changes per land cover class in terms of frequency according to different levels of fraction of disturbance. The reduction of *closed forest* is explained mostly by the increase of *Open forest* class. Other non-forest classes keep the same number of pixels, except for *Bare* (Ba) in both areas, and *NoData*, *Shrubland* (S), *Herbaceous vegetation* (HV) and *Water* (W) in Ucayali. The increase of these non-forest classes might be explained they represent a higher fraction at subpixel level than *Open forest*. It is important to highlight these replacements were less common than *Closed forest* replaced by *Open forest.*

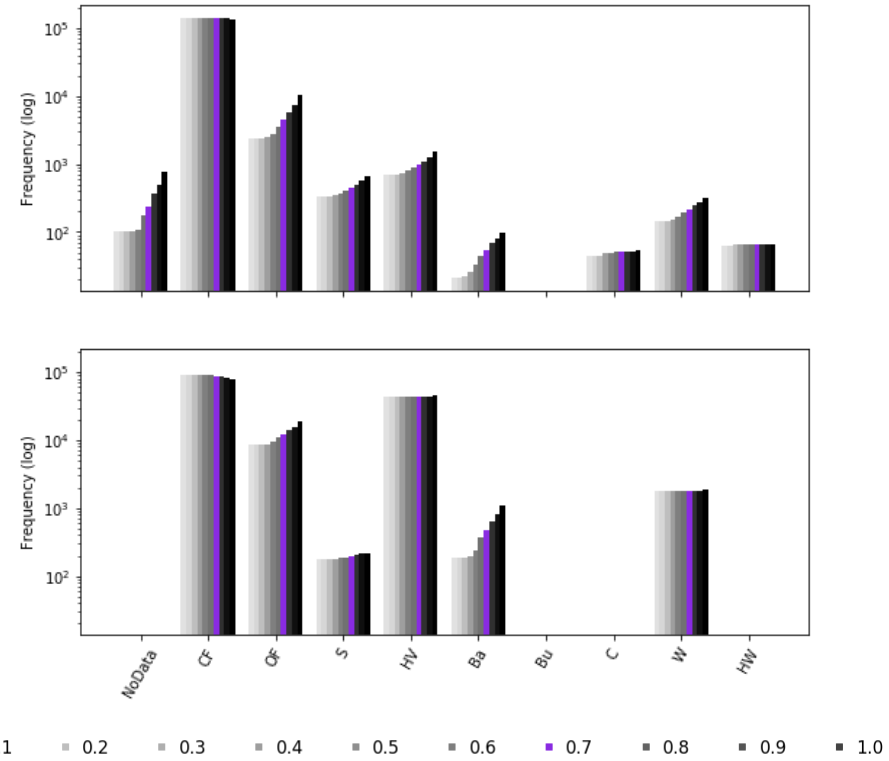


Figure 4.14 Frequencies in logarithmic scale of LC classes in the areas in Ucayali (above) and Mato Grosso (below) according to increasing proportions of disturbance determined from *closed forest*. The complete name by class is shown at the legend of **Figure 4.13**.

Temporal-driven derived hybrid datasets

These datasets were generated by pixels with LC types that had not changed from a considerable length of time. This particular procedure was only applied for two out of three LC schemes in the MODIS Land Cover Type product (2001-2018). Only the stable pixels from 2001 to 2015, which end year matches with the year of analysis of this work, were considered. These products are referred as 2001-2015 MODIS IGBP (M17*), 2001-2015 MODIS FAO-LCCS2 (M11*). The same procedure was also applied to the ESA-CCI FAO-LCCS dataset however it was

found lower performance than the MTLCC network trained using hybrid datasets generated from the MODIS Land Cover Type product.

4.5.3.3 Feature normalisation

A min-max normalisation per type of feature was applied prior to model training. Whilst other types of normalisation commonly used in remote sensing e.g. z-normalisation are prone to loss the temporal trend and significance of the magnitude of SITS (Pelletier *et al*., 2019), the min-max normalisation, which performs a subtraction of the minimum, then a division by the range, i.e. the maximum minus the minimum, preserves the relationships of the original data (Han *et al*., 2011). For this work, the min-max values were derived from the datasets' metadata which commonly indicate the range of valid values. Hence, in the MODIS product, this range goes between -100 to 16000. The normalised values were also scaled between 0 and 1. This scale ensured all features have the same probability of being taken during the learning process. The use of a global max and min might also facilitate the training and application of the model across regions different to the study area as other scholars have proposed (Pelletier *et al*., 2016) and shown in **Chapter 5**.

4.5.4 Data partition

The study area was subdivided into rectangular tiles of 384 pixels x 384 pixels of MODIS data at a spatial resolution of 250-m. The tile size is a multiple of the patch size of 24 pixels x 24 pixels, which was the size used to train the MTLCC architecture described in **Section 4.5.1**. Not all tiles covering the study area were considered for training and assessing the model performance. Those tiles with a large proportion (>70%) of *Evergreen broadleaf forest* class according to the IGBP classification from MODIS Land Cover Type product (MCD12Q1) in 2015 were discarded. This procedure reduced the dominance of this particular class, or its equivalent class in the other datasets. This ensured our focus on dynamic land use environments and avoided domination by static forest pixels over the majority of the region This reduction helps the MTLCC network classifier to distinguish other classes which are less dominant.

The input MODIS images and reference data were randomly split at tile level to generate partitions for network training, validation and model testing using a ratio

of 4:1:1 (see **Figure 4.15**) as recommended by previous work using the model (Rußwurm and Körner, 2018b). The tile-wise spatial separation enforces independence of the data partitions and support bias-reduced predictive performance results (Schratz *et al.*, 2018). The spatial-based partition design also attempts to best preserve the relative percentages of all classes, allowing for consistent class balances in all splits as shown in **Figure 4.16**.

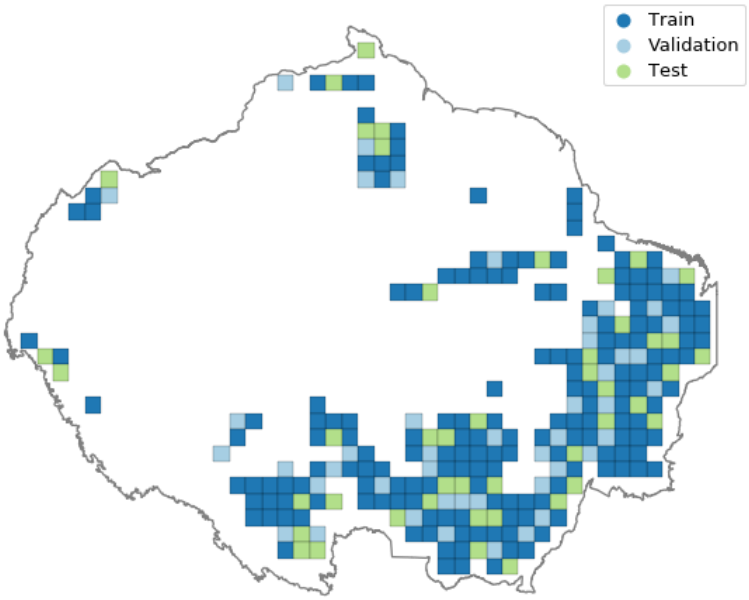


Figure 4.15 Illustration of an input/reference data partition using non-overlapping 384 pixels x 384 pixels tiles for training (blue), validation (light blue), and test (light green).

According to the split ratio, for 2015 some 26.1, 6.6 and 6.6 millions of pixels, represented in 267 tiles of 384 pixels x 384 pixels, were used for training, validation and testing of the network assessed. To evaluate the performance of the MTLCC network according to different datasets, the splitting of training and validation tiles was repeated five times as suggested in the literature (Lyons *et al.*, 2018). Hence, each algorithm or experiment was evaluated five times on different train/validation splits (also called folds).

4.5.5 Performance evaluation

The quality of the predictions by the MTLCC network according to different pre-existing and hybrid LC datasets was assessed over test tiles (45 in total) of 384 pixels x 384 pixels. A two-stage assessment approach was conducted to determine the most suitable labelled dataset to the study of post-loss LC change over Terra-i's areas. For the first stage, traditional metrics for assessing machine

learning classifiers (e.g. overall accuracy, recall, precision), f-score, and others with land change and remote sensing research foci (e.g. kappa, quantity disagreement and allocation disagreement) were computed according to the test partition from the labelled datasets (see **Section 4.5.5.1**).

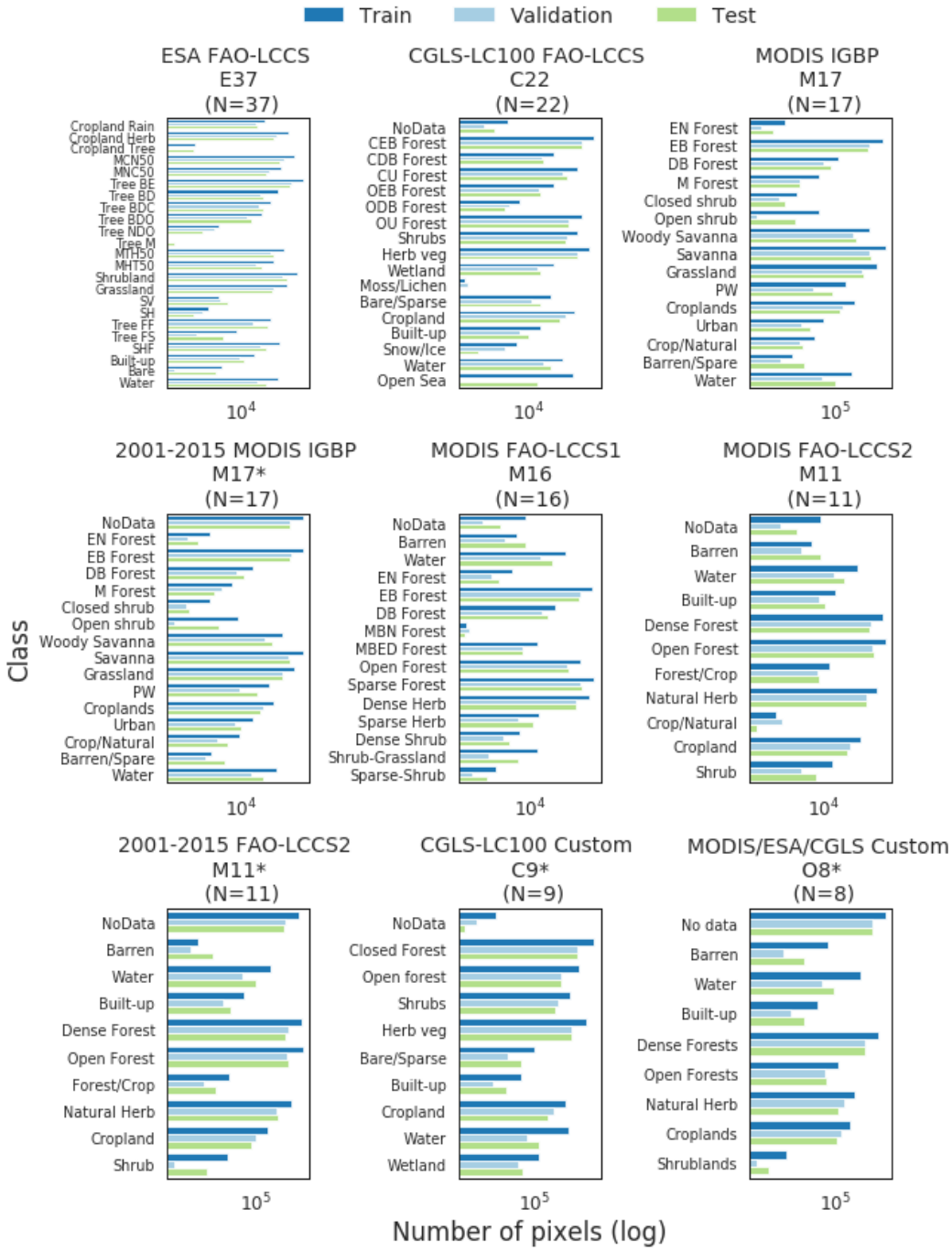


Figure 4.16 Class distribution of the training, validation and test partitions for the LC pre-existing and hybrid (*) reference datasets, ordered by total number of classes, assessed with the MTLCC network. Classes are represented as either; 1) a single word indicating the cover type; 2) an acronym indicating the cover type, in particular for the ESA-CCI dataset and forest/tree like cover types. Further details of the classes and respective acronyms and total number of pixels per dataset can be found in **Appendix II**.

A second stage consisted of using the trained models in 2015 to make predictions for all available years of MODIS imagery, from 2001 to 2018, across seven tiles sampled from the population of test tiles. The observed LC predictions were masked using Terra-i's deforested pixels. The post-loss LC data by trained model, which represents a given labelled dataset, was visually inspected i) by creating charts and maps of the temporal and spatial distribution of the LC predictions; and ii) by comparison with LC observations derived from MapBiomas Amazonia. From this visual verification, the potential labelled datasets for the study of post-loss LC were examined through a quantitative evaluation using an independent validation set as recommended by Olofsson *et al*. (2013) (see **Section 4.5.5.2**).

4.5.5.1 Confusion matrix and derived metrics

The confusion matrix is a double entry table where row entries are the actual classes (reference data) and column entries are the predicted classes (see **Table 4.4**). Each cell of the table usually contains the number of elements (pixels in pixel-based classification) of the row class predicted by the classifier as belonging to the column class. In this matrix, the diagonal elements represent the number of correctly classified pixels, also called true positive and true negative samples. The matrix also provides misclassification represented by the number of false positive and false negative samples.

Table 4.4 Confusion matrix for a binary classification problem.

	Predicted Class	**Predicted No-Class**
Actual Class	True Positive (TP)	False Positive (FP)
Actual No-Class	False Negative (FN)	True Negative (TN)

Five metrics were computed from the confusion matrix of the multi-class classifications generated by the trained MTLCC models according to pre-existing and hybrid datasets. The equations of these metrics are presented in **Table 4.5**. Overall accuracy (OA) is the sum of the diagonal elements divided by the sum of all elements of the confusion matrix. Precision and recall (also denoted in remote sensing as user and producer accuracy) are metrics capturing the errors of commission and omission. The F1-score is the harmonic mean between precision and recall. Being informative for imbalanced classes, the Kappa coefficient (κ) was also computed according to its feature of providing a less optimistic value than the confusion matrix derived indices. The κ coefficient is a statistical measure of interrater

agreement taking into account the agreement occurring by chance. All metrics were computed per class, and then averaged across all classes to give the reported average metric.

Table 4.5 Measures for performance assessment of classification derived from the confusion matrix. Acronyms are described as follows: TP are true positive, FP– false positive, FN false negative, TN– true negative, p_e- the hypothetical probability of chance agreement.

Metric	Formula	Range of values
Overall accuracy (OA)	$\frac{TP + TN}{TP + TN + FP + FN}$	[0, 1]
Precision	$\frac{TP}{TP + FP}$	[0, 1]
Recall	$\frac{TP}{TP + FN}$	[0, 1]
F1-score	$2 \times \frac{precision \times recall}{precision + recall}$	[0, 1]
Kappa (κ)	$\frac{OA - p_e}{TP + FN}$	[0, 1]

In addition to the metrics the above table, the complement of the total accuracy, or the total error was decomposed into quantity (or area) disagreement and allocation disagreement. Both metrics have been commonly used in the remote sensing community as an alternative to Kappa (Salk *et al*., 2018). The quantity disagreement measures the difference between the reference and classified maps attributable to the difference in proportions of categories. The allocation disagreement is defined as the difference between the reference and classified maps attributed to differences in matching spatial allocation of categories. Formulas for these metrics on category-specific and map-wide levels are detailed in Pontius and Millones (2011) and Pontius and Santacruz (2014). The values for both metrics ranges from 0 (no disagreement) to 100 (high disagreement).

4.5.5.2 Cross-year verification

This stage consisted of qualitative and quantitative assessments of annual predictions by the MTLCC network models trained using different labelled datasets with a particular focus on the study of long-term post-loss LC change. Due to demanding detailed information derived from the visual inspection of available high-resolution satellite imagery (e.g. Landsat), this assessment was conducted over a single tile sampled per MODIS scene across the area of study (see **Figure 4.17**).

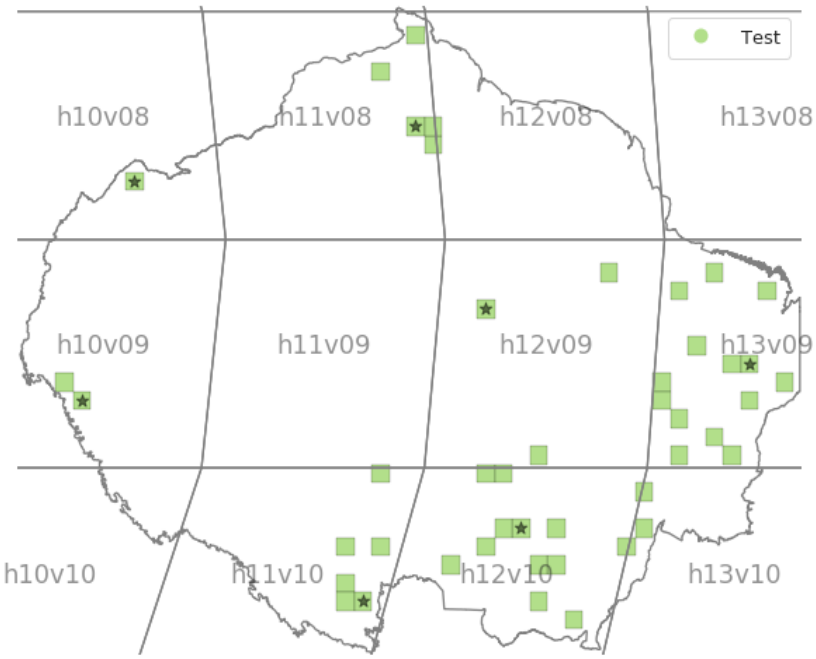


Figure 4.17 Distribution of the sampled tiles (★) for the cross-year evaluation in the study area. A total of 12 MODIS scenes, denoted as hXXvXX, cover the study area. The distribution of the sample tiles ensured that different landscapes are analysed across the study area ranging from large scale clearings in Mato Grosso, Brazil (e.g. MODIS scene h12v10) to small and medium-scale clearings in the Western Amazon (e.g. MODIS scenes h10v08 and h10v09).

4.5.5.3 Pre-processing

Prior to computing and assessing post-loss LC change per sampled tile, the initial predictions of all trained MTLCC network models and the reference dataset of deforested areas (Terra-i) were pre-processed. For the initial predictions, a smoothing time-series rectification technique proposed by Abercrombie and Friedl (2016) was used to reduce spurious year-to-year LC changes. The technique, which was successfully applied to enhance the consistency of the MODIS Land Cover product (MCD12Q1) (Sulla-Menashe *et al*., 2019), uses Hidden Markov models (HMMs). HMMs adjust posterior marginal probabilities for each class at each pixel according to initial probabilities and a transition class probabilities matrix derived from the supervised model and expert-knowledge, respectively. For this work the initial probabilities were obtained from the confidence maps per class derived from trained MTLCC networks. The transition matrix defines the probability of pixel change between LC classes from one year to another. The values of the transition matrix can be filled according to different categories as defined by Hermosilla *et al*. (2018) such as likely (0.90), probable (0.10), possible (0.025), or not likely (0.001). For this work all values were set to

0.10 (or probable) as it was difficult to define other categories for all possible changes between LC classes per labelled dataset. **Figure 4.18** depicts the effect of the smoothing pre-processing over a set of 2001-2018 LC maps predicted over a representative tile by one of the trained models using the 2001-2015 stable MCD12Q1 IGBP labelled dataset.

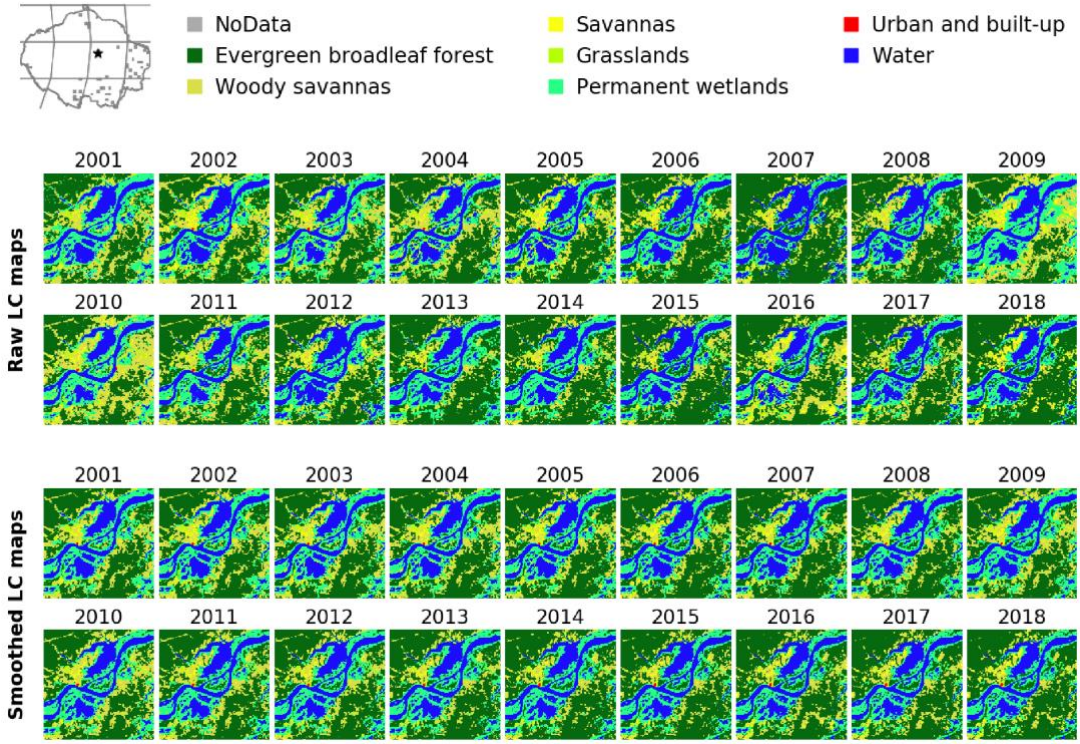


Figure 4.18 Raw and pre-processed (i.e. smoothed) 2001-2018 LC maps predicted by a trained model using 2001-2015 stable MCD12Q1 IGBP hybrid dataset over a sampled tile. Spurious changes between certain years in the raw predictions are reduced by the smoothing algorithm (see for example from LC maps from 2007 to 2011).

Regarding the Terra-i dataset (version 2004_01_01_to_2019_06_10), the pre-processing aimed at avoiding noise (pixels wrongly detected or non-related with a forest disturbance) by i) discarding pixels located within a 1-km buffer created from water bodies according to the Global Surface Water (GSW) dataset (Pekel *et al*., 2016), in which most of the cases represent changes in water level rather than deforestation; and ii) retaining only contiguous deforested areas >2 pixels (12.5 ha) as areas smaller than this are mostly false positives by the Terra-i system (Tang *et al*., 2019). Examples of 1-km buffer water mask as well as original and pre-processed Terra-i deforested data are depicted in **Figure 4.19**.

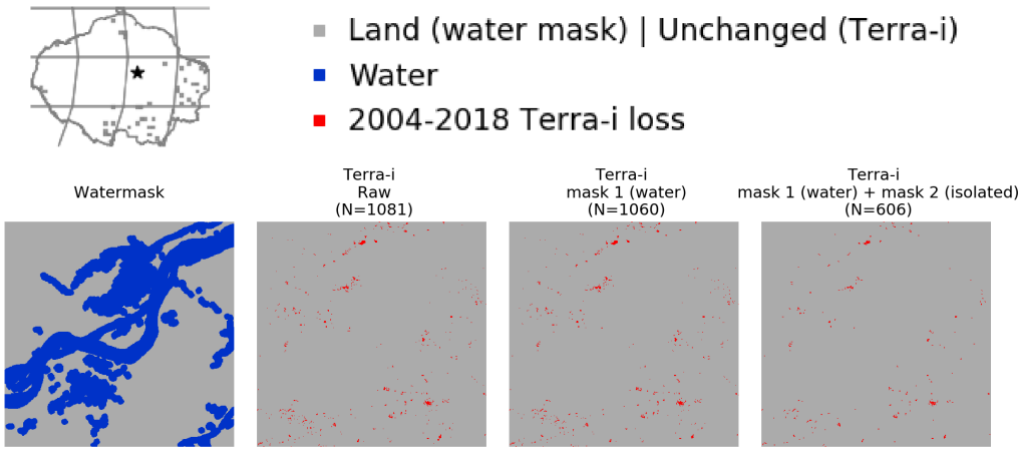


Figure 4.19 Example of the 1-km buffer water mask, raw and pre-processed 2004-2018 Terra-i data according to two masks (water and discarding isolated pixels) over a representative tile. The total number of Terra-i pixels prior to and after pre-processing is denoted in brackets as N.

4.5.5.4 Analysis

From the smoothed LC predictions and pre-processed Terra-i data, the population of target deforested pixels for validating post-loss LC change was derived from the single peak year of deforestation between 2004 and 2010 per representative tile in order to reduce processing time. This six-years window of deforestation ensured a considerable number of post-loss LC sequences, 8 to 14 years, can be analysed to the latest date of LC prediction, in this case 2018. Additionally, under the assumption that forest loss events might occur before the start year of detection by Terra-i, 2004, the LC predictions were projected from 2001 across all target pixels irrespective of when they were deforested. Land cover change before 2001 is not considered in this analysis since satellite image time series data are not available for that period. After generating the projected LC predictions, the selection of the potential suitable labelled datasets for the study of post-loss LC change using the MTLCC network was firstly conducted by visually comparing the spatial and temporal consistency of the LC predictions by the trained model against 2001-2017 regional-tuned observations of the MapBiomas Amazonia dataset. The MapBiomas dataset was also masked by the population of target deforested areas.

The candidate datasets for the study of post-loss LC change observed from the above qualitative assessment as well as the results of the single year performance assessment as described in **Section 4.5.5.1** were quantitatively validated using an

independent validation dataset. This independent dataset was collected over the most common LC sequences defined as those located at quantile of 95% according to the frequency values computed for all possible sequences per representative tile. This quantile value was defined due to the fact that it discriminates the most common sequences, whose proportions sum up to 60% of the total possible sequences across the areas assessed. The sampling strategy consisted of randomly collecting 1% of these most common sequences weighted per number of observations per sequence per representative tile. The sample per potential suitable labelled dataset was verified separately using visual inspection of high-resolution annual composites of Landsat 7 and 8 from 2001 to 2017. These composites were generated following the same procedure described in **Section 3.4.1.3** for exploring LC and LU over Terra-i's field campaigns' sites. For each pixel, which represented a given sequence, the LC type was classified annually according to the observed Landsat composites. The validation set generated by the potential labelled dataset was then used to compare with actual LC predictions. Metrics derived from the confusion matrix as previously depicted in **Table 4.5** were computed and analysed per year. Additionally, the classification of all years was aggregated and assessed as a single period, 2001-2017.

4.5.6 Software and implementation

The hardware and software used according to the main steps in this study are presented in **Table 4.6.** For the implementation of the MTLCC network, and data preparation scripts, python 3.6 was used. The MTLCC network was trained using TensorFlow version 1.7 with scripts publicly available in a repository provided by Rußwurm and Körner (2018b). Some adaptations and improvements of the original repository were made as it was designed for Sentinel-2 imagery. For instance, the model was adapted to read MODIS data in the native resolutions of their bands, 250-m (red and NIR) and 500-m (blue, green, SWIR1, SWIR2, SWIR3). The resampling operations of the bands at 500-m to 250-m and generation of hybrid datasets were made on-the-fly using the Dataset API of TensorFlow. Furthermore, the original implementation of the MTLCC network also presented a coding error related to adding the year as input band. The year was dropped as input due to the fact it was generating reductions to the prediction capability of the MTLCC network, in particular for years different to the trained year.

According to a previous 'trial and error' experimentation (Coca-Castro *et al.*, 2019), using the hardware resources available as described in **Table 4.6**, it was found that optimal results for training the MTLCC network were obtained based on Gated Recurrent Unit (GRU) in a bidirectional manner, single layer and 64 recurrent cells. The size of convolutional and classification kernels was set to 3 x 3 as was proposed by Rußwurm and Körner (2018b). The batch size was set to 32 and models were trained until 30 epochs when the models start converging. The computation of performance metrics as described in **Section 4.5.5.1** were computed using the library *scikit-learn* version 0.19.1.

Regarding the analysis of post-loss LC sequences, the R software v.3.4.3 was used, in particular the package *traMinerR* version 2.0-11.1 (Gabadinho *et al.*, 2011). This package is suited to the analysis of sequential data and has been used to assess chronological LC/LU data in tropical settings (Mas *et al.*, 2019). Different to the *lucCalculus* package in R used to assess LU trajectories in **Chapter 3** (see **Section 3.5.4**), the *traMineR* provided considerably more additional options to visualise and analyse large volumes of chronological LC predictions and thus it was suitable for the study of post-loss LC change trajectories.

Table 4.6 Lists of hardware and software used in this study organised by step.

Step	Hardware and/or software	Specifications
Data generation and creation of the validation set	Google Earth Engine (GEE) Google Cloud Platform (GCP)	GEE is a cloud-based platform used to access and download input datasets (see **Section 4.5.2**) exported as GZ TFrecords to GCP.
Data pre-processing and model training	Server	A server was used for data storage and preparation of input data as well as model training. Its main features are: Memory: 128 GB Processor: Intel Xeon E5-2667 v4 3.20 GHz×16 Operating system: Windows Server 2016 X64 GPU: (1) NVIDIA Tesla M60
Post-loss LC visualisation and validation	traMinerR Google Earth Engine (GEE) Google Forms	traMinerR was used to the visualisation and analysis of post-loss LC. The validation of post-loss LC change by the trained model was conducted in GEE. An independent validation set was created by visualising annual composites of Landsat (see **Section 4.5.5**) and annotating the corresponding LC per year in a Google form as created in Chapter 3 (see **Section 3.4.3**).

4.6 Results

4.6.1 Single year performance over the test partition

The summary of the performance of the MTLCC network trained using original and hybrid labelled datasets according to their respective test partitions is reported in **Table 4.7**. Overall, the models trained using hybrid datasets performed better than the original. In particular, the hybrid dataset MODIS/ESA/CGLS (O8*) returned the highest overall accuracy, 92.35±0.48 over the 45 test tiles. Other hybrid datasets related to multi-year stable pixels from MODIS dataset, 2001-2015 MODIS IGBP (M17*) and 2001-2015 MODIS FAO-LCCS2 (M11*) provided a notable gain in OA compared to the original datasets.

Table 4.7 Averaged values (± one standard deviation) of pixel-wise accuracy metrics over five folds of the MTLCC network model trained according to original and hybrid (*) labelled datasets described in **Section 4.5.2**. The best values per metric are highlighted in bold.

Labelled dataset	Overall accuracy (OA) (%)	Quantity Disagreement (%)	Allocation Disagreement (%)	Kappa (κ)
ESA FAO-LCCS (E37)	66.46 ± 0.26	13.61 ± 1.69	19.92 ± 1.87	0.57 ± 0.0
CGLS-LC100 FAO-LCCS (C22)	63.37 ± 1.68	7.38 ± 2.26	29.24 ± 0.89	0.51 ± 0.03
MODIS IGBP (M17)	79.96 ± 0.89	7.61 ± 2.15	12.43 ± 1.53	0.72 ± 0.01
2001-2015 MODIS IGBP (M17*)	86.36 ± 0.64	5.34 ± 0.39	8.31 ± 0.64	0.81 ± 0.01
MODIS FAO-LCCS1 (M16)	81.24 ± 0.59	7.23 ± 3.21	11.52 ± 2.64	0.74 ± 0.01
MODIS FAO-LCCS2 (M11)	84.92 ± 0.85	2.89 ± 1.42	12.19 ± 1.06	0.77 ± 0.01
2001-2015 MODIS FAO-LCCS2 (M11*)	89.17 ± 1.00	4.81 ± 1.56	6.02 ± 0.85	**0.83 ± 0.02**
CGLS-LC100 Custom (C9*)	69.48 ± 2.01	5.58 ± 1.06	24.94 ± 2.09	0.56 ± 0.03
MODIS/ESA/CGLS Custom (O8*)	**92.35 ± 0.48**	**3.68 ± 1.03**	**3.97 ± 1.44**	0.74 ± 0.01

The observed values of the disagreement related metrics, quantity and allocation, show they are more informative than OA and κ. For instance, while κ values of MODIS/ESA/CGLS (O8*) and MODIS LC-FAO (M16) are equal (κ=0.74±0.01), their values of quantity and allocation disagreement are not. **Figure 4.20** displays per-class F1-score values for three hybrid labelled datasets, MODIS/ESA/CGLS

(O8*), MODIS FAO-LCCS2 (M11*) and CGLS-LC100 (C9*). These datasets have lower total values of quantity disagreement and allocation disagreement than their corresponding original LC datasets.

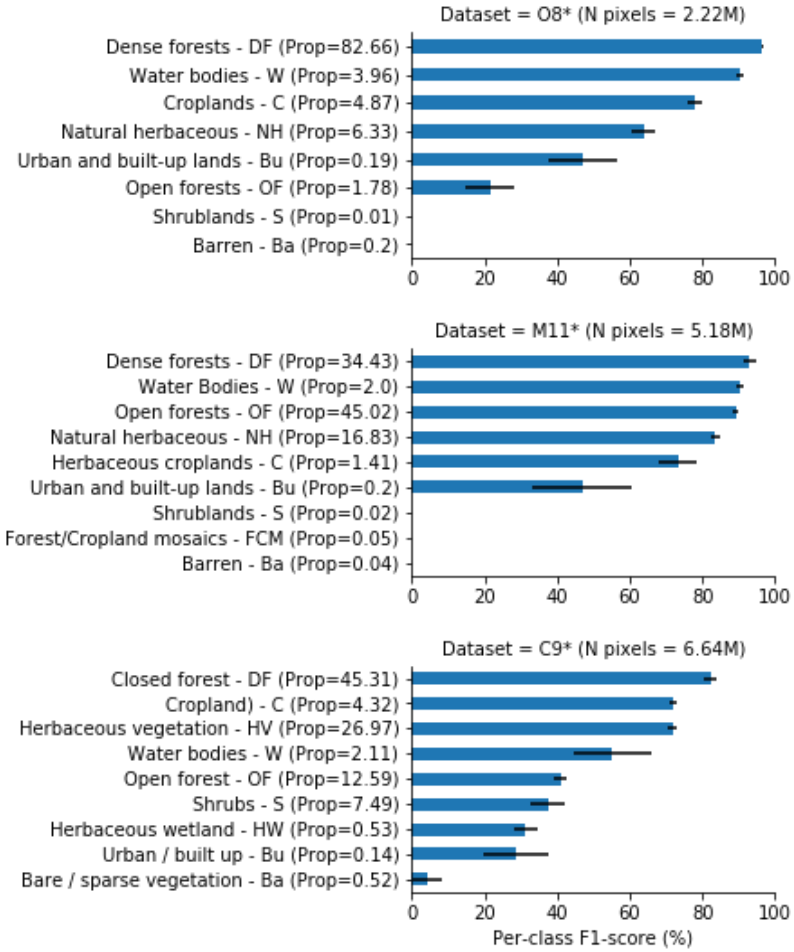


Figure 4.20 Class-wise F1-score of three hybrid labelled datasets. Datasets are ordered by size (number of reference pixels). Classes are ordered by F1-score value. The full and short-name of each class is provided followed by its respective proportion (Prop) according to the pixels in the test partition by dataset. The standard deviation per class corresponding to five runs of the model per dataset is represented by a black line.

A visual inspection of the predictions (see **Figure 4.21**) corroborated the hybrid datasets C9* and M11* datasets as having a better generalization than the O8* hybrid dataset. The dominance of *Dense Forest* class in the O8* dataset considerably misled the prediction capability of the MTLCC network for other classes, in particular the *Natural Herbaceous* class. Whereas the *Dense/Closed* and *Open Forest* classes in M11* and C9* are correctly classified for the majority of the areas, the cropland-like class was misclassified as herbaceous-like vegetation (Natural Herbaceous and Herbaceous Vegetation, respectively).

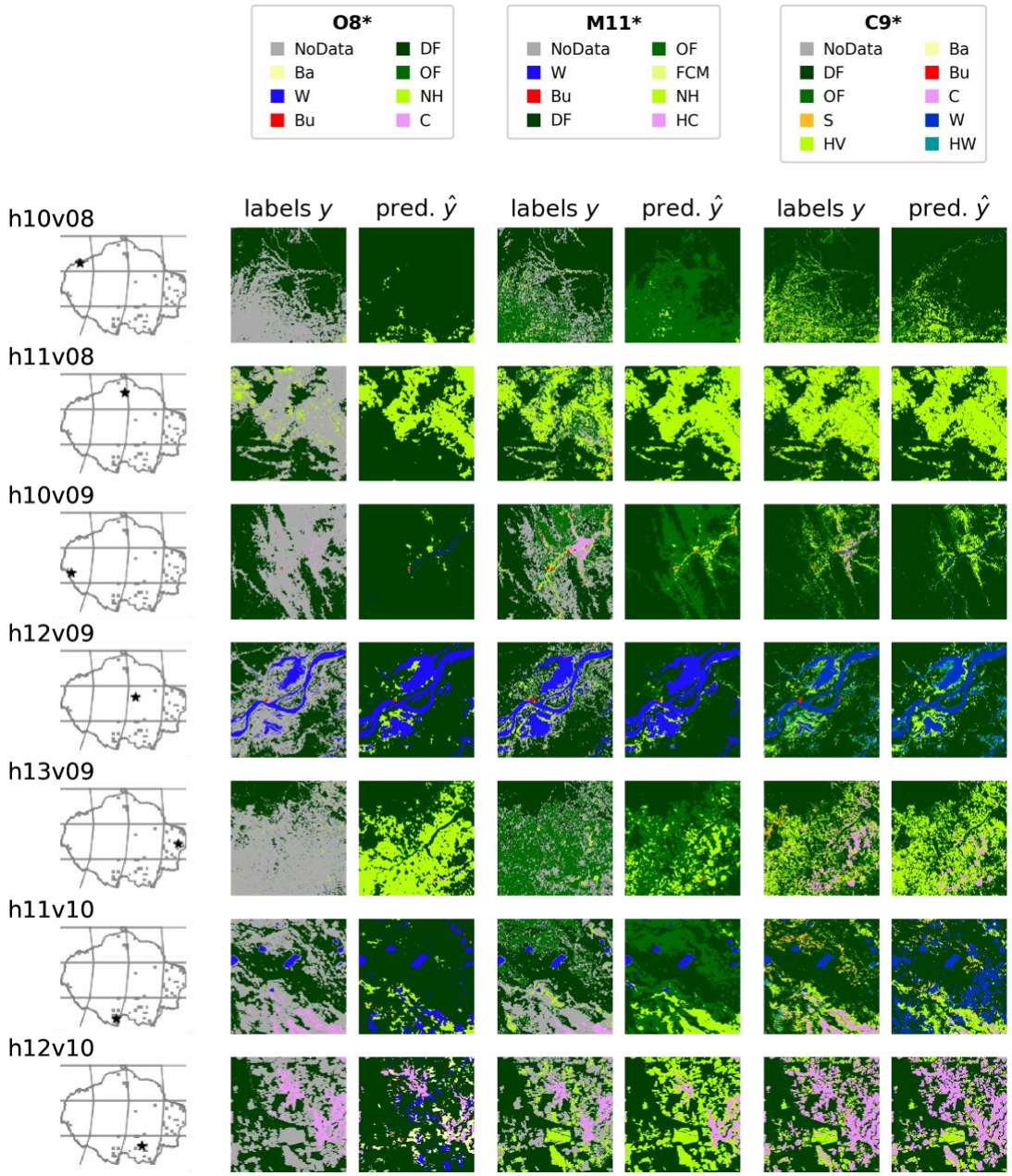


Figure 4.21 Reference labels y and predictions $\hat{y}$ for 2015 for trained MTLCC models using three different hybrid datasets over seven representative tiles. In O8*, the *Natural Herbaceous* (NH) is incorrectly classified as *Dense Forest* (DF) over certain tiles (see h10v08, h10v09, h11v10 and h12v10). The predictions in M11* and C9* misclassify *Cropland* like cover in the representative tiles in h10v09 and h11v10. Further details of the classes and respective acronyms per dataset can be found in **Appendix II**.

4.6.2 Cross-year performance over representative areas

4.6.2.1 Qualitative assessment

This section presents a visual inspection of the spatio-temporal consistency of the smoothed LC predictions by the trained models using C9* and M11* datasets across 17 years of MODIS imagery. To facilitate this task, the predictions are

studied over a single deforestation year, in this case the peak year of Terra-i's deforestation data from 2004 to 2010 per sampled tile. The analysis over three out of seven sampled tiles, which shows contrast in clearing sizes and locations across the study area, is presented as follows.

Tile h12v10: large area clearings in the eastern Amazon

Figure 4.22 illustrates the spatial distribution of peak year Terra-i's deforested pixels used to project post-loss LC change in a tile with large area clearings (greater than 3 Terra-i's deforested contiguous pixels). For the MapBiomas dataset, most post-loss LC is due to farmland which remained stable until 2017. In contrast, the observed post-loss LC data in the hybrid datasets are dominated by different classes. Whilst C9* shows conversion from *Dense Forest* to *Cropland* or *Herbaceous Vegetation*, post-loss LC in the M11* dataset is mostly dominated by a single class, *Natural Herbaceous*.

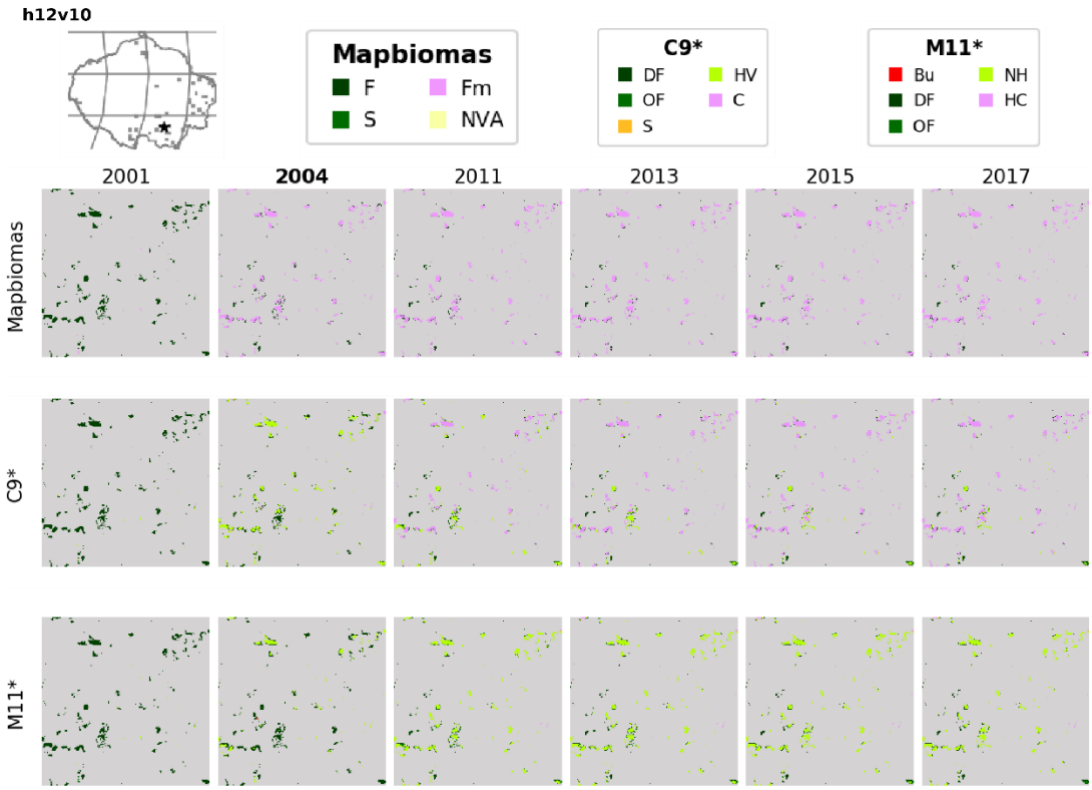


Figure 4.22 Subset of 2001-2017 annual LC extracted from the MapBiomas dataset and predictions by the MTLCC network trained using C9* and M11* datasets over deforested pixels in a representative tile (h12v10). The peak deforestation year is highlighted in bold.

Figure 4.23 presents the distribution of the classes as well as sequences partitioned by the stability of the observed LC class(es) i.e. remained the same (stable) or different (non-stable) from 2001 to 2017. Overall, according to the classes distribution per year chart, for all datasets, the proportion of dense/primary forest-like pixels is abruptly reduced from the peak deforestation year. After deforestation, the proportion of cropland/herbaceous-like cover steadily increases over time. The partition of the sequences shows a high proportion of non-stable sequences (greater than 89%) regarding stable sequences (less than 11%).

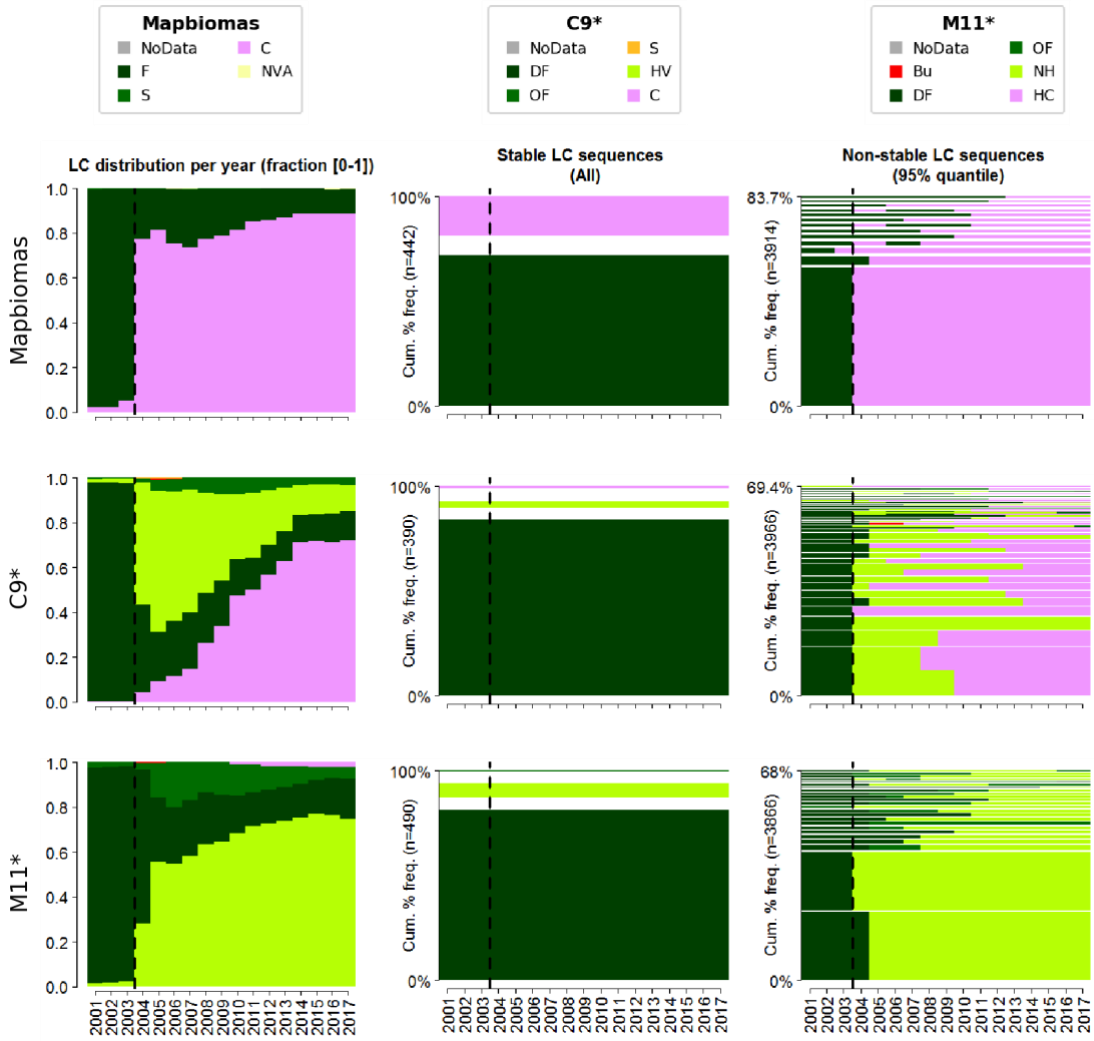


Figure 4.23 Comparison of 2001-2017 annual LC extracted from the MapBiomas dataset in comparison to predictions of the MTLCC model trained using C9* and M11* datasets over target deforested pixels in a representative tile located in h12v10. The peak year of the target deforested areas is highlighted by a dashed vertical line. It is important to note the most common non-stable sequence per dataset matches the deforestation year, except for M11* which shows changes from *Dense Forest* (DF) some years after flagged by Terra-i.

Tile h10v08: small area clearings in the northwest Amazon

The observed post-loss LC change trajectories over deforested areas, dominated by small clearings (less than 3 Terra-i's deforested contiguous pixels) (see **Figure 4.24**), are not as distinguishable as the previous tile with large area clearings. Overall, some pixels flagged as deforested are *Forest formation/Dense Forest* over the entire analysed period, 2001-2017. Whilst the converted *Dense Forest* pixels in C9* and M11* datasets in 2007 are due to *Open Forest*, the *Forest formation* in the MapBiomas dataset is due to *Farmland*.

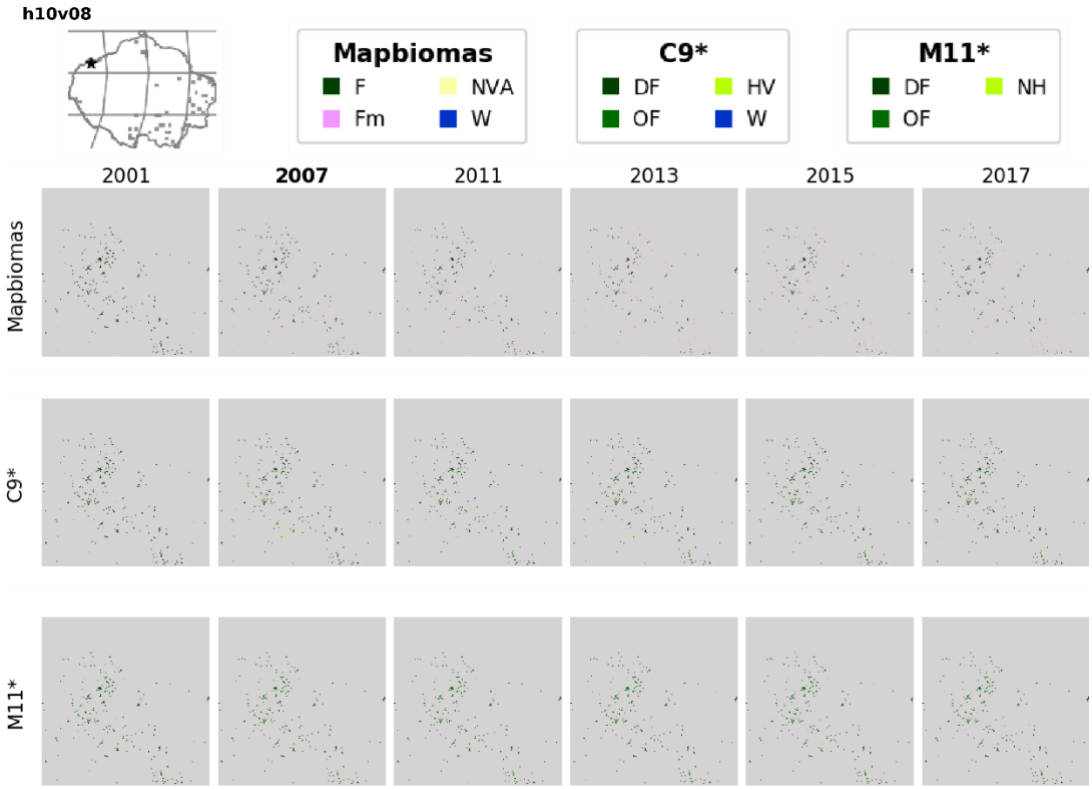


Figure 4.24 Subset of 2001-2017 annual LC extracted from the MapBiomas dataset and predictions by the MTLCC network trained using C9* and M11* datasets over deforested pixels in a representative tile (h10v08). The peak deforestation year is highlighted in bold.

The charts of the LC distribution in **Figure 4.25** highlight some similarities and differences between the datasets analysed. For instance, similar patterns are found between MapBiomas and C9* prior to or after the target deforestation year. Prior to deforestation (before 2007), *Forest formation* and *Dense Forest* pixels represented a large proportion of test partition in MapBiomas and C9*, respectively. In contrast, the predictions using M11* indicated *Open Forest* and *Dense Forest* share equal proportions. After deforestation, the proportion of the

classes associated with the reduction of *Forest formation/Dense Forest* are similar between MapBiomas and the C9* dataset. For M11*, the reduction of *Dense Forest* occurs abruptly after the deforestation year and does not follow a progressive reduction in comparison to the other datasets. The proportion of stable/non-stable is different between MapBiomas and hybrid datasets. While the proportion of stable (35%) is lower than non-stable pixels (65%) in the MapBiomas dataset, it is almost equal for the C9* (47%/53%) and M11* (51%/49%). The charts of non-stable sequences also indicate some forest loss occurs before the target deforestation year, being more evident for the MapBiomas dataset.

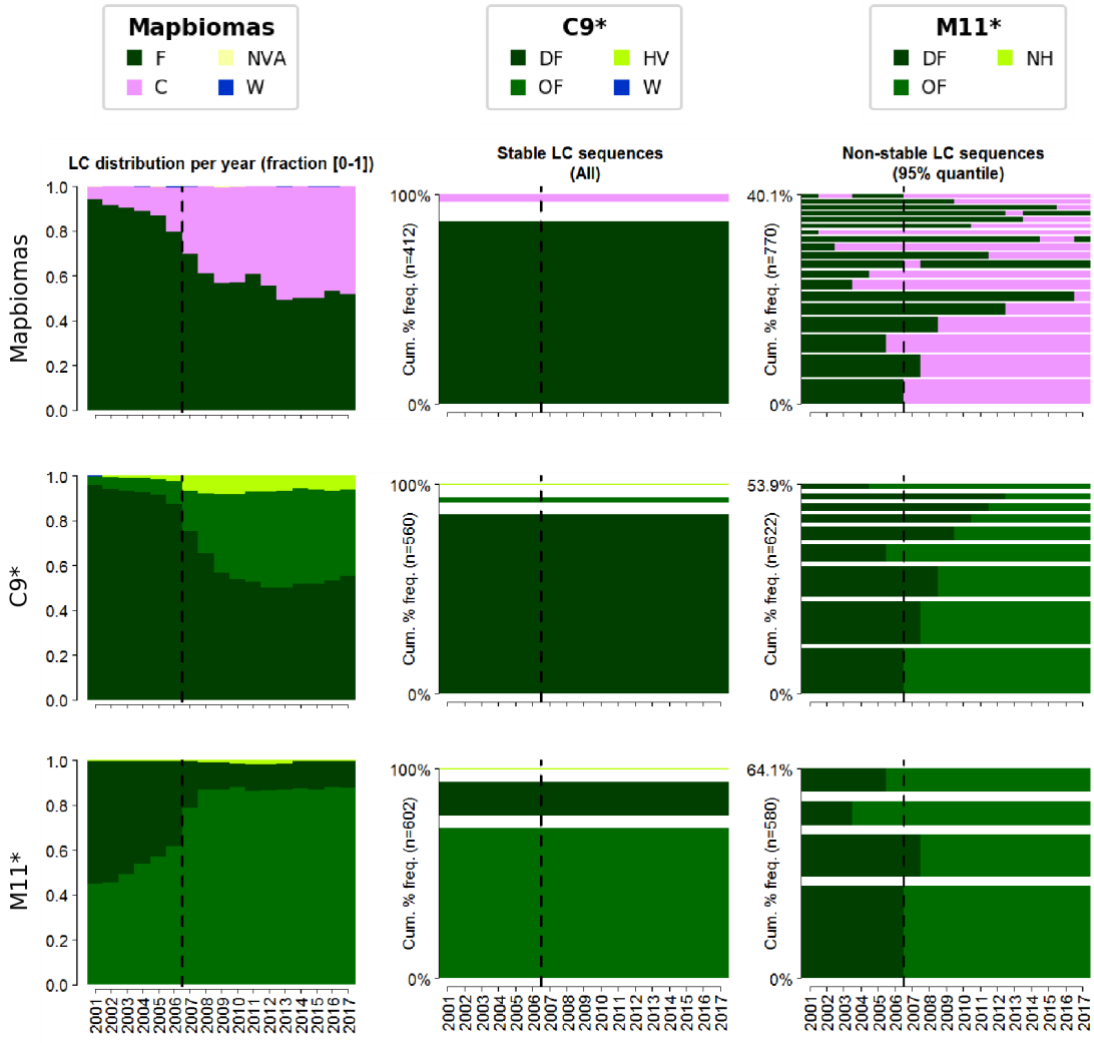


Figure 4.25 Comparison of 2001-2017 annual LC extracted from the MapBiomas dataset in comparison to predictions of the MTLCC model trained using C9* and M11* datasets over target deforested pixels in a representative tile located in h10v08. The peak year of the target deforested areas is highlighted by a dashed vertical line.

Tile h10v09: small area clearings in the western Amazon

The post-loss LC in this tile presents similarities to the tile located in the MODIS h10v08 scene i.e. small clearings (see **Figure 4.26**). Dense Forest in C9* and M11* datasets are converted to Open Forest, whereas Forest formation is due to Farmland in the MapBiomas.

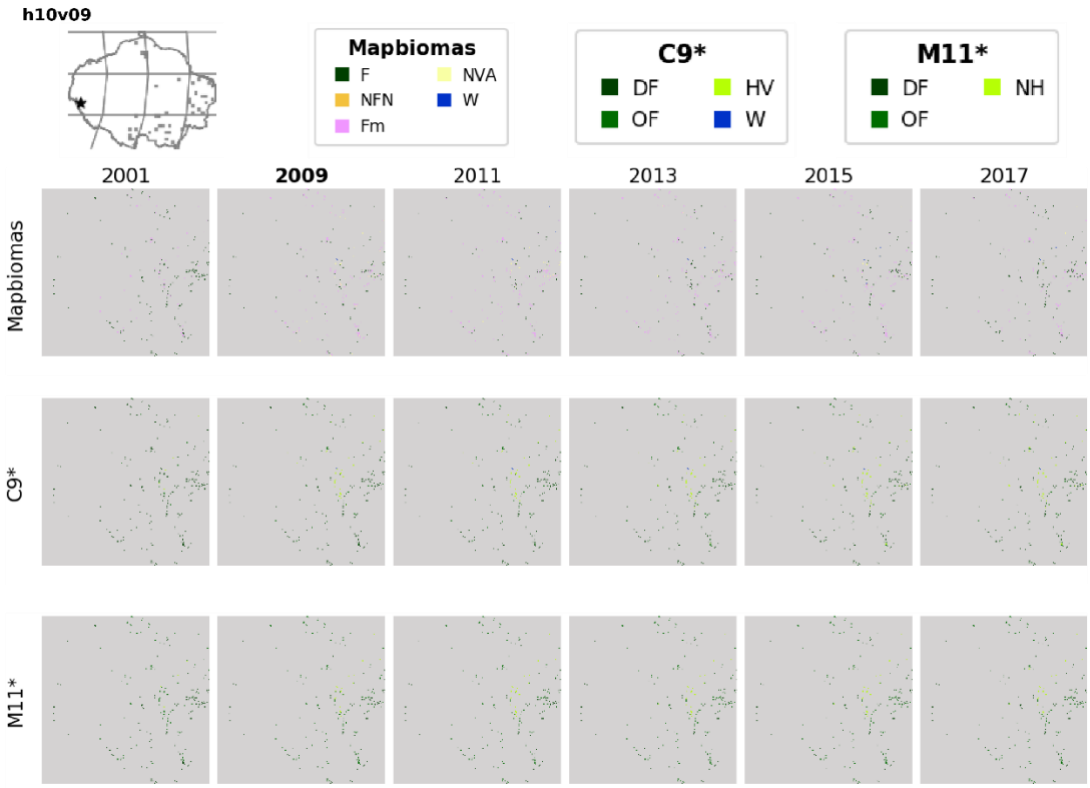


Figure 4.26 Subset of 2001-2017 annual LC extracted from the MapBiomas dataset and predictions by the MTLCC network trained using C9* and M11* datasets over deforested pixels in a representative tile (h10v08). The peak deforestation year is highlighted in bold.

Figure 4.27 illustrates the charts of LC distribution and stable/non-stable LC sequences per year for this tile. It is evident that the proportion of *Forest formation* in the MapBiomas dataset is reduced from 80% in the target deforestation year to 40% at the end of the analysed period. This proportion is lower for *Dense Forest* in the C9* dataset, which is reduced from 80% to 60%. After the target deforestation year, the proportion of classes different to *Dense Forest* in the M11* dataset is relatively similar to the proportions of *Forest formation* in the MapBiomas dataset. The proportion of stable/non-stable differs between MapBiomas and LC predictions by the MTLCC network trained using the hybrid datasets. While the proportion of stable pixels (14%) is considerably smaller than

non-stable (86%) in the MapBiomas dataset, it is almost equal for C9* (46%/54%) and M11* (45%/55%). Moreover, for the MapBiomas dataset, the quantile of 95% retains only 22.1% of the possible non-stable sequences. In contrast, the non-stable sequences seem to be less diverse in C9* and M11* predictions due to the 95% quantile retaining 40.8% and 44.9% of the potential sequences, respectively. The non-stable charts indicate forest loss also occurred before the target deforestation year, being more pronounced in the observed sequences of the MapBiomas dataset.

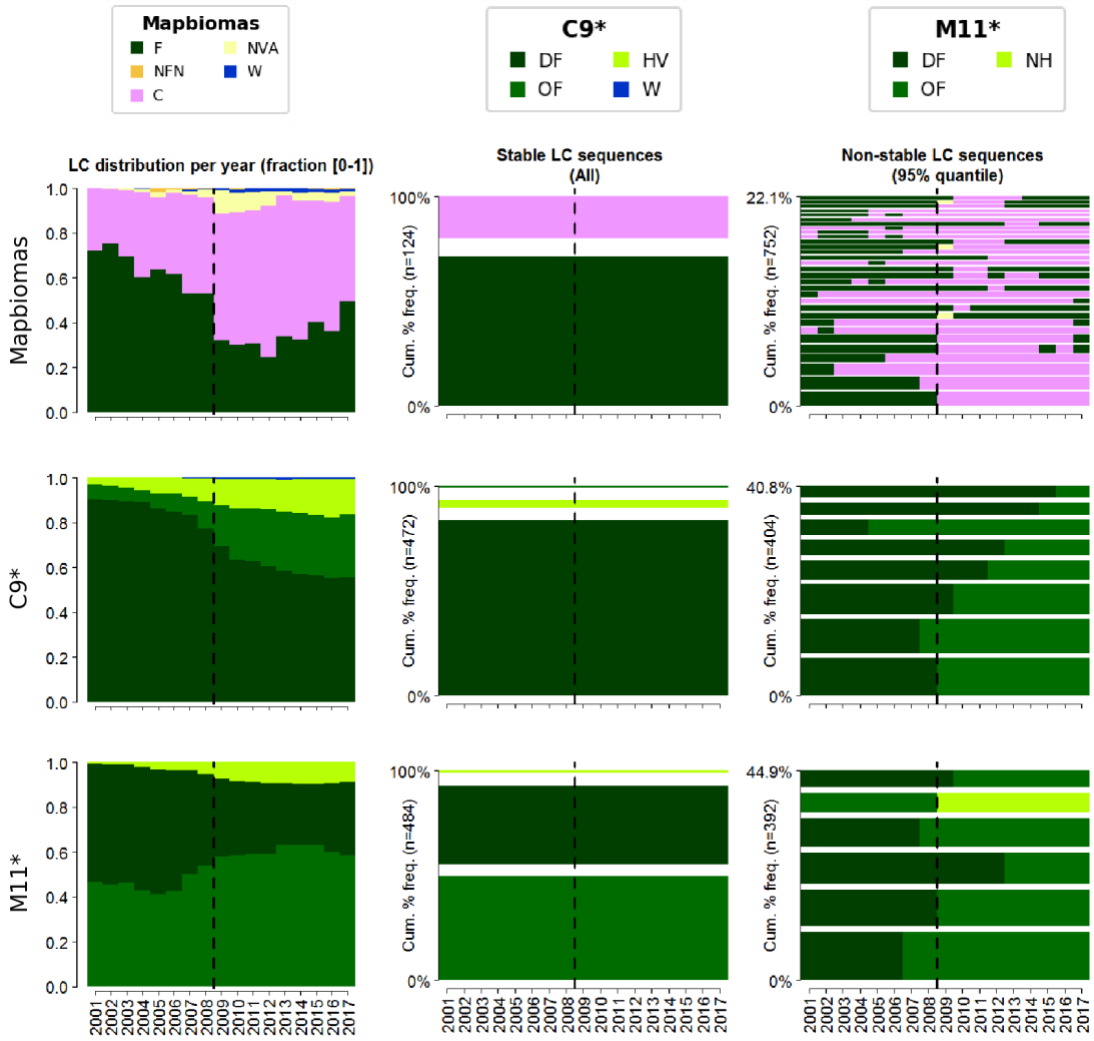


Figure 4.27 Comparison of 2001-2017 annual LC extracted from the MapBiomas dataset in comparison to predictions of the MTLCC model trained using C9* and M11* datasets over target deforested pixels in a representative tile located in h10v09. The peak year of the target deforested areas is highlighted by a dashed vertical line.

4.6.2.2 Quantitative assessment

This section shows results of the independent validation conducted over the C9*, M11* and MapBiomas datasets according to the sampling strategy as described in **Section 4.5.5.2**. **Table 4.8** presents the number of samples per dataset per tile which were verified using annual high-resolution (30-m) Landsat archives. From a population of 8885 deforested sites located across the representative tiles, a total of 136, 147 and 98 samples were quantitatively assessed in MapBiomas, C9* and M11* datasets, respectively.

Table 4.8 Number of samples per tile and totals per dataset.

Tile ID (MODIS Scene)	**Peak year (total pixels)**	**Number of samples (stable/non-stable)**		
		MapBiomas	CGLS-LC100 Custom (C9*)	2001-2015 MODIS FAO-LCCS2 (M11*)
h10v08	2007 (1182)	23 (4/19)	15 (6/9)	10 (5/5)
h10v09	2009 (876)	32 (1/31)	13 (5/8)	9 (3/6)
h11v08	2010 (85)	3 (1/2)	2 (1/1)	1 (1/1)
h11v10	2005 (2140)	26 (2/24)	43 (2/41)	21 (1/20)
h12v09	2009 (85)	3 (1/2)	3 (1/2)	6 (1/5)
h12v10	2004 (4356)	43 (4/39)	65 (4/61)	47 (5/42)
h13v09	2007 (161)	6 (1/5)	6 (1/5)	4 (1/3)
Total	2004-2010 (8885)	136 (14/122)	147 (20/127)	98 (17/81)

Table 4.9 shows accuracy metrics derived from 2001-2017 annual high-resolution image interpretation in all samples per dataset as described in **Table 4.8**. The observed annual LC in the MapBiomas dataset obtained a higher overall accuracy (OA) and a lower total disagreement (quantity and allocation) than the hybrid datasets for the stable sequences. For the non-stable sequences, metrics of the hybrid datasets approximate to the values of the MapBiomas dataset.

Complementary to the above metrics, the metrics per year for non-stable sequences indicated that most of lower values in OA (below 70%) are located within the window period, 2004 to 2010, used to select the target deforested sites per tile (see **Figure 4.28**). In particular, the lowest values are located in 2004 which coincides with the target year being the tile with the largest number of samples validated (see tile h12v10 in **Table 4.8**).

Table 4.9 Accuracy metrics per dataset derived from high-resolution image interpretation spanning the full set of samples for the analysed period, 2001-2017. The metrics are divided by stability (stable/non-stable) per dataset.

Labelled dataset	Stability	Overall accuracy (%)	Quantity Disagreement (%)	Allocation Disagreement (%)
MapBiomas	Stable	94.54	5.46	0.65
	Non-stable	73.92	4.68	21.41
CGLS-LC100 Custom (C9*)	Stable	84.12	15.88	1.20
	Non-stable	73.37	7.07	19.56
2001-2015 MODIS FAO-LCCS2 (M11*)	Stable	82.01	15.22	2.77
	Non-stable	72.88	3.23	23.89

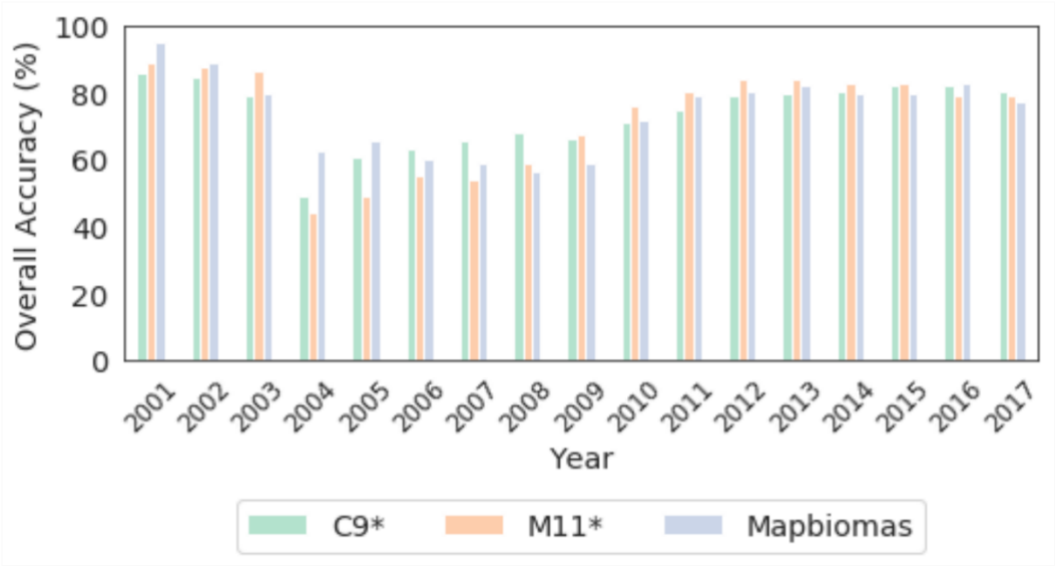


Figure 4.28 Overall accuracy per year per dataset of the samples verified via high-resolution image interpretation across seven representatives in the study area. The lowest values are located in 2004 which coincides with the target year of the tile with the largest number of samples validated. In general, M11* achieves slightly greater values over time.

4.7 Discussion and conclusions

The objective of this chapter was to study the effectiveness of the proposed DL model, the MTLCC network, according to multiple labelled datasets derived from pre-existing global LC products. These datasets allow for automatically creating a large volume of training samples (n > 10 million) to train the MLTCC network in rapidly classifying LC for a desired year. The most suitable labelled datasets for the study of post-loss LC change trajectories are discussed according to a two-stage assessment. Given the variety of labelled datasets processed here, rather than computing confidence interval and conducting probabilistic sampling (Wulder *et al.*, 2006), the accuracy of the MTLCC network trained was analysed

according to two procedures. Specifically, the first procedure referred to compute performance metrics over the test partition in the trained year (see **Section 4.7.1**); in the second procedure (see **Section 4.7.2**), a cross-year evaluation over selected representative sites was conducted, alongside a comparison with MapBiomas Amazonia, the most accurate LC product in the study area.

4.7.1 Single year performance over test partition

The validation exercise carried out using the test partition revealed the MTLCC network is flexible in generating LC predictions according to different reference labels derived from pre-existing global LC maps. This flexibility aligns with previous findings assessing the MTLCC network proposed by Rußwurm and Körner (2018b) using datasets in Europe (Pelletier *et al*., 2019) and Africa (Rustowicz *et al*., 2019). Overall, the capability of the model in predicting LC using dense earth-observation time series, in this case 46 multi-spectral MODIS images in 2015 at a spatial resolution of 250-m, is influenced by characteristics of the training dataset (size, number of classes, labels noise).

As shown in **Section 4.6.1**, amongst the set of metrics assessed, disagreement related metrics were more informative in comparing the performance of the labelled datasets. The separation of the overall disagreement into components of quantity and allocation is useful in assessing the quality of LC map(s) according to the end purpose. For capturing the net quantity of LC change during the multiple times, the quantity disagreement is of more relevance than the allocation disagreement (Pontius and Millones, 2011). This type of insight is not offered by the Kappa metric as two out of nine labelled datasets assessed presented equal values for this metric however their overall disagreement was different. According to the above premise, the values of quantity disagreement indicate the hybrid datasets offer a substantial gain in performance in comparison to the corresponding original datasets. This result can be partly explained by the reduction of noisy labels due to the exploitation of temporal, spatial and semantic (nomenclature) attributes from the original datasets in generating the hybrid datasets (see **Section 4.5.3.2**). In particular, the temporal-driven hybrid datasets derived from MODIS IGBP and MODIS FAO-LCCS2 datasets achieve lower total

disagreement values in comparison to the spatial attributes i.e. subpixel information exploited from the CGLS-LC100 dataset.

The per-class accuracies shown in **Figure 4.20** indicate that the MTLCC model is unable to identify *Barren* class amongst the hybrid datasets (F1-score below 5%). The prediction of *Shrubland* is also difficult, with the exception of the C9* which obtained a F1-Score greater than 30%. The capabilities of the MTLCC network to map this particular class in C9* can be explained by the area proportional allocation of training samples which was larger in this dataset (7%) in comparison with O8* (0.01%) and M11* (0.03%). The five runs of the model per dataset were useful in highlighting the performance variance for certain target LC classes. For example, the *Urban/Built-up* class shows the largest variation with standard values from 8.88% to 13.54%. In contrast, the standard deviation for *Dense/Closed forest* class ranged from 0.28% to 1.72%. These differences in standard deviation might be related to the low proportion of samples in *Urban/Built-up* class (lower than 0.2%). In addition, mixed pixels of vegetation and urban-like covers might hinder stable predictions by the MTLCC network. As the literature suggests (MacLachlan *et al*., 2017), the heterogeneity of urban areas can be handled by using sub-pixel classification, in particular for large-area classification using moderate resolution satellite data. Whilst this technique might reduce the variance for this particular class, the extent and occurrence of forest conversion to urban is smaller than other conversions. Consequently, larger and more predominant conversions are represented by LC classes such as cropland, natural herbaceous, open forest which had reasonable accuracy values, in particular in M11* and C9* datasets.

A visual inspection of the predictions (see **Figure 4.21**) corroborated the hybrid datasets C9* and M11* datasets as having a better generalization than the O8* hybrid dataset. Although the integration of multiple datasets has been a common and low-cost technique in creating hybrid datasets (Schepaschenko *et al*., 2015; Tsendbazar *et al*., 2015, 2017), the training labelled dataset derived from the O8* dataset did not achieve consistent predictions. A possible explanation might be the size of the training dataset which was considerably smaller (12.6M of pixels) in comparison with M11* (25.1M) and C9* (32.7M) datasets. In addition to the dataset size, the poor performance of the trained model using the O8* dataset

might be associated with a high-class imbalance. Class imbalance is a common issue in supervised classification and it is a subfield of study within machine learning, including deep learning models (Johnson and Khoshgoftaar, 2019). While the *Dense Forest* class represented some two-thirds (75%) of the O8* dataset, the proportion of *Dense/Closed Forest* in the C9* and M11* datasets was less dominant and not higher than 40%. Both issues, dataset size and class imbalance, have been tackled in previous deep learning studies for remote sensing classification by increasing the training time (number of epochs), applying augmentation techniques (Qiu *et al*., 2019) and training with specialised loss functions i.e. focal loss (Teimouri *et al*., 2019). Although these solutions might alleviate the poor performance of O8* in non-dominant classes, the scope of this study was to find optimal labelled dataset(s) which guarantee(s) a reasonable performance reducing any additional modification in the original dataset that interfere with its comparison.

4.7.2 Cross-year performance over representative areas

Following the results of the first stage assessment (see **Section 4.7.1**), the cross-year evaluation was carried out over the MTLCC network models trained using C9* and M11* datasets. To approach the scope of mapping post-loss LC change trajectories, the annual LC predictions from 2001 to 2017 in these models were studied over disturbed pixels as indicated by Terra-i. Although changes areas could be simply extracted using annual LC maps generated by the MTLCC model, this procedure is not the most optimal as it heavily depends on the accuracy of the predicted LC maps. In this regard, the approach taken in this research is to use separate but complementary methods for LC mapping (the MTLCC network) and change detection (Terra-i). It is worth to mention that during the development of this research, there have been an increase attention to develop robust systems supported by deep supervised learning method for detecting and classifying change in a joint manner or so called semantic change detection (see for example Caye Daudt *et al.* (2019)). However, this approach is limited to characterise the immediate change so further work is expected to provide long-term subsequent LC over change areas.

The inclusion of LC predictions from 2001 to 2003 allowed a complete track of all possible LC change trajectories per analysed pixel. The removal of target

pixels near water bodies guarantees that forest changes analysed represent mostly human-driven disturbances rather than natural e.g. river meandering. In addition, the smoothing of raw LC annual predictions using Hidden Markov models (HMMs) was fundamental to reduce illogical transitions. This particular step allows for better comparison with the observed LC annual data extracted from the MapBiomas Amazonia dataset. The producer of this particular datas*et* also removed illogical transitions but using a semi-automated method relying on rules defined by expert knowledge (Proyecto MapBiomas Amazonía, 2019).

The qualitative inspection of the smoothed LC predictions from 2001 to 2017 over three regions in the study area (see **Section 4.6.2.1**) indicate some challenges of the trained MTLCC models using the C9* and M11* datasets. Whilst the LC sequences of trained models using both datasets were spatially and temporally consistent in a tile located in the MODIS h12v10 scene with large-area disturbed sites (see **Figure 4.22** and **Figure 4.23**), they were unable to capture possible LC sequences in the other tiles (h10v08 and h10v09) as is observed in the MapBiomas. This means the coarser resolution of the hybrid datasets, which are resampled to 250-m, could reduce the capability of the MTLCC network to fully predict LC across areas, in particular with small-scale or fractional disturbance. The quantitative analysis using an independent dataset and separated by the stability of the sequence (see **Table 4.9**) shows the overall accuracy is higher in the LC change trajectories observed in MapBiomas, in particular stable sequences (overall accuracy of 94.5%). For the non-stable sequences of observed LC in MapBiomas and M11*, it was found that the *quantity disagreement* values as defined in **Section 4.5.5.1** were closer than the C9*.

The MTLCC network trained with C9* produced coherent spatio-temporal distributions of LC classes that had a general good agreement with the MapBiomas maps. This could be partially explained by the level of detail provided by native spatial resolution of the C9* (100-m) which after being resampled to 250-m remains closer to the resampled MapBiomas dataset. Additionally, taking into account that the training data derived from C9* already include mislabelled samples due to the error of the original maps of CGLS-LC100 (accuracy of 80%), the MTLCC network appears to show a grade of robustness to mislabelled training samples. Z. Sun *et al*. (2019) trained a LSTM-RNN model to predict LC

based on training labels extracted from an outdated LC map. The authors suggested the training samples should be as accurate as possible for reducing the propagation of errors in this particular model. Following this work, the M11* is therefore of greater suitability for the analysis of post-loss LC change trajectories. First, the number of noisy labels is considerably reduced by using the temporal attributes of its original dataset as is reported in previous work (Zhang and Roy, 2017; Huang *et al*., 2015). This means the predictions are less prone to errors than a single-time step map such as C9*. Secondly, the number of classes fits with a classification scheme (FAO-LCCS2) already implemented in other latitudes at national scale e.g. South Africa (Wessels *et al*., 2016).

5 Calibration and comparison of the proposed model for the study of post-loss LC change

5.1 Introduction

As large volumes of earth observation data are generated daily, an increase in cutting-edge methods are developed and deployed to ingest and analyse them. Emerging deep learning-based algorithms and cloud computing infrastructure have become available with the potential to impact image processing and analysis of big earth observation data (Parente *et al.*, 2019). Following the promising results obtained using the MTLCC network to study post-loss LC (see **Chapter 4**), this chapter aims to calibrate (also known as fine-tune) the MTLCC model using cloud computing infrastructure. This calibration procedure is necessary to maximize the performance of the model to ensure consistent spatio-temporal LC predictions suited to the study of post-loss LC change trajectories. The use of cloud infrastructure means conducting multiple experiments in parallel that might take longer using the hardware described in **Chapter 4** (i.e. a local server with a single GPU). Moreover, the particular cloud-infrastructure chosen provides a set of state-of-the-art techniques to calibrate machine learning models, some of them introduced in the sections below.

In addition to the above exercise, the fine-tuned MTLCC network is compared against two conventional machine learning algorithms in remote sensing classification, Random Forest and Support Vector Machine (SVM). Both algorithms were also calibrated in order to yield their maximum potential according to the input training data. Finally, some capabilities of the MTLCC network are explored such as discrimination of the LC classes and filtering of cloud observations from the input satellite images. Moreover, the performance of the calibrated model to predict over geographically diverse areas in the pantropical study area are presented and discussed.

5.2 Background

5.2.1 Hyper-parameter tuning

Most machine learning models are parameterized by a set of hyper-parameters which must be set appropriately to maximize the performance of the learning approach. Hyper-parameters differ from model parameters i.e. neuron weights in neural networks due to the fact they are not learned during the training phase. This means the settings of hyper-parameters are specified outside the training procedure and could control the capacity of the model in fitting the data. The task of finding an optimal setting of hyper-parameters is known as hyper-parameter tuning (Zheng, 2015).

Conceptually, model training and hyper-parameter tuning represent optimisation tasks but their approaches remain different in practice. Whilst in model training the quality of a proposed set of model parameters can be expressed as a mathematical formula, this is not the case in hyperparameters as their quality depends on the outcome of a black-box function (i.e. the model training process) (Zheng, 2015). A black-box function describes a lack of expression for the objective function and derivatives. The goal of hyper-parameter tuning is to find a set of hyper-parameters λ^* from a hyper-parameter configuration Λ that yields an optimal model by minimizing a loss function L. Given a dataset $\mathcal{D}$, the hyper-parameter optimisation can be formalised as follows (Feurer and Hutter, 2019).

$$\lambda^* = argmin_{\lambda \in \Lambda} \mathbb{E}_{(D_{train}, D_{valid}) \sim \mathcal{D}} V(\mathcal{L}, \mathcal{A}_\lambda, D_{train}, D_{valid}) \qquad \text{Equation 5.1}$$

where $V(\mathcal{L}, \mathcal{M}_\lambda, D_{train}, D_{valid})$ measures the loss of a model generated by algorithm $\mathcal{A}$ with hyperparameters on training data D_{train} and evaluated on validation data D_{valid}. Both partitions, D_{train} and D_{valid}, are assumed to be representative of the full dataset $\mathcal{D}$. The most common options for the validation protocol $V(\cdot,\cdot,\cdot)$ include the holdout and cross-validation error for a given loss function (Feurer and Hutter, 2019). Bischl *et al.* (2012) provides an overview of these and other various strategies related to the validation protocol.

In general, there are two ways in which hyper-parameters are tuned; manual and automated tuning. For the former, the modeller is in charge of searching the hyper-parameter settings to test different parameter combinations. In the latter,

the hyper-parameter search is automated and is made part of the training algorithm. An example of a manual search was conducted in **Chapter 4** related to the assessment of the pre-existing dataset. The hyper-parameters were set manually after experimentation in a trial and error manual search process limited by the available hardware resources, a single local GPU. This process is tedious and error-prone as described by scholars working on neural networks (Almeida and Ludermir, 2008). Since the analysis of this Chapter is supported by cloud computing infrastructure, the evaluation of automated hyper-parameter tuning is feasible, allowing for the assessment of main hyper-parameters of the MTLCC network applied to the study of post-loss LC change trajectories. The following sections describe search algorithms divided according to the methodological approach in two categories, model-based and model-free hyper-parameter optimisation. Whilst model-based search algorithms can use the knowledge gained during their processing to adapt and intensify the search in areas of the hyperparameter space with higher result potential, model-free do not use this information for adaptation. For the sake of simplicity, this research describes only the grid search and random search as examples of model-free, and Bayesian optimisation as examples of model-based optimisation.

5.2.1.1 Grid and Random Search

Grid search is one of the simplest methods for hyper-parameter optimisation. It primarily consists of an exhaustive search in the hyper-parameter space considering all possible combinations. This process is highly impacted by the curse of dimensionality (number of hyper-parameters). This means that the more dimensions and complexity are added, additional computational resources are demanded. The approach is commonly applied when the dimensions are less than or equal to 3 (Feurer and Hutter, 2019). In the context of deep neural networks, grid search is unfeasible as hyper-parameters are too numerous and thus computationally prohibitive (Bey, Díaz, *et al.*, 2016).

To explore a wider hyper-parameter space (more than 3 dimensions), Bergstra and Bengio (2012) proposed the random search method. These scholars created this method under the assumption that some hyper-parameters are significantly more important than others in a specific machine learning problem. This assumption considerably reduces the search of optimal hyperparameter settings

by choosing random points in the parameter space according to a marginal distribution per hyper-parameter. This distribution is defined generally by the modeller including Bernoulli for binary, multinoulli for discrete, uniform on a log-scale for positive real values.

Figure 5.1 illustrates Bergstra and Bengio's (2012) comparison between grid and random search approaches. Whilst the grid search solely tested three values of each hyper-parameter, the random search tested nine for the same cost. This gain in performance can be translated into a faster search which is supported by a simpler parallelization and flexible resource allocation than the grid search. Due to this characteristic, the random search is commonly used as a baseline for hyper-parameter optimisation (Feurer and Hutter, 2019).

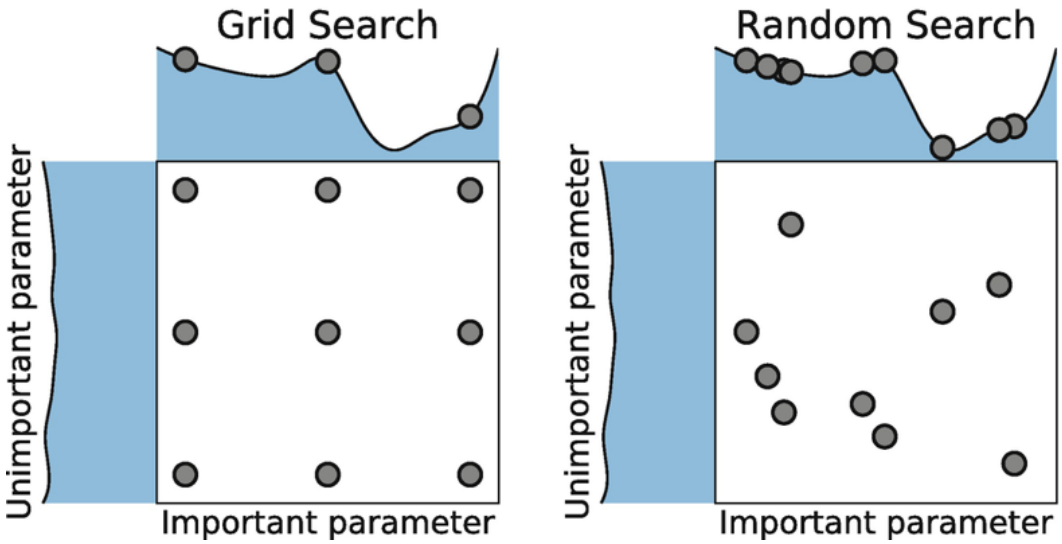


Figure 5.1 Comparison of grid search (left) and random search (right) for minimizing a function with one important and one unimportant parameter. The figure is based on the illustration by Bergstra and Bengio (2012) as adapted by Feurer and Hutter (2019).

5.2.1.2 Bayesian Optimisation

Bayesian optimisation is a well-established framework (Močkus, 1975) for global optimisation with a growing interest in noisy and expensive black-box functions such as hyper-parameter tuning of deep neural networks (Feurer and Hutter, 2019). Using the notation from **Section 5.2.1**, Bayesian optimisation incorporates a probabilistic model $\mathcal{M}$ for modelling the relationship between a hyper-parameter configuration Λ and its performance $f(\lambda)$. The model is fitted by using previously gathered data (prior belief) which are then applied in order to select the next point λ_{new} to evaluate, trading off exploitation and exploration in order to find the

minimum of f. The Bayesian optimisation then evaluates $f(\lambda_{new})$, updates $\mathcal{M}$ with the new data $(\lambda_{new}, f(\lambda_{new}))$ (Eggensperger *et al*., 2014).

The basic and variants of Bayesian optimisation suited to hyper-parameter tuning have been a subject of study by multiple scholars, see for example Snoek *et al*. (2012). So far, three state-of-the-art methods, which are selected based on their popularity (high number of citations) as described by Eggensperger *et al*. (2014) are:

- Spearmint (Snoek *et al*., 2012) is a basic form of Bayesian optimisation method that models $\rho\mathcal{M}(f \mid \lambda)$ with Gaussian process (GP) models. It supports continuous and discrete parameters, but no conditional parameters (Eggensperger *et al*., 2014).
- Sequential Model-based Algorithm Configuration (SMAC) (Hutter *et al*., 2011) approximates $\rho\mathcal{M}(f \mid \lambda)$ with Random Forests. When performing cross-validation, SMAC only evaluates as many folds as necessary to show that a configuration is worse than the best one observed thus far. SMAC can handle continuous, categorical, and conditional parameters (Eggensperger *et al*., 2014).
- Tree Parzen Estimator (TPE) (Bergstra *et al*., 2011), models $\rho\mathcal{M}(f \mid \lambda)$ indirectly. Whilst SMAC and Spearmint model $\rho\mathcal{M}(f \mid \lambda)$ explicitly, i.e. the estimation of the cost function value given the parameters, TPE factors it (Madrigal *et al*., 2019). It models $\rho(f < f*)$, $\rho(\lambda \mid f < f*)$, and $\rho(\lambda \mid f \geq f*)$, where $f*$ is defined as a fixed quantile of the function values observed thus far, and the latter two probabilities are defined by tree-structured Parzen density estimators (also referred to as kernel density estimation). TPE can handle continuous, categorical, and conditional parameters (Eggensperger *et al*., 2014).

In terms of performance, Bayesian optimisation with Gaussian process (GP) models have presented inferior performances than TPE and SMAC (Bergstra *et al*., 2011; Hoffman *et al*., 2014). Further information about Bayesian optimisation in the context of optimizing the hyper-parameters of machine learning models can be found in Bergstra *et al*. (2011).

5.2.2 Transfer learning

Transfer learning (TL) refers to a subfield of machine learning. The aim of TL is to apply knowledge learned in one situation to another situation. Under this approach, different domain fields have transferred knowledge to not only the target domain but also from other source domains and source tasks (Pan and Yang, 2010). This technique has gained more popularity with the emergence of deep learning (Bengio, 2012). Thanks to its focus on learning representation and in particular 'abstract' representations, deep learning is well suited to transfer learning (Bengio *et al*., 2011). This particular benefit is investigated in remote sensing studies. For instance, the knowledge and high-performance of deep neural networks in the optical domain classification have been transferred to optimise the Synthetic Aperture Radar (SAR) domain in which large training datasets are often scarce (Rostami *et al*., 2019). Transfer learning can be divided into three categories depending on how the domains and tasks are related (see **Figure 5.2**) (Pan and Yang, 2010).

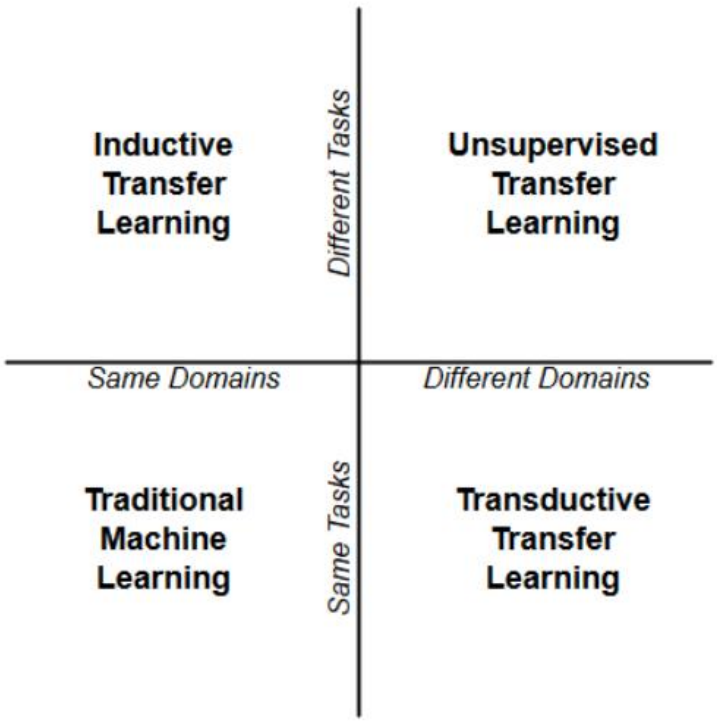


Figure 5.2 Quadrant diagram related to the categories of transfer learning according to the domain (x-axis) and task type (y-axis). Source: Bertrand (2019).

In terms of hyperparameter optimisation, increased attention has been focused on TL. The principle aim is to exploit evaluations from previous related tasks (e.g.

a neural network tuned on multiple datasets) to further accelerate the hyper-parameter search (Salinas *et al*., 2019). Several approaches have been proposed to induce TL in hyper-parameter optimisation. The main approach consists in modelling tasks jointly or via a conditional independence structure, which has been conducted through multi-output Gaussian Processes (GPs) (Swersky *et al*., 2013), weighted combination of GPs (Katakami *et al*., 2019) and neural networks (Feurer *et al*., 2018). An alternative approach focuses on settings where models need to be updated online as new problems e.g. data with different classes emerge. This procedure is achieved by fitting a sequence of surrogate models to residuals for the predictions of the previously fitted model (Golovin *et al*., 2017). Finally, a third approach has been explored by conducting Bayesian optimization (BO) with the solutions i.e. tuned hyper-parameters of the previous BO problems (also known as warm-starting) (Feurer *et al*., 2015). It is important to note other emerging methods exist in order to not only optimise the accuracy of a black-box function as common methods do but also its runtime or memory consumption.

5.2.3 Cloud-computing

Cloud-computing stands for the latest progress in distributed computing by offering 'computing as a service' to end-users in a 'pay-as-you-go' price model. Substantial cloud computing resources are now available under this model, from providers such as Amazon Web Services (AWS), Microsoft Azure, and Google's Compute Engine. Cloud-computing is proven to be convenient in terms of budget and energy consumption efficiencies (Marston *et al*., 2011). It is a technology which has expanded to different industries and knowledge domains. For geospatial sciences, Google Earth Engine (GEE) is thus far the most well-known example of a successful implementation of this technology. GEE is a cloud-based platform facilitating access and analysis to a multi-petabyte catalogue of satellite imagery and geospatial datasets (Gorelick *et al*., 2017). Further examples of the principles and implementation of this technology for geospatial sciences and applications can be found in Yang *et al*. (2011), and more recently in Yang *et al*. (2017) and; Yao *et al*. (2019).

Three types of services are offered within the cloud computing (Yang *et al*., 2011): Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). The descriptions of these categories are as follows:

- IaaS delivers computer infrastructure, including physical machines, network, storage and system software, as virtualized computing resources over computer networks. Users can configure, deploy, and run Operating Systems (OS) and applications based on the OS. The Amazon Elastic Compute Cloud (EC2) and Google's Compute Engine are the most notable examples of IaaSs.
- PaaS offers a higher level than IaaS by providing a platform service for software developers to develop applications. Apart from computing platforms, PaaS provides a layer of cloud-based software and Application Programming Interface (API) that can be used to build higher-level services. GEE and Microsoft Azure are examples of PaaS. Users can develop or run applications on such a platform without maintaining the OS, server hardware or computing capacity.
- SaaS provides a wide range of capabilities in applications that are provided through Web browsers to the end-user. Examples of this category are Google's Gmail and apps. The implementation of ESRI's ArcGIS on the cloud is another example of geospatial-focused SaaS.

5.3 Study area and context

The study area corresponds to the Amazon region which represents the largest continuous region of tropical forest in the world. The Amazon as delineated by the Amazonian Network of Georeferenced Socio-Environmental Information (RAISG) was chosen (the relevance of the area and rationale of the boundary chosen are discussed in **Chapter 4** (see **Section 4.4**). This delineation encompasses approximately 7.8 million km^2 using a biogeographical criterion, with the exception of Ecuador and Brazil where legal-administrative criteria are used.

5.4 Methodology

Figure 5.3 describes workflow with an aim to calibrate the MTLCC network and traditional models according to the optimal source of reference data to generate LC predictions suited to the study of post-loss LC change trajectories. This optimal source was derived from the assessment of original and hybrid LC datasets as described in **Chapter 4**. The analysis of post-loss LC data according to predictions from the calibrated MTLCC network is presented in **Chapter 6**.

The inputs of the workflow are:

1. Multi-Temporal-Spectral Data (satellite data) (see **Section 5.4.1**).

2. A reference data with labelled samples (geographical positions for which LC is known) (see **Section 5.4.1**).

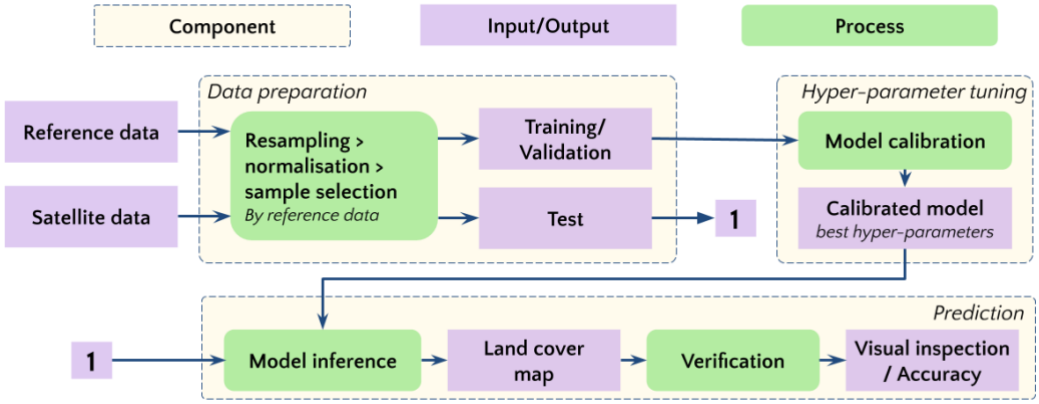


Figure 5.3 Diagram of the procedure for generating and validating LC maps and derived post-loss LC change trajectories.

Following the workflow described in **Chapter 4**, some data preparation operations, which are further described in **Appendix III,** were conducted prior to model training including i) resampling reference and satellite data to the scale of interest, in this case at the spatial resolution of the reference deforestation dataset; ii) creation of a suitable labelled dataset derived from the assessment in Chapter 4 and iii) normalisation of the satellite data. A data partition is then performed according to a tile-wise design in which tiles are distributed over specific regions of the study area.

After above operations, the proposed and traditional supervised learning algorithm as described in **Section 5.4.2** are calibrated according to a hyper-parameter configuration per algorithm. The best models per algorithm, which were determined according to a model selection approach (see **Section 5.4.3**), are evaluated using quantitative (i.e. overall accuracy) and qualitative aspects (i.e. visual inspection) (see **Section 5.4.4**) over test areas not seen during models' training.

5.4.1 Datasets

In this section, the datasets used for the experiments are described. The datasets are composed of multispectral optical satellite images (input data) and reference LC data (so called labelled dataset). Unlike the settings used in **Chapter 4**, the assessment of the best hyper-parameters for the MTLCC network and conventional algorithms in this chapter was conducted using satellite observations from three consecutive years, 2001, 2002 and 2003. This multi-year training strategy aims to increase generalisation of the model to predict over years with changing conditions (e.g. weather) (Bailly *et al.*, 2017). In addition, these years match the period before 2004 which is the first year of Terra-i's deforestation predictions.

Table 5.1 summarises the main features of input and reference LC datasets used in this work. Further description of the pre-processing conducted over these datasets can be found in **Appendix III**.

Table 5.1 Description of the input datasets.

Short name	Name	Number of layers	Source (and version)	Spatial and temporal resolution	Values range
X_{250m}	MODIS 250-m	2	MOD09Q1 v6	250-m 8-days	[-100, 16000]
X_{500m}	MODIS 500m-	5	MOD09A1 v6	500-m 8-days	[-100, 16000]
X_{doy}	Day of the year	1	MOD09Q1 v6	250-m 8-days	[1,361]
Y	MODIS FAO-LCCS2	1	MCD12Q1 v6	500-m Annual	[0,11]

5.4.2 Models

This section introduces the MTLCC network and corresponding hyper-parameters. In addition, the hyper-parameters explored for Random Forest (RF) and the Support Vector Machine (SVM), two shallow machine learning methods widely used in LC classification using remote sensing data (Khatami *et al.*, 2016; Wulder *et al.*, 2018), are described.

It is worth to mention that the above shallow models are not the most optimal to compare as they do not consume and process the information at the same manner as the MTLCC network. For instance, whilst the MTLCC is designed to extract in a joint manner spatiotemporal information from SITS patches with dimensions of D pixels x D pixels, the original configuration of RF and SVM models requires to manually configure this information (so called feature engineering) in the input vector. Moreover, the framework (TensorFlow) to deploy the MTLCC network allows processing a larger amount of data which is commonly a requirement to train this sort of deep neural network model. Due to limitations of this research to explore other learners and infrastructure to process them, it is suggested that future work includes comparison against comparable deep neural network algorithms as it is reflected in **Section 7.2.4.7.2.4**

5.4.2.1 MTLCC network

The MTLCC network as introduced in **Section 4.5.1** encodes a satellite time series to increasingly higher level D dimensional representations through N cascaded bidirectional convolutional recurrent layers. The hidden states are initialized as zero-vectors. The last hidden states from forward h_T^{seq} and backward passes h_T^{rev} are concatenated $h_T = [h_T^{seq} \| h_T^{rev}]$. The concatenated representation h_T is then projected to softmax-normalized feature maps for each class using a convolutional layer.

The architecture has a large number of hyperparameters that can be changed and adapted. In this work, the hyper-parameters tuned were 1) batch size $B \in \{8,16,24,36,40\}$ 2) number of hidden states $D_h \in \{24,48,64,128,256\}$3) learning rate $lr \in \{0.1,\cdots,0.000001\}$and 4) optimiser type $O \in \{adam, nadam\}$. The space explored by hyper-parameter is based on previous experimentation i.e. trial and error as indicated in **Section 4.5.6**. For instance, the settings of the MTLCC network trained in **Chapter 4** were $B = 32$, $D_h = 64$, $lr = 0.01$ and $O = adam$. Due to optimisation playing an important role in the training of neural networks such as the MTLCC model, the Nadam optimizer (Dozat, 2016) was added to the hyper-parameter space of optimisers. Nadam is an adapted version of Adam which uses a Nesterov accelerated gradient to calculate adaptive learning rates for different parameters. Further information of the differences of Nadam and Adam can be

found in Kingma and Ba (2014). Nadam has been used in remote sensing studies, including LC classification in the tropics (Parente *et al.*, 2019).

Other parameters of the MTLCC network such as number of layers N and the sizes of the convolutional k_{rnn} and classification kernel k_{class} were not explored. According to previous work by Rußwurm and Körner (2018b), the number of hidden states seems to be more relevant than the depth (number of layers) for the task of classification. To the same extent, kernel values higher than 3x3 for either k_{rnn} or k_{class} do not provide any gain in terms of accuracy but do require more computation resources (Rußwurm and Körner, 2018b).

5.4.2.2 Random Forest (RF)

RF is a non-parametric ensemble algorithm which combines many decision tree classification models (Breiman, 2001). Each tree is built by performing an individual learning algorithm splitting the input variable set into subsets based on an attribute values test e.g. the Gini coefficient. Compared with classical Decision Trees (DT), there are no pruning procedures in RF trees. Instead, they are built by randomly selecting at each node a subset of input variables. As a result, the computational complexity of the algorithm is reduced, and the correlation between trees also decreases. The split process decorrelates the various RF trees and consequently it makes the classification results less variable and more reliable (Pelletier *et al.*, 2016).

The hyper-parameters tuned for RF were the number of trees $n_{tree} \in \{20,\cdots,2000\}$, the number of features to be considered at every split $N_f \in \{auto, sqrt\}$, the maximum depth $max_{depth} \in \{10,\cdots,110\}$, the minimum number of samples required to split a node $min_{node} \in \{2,\cdots,10\}$, the minimum number of samples required at each leaf node $min_{leaf} \in \{1,\cdots,4\}$, and whether or not to use bootstrapping $bootstrap \in \{true, false\}$.

5.4.2.3 Support Vector Machine (SVM)

SVM is a non-parametric algorithm which aims to determine the location of decision boundaries (also known as hyperplane) producing optimal separation of classes (Cortes and Vapnik, 1995). To achieve it, SVM tries to maximize the distance between the hyperplane and the closest instances to this hyperplane.

The resulting closest instances are called support vectors. The SVM classifier is often formulated as a quadratic optimization problem. When searching for the best hyperplane, the algorithm accepts some misclassification errors. This procedure is controlled by slack variables which tolerate misclassification and more specifically by a regularization parameter C, also named penalty error (Pelletier *et al.*, 2017).

In addition to C, multiple kernel functions exist for SVM in order to embed the data in higher dimensional space where a linear separation can be found. These functions allow computation of dot products in a higher dimensional space without transforming input data. Amongst types of kernels, the Radial Basis Function (RBF) is the most common in remote sensing classification. Connected to this particular kernel, the parameter γ controls the width of the Gaussian kernel function in the SVM-RBF model (Pelletier *et al.*, 2017).

According to the hyper-parameters introduced above, the search space of the SVM-RBF model was $C \in \{0.01, \cdots, 1000000\}$ and $\gamma \in \{0.01, \cdots, 100\}$.

5.4.3 Model selection

The selection of the best model varied between the MTLCC model and shallow learners in terms of dataset size, techniques and hardware used. For the MTLCC network, due to the large number of pixels in training and validation splits (26.1 and 6.6 millions respectively) in the full dataset, the hyper-parameter search was conducted using a two-stage strategy. This strategy was implemented by using notions of transfer learning as introduced in **Section 5.2.2** with an aim to reduce the computation time, and consequently associated costs of cloud services. The first stage was aimed at finding the optimal hyper-parameters using a representative sample from the full dataset. Subsequently, the knowledge acquired from the first stage was used to refine the best hyper-parameters using the full dataset. The representative dataset was obtained by testing different proportions of samples randomly selected from the full dataset's training and test tiles. It was found that by using at least half of the tiles per split, a close distribution to the full dataset in terms of the input features per LC class is ensured. **Table 5.2** compares the number of pixels and respective proportions per LC class in the full and sampled dataset over training and validation tiles. **Figure** 5.4 shows the

spatial distribution and example of the distribution of one of the features per class from pixels of sampled and full dataset.

Table 5.2 Distribution of the pixels in the full and sampled dataset in the training and validation tiles used to find the best hyper-parameter for the MTLCC network.

Class	Full dataset		Sampled dataset	
	N pixels	Proportion (%)	N pixels	Proportion (%)
Barren	687	0.00	376	0.00
Water Bodies	443,179	1.77	283,403	2.13
Urban/Built-up Lands	41,179	0.16	28,008	0.21
Dense Forests	9,609,537	38.30	4,827,641	36.34
Open Forests	10,794,104	43.02	5,783,581	43.54
Natural Herbaceous	3,759,148	14.98	2,067,666	15.57
Herbaceous Croplands	436,815	1.74	285,660	2.15
Shrublands	7,272	0.03	6,716	0.05

The implementation of the above strategy was conducted using a cloud infrastructure as detailed in **Section 5.4.5**. This particular infrastructure provides a Bayesian optimisation algorithm based on Gaussian Process (GP) model $F(x)$ of the objective function $f(x)$ (Golovin *et al*., 2017). In terms of transfer learning, $F(x)$ are a stack of GP models where each layer predicts residuals between its training data and the estimate produced by lower layers. This allows the correcting and enriching of each layer and the estimate produced by lower layers. The particular implementation of this approach by the cloud provider is also flexible in transferring learning between different datasets which commonly present shifts and scalings of the objective value (Kochanski *et al*., 2017).

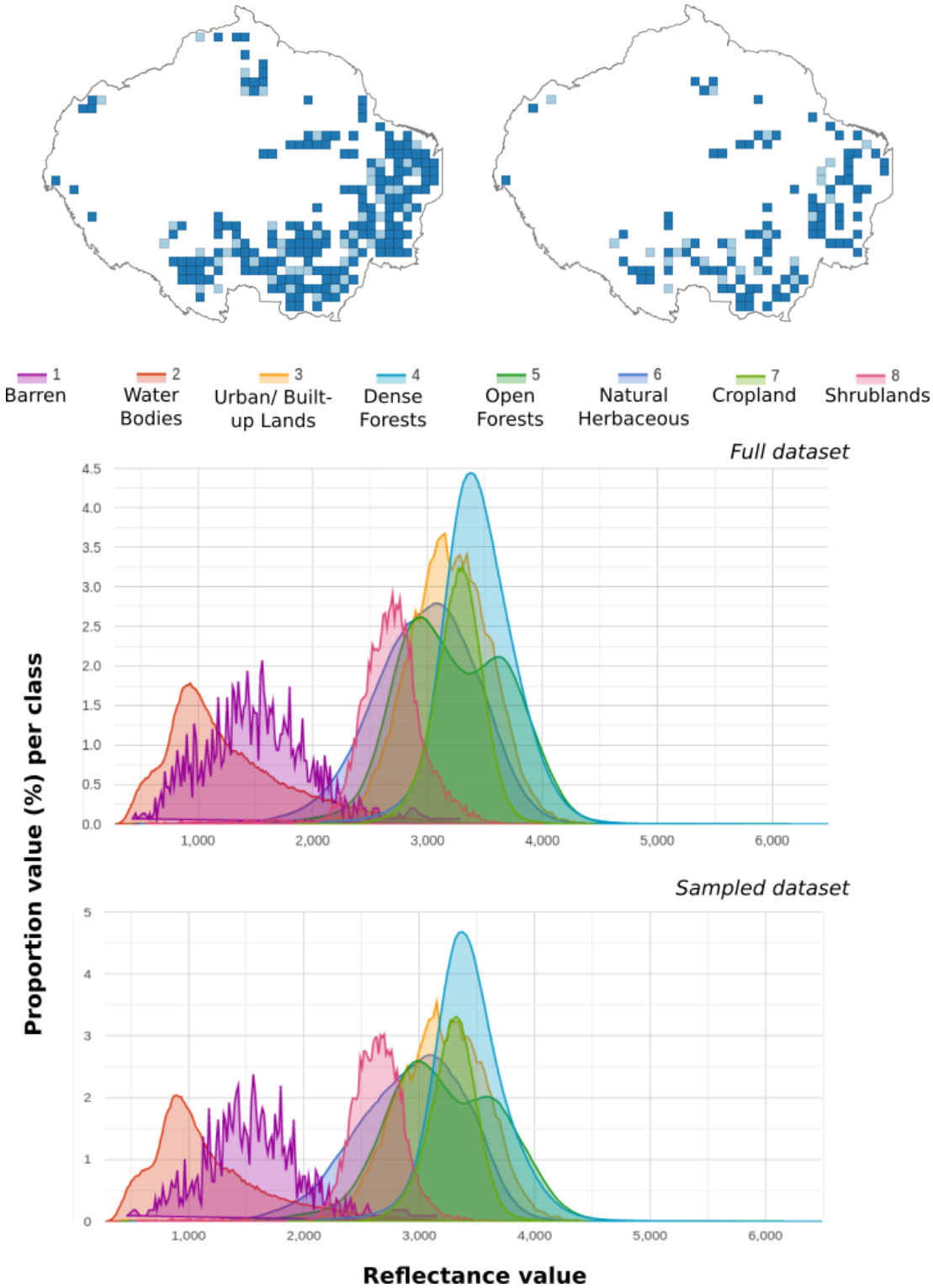


Figure 5.4 (top) Spatial distribution of the training (blue) and validation (light blue) tiles in the full (left) and sampled (right) dataset; and (bottom) distribution of the mean NIR computed per class from 46 MODIS images in 2003 according to the training split in the full and sampled 2001-2015 FAO-LCCS2 dataset.

In contrast to the approach for the MTLCC network, the hyper-parameters of the RF and SVM models were determined through a Bayesian optimisation method

based on tree-structured Parzen density estimators (TPE) as introduced in **Section 5.2.1.2**. The rationale of utilising different methods between the MTLCC and shallow learners was due to limited resources for cloud computing. Therefore, the available credits were exclusively allocated for conducting the experiments related to the MTLCC network. The implementation of TPE for RF and SVM models was performed in a local machine using a hyper-parameter optimization library as described in **Section 5.4.5**. To reduce tuning time and class imbalance in the data, only up to 3000 samples per class were used. These samples were randomly taken from training and validation tiles from the full dataset of three years, 2001, 2002 and 2003. The input feature for shallow learners consists of a vector representing flattened satellite time series of the spectral bands of MODIS at a spatial resolution of 250-m. To this vector, the date related to the MODIS images for each year was added. As SVM models are sensitive to the scale of the input variables (Graf *et al*., 2003), these input features were normalised and scaled as described in **Section 0**. Scaling is mostly important in algorithms that are distance based and require Euclidean Distance such as SVM but not for RF which is unaffected by monotonic transformations of individual features (Hastie *et al*., 2009).

5.4.4 Performance evaluation

The performance of the MTLCC network and shallow learners was compared by computing traditional metrics for assessing machine learning classifiers – e.g. overall accuracy, recall, precision, f-score –and others with a more remote sensing focus – e.g. quantity disagreement and allocation disagreement – as conducted in **Chapter 4** (see **Section 4.5.5.1**).

The optimal hyper-parameters were determined separately by classifier using the multi-spectral MODIS time series and the labelled datasets with eight LC classes, as described in **Section 5.4.1**. The Overall Accuracy (OA) was used to rank the classifiers' performance according to different hyper-parameter configurations. This metric was computed using the validation partition (tiles) during training. For each classifier, the best, second-best and third-best models according to hyper-parameter optimisation were evaluated using the test partition (tiles) as shown in **Appendix III**. In addition to OA, F1-score, quantity disagreement and allocation disagreement metrics were used to determine the performance of the best

models per classifier for three consecutive years, 2001, 2002 and 2003 which were the same years used for training.

5.4.5 Software and implementation

For the implementation of the MTLCC network, shallow learners and data preparation scripts, python 3.6 was used. Thanks to a research grant awarded by the Google Cloud Platform, £800 in cloud credits were used to deploy the MTLCC network using Google's AI platform and TensorFlow (TF) Estimator API version 1.14.

The AI platform can be categorised as an Infrastructure as a Service (IaaS) according to the cloud services introduced in **Section 5.2.3**. This platform facilitates plugin ML-based workflows to preconfigured virtual machines. This allows a focus on model development and deployment rather than infrastructure management. The charging cost of the AI platform is primarily based on the consumption of training units measured by the combination of computing resources such as CPU, GPU and memory. Predefined SCALE TIERS are offered by the platform and their number of training units varies according to the complexity of resources. For instance, whilst a BASIC scale tier consists of 1 training unit i.e. a single worker, a PREMIUM_1 scale tier consists of 75 ML training units i.e., many workers and parameter servers (Zhang *et al.*, 2018). CUSTOM scale tiers can be also defined according to the user needs. For example, the training of the MTLCC network models for the hyper-parameter optimisation exercises was conducted using a CUSTOM scale tier configured with a single master, 8 virtual CPUs, 52GB of RAM and a single GPU NVIDIA Tesla P100.

The adapted repository of the MTLCC network described in **Chapter 4** was ported to TF Estimator API version 1.14. The integration with the Estimator API facilitates the training and assessment of the target hyper-parameters of the MTLCC network in Google's AI platform. **Figure 5.5** presents a highlight-level overview of the train/deploy process on the AI platform.

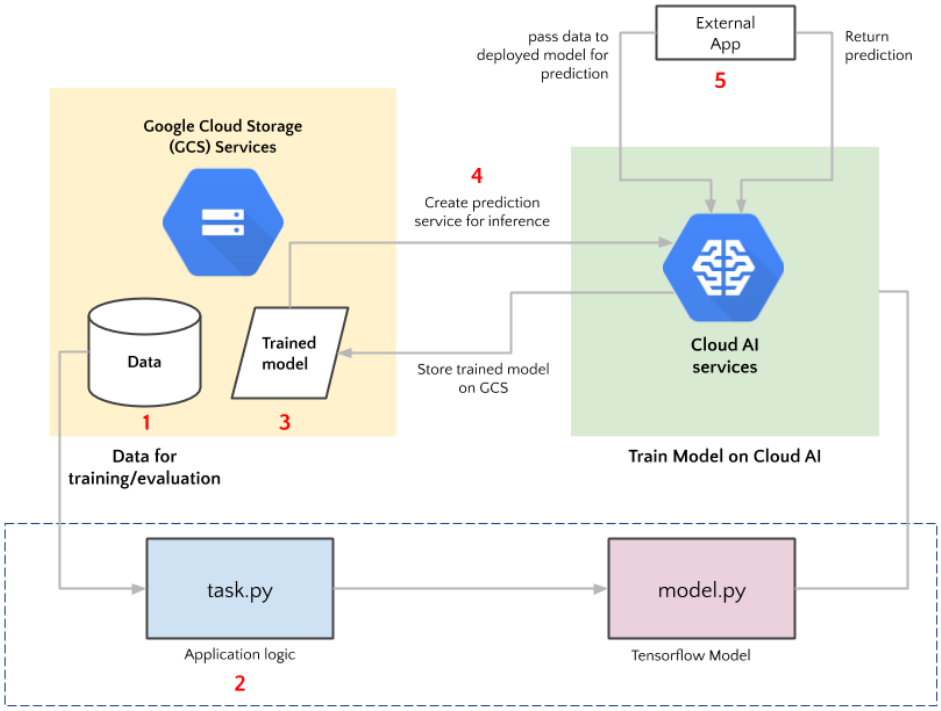


Figure 5.5 Illustration of the train/deploy process on Google AI platform. The numbers relate to main steps in the process including 1) store the data for training and inference on Google Cloud Storage (GCS); 2) execute scripts related to the model based on an application logic using the training data as specified in the task.py file; 3) store the trained model on GCS; 4) serve the trained model created on Cloud AI; and 5) predict using the served model from an external application sending data. Adapted from Bisong (2019).

For hyper-parameter optimisation, the AI platform is based on Google's Vizier as internal service. This service was built to be scalable with minimal user configuration and setup (Golovin *et al.*, 2017). To set a study, a configuration file in a human-readable format is created and added to the package of training scripts uploaded to Google's AI platform. This file indicates the target hyper-parameters, their ranges of values and the metric to optimise. In addition, the number of tests i.e. trials for the hyper-parameter optimisation must be specified in the *maxTrials* field. To reduce the time in the study the trials can be run in parallel by specifying the *maxParallelTrials*. In sequential algorithms such as Bayesian optimisation, it is preferred to use small values in this particular field due to the algorithm benefitting from a previous run. In contrast, for non-sequential algorithms such as grid search as described in **Section 5.2.1.1**, multiple runs in parallel are preferred as optimal hyper-parameters are tuned irrespective of the results of previous runs. A further parameter that reduces computing time is *enableTrialEarlyStopping* which as the name indicates stop unpromising trials based on the approach described in Golovin *et al.* (2017).

The two-stage strategy described in **Section 5.4.3** for the MTLCC network was implemented, modifying the above configuration file. In the first stage, models were trained up to 10 epochs using the sample dataset. The *maxTrials*, *maxParallelTrials* and *enableTrialEarlyStopping* were set to 10, 1 and FALSE, respectively. Then, for the second stage using the full dataset, the *maxTrials*, *maxParallelTrials* and *enableTrialEarlyStopping* were set as the first stage, except maxTrials which was changed to 5 and models were trained up to 7 epochs. This reduction in the number of training epochs was compensated by the number of instances i.e. the full dataset contains two times the instances (pixels) than the sampled dataset. The knowledge between studies was specified by adding the *resumePreviousJobId* field indicating the job ID of the previous trial. Below is an example of a file configured for the hyper-parameter optimisation of the MTLCC network at the first stage.

```
trainingInput:
 hyperparameters:
   maxTrials: 10
   maxParallelTrials: 1
   enableTrialEarlyStopping: False
   goal: MAXIMIZE
   hyperparameterMetricTag: accuracy
   params:
   - parameterName: batchsize
     type: DISCRETE
     discreteValues: [8, 16, 24, 32, 40]
   - parameterName: learning_rate
     type: DOUBLE
     minValue: 0.000001
     maxValue: 0.1
     scaleType: UNIT_LOG_SCALE
   - parameterName: convrnn_filters
     type: DISCRETE
     discreteValues: [24, 48, 64, 128, 256]
   - parameterName: optimizertype
     type: CATEGORICAL
     categoricalValues: [adam, nadam]
```

The hyper-parameter optimisation for RF and SVM models was conducted using *scikit-learn* version 0.19.1 and *hyperopt* version 0.2.2, both open-source packages in python. *scikit-learn* is a well-established package for training supervised and unsupervised ML-based methods. *hyperopt* provides algorithms and parallelisation infrastructure for hyper-parameter optimisation. In particular, the Bayesian optimisation method based on tree-structured Parzen density estimators (TPE) was used to find the optimal hyper-parameters for RF and SVM models. The deployment of a study retains similarities to the Google AI platform due to the ability to define some parameters such as the number of trials, hyper-parameter configuration and resume of previous job. Further description of the principles and steps of *hyperopt* can be found in Bergstra *et al*. (2013). The number of trials for RF and SVM was defined as 100 and 25, respectively with three-fold cross-validation. More trials were allocated to RF due to it having considerably more target hyper-parameters (six) than SVM (two).

Once training was completed, the evaluation was run locally on a server. The performance metrics as described in **Section 5.4.4** were computed using the library *scikit-learn* version 0.19.1 The details of the resources and the packages used by step are summarised in **Table 5.3**.

Table 5.3 Lists of hardware and software used in this study organised by step.

Step	Hardware and/or software	Specifications
Data generation and creation of the validation set	Google Earth Engine (GEE) Google Cloud Platform (GCP)	GEE is a cloud-based platform used to access and download input datasets (see **Section 5.4.1**) exported as GZ TFrecords to GCP.
Data pre-processing Model training for RF and SVM models Generation and assessment of the predictions	Server scikit-learn hyperopt TensorFlow	A server was used for data storage and input data preparation. Its main features are: Memory: 128 GB, Processor: Intel Xeon E5-2667 v4 3.20, Operating system: Windows Server 2016 X64, GPU: (1) NVIDIA Tesla M60 This server was used to run locally the hyper-parameter optimisation method for RF and SVM models using scikit-learn and hyperopt
Training of the MTLCC network	Google AI TensorFlow	Google AI was used to find the optimal parameters. The settings configured are: A single master type using n1-highmem-8 (8 virtual cores and 52GB of memory) GPU: (1) NVIDIA Tesla P100

5.5 Results

In this section, calibrated MTLCC networks are compared with the two calibrated shallow learners, RF and SVM. The results are presented in four parts. Firstly, in **Section 5.5.1**, the results of the hyper-parameter optimisation are described per classifier. Secondly, in **Section 5.5.2,** the quantitative results derived from the performance metrics for the best, second- and third-best models per classifier are presented. Thirdly, in **Section 5.5.3** a visual inspection of predictions by the best model per classifier are compared, in particular across areas with high and low disagreement between the MTLCC network and shallow learners. Fourthly, **Section 5.5.4** aims to highlight the ability of the MTLCC model to ignore noise, e.g., induced by clouds, from raw satellite time series data. This procedure is performed by visualising the internal gates of the hidden units of the network as proposed in previous work (Rußwurm and Körner, 2018b). Finally, the capability of using the model trained in the study area in order to predict in other areas of the pantropics is presented in **Section 5.5.5**.

5.5.1 Hyper-parameter configurations

The optimal hyperparameters for model selection as described in **Section 5.4.3** are presented separately by classifier. For the MTLCC network, the two-stage strategy facilitates an efficient usage of the cloud resources through a Bayesian optimisation technique provided by Google's AI platform. **Table 5.4** describes the optimal hyper-parameters and corresponding overall accuracy measured over the validation partition at the end of the training process according to the trials set at each stage.

Table 5.4 Results of the trials i.e. simulations for the first and second stages designed to find optimal hyper-parameters to the MTLCC network. Trials per stage are organised by overall accuracy (OA).

(a) First stage: 10 trials run over 10 epochs using the sampled dataset.

Rank	OA	Batch size	Learning rate	Hidden states	Optimiser
1	0.89	40	6.5E-02	128	nadam
2	0.88	40	3.9E-02	128	nadam
3	0.85	8	3.9E-05	128	nadam
4	0.84	32	5.0E-06	256	nadam
5	0.83	16	1.7E-05	256	nadam
6	0.82	32	6.3E-02	128	nadam

7	0.80	16	2.0E-05	64	adam
8	0.78	8	7.7E-02	24	nadam
9	0.75	32	2.0E-06	24	adam
10	0.71	24	2.6E-02	24	adam

(b) Second stage: 5 trials run over 7 epochs using the full dataset.

Rank	OA	Batch size	Learning rate	Hidden states	Optimiser
1	0.90	40	5.7E-02	128	nadam
2	0.90	24	9.0E-06	128	nadam
3	0.87	16	4.0E-06	128	nadam
4	0.87	40	9.8E-02	128	nadam
5	0.85	40	6.1E-02	128	adam

The results above show the logic of the Bayesian hyper-parameter optimisation technique provided by Google's AI platform. Whilst the first stage explores a broad hyper-parameter space using the sample dataset, the second stage refines the search using the full dataset. The best model in the first stage yields an overall accuracy (OA) of 0.89 using a batch size of 40, 128 hidden states, learning rate of 6.5E-02 and nadam optimiser. The best trial at the second stage yields a slightly better OA (0.90) than the best model at the first stage (0.89). The hyper-parameter search suggests 128 as the optimal number of hidden states to train the MTLCC network. The optimal values for the learning rate and batch size are more variable without a clear pattern. For the optimiser, Nadam is dominant indicating it yields better results than the Adam optimiser.

Due to the large number of trials tested for the shallow learners, Random Forest (RF) and Support Vector Machine (SVM), 25 and 100, respectively, only the three best hyperparameter sets are reported in **Table 5.5**.

Table 5.5 Results of the trials i.e. runs for finding the best hyper-parameters for the shallow learners, RF and SVM.

(a) RF

Rank	OA	Number of trees	Max features	Max depth	Min samples split	Min samples leaf node	Bootstrap
1	0.76	1240	sqrt	11	6	3	True
2	0.76	550	sqrt	11	3	3	True
3	0.76	980	sqrt	11	2	3	True

(b) SVM

Rank	OA	C	gamma
1	0.74	1	1
2	0.74	1	0.1
3	0.71	1000000	1

In addition to the above tables, **Figure 5.6** presents the distribution of all trials tested for RF and SVM. These charts confirm certain patterns in the optimal values per hyper-parameter. For RF, sqrt and true are optimal settings for the maximum number of features $N_f \in \{auto, sqrt\}$ and Bootstrap $bootstrap \in \{true, false\}$, respectively. These settings were prevalent in the best trials irrespective of the dataset. From the search space explored for the number of trees $n_{tree} \in \{20,\cdots,2000\}$, the optimal values are located from 500 to 1500. For the values explored in the maximum depth $max_{depth} \in \{10,\cdots,110\}$ and the minimum number of samples $min_{node} \in \{2,\cdots,10\}$, the models favour small values.

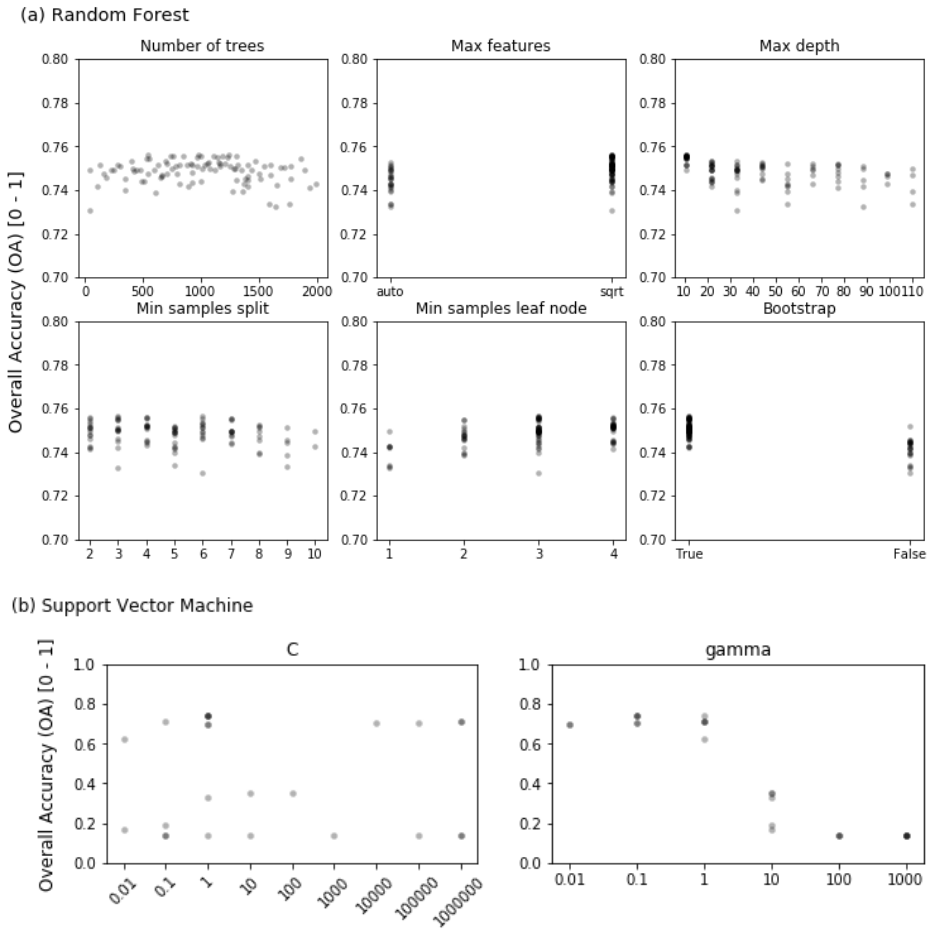


Figure 5.6 Overall accuracy (OA) of all trials experimented for (a) Random Forest; and (b) Support Vector Machine (SVM). Each black dot corresponds to a trial i.e. run.

5.5.2 Quantitative evaluation

The previous section described the optimal hyper-parameters per classifier according to the validation partition. In this section, the best, second-best, and third-best trials per classifier are evaluated using the test partition. In **Table 5.6**, the mean and standard deviation of four performance metrics as described in **Section 5.4.4** are presented. These metrics were computed by comparing the reference data derived from the 2001-2015 MODIS FAO-LCCS2 dataset and predictions in 2001, 2002 and 2003 by classifier.

Table 5.6 Comparison of the performance of two shallow learners, Random Forest (RF) and Support Vector Machine (SVM), and the MTLCC network trained according to 2001-2003 multi-spectral MODIS satellite images and labels extracted from the 2001-2015 MODIS FAO-LCCS2 hybrid dataset over the test tiles. The mean and standard deviation of three models with the best hyperparameter sets are reported per classifier per year. The best values per metric per year are highlighted in bold.

Metric	RF			SVM			MTLCC		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
OA	0.81 ±0.00	0.78 ±0.00	0.79 ±0.00	**0.82 ±0.01**	**0.82 ±0.01**	**0.83 ±0.01**	0.76 ±0.05	0.75 ±0.04	0.74 ±0.03
F1-Score	0.53 ±0.00	0.51 ±0.00	0.52 ±0.00	**0.56 ±0.02**	**0.56 ±0.02**	**0.57 ±0.02**	0.46 ±0.04	0.45 ±0.03	0.44 ±0.02
Quant. disagm	9.36 ±0.01	11.29 ±0.05	10.17 ±0.05	7.93 ±1.00	7.96 ±1.12	**6.99 ±1.12**	**6.19 ±0.71**	**7.16 ±1.86**	7.00 ±2.39
Alloc. disagm	**9.89 ±0.02**	10.55 ±0.01	10.93 ±0.01	9.95 ±0.57	**9.92 ±0.53**	**10.02 ±0.53**	17.48 ±6.87	17.83 ±6.22	18.51 ±6.22

According to the values reported above, there are no remarkable differences in the performance of the three classifiers during the years assessed. The standard deviation denoted per metric per model shows predictions by RF followed by SVM are more stable irrespective of different hyper-parameter values. Overall, SVM followed by RF models yield the best values in overall accuracy, F1-Score and allocation disagreement. The quantity disagreement is the only metric in which the MTLCC network demonstrates the smallest values in two (2001 and 2002) out of the three years. It is important to note that the results of the MTLCC are not satisfactory mainly due to the three best trials being trained over too few epochs. This is evident as MTLCC's standard deviation of most metrics, in particular allocation disagreement, is considerably higher than in RF and SVM models. Although the number of training instances (pixels) per epoch is large (20.1 million pixels per year), 7 epochs are not enough to show stable predictions

in the MTLCC network. To support the relevance of the impact due to number of training epochs used, the best trial of the MTLCC network according to the test partition was trained and monitored with up to 30 epochs (see **Table 5.7**). This number of epochs was defined as being optimal as results showed no improvements in terms of overall accuracy and cross-entropy loss were seen by increasing epochs to more than 30.

Table 5.7 Performance metrics monitored for the best model of the MTLCC network (128 hidden units, learning rate of 9E-06, batch size of 24, and Nadam optimiser) trained over 30 epochs.

Metric	10 Epochs			20 Epochs			30 Epochs		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
OA	0.90	0.90	0.90	0.91	0.91	0.92	0.91	0.91	0.92
F1-Score	0.55	0.56	0.57	0.58	0.58	0.59	0.60	0.59	0.60
Quant. disagm	3.53	3.31	2.12	1.18	1.58	1.38	0.93	1.61	1.07
Alloc. disagm	6.93	7.03	7.40	7.90	7.67	7.10	8.15	7.36	7.37

A comparison of the per-class mapping accuracies according to F1-score between the best models of RF, SVM and MTLCC network is presented in **Table 5.8** . The best model of the MTLCC network trained over 30 epochs reports the most accurate results according to the class-wise F1-score in six out of eight LC classes. For most of these classes, the MTLCC network yields superior values than the shallow learners. The most notable increase in F1-score can be seen in *Water Bodies and Open Forests* with values above 91%, being much higher than the shallow learners which have values below 85%. For *Natural Herbaceous*, *Herbaceous Croplands and Urban/Built-up*, SVM presents higher values than RF but they are still remarkably lower than MTLCC. Whilst RF and SVM show some capabilities to map minor LC classes such as *Barren* and *Shrublands*, the MTLCC network is not capable of distinguishing them. The low discrimination of these minor classes might be attributed to the class imbalance and their small proportion within the training dataset (less than 0.05%).

In terms of the global metrics, the MTLCC network yields the highest F1-Score and smallest values in both disagreement components. From **Table 5.8**, the MTLCC has a difference with SVM and RF in average F1-score between 2001 and

2003 of 2.23% and 7.89%, respectively. For the disagreement components, the MTLCC model has reductions in quantity disagreement in comparison with SVM and RF of 5.59% and 9.03%, respectively, and 1.79% and 2.83% in allocation disagreement, respectively.

Table 5.8 Per-class and average F1-score comparison between RF, SVM, and the MTLCC network method for mapping eight-LC target classes. To facilitate the discussion between models, per-class F1-score, average F1-score and disagreement components are reported in percentage. The best values per class and global metrics are highlighted in bold.

Land cover	RF			SVM			MTLCC		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
Barren	22.06	21.96	24.45	**34.22**	**35.15**	**35.37**	0	0	0
Water Bodies	83.61	82.68	84.38	82.8	81.69	83.27	**91.50**	**92.39**	**91.90**
Urban and Built-up	5.1	4.27	4.7	11.42	11.88	11.58	**41.12**	**30.49**	**36.29**
Dense Forests	92.42	91.47	91.24	92.25	91.57	91.87	**95.72**	**95.31**	**95.72**
Open Forests	84.22	82.69	82.73	83.45	83.17	84.45	**90.90**	**90.94**	**91.61**
Natural Herbaceous	67.54	61.97	64.83	76.02	76.18	77.54	**82.46**	**83.79**	**83.90**
Herbaceous Croplands	67.32	63.36	63.92	75.86	77.21	76.13	**79.62**	**81.63**	**84.08**
Shrublands	0.97	1.07	1.15	**3.34**	**5.42**	**4.18**	0	0	0
Average F1-Score	52.9	51.19	52.17	57.42	57.78	58.05	**60.17**	**59.32**	**60.44**
Quantity disagm.	9.35	11.23	10.12	7	7.18	6.19	**0.93**	**1.61**	**1.07**
Allocation disagm.	9.9	10.56	10.92	9.4	9.34	9.51	**8.15**	**7.36**	**7.37**

5.5.3 Visual inspection

LC classification results for seven test tiles by the best models of RF, SVM and MTLCC network are shown in **Figure 5.7**. The white and yellow circles highlight incorrect and correct classification, respectively. The best model of RF misclassified *Dense Forests* in particular over areas with high recurrence of clouds as shown in h10v08 and h10v09 tiles in **Figure 5.7**. Despite the best model of SVM seeming to be more robust to the noise than RF in these areas, it notably confuses *Herbaceous Cropland*, *Natural Herbaceous* and *Open Forests* with *Urban/Built-up Lands*. Regarding the MTLCC network, it provides in general, the closest match to the reference data observations including accurate LC classification over noisy, i.e. cloudy, areas. Most of the misclassifications by the

MTLCC network are associated with *Herbaceous Cropland* which is confused with *Open Forests* and *Natural Herbaceous*. Two areas, h10v09 and h13v09 tiles in **Figure 5.7**, show this confusion. In both tiles, the reference data observations indicate the dominance of *Dense Forests* and *Open Forests*. This dominance might cause difficulty in differentiating *Herbaceous Cropland* over these areas.

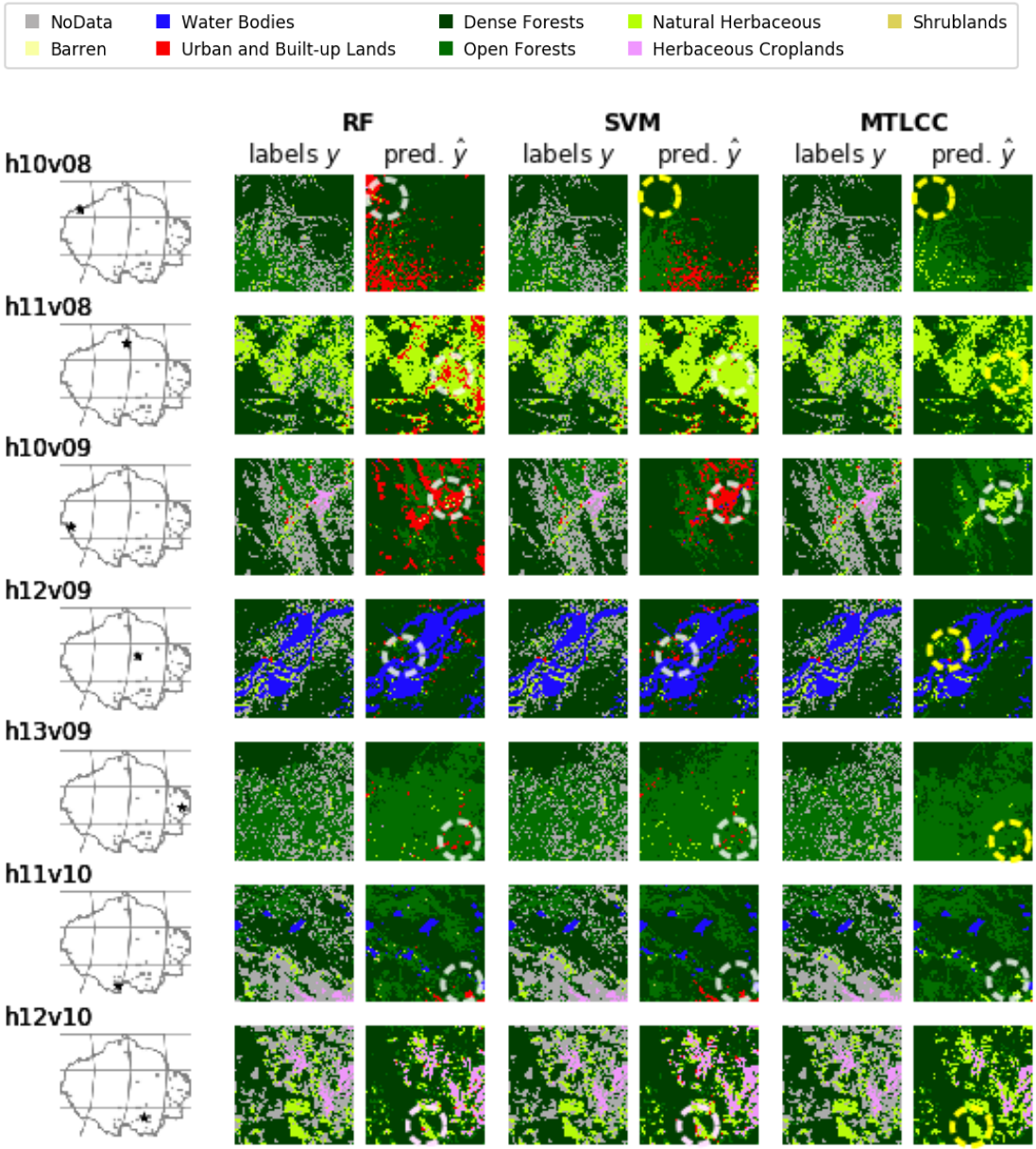


Figure 5.7 Reference labels y and predictions $\hat{y}$ by the best models of RF, SVM and MTLCC for year 2001 across seven test tiles distributed across the study area. *NoData* represents pixels that were not analysed during either training or evaluation of the models. The white and yellow circles denote incorrect and correct classification, respectively. For instance, in RF and SVM, *Open Forests* pixels are incorrectly classified as *Urban and Built-up* over most of the areas particularly h10v08, h10v09, h11v10 and h12v10. The presence of clouds in h10v08 as highlighted by the white circle was discriminated by SVM and MTLCC network but not RF.

5.5.4 Visualisation and analysis of the internal gates

In previous work, the visualisation of the internal gates of LSTM cells provided a better approximation of how these units contribute to the classification capability of the MTLCC network using Sentinel-2 data (Rußwurm and Körner, 2018b). This section aims to corroborate how the internal gates of GRU cells encode MODIS satellite image time series for the task of classifying 8 target LC classes. In contrast to LSTMs, GRUs formulate a single, but more sophisticated, output vector. Furthermore, it uses only two gating units, namely the reset r_t and update u_t gates, to adaptively capture dependencies over different time scales. In addition, both structures control how much memory and information is read and written to the output (Rußwurm and Körner, 2018b).

The inspection of the above gates is conducted over a subset of 192 pixels per 192 pixels from the tile h10v08 shown in **Figure 5.7**. This particular tile and related subset have a high cloud cover which can affect the performance of a classifier as demonstrated with the best model of RF. For sake of simplicity, the activations of the internal components of 3, which are denoted by the raised index $i \in \{12,90,123\}$, out of 128 hidden units of the best MTLCC model, are analysed over a sequence of 45 MODIS satellite images $t \in \{1,\cdots,45\}$ in 2001 (see **Figure 5.8**). Higher values of either r_t and u_t gates are directly related to increases in the influence of the previous state. If both values are low, the state in the corresponding locations is rewritten with a new value. After iteratively processing the sequence, the final state h_T is used to produce activations for each class, as described in **Section 4.5.1**.

The activations of two out of the three hidden units $i \in \{12,90\}$ are uniformly changing over time. In agreement with previous work visualising LSTMs (Rußwurm and Körner, 2018a, 2018b), it remains difficult to associate relationships of the activations of the majority of the hidden units. It is assumed that the combination of all activations of the network structure contribute to the extraction of classification-relevant features. Furthermore, certain spatial patterns of the clouds were notably identified along the span of MODIS images for certain cells. For instance, one hidden unit $i \in \{12\}$ clearly shows how the update gate u_t approaches zero at certain time steps over the entire tile (e.g. $t \in \{20\}$) or certain areas related to clouds (e.g. $t \in \{2,3,18,29,30,31,43,44\}$). In addition, the

changes in the sign of the state or output at certain observation times is an indicator of the direction in which information is rewritten, and consequently contributing to producing activations for each class.

Figure 5.8 Internal GRU cell activations of the reset r_t and update u_t gates and output at three selected hidden units $i \in \{30,90,123\}$ given the current input $x_{(t)}$ over the sequence of observations $t \in \{1, \cdots, 45\}$. The detail of features at the cell states increased gradually, which indicated the aggregation of information over the sequence. While most cells are likely to contribute to the classification decision, only some cells are visually interpretable regarding the current input $x_{(t)}$as shown by $i \in \{30,90\}$.

5.5.5 Transfer learning over other pantropical areas

As previously discussed in **Section 5.2.2**, transfer learning refers to re-using knowledge learned in one situation for another situation. This section aims to present the feasibility in transferring knowledge of the pre-trained MTLCC network over regions different to the study area. Therefore, the best calibrated MTLCC model, which was trained over 30 epochs, was retrained using 2001, 2002 and 2003 MODIS satellite observations and LC reference data derived from 2001-2015 MODIS FAO-LCCS2 layer obtained from three geographically different areas in the pantropics. The technicalities of this particular procedure are described further in **Appendix IV**. In total, six tiles of 384 pixels per 384 pixels were distributed in surrounding areas of (1) Bogota, Colombia in South America; (2) Lisala, Democratic Republic of the Congo in Africa; and (3) Pontianak, Indonesia in Asia (1). Whilst, the areas in Africa and Asia shared a similar type of *Dense Forests*, mostly humid evergreen forest, and low urban density with the Amazon region, the area in South America is dominated by evergreen mountain forests and dense urban areas. The number of tiles per location was defined according to a split ratio of 4:1:1 used to create the partitions for training the MTLCC model across the Amazon region as described in **Appendix III**. This ratio means 4 tiles were used for training, one for validation and one for testing. Both quantitative and qualitative (i.e. visual inspection) assessments of the predictions were conducted over the test tile per location.

To highlight the benefits of training the pre-trained MTLCCC network using new data, the predictions of the pre-trained MTLCC model with/without retraining and the MTLCC network trained from scratch over 30 epochs using the new data were compared. The number of epochs was chosen as no improvements in terms of overall accuracy and cross-entropy loss were seen by increasing the epochs to more than 30, in particular for the pre-trained MTLCC network run using new data. **Appendix IV** contains further details of the quantitative results of four performance metrics, which are described in **Section 5.4.4**, and assessed in three consecutive years per location. **Figure 5.9** presents the visual inspection over the test tile per target area. Overall, the strategy of training the pre-trained MTLCC network achieves the best performance in comparison with the other strategies. This was reflected in higher values in the performance metrics (see **Appendix IV**).

In addition, the spatial configuration of predicted LC classes is more accurate in the retrained MTLCC network than in the other strategies with closer spatial patterns to the reference labels derived from 2001-2015 MODIS FAO-LCCS2 layer.

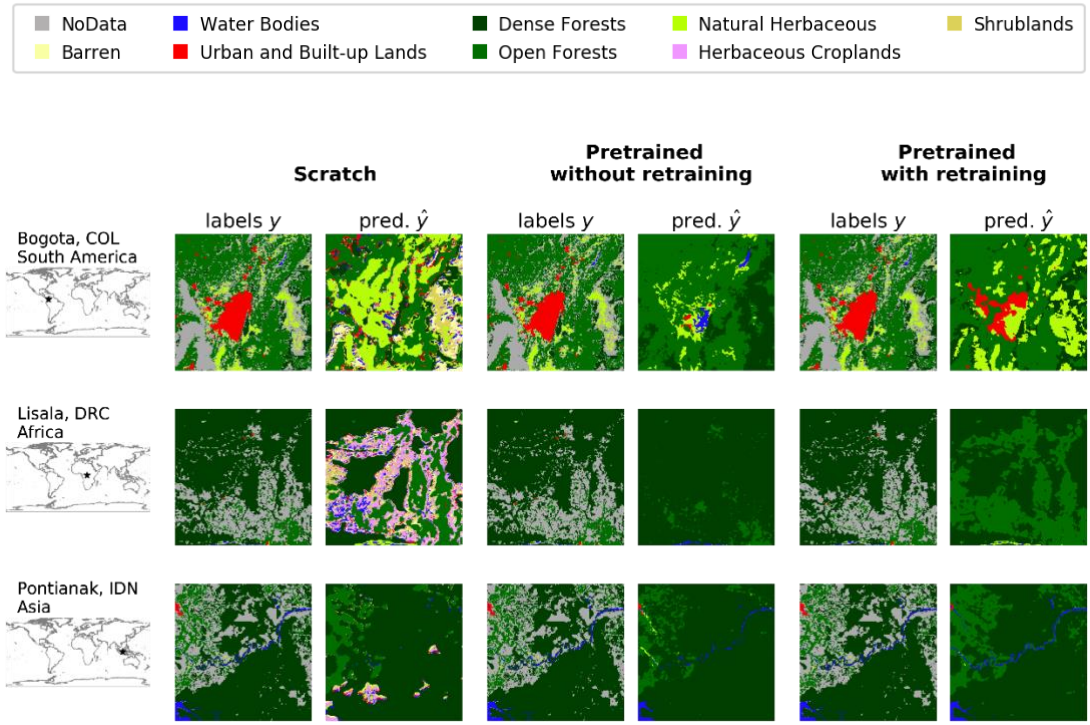


Figure 5.9 Reference labels y and predictions $\hat{y}$ according to three strategies to train the MTLCC network for year 2003 across three tiles distributed in different geographical areas in the pantropics. *NoData* represents pixels that were not analysed during either training or evaluation of the models.

According to the figure above, some differences are evident according to the target area. For instance, challenging areas can be observed in the tiles in South America and Africa where predictions of the MTLCC network trained from the scratch obtained poor accuracy as the spatial patterns of the predicted LC classes are remarkably different to the reference labels. Although the predictions of the retrained MTLCC network are satisfactory in the tile in South America, *Urban and Built-up* is incorrectly classified as *Natural Herbaceous.* The tile in Asia seems to be less challenging, however the MTLCC network trained from scratch fails to distinguish *Water Bodies,* which is captured by the pre-trained MTLCC network with or without retraining.

5.5.6 Discussion and conclusions

The objective of this chapter was to calibrate the MTLCC network to generate accurate LC predictions for the study of post-loss LC change trajectories. Four hyper-parameters of the network were explored using a Bayesian-based optimisation technique available through Google's AI platform. In addition, two shallow learners, Random Forest (RF) and Support Vector Machine (SVM) were also calibrated using a Bayesian-based approach but instead, by using a local server. The three best models of the MTLCC network, RF and SVM were compared according to standard ML and remote sensing foci metrics. These metrics were computed according to the test partitions with multispectral MODIS satellite image time series and reference data of three consecutive years, 2001, 2002 and 2003. Overall, the MTLCC network outperforms the shallow learners in all three years achieving an average F1-score of 59.98±0.58% quantity disagreement of 1.20±0.36% and allocation disagreement of 7.63±0.45%. In contrast, SVM, the best shallow learner model, obtained values of 57.75±0.32%, 6.79±0.53% and 9.42±0.09%, respectively. Furthermore, the analysis of F1-score per-class indicates that there is substantial gain in accuracy for six out of eight target-LC classes. Beside capturing noise patterns in the MODIS satellite observations as demonstrated in **Section 5.5.4**, the MTLCC network demonstrates its suitability for operation with large volumes of training data which is a primary constraint with the shallow learners.

The approach to find the optimal values of four hyper-parameters of the MTLCC network was implemented successfully using cloud-computing resources. Specifically, a two-stage approach with increasing amounts of training data (sample and full dataset) was conducted. During this procedure, it was found that the number of hidden states and optimiser type are the most important hyper-parameters influencing the performance of the network. In this regard, 128 hidden units and Nadam optimiser are the optimal settings according to 15 trials (10 and 5 at the first and second stage, respectively) run through Google's AI platform. The optimal number of hidden units is two times larger than the value derived from previous 'trial and error' experimentation (Coca-Castro *et al.*, 2019) as used in **Chapter 4** (see **Section 4.5.6**). Although values greater than 128 were also explored in two out of ten trials at the first stage, they were ranked as fourth and

fifth according to their overall accuracy. At the second stage, the Bayesian-based search algorithm focused exclusively on 128 hidden units searching the optimal values of the other hyper-parameters. In terms of the optimiser type, most of the trials indicated Nadam is the most suitable for the network and input dataset. This type of optimiser is an adapted version of Adam (Kingma and Ba, 2015) and uses Nesterov momentum to calculate adaptive learning rates. Although the Adam algorithm is normally predominant in remote sensing studies using deep learning techniques, some scholars have reported a superior performance by the Nadam optimiser. For instance, Sameen *et al*. (2018) reported a slightly better overall accuracy using Nadam (0.974) in comparison to Adam (0.970) in trained Convolutional Neural Networks (CNNs) for classifying LC using high resolution satellite images. Although this difference is not very large, by incorporating the Nesterov momentum into Adam, Nadam has a faster convergence, and therefore faster training than the Adam optimiser (Dozat, 2016). Regarding the batch size and learning rate, they did not show a clear pattern at the first stage. This can be primarily attributed to the different number of hidden units tested. However, some evidence was captured at the second stage with all trials set with 128 hidden states. Whilst three out of five trials at the second stage had learning rates above 5E-02 using a batch size of 40, the remaining ones with smaller batch sizes of 16 and 24 had learning rates below 9E-06. The learning rate can influence training time and finding the optimal point to set the network weights. If the steps are too small, training will take a long time. On the other hand, if the steps are too large, it will bounce around and could even miss the optimal point (Goodfellow *et al*., 2016).

In regards to the shallow learners, RF and SVM with the Radial Basis Function (RBF) kernel (SVM-RBF), the patterns observed for the hyper-parameters in all trials as shown in **Figure 5.6** are in agreement with remote sensing studies for LC classification. In these exercises, RF and SVM were trained with a standard size of 3000 samples per class per year which are in the range suggested by Pelletier *et al*. (2016). According to this size, RF takes considerably less time than SVM. The SVM classifier has generally been found to have a longer training time than RF (Pelletier *et al*., 2016). In RF, there are not significant increases in overall accuracy with a number of trees exceeding 500. This result is consistent with other large area LC classification studies shown using a number of trees between 100 (Inglada *et al*., 2017) and 500 (Zhang and Roy, 2017). The values explored for

the remaining hyper-parameter showed a minor influence on the classification accuracy. This outcome is in agreement with Pelletier *et al*. (2016) who concluded RF as an easily parameterised model, explaining its wide adoption for LC classification. For SVM-RBF, the hyper-parameters show C had optimal values below 1 and greater to 10000. For the hyper-parameter gamma γ, the higher overall accuracy values were located in values lower than 1. Many studies have compared SVM and RF showing comparable performances for LC mapping (Löw *et al*., 2013; Pelletier *et al*., 2016). In this study, SVM-RBF gave a better performance in some specific cases where RF generates poor accuracy. For instance, whilst the predictions by RF in the test tile of h10v08 as displayed in **Figure 5.7** were affected by noise in the input MODIS satellite image time series (so called attribute noise), SVM-RBF appeared to be capable of coping with it. This capability is consistent with the study by Pelletier *et al*. (2017) who indicated SVM-RBF is robust to the presence of noise with a large feature vector size. In this work, the large feature size is represented by a concatenate vector of surface reflectance values of 7 spectral bands for all MODIS images available per year plus their corresponding day of the year. Despite the robustness of the SVM, the algorithm confused *Herbaceous Cropland, Natural Herbaceous* and *Open Forests* with *Urban/Built-up Lands*.

Whilst this work shows the feasibility of transferring the knowledge of the fine-tuned MTLCC network trained in the Amazon to other geographical areas with a relatively small dataset, further investigation is needed. In this regard, it would be advantageous to conduct a systematic assessment following the guidelines of previous studies (Bailly *et al*., 2017; S. Wang *et al*., 2019). For instance, S. Wang *et al*. (2019) assessed the performance of training supervised models across areas with poor reference data including the effect of the geographic distance and the impact of different years. It is worth mentioning most transfer learning studies in DL for remote sensing classification are focused on pre-trained CNN and little work exists for RNN-based models (Ienco *et al*., 2019) as presented in this work using the MTLCC network.

6 Understanding the patterns of post-loss LC change trajectories over newly deforested areas across the Amazon region through data-mining

6.1 Introduction

The study of spatio-temporal dynamics of land cover (LC) and land use (LU) change such as post-loss LC/LU is important in understanding their resulting effects on ecosystems. New perspectives in investigating these dynamics have emerged through analysis of high-temporal LC/LU data. Traditional LC/LU modelling approaches depend upon comparisons of a limited number of time periods e.g. a pair of dates, or thematic classes e.g. forest/non-forest. The availability of high-temporal and thematic rich LC/LU data has however facilitated a more complex multi-temporal analysis, though these are coupled with new challenges. These challenges include the development and evaluation of novel methods able to process high volumes of data in order to extract insights describing and explaining the succession of LC/LU types over time. In this regard, inspired by methods originally developed to assess life course trajectories, Mas *et al*. (2019) proposed sequence analysis (SA) as a potential approach to describing and analysing chronological LC/LU change successions as observed in long-term post-loss LC change.

Based on the above premises, this chapter aims to investigate whether data from 2001-2019 high-temporal LC data generated by the fine-tuned model from **Chapter 5** and 2004-2018 Terra-i's deforestation detections allow mining of post-loss LC change trajectories. The chapter ultimately produces and explains factors behind typologies of post-loss LC. First, the suitability of sequence analysis is rigorously investigated through the study of post-loss LC change data over a large area, in this case, the Amazon region. Similar groups of post-loss LC sequences (i.e. similar changes of land cover) amongst newly deforested areas detected by Terra-i are determined, the sequence of their changes identified and discussed accordingly. The identification of common groups allows for discovering important aspects related to the study of post-loss LC change, such as dominant

pathways following deforestation, and the number of years which post-loss LC change data are necessary for their successful characterisation. Furthermore, as described in **Chapter 2**, several spatial determinants can influence the post-loss LC pathways. Observed groups of post-loss LC change trajectories were tested against some of these determinants to quantify their influence. The outcomes are discussed according to the narrative of forest change in the pantropics. Finally, an evaluation of the suitability and limitations of data mining techniques for determining and explaining trajectories of post-loss LC change are provided.

6.2 Background

The identification and characterisation of historical, current and future dynamics of LC/LU is a key subfield of interest amongst scholars in land system science (Verburg *et al*., 2015). Several approaches have been proposed to systematically study the dynamics of LC/LU change, primarily based on data derived from remote sensing (Mas *et al*., 2019). The most common approaches consist of computing changes rates, the elaboration of change and Markov matrices (Mirkatouli *et al*., 2015; Liping *et al*., 2018; Labán and Sturzenegger, 1994) and metrics such as persistence, gains, losses, swaps and change density (Pontius *et al*., 2004; Baral *et al*., 2018; De Alban *et al*., 2019). These approaches are chiefly limited to the study of relationships between a pair of dates which are then repeated for each period over the entire target time (Mas *et al*., 2019).

In contrast to the approaches above, few authors have analysed the chronological succession of LC/LU over time. Rather than a separate analysis of periods by a pair of dates, this approach aims to analyse the full sequence of LC/LU observations over the entire duration. Initial work on this direction was conducted by analysing a reduced number of combinations using a small number of LC/LU categories and time steps (B. Zhang *et al*., 2017; Baral *et al*., 2018). To a considerable degree, this has been enabled by; advances in computational power of computers; the reduction in the price of storage; the opening up of satellite remote sensing products; the release of higher-level (pre-processed) satellite products; and lately, distributed cloud-computing infrastructure e.g. Google Earth Engine. More recently, inspired by sequence analysis, Mas *et al*. (2019) has shown the potential of a set of procedures that allowed the definition of key LC/LU categories and sequences, comparison of sequences, and the search and linkage

of sequence patterns with drivers of LC/LU change. This procedure was tested using 1985-2017 regionally-tuned LC/LU annual maps extracted from Mapbiomas dataset collection 3 across a relatively small area of 340,500 km$_2$ in the Northeastern Region of Brazil. This chapter aims to extend Mas *et al.*'s (2019) investigation, but in a considerably larger area, the Amazon rainforest, with an approximate extent of 7.8 million km$_2$, targeting deforested areas detected by Terra-i from 2004 to 2018. The following sections introduce the fundamentals of the sequence analysis.

6.2.1 Sequence analysis

The sequence analysis (SA) comprises a set of techniques for describing and summarising sequential data (Ritschard and Studer, 2018a). SA was introduced firstly in computer science (Levenshtein, 1966), then in molecular biology for the study of DNA and RNA sequences (Levitt, 1969), and later in social sciences (Abbott, 1983). Amongst these areas, SA in the study of life course trajectories (Elder *et al.*, 2003), a subfield of social sciences focused on investigating the progression of individuals over their life course, has became a popular tool with powerful dedicated software (Brzinsky-Fay *et al.*, 2006; Gabadinho *et al.*, 2011; Halpin, 2017). The wide adoption of SA, in particular in social sciences, can be explained due to it being a convenient way of coding individual sequences into a form suitable for quantitative analysis (Ritschard and Studer, 2018a).

State and event sequences are two particular components subject to different interpretations and techniques in SA. In terms of LC/LU, a state, such as 'Forest' lasts the entire unit of time while an event, as a conversion in LC e.g. 'forest loss', occurs at a particular time point. Events can occur at the same time and are likely to generate a state change. On the other hand, states are mutually exclusive (Ritschard and Studer, 2018a). Another distinction between both components relates to their focus. State sequences aim to investigate duration and timing, whereas event sequences target the order in which events occur. In social sciences, state sequences are remarkably further studied than event sequences as a broader number of methods exist to process state sequence datasets (Abbott and Forrest, 1986; Ritschard and Studer, 2018b). Based on this premise, this work will limit its focus on the analysis of state sequences which, in the case of post-loss LC, can be analysed using high-temporal LC/LU data. Information

that can be extracted from state sequences include i) variety of states present in a set of sequences; ii) within-sequence state distribution; iii) timing (the time at which each state occurs); iv) duration (the cumulative and total time spent in the different successive states); and v) sequencing (the order of the different successive states) (Studer and Ritschard, 2016).

SA can be conducted in two phases: exploratory, and explanatory (Mas *et al*., 2019). The first phase aims to identify patterns in the sequences by finding the most typical sequence, which can be described as the succession states that best represent the entire group. Basic and advanced data-driven methods are available to determine the most typical sequence. Two primary methods exist, one which consists of determining the modal sequence, being the sequence obtained by computing the modal state at each position, and the other which determines the medoid, which identifies the sequence that minimises a given function of the distances to all other sequences. To tackle inconsistencies in the basic methods in which they do not necessarily return an observed sequence, advanced methods based on the pairwise distance, or dissimilarity have been developed (Studer and Ritschard, 2016). These methods provide a greater consistency in mapping the most common sequence representing the entire set of sequences. Furthermore, the use of pairwise dissimilarities between observed sequences can be employed to identify the principal types of patterns by clustering sequences using these measures. Different dissimilarity measures have been proposed and those relevant to the study of LC/LU trajectories as described in Mas *et al*. (2019) are illustrated in **Figure 6.1.**

Regarding the explanatory phase, this aims to assess the relationship between observed patterns in the sequences and covariates. In the study of post-loss LC, this can be translated to explaining the association of certain trajectories to environmental (e.g. terrain, climate, soil suitability) and socio-economic variables (e.g. accessibility, population).

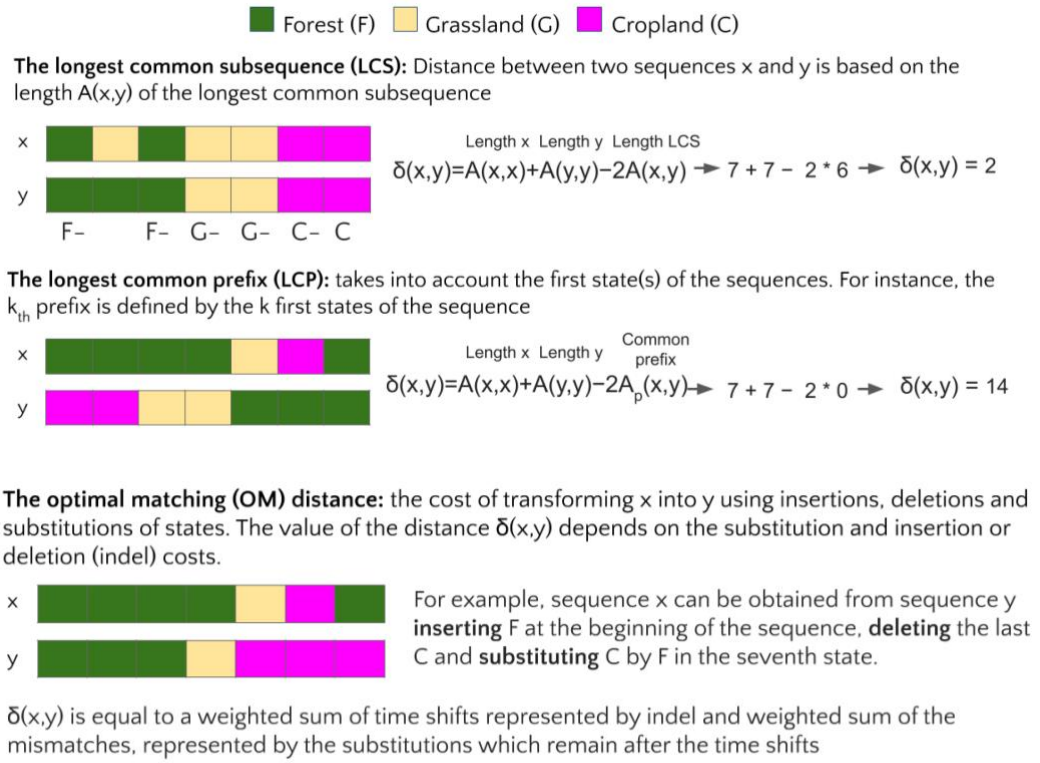


Figure 6.1 Representation of three distinct types of dissimilarity measures suitable to the study of chronological LC/LU observations as identified by Mas *et al*. (2019).

6.3 Study area and context

The study area corresponds to the Amazon region which represents the largest continuous region of tropical forest in the world. The delineation of the Amazon chosen is the Amazonian Network of Georeferenced Socio-Environmental Information (RAISG). The relevance of the area and rationale of the boundary chosen are described in **Chapter 4**, see **Section 4.4**. This delineation encompasses approximately 7.8 million km2 using a biogeographical criterion, with the exception of Ecuador and Brazil where legal-administrative criteria are used.

6.4 Methodology

Figure 6.2 describes the workflow aiming to determine the patterns of post-loss LC across the Amazon region. This workflow is principally based on the sequence analysis approach proposed by Mas *et al*. (2019). In this regard, the approach is conducted according to i) exploratory and ii) explanatory stages as described in **Section 6.2.1**. The main inputs for both stages are:

1. Annual LC maps (see **Section 6.4.1.1**).

2. A reference dataset indicating the location of deforested areas (see **Section 6.4.1.1**).

3. Auxiliary data (e.g. water bodies, topography, climate and accessibility) (see **Section 6.4.1.2**).

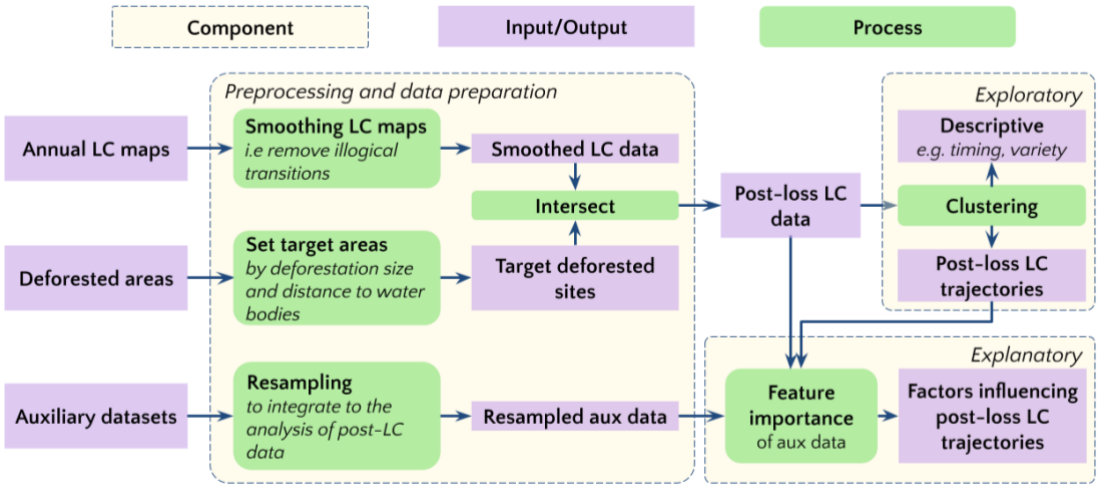


Figure 6.2 Diagram summarizing the main components for the study of post-loss LC change trajectories in the study area.

According to the scheme above, some pre-processing operations are performed (see **Section 6.4.1.1**) including: i) removing illogical transitions, i.e. smoothing from 2001-2019 annual LC maps generated by the calibrated MTLCC model in **Chapter 5**, and; ii) defining the target deforested sites according to their size and proximity to water bodies, to ensure the target pixels represent change due to anthropogenic activities rather than natural causes. Both pre-processed datasets are then merged to generate post-loss LC data which is then analysed through explanatory and exploratory stages (see **Section 6.4.2**). Both stages are conducted using a tool suited to capture LC change trajectories (see **Section 6.5**). Finally, the observed patterns are described and discussed according to narratives of post-loss LC extracted from the existing literature (see **Sections 6.6 and 6.7**).

6.4.1 Input data

6.4.1.1 Post-loss LC data

Post-loss LC data were generated following the protocol for cross-year evaluation across seven tiles of 384 pixels per 384 pixels as described earlier in **Chapter 4** (see **Section 4.5.5.2**). For the purpose of this work, the protocol was implemented over all tiles covering the study area (554 tiles of 384 pixels per 384 pixels in total). Briefly, as part of the protocol, two input datasets are required, annual LC data and a reference dataset of deforested areas. **Table 6.1** describes the main characteristic of the input datasets to generate post-loss LC across the entire Amazon region.

Table 6.1 Summary of the input datasets to generate post-loss LC data.

Product	Description	Temporal coverage	Valid range
LC annual maps	LC dataset generated according to a fine-tuned MTLCC network trained using 2001-2003 MODIS spectral data and LC reference data derived from 2001-2015 MODIS FAO-LCCS2 layer.	2001-2019	1-8
Terra-i	Terra-i is based on MODIS MOD13Q1 satellite data product with a 250-m spatial resolution and a 16-day temporal resolution. The version 2004_01_01_2019_06_10 was used.	January 2004 - June 2019	0 (*NoData*) - 2019

Prior to computing and assessing post-loss LC, the initial predictions of the fine-tuned model in **Chapter 5** and the reference dataset of deforested areas (Terra-i) were pre-processed. This preprocessing followed the same logic as the procedure applied in **Section 4.5.5.3**. A visualisation of the raw and post-processed 2001-2019 annual LC maps can be inspected in **Appendix V**. Examples of 1-km buffer water mask as well as original and pre-processed Terra-i deforested data are depicted in **Figure 6.3**.

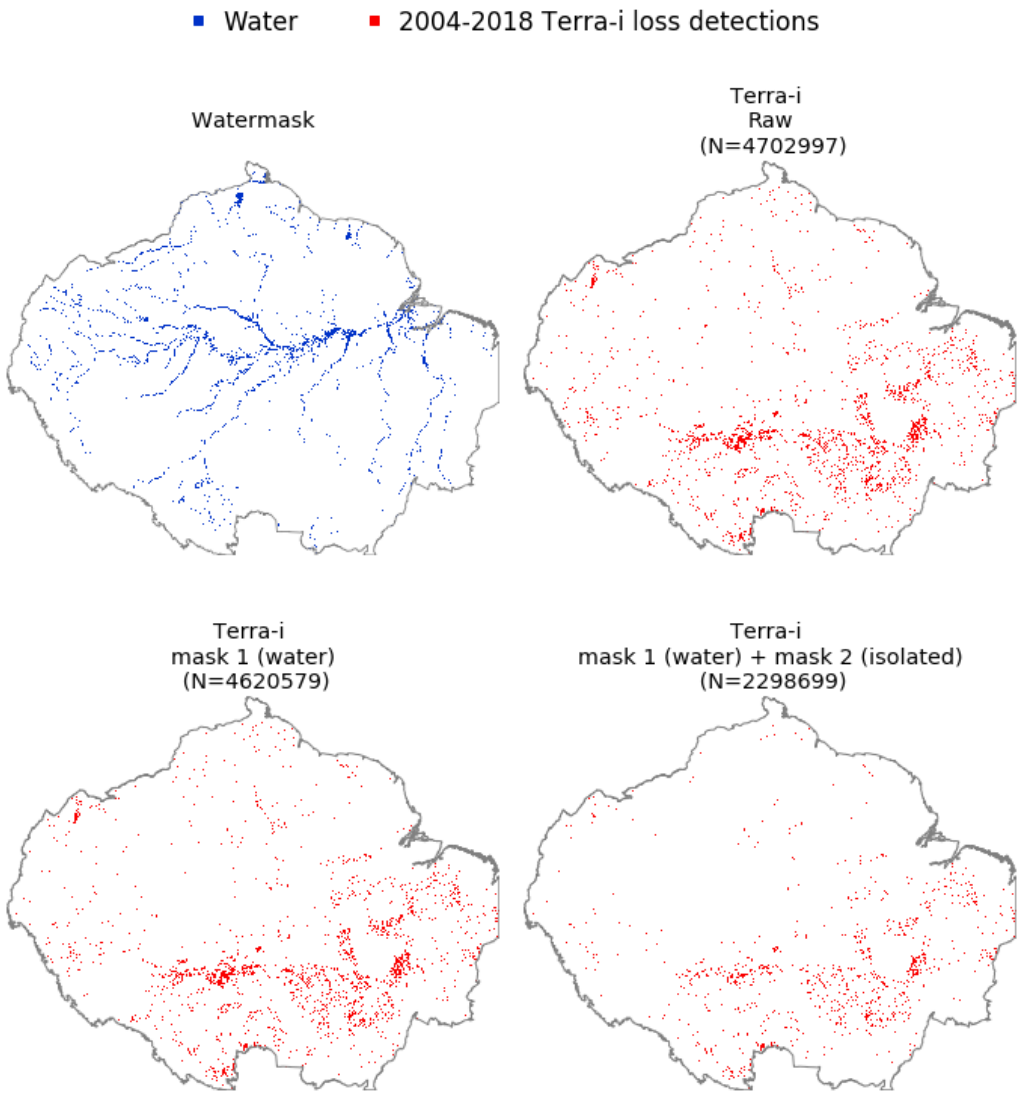


Figure 6.3 Example of the 1-km buffer water mask, raw and pre-processed 2004-2019 Terra-i data according to two masks (water and discarding isolated pixels) over a subset area in the study area. The total number of Terra-i pixels prior and after pre-processing is denoted in brackets as N.

6.4.1.2 Auxiliary datasets

A pool of potential explanatory variables was explored according to factors identified in previous studies as drivers of LC change and forest succession (Rutherford *et al*., 2007). These factors are predominantly spatial datasets of accessibility, topography, climate and protected areas which are assumed to be determinants of observed post-loss LC change trajectories. The sources and main features of the datasets for deriving explanatory variables are summarised in **Table 6.2**.

Table 6.2 Description of the representative external auxiliary datasets grouped by category used for extracting explanatory variables as part of the sequence analysis of post-loss LC data.

Category	Layer	Source	Units	Spatial resolution	Format	Period
Accessibility	Accessibility	The Malaria Atlas Project[1]	min	multiple	raster	2015
Biophysical	Elevation	CGIAR SRTM V4[2]	m	30 m	raster	2001
	Slope	CGIAR SRTM V4[2]	angle	30 m	raster	2001
	Temperature	Bioclim V2[3]	°C	1 km	raster	1970-2000
	Precipitation	Bioclim V2[3]	mm	1 km	raster	1970-2000
Policy	Protected areas (PAs)	UNEP-WCMC World Database on PAs[4]	N/A	N/A	vector	2019

[1] (Weiss *et al*., 2018) [2] (Jarvis *et al*., 2008) [3] (Fick and Hijmans, 2017) [4] (UNEP-WCMC, 2020)

The above layers were pre-processed to integrate them to the post-loss LC sequence analysis. The layer of protected areas (PAs), which is originally provided in vector format, contains their location and associated information including the International Union for Conservation of Nature (IUCN) category and year of implementation. The IUCN classification scheme classifies PAs in six categories related to their conservation goals, management objectives, and protection levels. The IUCN categories provide a gradient of naturalness varying from the most natural (category I) to the least natural condition (category V). Based on the IUCN category, PAs were aggregated into two major groups, conservation (I to III) and exploitation (IV to V) categories as proposed by Leberger *et al*. (2020). These categories were defined by the authors to test whether 'exploitation categories' had higher forest loss than 'conservation categories'. For the purposes of the study of post-loss LC, the test was reframed to capture the extent to which observed sequences are related to the proximity of these major categories. The proximity was defined by computing the euclidean distance using a target spatial resolution equal to Terra-i's deforestation dataset, 250-m. The proximity procedure was performed for the extent of South America. This reduces potential boundary effects taking into account the proximity of PAs outside the study area, in this case, the Amazon region. Only PAs reported as established before 2004

were considered, as the period of Terra-i's detection starts from 2004. This procedure is in agreement with Leberger *et al*. (2020) who discarded PAs before their period of target forest loss to avoid confusion between years of loss and eventual newly created areas. Additionally, before computing the proximity, some PAs in the study area were removed as the WDPA database has several PAs with missing information (e.g., unreported categories, unverified PAs). Regarding layers in raster format, these were resampled to the same spatial resolution as the proximity datasets of PAs (250-m) using the nearest neighbour algorithm.

Figure 6.4 displays maps of the spatial distribution of above covariates across the study area. Some patterns are evident from these maps. For example, the inaccessibility shows that less accessible areas are located on the western side of the Amazon region. Elevation and slope maps indicate the highest values are located across the Andean region. Although the average precipitation seems to be uniformly distributed in the area, most of the highest values are located on the northern side. The patterns observed in the proximity to PAs are related to their number, there being more and larger conservation PAs (N=110) than the exploitation PAs (N=51).

Additional to the above covariate layers, the proximity to each class from predicted LC maps were computed to describe and explain the observed patterns of post-loss LC. As detailed in the sections below (see section 6.4.2), *Natural Herbaceous* and *Herbaceous Croplands* classes were aggregated into a single class, *Farmland*, reducing the target LC categories from 8 to 7. For the purposes of this analysis, the proximity values were computed from the aggregated 2019 LC map as it represents the last date in the series of 2001-2019 LC maps. Before computing the proximity, only contiguous pixels >20 pixels were retained per class as areas smaller than this are mostly noise by the MTLCC model. **Figure 6.5** displays the spatial distribution of the proximity-related layers derived from the aggregated 2019 LC map. Distinct from PAs' proximity layers, the proximity calculation procedure for the LC maps was conducted according to the extent of the study area as predictions from the fine-tuned model were solely generated across this region.

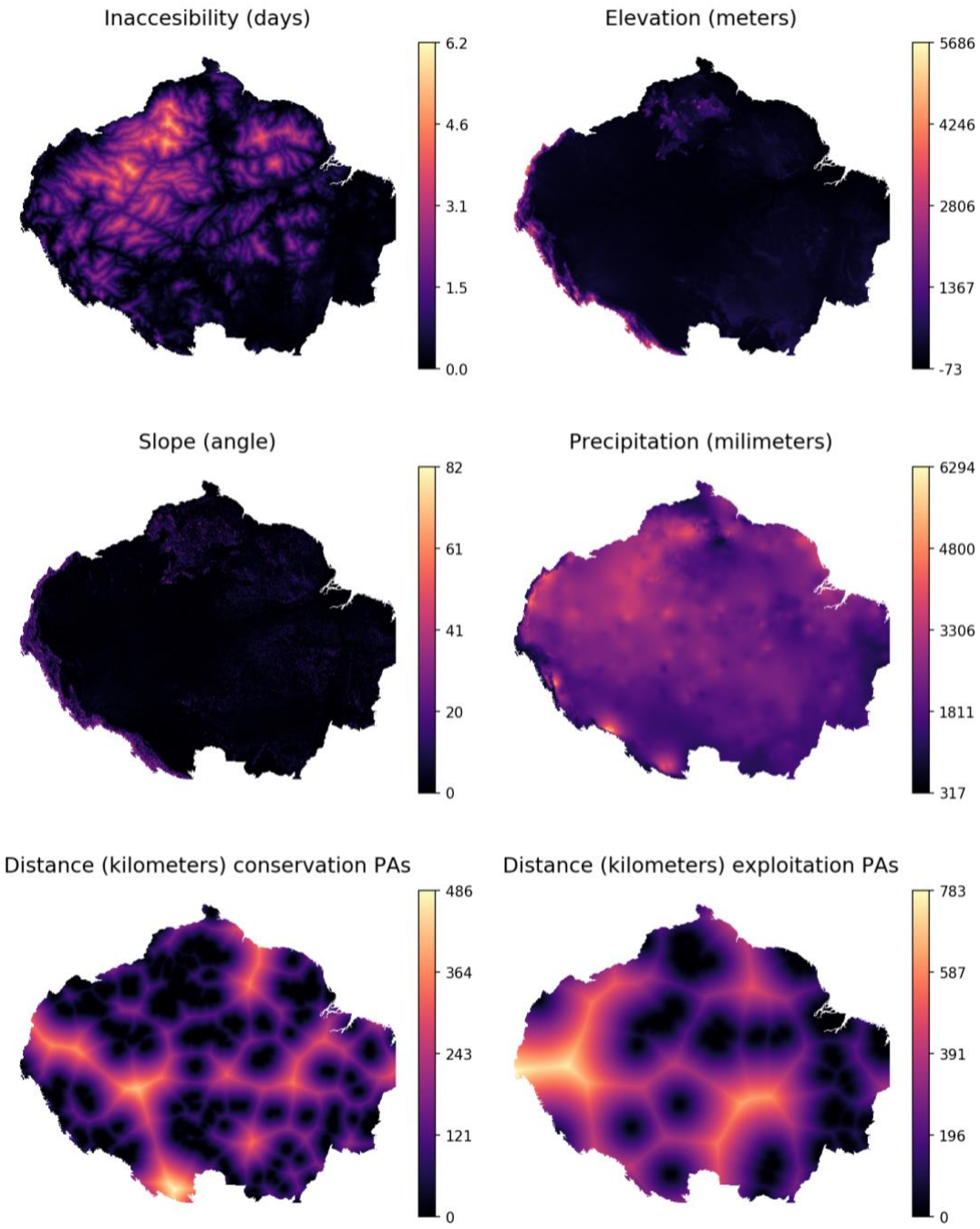


Figure 6.4 Spatial distribution of covariates used to study observed post-loss LC data.

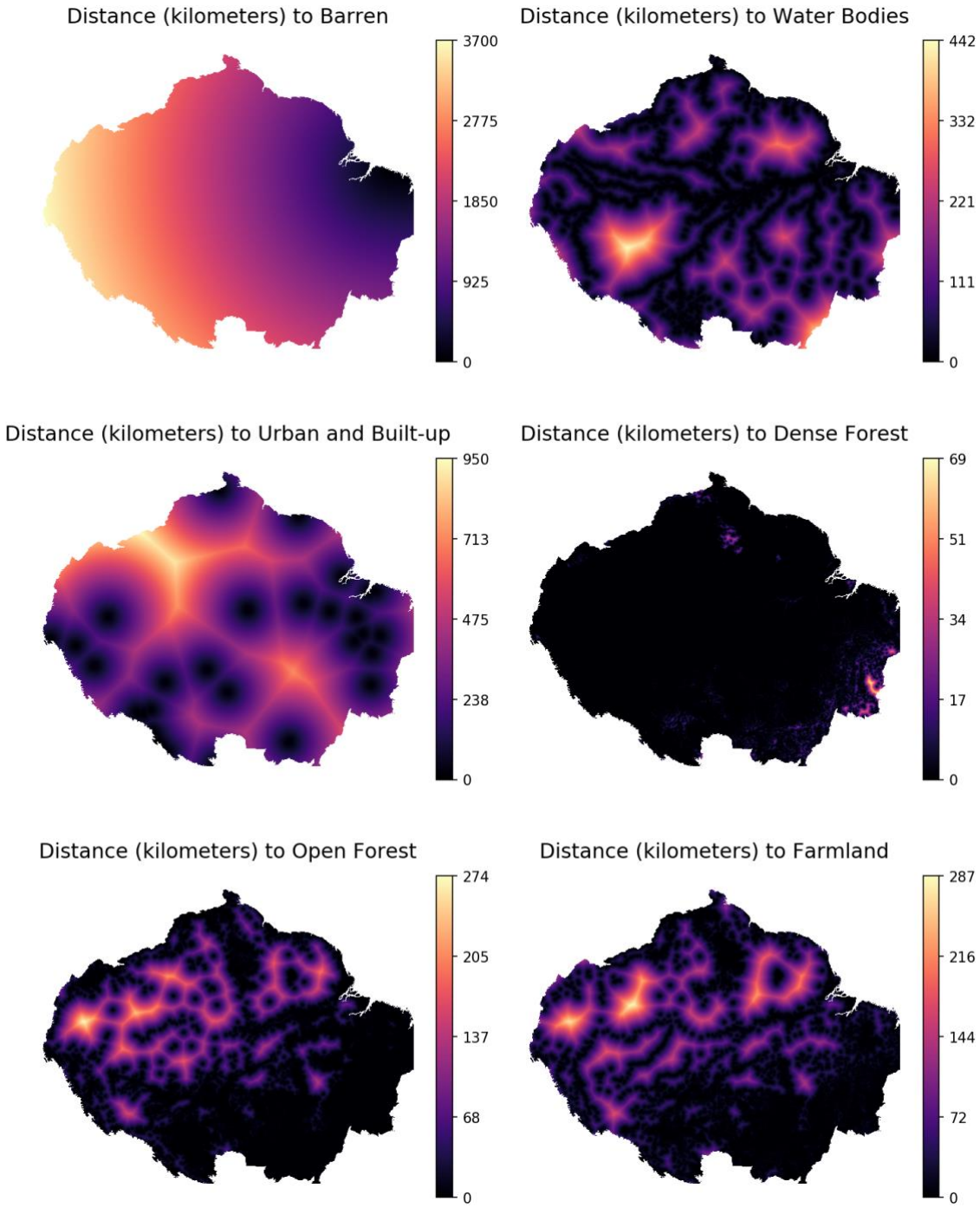


Figure 6.5 Spatial distribution of proximity-based layers computed from the 2019 LC predictions map.

6.4.2 Post-loss LC sequence analysis

A raster database of post-loss LC change was generated by extracting 2001-2019 smoothed LC maps according to pre-processed 2004-2018 Terra-i detections per year. The resulting post-loss LC change data were then partitioned by major habitat type. This partition was conducted under the assumption that

different post-loss LC trajectories and impacts might be observed according to habitat type. **Table 6.3** contains the total number of target deforested pixels for the period 2004-2018 grouped by major habitat type. According to the table below, a large proportion (86.5%) of deforested pixels occur from Tropical Humid Broadleaf Forests. Consequently, the implementation of exploratory and explanatory stages, as detailed in the sections below, focuses on this particular habitat in order to test the appropriateness of sequence analysis within the habitat most commonly associated with the Amazon region.

Table 6.3 Distribution of the target i.e. pre-processed Terra-i's deforested pixels for the period 2004-2018 according to the major habitat types across the study area.

Major habitat	N pixels	Proportion (%)
Flooded Grasslands & Savannas	2,698	0.12
Mangroves	333	0.01
Tropical Dry Broadleaf Forests	99,868	4.34
Tropical Grasslands, Savannas & Shrublands	206,926	9.00
Tropical Humid Broadleaf Forests	1,988,874	86.52
Total	2,298,699	

6.4.2.1 Exploratory

According to findings in **Chapter 3** and **Chapter 4**, an exploratory analysis was conducted by investigating LC sequences via a two-stage analysis. For the first stage, charts of the distribution of LC classes and sequences partitioned by the stability of the observed LC class(es) i.e. remained the same (stable) or different (non-stable), from 2001 to 2019 were created per deforestation year as conducted in **Chapter 4**. A preliminary analysis of these charts was important in order to perform additional pre-processing to the 2001-2019 smoothed LC data. This procedure aggregated *Natural Herbaceous* and *Herbaceous Croplands* into a single class, *Farmland*, reducing the target LC categories from 8 to 7 (see **Table 6.4**). This process facilitated the reduction of sequences which might be noise in the calibrated MTLCC model rather than actual LC transitions.

Table 6.4 Aggregated classes from 8-classes 2001-2009 LC maps derived from the fine-tuned MTLCC model in **Chapter 5** used in this study. Numbers in parentheses represent the original class codes. A broad definition per class is provided according to the original source, in this case, MODIS FAO-LCCS2 layer.

Aggregated class	8-classes LC map	Broad definition
Barren	Barren (1)	At least 60% of the area is non-vegetated barren (sand, rock, soil) or permanent snow/ice with less than 10% vegetation
Water Bodies	Water Bodies (2)	At least 60% of the area is covered by permanent water bodies
Urban and Built-up Lands	Urban/Built-up Lands (3)	At least 30% of the area is made up of impervious surfaces including building materials, asphalt, and vehicles
Dense Forests	Dense Forests (4)	Tree cover >60% (canopy >2-m)
Open Forests	Open Forests (5)	Tree cover 10-60% (canopy >2-m)
Farmland	Natural Herbaceous (6) Herbaceous Croplands (7)	Dominated by herbaceous annuals (<2-m) Grazing or cultivated fraction >60%
Shrublands	Shrublands (8)	Shrub cover >60% (1-2m)

Figure 6.6 compares charts according to the original and aggregated LC data extracted over deforested areas in 2007 located across the Tropical Humid Broadleaf Forests habitat. **Appendix VI** provides further visualisation of the resulting charts according to the aggregated class over the remaining deforested years in the target habitat. From the charts below it is evident most of the non-stable sequences start with *Dense Forests*. This pattern, which is systematically observed for the remaining deforested years (see **Appendix VI**), confirms Terra-i's detections which mostly capture conversion from this particular LC category. It is important to note the conversion at the cost of *Dense Forests* exists in years before/after Terra-i's deforestation year. This observation coincides with early/late detections by the Terra-i system as characterised in **Chapter 3** with the analysis of field records across the Peruvian Amazon.

Regarding the second stage, its aim was to identify typologies of post-loss LC change across the study area. These typologies can be translated as distinct pathways following forest loss. To ensure the observed typologies correspond to conversions related to forest loss, 2001-2019 LC data per deforestation year were pre-processed. In specific, conversions from other LC types such as

Farmland or *Open Forest*, which might be captured as change by Terra-i, were discarded. This particular rule was injected by retaining only those sequences with the presence of *Dense Forests* between two years before and two years after deforestation. It is important to note that the start year was redefined as two years before the deforestation date indicated by Terra-i. According to the last year of LC data, 2019, this scheme was only applicable to deforested areas in 2004 and 2016.

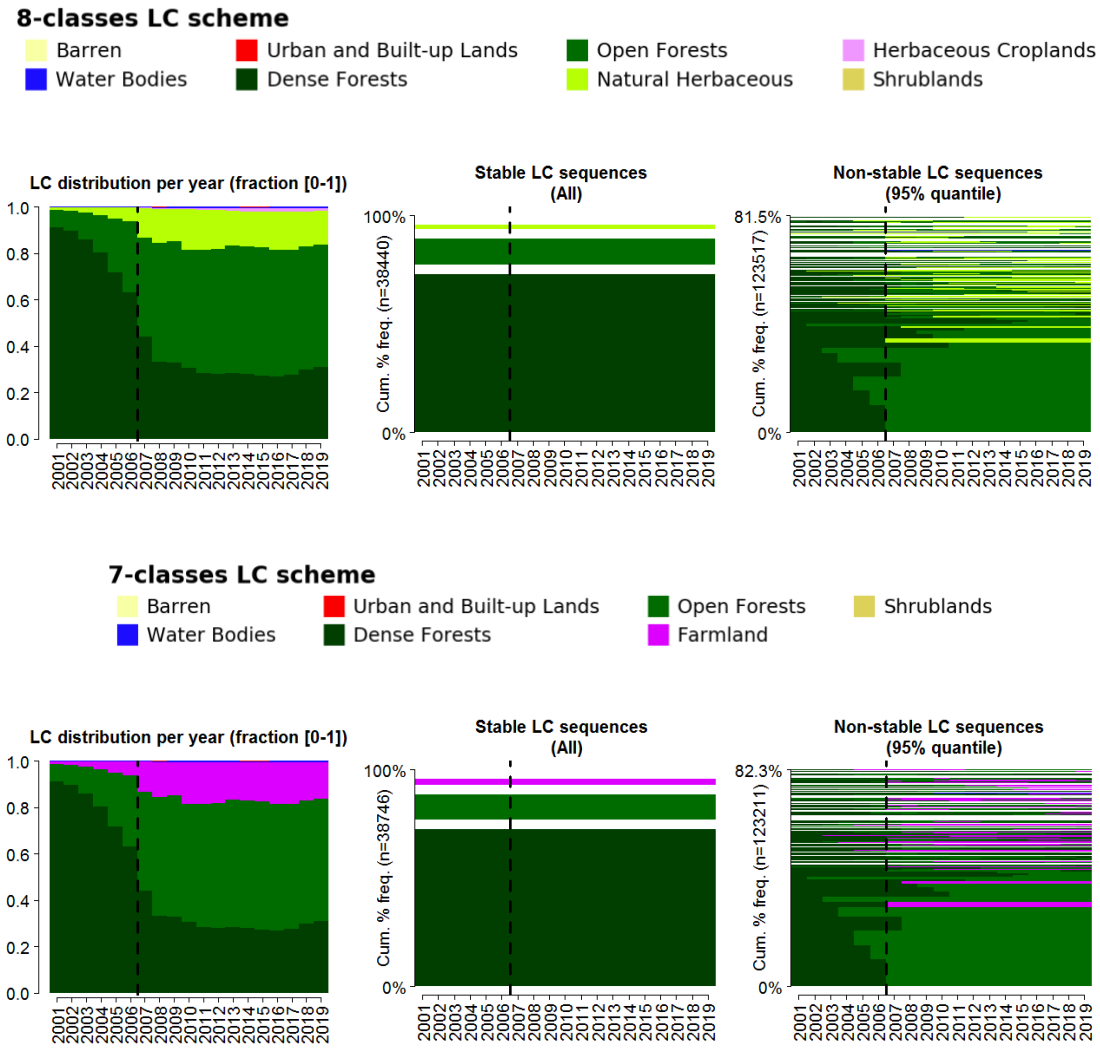


Figure 6.6 Comparison of the LC distribution, stable and unstable sequences over the period 2001-2019 for deforested areas in 2007 (highlighted by a dashed vertical line) located across the Tropical Humid Broadleaf Forests habitat in the Amazon region.

After the above pre-processing, Ward's agglomerative hierarchical clustering (Ward, 1963) was conducted to characterise typologies of post-loss LC. Ward's method iteratively merges all possible pairs of observed sequences based on the sum of square distances to identify the group configuration that minimizes in-

group variance (Johnson, 1967). This method was chosen, as is common in sequence analysis, with easily visualisable results in the form of a dendrogram. For sequence analysis, a distance-based matrix is usually used as input to return agglomerative groups i.e. clusters. To generate this matrix, different pairwise dissimilarity metrics exist being those relevant for characterising chronological LC/LU changes summarised in Mas *et al*. (2019). For this work, the optimal matching (OM) distance based on transition rates (TR) was selected as being optimal for distinguishing typologies of post-loss LC. Briefly, the OM distance between two sequences x and y is defined according to the costs of two possible operations, substitution and insertion or deletion (indel), required to match the states of x and y. Based on the example of OM in **Figure 6.1**, sequence x can be obtained from sequence y inserting *F* at the beginning of the sequence, deleting the last *C* and substituting *C* by *F* in the seventh state. OM is then the sum of two terms: a weighted sum of time shifts represented by indels and a weighted sum of the mismatches, represented by the substitutions which remain after the time shifts (Mas *et al*., 2019). Various methods exist to set the values of these costs (Studer and Ritschard, 2016). Following the OM-TR implemented by Mas *et al*. (2019), the transition rates between LC types were used to set indel and substitution costs. A OM-TR based distance matrix was then derived from the observed LC sequences in deforested areas between 2004 to 2016.

Figure 6.7 shows the resulting dendrogram derived from the Ward's agglomerative hierarchical clustering using OM-TR distance matrix extracted from LC sequences of 2004-2016 Terra-i's pre-processed deforested areas. It appears that the different sequences are primarily grouped according to length rather than order. The identification of typologies of post-loss LC is therefore not optimal using sequences of various lengths. A solution consists of retaining only the sequences observed for at least t_{min} years and analysing their trajectory over only t_{min} years. Although a high t_{min} can provide enough LC observations to characterise typologies of post-loss LC, this is not optimal as the analysis will be reduced to few deforested years i.e. 2004 and 2005 having 17 and 16 years of LC data, respectively. In contrast, if a small t_{min} is set, more deforested years can be included, though the sequences may be too short to highlight the variety of trajectories. Therefore, the optimal typologies of post-loss LC were determined by a joint assessment of t_{min} values from 6 to 16, in combination with increasing

levels in a number of clusters k from 2 to 10. It is important to note the values tested were defined empirically as no previous work based on sequence analysis exists for determining typologies of post-loss LC change.

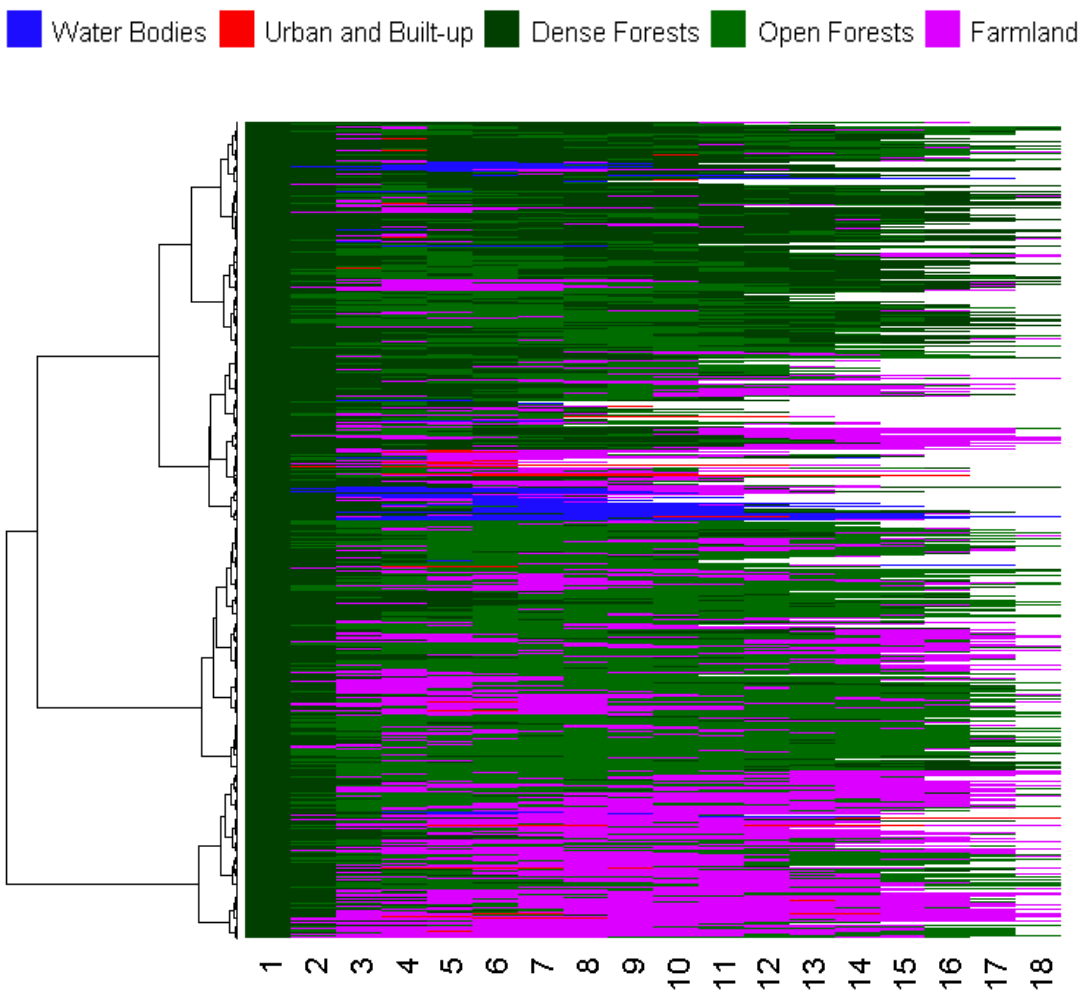


Figure 6.7 Dendrogram (left) created using a OM-TR based distance matrix derived from observed post-loss LC sequences (right) for deforested areas between 2004 and 2016. The length of each sequence, N, is related to the number of LC data available according to the deforestation year analysed. The sequences are ordered according to their similarity and related sub-groups (dendrogram's branches). For instance, a defined sub-group of similar LC sequences with continuous presence of *Water Bodies* can be seen in the central part of the plot.

The best combination of minimum period t_{min} and the number of clusters k was then determined by well-known measures related to cluster quality and stability as introduced by Han *et al.* (2017). Three statistics were used to empirically determine cluster quality and to describe various attributes of the clustering: Average Silhouette Width (ASW) (Rousseeuw, 1987), Hubert's C index (Hubert and Levin, 1976) and the Point Bi-serial Correlation (PBC) (Milligan and Cooper,

1985). ASW, with values ranging from -1 to 1, measures the coherence of assignments to clusters: high coherence (close to 1) implies that clusters are homogeneous and well separated from each other. The HC index shows the gap between the partition obtained and the best partition theoretically possible with this number of groups. HC values vary from 0 to 1 with low values (close to 0) indicating good clustering. Finally, PBC with values between -1 to 1 measures the capacity of the cluster solution to reproduce the original distance matrix. A high PBC value is preferred (Han *et al*., 2017). The various descriptive statistics of these classifications are provided as they possess distinct optimal solutions and value differing characteristics of a classification. Thus, to fully explore the cluster results, it is beneficial to retain all of these statistics.

Determining cluster stability is important for determining how robust the clusters observed are with changes in the input dataset (Hennig, 2007). For each combination of minimum period and number of clusters, the stability of the clusters was determined according to data sampling fluctuations by using bootstrap methods (Han *et al*., 2017). The clusterwise Jaccard Bootstrap Mean (JBM) (Hennig, 2007) was used as a measure to conduct the bootstrap by re-sampling the data and computing the Jaccard similarities of the original clusters to the most similar clusters in the re-sampled data. When JBM is below 0.6, the cluster solution should not be relied upon. JBM values greater than 0.85 indicate highly stable clusters. A JBM between 0.6 and 0.85 suggests some structure, but exact cluster membership is uncertain (Han *et al*., 2017).

From the optimal combination of minimum period and number of clusters, a meaningful interpretation of post-loss LC typologies was conducted. This step was facilitated through visualisation tools such as sequence index, sequence modal state and tree-structure plots (Gabadinho *et al*., 2011; Studer, 2013). In sequence index plots each sequence is represented by a line composed of different segments representing states, in this case LC types, and the length of the segments being proportional to time spent in a state. Therefore, large amounts of information are summarised in sequence index plots such as order, prevalence and timing of states and overall variability within and between sequences. In the sequence modal plots, the sequence state i.e. LC type with the highest frequency at each point, is observed. Finally, tree-structure plots are

useful to graphically represent the agglomeration procedure, in this case the Ward's hierarchical clustering to identify clusters i.e. trajectories from post-loss LC sequences.

6.4.2.2 Explanatory

Several statistical modelling methods are currently available to analyse, describe and understand the typologies characterised from SA, multinomial logistic regression (MLR) being one of the most common (Darak *et al*., 2015). MLR applies a non-linear log transformation that allows the calculation of probability of occurrence of two or more unordered classes of a dependent variable (in this case, post-loss LC change trajectory) based on explanatory variables (Hosmer Jr. *et al*., 2013). It makes no statistical assumptions concerning normality, linearity and homogeneity of variance for independent variables (Rutherford *et al*., 2007). In addition to the MLR, scholars have explored Random Forest (RF) (Breiman, 2001) as an alternative, to model a given phenomenon according to related factors. Briefly, as it was introduced in **Section 5.4.2.2**, RF forms a combination of multiple decision trees, where each tree contributes a single vote to the final output, which is the most frequent class e.g. post-loss LC change trajectory.

Based on the above premises, the auxiliary layers described in **Section 6.4.1.2** were investigated in order to understand their relationships with observed post-loss LC trajectories. To identify the variables with the strongest influence, the Recursive Feature Elimination (RFE) approach (Guyon and Elisseeff, 2003) was implemented. RFE is a backward selection algorithm that iteratively removes uninformative features from a given model based on variable importance scores (Kuhn and Johnson, 2013). The RFE was run separately with the MLR and RF models as base-learners, using a nested cross-validation approach as proposed by Anderegg *et al*. (2019). The dataset containing post-loss LC trajectories and related auxiliary layers was resampled 30 times with an 80:20 ratio. This ratio means, for each resample, feature elimination was conducted on 80% of the data, and model performance was assessed on the remaining 20% in 11 decreasing steps with values corresponding to the total number of auxiliary layers. Due to the sensitivity of the MLR and RF models to class imbalance, the dataset was iteratively downsampled in each resample to the maximum number of observations of the minor class (in this case, the water persistent trajectory). In

each iteration, the base-learner hyper-parameters were tuned using a 10-fold cross-validation. The models were run in parallel using default tuning for selecting the best model, considering *mtry* and *decay* for RF and MLR, respectively as the main hyper-parameters. For interpreting RFE, removed features were assigned a rank corresponding to an iteration after which they were eliminated i.e. those removed first were ranked 11, whereas feature retained as the last was ranked 1.

6.5 Software and implementation

The hardware and software used according to the main steps in this study are presented in **Table 6.5**. For the generation of LC predictions using the fine-tuned MTLCC model, and data preparation scripts, python 3.6 was used. The parallel implementation in python by Abercrombie and Friedl (2016) was used to generate efficiently smoothed 2001-2019 LC maps. The sequence analysis of post-loss LC change data was implemented in the R software v.3.6.2 by using the package *traMinerR* v2.0-11.1 (Gabadinho *et al*., 2011) and *WeightedCluster* v1.4 (Studer, 2013). The *traMineR* provides a wide variety of functions to visualise and analyse sequential data as post-loss LC observations. Due to the fact that 2_{31}–1 elements is the maximum vector size that R software can process, the *traMinerR* package can only compute pairwise dissimilarities up to around 35,000 sequences, which could result in the sequence limit being reached before all possible sequences were determined (Durrant *et al*., 2019). The *WeightedCluster* package was therefore used, being created to overcome this problem by solely analysing unique sequences and then applying the appropriate weights to subsequent analysis. For instance, in the post-loss LC change data of deforested areas in 2004, the number of observed sequences was greater than 100,000 but only some 9,000 sequences were unique. For the clustering procedure, the *hclust* function of the stats package v.3.6.2 (Core Team and Others, 2013) was used taking into account the weighted information.

Regarding the explanatory phase, the *nnet* v7.13-12 (Venables and Ripley, 2002) and *ranger* v.0.12.1 (Wright and Ziegler, 2017) packages were used to implement the MLR and RF models, respectively. The RFE approach was conducted in the *caret* package v6.0-85 (Kuhn, 2008) using parallel processing functions provided by the snowfall package v1.84-6.1 (Knaus, 2010).

Table 6.5 Lists of hardware and software used in this study organised by step.

Step	Hardware and/or Software	Specifications
Data downloading	Google Earth Engine (GEE) Google Cloud Platform (GCP)	GEE was used to access and download 2001 to 2019 MODIS satellite images and auxiliary datasets (see **Section 6.4.1**) exported as GZ TFrecords and GeoTIFF, respectively to GCP.
Data pre-processing and generation predictions	Server Tensorflow	The same Windows OS server used in previous chapters with a memory of 128 GB, Intel Xeon E5-2667 v4 @ 3.20 GHz ×16 cores and (1) GPU (NVIDIA Tesla M60)
Post-loss LC visualisation and analysis	R version v3.6.2 Exploratory stage (traMinerR WeightedCluster, hclust) Explanatory stage (nnet, ranger, caret)	The pre-processing and analysis of the post-loss LC data was conducted in R with specific packages for each stage, exploratory and explanatory.

6.6 Results

This section presents the main outcomes of the sequence analysis (SA) applied, to understand the chronological LC change over deforested areas (so-called post-loss LC change trajectories). The pre-processing operations were fundamental in preparing the input data for SA. For instance, smoothing 2001-2019 LC predictions generated by the fine-tuned model in **Chapter 5** allowed reducing illogical transitions. Additionally, the filtering operations in 2004-2018 Terra-i detections i.e. removing pixels located within a 1-km buffer created from water bodies and single deforested pixels ensured the target pixels represent change due to anthropogenic activities rather than natural causes. Amongst the LC types from the smoothed LC maps, *Dense Forests* was used as a proxy to limit focus to those sequences showing successions related to forest conversion. Moreover, the aggregation of *Natural Herbaceous* and *Herbaceous Croplands*, both defined originally as dominated by herbaceous annuals (<2m) (see **Table 6.4**), into a single class, *Farmland*, facilitated a reduction in sequences which might be noise in the calibrated MTLCC model rather than real LC transitions.

From the above assumptions, the following sections present the findings of SA applied to the study of post-loss LC change trajectories according to exploratory

and explanatory stages. The results are focused on Terra-i's deforested areas located within the Tropical Humid Broadleaf Forests in the Amazon region as indicated in **Section 6.4.2**.

6.6.1 Exploratory

The exploratory stage provided valuable information of the chronological LC change over Terra-i's deforested areas and typologies of post-loss LC change trajectories. In the first period of analysis, in which 2001-2019 LC data were projected per deforested year, stable and non-stable LC sequences were identified according to the presence/absence of change amongst LC types. **Figure 6.8** presents the partition of both categories per deforested year. It is important to note that the proportion of non-stable LC was above 70% between 2004 and 2009. From 2010, this proportion was reduced below 70%. The implication of this reduction is that fewer sequences were analysed in the SA approach because only unstable sequences show changes in contrast to a single LC type over time. Although some of the stable sequences might be attributed to noise by the Terra-i system (false detections), some other changes between LC types might be hindered by the smoothing algorithm or misclassifications from the fine-tuned MTLCC network itself.

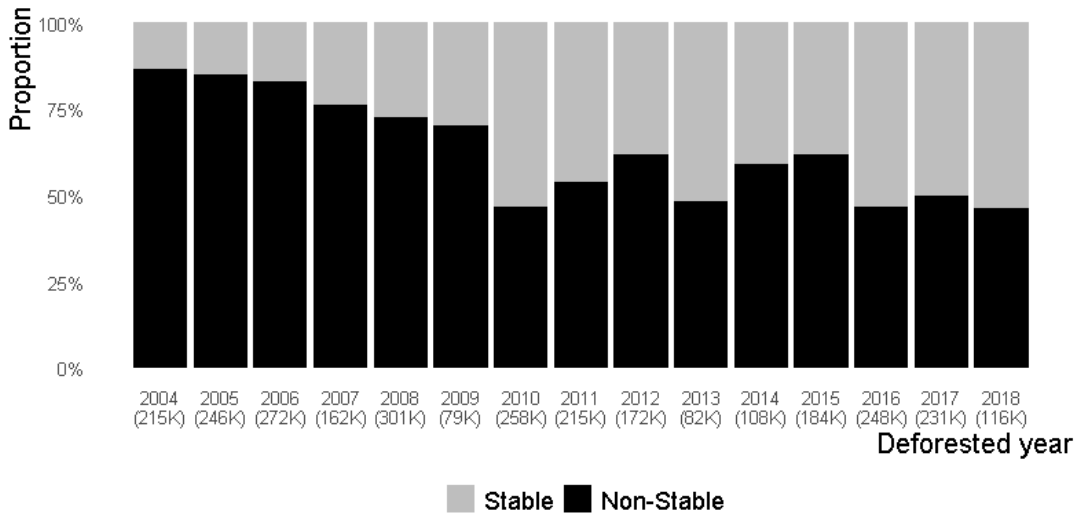


Figure 6.8 Proportion of 2001-2019 stable and non-stable LC sequences by deforested year. The number of total sequences per year is expressed in thousand (K). We can see the stable are in average above 50% after 2010. This lastly affects the study of post-loss LC as only pixels defined as non-stable favoured observing at least two LC types over time.

Rather than analysing observed LC sequences separately per deforested year, the strategy implemented grouped all possible LC sequences for most of the deforested years from the pre-processed 2004-2018 Terra-i dataset. This strategy joined all observed LC sequences by their relative start year of deforestation which was redefined as two years before the date indicated by Terra-i. **Figure 6.9** presents the LC partition over time, in this case, year following deforestation.

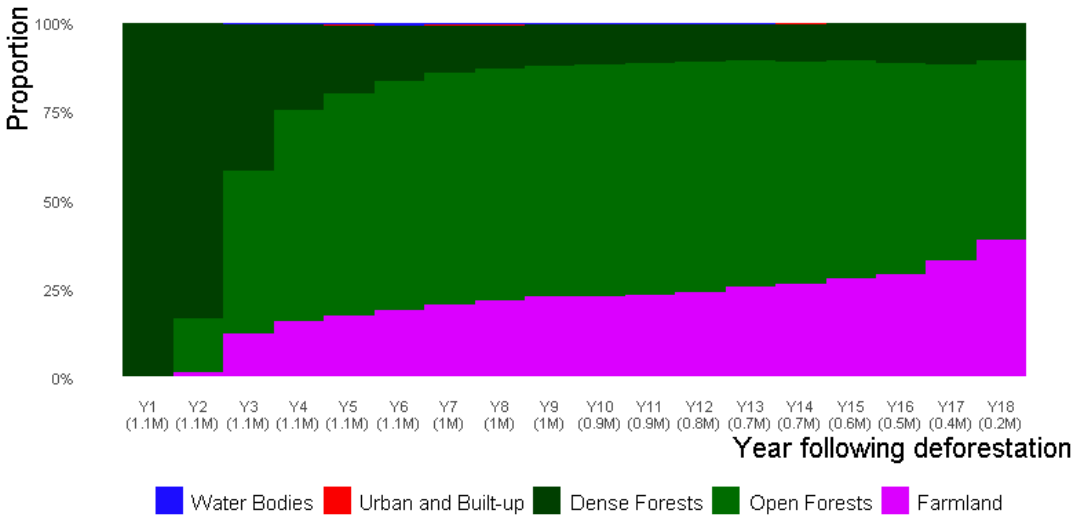


Figure 6.9 LC distribution per year following deforestation according to pre-processed 2004-2016 Terra-i data. LC sequences were merged by their relative start year of deforestation. The number of total sequences per year, expressed in millions (M). The diagram shows that most of the conversion of *Dense Forests* appears to *Open Forests* followed by Farmland. The first and second year have proportion of forest above 75%.

Whilst the previous chart is informative about the distributions of LC types following deforestation, it remains unclear which different pathways LC change could take. In this regard, the assessment of the minimum period and number of clusters provided a data-driven guidance of the optimal number of post-loss LC typologies. **Figure 6.10** presents the results of three quality metrics, Average Silhouette Width (ASW), Hubert's C index (HC) and the Point Bi-serial Correlation (PBC) plus a stability metric, the clusterwise Jaccard Bootstrap Mean (JBM). Although metrics show different optimal solutions, some patterns are evident. For instance, ASW and HC values are contrasting, meaning high values of ASW are related to low values of HC. PBC shows higher values with the increase of minimum period and a decrease in the number of clusters. The stability, measured by JBM, is proportionally decreased by increases in the number of clusters.

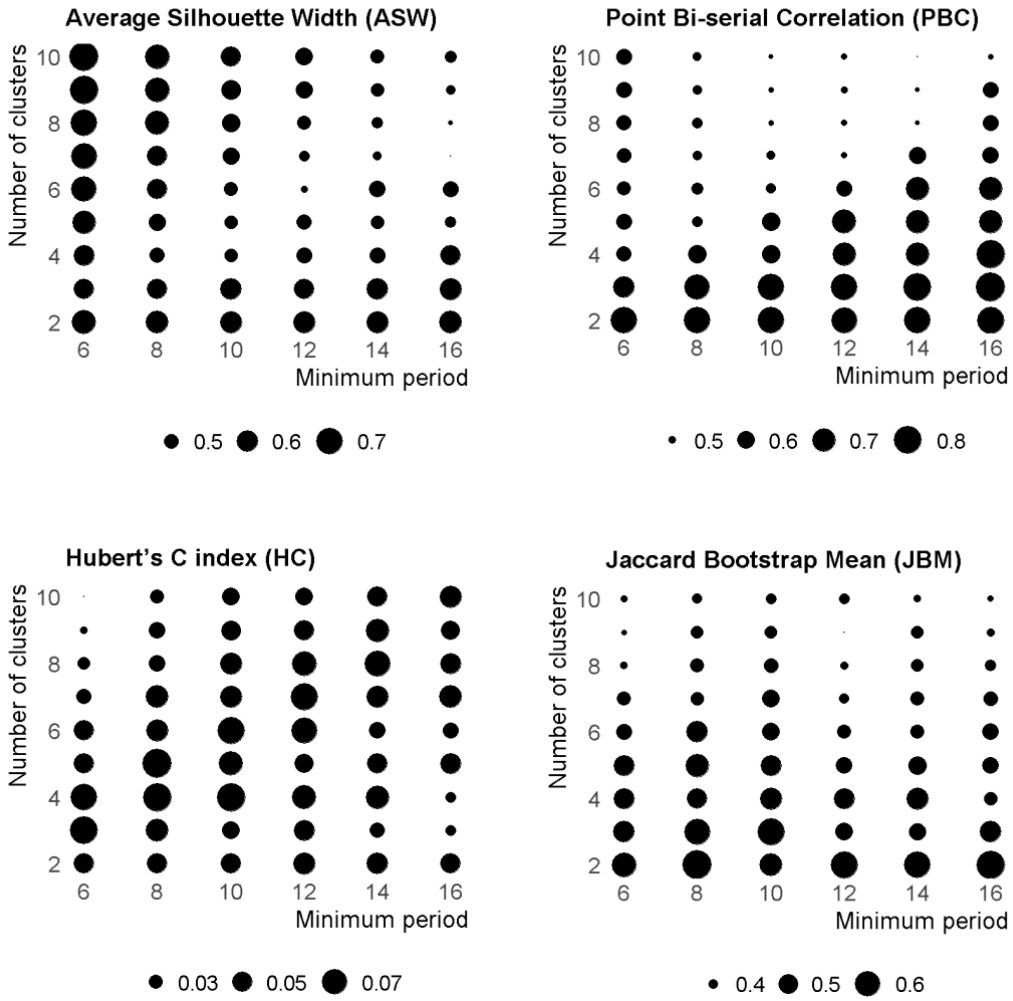


Figure 6.10 Values of four cluster-related performance metrics used to guide the optimal configuration of number of clusters and minimum period to determine homogeneous groups (trajectories) of post-loss LC change. Larger points should be interpreted as optimal by metric, except for the HC metric where small values are preferred. The diagram shows ASW and HC values are contrasting, meaning high values of ASW are related to low values of HC. PBC shows higher values with the increase of minimum period and a decrease in the number of clusters. The stability, measured by JBM, is proportionally decreased by increases in the number of clusters. The optimal configuration was defined as a minimum period and number of clusters equal to 12 and 6. Apart from being visually selected according to the performance metrics, the optimal configuration was aided by sequence index plots as it is described below.

A visual inspection of the most promising configurations according to the quality and stability metrics was aided by sequence index plots. The configuration of a minimum period and number of clusters equal to 12 and 6, respectively was considered as the optimal solution (see the **Appendix VII** for details). This solution yields a total of 802,975 post-loss LC sequences grouped in six clusters as depicted in **Figure 6.11**.

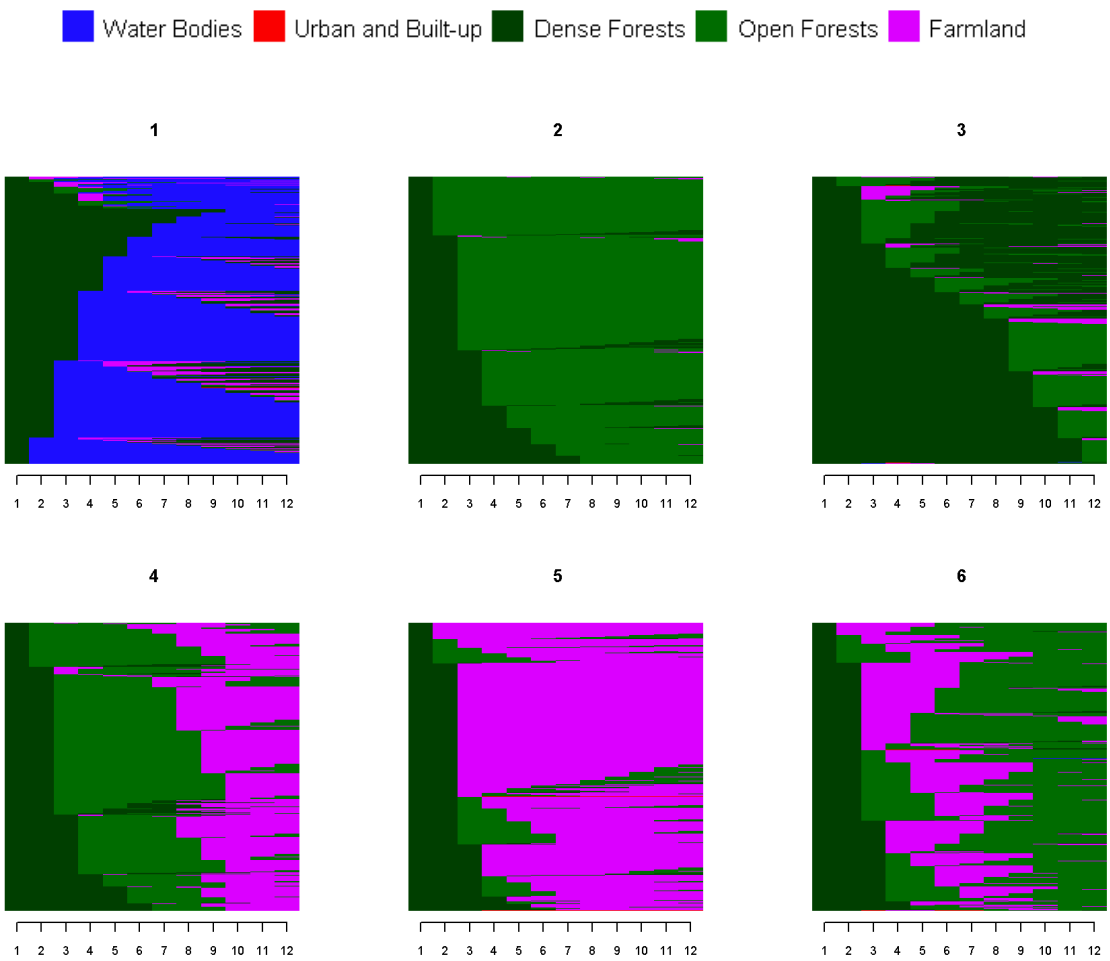


Figure 6.11 Sequence index plots showing post-loss LC sequences grouped in six representative clusters derived from a systematic assessment of different minimum periods and number of clusters. Further interpretation of the meaning of clusters is depicted by using the modal plots as illustrated in **Figure 6.13**.

Complementary to the charts above, **Figure 6.12** presents a tree-structure diagram showing clusters according to the most promising configuration. Starting from the root cluster, which gathers the whole set of post-loss LC sequences N=802,975, clusters are split according to the number of clusters set in the optimal configuration. For instance, cluster 1 identified at the 2nd split (that gathers 631,363 sequences with dominant transitions from *Dense Forests* to *Open Forests*) is divided at the 3rd split into two clusters that contain 88,097 and 543,266 sequences. As visually noticeable, the algorithm distinguished sequences with a tendency towards *Dense Forests* to *Water Bodies* (cluster 1) at the 4th split. Both types of representations, sequence index and tree-structure plots, were fundamental to guide the selection of the optimal combination of minimum period and number of clusters. Other solutions close to the optimal are detailed in **Appendix VII**.

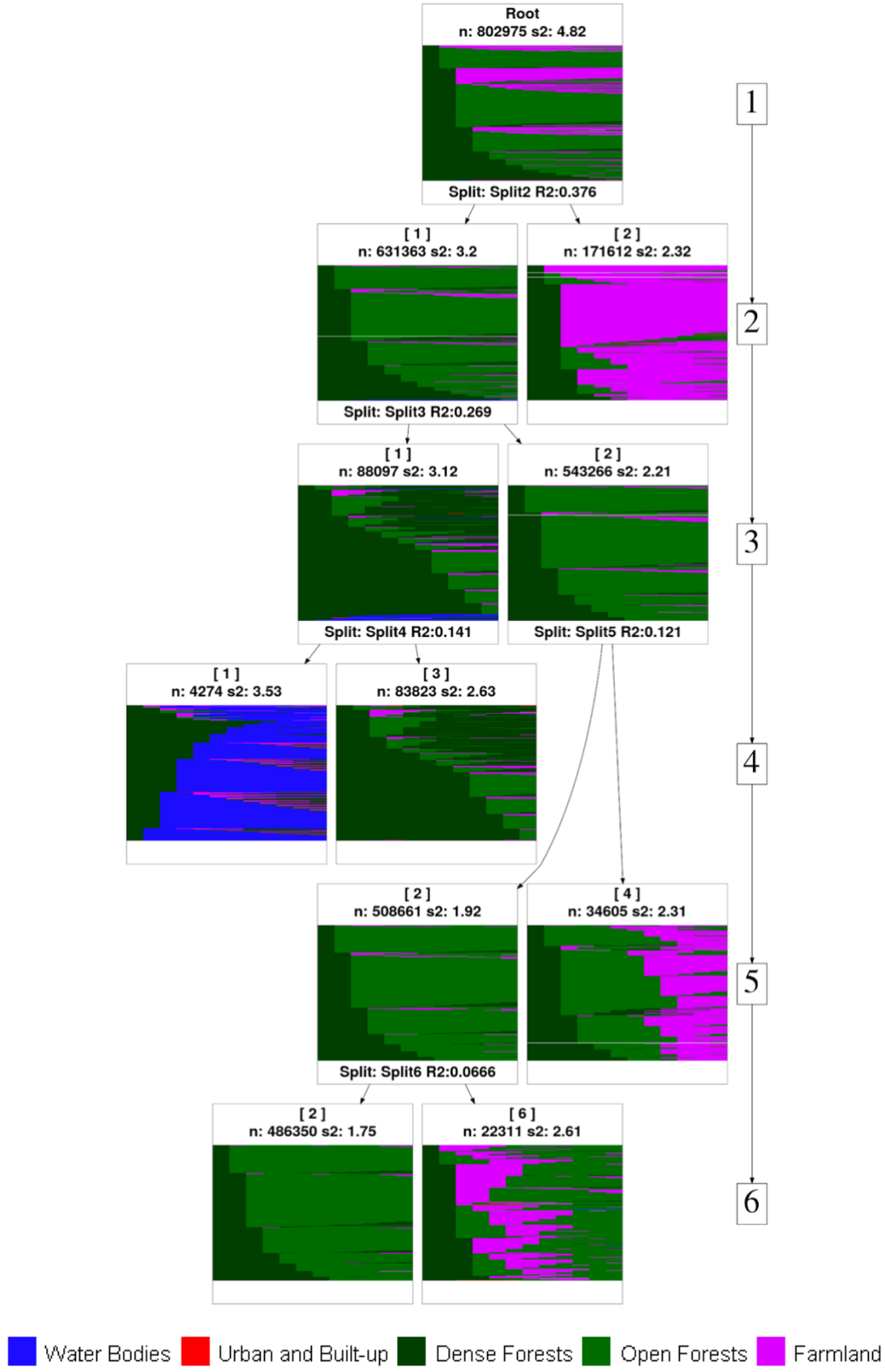


Figure 6.12 Tree-structure diagram showing how cluster partitions for the selected solution (minimum period of 12 and 6 clusters) were generated. The number of sequences per partition is denoted by n. From the first level, denoted as 1, which contains the whole population of post-loss LC sequences (n=802975), two groups are derived. One of this groups contains a variety of multiple post-loss LC groups which are more clearly defined until level 6. This level contains a cluster with the largest number of pixels (n=486350).

The modal state sequence plots as represented in **Figure 6.13** supported naming and describing the clusters. The first post-loss LC change trajectory was called Water Persistent (n=4,274 or 0.5% of the total number of post-loss LC observations) because sequences were most often in the *Water Bodies* class. The second and largest group of trajectories was classified as Open Forest Persistent (n=486,350 or 60.6%). This particular group characterises a uniform presence of *Open Forests* from the initial years (1-2 years) following deforestation. The third group was called Dense Forest Persistent (n=83,823 or 10.4%) as the timing of other LC types is very short (less than 4 years) and not as predominant following deforestation as the other typologies. The fourth and sixth groups were identified as Open Forest-Farmland (n=22,311 or 2.7%) and Farmland-Open Forest (n=34,605 or 4.3%) because both categories alternated following conversion of *Dense Forests*. The fifth group was described as Farmland Persistent (n=171,612 or 21.4%) as *Farmland* was mostly present over time.

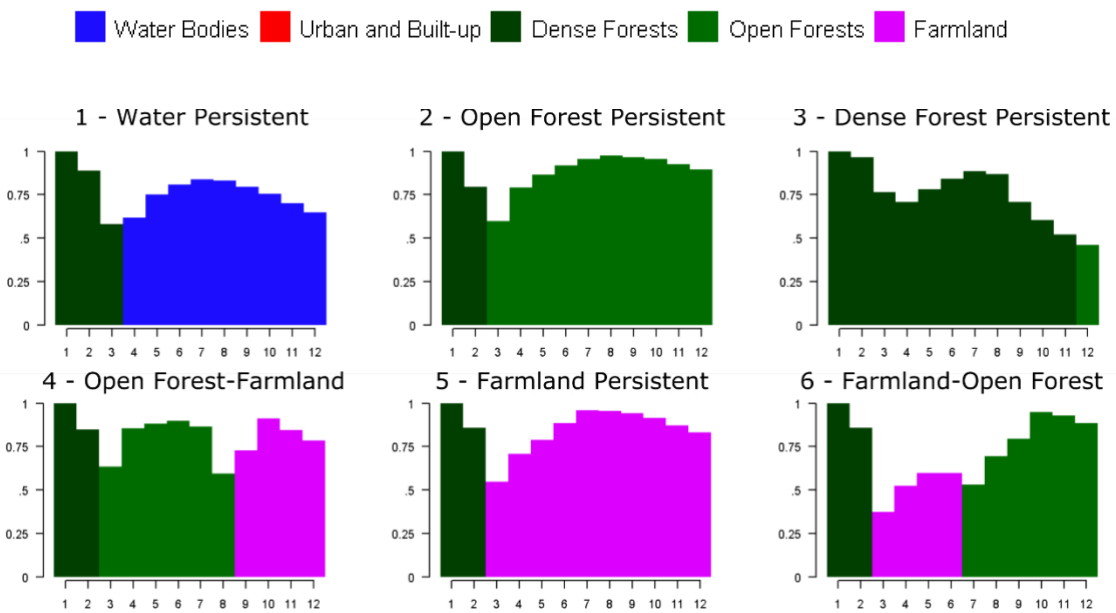


Figure 6.13 Modal state sequence plots for the observed post-loss trajectories. The sequence state i.e. LC type with the highest proportion from 0 to 1 (y-axis) at each point is plotted per year. For instance, the trajectory of Water Persistent show a dominance of water following forest loss.

The six typologies of post-loss LC change trajectories can be further described according to their diversity of LC types per year following deforestation. For this purpose, the Shannon entropy index (SEI) was measured for all observed sequences per post-loss LC trajectory, which provides a measure of how stable

produced land cover remains. Entropy values vary from 0-1, with a value of 0 interpreted as the presence of a single LC type, whereas values of 1 indicate the presence of all possible LC types in a given time. **Figure 6.14** presents the evolution of SEI values over time according to six trajectories of post-loss LC. The values of SEI in Year 1 are equal to 0 as only Dense Forest is present in all sequences irrespective of the post-loss LC trajectory. The first peak (SEI > 0.4) between Year 3 and 4 coincides with the date of Terra-i's detection of forest conversion. After this year, SEI values have different patterns by post-loss LC typology. An interesting pattern is in the Open Forest-Farmland (cluster 4) which has a second peak. This can be interpreted as an abrupt change between Open Forest and Farmland. In contrast, this second peak is absent in the Farmland-Open Forest trajectory indicating a more gradual change between both LC types. For the remaining trajectories, SEI values decrease after the first peak however gradual to sharp increases which are observed from the seventh year following deforestation.

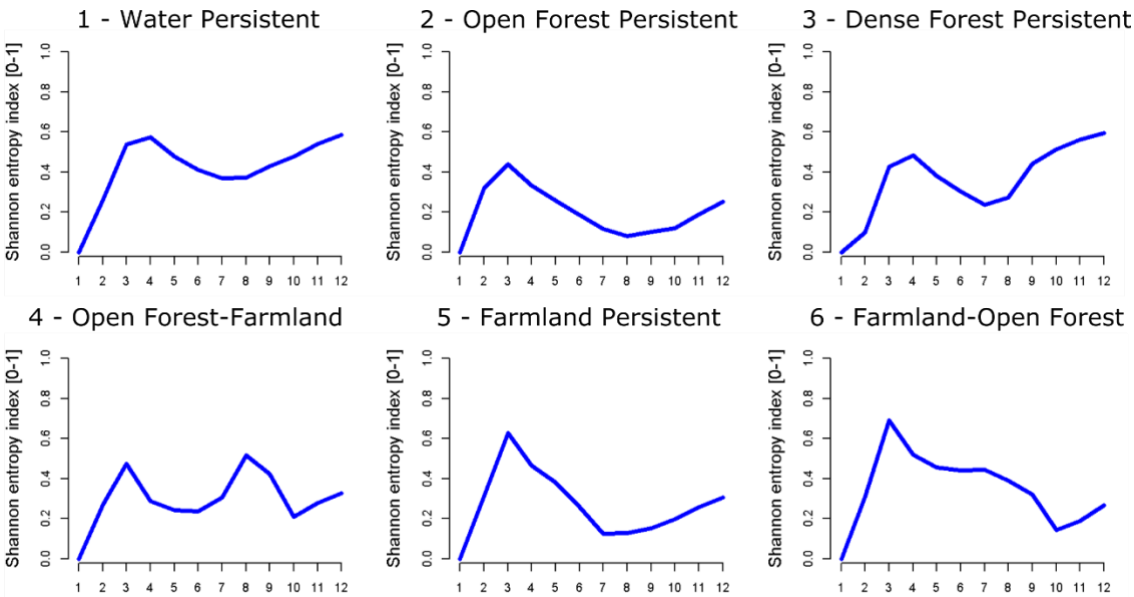


Figure 6.14 Values of the Shannon entropy index (SEI) are displayed for each post-loss LC trajectory identified by Ward's hierarchical cluster analysis over a minimum period of 12 years.

A final exercise consisted of visualising the spatial patterns of the observed post-loss LC change trajectories across the study area. **Figure 6.15** and **Figure 6.16** depict zoom-in maps showing the spatial distribution of post-loss LC trajectories separated by relative size of the deforested areas as large to medium (A) and medium to small (B) across the study area.

Figure 6.15 Spatial distribution of large to medium post-loss LC trajectories over five areas across the target habitat, the Tropical Humid Broadleaf Forests, in the Amazon region. Planet Imagery (March-Oct 2019) and Open Street Maps are used as base maps.

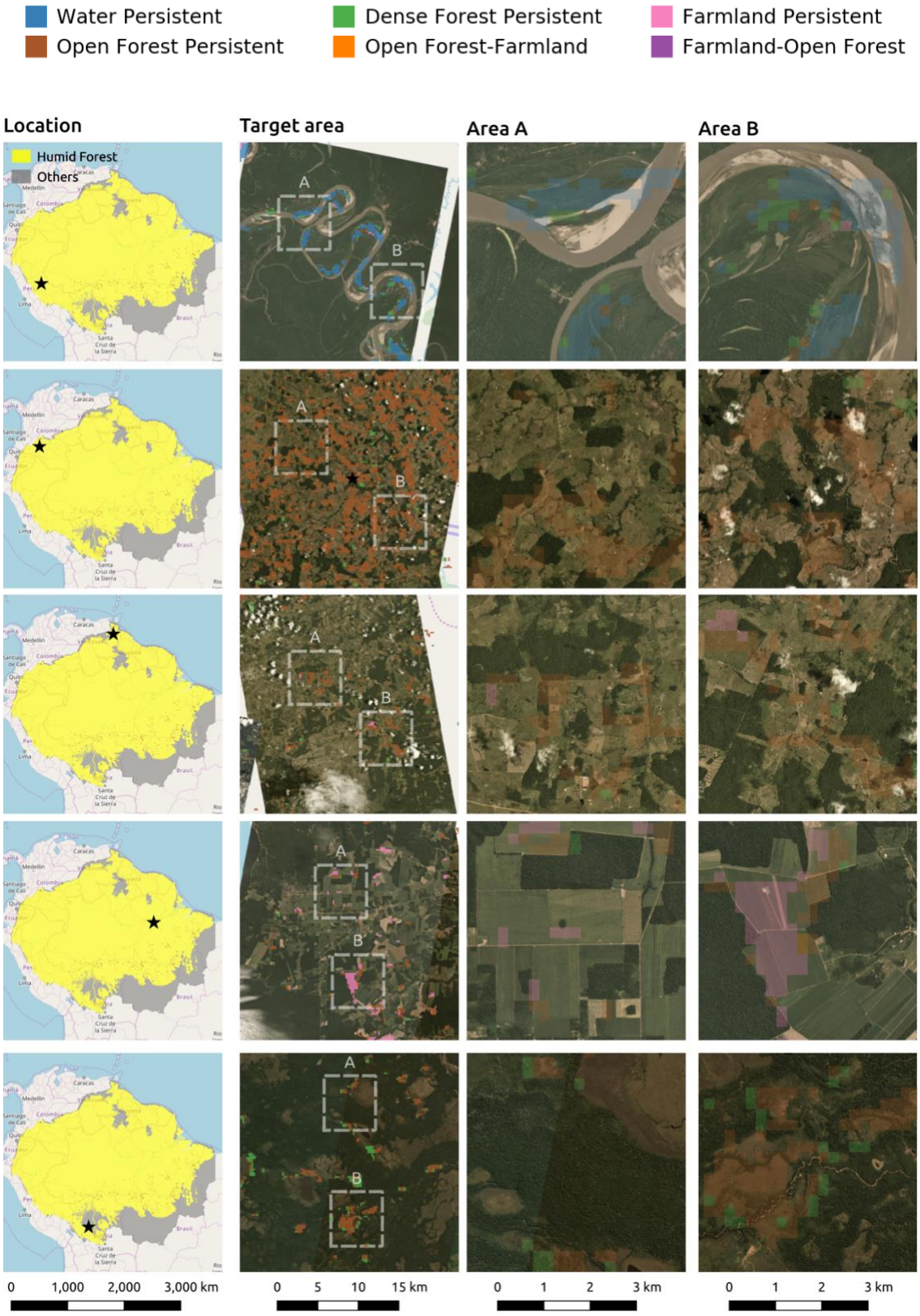


Figure 6.16 Spatial distribution of medium to small post-loss LC trajectories over five areas across the target habitat, the Tropical Humid Broadleaf Forests, in the Amazon region. Planet Imagery (March-Oct 2019) and Open Street Maps are used as base maps.

6.6.2 Explanatory

All auxiliary datasets described in **Section 6.4.1.2** except proximity to Barren were explored in order to understand their relationship amongst post-loss LC change trajectories characterised in **Section 6.6.1**. The exclusion of the proximity to Barren was due to its odd spatial distribution even after removing salt and pepper pixels (see **Figure 6.5**). **Figure 6.17** provides a series of boxplots indicating the values of the auxiliary layers by post-loss LC trajectory. The layers are grouped as internal or external. The former group refers to proximity-based layers generated from the 2019 LC map, which was also used to derive the observed post-loss LC trajectories. The external group indicates those layers which are independent of the generation of the post-loss LC data. Only post-loss LC sequences between the 5th and the 95th percentile per variable were retained. This procedure allows the discarding of extreme values affecting interpretation and posterior analysis. As a result, 783,315 out of the original post-loss LC sequences (N=802,975) were retained. The percentage of observations retained by post-loss LC change trajectory were 72.3% (Water Persistent), 97.9% (Open Forest Persistent), 92.4% (Dense Forest Persistent), 99.2% (Open Forest-Farmland), 99.1% (Farmland Persistent) and 99.2% (Farmland-Open Forest).

After the above pre-processing, some notable patterns can be observed in the boxplots amongst the auxiliary layers and post-loss LC change trajectories. In the internal variables, as expected, the trajectory of Water Persistent is closer to *Water bodies*. This trajectory also has considerably higher median values, shown by the line that divides the box into two parts, than the remaining trajectories in the layer of proximity to *Open Forest*. The trajectories related to the presence of Farmland (Open Forest-Farmland, Farmland-Open Forest and Farmland Persistent) are further distanced from *Dense Forests* than from Water Persistent, Open Forest Persistent and Dense Forest Persistent trajectories. For the external layers, the Water Persistent trajectory is slightly differentiated from other trajectories in four (elevation, slope, precipitation and proximity to exploitation PAS) out of six variables. Apart from this difference, there are no remarkable patterns amongst trajectories not linked to Water. The median values and interquartile range (IQR) of the internal and external variables according to the post-loss LC change trajectories are further detailed in **Appendix VIII**.

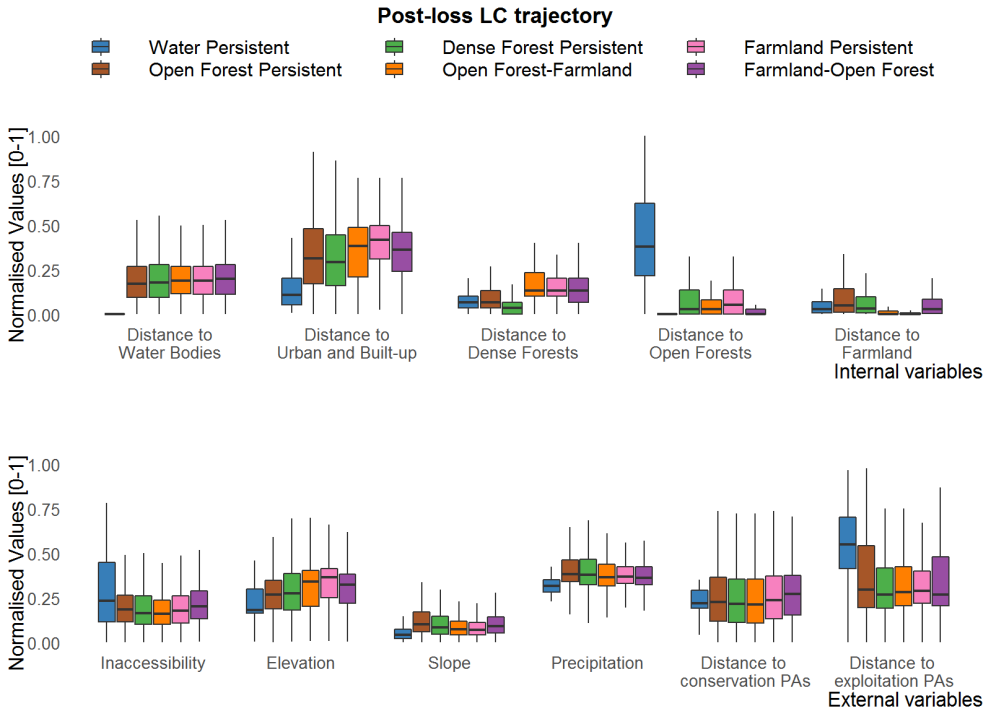


Figure 6.17 Boxplots indicating the values of internal and external auxiliary layers according to six post-loss LC trajectories distributed across the Humid Forest Habitat in the Amazon. For visualisation purposes, the values of all variables are normalised and scale between 0 and 1 using global feature min/max normalisation.

To statistically determine differences in the distribution of each variable amongst the post-loss LC change trajectories, the Kruskal-Wallis (for multiple comparisons) or Wilcoxon-Mann–Whitney U (for single comparison) tests were used. These tests were chosen after assessing the main assumptions on the properties of each variable. This assessment indicated all variables were not normally distributed and had outliers. Consequently, both non-parametric (distribution free) tests were the most suitable to conduct the statistical analysis. The Kruskal-Wallis reveals significant differences ($p<0.0001$) for all variables (see **Table 6.6**). The post-hoc test using Wilcoxon-Mann–Whitney U with the Bonferroni correction indicated all pair-wise group comparisons for all set of variables were significantly different ($p<0.05$ to $p<0.0001$) except two, the comparison of Water Persistent vs Farmland-Open Forests for the variable of proximity to Farmland and Water Persistent vs Open Forests Persistent for the variable of proximity to conservation PAs (see **Table 6.7** and **Table 6.8**).

Table 6.6 Kruskal-Wallis test conducted to compare differences in distribution of internal and external variables amongst groups of post-loss LC change trajectories. **** p<0.0001.

Variable	Type	Chi-square	Degrees of freedom	p-value significance
Proximity to Water	Internal	8369	5	****
Proximity to Urban/Built-up	Internal	26229	5	****
Proximity to Dense Forests	Internal	116190	5	****
Proximity to Open Forests	Internal	143339	5	****
Proximity to Farmland	Internal	211091	5	****
Inaccessibility	External	3296	5	****
Elevation	External	47360	5	****
Slope	External	30153	5	****
Precipitation	External	9850	5	****
Proximity to conservation	External	1806	5	****
Proximity to exploitation	External	2788	5	****

Table 6.7 Post-hoc comparison of internal variables between the post-loss LC trajectories groups according Mann–Whitney U test. Each post-loss LC change group is represented as an acronym as detailed in the table footer. Non-significant (ns) p-values are highlighted in bold. Levels of statistical significance are denoted as * p<0.05; ** p<0.01; *** p<0.001; **** p<0.0001.

Pair-wise Post-loss LC comparison	Internal variable and stats									
	Prox. to Water		Prox. to Urban/Built-up		Prox. to Dense Forests		Prox. to Open Forests		Prox. to Farmland	
	Z-value	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*
WP vs OFP	-599	****	-597	****	-595	****	-589	****	-595	****
WP vs DFP	-246	****	-242	****	-231	****	-224	****	-235	****
WP vs OF-F	-165	****	-161	****	-158	****	-131	****	-137	****
WP vs FP	-357	****	-356	****	-354	****	-342	****	-343	****
WP vs F-OP	-135	****	-131	****	-123	****	-93	****	**-113**	**ns**
OFP vs DFP	-491	****	-481	****	-435	****	-550	****	-467	****
OFP vs OF-F	-545	****	-550	****	-573	****	-569	****	-491	****
OFP vs FP	-445	****	-488	****	-541	****	-556	****	-240	****
OFP vs F-OP	-564	****	-566	****	-574	****	-562	****	-548	****
DFP vs OF-F	-171	****	-188	****	-248	****	-158	****	-100	****
DFP vs FP	-251	****	-296	****	-367	****	-256	****	-130	****
DFP vs F-OP	-187	****	-197	****	-233	****	-148	****	-172	****
OF-F vs FP	-281	****	-293	****	-282	****	-299	****	-271	****
OF-F vs F-OP	-112	****	-106	**	-97	****	-78	****	-154	****
FP vs F-OP	-302	****	-288	****	-290	****	-267	****	-341	****

Water Persistent (WP); Open Forest Persistent (OFP); Dense Forest Persistent (DFP); Open Forest-Farmland (OF-F); Farmland Persistent (FP); Farmland-Open Forest (F-OF).

Table 6.8 Post-hoc comparison of external variables between the post-loss LC trajectories groups according Mann–Whitney U test. Each post-loss LC change group is represented as an acronym as detailed in the table footer. Non-significant (ns) p-values are highlighted in bold. Levels of statistical significance are denoted as * p<0.05; ** p<0.01; *** p<0.001; **** p<0.0001.

Pair-wise Post-loss LC comparison	External variable and stats											
	Inaccess.		Elevation		Slope		Precip.		Prox. to conser. PAs		Prox. to exploit. PAs	
	Z-value	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*	*Z-value*	*p-value sign.*
WP vs OFP	-593	****	-596	****	-597	****	-596	****	**-594**	**ns**	-592	****
WP vs DFP	-231	****	-239	****	-239	****	-239	****	-233	****	-227	****
WP vs OF-F	-142	****	-157	****	-154	****	-154	****	-146	****	-138	****
WP vs FP	-346	****	-354	****	-351	****	-352	****	-349	**	-344	****
WP vs F-OP	-110	****	-126	****	-123	****	-122	****	-117	****	-104	****
OFP vs DFP	-477	****	-496	****	-467	****	-480	****	-483	****	-476	****
OFP vs OF-F	-532	****	-558	****	-524	****	-531	****	-538	****	-537	****
OFP vs FP	-418	****	-520	****	-357	****	-391	****	-437	****	-419	****
OFP vs F-OP	-564	****	-570	****	-554	****	-552	****	-565	****	-558	****
DFP vs OF-F	-162	****	-185	****	-159	****	-159	****	-167	***	-172	****
DFP vs FP	-249	****	-291	****	-221	****	-231	****	-256	****	-255	****
DFP vs F-OP	-195	****	-191	****	-184	****	-173	****	-193	****	-186	****
OF-F vs FP	-290	****	-294	****	-275	****	-282	****	-289	****	-284	****
OF-F vs F-OP	-128	****	-103	****	-119	****	-107	*	-122	****	-110	****
FP vs F-OP	-309	****	-285	****	-312	****	-298	****	-306	****	-299	****

Water Persistent (WP); Open Forest Persistent (OFP); Dense Forest Persistent (DFP); Open Forest-Farmland (OF-F); Farmland Persistent (FP); Farmland-Open Forest (F-OF).

Complementary to the exploration and statistical analysis of the variables, the Recursive Feature Elimination (RFE) approach allowed for i) identifying the most important external and internal auxiliary features and ii) determining the benefit of adding additional features. For the first outcome, **Table 6.9** reports the features ordered by rank and base learner. The proximity to *Water* derived from the 2019 LC map was consistently the most informative auxiliary variable. This feature was retained in all 30 resamples in both MLR and RF models. In terms of the two types of variables, the internal group are the most influential however their order vary according to the base learner model. Amongst the external variables, Elevation and proximity to exploitation PAs appear as relevant for distinguishing the trajectories. In contrast, Slope was removed at the first iteration in both RF

and MRF which means it is not a relevant predictor according to interpretation described in **Section 6.6.2**.

Table 6.9 Ranks of the auxiliary variables as determined by recursive feature elimination (RFE) using multinomial logistic regression and random forest as base learners. Mean feature rank and standard deviation (sd) are reported based on 30 resamples of the post-loss LC dataset. The type of variable, internal or external are also reported in the last column. The ranks are based on twelve increasing levels starting from 1 to 11 by 1. For interpretation, features with major importance are ranked close to 1, whereas less relevant features are ranked close to 11.

(a) Multinomial logistic regression

Feature	Mean	sd	Type
Proximity to Water	1.0	0.0	Internal
Proximity to Dense Forests	2.1	0.3	Internal
Proximity to Urban/Built-up	2.9	0.3	Internal
Proximity to Farmland	4.5	0.5	Internal
Proximity to exploitation PAs	4.5	0.5	External
Elevation	6.2	0.5	External
Proximity to Open Forests	7.7	0.9	Internal
Inaccessibility	8.0	1.1	External
Proximity to conservation PAs	8.2	1.0	External
Slope	10.3	0.6	External
Precipitation	10.6	0.6	External

(b) Random Forest

Feature	Mean	sd	Type
Proximity to Water	1.00	0.00	Internal
Proximity to Farmland	2.00	0.00	Internal
Proximity to Open Forests	3.00	0.00	Internal
Proximity to Urban/Built-up	4.13	0.35	Internal
Elevation	5.60	0.68	External
Proximity to Dense Forests	6.03	1.54	Internal
Proximity to exploitation PAs	6.53	0.63	External
Precipitation	8.23	0.63	External
Proximity to conservation PAs	8.47	0.63	External
Inaccessibility	10.00	0.00	External
Slope	11.00	0.00	External

In the second outcome, there is strong evidence of the existence of complementary information amongst the features, as models' performance changed by the sequential removal of features (see **Figure 6.18**). The RF model achieved in general, higher values of overall accuracy (OA) in comparison with the MRF model indicating a better prediction capability for identifying the six types of post-loss LC trajectories. In the RF model, OA values reach a plateau after eight features with a slight decrease at 11 features. In the MLR model, OA values gradually decreased after 3 features. This can be explained as the MLR model is highly affected by collinearity which potentially, might be present with increases in the number of variables. In contrast, the RF model seems to be more robust as superior values were obtained with the addition of more features.

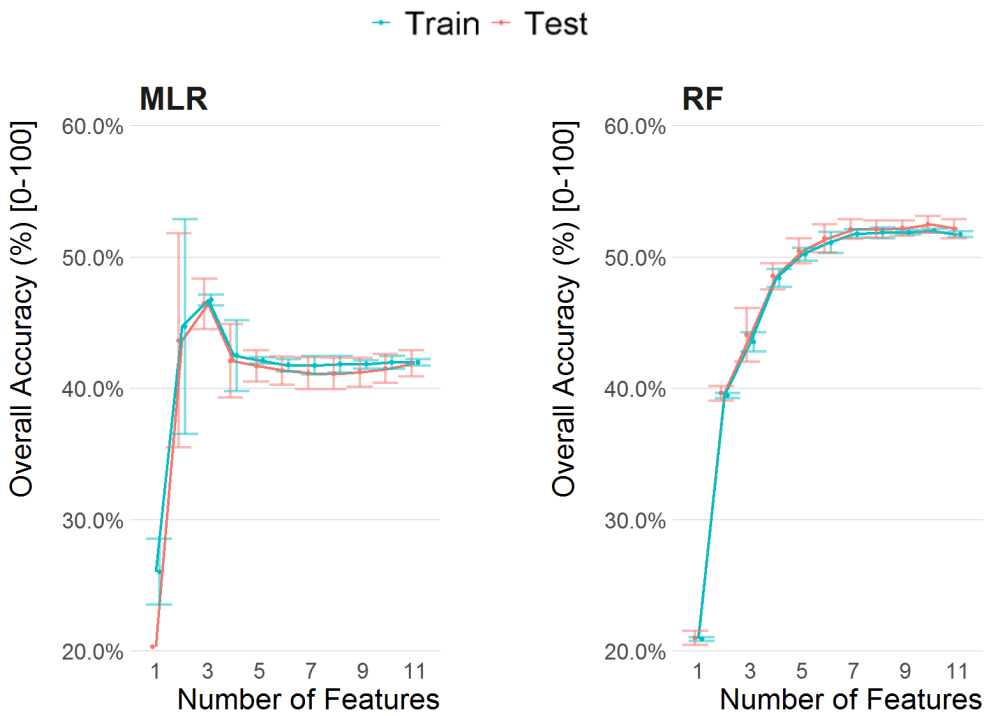


Figure 6.18 Performance profile of multinomial logistic regression (MLR) and random forest (RF) based on auxiliary features to predict post-loss LC trajectories as a function of the number of features used. Mean performance and standard deviation are shown based on 30 resamples of the data.

Under the assumption that the Water persistent trajectory can be attributed to natural causes, the observations under this typology were excluded in a second run of the RFE algorithm. This allows a focus on determining the most important factors of post-loss LC change trajectories induced by human activities. As RFE using the RF model provides better prediction capability than the RFE-MLR, the

analysis was solely run with the former model using 30 resamples and twelve increasing levels of features. **Table 6.10** summarises the ranks of the variables. It appears internal variables remain important however the proximity to Water is no longer a critical factor. The proximity to Farmland, proximity to Urban/Built-up and proximity to exploitation PAs are ranked as the top three most important variables. Additionally, it is confirmed that slope is not a key factor determining the post-loss LC change trajectories.

Table 6.10 Ranks of auxiliary variables as determined by RFE using random forest from post-loss LC data excluding observations of the Water persistent trajectory. Mean feature rank and standard deviation (sd) are reported based on 30 resamples. The ranks are based on twelve increasing levels starting from 1 to 11 by 1. For interpretation, features with major importance are ranked close to 1, whereas less relevant features are ranked close to 11.

Feature	Mean	sd	Type
Proximity to Farmland	1.1	0.5	Internal
Proximity to Urban/Built-up	1.9	0.3	Internal
Proximity to Exploitation PAs	2.9	0.3	Internal
Elevation	5.3	0.9	Internal
Proximity to Water	5.7	1.9	External
Proximity to Dense Forests	6.0	0.0q	External
Proximity to Conservation PAs	6.0	2.1	Internal
Proximity to Open Forests	7.0	0.0	External
Inaccessibility	9.0	0.3	External
Precipitation	9.9	0.3	External
Slope	11.0	0.0	External

Regarding the influence of the number of features, a similar pattern to the first run with the Water trajectory displaying a slight decrease of accuracy at 10 features was observed (see **Figure 6.19**). Higher accuracy values than the RF model including Water persistent can be attributed to more observations for each training iteration i.e. resample.

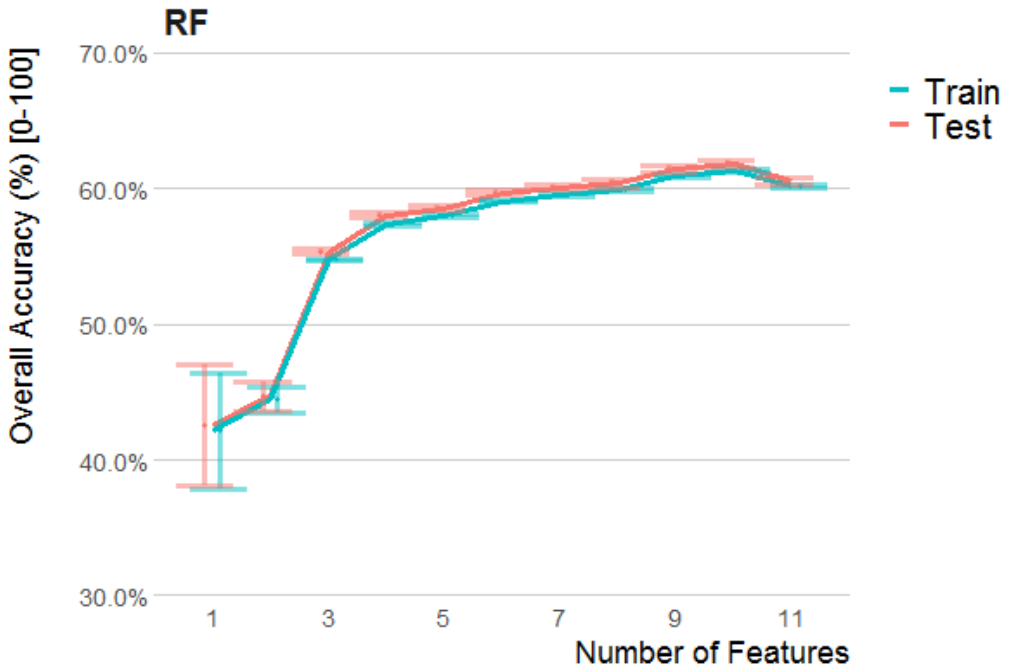


Figure 6.19 Performance profile of random forest (RF) based on auxiliary features to predict post-loss LC trajectories, excluding Water persistent, as a function of the number of features used. Mean and standard deviation are based on 30 resamples of the data.

6.7 Discussion and conclusions

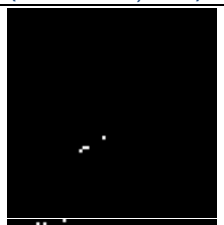
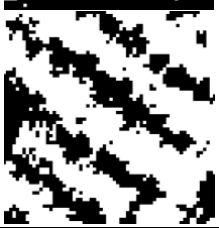
Application of sequence analysis (SA) and other data-driven techniques to long-term (>2 years) post-loss LC change data extracted over a large area, such as the Amazon humid forest, yielded novel insights into the study of LC/LU dynamics following deforestation. For instance, it was found that a minimum period of 12 years after forest loss is needed in order to reasonably characterise different typologies of post-loss LC change trajectories. Combining SA with cluster analysis allowed for grouping post-loss LC sequences with similar features (in this case, based on transition rates). Both exploratory and explanatory stages guided the outcomes of this work.

With the first stage, distinct typologies of post-loss LC change trajectories were defined and characterised. As a result, six distinct typologies were derived according to SA analysis of 12-years of LC data over deforested areas delineated by a well-established deforested product (Terra-i). Distinct from earlier studies mapping immediate (1-2 years) and long-term (>2 years) post-loss LC change as indicated in the literature review (see **Section 2.3.2**), the study of the chronological LC sequences using SA offers a wider comprehensive picture of

the pathways of forest change. This is particularly the case over a large and highly dynamic area such as the Amazon region.

The finding of different trajectories implies estimates of carbon emissions and uptake should go beyond traditional assumptions of a full recovery or complete conversion to non-forest after disturbance. This is in agreement with Tang *et al*. (2020) who claim the benefits of injecting complex post-loss LC dynamics into a spatially explicit carbon bookkeeping model. Moreover, the characterisation of typologies of post-loss LC change trajectories is complementary to previous work describing deforestation spatial patterns over a similar geographical extent (Coca-Castro, 2015) as well as smaller extents in the Brazilian Amazon (de Filho and Metzger, 2006; Saito *et al*., 2012). From total post-loss LC sequences (N= 802,975), some 81% is dominated by two typologies, Open Forest Persistent (60%) and Farmland Persistent (21.4%). The former typology represents the conversion from *Dense Forests* to *Open Forests*. According to the definitions of the LC types defined in **Table 6.4**, this conversion means tree cover changed from >60% to 10-60%. A visual inspection of the post-loss LC change trajectories reveals most of the deforested areas with the Open Forest Persistent trajectory coincide with three typologies of deforestation spatial patterns named as diffuse extensive, diffuse intensive and multidirectional (see **Table 6.11**). These typologies are mainly characterised by small to medium scale clearings, the former being mostly dominant in the non-Brazilian Amazonian countries (Kalamandeen *et al*., 2018). In contrast, a geometric pattern is associated with the Farmland Persistent trajectory. In this trajectory, *Dense Forests* are converted to areas dominated by herbaceous annuals (<2-m) with at least 60% cultivated. In terms of carbon emission, both dominant pathways can have different implications. Carbon uptake from the remaining tree cover in areas with the Open Forest Persistent trajectory can compensate for carbon emissions produced by the clearings. This balance is less likely to be probable in Farmland Persistent areas which are dominated by herbaceous annuals (<2-m).

Table 6.11 Description of four deforestation spatial pattern typologies over a grid size of 15 km x 15 km found in the Amazon Humid Forest according to a non-forest/forest map derived from 2004-2013 aggregated Terra-i data. Adapted from Coca-Castro (2015).

Pattern	Visualisation Forest Non-Forest (scale 1:75,000)	Description	Land-use agents associated / Spatial distribution / Accessibility
Diffuse extensive		Small-scale clearings	Smallholder subsistence agriculture / Dispersed or scatter distribution / Low accessibility
Diffuse extensive		Small to medium scale (irregular or geometric) clearings	Roadside or riverside colonization by spontaneous migrants / Clustered distribution / Low to medium accessibility
Geometric		Large-scale (geometric) clearings	Modern and industrial sector activities / Clustered distribution / Medium to high accessibility
Multi-directional		Corridor-like clearings (irregular or geometric) perpendicular to a main corridor clearing	Planned resettlement schemes (mostly in the Brazilian Amazon) / Clustered and/or scatter distribution / High accessibility

Next to these dominant post-loss LC change trajectories, Dense Forest Persistent represents 10% of observed post-loss LC sequences. Whilst this trajectory can be associated with tree-like vegetation regrowth, a visual inspection indicates that some areas are related to the establishment of tree plantations e.g. oil palm rather than natural regrowth. The separation of both types is critical as each land use has different implications in carbon emissions and biodiversity (Fagan *et al*., 2015). In terms of spatial patterns, irregularly shaped clearings are more likely to be related to forest regrowth so this feature could be useful in separating it from tree plantations. This separation could contribute to studies focused on the forest transition curve as defined in **Section 2.2.2**.

Compare to the current literature on post-deforestation land use in the Amazon, the Open Forest-Farmland and Farmland-Open Forest, which share 2.8% and

4.3% of the total number of post-loss LC sequences respectively, support the relevance to study and monitor the complexity of the post-deforestation dynamics. According to the literature related to the approximation to the causes of pantropical forest change (see **Section 2.2.3**), scholars suggest, in general, LC/LU conversions/transitions following deforestation do not follow a fixed pattern (van Vliet et al., 2012). This non-fixed pattern is true in a large and socio-environmental heterogenous area as the Amazon. For instance, Tang et al. (2020) used 16 years of Landsat imagery to identify post-deforestation land cover dynamics in the Colombian Amazon. The authors claimed as a common land conversion the sequence of primary forest (5 years) followed by secondary forest (6 years) and then pasture/cropland (remaining years). The identification of this particular pattern motivated the authors to propose a spatially-explicit carbon bookkeeping model with a more accurate approximation to the carbon emissions and uptake than traditional models assuming a full recovery or complete conversion to non-forest after disturbance. Land abandonment in the Brazilian Amazon region, where most of the Farmland-Open Forest post-loss LC pattern were distributed, has been attributed to a range of factors reduced crop productivity, lack of financial incentives, migration patterns, non-traditional land uses and market fluctuations (Perz and Skole, 2003). Moreover, it has been suggested the land-use practices prior forest regeneration have a determinant influence on the vegetation regenerating following land abandonment in the Amazon (Chazdon et al., 2007).

For the areas with the Open Forest-Farmland post-loss LC pattern, it might be caused by land use activities such as soil preparation for suitability for agricultural or pasture production (Galford *et al*., 2010). In addition, according to their proximity to intersecting areas between Open Forest Persistent and Farmland Persistent, some of these areas might relate to forest edges. A recent study by Brinck *et al*. (2017) indicated edge-related carbon emission accounts for 31% of current estimated annual carbon emissions from the pantropical forest. Therefore, the characterisation of post-disturbance LC/LU change trajectories in edge-forest areas is relevant for improving baselines for emission limitations schemes e.g. REDD+ (Numata *et al*., 2011), because, for example, transitions from forest to permanent cropland have much greater emissions than transitions from forest back to forest regeneration. The minor post-loss LC trajectory, the Water

Persistent (0,5%), is the result of natural processes such as meandering rivers. This is inferred due to the location of this trajectory at the Western Amazon where hydrologically dynamic wetlands have been characterised (Reis *et al*., 2019).

Whilst the interactions of trajectories of post-loss LC change with carbon dynamics were described previously, many other ecosystem services to which these are also relevant including pollination, pest control, water quality, quantity and regulation. For instance, Gutierrez-Arellano and Mulligan (2018) indicated that some key regulation services provided by fauna, so-called faunal ecosystem services, can be negatively impacted globally by the transformation of forests/woodlands to agricultural land (or agriculturisation). However, agriculturisation can present diverse dynamics in forest-agriculture frontiers and consequently impacts in such services. For example, slash and burn, a low-intensive agriculture, continuously create areas of successional growth which might sustain pollinisation services (Laso Bayas *et al*., 2017), whereas intensive monoculture agriculture do not due to increased habitat isolation, and reduction of floral resources and nesting areas in remnant habitats (Kennedy *et al*., 2013). It is expected that further information of post-loss LC/LU dynamics assists scholars in providing greater accuracy in estimates of the impacts of different trajectories like agriculturisation on pollinisation and other key ecosystem services.

At the second stage of this work, the statistical difference and influence of 11 spatial layers representing inaccessibility, biophysical and policy factors were examined over the observed typologies of post-loss LC change trajectories. These layers, grouped as internal and external, were described and analysed under statistical- and machine learning-based approaches. According to the statistical analysis, the Kruskal-Wallis test showed the differences in the distribution in each variable, internal or external, by post-loss LC group are statistically significant ($p<0.00001$). These differences were assessed in more detail by the post-hoc tests of Mann–Whitney U which indicated only two out 30 per 165 pair-wise group comparisons were not significant ($p>0.05$). This outcome reinforces the groups are mostly distinguishable according to the set of 11 variables analysed. To determine to what extent these variables are useful to classify the observed types of post-loss LC change trajectories, the machine learning-based algorithm of Recursive Feature Elimination (RFE) was

implemented. This approach, which was found to be optimal, using Random Forest as reported by previous studies (Jeune *et al*., 2018), allowed determining internal variables contributed more consistently than external variables to distinguish the observed post-loss LC change typologies. The exercises excluding the minor trajectory, Water Persistent, show proximity to Farmland, Urban/Built-up and Water as being the three most important internal variables determining non-water related post-loss LC sequences. As expected, when observations of the Water Persistent trajectory are included, proximity to Water ranks as more important than proximity to Farmland and proximity to Urban/Built-up. From the external variables, only proximity to exploitation PAs and elevation seem to be relevant to the observed typologies. Whilst the exploitation PAs refer to areas with sustainable use, a recent study in the Brazilian Amazon suggests such areas have not avoided deforestation (Jusys, 2018). Therefore, the characterisation of post-loss LC change trajectories as provided in this work are fundamental to a better understanding of potential land uses proximate to these areas. Regarding elevation, the Farmland related trajectories tend to be clustered in areas with higher elevation, as in the Eastern Amazon, than Open Forest Persistent, Dense Forest Persistent and Water Persistent typologies, most of them located in lowlands in the Western Amazon. Higher values of elevation and slope have been seen as a constraint factor in deforestation studies (Busch and Ferretti-Gallon, 2017), however in this study only elevation was relevant to the observed post-loss LC trajectories.

7 Conclusions and future work

This chapter presents the concluding remarks of this thesis. **Section 7.1** briefly states the rationale, research aim and contributions. This is followed by **Section 7.2** synthesizing results, limitations and future work of **Chapters 2-6** in addressing five research objectives. **Section 7.3** concludes with reflections on how the research questions were addressed and the implications of the findings for understanding post-loss LC/LU change trajectories and impacts on ecosystem services in the pantropics.

7.1 Review of purpose and contributions

Forests play a substantial role in biodiversity protection, climate regulation and human livelihoods. Despite efforts in measuring and modelling forest cover and forest change, few advances exist in the study of LC/LU dynamics following deforestation (so called post-loss LC/LU change). Existing baselines for the calculation of global/regional carbon fluxes are primarily based on the immediate (1 or 2 years) post-loss LC/LU change. Therefore, these fluxes are based on simplistic assumptions such as a full recovery or complete conversion to non-forest after disturbance. Same naive assumptions lead to biased conclusions of the impact of forest loss in biodiversity and ecosystem services. The current plethora of multi-source datasets with considerable records (more than 10 years) and new methods to analyse them now makes it feasible to study complex long-term LC/LU dynamics over deforested areas. The study of such dynamics poses new opportunities and challenges for the development of theory and application of these methods to land system research.

This thesis was developed from previous work related to field validation of Terra-i's deforestation alerts in the Peruvian Amazon (Coca-Castro *et al.*, 2014) and mapping of spatial patterns of deforestation across the Amazon humid forest (Coca-Castro, 2015). Both early investigations were valuable as they indicated a need for further contribution to the study of complex LC/LU dynamics in post-disturbed areas. To address this need, this thesis aimed to develop models for mapping and analysing long-term trajectories of post-loss land cover and land-use change pantropically, based on time series analysis of earth observation (EO) and contextual data, for use in environmental and agricultural studies. This aim

was addressed by five research objectives set in **Chapter 1**. In summary, the main contributions involve:

1) Deployment of an end-to-end deep learning architecture capable of ingesting and processing large volumes of high frequency multispectral time series for generating rapid annual LC predictions suited to large-area study of post-loss LC change trajectories.
2) Understanding and explaining the patterns observed in long-term LC dynamics following deforestation with a novel set of techniques for mining chronological information.

7.2 Summary of findings, limitations and future work

This thesis consisted of five objectives, one related to a state-of-the-art review and the others being empirical. The preliminary results of some of these objectives were socialised in multiple academic events as described in **Appendix IX**. The main findings, conclusions and future work by objective are described below. The section ends with a reflection of how errors and uncertainties may have propagated through the analysis chain.

7.2.1 State-of-art review in the study of post-loss LC/LU change

The first objective aims to explore existing knowledge and identify gaps regarding post-loss LC/LU change modelling and monitoring for tropical areas. This objective was fulfilled in **Chapter 2** by a state-of-art review of theoretical, methodological and policy considerations in forest change research. This review allowed identification of a rich body of theory and definitions pertinent to the study of post-loss LC/LU. For instance, the definitions of forest, land cover and land use (see **Section 2.2.1**), which are ambiguously used in theory building and methodological developments, were key to the delineation of areas of interest and dynamics targeted in this thesis. In this regard, Terra-i's detections, which were used as reference for deforested areas, include abrupt disturbances in non-forest land surfaces (e.g. burning pasture, changes in tree plantations, among others). For the purpose of this thesis, only detections representing *forest loss*, defined as the replacement of pristine and/or secondary tree cover by anthropogenic land uses (e.g. farmlands, mining and/or human settlements) or by natural events (e.g. flooding, wildfires), were analysed. The distinction between land cover and land

use was also relevant for the purpose of indicating only post-loss LC/LU dynamics which are able to be measured in **Chapter 2**, whereas, due to limitations in the availability of LU data at large scale, the remaining chapters are limited to the analysis of post-loss LC change trajectories.

The revision of existing methodological advances in **Section 2.3.2** confirmed the lack of post-loss LC/LU research. Whilst records of pantropical forest change start from 2000 with different temporal resolutions, the study of post-loss LC/LU change has a limited temporal scope and is predominantly restricted to the immediate change (1-2 years). Only few methodological advances exist in mapping long-term (>2 years) post-loss LC/LU change at annual (Fagua and Ramsey, 2019; MapBiomas Project, 2019) or biennial (Almeida *et al*., 2016; Arévalo *et al*., 2019) time steps. For future work, it is suggested a conduction of a systematic review focused on long-term post-loss LC/LU change at multiple scales from local to global would be valuable. The increased interest in multi-temporal analysis of LC data fostered by historical records of satellite images, see for example the GLanCE project (Tarrio *et al*., 2019), is likely to lead to a greater volume of publications on post-loss LC/LU change trajectories during this decade.

7.2.2 Post-loss LC/LU change trajectories at local scale

Chapter 3 accomplishes objective 2: "To conduct an exploratory analysis over deforested sites with high-levels of LC/LU information from in-situ observations to multi-temporal satellite data". The inspection of this information over some 120 sites detected as disturbed between 2011 and 2015 by the Terra-i system provided different insights into the study of post-loss LC/LU change. Firstly, it was found that change detections captured by the Terra-i system have limitations regarding the size of disturbance i.e. changes below 30% of the MODIS pixel are hardly detectable, and time-shift e.g. early or late detections (see section 3.5.1). The identification of these limitations, which are also reported in Tang *et al*. (2019), was fundamental for conducting and interpreting posterior analyses at large scale. Secondly, the visual interpretation of high to very-high satellite data allowed the identification of both LC and LU dynamics over the target deforested areas. In terms of LC, it was found *Tree cover* is gradually reduced prior to forest loss related to increases in *Cultivated and managed* and *Mosaic of cultivated and*

managed/natural vegetation. Regarding LU, contrasting dynamics were observed according to the immediate LU type following deforestation. Those forested areas converted to cropland maintained this class for more than 2 years, whereas areas with immediate conversion to grassland were less persistent over time with non-grassland LU types following deforestation. Thirdly, the dynamics observed within the 500-m buffer areas (also denoted as landscapes), indicate LC/LU change in deforested areas can be described in larger mapping units than pixels. This is in agreement with landscape approaches as described in Bourgoin *et al*. (2020) who mapped 45 years of fragmentation trajectories in a district in Vietnam. In addition to LC/LU dynamics, the study of fire activity offers a complementary perspective of landscape change over deforested areas. For instance, a high number of active fires were observed over 500-m buffer areas with >50% forest/natural vegetation and mosaics.

Whilst the above findings aided in providing a general picture of the complexity in the study of post-loss LC/LU change, the time-consuming nature and small-area coverage of the methods and tools used in this objective are not feasible for large-area analysis. For future work, it is recommended that further exploration of the sample interpretation e.g. GeoWiki, Collect Earth, LACO-Wiki, is conducted, as they are in constant improvement. For instance, GeoWiki's recent campaign 'Human Impact on Forests' could be extended to analyse not only the immediate LC/LU following forest disturbance but also long-term changes.

7.2.3 Pre-existing global LC classifications and high frequency multispectral time series suited to large-area study of post-loss LC change

The existing work in the study of post-loss LC/LU change trajectories at large scale is based on LC/LU data generated at annual or biennial time steps. These data and related nomenclature *i.e.* classes are generally customised according to the spatial unit of analysis (see **Section 2.3.2**). For this thesis, the unit was defined by the nominal spatial resolution of Terra-i's deforested areas (250-m). Building upon initial results of an early investigation assessing a deep learning (DL) architecture called Multitemporal Land Cover Classification (MTLCC) network for large-area annual LC predictions across the Brazilian Amazon (Coca-Castro *et*

al., 2019), the third objective of this thesis aimed to "assess pre-existing global LC classifications and high frequency multispectral time series suited to large-area study of post-loss LC change trajectories". In this regard, **Chapter 4** introduced the MTLCC network and related notions in DL for remote sensing classification. This was followed by an assessment of its effectiveness according to nine LC classification schemes (also referred to as labelled datasets) derived from pre-existing global LC products. Some of these classification schemes, denoted as hybrid datasets, were proposed under the assumption they lead to greater accuracy by exploiting spatial, temporal and nomenclature attributes from the original LC products. The effectiveness of the trained MTLCC network models using original and hybrid datasets was assessed according to a two-stage procedure particularly focused in their suitability to the study of post-loss LC change.

The assessment confirmed that the MTLCC network is efficient enough to handle large volumes of multispectral time series with minimal pre-processing. However, this capability is influenced by characteristics of the labelled dataset (size, number of classes, labels noise). Overall, the models trained using hybrid datasets yielded better performance than those using the original pre-existing LC products. In particular, M11* hybrid dataset derived from temporal attributes (multitemporal) of 2001-2015 MODIS FAO-LCCS2 layer achieved lower total disagreement values in comparison to C9* hybrid dataset created using spatial attributes i.e. subpixel information from 2015 CGLS-LC100 dataset (see **Section 4.6**). This can be explained as the number of misclassified pixels is considerably reduced by exploiting the temporal attributes as reported in previous work (Zhang and Roy, 2017; Huang *et al*., 2015). In contrast, the spatial attributes of the single-time step CGLS-LC100 map to derive C9* are less effective to reduce this noise. This suggested the hybrid dataset generated from the multi-temporal MODIS-FAO-LCCS2 layer provides the best suited labelled dataset to train the MTLCC network. Furthermore, the trained model with this dataset provided consistent predictions aimed at the study of long-term post-loss LC change as reinforced in the cross-year comparison against the regionally tuned Mapbiomas Amazonia.

The analysis covered in this objective provided some interesting results of the impacts of multiple classification schemes on the MTLCC network. For future

work, it is recommended to follow-up new versions of the CGLS-LC100 FAO-LCCS dataset as the hybrid dataset, C9* yielded also reasonable results. According to the providers of this dataset, annual updates of this product will be provided at a global scale as it was released and validated to Africa (Tsendbazar *et al.*, 2017). Consequently, the semantic-spatial rules for producing C9* could be complemented in the near future with temporal attributes, as was successfully done with MODIS Land Cover Type products.

7.2.4 Calibration and deployment of the proposed data analytics technique method for the study of post-loss LC change

Chapter 5 fulfilled objective 4: "To calibrate and implement the proposed data analytics technique for generating 19-years of LC information using the best suited satellite and labelled data across a large area in the tropics". The MTLCC network was calibrated according to four main hyper-parameters (batch size, number of hidden states, learning rate and optimiser type) and compared against conventional shallow machine learning algorithms in remote sensing classification, Random Forest (RF) and Support Vector Machine (SVM). Both algorithms were also calibrated in order to yield their maximum potential according to the input training data. Following the outcomes of **Chapter 4**, the best hyper-parameter configurations of the MTLCC, RF and SVM models were obtained by using LC reference data derived from 2001-2015 MODIS FAO-LCCS2 layer and 2001, 2002 and 2003 MODIS satellite observations (see **Section 5.5.1**).

Overall, the calibrated MTLCC network achieved the highest classification accuracy compared with the shallow learners. The analysis of F1-score per-class indicates that there is substantial gain in accuracy for six out of eight target-LC classes (see **Section 5.5.2**). A visual inspection of the predictions indicated the MTLCC network provides in general, the closest match to the reference data observations including over noisy, i.e. cloudy, areas (see **Section 5.5.3**). This capability to discard noise is important in pantropical areas with high percentage of cloud cover such as the Chocó-Darien Global Ecoregion (Fagua and Ramsey, 2019).

The work in **Chapter 5** is not free of limitations. It was observed that RF followed by SVM requires less computation time for training than the MTLCC network. In terms of inference, the difference between the classifiers is notably reduced. Ideally, future work should compare the performance of the proposed method against other classifiers on similar datasets, reporting the time for training and inference as well as disk usage as proposed by Garnot *et al*. (2019b) and Pelletier *et al*. (2019). Furthermore, the role played by other additional input features in the MTLCC network (i.e. vegetation indexes, slope, terrain) needs some further investigation as preliminary work demonstrated some influence in the output (Coca-Castro *et al*., 2019). In terms of classifiers, the MTLCC network needs further comparison with other established end-to-end trainable architectures capable of handling the temporal dimension, and where possible the spatial dimension of satellite image time series. Amongst these architectures, it would be interesting to compare the performance against emerging deep neural networks suited to satellite image time series analysis such as self-attention models as proposed by Brandt (2019), Garnot *et al*. (2019b), and Rußwurm and Körner (2019). These particular networks seem to be faster and more parsimonious in memory than the MTLCC network. Finally, the use of cloud-computing should be explored further as it facilitated running the large-scale experiments of this thesis. In this regard, thanks to a research grant awarded by the Microsoft AI for Earth programme, £5500 in cloud credits will be used to explore future work suggested in this thesis.

7.2.5 Understanding the spatial and temporal patterns of the observed post-loss LC change trajectories

After the deployment of the MTLCC network, 19-years of LC data were successfully generated across the Amazon. **Chapter 6** aimed at the discovery of patterns in such records through the analysis of long-term post-loss LC change. In this regard objective 5: “To understand the spatial and temporal patterns of observed post-loss LC change trajectories” was addressed in this chapter through characterising and mapping typologies of post-loss LC change trajectories using the sequence analysis (SA) approach.

Thanks to the integration of SA and other data-driven techniques e.g. clustering, it was possible to determine the optimal period to identify distinct typologies or pathways of post-loss LC change. As a result, 12 years of annual LC data was identified as the optimal period for a large area such as the Amazon humid forest. For this period and geographical extent, some 81% of the observed post-loss LC sequences (N=802,975) have two marked pathways from *Dense Forests* defined as tree cover >60% to i) partial conversion (60%) with tree cover between 10-60% and ii) complete removal of tree cover (21%). Interestingly, two additional pathways sharing 7.1% shows alternation between partial and complete removal. The distinction of such typologies and others described in **Section 6.6.1** demonstrated the complexity of LC dynamics in post-disturbed areas. This complexity is therefore relevant in on-going efforts of the estimation of carbon emission and impacts of deforestation in biodiversity and ecosystem services in the pantropics. To provide better estimates of carbon fluxes in observed post-loss LC change trajectories is recommended to follow guidelines of previous research quantifying biomass change and land cover change processes, see for example McNicol *et al.* (2018) and De Sy et al. (2019). The results derived from SA did not provide a direct evidence of the associated land-use to observed post-loss LC change trajectories. However, as shown in **Section 6.7** some spatial patterns of deforestation could help to link LU types to observed typologies of post-loss LC change.

The methodology based on the Recursive Feature Elimination (RFE) approach allowed determining the influence of 11 spatial layers representing inaccessibility, biophysical and policy factors on the observed typologies of post-loss LC change. It was found that the internal variables, which refer to proximity-based layers generated from the 2019 LC map, contribute more consistently than the external variables to the observed post-loss LC change typologies. Although these results are not shown in this thesis, the removal of the internal variables reduced the accuracy of the RF model below 35%. Therefore, further work is needed to explore other variables used in the study of disturbances e.g. fire activity, soil suitability, biomass (Laurance *et al*., 2014; Nepstad *et al*., 2008; Le Page *et al*., 2010; Wright and Ziegler, 2017). Additionally, discretisation, which stands for the process of transforming a continuous variable into a discrete one by creating a set of contiguous intervals, might contribute to achieve a better prediction

capability by the RF model (Lustgarten *et al*., 2008). Also, discretisation could help to increase the interpretability of the model (Lustgarten *et al*., 2011), in this case, for a better understanding of the factors associated with long-term post-loss LC change.

It is important to emphasise this study is not free of limitations, mainly associated with the uncertainty inherent to the datasets employed. For instance, although the typologies of post-loss LC change trajectories were described in terms of the timing and variability over time i.e. entropy, they were not validated in a systematic manner. Emerging methodologies have been proposed to monitor and validate LC/LU dynamics in post-disturbance landscapes, see for example Arévalo *et al*. (2019), however most of them are still in their infancy, in particular to validate long-term post-loss LC change data over large areas as the Amazon region. As an alternative, the study of the spatial patterns linked to the observed post-loss LC trajectories could contribute to reduce uncertainties as some preliminary relationships were found in this investigation. In future work, it would be interesting to ingest the information used to derive spatial patterns i.e. landscape metrics as part of sequence analysis, either during the clustering or post-processing. Alternatively, future work can also consider exploring emerging models capable of providing joint-predictions of LC/LU using satellite imagery as proposed by Zhang *et al*. (2019) which synergistic information has been a topic of debate amongst scientist in land system science (Joshi *et al*., 2016).

Finally, whilst SA was useful to the study of post-loss LC change trajectories, it can also have limitations as discussed in Durrant *et al*. (2019). SA does not depend on explicit distributional assumptions, however a range of choices have to be made including a pairwise dissimilarity metric, number of clusters and cluster analysis technique (Han *et al*., 2017; Ritschard and Studer, 2018a). In this work, different settings were implemented and assessed mainly based on previous work by Mas *et al*. (2019). The number of clusters and minimum period were rigorously determined using cluster quality and stability measures aided by the visual exploration of the most optimal solutions. Hierarchical clustering was performed as they have the advantage of easily visualisable results in the form of a dendrogram. The choices presented in this study can be seen as a contribution

to future studies aiming to use and assess SA and related techniques for the characterisation and mapping of long-term post-loss LC change.

7.2.6 Errors and uncertainties

Whilst the findings of this research are encouraging and suggest the suitability of the deep learning method and subsequent sequence/clustering analysis to the study of long-term post-loss LC, it is worth to reflect about the propagation of errors and uncertainties in the analysis chain. This propagation, which is represented in input data themselves as well as assumptions made by the proposed methods for spatial scaling, is discussed as follows.

For the input data, the key uncertainties and errors are related to Terra-i's alerts, MODIS satellite data and reference global LC products. The local and regional scale analysis of **Chapter 2** and **Chapter 6** showed Terra-i's alerts contain false positives i.e. non-forest related change or are lagged i.e. early/late detections. These particular errors might hinder a complete and consistent mapping of post-loss LC in the study area. The propagation of this error was reduced by setting rules to focus exclusively on forest change pixels and tolerate a lagged period of 2 years (see **Section 6.4.2.1**). Equally important are the errors in the generated LC maps due to implicit limitations of the input MODIS satellite data. Although the high-temporal resolution of the products used, MOD09Q1 and MOD09A1, was key to train the proposed deep learning model, the spatial resolution is not enough to fully distinguish certain classes such as mosaics of distinct LC types (e.g., cropland/natural vegetation mosaics) or classes that occur along a climate gradient with mixtures of life forms at sub-pixel scales (e.g., shrubland, sparse vegetation). This is a general problem amongst the existing global LC products derived from coarse satellite imagery such as ESA-CCI (Li *et al.*, 2018) and MCD12Q1 (Sulla-Menashe *et al.*, 2019). Some scholars have proposed to use subpixel level information to more accurately represent the mixture of land covers within a pixel (Kumar et al., 2017). Given the inherent spatio-temporal errors of the global LC maps and other listed in **Section 4.5.2.2**, which were used as sources of labelled data, the propagation of these errors were reduced in this research by generating hybrid maps. In particular, it was shown that the model trained with temporal-driven hybrid maps obtained the best performance (see **Section 4.6.1**).

Regarding the spatial scaling, it was the main operation to resample both the surface reflectance satellite image and labelled data at same spatial resolution of 250-m as the target deforested areas. However, some caveats should be mentioned regarding the main methods used, bilinear interpolation for downscaling surface reflectance satellite image and categorical scaling (downscaling/upscaling) for labelled data. For the bilinear interpolation, it does not require any auxiliary information, however, it suffers from smoothing which means that some of the variance in the original variable is lost in the resample variable (Atkinson, 1988). Other more sophisticated methods for downscaling continuous geospatial data based on fine-scale auxiliary information and/or a process model as compiled by (Ge et al., 2019) should be further explored. A few studies have increased the resolution of the 500-m bands in MODIS using some of these sophisticated methods. Amongst these studies, it is important to highlight Wang *et al.* (2015)'s study which conceptualized a new geospatial method termed ATPRK (area-to-point regression kriging) validating its benefits for LC/LU classification in the Brazilian Amazon. ATPRK takes advantages of the fine spatial resolution information of 250-m in MODIS bands 1 and 2 by regression modelling to downscale the coarse residuals from the regression. The same method (ATPRK) has been applied to sharpen the 20-m bands of Sentinel-2 with the help of the 10-m bands (Wang et al., 2016). For the categorical scaling, the simplest approaches were used for downscaling/upscaling the global LC products. These approaches fail mainly to capture the spatial patterns of LC/LU features. This means some linear patterns of certain classes e.g. water bodies are lost in the resampling operation. More sophisticated methods based on machine-learning and/or auxiliary data as described (Ge et al., 2019) should be further explored.

7.3 Conclusions

This thesis investigated and mapped long-term (>2 years) post-loss LC/LU change in a representative large area of the pantropics (the Amazon humid forest region) by using multiple sources of georeferenced data and novel methods to analyse them. Research objectives were successfully accomplished as shown previously meaning the research questions could be answered as follows:

(1) *Can **land cover** change trajectories (e.g. tree, herbaceous, shrubland, bare areas) following deforestation be monitored at the large scale, based on pre-*

existing contextual data and time series analysis of freely available multispectral EO data?

During the course of writing this thesis there has been an increase in the number of datasets and initiatives aiming to capture post-loss LC/LU change at large scale. However, most of them are still limited to immediate or direct conversion following forest loss, see for example Curtis *et al*. (2018), or rely on conventional machine learning methods to process satellite data, see for example Arévalo *et al*. (2019). By integrating lessons learned in the state of art-review (**Objective 1**) and a local scale analysis (**Objective 2**), it was possible through **Objective 3 and 4** to deploy an end-to-end supervised deep learning architecture capable of ingesting and processing large volumes of MODIS satellite image time series. The deployment of this particular architecture, locally or in the cloud, at the scale of the Amazon region is a major contribution of this research. Key aspects for the large-scale generation of LC data such as the reduction of traditional pre-processing (i.e. cloud filtering) and postprocessing (i.e. removal of illogical transitions) steps contributed towards a dedicated spatially explicit multi-temporal LC dataset suited to the study of post-loss LC change trajectories in the pantropics. In this regard, most of existing advances commonly cover smaller geographical areas and/or analyse shorter time spans (<5 years). Moreover, different to previous implementations of the MTLCC model using Sentinel data, the use of MODIS allowed testing the capability of this particular deep learning model to predict over longer periods. To reduce illogical transitions, the application of post-classification techniques such as Hidden Markov Models allowed a fast and automate stabilisation of the annual LC maps for their posterior analysis in **Chapter 6**. The application of this post-processing technique differs to other initiatives in the region such as the MapBiomas Amazonia Initiative which relies in experts-based rules. These rules commonly reduce the scalability of the approach to other regions in the tropics.

After a successful calibration of the MTLCC model in the cloud, this research generated consistently 19-years of LC predictions suited to the study of LC change trajectories over almost 2 million deforested pixels between 2004 and 2018 detected by Terra-i across the Amazon (see **Section 6.4.2**). Furthermore, the transfer learning exercises with the calibrated predictive model showed

reasonable LC predictions for other areas in the pantropics, compared visually to reference data (see **Section 5.5.5**). This particular outcome was important as it shows the flexibility of the network to provide long-term LC data where studies in post-loss LC change are also needed.

(2) *To what extent can mapped land cover change trajectories be characterised and explained by socio-environmental conditions and indicate which **land uses** are the proximate causes of deforestation?*

The analysis of high-temporal LC/LU data as generated in **Objective 3 and 4** came with new challenges. These challenges included exploring methods suitable to process high volumes of data, in this case, 2 million deforested pixels with 19-years of LC data. Inspired by the initial results by Mas *et al*. (2019) in LC/LU change research, **Objective 5** showed the feasibility of the sequence analysis (SA) approach for characterising and analysing large volumes of chronological information as analysed in this thesis. SA allowed the definition of minimum period to reasonably characterise different typologies of post-loss LC change trajectories over the study area (see **Section 6.6.1**). Moreover, the observed typologies were explained according to 11 spatial covariate layers representing inaccessibility, biophysical and policy factors (see **Section 6.6.2**). Key factors in determining particular trajectories related to human-driven disturbances were proximity to Farmland, proximity to Urban/Built-up, proximity to exploitation protected areas and elevation (see **Table 6.10**). This thesis partially associated potential land-use agents related to the observed post-loss LC trajectories according to previous work in mapping of spatial patterns of deforestation across the same study area (Coca-Castro, 2015) (see **Table 6.11**). For complete evidence of LC/LU dynamics in deforested areas, this research suggests the exploration of novel techniques capable of providing joint-predictions of LC/LU using satellite imagery as proposed by Zhang *et al*. (2019).

By contributing to answering the above research questions, this PhD project supports modellers in ecosystem services, biodiversity, and other deforestation-relevant topics by going beyond the immediate impacts of deforestation to understanding the typologies and dynamics of long-term LC/LU change trajectories which have impacts on these areas of research. Whilst we have

discussed interactions of trajectories of land cover change with carbon dynamics there are many other ecosystem services to which these are also relevant including pollination, pest control, water quality, quantity and regulation. In these cases, as well as deforestation itself the trajectory of LC/LU change may have an important impact on the demand and supply of these services. Moreover, the spatio-temporal configuration of LC/LU can impact further transitions through for example fire dynamics and suppression by different patterns of cover and use.

References

Abbott, A. (1983) Sequences of Social Events: Concepts and Methods for the Analysis of Order in Social Processes. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 16 (4), 129–147.

Abbott, A. & Forrest, J. (1986) Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History*, 16 (3), 471.

Abercrombie, S. P. & Friedl, M. A. (2016) Improving the Consistency of Multitemporal Land Cover Maps Using a Hidden Markov Model. *IEEE Transactions on Geoscience and Remote Sensing*, 54 (2), 703–713.

Abood, S. A., Lee, J. S. H., Burivalova, Z., Garcia-Ulloa, J. & Koh, L. P. (2015) Relative Contributions of the Logging, Fiber, Oil Palm, and Mining Industries to Forest Loss in Indonesia. *Conservation Letters*, 8 (1), 58–67.

Achard, F., Beuchle, R., Mayaux, P., Stibig, H.-J., Bodart, C., Brink, A., Carboni, S., Desclée, B., Donnay, F., Eva, H. D., Lupi, A., Raši, R., Seliger, R. & Simonetti, D. (2014) Determination of tropical deforestation rates and related carbon losses from 1990 to 2010. *Global Change Biology*, 20 (8), 2540–2554.

Aguiar, A. P. D., Câmara, G. & Escada, M. I. S. (2007) Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. *Ecological Modelling*, 209 (2–4), 169–188.

De Alban, J., Prescott, G., Woods, K., Jamaludin, J., Latt, K., Lim, C., Maung, A. & Webb, E. (2019) Integrating Analytical Frameworks to Investigate Land-Cover Regime Shifts in Dynamic Landscapes. *Sustainability*, 11 (4), 1139.

Allen, J. F. (1984) Towards a general theory of action and time. *Artificial Intelligence*, 23 (2), 123–154.

Almeida, C. A. de, Coutinho, A. C., Esquerdo, J. C. D. M., Adami, M., Venturieri, A., Diniz, C. G., Dessay, N., Durieux, L. & Gomes, A. R. (2016) High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. *Acta Amazonica*, 46 (3), 291–302.

Almeida, L. M. & Ludermir, T. (2008) 'An improved method for automatically searching near-optimal artificial Neural Networks', in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*.

Amaral, S. & D'Alge, J. C. L. (2009) Spatial data availability and its implications for sustainable development of the Brazilian Amazon. *Earth Science Informatics*, 2 (4), 193–203.

Anderegg, J., Hund, A., Karisto, P. & Mikaberidze, A. (2019) In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features. *Frontiers in Plant Science*, 10.

Angelsen, A. & Rudel, T. K. (2013) Designing and Implementing Effective REDD + Policies: A Forest Transition Approach. *Review of Environmental Economics*

and Policy, 7 (1), 91–113.

Antoniou, V., Fonte, C., See, L., Estima, J., Arsanjani, J., Lupia, F., Minghini, M., Foody, G. & Fritz, S. (2016) Investigating the Feasibility of Geo-Tagged Photographs as Sources of Land Cover Input Data. *ISPRS International Journal of Geo-Information*, 5 (5), 64.

Arévalo, P., Olofsson, P. & Woodcock, C. E. (2019) Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series: A test methodology for REDD+ reporting. *Remote Sensing of Environment*, (January 2018), 1–14.

Arima, E. Y. (2016) A Spatial Probit Econometric Model of Land Change: The Case of Infrastructure Development in Western Amazonia, Peru. *PLOS ONE*, 11 (3), e0152058.

Arima, E. Y., Richards, P., Walker, R. & Caldas, M. M. (2011) Statistical confirmation of indirect land use change in the Brazilian Amazon. *Environmental Research Letters*, 6 (2), 024010.

Arima, E. Y., Walker, R. T., Perz, S. G. & Caldas, M. (2005) Loggers and Forest Fragmentation: Behavioral Models of Road Building in the Amazon Basin. *Annals of the Association of American Geographers*, 95 (3), 525–541.

Armenteras, D., Espelta, J. M., Rodríguez, N. & Retana, J. (2017) Deforestation dynamics and drivers in different forest types in Latin America: Three decades of studies (1980–2010). *Global Environmental Change*, 46139–147.

Arvor, D., Jonathan, M., Meirelles, M. S. P., Dubreuil, V. & Durieux, L. (2011) Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*, 32 (22), 7847–7871.

Asner, G. P., Llactayo, W., Tupayachi, R. & Luna, E. R. (2013) Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences*, 110 (46), 18454–18459.

Asner, G. P., Rudel, T. K., Aide, T. M., DeFries, R. & Emerson, R. (2009) A Contemporary Assessment of Change in Humid Tropical Forests. *Conservation Biology*, 23 (6), 1386–1395.

van Asselen, S. & Verburg, P. H. (2012) A Land System representation for global assessments and land-use modeling. *Global Change Biology*, 18 (10), 3125–3148.

Atkinson, P. (1988) 'The Interrelationship Between Resampling Method And Information Extraction Technique', in *International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Moving Toward the 21st Century'.*

Atkinson, P. M. & Tatnall, A. R. L. (1997) Introduction Neural networks in remote sensing. *International Journal of Remote Sensing*, 18 (4), 699–709.

Azzari, G. & Lobell, D. B. (2017) Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. *Remote Sensing of Environment*, 20264–74.

Bahdanau, D., Cho, K. & Bengio, Y. (2015) 'Neural Machine Translation by Jointly Learning to Align and Translate', in *3rd International Conference on Learning Representations*.

Bailly, A., Chapel, L., Tavenard, R. & Camps-Valls, G. (2017) Nonlinear Time-Series Adaptation for Land Cover Classification. *IEEE Geoscience and Remote Sensing Letters*, 14 (6), 896–900.

Ballas, N., Yao, L., Pal, C. & Courville, A. (2016) *Delving Deeper into Convolutional Networks for Learning Video Representations*. Available from: http://arxiv.org/abs/1511.06432.

Baral, P., Wen, Y. & Urriola, N. (2018) Forest Cover Changes and Trajectories in a Typical Middle Mountain Watershed of Western Nepal. *Land*, 7 (2), 72.

Bastin, L., Buchanan, G., Beresford, A., Pekel, J.-F. & Dubois, G. (2013) Open-source mapping and services for Web-based land-cover validation. *Ecological Informatics*, 149–16.

Baumann, M., Ozdogan, M., Richardson, A. D. & Radeloff, V. C. (2017) Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. *International Journal of Applied Earth Observation and Geoinformation*, 5472–83.

Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G. & Dupuy, S. (2018) M3 Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (12), 4939–4949.

Bengio, Y. (2012) 'Deep learning of representations for unsupervised and transfer learning', in *Proceedings of ICML workshop on unsupervised and transfer learning*.

Bengio, Y., Bastien, F., Bergeron, A., Boulanger--Lewandowski, N., Breuel, T., Chherawala, Y., Cisse, M., Côté, M., Erhan, D., Eustache, J. & Others (2011) 'Deep learners benefit more from out-of-distribution examples', in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*.

Bengio, Y., Simard, P. & Frasconi, P. (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5 (2), 157–166.

Bergstra, J. & Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13 (Feb), 281–305.

Bergstra, J. S., Bardenet, R., Bengio, Y. & Kégl, B. (2011) 'Algorithms for Hyper-Parameter Optimization', in *Advances in Neural Information Processing Systems 24*.

Bergstra, J., Yamins, D. & Cox, D. (2013) 'Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms', in *Proceedings of the 12th Python in science conference*.

Bey, A., Díaz, A. S. P., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J. F., Moore, R., Federici, S., Rezende, M., Patriarca, C., Turia, R., Gamoga, G., Abe, H., Kaidong, E. & Miceli, G. (2016) Collect earth: Land use and land cover assessment through augmented visual interpretation. *Remote Sensing*, 8 (10), 807.

Bey, A., Sánchez-Paus Díaz, A., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.-F., Moore, R., Federici, S., Rezende, M., Patriarca, C., Turia, R., Gamoga, G., Abe, H., Kaidong, E. & Miceli, G. (2016) Collect Earth: Land Use and Land Cover Assessment through Augmented Visual Interpretation. *Remote Sensing*, 8 (10), 807.

Bischl, B., Mersmann, O., Trautmann, H. & Weihs, C. (2012) Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. *Evolutionary Computation*, 20 (2), 249–275.

Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*. Clarendon Press.

Bisong, E. (2019) 'Google Cloud Machine Learning Engine (Cloud MLE)', in Ekaba Bisong (ed.) *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*.

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I. & Defourny, P. (2012) Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, 9 (6), 2145–2157.

Bourgoin, C., Oszwald, J., Bourgoin, J., Gond, V., Blanc, L., Dessard, H., Phan, T. Van, Sist, P., Läderach, P. & Reymondin, L. (2020) Assessing the ecological vulnerability of forest landscape to agricultural frontier expansion in the Central Highlands of Vietnam. *International Journal of Applied Earth Observation and Geoinformation*, 84101958.

Brandt, J. (2019) *Spatio-temporal crop classification of low-resolution satellite imagery with capsule layers and distributed attention*. Available from: http://arxiv.org/abs/1904.10130.

Breiman, L. (2001) Random Forests. *Machine Learning*, 45 (1), 5–32.

Brinck, K., Fischer, R., Groeneveld, J., Lehmann, S., Dantas De Paula, M., Pütz, S., Sexton, J. O., Song, D. & Huth, A. (2017) High resolution analysis of tropical forest fragmentation and its impact on the global carbon cycle. *Nature Communications*, 8 (1), 14855.

Brzinsky-Fay, C., Kohler, U. & Luniak, M. (2006) Sequence Analysis with Stata. *The Stata Journal: Promoting communications on statistics and Stata*, 6 (4), 435–460.

Buchhorn, M., Smets, B., Bertels, L., Lesiv, M., Tsendbazar, N.-E., Herold, M. & Fritz, S. (2019) *Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe*. Available from: https://doi.org/10.5281/zenodo.3243509.

Busch, J. & Ferretti-Gallon, K. (2017) What Drives Deforestation and What Stops It?

A Meta-Analysis. *Review of Environmental Economics and Policy*, 11 (1), 3–23.

Cai, S., Liu, D., Sulla-Menashe, D. & Friedl, M. A. (2014) Enhancing MODIS land cover product with a spatial–temporal modeling algorithm. *Remote Sensing of Environment*, 147243–255.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B. & Li, Z. (2018) A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, 21035–47.

Camara, G., Assis, L. F., Ribeiro, G., Ferreira, K. R., Llapa, E. & Vinhas, L. (2016) 'Big earth observation data analytics', in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data - BigSpatial '16*.

Carlson, K. M., Curran, L. M., Ratnasari, D., Pittman, A. M., Soares-Filho, B. S., Asner, G. P., Trigg, S. N., Gaveau, D. A., Lawrence, D. & Rodrigues, H. O. (2012) Committed carbon emissions, deforestation, and community land conversion from oil palm plantation expansion in West Kalimantan, Indonesia. *Proceedings of the National Academy of Sciences*, 109 (19), 7559–7564.

Caye Daudt, R., Le Saux, B., Boulch, A. & Gousseau, Y. (2019) Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, 187102783.

Chadid, M., Dávalos, L., Molina, J. & Armenteras, D. (2015) A Bayesian Spatial Model Highlights Distinct Dynamics in Deforestation from Coca and Pastures in an Andean Biodiversity Hotspot. *Forests*, 6 (12), 3828–3846.

Chazdon, R. L., Brancalion, P. H. S., Laestadius, L., Bennett-Curry, A., Buckingham, K., Kumar, C., Moll-Rocek, J., Vieira, I. C. G. & Wilson, S. J. (2016) When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. *Ambio*, 45 (5), 538–550.

Chazdon, R. L., Letcher, S. G., Van Breugel, M., Martínez-Ramos, M., Bongers, F. & Finegan, B. (2007) Rates of change in tree communities of secondary Neotropical forests following major disturbances. Philosophical Transactions of the Royal Society B: Biological Sciences

Cheng, H., Sinha, A., Cruz, F. W., Wang, X., Edwards, R. L., D'Horta, F. M., Ribas, C. C., Vuille, M., Stott, L. D. & Auler, A. S. (2013) Climate change patterns in Amazonia and biodiversity. *Nature Communications*, 4 (1), 1411.

Chi, M., Sun, Z., Qin, Y., Shen, J. & Benediktsson, J. A. (2017) A Novel Methodology to Label Urban Remote Sensing Images Based on Location-Based Social Media Photos. *Proceedings of the IEEE*, 105 (10), 1926–1936.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. (2014) *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. Available from: http://arxiv.org/abs/1406.1078.

Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K. & Bengio, Y. (2015) 'Attention-

Based Models for Speech Recognition', in *Advances in Neural Information Processing Systems 28*.

Coca-Castro, A. (2015) *Mapping deforestation stages and spatial patterns in the Amazon rainforest using fractal analysis and data mining techniques*. King's College London. [Online Thesis] http://babel.banrepcultural.org/cdm/ref/collection/p17054coll23/id/1173

Coca-Castro, A., Ruswurm, M., Reymondin, L. & Mulligan, M. (2019) 'Sequential Recurrent Encoders for Land Cover Mapping in The Brazilian Amazon Using Modis Imagery and Auxiliary Datasets', in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*.

Coca-Castro, A., Tello-Dagua, J., Reymondin, L., Paz-Garcia, P. & Hyman, G. (2014) *Reporte de la validación en campo del Sistema Terra-i (CIAT) en las comunidades del proyecto ASSETS en la región de Ucayali, Perú*. Available from: http://www.terra-i.org/dam/jcr:5aa36ebd-f93b-40d1-b334-1c849ba6f374/FinalReport-ValidationTerrai_ASSETScommunities_SP.pdf.

Coe, M. T., Costa, M. H. & Soares-Filho, B. S. (2009) The influence of historical and potential future deforestation on the stream flow of the Amazon River – Land surface processes and atmospheric feedbacks. *Journal of Hydrology*, 369 (1–2), 165–174.

Comber, A. J. (2008) The separation of land cover from land use using data primitives. *Journal of Land Use Science*, 3 (4), 215–229.

Core Team, R. & Others (2013) R: a language and environment for statistical computing. *R Foundation for statistical computing, Vienna*,

Cortes, C. & Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20 (3), 273–297.

Crespo Cuaresma, J., Danylo, O., Fritz, S., McCallum, I., Obersteiner, M., See, L. & Walsh, B. (2017) Economic Development and Forest Cover: Evidence from Satellite Data. *Scientific Reports*, 7 (1), 40678.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A. & Hansen, M. C. (2018) Classifying drivers of global forest loss. *Science*, 361 (6407), 1108–1111.

Darak, S., Mills, M., Kulkarni, V., Kulkarni, S., Hutter, I. & Janssen, F. (2015) Trajectories of Childbearing among HIV Infected Indian Women: A Sequence Analysis Approach. *PLOS ONE*, 10 (4), e0124537.

DeFries, R. S., Rudel, T., Uriarte, M. & Hansen, M. (2010) Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*, 3 (3), 178–181.

DeFries, R. S. & Townshend, J. R. G. (1999) Global land cover characterization from satellite data: from research to operational implementation? *Global Ecology and Biogeography*, 8 (5), 367–379.

Dirzo, R. & Raven, P. H. (2003) Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, 28 (1), 137–167.

Dozat, T. (2016) *Incorporating Nesterov Momentum into Adam*. Available from: http://cs229.stanford.edu/proj2015/054_report.pdf.

Drummond, M. A. & Loveland, T. R. (2010) Land-use Pressure and a Transition to Forest-cover Loss in the Eastern United States. *Bioscience*, 60 (4), 286–298.

Durrant, G. B., Maslovskaya, O. & Smith, P. W. F. (2019) Investigating call record data using sequence analysis to inform adaptive survey designs. *International Journal of Social Research Methodology*, 22 (1), 37–54.

Eckner, A. (2012) A framework for the analysis of unevenly spaced time series data. Preprint. Available at: http://www. eckner. com/papers/unevenly_spaced_time_series_analysis. pdf Available from: http://eckner.com/papers/unevenly_spaced_time_series_analysis.pdf.

Edwards, D. P., Sloan, S., Weng, L., Dirks, P., Sayer, J. & Laurance, W. F. (2014) Mining and the African Environment. *Conservation Letters*, 7 (3), 302–311.

Eggensperger, K., Hutter, F., Hoos, H. H. & Leyton-Brown, K. (2014) 'Surrogate Benchmarks for Hyperparameter Optimization', in *MetaSel 2014: Meta-learning and Algorithm Selection*.

Elder, G. H., Johnson, M. K. & Crosnoe, R. (2003) 'The Emergence and Development of Life Course Theory', in Jeylan T Mortimer & Michael J Shanahan (eds.) *Handbook of the Life Course*.

ESA (2017) *Land Cover CCI Product User Guide Version 2.0*. Available from: https://www.esa-landcover-cci.org/?q=webfm_send/84.

Eva, H. D., Huber, O., Achard, F., Balslev, H., Beck, S. & others (2005) *A proposal for defining the geographical boundaries of Amazonia; synthesis of the results from an expert consultation workshop organized by the European Commission in collaboration with the Amazon Cooperation Treaty Organization - JRC Ispra, 7-8 June 2005*. Available from: http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/23658/1/eur 21808 en.pdf.

Eva, H. & Lambin, E. F. (2000) Fires and land-cover change in the tropics:a remote sensing analysis at the landscape scale. *Journal of Biogeography*, 27 (3), 765–776.

Ewers, R. M. & Laurance, W. F. (2006) Scale-dependent patterns of deforestation in the Brazilian Amazon. *Environmental Conservation*, 33 (3), 203–211.

Fagan, M., DeFries, R., Sesnie, S., Arroyo-Mora, J., Soto, C., Singh, A., Townsend, P. & Chazdon, R. (2015) Mapping Species Composition of Forests and Tree Plantations in Northeastern Costa Rica with an Integration of Hyperspectral and Multitemporal Landsat Imagery. *Remote Sensing*, 7 (5), 5660–5696.

Fagua, J. C. & Ramsey, R. D. (2019) Geospatial modeling of land cover change in the Chocó-Darien global ecoregion of South America; One of most biodiverse and rainy areas in the world. *PLOS ONE*, 14 (2), e0211324.

FAO (2000) *FRA 2000 on definitions of forest and forest change*. (Working Paper

Series 33). Available from: http://www.fao.org/3/ad665e/ad665e04.htm.

Feurer, M. & Hutter, F. (2019) 'Hyperparameter Optimization', in Frank Hutter, Lars Kotthoff, & Joaquin Vanschoren (eds.) *Automated Machine Learning: Methods, Systems, Challenges*.

Feurer, M., Letham, B. & Bakshy, E. (2018) *Scalable Meta-Learning for Bayesian Optimization*. Available from: http://arxiv.org/abs/1802.02219.

Feurer, M., Springenberg, J. T. & Hutter, F. (2015) 'Initializing bayesian hyperparameter optimization via meta-learning', in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Fick, S. E. & Hijmans, R. J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37 (12), 4302–4315.

de Filho, F. J. B. O. & Metzger, J. P. (2006) Thresholds in landscape structure for three common deforestation patterns in the Brazilian Amazon. *Landscape Ecology*, 21 (7), 1061–1073.

Finer, M. & Novoa, S. (2016) *Large-scale vs. Small-scale Deforestation in the Peruvian Amazon. MAAP: 32.* Available from: https://maaproject.org/2016/scale/.

Fornacca, D., Ren, G. & Xiao, W. (2017) Performance of Three MODIS Fire Products (MCD45A1, MCD64A1, MCD14ML), and ESA Fire_CCI in a Mountainous Area of Northwest Yunnan, China, Characterized by Frequent Small Fires. *Remote Sensing*, 9 (11), 1131.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A. & Huang, X. (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114 (1), 168–182.

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., Kraxner, F. & Obersteiner, M. (2009) Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, 1 (3), 345–354.

Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D., Duerauer, M., Karner, M., Dresel, C., Laso-Bayas, J.-C., Lesiv, M., Moorthy, I., Salk, C. F., Danylo, O., Sturn, T., Albrecht, F., You, L., Kraxner, F. & Obersteiner, M. (2017) A global dataset of crowdsourced land cover and land use reference data. *Scientific Data*, 4 (1), 170075.

Fritz, S., See, L., van der Velde, M., Nalepa, R. A., Perger, C., Schill, C., McCallum, I., Schepaschenko, D., Kraxner, F., Cai, X., Zhang, X., Ortner, S., Hazarika, R., Cipriani, A., Di Bella, C., Rabia, A. H., Garcia, A., Vakolyuk, M., Singha, K., Beget, M. E., Erasmi, S., Albrecht, F., Shaw, B. & Obersteiner, M. (2013) Downgrading Recent Estimates of Land Available for Biofuel Production. *Environmental Science & Technology*, 47 (3), 130128103203003.

Fujisaka, S. & White, D. (1998) Pasture or permanent crops after slash-and-burn cultivation? Land-use choice in three Amazon colonies. *Agroforestry Systems*

volume, 42 (1), 45–59.

Fulcher, B. D. (2018) Feature-based time-series analysis. Feature Engineering for Machine Learning and Data Analytics Available from: https://arxiv.org/abs/1709.08055.

van der Gaast, W., Sikkema, R. & Vohrer, M. (2018) The contribution of forest carbon credit projects to addressing the climate change challenge. *Climate Policy*, 18 (1), 42–48.

Gabadinho, A., Ritschard, G., Müller, N. S. & Studer, M. (2011) Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40 (4), 1–37.

Gakhov, A. (2015) *Recurrent Neural Networks. Part 1: Theory*. Available from: https://www.slideshare.net/gakhov/recurrent-neural-networks-part-1-theory.

Galford, G. L., Melillo, J., Mustard, J. F., Cerri, C. E. P. & Cerri, C. C. (2010) The Amazon Frontier of Land-Use Change: Croplands and Consequences for Greenhouse Gas Emissions. *Earth Interactions*, 14 (15), 1–24.

Garnot, V. S. F., Landrieu, L., Giordano, S. & Chehata, N. (2019a) *Satellite Image Time Series Classification with Pixel-Set Encoders and Temporal Self-Attention*. Available from: http://arxiv.org/abs/1911.07757.

Garnot, V. S. F., Landrieu, L., Giordano, S. & Chehata, N. (2019b) *Time-Space tradeoff in deep learning models for crop classification on satellite multi-spectral image time series*. Available from: http://arxiv.org/abs/1901.10503.

GBDX (2019) *Datasets on GBDX*. Available from: https://gbdxdocs.digitalglobe.com/docs/datasets-on-gbdx.

Ge, Y., Jin, Y., Stein, A., Chen, Y., Wang, Jianghao, Wang, Jinfeng, Cheng, Q., Bai, H., Liu, M. & Atkinson, P. M. (2019) Principles and methods of scaling geospatial Earth science data. *Earth-Science Reviews*, 197102897.

Geist, H. J. & Lambin, E. F. (2002) Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *Bioscience*, 52 (2), 143.

Golovin, D., Solnik, B., Moitra, S., Kochanski, G., Karro, J. & Sculley, D. (2017) 'Google vizier: A service for black-box optimization', in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Gómez, C., White, J. C. & Wulder, M. A. (2016) Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 11655–72.

Goodfellow, I., Bengio, Y. & Courville, A. (2016) *Deep Learning*. MIT Press.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. & Moore, R. (2017) Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 20218–27.

Graf, A. B. A., Smola, A. J. & Borer, S. (2003) Classification in a normalized feature

space using support vector machines. *IEEE Transactions on Neural Networks*, 14 (3), 597–605.

Grau, H. R., Gasparri, N. I. & Aide, T. M. (2005) Agriculture expansion and deforestation in seasonally dry forests of north-west Argentina. *Environmental Conservation*, 32 (2), 140–148.

Graves, A., Mohamed, A. & Hinton, G. (2013) 'Speech recognition with deep recurrent neural networks', in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*.

Gullison, R. E. & Hardner, J. (2018) Progress and challenges in consolidating the management of Amazonian protected areas and indigenous territories. *Conservation Biology*, 32 (5), 1020–1030.

Gutierrez-Arellano, C. & Mulligan, M. (2018) A review of regulation ecosystem services and disservices from faunal populations and potential impacts of agriculturalisation on their provision, globally. *Nature Conservation*, 301–39.

Gutiérrez-Vélez, V. H. & DeFries, R. (2013) Annual multi-resolution detection of land cover conversion to oil palm in the Peruvian Amazon. *Remote Sensing of Environment*, 129154–167.

Guyon, I. & Elisseeff, A. (2003) An Introduction to Variable and Feature Selection. *The Journal of Machine Learning Research*, 3 (Mar), 1157–1182.

Halpin, B. (2017) SADI: Sequence Analysis Tools for Stata. *The Stata Journal: Promoting communications on statistics and Stata*, 17 (3), 546–572.

Hamida, A. B., Benoit, A., Lambert, P. & Amar, C. B. (2018) 3-D Deep Learning Approach for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56 (8), 4420–4434.

Hammer, D., Kraft, R. & Wheeler, D. (2009) *FORMA: Forest Monitoring for Action--Rapid Identification of Pan-tropical Deforestation Using Moderate-Resolution Remotely Sensed Data - Working Paper 192*. Available from: https://www.cgdev.org/publication/forma-forest-monitoring-action-rapid-identification-pan-tropical-deforestation-using.

Han, J., Kamber, M. & Pei, J. (2011) *Data mining concepts and techniques third edition*.

Han, Y., Liefbroer, A. & Elzinga, C. (2017) Comparing methods of classifying life courses: sequence analysis and latent class analysis. *Longitudinal and Life Course Studies*, 8 (4), 319–341.

Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V, Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F. & Moore, R. (2016) Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, 11 (3), 034008.

Hansen, M. C. & Loveland, T. R. (2012) A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 12266–74.

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V, Goetz, S. J., Loveland, T. R., Kommareddy, A.,

Egorov, A., Chini, L., Justice, C. O. & Townshend, J. R. G. (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342 (6160), 850–853.

Hargita, Y., Giessen, L. & Günter, S. (2020) Similarities and Differences between International REDD+ and Transnational Deforestation-Free Supply Chain Initiatives – A Review. *Sustainability*, 12 (3), 896.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Science & Business Media.

Helber, P., Gram-Hansen, B., Varatharajan, I., Azam, F., Coca-Castro, A., Kopackova, V. & Bilinski, P. (2018) *Mapping Informal Settlements in Developing Countries with Multi-resolution, Multi-spectral Data*. Available from: http://arxiv.org/abs/1812.00812.

Hennig, C. (2007) Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, 52 (1), 258–271.

Hepner, G., Logan, T., Ritter, N. & Bryant, N. (1990) Artificial neural network classification using a minimal training set- Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, 56 (4), 469–473.

Hermosilla, T., Wulder, M. A., White, J. C., Coops, N. C. & Hobart, G. W. (2018) Disturbance-Informed Annual Land Cover Classification Maps of Canada's Forested Ecosystems for a 29-Year Landsat Time Series. *Canadian Journal of Remote Sensing*, 44 (1), 67–87.

Hochreiter, S. & Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, 9 (8), 1735–1780.

Hoen, H. F., Roos, A., Lönnstedt, L., Nord, T., Gong, P. & Stendahl, M. (2014) 'The forest-based sector and research--some thoughts on relevance and future prospects', in *Scandinavian Forest Economics: Proceedings of the Biennial Meeting of the Scandinavian Society of Forest Economics*.

Hoffman, M., Shahriari, B. & Freitas, N. (2014) 'On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning', in *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Hosmer Jr., D. W., Lemeshow, S. & Sturdivant, R. X. (2013) *Applied Logistic Regression*. John Wiley & Sons.

Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A. & Romijn, E. (2012) An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters*, 7 (4), 044009.

Hu, J., Ghamisi, P. & Zhu, X. (2018) Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification. *ISPRS*

International Journal of Geo-Information, 7 (9), 379.

Huang, W., Xiao, L., Wei, Z., Liu, H. & Tang, S. (2015) A New Pan-Sharpening Method With Deep Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 12 (5), 1037–1041.

Huang, X. & Friedl, M. A. (2014) Distance metric-based forest cover change detection using MODIS time series. *International Journal of Applied Earth Observation and Geoinformation*,

Hubert, L. J. & Levin, J. R. (1976) A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin*, 83 (6), 1072–1080.

Hughes, L., Schmitt, M. & Zhu, X. (2018) Mining Hard Negative Samples for SAR-Optical Image Matching Using Generative Adversarial Networks. *Remote Sensing*, 10 (10), 1552.

Hutter, F., Hoos, H. H. & Leyton-Brown, K. (2011) 'Sequential Model-Based Optimization for General Algorithm Configuration', in *Lecture Notes in Computer Science*.

Hyman, G. & Barona, E. (2010) *Roads and deforestation in the Central Peruvian Amazon*. Available from: https://hdl.handle.net/10568/67279.

Ienco, D., Gaetano, R., Dupaquier, C. & Maurel, P. (2017) Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, 14 (10), 1685–1689.

Ienco, D., Interdonato, R. & Gaetano, R. (2019) *Supervised level-wise pretraining for recurrent neural network initialization in multi-class classification*. Available from: http://arxiv.org/abs/1911.01071.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D. & Rodes, I. (2017) Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series. *Remote Sensing*, 9 (1), 95.

Interdonato, R., Ienco, D., Gaetano, R. & Ose, K. (2019) DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn. *ISPRS J. Photogramm. Remote Sens.*, 14991–104.

Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. (2008) *Hole-filled SRTM for the globe Version 4*. Available from: http:/srtm.csi.cgiar.org.

Jeune, W., Francelino, M. R., Souza, E. de, Fernandes Filho, E. I. & Rocha, G. C. (2018) Multinomial Logistic Regression and Random Forest Classifiers in Digital Mapping of Soil Classes in Western Haiti. *Revista Brasileira de Ciência do Solo*, 42.

Ji, S., Zhang, C., Xu, A., Shi, Y. & Duan, Y. (2018) 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. *Remote Sensing*, 10 (2), 75.

Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B. & Wang, X. (2014) Land Cover Classification of Landsat Data with Phenological Features Extracted from Time

Series MODIS NDVI Data. *Remote Sensing*, 6 (11), 11518–11532.

Jia, X., Khandelwal, A., Nayak, G., Gerber, J., Carlson, K., West, P. & Kumar, V. (2017) 'Predict Land Covers with Transition Modeling and Incremental Learning', in *Proceedings of the 17th SIAM International Conference on Data Mining*. Proceedings.

Jin, Y.-Q. & Liu, C. (1997) Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks. *International Journal of Remote Sensing*, 18 (4), 971–979.

Johnson, J. M. & Khoshgoftaar, T. M. (2019) Survey on deep learning with class imbalance. *Journal of Big Data*, 6 (1), 27.

Jönsson, P. & Eklundh, L. (2004) TIMESAT – a program for analyzing time-series of satellite sensor data. *Computers & Geosciences*, 30 (8), 833–845.

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C. & Waske, B. (2016) A Review of the Application of Optical and Radar Remote Sensing Data Fusion to Land Use Mapping and Monitoring. *Remote Sensing*, 8 (1), 70.

Jurjonas, M., Crossman, K., Solomon, J. & Baez, W. L. (2016) Potential Links Between Certified Organic Coffee and Deforestation in a Protected Area in Chiapas, Mexico. *World Development*, 7813–21.

Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J. T., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F. & Kaufman, Y. (2002) The MODIS fire products. *Remote Sensing of Environment*, 83 (1–2), 244–262.

Justice, C. O., Townshend, J. R. G., Vermote, E. F., Masuoka, E., Wolfe, R. E., Saleous, N., Roy, D. P. & Morisette, J. T. (2002) An overview of MODIS Land data processing and product status. *Remote Sensing of Environment*, 83 (1–2), 3–15.

Jusys, T. (2018) Changing patterns in deforestation avoidance by different protection types in the Brazilian Amazon Asim Zia (ed.). *PLOS ONE*, 13 (4), e0195900.

Kalamandeen, M., Gloor, E., Mitchard, E., Quincey, D., Ziv, G., Spracklen, D., Spracklen, B., Adami, M., Aragão, L. E. O. C. & Galbraith, D. (2018) Pervasive Rise of Small-scale Deforestation in Amazonia. *Scientific Reports*, 8 (1), 1600.

Kamilaris, A. & Prenafeta-Boldú, F. X. (2018) Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 14770–90.

Katakami, S., Sakamoto, H. & Okada, M. (2019) Bayesian Hyperparameter Estimation using Gaussian Process and Bayesian Optimization. *Journal of the Physical Society of Japan*, 88 (7), 074001.

Kennedy, C. M., Lonsdorf, E., Neel, M. C., Williams, N. M., Ricketts, T. H., Winfree, R., Bommarco, R., Brittain, C., Burley, A. L., Cariveau, D., Carvalheiro, L. G., Chacoff, N. P., Cunningham, S. A., Danforth, B. N., Dudenhöffer, J.-H., Elle, E., Gaines, H. R., Garibaldi, L. A., Gratton, C., Holzschuh, A., Isaacs, R., Javorek, S. K., Jha, S., Klein, A. M., Krewenka, K., Mandelik, Y., Mayfield, M. M., Morandin, L., Neame, L.

A., Otieno, M., Park, M., Potts, S. G., Rundlöf, M., Saez, A., Steffan-Dewenter, I., Taki, H., Viana, B. F., Westphal, C., Wilson, J. K., Greenleaf, S. S. & Kremen, C. (2013) A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. *Ecology Letters*, 16 (5), 584–599.

Khatami, R., Mountrakis, G. & Stehman, S. V. (2016) A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, 17789–100.

Kingma, D. P. & Ba, J. (2015) 'Adam: A Method for Stochastic Optimization', in *3rd International Conference on Learning Representations*.

Knaus, J. (2010) snowfall: Easier cluster computing (based on snow). *R package version*, 1.

Kobayashi, S., Soudre, M. & Ricse, A. (2014) Secondary succession of mixed plantations established to rehabilitate abandoned pasture in the Peruvian Amazon. *Tropics*, 23 (2), 83–90.

Kochanski, G., Golovin, D., Karro, J., Solnik, B., Moitra, S. & Sculley, D. (2017) 'Bayesian optimization for a better dessert', in *NIPS, workshop on Bayesian optimization*.

Krenker, A., Bester, J. & Kos, A. (2011) 'Introduction to the Artificial Neural Networks', in *Artificial Neural Networks - Methodological Advances and Biomedical Applications*.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017) ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60 (6), 84–90.

Krylov, A., Steininger, M. K., Hansen, M. C., Potapov, P. V, Stehman, S. V, Gost, A., Noel, J., Talero Ramirez, Y., Tyukavina, A., Di Bella, C. M., Ellis, E. A. & Ellis, P. (2018) Contrasting tree-cover loss and subsequent land cover in two neotropical forest regions: sample-based assessment of the Mexican Yucatán and Argentine Chaco. *Journal of Land Use Science*, 13 (6), 549–564.

Kuhn, M. (2008) Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28 (5), 1–26.

Kuhn, M. & Johnson, K. (2013) *Applied Predictive Modeling*. Springer, New York, NY.

Kumar, U., Ganguly, S., Nemani, R. R., Raja, K. S., Milesi, C., Sinha, R., Michaelis, A., Votava, P., Hashimoto, H., Li, S., Wang, W., Kalia, S. & Gayaka, S. (2017) Exploring subpixel learning algorithms for estimating global land cover fractions from satellite data using high performance computing. *Remote Sensing*,

Kyere, I., Astor, T., Graß, R. & Wachendorf, M. (2019) Multi-Temporal Agricultural Land-Cover Mapping Using Single-Year and Multi-Year Models Based on Landsat Imagery and IACS Data. *Agronomy*, 9 (6), 309.

Labán, R. & Sturzenegger, F. (1994) Fiscal conservatism as a response to the debt crisis. *Journal of Development Economics*, 45 (2), 305–324.

Lambin, E. F., Geist, H. J. & Lepers, E. (2003) Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources*, 28 (1), 205–241.

Lambin, E. F., Rounsevell, M. D. A. & Geist, H. J. (2000) Are agricultural land-use models able to predict changes in land-use intensity? *Agriculture, Ecosystems & Environment*, 82 (1–3), 321–331.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., Coomes, O. T., Dirzo, R., Fischer, G., Folke, C., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., Skånes, H., Steffen, W., Stone, G. D., Svedin, U., Veldkamp, T. A., Vogel, C. & Xu, J. (2001) The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11 (4), 261–269.

Laso Bayas, J. C., Lesiv, M., Waldner, F., Schucknecht, A., Duerauer, M., See, L., Fritz, S., Fraisl, D., Moorthy, I., McCallum, I., Perger, C., Danylo, O., Defourny, P., Gallego, J., Gilliams, S., Akhtar, I. ul H., Baishya, S. J., Baruah, M., Bungnamei, K., Campos, A., Changkakati, T., Cipriani, A., Das, Krishna, Das, Keemee, Das, I., Davis, K. F., Hazarika, P., Johnson, B. A., Malek, Z., Molinari, M. E., Panging, K., Pawe, C. K., Pérez-Hoyos, A., Sahariah, P. K., Sahariah, D., Saikia, A., Saikia, M., Schlesinger, P., Seidacaru, E., Singha, K. & Wilson, J. W. (2017) A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform. *Scientific Data*, 4 (1), 170136.

Laurance, W. F., Clements, G. R., Sloan, S., O'Connell, C. S., Mueller, N. D., Goosem, M., Venter, O., Edwards, D. P., Phalan, B., Balmford, A., Van Der Ree, R. & Arrea, I. B. (2014) A global strategy for road building. *Nature*, 513 (7517), 229–232.

Laurent, C., Pereyra, G., Brakel, P., Zhang, Y. & Bengio, Y. (2016) 'Batch normalized recurrent neural networks', in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Leberger, R., Rosa, I. M. D., Guerra, C. A., Wolf, F. & Pereira, H. M. (2020) Global patterns of forest loss across IUCN categories of protected areas. *Biological Conservation*, 241108299.

LeCun, Y., Bengio, Y. & Hinton, G. (2015) Deep learning. *Nature*, 521 (7553), 436–444.

Lee, J., Weger, R. C., Sengupta, S. K. & Welch, R. M. (1990) A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, 28 (5), 846–855.

Lesiv, M., See, L., Laso Bayas, J., Sturn, T., Schepaschenko, D., Karner, M., Moorthy, I., McCallum, I. & Fritz, S. (2018) Characterizing the Spatial and Temporal Availability of Very High Resolution Satellite Imagery in Google Earth and Microsoft Bing Maps as a Source of Reference Data. *Land*, 7 (4), 118.

Levenshtein, V. I. (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10 (8), 707–710.

Levitt, M. (1969) Detailed Molecular Model for Transfer Ribonucleic Acid. *Nature*, 224 (5221), 759–763.

Li, C., Gong, P., Wang, J., Zhu, Z., Biging, G. S., Yuan, C., Hu, T., Zhang, H., Wang, Q., Li, X., Liu, X., Xu, Y., Guo, J., Liu, C., Hackman, K. O., Zhang, M., Cheng, Y., Yu, L., Yang, J., Huang, H. & Clinton, N. (2017) The first all-season sample set for mapping global land cover with Landsat-8 data. *Science Bulletin*, 62 (7), 508–515.

Li, H., Zhang, C., Zhang, S. & Atkinson, P. (2019) A hybrid OSVM-OCNN Method for Crop Classification from Fine Spatial Resolution Remotely Sensed Imagery. *Remote Sensing*, 11 (20), 2370.

Li, M., Zang, S., Zhang, B., Li, S. & Wu, C. (2014) A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information. *European Journal of Remote Sensing*, 47 (1), 389–411.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A. & Peng, S. (2018) Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015). *Earth System Science Data*, 10 (1), 219–234.

Li, Y., Zhang, H. & Shen, Q. (2017) Spectral–Spatial Classification of Hyperspectral Imagery with 3D Convolutional Neural Network. *Remote Sensing*, 9 (1), 67.

Liang, H. & Li, Q. (2016) Hyperspectral Imagery Classification Using Sparse Representations of Convolutional Neural Network Features. *Remote Sensing*, 8 (2), 99.

Lieth, H. & Werger, M. J. A. (2012) *Tropical Rain Forest Ecosystems: Biogeographical and Ecological Studies*. Elsevier.

Lima, A., Silva, T. S. F., Aragão, L. E. O. e C. de, Feitas, R. M. de, Adami, M., Formaggio, A. R. & Shimabukuro, Y. E. (2012) Land use and land cover changes determine the spatial relationship between fire and deforestation in the Brazilian Amazon. *Applied Geography*, 34239–246.

Lima, M. G. B., Ashely-Cantello, W., Visseren-Hamakers, I. J., Gupta, A. & Braña, J. (2015) *Forests Post-2015: Maximizing Synergies between the Sustainable Development Goals and REDD+*. Available from: https://bit.ly/2QTsxbR.

Liping, C., Yujun, S. & Saeed, S. (2018) Monitoring and predicting land use and land cover changes using remote sensing and GIS techniques – A case study of a hilly area, Jiangle, China Andreas Westergaard-Nielsen (ed.). *PLOS ONE*, 13 (7), e0200493.

Loveland, T. R. & Belward, A. S. (1997) The IGBP-DIS global 1km land cover data set, DISCover: First results. *International Journal of Remote Sensing*, 18 (15), 3289–3295.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L. & Merchant, J. W. (2000) Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing*, 21 (6–7), 1303–1330.

Löw, F., Michel, U., Dech, S. & Conrad, C. (2013) Impact of feature selection on the

accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, 85102–119.

Lustgarten, J. L., Gopalakrishnan, V., Grover, H. & Visweswaran, S. (2008) 'Improving classification performance with discretization on biomedical datasets', in *AMIA Symposium Proocedings*.

Lustgarten, J. L., Visweswaran, S., Gopalakrishnan, V. & Cooper, G. F. (2011) Application of an efficient Bayesian discretization method to biomedical data. *BMC Bioinformatics*, 12 (1), 309.

Lyons, M. B., Keith, D. A., Phinn, S. R., Mason, T. J. & Elith, J. (2018) A comparison of resampling methods for remote sensing classification and accuracy assessment. *Remote Sensing of Environment*, 208145–153.

Lyu, H., Lu, H., Mou, L., Li, W., Wright, J., Li, Xuecao, Li, Xinlu, Zhu, X., Wang, J., Yu, L. & Gong, P. (2018) Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data. *Remote Sensing*, 10 (3), 471.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. & Johnson, B. A. (2019) Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152166–177.

MacDicken, K. G. (2015) Global Forest Resources Assessment 2015: What, why and how? *Forest Ecology and Management*, 3523–8.

Macedo, M. N., DeFries, R. S., Morton, D. C., Stickler, C. M., Galford, G. L. & Shimabukuro, Y. E. (2012) Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. *Proceedings of the National Academy of Sciences*, 109 (4), 1341–1346.

Maciel, A. M. (2017) *Spatiotemporal interval logic for reasoning about land use change dynamics*. Instituto Nacional de Pesquisas Espaciais (INPE). [Online Thesis] http://urlib.net/sid.inpe.br/mtc-m21b/2017/11.23.13.04

Maciel, A. M., Camara, G., Vinhas, L., Picoli, M. C. A., Begotti, R. A. & Assis, L. F. F. G. de (2019) A spatiotemporal calculus for reasoning about land-use trajectories. *International Journal of Geographical Information Science*, 33 (1), 176–192.

MacLachlan, A., Roberts, G., Biggs, E. & Boruff, B. (2017) Subpixel land-cover classification for improved urban area estimates using Landsat. *International Journal of Remote Sensing*, 38 (20), 5763–5792.

Madrigal, F., Maurice, C. & Lerasle, F. (2019) Hyper-parameter optimization tools comparison for multiple object tracking applications. *Machine Vision and Applications*, 30 (2), 269–289.

Man, C. D., Nguyen, T. T., Bui, H. Q., Lasko, K. & Nguyen, T. N. T. (2018) Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers. *International Journal of Remote Sensing*, 39 (4), 1243–1255.

MapBiomas Project (2019) MapBiomas Project. Collection V4 of Brazilian Land Cover & Use Map Series Available from: https://mapbiomas.org/.

Marcos, D., Volpi, M., Kellenberger, B. & Tuia, D. (2018) Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 14596–107.

Marquardt, K., Pain, A., Bartholdson, Ö. & Rengifo, L. R. (2019) Forest Dynamics in the Peruvian Amazon: Understanding Processes of Change. *Small-scale Forestry*, 18 (1), 81–104.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J. & Ghalsasi, A. (2011) Cloud computing — The business perspective. *Decision Support Systems*, 51 (1), 176–189.

Mas, J.-F., Nogueira de Vasconcelos, R. & Franca-Rocha, W. (2019) Analysis of High Temporal Resolution Land Use/Land Cover Trajectories. *Land*, 8 (2), 30.

Mather, A. S. (1992) The Forest Transition. *Area*, 24 (4), 367–379.

Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F. M. & de Queiroz, G. R. (2016) A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9 (8), 3729–3739.

McNicol, I. M., Ryan, C. M. & Mitchard, E. T. A. (2018) Carbon losses from deforestation and widespread degradation offset by extensive growth in African woodlands. *Nature Communications*, 9 (1), 3045.

Meijer, K. S. (2015) A Comparative Analysis of the Effectiveness of Four Supply Chain Initiatives to Reduce Deforestation. *Tropical Conservation Science*, 8 (2), 583–597.

Menton, M. & Cronkleton, P. (2019) *Migration and forests in the Peruvian Amazon: A review*. [Online Thesis] https://www.cifor.org/library/7305. Center for International Forestry Research (CIFOR).

Merkle, N., Auer, S., Muller, R. & Reinartz, P. (2018) Exploring the Potential of Conditional Adversarial Networks for Optical and SAR Image Matching. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (6), 1811–1820.

Mertens, B. (2002) Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: the case of São Félix do Xingú in South Pará. *Agricultural Economics*, 27 (3), 269–294.

Meyfroidt, P. (2016) Approaches and terminology for causal analysis in land systems science. *Journal of Land Use Science*, 11 (5), 501–522.

Meyfroidt, P., Carlson, K. M., Fagan, M. E., Gutiérrez-Vélez, V. H., Macedo, M. N., Curran, L. M., DeFries, R. S., Dyer, G. A., Gibbs, H. K., Lambin, E. F., Morton, D. C. & Robiglio, V. (2014) Multiple pathways of commodity crop expansion in tropical forest landscapes. *Environmental Research Letters*, 9 (7), 074012.

Meyfroidt, P. & Lambin, E. F. (2011) Global Forest Transition: Prospects for an End to Deforestation. *Annual Review of Environment and Resources*, 36 (1), 343–371.

Meyfroidt, P., Lambin, E. F., Erb, K.-H. & Hertel, T. W. (2013) Globalization of land use: distant drivers of land change and geographic displacement of land use. *Current Opinion in Environmental Sustainability*, 5 (5), 438–444.

Milligan, G. W. & Cooper, M. C. (1985) An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50 (2), 159–179.

Minh, D. H. T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F. & Maurel, P. (2018) Deep Recurrent Neural Networks for Winter Vegetation Quality Mapping via Multitemporal SAR Sentinel-1. *IEEE Geoscience and Remote Sensing Letters*, 15 (3), 464–468.

Mirkatouli, J., Hosseini, A. & Neshat, A. (2015) Analysis of land use and land cover spatial pattern based on Markov chains modelling. *City, Territory and Architecture*, 2 (1), 4.

Mitchard, E. (2016) *A Review of Earth Observation Methods for Detecting and Measuring Forest Change in the Tropics*. Available from: https://www.geos.ed.ac.uk/homes/emitchar/Mitchard_EO_Methods_Forests.pdf.

Močkus, J. (1975) 'On bayesian methods for seeking the extremum', in *Optimization Techniques IFIP Technical Conference*.

MOCUPP (2017) *Monitoring land use change within production landscapes*. Available from: http://mocupp.org/sites/default/files/documento-mocupp-en.pdf.

Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Arai, E., del Bon Espirito-Santo, F., Freitas, R. & Morisette, J. (2006) Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 103 (39), 14637–14641.

Mou, L., Ghamisi, P. & Zhu, X. X. (2017) Deep Recurrent Neural Networks for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (7), 3639–3655.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, 403 (6772), 853–858.

Nallapati, R., Zhou, B., dos Santos, C., Gulcehre, C. & Xiang, B. (2016) 'Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond', in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*.

Ndikumana, E., Minh, D. H. T., Baghdadi, N., Courault, D. & Hossard, L. (2018) Deep Recurrent Neural Network for Agricultural Classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10 (8), 1217.

Nepstad, D. C., Stickler, C. M., Filho, B. S.- & Merry, F. (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 (1498), 1737–1746.

Numata, I., Cochrane, M. A., Souza Jr, C. M. & Sales, M. H. (2011) Carbon emissions from deforestation and forest fragmentation in the Brazilian Amazon. *Environmental Research Letters*, 6 (4), 044003.

O'Reilly, C., Moessner, K. & Nati, M. (2017) Univariate and Multivariate Time Series Manifold Learning. *Knowledge-Based Systems*, 1331–16.

Oliveira, P. J. C., Asner, G. P., Knapp, D. E., Almeyda, A., Galvan-Gildemeister, R., Keene, S., Raybin, R. F. & Smith, R. C. (2007) Land-Use Allocation Protects the Peruvian Amazon. *Science*, 317 (5842), 1233–1236.

Olofsson, P., Foody, G. M., Stehman, S. V. & Woodcock, C. E. (2013) Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129122–131.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C. & Others (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience*, 51 (11), 933–938.

Le Page, Y., van der Werf, G. R., Morton, D. C. & Pereira, J. M. C. (2010) Modeling fire-driven deforestation potential in Amazonia under current and projected climate conditions. *Journal of Geophysical Research*, 115 (G3), G03012.

Palsson, F., Sveinsson, J. R. & Ulfarsson, M. O. (2017) Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. *IEEE Geoscience and Remote Sensing Letters*, 14 (5), 639–643.

Pan, S. J. & Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22 (10), 1345–1359.

Parente, L., Taquary, E., Silva, A. P., Souza, C. & Ferreira, L. (2019) Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data. *Remote Sensing*, 11 (23), 2881.

Paz, P. A., Reymondin, L. & Tello, J. J. (2018a) *Reporte de la validación en campo del Sistema Terra-i (CIAT) utilizando tecnología UAV en Yurimaguas, Perú*.

Paz, P. A., Reymondin, L. & Tello, J. J. (2018b) *Reporte de trabajo de validación en campo de datos de cambios de cobertura, detectados por Terra-i en periodo 2013, 2014 y 2015*.

Paz, P. A., Tello, J. J., Leveron, S. & Reymondin, L. (2018) *Reporte de validación en campo de alertas de pérdida de cobertura dadas por el sistema Terra-i, utilizando tecnología UAV y recolección de datos con GPS, en el occidente de Honduras*.

Pekel, J.-F., Cottam, A., Gorelick, N. & Belward, A. S. (2016) High-resolution mapping

of global surface water and its long-term changes. *Nature*, 540 (7633), 418–422.

Pelletier, C., Valero, S., Inglada, J., Champion, N. & Dedieu, G. (2016) Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment*, 187156–168.

Pelletier, C., Valero, S., Inglada, J., Champion, N., Marais Sicre, C. & Dedieu, G. (2017) Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series. *Remote Sensing*, 9 (2), 173.

Pelletier, C., Webb, G. & Petitjean, F. (2019) Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing*, 11 (5), 523.

Pendrill, F. & Persson, U. M. (2017) Combining global land cover datasets to quantify agricultural expansion into forests in Latin America: Limitations and challenges. *PLOS ONE*, 12 (7), e0181202.

Pendrill, F., Persson, U. M., Godar, J. & Kastner, T. (2019) Deforestation displaced: trade in forest-risk commodities and the prospects for a global forest transition. *Environmental Research Letters*, 14 (5), 055003.

Pendrill, F., Persson, U. M., Godar, J., Kastner, T., Moran, D., Schmidt, S. & Wood, R. (2019) Agricultural and forestry trade drives large share of tropical deforestation emissions. *Global Environmental Change*, 561–10.

Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K. & Others (2003) Good practice guidance for land use, land-use change and forestry. Good practice guidance for land use, land-use change and forestry. Available from: https://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf_files/GPG_LULUCF_FULL.pdf.

Perz, S., Chavez, A. B., Cossio, R., Hoelle, J., Leite, F. L., Rocha, K., Rojas, R. O., Shenkin, A., Araujo Carvalho, L., Castillo, J. & Rojas Cespedes, D. (2015) Trans-boundary infrastructure, access connectivity, and household land use in a tri-national frontier in the Southwestern Amazon. *Journal of Land Use Science*, 10 (3), 342–368.

Perz, S. G., Caldas, M. M., Arima, E. & Walker, R. J. (2007) Unofficial Road Building in the Amazon: Socioeconomic and Biophysical Explanations. *Development and Change*, 38 (3), 529–551.

Perz, S. G., Rosero, M., Leite, F. L., Araujo Carvalho, L., Castillo, J. & Vaca Mejia, C. (2013) Regional Integration and Household Resilience: Infrastructure Connectivity and Livelihood Diversity in the Southwestern Amazon. *Human Ecology*, 41 (4), 497–511.

Perz, S. G. & Skole, D. L. (2003) Social determinants of secondary forests in the Brazilian Amazon. *Social Science Research*,

Petitjean, F., Inglada, J. & Gancarski, P. (2012) Satellite Image Time Series Analysis

Under Time Warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50 (8), 3081–3095.

Petitjean, F., Kurtz, C., Passat, N. & Gançarski, P. (2012) Spatio-temporal reasoning for the classification of satellite image time series. *Pattern Recognition Letters*, 33 (13), 1805–1815.

Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V. & Justice, C. O. (2010) Estimating Global Cropland Extent with Multi-year MODIS Data. *Remote Sensing*, 2 (7), 1844–1863.

Piu, H. C. & Menton, M. (2014) The context of REDD+ in Peru: Drivers, agents and institutions. The context of REDD+ in Peru: Drivers, agents and institutions Available from: http://www.cifor.org/publications/pdf_files/OccPapers/OP-106.pdf.

Pontius, R. G. & Millones, M. (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32 (15), 4407–4429.

Pontius, R. G. & Santacruz, A. (2014) Quantity, exchange, and shift components of difference in a square contingency table. *International Journal of Remote Sensing*, 35 (21), 7543–7554.

Pontius, R. G., Shusas, E. & McEachern, M. (2004) Detecting important categorical land changes while accounting for persistence. *Agriculture, Ecosystems & Environment*, 101 (2–3), 251–268.

Pradhan, P. (2005) *Multiresolution based, multisensor, multispectral image fusion*. Mississippi State University. [Online Thesis] https://ir.library.msstate.edu/bitstream/handle/11668/15457/etd-07082005-140541.pdf?sequence=1

Prestele, R., Arneth, A., Bondeau, A., de Noblet-Ducoudré, N., Pugh, T. A. M., Sitch, S., Stehfest, E. & Verburg, P. H. (2017) Current challenges of implementing anthropogenic land-use and land-cover change in models contributing to climate change assessments. *Earth System Dynamics*, 8 (2), 369–386.

Proyecto MapBiomas Amazonía (2019) - *Colección 1 de los mapas anuales de cobertura y uso del suelo*. Available from: http://amazonia.mapbiomas.org/.

Qiu, C., Mou, L., Schmitt, M. & Zhu, X. X. (2019) Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154151–162.

RAISG (2012) *Amazonia Under Pressure*. Available from: https://www.amazoniasocioambiental.org/en/publication/amazonia-under-pressure/.

Ramankutty, N. & Graesser, J. (2017) Latin American oil palm follows an unfamiliar route to avoid deforestation. *Environmental Research Letters*, 12 (4), 041001.

Ravikumar, A., Sears, R. R., Cronkleton, P., Menton, M. & Pérez-Ojeda del Arco, M.

(2017) Is small-scale agriculture really the main driver of deforestation in the Peruvian Amazon? Moving beyond the prevailing narrative. *Conservation Letters*, 10 (2), 170–177.

Redo, D. J., Grau, H. R., Aide, T. M. & Clark, M. L. (2012) Asymmetric forest transition driven by the interaction of socioeconomic development and environmental heterogeneity in Central America. *Proceedings of the National Academy of Sciences*, 109 (23), 8839–8844.

Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., Kellndorfer, J. M., Rosenqvist, A., Lehmann, E. A., Woodcock, C. E., Seifert, F. M. & Herold, M. (2016) Combining satellite data for better tropical forest monitoring. *Nature Climate Change*, 6 (2), 120–122.

Reis, V., Hermoso, V., Hamilton, S. K., Bunn, S. E., Fluet-Chouinard, E., Venables, B. & Linke, S. (2019) Characterizing seasonal dynamics of Amazonian wetlands for conservation and decision making. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 29 (7), 1073–1082.

Reymondin, L., Jarvis, A., Pérez-Uribe, A., Touval, J., Argote, K., Rebetez, J., Guevara, E. & Mulligan, M. (2012) A methodology for near real-time monitoring of habitat change at continental scales using MODIS-NDVI and TRMM. *Submitted Remote Sensing of Environment*,

Ripley, B. D. (2007) *Pattern Recognition and Neural Networks*. Cambridge University Press.

Ritschard, G. & Studer, M. (2018a) 'Sequence Analysis: Where Are We, Where Are We Going?', in Gilbert Ritschard & Matthias Studer (eds.) *Sequence Analysis and Related Approaches: Innovative Methods and Applications*.

Ritschard, G. & Studer, M. (2018b) *Sequence Analysis and Related Approaches: Innovative Methods and Applications*. Springer.

Roerink, G. J., Menenti, M. & Verhoef, W. (2000) Reconstructing cloudfree NDVI composites using Fourier analysis of time series. *International Journal of Remote Sensing*, 21 (9), 1911–1917.

Romijn, E., Ainembabazi, J. H., Wijaya, A., Herold, M., Angelsen, A., Verchot, L. & Murdiyarso, D. (2013) Exploring different forest definitions and their impact on developing REDD+ reference emission levels: A case study for Indonesia. *Environmental Science & Policy*, 33246–259.

Rostami, M., Kolouri, S., Eaton, E. & Kim, K. (2019) Deep Transfer Learning for Few-Shot SAR Image Classification. *Remote Sensing*, 11 (11), 1374.

Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 2053–65.

Rudel, T. K. (2013) *Tropical Forests: Paths of Destruction and Regeneration*. Columbia University Press.

Rudel, T. K., Coomes, O. T., Moran, E., Achard, F., Angelsen, A., Xu, J. & Lambin, E.

(2005) Forest transitions: towards a global understanding of land use change. *Global Environmental Change*, 15 (1), 23–31.

Rudel, T. K., De Fries, R. S., Asner, G. P. & Laurance, W. F. (2009) Changing Drivers of Deforestation and New Opportunities for Conservation. *Conservation Biology*, 23 (6), 1396–1405.

Rudel, T. K., Meyfroidt, P., Chazdon, R., Bongers, F., Sloan, S., Grau, H. R., Van Holt, T. & Schneider, L. (2020) Whither the forest transition? Climate change, policy responses, and redistributed forests in the twenty-first century. *Ambio*, 49 (1), 74–84.

Rußwurm, M. & Körner, M. (2018a) 'Convolutional LSTMs for Cloud-Robust Segmentation of Remote Sensing Imagery', in *NeurIPS 2018 Spatiotemporal Workshop*.

Rußwurm, M. & Körner, M. (2018b) Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. *ISPRS International Journal of Geo-Information*, 7 (4), 129.

Rußwurm, M. & Körner, M. (2019) *Self-Attention for Raw Optical Satellite Time Series Classification*. Available from: http://arxiv.org/abs/1910.10536.

Rustowicz, R., Cheong, R., Wang, L., Ermon, S., Burke, M. & Lobell, D. (2019) 'Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods', in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Rutherford, G. N., Guisa, A. & Zimmermann, N. E. (2007) Evaluating sampling strategies and logistic regression methods for modelling complex land cover changes. *Journal of Applied Ecology*, 44 (2), 414–424.

Ryan, C. M., Berry, N. J. & Joshi, N. (2014) Quantifying the causes of deforestation and degradation and creating transparent REDD+ baselines: A method and case study from central Mozambique. *Applied Geography*, 5345–54.

Saah, D., Tenneson, K., Matin, M., Uddin, K., Cutter, P., Poortinga, A., Nguyen, Q. H., Patterson, M., Johnson, G., Markert, K., Flores, A., Anderson, E., Weigel, A., Ellenberg, W. L., Bhargava, R., Aekakkararungroj, A., Bhandari, B., Khanal, N., Housman, I. W., Potapov, P., Tyukavina, A., Maus, P., Ganz, D., Clinton, N. & Chishtie, F. (2019) Land Cover Mapping in Data Scarce Environments: Challenges and Opportunities. *Frontiers in Environmental Science*, 7150.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M. & Morel, A. (2011) Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, 108 (24), 9899–9904.

Saito, A., Fonseca, L. M. G., Escada, M. I. S. & Korting, T. S. (2012) Efeitos da mudança de escala em padrões de desmatamento na Amazônia. *Revista Brasileira de Cartografia*, 63 (3), 401–414.

Salinas, D., Shen, H. & Perrone, V. (2019) *A Copula approach for hyperparameter transfer learning*. Available from: http://arxiv.org/abs/1909.13595.

Salk, C., Fritz, S., See, L., Dresel, C. & McCallum, I. (2018) An Exploration of Some Pitfalls of Thematic Map Assessment Using the New Map Tools Resource. *Remote Sensing*, 10 (3), 376.

Salvini, G., Herold, M., De Sy, V., Kissinger, G., Brockhaus, M. & Skutsch, M. (2014) How countries link REDD+ interventions to drivers in their readiness plans: implications for monitoring systems. *Environmental Research Letters*, 9 (7), 074004.

Sameen, M. I., Pradhan, B. & Aziz, O. S. (2018) Classification of Very High Resolution Aerial Photos Using Spectral-Spatial Convolutional Neural Networks. *Journal of Sensors*, 20181–12.

Schepaschenko, D., See, L., Lesiv, M., Bastin, J.-F., Mollicone, D., Tsendbazar, N.-E., Bastin, L., McCallum, I., Laso Bayas, J. C., Baklanov, A., Perger, C., Dürauer, M. & Fritz, S. (2019) Recent Advances in Forest Observation with Visual Interpretation of Very High-Resolution Imagery. *Surveys in Geophysics*, 40 (4), 839–862.

Schepaschenko, D., See, L., Lesiv, M., McCallum, I., Fritz, S., Salk, C., Moltchanova, E., Perger, C., Shchepashchenko, M., Shvidenko, A., Kovalevskyi, S., Gilitukha, D., Albrecht, F., Kraxner, F., Bun, A., Maksyutov, S., Sokolov, A., Dürauer, M., Obersteiner, M., Karminov, V. & Ontikov, P. (2015) Development of a global hybrid forest mask through the synergy of remote sensing, crowdsourcing and FAO statistics. *Remote Sensing of Environment*, 162208–220.

Schmidhuber, J. (2015) Deep learning in neural networks: An overview. *Neural Networks*, 6185–117.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J. & Brenning, A. (2018) *Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data*. Available from: http://arxiv.org/abs/1803.11266.

Schwartz, N. B., Uriarte, M., DeFries, R., Gutierrez-Velez, V. H. & Pinedo-Vasquez, M. A. (2017) Land-use dynamics influence estimates of carbon sequestration potential in tropical second-growth forest. *Environmental Research Letters*, 12 (7), 074023.

Schwartz, N. B., Uriarte, M., Gutiérrez-Vélez, V. H., Baethgen, W., DeFries, R., Fernandes, K. & Pinedo-Vasquez, M. A. (2015) Climate, landowner residency, and land cover predict local scale fire activity in the Western Amazon. *Global Environmental Change*, 31144–153.

Scouvart, M., Adams, R. T., Caldas, M., Dale, V., Mertens, B., Nedelec, V., Pacheco, P., Rihoux, B. & Lambin, E. F. (2008) Causes of deforestation in the Brazilian Amazon: a qualitative comparative analysis. *Journal of Land Use Science*, 2 (4), 257–282.

See, L., Laso Bayas, J., Schepaschenko, D., Perger, C., Dresel, C., Maus, V., Salk, C., Weichselbaum, J., Lesiv, M., McCallum, I., Moorthy, I. & Fritz, S. (2017) LACO-

Wiki: A New Online Land Cover Validation Tool Demonstrated Using GlobeLand30 for Kenya. *Remote Sensing*, 9 (7), 754.

Segal-Rozenhaimer, M., Li, A., Das, K. & Chirayath, V. (2020) Cloud detection algorithm for multi-modal satellite imagery using convolutional neural-networks (CNN). *Remote Sensing of Environment*, 237111446.

Sexton, J. O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., Anand, A., Huang, C., Channan, S., Pimm, S. L. & Townshend, J. R. (2016) Conservation policy and the measurement of forests. *Nature Climate Change*, 6 (2), 192–196.

Sharma, A., Liu, X. & Yang, X. (2018) Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Networks*, 105346–355.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K. & Woo, W.-C. (2015) 'Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting', in *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*.

da Silva, J. M. C., Rylands, A. B. & da Fonseca, G. A. B. (2005) The Fate of the Amazonian Areas of Endemism. *Conservation Biology*, 19 (3), 689–694.

Sloan, S. & Sayer, J. A. (2015) Forest Resources Assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. *Forest Ecology and Management*, 352134–145.

Snoek, J., Larochelle, H. & Adams, R. P. (2012) 'Practical Bayesian Optimization of Machine Learning Algorithms', in *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*.

Soliman, A., Soltani, K., Yin, J., Padmanabhan, A. & Wang, S. (2017) Social sensing of urban land use based on analysis of Twitter users' mobility patterns. *PLOS ONE*, 12 (7), e0181657.

Southworth, J., Marsik, M., Qiu, Y., Perz, S., Cumming, G., Stevens, F., Rocha, K., Duchelle, A. & Barnes, G. (2011) Roads as Drivers of Change: Trajectories across the Tri-National Frontier in MAP, the Southwestern Amazon. *Remote Sensing*, 3 (5), 1047–1066.

Stéphenne, N. & Lambin, E. F. (2001) A dynamic simulation model of land-use changes in Sudano-sahelian countries of Africa (SALU). *Agriculture, Ecosystems & Environment*, 85 (1–3), 145–161.

Stoian, A., Poulain, V., Inglada, J., Poughon, V. & Derksen, D. (2019) Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems. *Remote Sensing*, 11 (17), 1986.

Studer, M. (2013) *WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R*.

Studer, M. & Ritschard, G. (2016) What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. *Journal*

of the Royal Statistical Society: Series A (Statistics in Society), 179 (2), 481–511.

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P. & Friedl, M. A. (2019) Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sensing of Environment*, 222183–194.

Sulla-Menashe, D., Kennedy, R. E., Yang, Z., Braaten, J., Krankina, O. N. & Friedl, M. A. (2014) Detecting forest disturbance in the Pacific Northwest from MODIS time series using temporal segmentation. *Remote Sensing of Environment*, 151114–123.

Sun, Y., Luo, J., Wu, T., Zhou, Y., Liu, H., Gao, L., Dong, W., Liu, W., Yang, Y., Hu, X., Wang, L. & Zhou, Z. (2019) Synchronous Response Analysis of Features for Remote Sensing Crop Classification Based on Optical and SAR Time-Series Data. *Sensors*, 19 (19), 4227.

Sun, Z., Di, L. & Fang, H. (2019) Using long short-term memory recurrent neural network in land cover classification on Landsat and Cropland data layer time series. *International Journal of Remote Sensing*, 40 (2), 593–614.

Swersky, K., Snoek, J. & Adams, R. P. (2013) 'Multi-Task Bayesian Optimization', in *Advances in Neural Information Processing Systems 26*.

De Sy, V., Herold, M., Achard, F., Avitabile, V., Baccini, A., Carter, S., Clevers, J. G. P. W., Lindquist, E., Pereira, M. & Verchot, L. (2019) Tropical deforestation drivers and associated carbon emission factors derived from remote sensing data. *Environmental Research Letters*, 14 (9), 094022.

De Sy, V., Herold, M., Achard, F., Beuchle, R., Clevers, J. G. P. W., Lindquist, E. & Verchot, L. (2015) Land use patterns and related carbon losses following deforestation in South America. *Environmental Research Letters*, 10 (12), 124004.

Tang, X., Bullock, E. L., Olofsson, P., Estel, S. & Woodcock, C. E. (2019) Near real-time monitoring of tropical forest disturbance: New algorithms and assessment framework. *Remote Sensing of Environment*, 224202–218.

Tang, X., Hutyra, L. R., Arévalo, P., Baccini, A., Woodcock, C. E. & Olofsson, P. (2020) Spatiotemporal tracking of carbon emissions and uptake using time series analysis of Landsat data: A spatially explicit carbon bookkeeping model. *Science of The Total Environment*, 720137409.

Tansey, K., Beston, J., Hoscilo, A., Page, S. E. & Paredes Hernández, C. U. (2008) Relationship between MODIS fire hot spot count and burned area in a degraded tropical peat swamp forest in Central Kalimantan, Indonesia. *Journal of Geophysical Research*, 113 (D23), D23112.

Tao, X., Liang, S. & Wang, D. (2015) Assessment of five global satellite products of fraction of absorbed photosynthetically active radiation: Intercomparison and direct validation against ground-based data. *Remote Sensing of Environment*, 163270–285.

Tardy, B., Inglada, J. & Michel, J. (2019) Assessment of Optimal Transport for

Operational Land-Cover Mapping Using High-Resolution Satellite Images Time Series without Reference Data of the Mapping Period. *Remote Sensing*, 11 (9), 1047.

Tarrio, K., Friedl, M. A., Woodcock, C. E., Olofsson, P., Turlej, K., Zhu, Z., Loveland, T. R., Bullock, E., Arevalo, P. A., Holden, C., Pasquarella, V. J. & Zhang, Y. (2019) 'Global Land Cover mapping and Estimation (GLanCE): a multitemporal Landsat-based data record of 21st century global land cover, land use and land cover change', in *AGU Fall Meeting 2019*.

Tatem, A., Goetz, S. & Hay, S. (2008) Fifty Years of Earth-observation Satellites. *American Scientist*, 96 (5), 390.

Teimouri, N., Dyrmann, M. & Jørgensen, R. N. (2019) A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. *Remote Sensing*, 11 (8), 990.

Tsendbazar, N.-E., de Bruin, S., Fritz, S. & Herold, M. (2015) Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sensing*, 7 (12), 15804–15821.

Tsendbazar, N.-E., de Bruin, S. & Herold, M. (2017) Integrating global land cover datasets for deriving user-specific maps. *International Journal of Digital Earth*, 10 (3), 219–237.

Tsendbazar, N.-E., Herold, M., de Bruin, S., Lesiv, M., Fritz, S., Van De Kerchove, R., Buchhorn, M., Duerauer, M., Szantoi, Z. & Pekel, J.-F. (2018) Developing and applying a multi-purpose land cover validation dataset for Africa. *Remote Sensing of Environment*, 219298–309.

Tyukavina, A., Hansen, M. C., Potapov, P., Parker, D., Okpa, C., Stehman, S. V., Kommareddy, I. & Turubanova, S. (2018) Congo Basin forest loss dominated by increasing smallholder clearing. *Science Advances*, 4 (11), eaat2993.

Tyukavina, A., Hansen, M. C., Potapov, P. V, Stehman, S. V, Smith-Rodriguez, K., Okpa, C. & Aguilar, R. (2017) Types and rates of forest disturbance in Brazilian Legal Amazon, 2000–2013. *Science Advances*, 3 (4), e1601047.

UNEP-WCMC (2020) Protected Planet: The World Database on Protected Areas (WDPA). UNEP-WCMC and IUCN: Cambridge UK and Gland, Switzerland. Available from: https://www.protectedplanet.net/.

UNFCCC (2001) *Report of the conference of the parties on its seventh session, held at Marrakesh from 29 October to 10 November 2001*. Available from: https://unfccc.int/resource/docs/cop7/13a01.pdf.

Uriarte, M., Pinedo-Vasquez, M., DeFries, R. S., Fernandes, K., Gutierrez-Velez, V., Baethgen, W. E. & Padoch, C. (2012) Depopulation of rural landscapes exacerbates fire activity in the western Amazon. *Proceedings of the National Academy of Sciences*, 109 (52), 21546–21550.

Valqui, M., Feather, C. & Llanos, R. E. (2015) Revealing the hidden-indigenous perspectives on deforestation in the Peruvian Amazone; the causes and

solutions. Lima, Perú Available from: https://rightsanddeforestation.org/resources/revealing-the-hidden/.

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*.

Verburg, P. H., Crossman, N., Ellis, E. C., Heinimann, A., Hostert, P., Mertz, O., Nagendra, H., Sikor, T., Erb, K.-H., Golubiewski, N., Grau, R., Grove, M., Konaté, S., Meyfroidt, P., Parker, D. C., Chowdhury, R. R., Shibata, H., Thomson, A. & Zhen, L. (2015) Land system science and sustainable development of the earth system: A global land project perspective. *Anthropocene*, 1229–41.

Verburg, P. H., Ellis, E. C. & Letourneau, A. (2011) A global assessment of market accessibility and market influence for global environmental change studies. *Environmental Research Letters*, 6 (3), 034019.

Vijay, V., Reid, C. D., Finer, M., Jenkins, C. N. & Pimm, S. L. (2018) Deforestation risks posed by oil palm expansion in the Peruvian Amazon. *Environmental Research Letters*, 13 (11), 114010.

van Vliet, N., Mertz, O., Heinimann, A., Langanke, T., Pascual, U., Schmook, B., Adams, C., Schmidt-Vogt, D., Messerli, P., Leisz, S., Castella, J.-C., Jørgensen, L., Birch-Thomsen, T., Hett, C., Bech-Bruun, T., Ickowitz, A., Vu, K. C., Yasuyuki, K., Fox, J., Padoch, C., Dressler, W. & Ziegler, A. D. (2012) Trends, drivers and impacts of changes in swidden cultivation in tropical forest-agriculture frontiers: A global assessment. *Global Environmental Change*, 22 (2), 418–429.

Waldner, F., Schucknecht, A., Lesiv, M., Gallego, J., See, L., Pérez-Hoyos, A., D'Andrimont, R., de Maet, T., Bayas, J. C. L., Fritz, S., Leo, O., Kerdiles, H., Díez, M., Van Tricht, K., Gilliams, S., Shelestov, A., Lavreniuk, M., Simões, M., Ferraz, R., Bellón, B., Bégué, A., Hazeu, G., Stonacek, V., Kolomaznik, J., Misurec, J., Verón, S. R., de Abelleyra, D., Plotnikov, D., Mingyong, L., Singha, M., Patil, P., Zhang, M. & Defourny, P. (2019) Conflation of expert and crowd reference data to validate global binary thematic maps. *Remote Sensing of Environment*, 221235–246.

Wang, H., Zhao, X., Zhang, X., Wu, D. & Du, X. (2019) Long Time Series Land Cover Classification in China from 1982 to 2015 Based on Bi-LSTM Deep Learning. *Remote Sensing*, 11 (14), 1639.

Wang, Q., Shi, W., Atkinson, P. M. & Zhao, Y. (2015) Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, 166191–204.

Wang, Q., Shi, W., Li, Z. & Atkinson, P. M. (2016) Fusion of Sentinel-2 images. *Remote Sensing of Environment*,

Wang, S., Azzari, G. & Lobell, D. B. (2019) Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*, 222303–317.

Wang, Y. & Dong, D. (1997) Retrieving forest stand parameters from SAR backscatter data using a neural network trained by a canopy backscatter model. *International Journal of Remote Sensing*, 18 (4), 981–989.

Wang, Z., Schaaf, C. B., Sun, Q., Shuai, Y. & Román, M. O. (2018) Capturing rapid land surface dynamics with Collection V006 MODIS BRDF/NBAR/Albedo (MCD43) products. *Remote Sensing of Environment*, 20750–64.

Wang, Z., Yan, W. & Oates, T. (2017) 'Time series classification from scratch with deep neural networks: A strong baseline', in *2017 International Joint Conference on Neural Networks (IJCNN)*.

Ward, J. H. (1963) Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58 (301), 236–244.

Waske, B. & Braun, M. (2009) Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64 (5), 450–457.

Watanabe, M., Koyama, C. N., Hayashi, M., Nagatani, I. & Shimada, M. (2018) Early-Stage Deforestation Detection in the Tropics With L -band SAR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11 (6), 2127–2133.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., Battle, K. E., Bhatt, S. & Gething, P. W. (2018) A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553 (7688), 333–336.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y. & van Leeuwen, T. T. (2010) Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics*, 10 (23), 11707–11735.

Wessels, K., van den Bergh, F., Roy, D., Salmon, B., Steenkamp, K., MacAlister, B., Swanepoel, D. & Jewitt, D. (2016) Rapid Land Cover Map Updates Using Change Detection and Robust Random Forest Classifiers. *Remote Sensing*, 8 (11), 888.

White, J. C., Wulder, M. A., Hobart, G. W., Luther, J. E., Hermosilla, T., Griffiths, P., Coops, N. C., Hall, R. J., Hostert, P., Dyk, A. & Guindon, L. (2014) Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Canadian Journal of Remote Sensing*, 40 (3), 192–212.

Wilson, S. J., Schelhas, J., Grau, R., Nanni, A. S. & Sloan, S. (2017) Forest ecosystem-service transitions: the ecological dimensions of the forest transition. *Ecology and Society*, 22 (4), art38.

Wright, M. N. & Ziegler, A. (2017) ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77 (1), art38.

Wu, Z., Wang, X., Jiang, Y.-G., Ye, H. & Xue, X. (2015) 'Modeling Spatial-Temporal Clues in a Hybrid Deep Learning Framework for Video Classification', in

Proceedings of the 23rd ACM International Conference on Multimedia.

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C. & Hermosilla, T. (2018) Land cover 2.0. *International Journal of Remote Sensing*, 39 (12), 4254–4284.

Wulder, M. A., Franklin, S. E., White, J. C., Linke, J. & Magnussen, S. (2006) An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data. *International Journal of Remote Sensing*, 27 (4), 663–683.

Xing, H., Meng, Y., Wang, Z., Fan, K. & Hou, D. (2018) Exploring geo-tagged photos for land cover validation with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 141237–251.

Xue, Z., Du, P. & Feng, L. (2014) Phenology-Driven Land Cover Classification and Trend Analysis Based on Long-term Remote Sensing Image Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7 (4), 1142–1156.

Yang, C., Goodchild, M., Huang, Q., Nebert, D., Raskin, R., Xu, Y., Bambacus, M. & Fay, D. (2011) Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing? *International Journal of Digital Earth*, 4 (4), 305–329.

Yang, C., Yu, M., Hu, F., Jiang, Y. & Li, Y. (2017) Utilizing Cloud Computing to address big geospatial data challenges. *Computers, Environment and Urban Systems*, 61120–128.

Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L. & Zhu, D. (2019) Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges. *Remote Sensing*, 12 (1), 62.

Zhang, B., Zhang, Q., Feng, C., Feng, Q. & Zhang, S. (2017) Understanding Land Use and Land Cover Dynamics from 1976 to 2014 in Yellow River Delta. *Land*, 6 (1), 20.

Zhang, C., Pan, X., Zhang, S. Q., Li, H. P. & Atkinson, P. M. (2017) A Rough Set Decision Tree Based MLP-CNN For Very High Resolution Remotely Sensed Image Classification. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W71451–1454.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J. & Atkinson, P. M. (2019) Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, 221173–187.

Zhang, H. K. & Roy, D. P. (2017) Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of Environment*, 19715–34.

Zhang, X., Wu, C., Huang, Z. & Li, Z. (2018) 'Occupation-Oblivious Pricing of Cloud Jobs via Online Learning', in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*.

Zheng, A. (2015) *Evaluating machine learning models: a beginner's guide to key*

concepts and pitfalls. O'Reilly Media.

Zhong, L., Hu, L. & Zhou, H. (2019) Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, 221430–443.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F. & Fraundorfer, F. (2017) Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5 (4), 8–36.

Zhu, Z. & Woodcock, C. E. (2014) Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144152–171.

Appendices

Appendix I: Characteristics of the different LC classification schemes and global products

MODIS Land Cover Type product (MCD12Q1)

MCD12Q1 is a global land cover product produced by NASA and collaborators (Friedl *et al*., 2010). This product, derived mainly from MODIS Terra and Aqua satellite data, provides 500-m resolution LC maps from 2001 annually. The most recent update, Collection 6, includes annual maps in multiple LC classification schemes from 2001 to 2017.

Whilst the Collection 5 product was mainly based on the 17-class International Geosphere-Biosphere Programme (IGBP) legend, the Collection 6 algorithm uses a hierarchical classification model where classes included in each level of the hierarchy reflect structured distinctions between land cover properties (Sulla-Menashe *et al*., 2019). The nested classifications are combined to create eight distinct classification schemes including the five legacy schemes included in Collection 5 (IGBP, UMD, LAI, BGC, and PFT), and three new legends based on the FAO-Land Cover Classification System (LCCS) (see **Table 1**).

The MCD12Q1 product is generated using supervised classification of MODIS reflectance data. In Collection 6, the nested classification was created based on smoothed and gap-filled MODIS Nadir BRDF-Adjusted Reflectance (NBAR) time series data obtained from MC43A2 and MC43A4 products (Wang *et al*., 2018). The smoothed time series were used to generate metrics including annual quantiles and variances for the spectral bands and several band combinations. These annual metrics were used as inputs to a Random Forest classifier for each layer of the hierarchy (Sulla-Menashe *et al*., 2019).

Following the supervised classification of smoothed NBAR data, the outputs were adjusted using ancillary information including water masks, vegetation continuous fields, climate, among other layers described in Sulla-Menashe *et al*. (2019). Then, to reduce interannual variability caused by classifier instability, Hidden Markov Models as proposed by Abercrombie and Friedl (2016) were used to post-process

map results for each year. After stabilization, the classifications were condensed into the final set of classification schemes previously described in **Table 1.**

Table 1 Description of the eight distinct LC classification schemes embedded in the Science Data Sets (SDS) of the MCD12Q1 product Collection 6. The schemes used in this work are highlighted in grey. Table adapted from Sulla-Menashe *et al.* (2019).

SDS Full Name	Short name	Description	Number of classes	Valid range
Land Cover Type 1	LC Type1	Annual IGBP classification	17	[1,17]
Land Cover Type 2	LC Type2	Annual University of Maryland (UMD) classification	16	[0,15]
Land Cover Type 3	LC Type3	Annual Leaf Area Index (LAI) classification	11	[0,10]
Land Cover Type 4	LC Type4	Annual BIOME-Biogeochemical Cycles (BGC) classification	9	[0,8]
Land Cover Type 5	LC Type5	Annual Plant Functional Type (PFT) classification	12	[0,11]
Land Cover Property 1	LC Prop1	FAO-Land Cover Classification System 1 (LCCS1) land cover layer	16	[1,43]
Land Cover Property 2	LC Prop2	FAO-LCCS2 land use layer	11	[1,40]
Land Cover Property 3	LC Prop3	FAO-LCCS3 surface hydrology layer	10	[1,51]

The reported accuracy of the MCD12Q1 product collection goes between 67% (IGBP) to 87% (FAO-LCCS3). The performance per class is different according to the classification scheme. Overall, for LC schemes such as IGBP and LCCS1 classifications, mosaics of distinct LC types (e.g., cropland/natural vegetation mosaics) or classes that occur along a climate gradient with mixtures of life forms at sub-pixel scales (e.g., shrublands, grasslands, and woodlands) are the most problematic (Sulla-Menashe *et al.*, 2019).

ESA-CCI LC

The European Space Agency-Climate Change Initiative (ESA-CCI) contains a set of multi-sensor global LC datasets. The last release, v2.0.7, provides a long-term series (1992-2015) on a yearly basis at a spatial resolution of 300-m. The product is derived from a combination and processing of global daily surface reflectance

of different observation systems: 300-m ENVISAT MERIS, 1-km SPOT-VEGETATION (SPOT-VGT), 1-km PROBA-V and 1-km AVHRR images (ESA, 2017).

The original dataset classifies the global LC into 37 classes (22 global classes and 15 regional classes) based on FAO's Land Cover Classification System (FAO-LCCS). The classification process applies machine learning and unsupervised algorithms on the whole ENVISAT-MERIS (2003-2012) archive using the majority of MERIS spectral bands to produce a LC baseline. The 1-km time series recorded by AVHRR (1992-1999), SPOT-VGT (1999-2013) and PROBA-V (2014-2015) were used to detect and confirm the change which was delineated more precisely at the 300-m spatial resolution according to 300-m data availability. This last step results in both back- and forward-dating the baseline LC map (10 years) to produce the 24 annual LC maps from 1992 to 2015. In order to avoid false change detections due to the interannual variability in classifications, each change has to persist over more than 2 successive years in the classification time series to be confirmed (Li *et al*., 2018).

The reported accuracy of the ESA-CCI product, which was assessed only for the map of 2015, varies between 71.5% and 75.4% for homogeneous (i.e. made of a single LC class) or heterogeneous (i.e. made of several or mosaic LC classes), respectively (Li *et al*., 2018). The highest use accuracy (83–97%) is for croplands (irrigated and rainfed), broadleaved evergreen forests, urban areas, bare areas, water bodies, and permanent snow and ice. Low user accuracies (<60%) are observed for mosaic classes of natural vegetation, lichens and mosses, sparse vegetation, and flooded forest with fresh water (Li *et al*., 2018).

CGLS-LC100

The Copernicus Global Land Service (CGLS) is a component of the Land Monitoring Service of Copernicus managed by the European Commission in partnership with the European Space Agency, the EU Member States and EU Agencies. Amongst the products of the CGLS catalogue, a global land cover product (CGLS-LC100) for 2015 at spatial resolution of 100-m was released recently including a discrete map and 10 continuous cover fractions. The product is mainly derived from processed surface reflectance of 5-daily and daily PROBA-

V multispectral image data with a spatial resolution of 100-m and 300-m, respectively (Buchhorn *et al*., 2019).

The discrete dataset classifies the global LC into 23 classes based on FAO-LCCS. The classification process applies Random Forest algorithms on 270 metrics stratified by biome clusters extracted from whole fusioned archives of PROBA-V. Expert rules are applied to combine the existing knowledge represented by the ancillary datasets (i.e. urban and water masks) and the classification results. The predicted class probabilities are used as thresholds in the decision tree designed expert rules in order to generate the discrete LC map with 23 classes (Buchhorn *et al*., 2019).

The reported accuracy of the CGLS-LC100 discrete LC map for 2015 varies between 75% and 80% for level 1 and 2 (when open and closed forests are separated), respectively. For level 1, the highest use accuracy (89–95%) is for forest, bare/sparse vegetation, snow/ice and water classes are mapped with high accuracies. Low user accuracies (<62%) are observed for the classes of wetlands, lichen/moss and shrubs. The overall accuracies and confusion matrices are also available for each continent. For South America, the LC map reaches an overall accuracy between 74 and 79.4% for classification schemes on level 1 and 2 based on 2734 independent samples taken for the validation campaign (Buchhorn *et al*., 2019).

MapBiomas

MapBiomas is a multi-institutional initiative to generate annual LC/LU maps from automatic classification cloud processing applied to satellite imagery. The complete project description can be found at http://mapbiomas.org. In addition to MapBiomas Brazil (MapBiomas Project, 2019), the initiative provides annual maps of LC/LU for the entire Amazon (http://amazonia.mapbiomas.org/). The latest version of product (Collection 1 - Beta) across the Amazon as defined by RAISG (also known as the MapBiomas Amazonia) has annual maps of LC/LU from 2001 to 2017 at a spatial resolution of 30m. The product is derived from Landsat data archive accessed and processed through the Google Earth Engine platform (Proyecto MapBiomas Amazonía, 2019).

The original dataset classifies LC/LU into 9 classes based on user-defined needs (countries). The annual classification is generated using supervised classification of Landsat reflectance data. In version 1.0, the classification was created based on Landsat-based observations from Landsat 5-TM, 7-ETM+, and 8-OLI products. These observations were used to generate metrics including annual maximum, minimum and variances for spectral bands and several band combinations. These metrics were used as inputs to a Random Forest classifier. Spatial and temporal filters were applied to the classified images to eliminate the salt and pepper effect and reduce unlikely transitions (Proyecto MapBiomas Amazonía, 2019).

The accuracy of the MapBiomas product for the Amazon does not yet have an accuracy assessment of any type.

Appendix II: Description of original and hybrid LC reference datasets

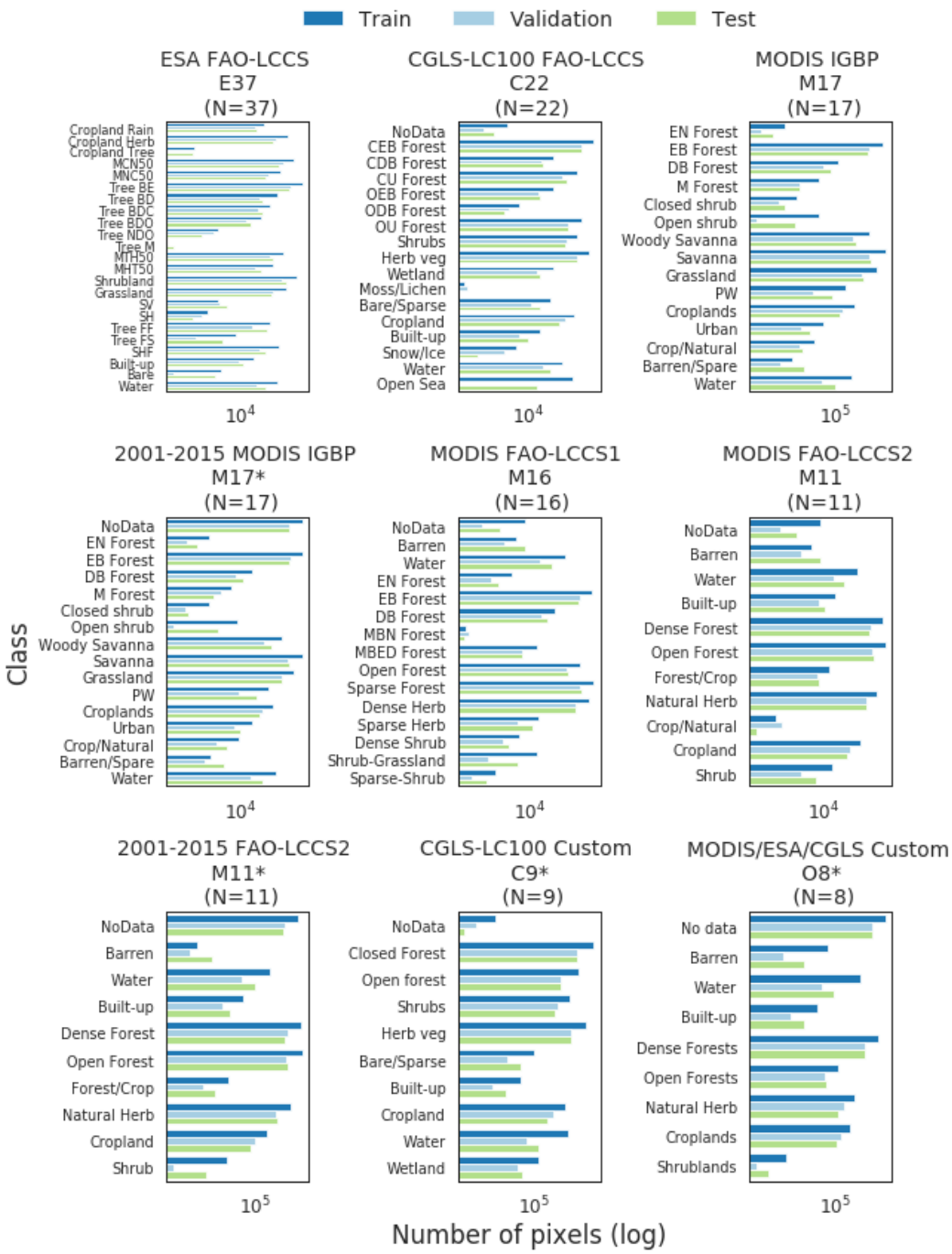


Figure 1 Class distribution of the training, validation and test partitions for the LC pre-existing and hybrid (*) reference datasets, ordered by total number of classes, assessed with the MTLCC network. Classes are represented as either; 1) a single word indicating the cover type; 2) an acronym indicating the cover type, in particular for the ESA-CCI dataset and forest/tree like cover types. Details of the classes and respective acronyms and total number of pixels per dataset can be found in the tables below.

Dataset: ESA FAO-LCCS - E37 (N=37)

Class	Acronym	Number pixels (% total)		
		Train	Validation	Test
Cropland rainfed	Cropland Rain	126,919 (0.49)	46,921 (0.71)	60,251 (0.91)
Cropland rainfed - Herbaceous cover	Cropland Herb	1,857,878 (7.12)	519,398 (7.83)	381,144 (5.74)
Cropland rainfed - Tree or shrub cover	Cropland Tree	59 (0)	0 (0)	50 (0)
Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	MCN50	3,576,535 (13.7)	1,090,892 (16.44)	681,806 (10.28)
Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	MNC50	756,073 (2.9)	204,004 (3.07)	169,477 (2.55)
Tree cover, broadleaved, evergreen, closed to open (>15%)	Tree BE	9,332,191 (35.76)	2,716,240 (40.93)	2,185,505 (32.94)
Tree cover, broadleaved, deciduous, closed to open (>15%)	Tree BD	608,171 (2.33)	82,805 (1.25)	116,664 (1.76)
Tree cover, broadleaved, deciduous, closed (>40%)	Tree BDC	273,850 (1.05)	71,588 (1.08)	118,481 (1.79)
Tree cover, broadleaved, deciduous, open (15-40%)	Tree BDO	98,608 (0.38)	19,468 (0.29)	31,065 (0.47)
Tree cover, needleleaved, deciduous, closed to open (>15%)	Tree NDO	877 (0)	510 (0.01)	150 (0)
Tree cover, mixed leaf type (broadleaved and needleleaved)	Tree M	0 (0)	0 (0)	6 (0)
Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	MTH50	1,143,007 (4.38)	253,437 (3.82)	332,448 (5.01)
Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	MHT50	349,944 (1.34)	54,053 (0.81)	92,637 (1.4)
Shrubland	Shrubland	4,723,376 (18.1)	1,025,056 (15.45)	1,605,219 (24.19)
Grassland	Grassland	1,688,442 (6.47)	360,007 (5.43)	314,141 (4.73)
Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	SV	835 (0)	1,051 (0.02)	2,383 (0.04)
Sparse herbaceous cover (<15%)	SH	275 (0)	139 (0)	49 (0)
Tree cover flooded fresh or brakish water	Tree FF	240,589 (0.92)	38,944 (0.59)	191,825 (2.89)
Tree cover flooded saline water	Tree FS	5,606 (0.02)	68 (0)	1,449 (0.02)

Shrub or herbaceous cover, flooded, fresh/saline/brakish water	SHF	680,022 (2.61)	83,138 (1.25)	164,438 (2.48)
Urban areas	Built-up	43,109 (0.17)	7,895 (0.12)	14,167 (0.21)
Bare areas	Bare	1,229 (0)	6 (0)	622 (0.01)
Water bodies	Water	592,117 (2.27)	59,900 (0.9)	171,543 (2.59)
TOTALS		**26,099,712**	**6,635,520**	**6,635,520**

Dataset: CGLS-LC100 FAO-LCCS - C22 (N=22)

Class	**Acronym**	**Number pixels (% total)**		
		Training	Validation	Test
Non-reliable	*NoData*	1,129 (0)	87 (0)	269 (0)
Closed forest, evergreen broadleaf	CEB Forest	10,625,337 (40.71)	2,637,562 (39.75)	2,666,731 (40.19)
Closed forest, deciduous broadleaf	CDB Forest	152,280 (0.58)	42,477 (0.64)	46,495 (0.7)
Closed forest, unknown	CU Forest	1,892,804 (7.25)	391,316 (5.9)	554,322 (8.35)
Open forest, evergreen broadleaf	OEB Forest	151,606 (0.58)	31,562 (0.48)	34,395 (0.52)
Open forest, deciduous broadleaf	ODB Forest	3,954 (0.02)	1,251 (0.02)	774 (0.01)
Open forest, unknown	OU Forest	2,705,226 (10.36)	659,158 (9.93)	649,226 (9.78)
Shrubs	Shrubs	1,734,300 (6.64)	619,005 (9.33)	500,008 (7.54)
Herbaceous vegetation	Herb veg	5,957,365 (22.83)	1,695,217 (25.55)	1,689,470 (25.46)
Herbaceous wetland	Wetland	140,629 (0.54)	24,932 (0.38)	36,024 (0.54)
Moss and lichen	Moss/ Lichen	12 (0)	15 (0)	0 (0)
Bare / sparse vegetation	Bare/ Sparse	97,055 (0.37)	12,590 (0.19)	35,440 (0.53)
Cultivated and managed vegetation/agriculture	Cropland	1,202,077 (4.61)	468,063 (7.05)	280,376 (4.23)
Urban / built up	Built-up	32,171 (0.12)	3,561 (0.05)	9,173 (0.14)
Snow and Ice	Snow/Ice	2,923 (0.01)	789 (0.01)	47 (0)
Permanent water bodies	Water	370,183 (1.42)	47,935 (0.72)	106,711 (1.61)
Open sea	Open Sea	1,030,661 (3.95)	0 (0)	26,059 (0.39)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**26,098,583**	**6,635,433**	**6,635,251**

Dataset: MODIS IGBP - M17 (N=17)

Class	Acronym	Number pixels (% total)		
		Training	Validation	Test
Evergreen needleleaf forest	EN Forest	1,112 (0)	127 (0)	373 (0.01)
Evergreen broadleaf forest	EB Forest	7,900,335 (30.27)	2,286,420 (34.46)	1,877,124 (28.29)
Deciduous broadleaf forest	DB Forest	145,475 (0.56)	35,317 (0.53)	68,921 (1.04)
Mixed forest	M Forest	22,836 (0.09)	4,521 (0.07)	4,332 (0.07)
Closed shrublands	Closed shrub	3,129 (0.01)	620 (0.01)	1,078 (0.02)
Open shrublands	Open shrub	22,949 (0.09)	87 (0)	2,915 (0.04)
Woody savannas	Woody Savanna	2,277,765 (8.73)	523,592 (7.89)	649,748 (9.79)
Savannas	Savanna	9,778,116 (37.46)	2,300,256 (34.67)	2,425,764 (36.56)
Grasslands	Grassland	4,569,894 (17.51)	1,213,178 (18.28)	1,250,993 (18.85)
Permanent wetlands	PW	267,375 (1.02)	13,781 (0.21)	79,963 (1.21)
Croplands	Croplands	623,125 (2.39)	215,899 (3.25)	146,400 (2.21)
Urban and built-up	Urban	36,659 (0.14)	5,134 (0.08)	10,502 (0.16)
Cropland/natural vegetation mosaic	Crop/ Natural	17,120 (0.07)	4,357 (0.07)	5,381 (0.08)
Barren or sparsely vegetated	Barren/ Spare	2,166 (0.01)	731 (0.01)	6,323 (0.1)
Water	Water	431,656 (1.65)	31,500 (0.47)	105,703 (1.59)
TOTALS		**26,099,712**	**6,635,520**	**6,635,520**

Dataset: 2001-2015 MODIS IGBP - M17 (N=17)*

Class	Acronym	Number pixels (% total)		
		Training	Validation	Test
Non-reliable	*NoData*	6,888,211 (26.39)	1,856,889 (27.98)	1,704,538 (25.69)
Evergreen needleleaf forest	EN Forest	345 (0)	37 (0)	109 (0)
Evergreen broadleaf forest	EB Forest	7,307,472 (28)	2,145,589 (32.33)	1,723,673 (25.98)
Deciduous broadleaf forest	DB Forest	35,108 (0.13)	6,558 (0.1)	14,507 (0.22)
Mixed forest	M Forest	3,833 (0.01)	1,195 (0.02)	596 (0.01)
Closed shrublands	Closed shrub	346 (0)	28 (0)	40 (0)

Open shrublands	Open shrub	6,796 (0.03)	8 (0)	996 (0.02)
Woody savannas	Woody Savanna	863,456 (3.31)	129,130 (1.95)	291,231 (4.39)
Savannas	Savanna	7,095,333 (27.19)	1,574,454 (23.73)	1,801,119 (27.14)
Grasslands	Grassland	2,900,453 (11.11)	774,384 (11.67)	853,081 (12.86)
Permanent wetlands	PW	206,045 (0.79)	8,678 (0.13)	54,199 (0.82)
Croplands	Croplands	335,634 (1.29)	101,181 (1.52)	72,967 (1.1)
Urban and built-up	Urban	36,053 (0.14)	5,126 (0.08)	10,385 (0.16)
Cropland/natural vegetation mosaic	Crop/ Natural	8,167 (0.03)	857 (0.01)	2,539 (0.04)
Barren or sparsely vegetated	Barren/ Spare	449 (0)	238 (0)	1,865 (0.03)
Water	Water	412,011 (1.58)	31,168 (0.47)	103,675 (1.56)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**19,211,501**	**4,778,631**	**4,930,982**

Dataset: MODIS FAO-LCCS1 - M16 (N=16)

Class	**Acronym**	**Number pixels (% total)**		
		Training	Validation	Test
Non-reliable	*NoData*	6,670 (0.03)	68 (0)	442 (0.01)
Barren	Barren	2,319 (0.01)	731 (0.01)	6,311 (0.1)
Water Bodies	Water	431,664 (1.65)	31,500 (0.47)	105,703 (1.59)
Evergreen Needleleaf Forests	EN Forest	1,482 (0.01)	167 (0)	382 (0.01)
Evergreen Broadleaf Forests	EB Forest	7,959,600 (30.5)	2,290,226 (34.51)	1,903,605 (28.69)
Deciduous Broadleaf Forests	DB Forest	145,567 (0.56)	35,325 (0.53)	69,077 (1.04)
Mixed Broadleaf-Needleleaf Forests	MBN Forest	12 (0)	16 (0)	10 (0)
Mixed Broadleaf Evergreen-Deciduous Forests	MBED Forest	22,828 (0.09)	4,505 (0.07)	4,322 (0.07)
Open Forests	Open Forest	2,314,138 (8.87)	525,341 (7.92)	664,796 (10.02)
Sparse Forests	Sparse Forest	9,848,036 (37.73)	2,307,543 (34.78)	2,442,230 (36.81)
Dense Herbaceous	Dense Herb	5,313,891 (20.36)	1,436,510 (21.65)	1,420,228 (21.4)
Sparse Herbaceous	Sparse Herb	26,948 (0.1)	2,824 (0.04)	14,312 (0.22)
Dense Shrublands	Dense Shrub	3,193 (0.01)	620 (0.01)	1,102 (0.02)

Shrubland-Grassland Mosaics	Shrub-Grassland	23,101 (0.09)	124 (0)	2,899 (0.04)
Sparse Shrublands	Sparse-Shrub	263 (0)	20 (0)	101 (0)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**26,093,042**	**6,635,452**	**6,635,078**

Dataset: MODIS FAO-LCCS2 - M11 (N=11)

Class	**Acronym**	**Number pixels (% total)**		
		Training	Validation	Test
Non-reliable	*NoData*	6,646 (0.03)	68 (0)	442 (0.01)
Barren	Barren	2,166 (0.01)	731 (0.01)	6,311 (0.1)
Water Bodies	Water	431,656 (1.65)	31,500 (0.47)	105,703 (1.59)
Urban and Built-up Lands	Built-up	36,659 (0.14)	5,134 (0.08)	10,502 (0.16)
Dense Forests	Dense Forest	8,129,022 (31.15)	2,330,159 (35.12)	1,977,331 (29.8)
Open Forests	Open Forest	12,129,053 (46.47)	2,826,638 (42.6)	3,098,060 (46.69)
Forest/Cropland Mosaics	Forest/ Crop	17,071 (0.07)	4,275 (0.06)	5,377 (0.08)
Natural Herbaceous	Natural Herb	4,697,720 (18)	1,220,274 (18.39)	1,281,312 (19.31)
Natural Herbaceous-Croplands Mosaics	Crop/ Natural	37 (0)	82 (0)	4 (0)
Herbaceous Croplands	Cropland	623,125 (2.39)	215,899 (3.25)	146,400 (2.21)
Shrublands	Shrub	26,557 (0.1)	760 (0.01)	4,078 (0.06)
Barren	Barren	6,646 (0.03)	68 (0)	442 (0.01)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**26,093,066**	**6,635,452**	**6,635,078**

Dataset: 2001-2015 MODIS FAO-LCCS2 - M11 (N=11)*

Class	**Acronym**	**Number pixels (% total)**		
		Training	Validation	Test
Non-reliable	*NoData*	5,987,965 (22.94)	1,646,322 (24.81)	1,459,356 (21.99)
Barren	Barren	449 (0)	238 (0)	1,865 (0.03)
Water Bodies	Water	412,011 (1.58)	31,168 (0.47)	103,675 (1.56)
Urban and Built-up Lands	Built-up	36,053 (0.14)	5,126 (0.08)	10,385 (0.16)

Dense Forests	Dense Forest	7,441,023 (28.51)	2,168,514 (32.68)	1,782,311 (26.86)
Open Forests	Open Forest	8,890,281 (34.06)	1,903,823 (28.69)	2,330,114 (35.12)
Forest/Cropland Mosaics	Forest/Crop	8,167 (0.03)	857 (0.01)	2,539 (0.04)
Natural Herbaceous	Natural Herb	2,980,906 (11.42)	778,242 (11.73)	871,252 (13.13)
Natural Herbaceous-Croplands Mosaics	Crop/Natural	335,634 (1.29)	101,181 (1.52)	72,967 (1.1)
Herbaceous Croplands	Cropland	7,223 (0.03)	49 (0)	1,056 (0.02)
Shrublands	Shrub	5,987,965 (22.94)	1,646,322 (24.81)	1,459,356 (21.99)
Barren	Barren	449 (0)	238 (0)	1,865 (0.03)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**20,111,747**	**4,989,198**	**5,176,164**

Dataset: CGLS-LC100 Custom C9 (N=9)*

Class	**Acronym**	**Number pixels (% total)**		
		Training	Validation	Test
Non-reliable	*NoData*	4,413 (0.02)	907 (0.01)	379 (0.01)
Closed forest	Closed Forest	11,662,219 (44.68)	2,835,203 (42.73)	3,006,293 (45.31)
Open forest	Open forest	3,487,791 (13.36)	831,710 (12.53)	835,173 (12.59)
Shrubs	Shrubs	1,706,816 (6.54)	617,219 (9.3)	497,211 (7.49)
Herbaceous vegetation	Herb veg	6,317,033 (24.2)	1,778,886 (26.81)	1,789,824 (26.97)
Bare / sparse vegetation	Bare/Sparse	93,485 (0.36)	11,946 (0.18)	34,769 (0.52)
Urban / built up	Built-up	33,451 (0.13)	3,559 (0.05)	9,586 (0.14)
Cultivated and managed vegetation/agriculture	Cropland	1,227,173 (4.7)	477,222 (7.19)	286,726 (4.32)
Water bodies	Water	1,423,880 (5.46)	52,794 (0.8)	140,093 (2.11)
Herbaceous wetland	Wetland	143,451 (0.55)	26,074 (0.39)	35,466 (0.53)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**26,095,299**	**6,634,613**	**6,635,141**

Dataset: MODIS/ESA/CGLS Custom O8 (N=8)*

Class	Acronym	Number pixels (% total)		
		Training	Validation	Test
Non-reliable	*NoData*	15,968,069 (61.18)	4,196,412 (63.24)	4,420,263 (66.62)
Barren	Barren	43,513 (0.17)	519 (0.01)	4,329 (0.07)
Water Bodies	Water	1,278,452 (4.9)	27,003 (0.41)	87,796 (1.32)
Urban and Built-up Lands	Built-up	15,471 (0.06)	1,062 (0.02)	4,311 (0.06)
Dense Forests	Dense Forests	7,502,383 (28.75)	1,945,675 (29.32)	1,831,085 (27.6)
Open Forests	Open Forests	134,421 (0.52)	31,596 (0.48)	39,542 (0.6)
Natural Herbaceous	Natural Herb	691,769 (2.65)	251,144 (3.78)	140,157 (2.11)
Croplands	Croplands	464,967 (1.78)	182,076 (2.74)	107,918 (1.63)
Shrublands	Shrublands	667 (0)	33 (0)	119 (0)
TOTALS including *NoData*		26,099,712	6,635,520	6,635,520
TOTALS Excluding *NoData*		**10,131,643**	**2,439,108**	**2,215,257**

Appendix III: Methodological checking

Datasets

Optical Satellite Data

MODIS data were the principal source of satellite imagery where surface reflectance time series were extracted. MODIS provides the electromagnetic radiation of the Earth in 36 spectral channels (C. O. Justice *et al*., 2002). MODIS, onboard NASA's Terra (since 1999) and Aqua (since 2002) platforms, collect images at three different spatial resolutions: 250-m, 500-m and 1-km. The first seven bands are used to observe LC features in addition to cloud and aerosol properties.

For this work, MOD09Q1 and MOD09A1 products, at a spatial resolution of 250-m and 500-m were used. Both products are composites of 8-day periods which sum a total of 46 observations per year, except 2001 which has 45. Each pixel of the MOD09Q1 or MOD09A1 datasets contains the best possible Level 2 Gridded observation during an 8-day period and is usually selected based on high-observation coverage, low-view angle, absence of clouds or cloud shadow, and aerosol loading. MOD09Q1 provides surface reflectance in two bands (red = 620–670 nm; and Near-Infrared (NIR) = 841–876 nm). MOD09A1 also includes these bands plus five bands (blue = 459–479 nm; green = 545–565 nm; Short-Wave Infrared 1 (SWIR1) 86 = 1230–1250 nm; SWIR2 = 1628–1652 nm; SWIR3 = 2105–2155 nm) (C. O. Justice *et al*., 2002).

Beside the surface reflectance bands, the day of the year was extracted and included as input band. This particular feature has been used by the original MTLCC network and other similar models (Rustowicz *et al*., 2019) to guide a better capture of the temporal aspects of satellite image observations.

Data

According to the results from the assessment of multiple LC datasets for the study of post-loss LC change (see **Chapter 4**), LC reference data for this work corresponds to FAO's Land Cover Classification System 2 (LCCS2) layer classification (hereafter referred to as FAO-LCCS2) that contains 11 different classes. Labels are extracted from the MCD12Q1 MODIS product collection 6

provided by the NASA Land Process Distributed Active Archive Center, which contains yearly LC information. The next section describes the pre-processing conducted for the extraction of labels from the MCD12Q1 MODIS FAO-LCCS2 layer. As part of this procedure, the original 11 LC classes were reduced to 8 target classes which are relevant to the study of post-loss LC change.

The reported accuracy of MODIS FAO-LCCS2 is almost stable in the period between 2001 and 2016 (80% on average) (see **Figure 1**). For this particular layer, a high proportion of misclassification was reported for *Forest/cropland mosaics* which was erroneously classified as *Open Forest*. In addition, none of the samples of *Natural Herbaceous-Croplands Mosaics* were classified in this class in the test partition and was assigned as either *Natural Herbaceous* or *Herbaceous Cropland*.

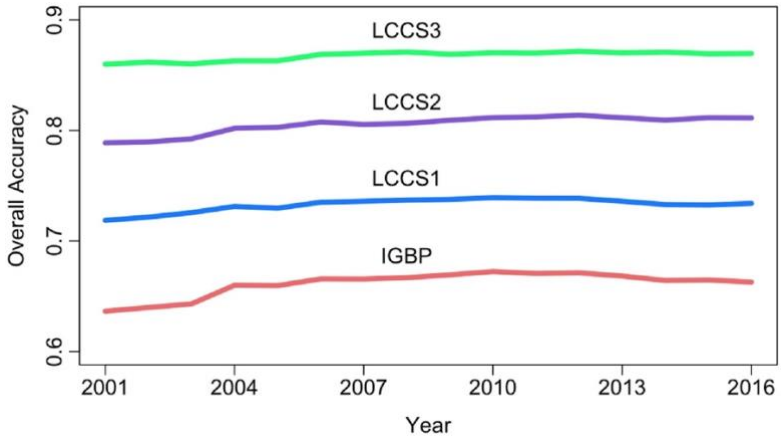


Figure 1 2001-2016 annual overall accuracy values for four classification schemes of the MCD12Q1 collection 6 product. The water class is not included in the overall accuracy statistics. Source: Sulla-Menashe *et al*. (2019).

Data preparation

Resampling

Scaling techniques were used prior to training to prepare the information required at the scale of interest, in this case at spatial resolution of the target deforested dataset, Terra-i (250-m). Downscale/upscale operations were carried out according to the type of variable, continuous or categorical (Ge *et al*., 2019). For continuous values of reflectance data, all bands, except NIR and red which are provided by MOD09Q1 product at 250-m, were resampled using bilinear interpolation which is the same used by the original MTLCC model (Rußwurm and Körner, 2018b). This operation is made on the fly according to the library used to

train the proposed model. For the labelled dataset, categorical downsampling was conducted by assigning a given land cover to a pixel of 250-m according to the overlapping class of the coarser LC map, in this case MODIS FAO-LCCS2. This operation was conducted through the cloud-based platform from which the MCD12Q1 MODIS product was accessed.

Hybrid LC dataset

Following the outcomes of **Chapter 4**, a hybrid LC dataset was generated from pixels with LC types that remain unchanged from 2001 to 2015 according to the MCD12Q1 MODIS FAO-LCCS2 layer. Due to *Forest/Cropland Mosaics* and *Natural Herbaceous/ Croplands Mosaics* being reported by the dataset producer as classes with high proportion of misclassification, both classes were reclassified as *NoData*. The implication of this procedure is that only pixels with LC types different to *NoData* are used to train and calibrate the models tested in this work. The reclassified hybrid dataset contains 8 classes: 1) Barren (Ba); 2) Water Bodies (W); 3) Urban and Built-up Lands (Bu); 4) Dense Forests (DF); 5) Open Forests (OF); 6) Natural Herbaceous (NH); 7) Herbaceous Croplands (HC); and 8) Shrublands (S). **Figure 2** illustrates the spatial distribution of eight classes of the hybrid dataset according to the study area.

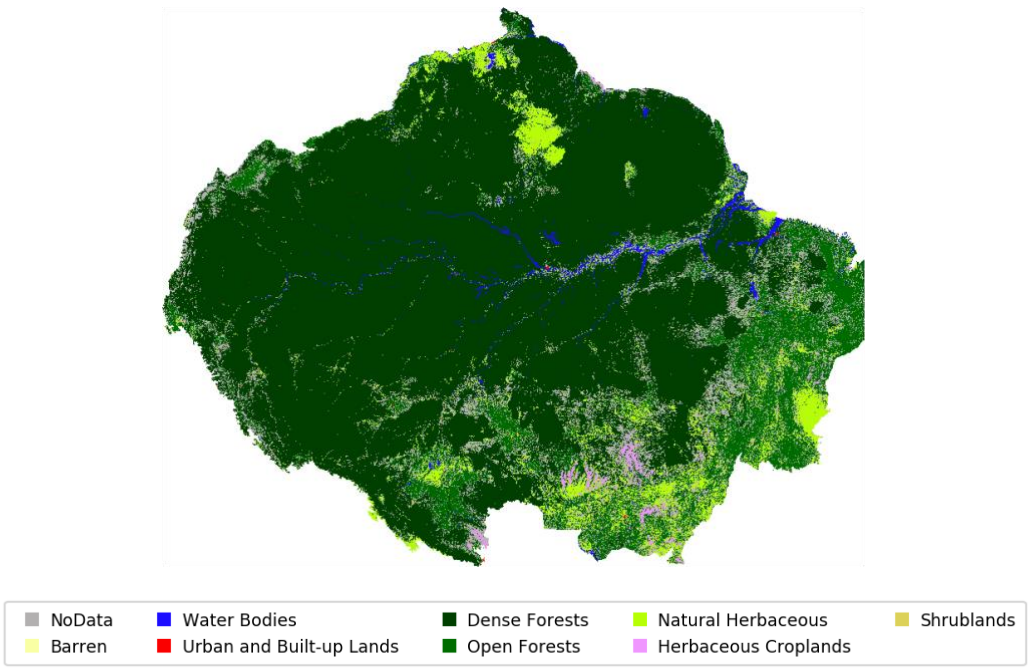


Figure 2 Map showing the hybrid dataset derived from pixels with LC types that remain unchanged according to 2001-2015 annual maps of MCD12Q1 MODIS FAO-LCCS2 layer.

The total number of pixels and proportions are reported in **Table 1**.

Table 1 Class distribution of the hybrid dataset according to the extent of the study area. Each pixel is at the scale of interest, in this case at the spatial resolution of the target deforested dataset, Terra-i (250-m).

Class ID	Class name	Shortname	N pixels	Proportion (%)
0	No data	NoData	14,755,308	10.026
1	Barren	Ba	3,054	0.002
2	Water Bodies	W	1,669,258	1.134
3	Urban/Built-up Lands	Bu	95,239	0.065
4	Dense Forests	Df	105,364,931	71.595
5	Open Forests	OF	17,687,336	12.018
6	Natural Herbaceous	NH	6,929,894	4.709
7	Herbaceous Croplands	C	647,576	0.440
8	Shrublands	S	15,321	0.010
Total			147,167,917	

Feature normalisation

A min-max normalisation per type of feature was applied prior to model training. Whilst other types of normalisation commonly used in remote sensing e.g. z-normalisation are prone to loss the temporal trend and significance of the magnitude of SITS (Pelletier *et al.*, 2019), the min-max normalisation, which performs a subtraction of the minimum, then a division by the range, i.e. the maximum minus the minimum, preserves the relationships of the original data (Han *et al.*, 2011). For this work, the min-max values were derived from the datasets' metadata which commonly indicate the range of valid values. Hence, in the MODIS product, this range goes between -100 to 16000. The normalised values were also scaled between 0 and 1. This scale ensured all features have the same probability of being taken during the learning process. The use of a global max and min might also facilitate the training and application of the model across regions different to the study area as other scholars have proposed (Pelletier *et al.*, 2016).

Data partition

For this work, rectangular tiles of 384 pixels x 384 pixels as defined in the experiments of **Chapter 4** (see **Section 4.5.4**) were used. The tile size is a multiple of the patch size of 24 pixels x 24 pixels, which is the size ingested to train the

MTLCC network. Not all tiles covering the study area were considered for use in the training and assessing of the performance of the MTLCC model. As detailed in **Chapter 4**, only those tiles with a large proportion (>70%) of *Evergreen broadleaf forest* class according to the IGBP classification from the MCD12Q1 product in 2015 were discarded to mitigate the dominance of this particular class, which is equivalent to the *Dense Forest* class in the MODIS FAO-LCCS2 layer.

The multispectral MODIS data and labels from the hybrid dataset were paired and randomly split at tile level to partitions for network training, hyper-parameter validation and model testing using a ratio of 4:1:1 (see **Figure 3a**). The tile-wise spatial separation enforces independence of the data partitions and supports bias-reduced predictive performance results (Schratz *et al.*, 2018). This spatial partition design also preserves the relative percentages of each land cover class over all splits as shown in **Figure 3b.**

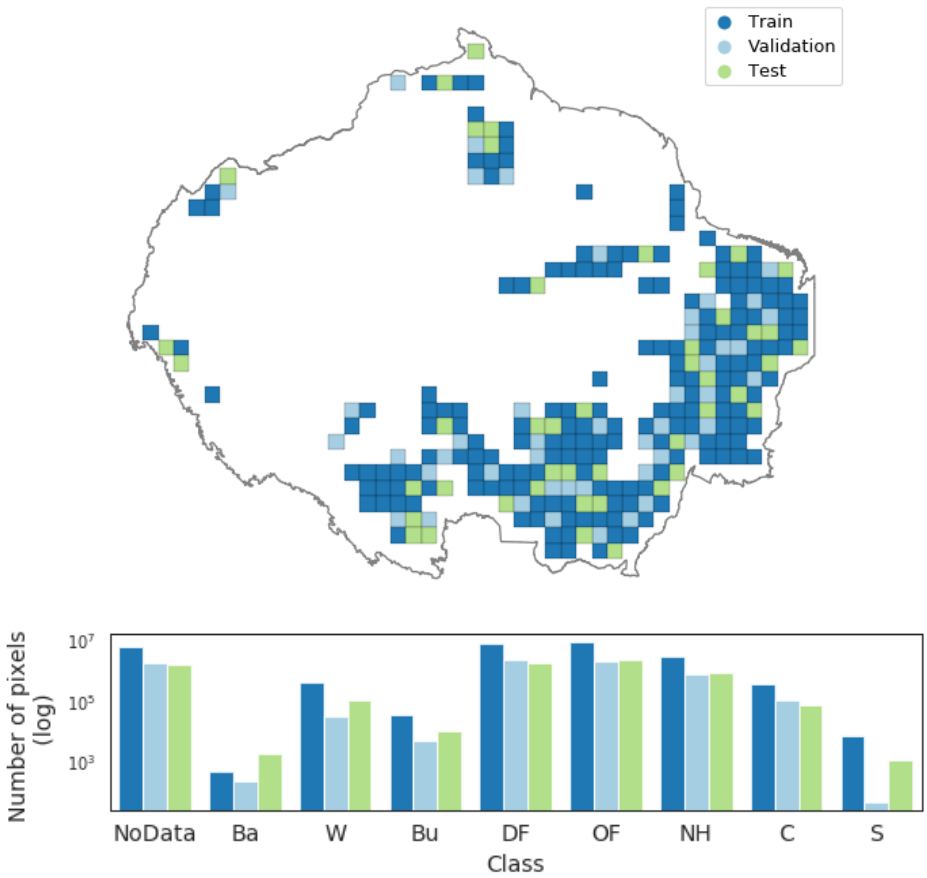


Figure 3 (a) Illustration of the tile-wise data partition using non-overlapping 384 pixels x 384 pixels tiles for training, validation, and testing; and (b) the class distribution of the 2001-2015 MODIS FAO-LCCS2 dataset per split. Classes are represented as an acronym indicating a cover type as described in **Table** .

Table 2 indicates the number of pixels per partition. Excluding *NoData*, some 20.1, 5.0 and 5.2 millions of pixels are allocated in the training, validation and test partitions. The majority of these pixels correspond to *Open Forests* followed by *Dense Forests*. In contrast, *Barren* followed by *Shrublands* are the classes in the minority.

Table 2 Class distribution of the hybrid dataset according to the tile-based partitions. Each pixel is at the scale of interest, in this case at spatial resolution of the target deforested dataset, Terra-i (250-m).

Class ID	Class name	Train	Validation	Test
0	No data	5,996,132	1,647,179	1,461,895
1	Barren	449	238	1,865
2	Water Bodies	412,011	31,168	103,675
3	Urban/Built-up Lands	36,053	5,126	10,385
4	Dense Forests	7,441,023	2,168,514	1,782,311
5	Open Forests	8,890,281	1,903,823	2,330,114
6	Natural Herbaceous	2,980,906	778,242	871,252
7	Herbaceous Croplands	335,634	101,181	72,967
8	Shrublands	7,223	49	1,056
Total (N pixels)		26,099,712	6,635,520	6,635,520
Total (N pixels) excl. No data		20,103,580	4,988,341	5,173,625

Appendix IV: Transfer learning experiments

Transfer learning experiments were conducted by retraining the best calibrated MTLCC network described in **Section 5.5.2**. The experiments were based on training data, 2001, 2002 and 2003 MODIS satellite observations and LC reference data derived from 2001-2015 MODIS FAO-LCCS2 layer obtained from three geographically different areas in the pantropics. In total, six tiles of 384 pixels per 384 pixels were distributed in surrounding areas of (1) Bogota, Colombia in South America; (2) Lisala, Democratic Republic of the Congo in Africa; and (3) Pontianak, Indonesia in Asia (1). Whilst, the areas in Africa and Asia shared a similar type of *Dense Forest*, mostly humid evergreen forest, and low urban density with the Amazon region, the area in South America is dominated by evergreen mountain forests and dense urban areas. The number of tiles per location was defined according to a split ratio of 4:1:1 used to create the partitions for training the MTLCC model across the Amazon region as described in **Section 0**. This ratio means 4 tiles were used for training, one for validation and one for testing. Both quantitative and qualitative (i.e. visual inspection) assessments of the predictions were conducted over the test tile per location. **Figure 1** illustrates the distribution of the partition tiles per location overlapped over a map of LC reference data derived from 2001-2015 MODIS FAO-LCCS2 layer. For the strategies of training the MTLCC network from scratch and retrain the MTLCC network fine-tuned in the Amazon, the train and validation tiles were used for training the model over 30 epochs. The test tile was used then to evaluate the performance of the trained models, including the MTLCC network fine-tuned in the Amazon without retraining. Both quantitative and qualitative (i.e. visual inspection) assessments of the predictions were carried out over the test tile per location. **Table 1** summarises the quantitative results for four metrics assessed per location for three consecutive years, 2001, 2002 and 2003. Overall, the strategy of retraining the pre-trained MTLCC network achieves the best performance in comparison with the other strategies. This was reflected in higher values in the performance metrics. The strategy to retrain the MTLCC calibrated model provided better results than training the MTLCC network from scratch. This result needs further investigation as the values of F1-score for the retrained MTLCC network varied considerably by location. For instance, the test tile in Bogota, Colombia obtained lower F1-score values (less than 0.44) than the other locations

(above 0.50). This result can be explained due to the landscape in Bogota, Colombia is remarkably different i.e. mountain evergreen forest with high urban density to the area where the pre-trained MTLCC network was trained.

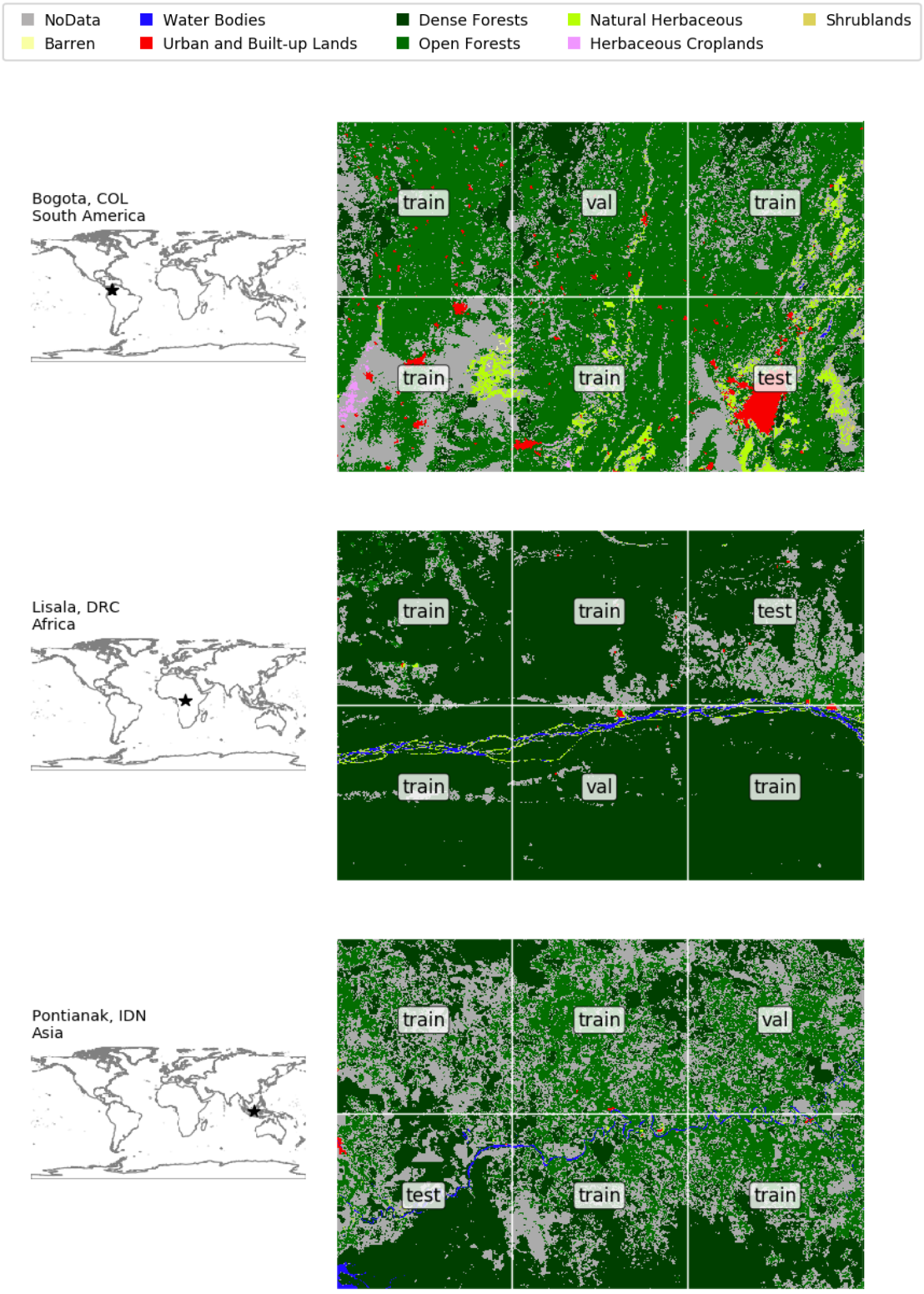


Figure 1 Distribution of the partition tiles across three geographical locations. *NoData* represents pixels not analysed during either training or evaluation of the MTLCC model.

Table 1 Performance metrics the pre-trained MTLCC model with/without retraining and the MTLCC network trained from scratch over 30 epochs. Bold highlights the best values per metric per location.

a) Bogota, Colombia in South America

Metric	Scratch			Pre-trained without retraining			Pre-trained with retraining		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
OA	0.41	0.38	0.49	0.69	0.71	0.71	**0.77**	**0.80**	**0.79**
F1-Score	0.16	0.15	0.18	0.39	0.33	0.31	**0.38**	**0.44**	**0.38**
Quant. disagm	40.17	51.99	36.67	14.84	13.17	12.01	**12.01**	**10.21**	**8.58**
Alloc. disagm	18.56	9.89	14.81	16.40	15.53	16.75	**10.84**	**9.75**	**12.35**

b) Lisala, Democratic Republic of the Congo in Africa

Metric	Scratch			Pre-trained without retraining			Pre-trained with retraining		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
OA	0.32	0.54	0.52	0.78	0.89	0.78	**0.94**	**0.94**	**0.95**
F1-Score	0.14	0.19	0.18	0.49	0.53	0.48	**0.55**	**0.50**	**0.51**
Quant. disagm	67.05	43.75	45.74	21.92	10.25	21.15	**3.54**	**5.37**	**4.22**
Alloc. disagm	0.85	1.81	2.19	**0.48**	1.16	0.85	2.61	**0.42**	**0.72**

c) Pontianak, Indonesia in Asia

Metric	Scratch			Pre-trained without retraining			Pre-trained with retraining		
	2001	**2002**	**2003**	**2001**	**2002**	**2003**	**2001**	**2002**	**2003**
OA	0.87	0.92	0.88	0.92	0.94	0.93	**0.95**	**0.95**	**0.95**
F1-Score	0.26	0.27	0.24	0.52	0.57	0.62	**0.55**	**0.63**	**0.63**
Quant. disagm	4.26	4.73	4.23	7.50	4.65	5.90	**0.78**	**3.24**	**2.39**
Alloc. disagm	8.77	3.42	7.94	**0.86**	**1.54**	**1.21**	4.12	1.99	2.52

Appendix V: Raw and post-processed 2001-2019 LC data

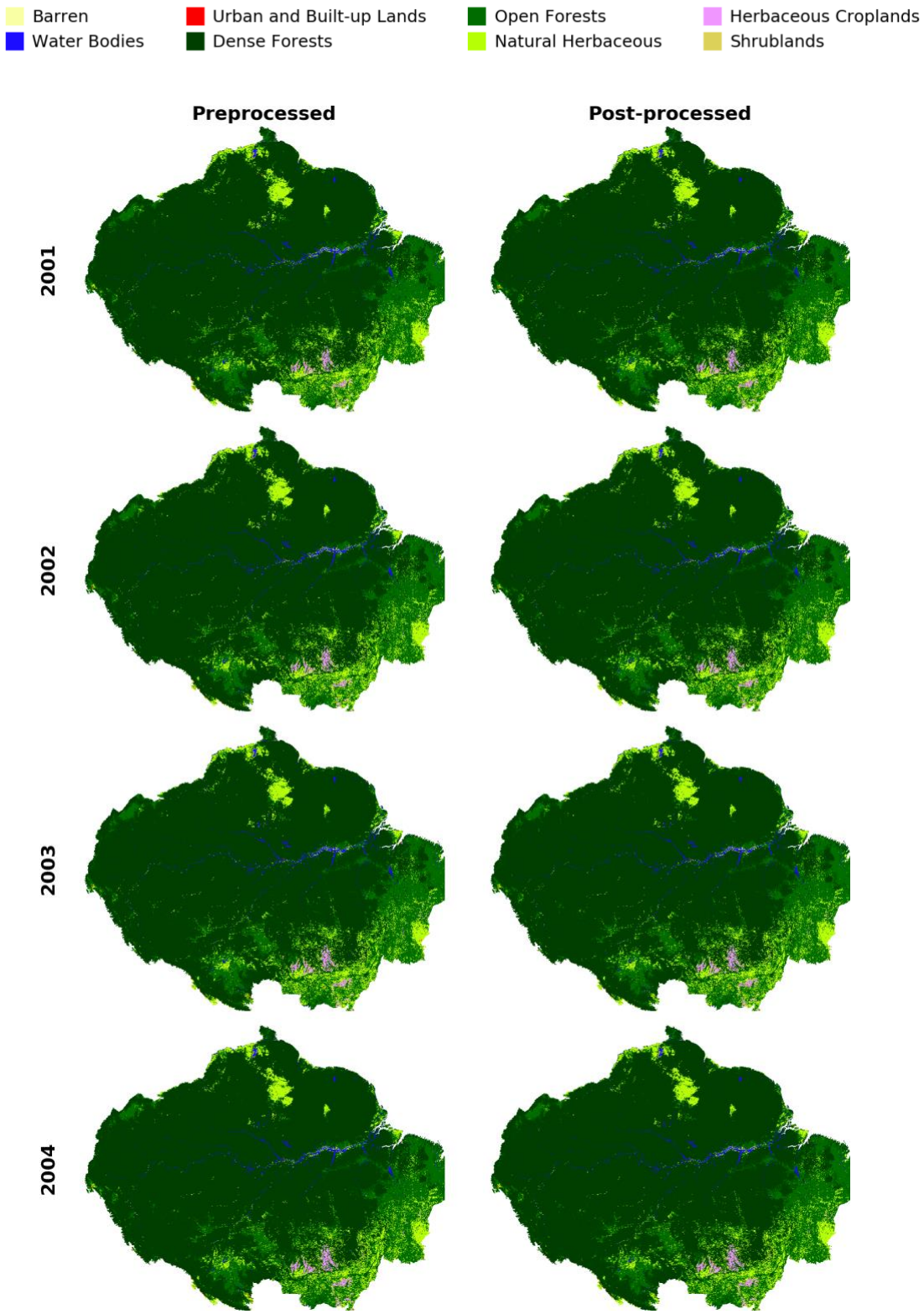


Figure 1 Raw and post-processed 2001-2004 LC maps over the Amazon region.

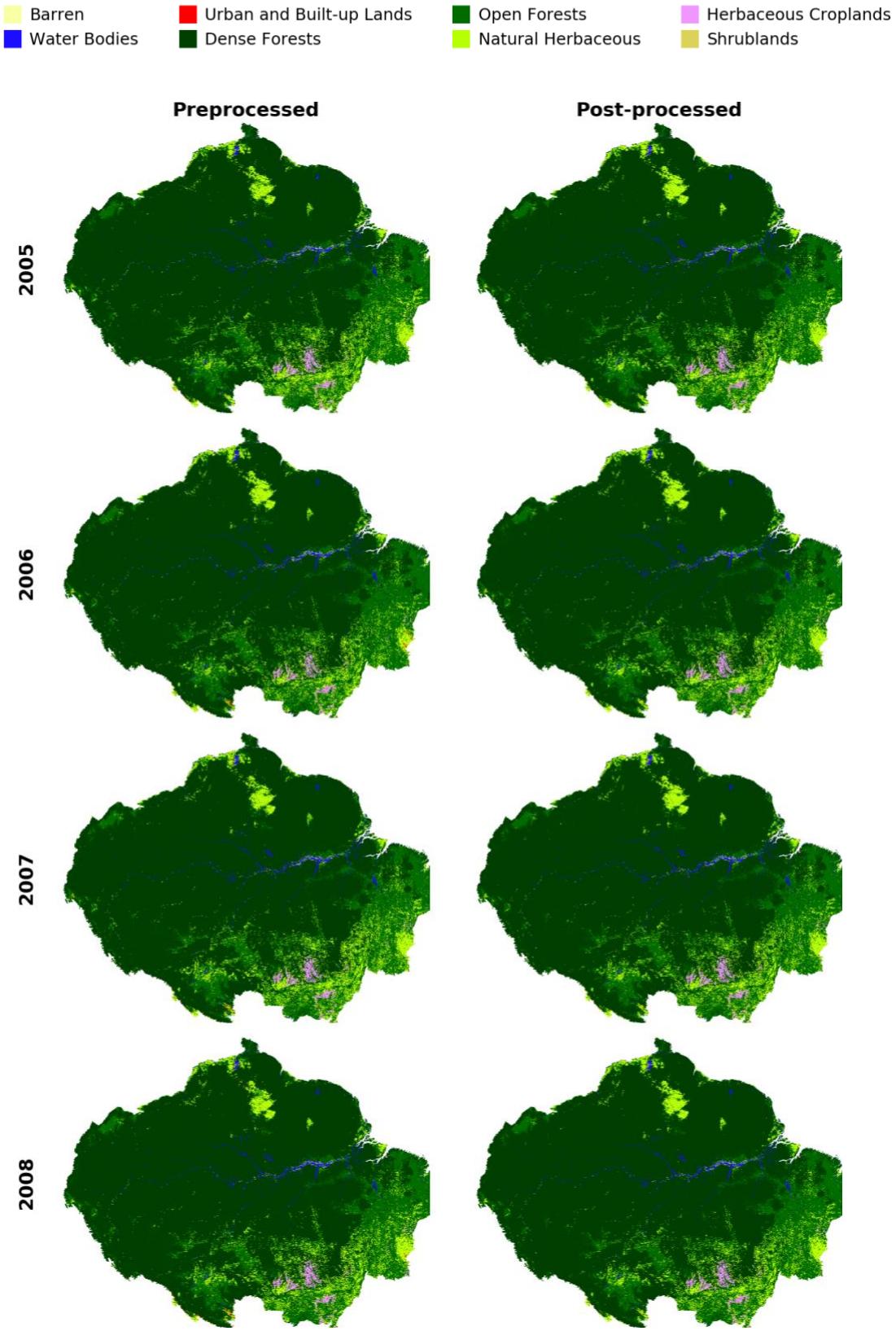


Figure 2 Raw and post-processed 2005-2008 LC maps over the Amazon region.

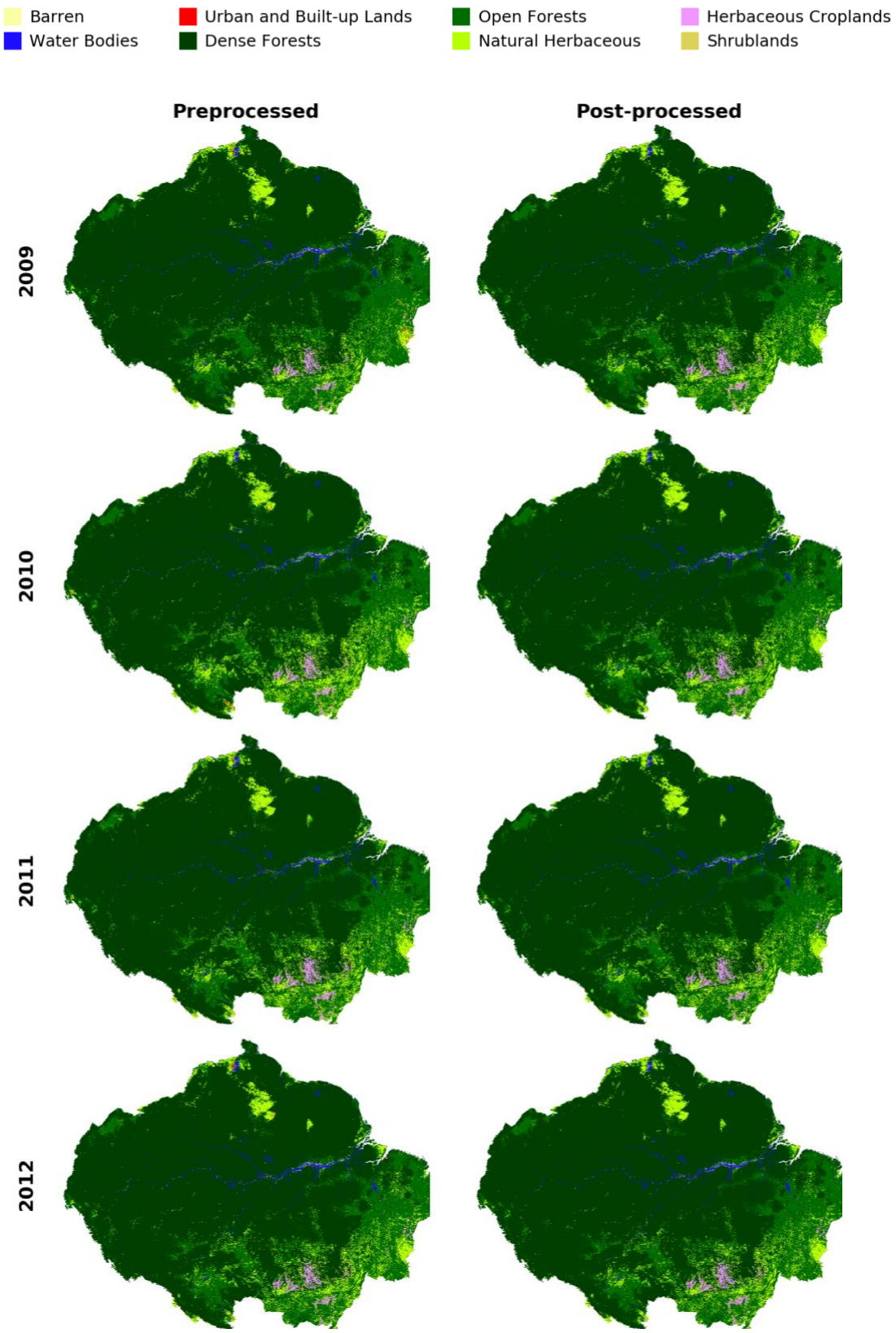


Figure 3 Raw and post-processed 2009-2012 LC maps over the Amazon region.

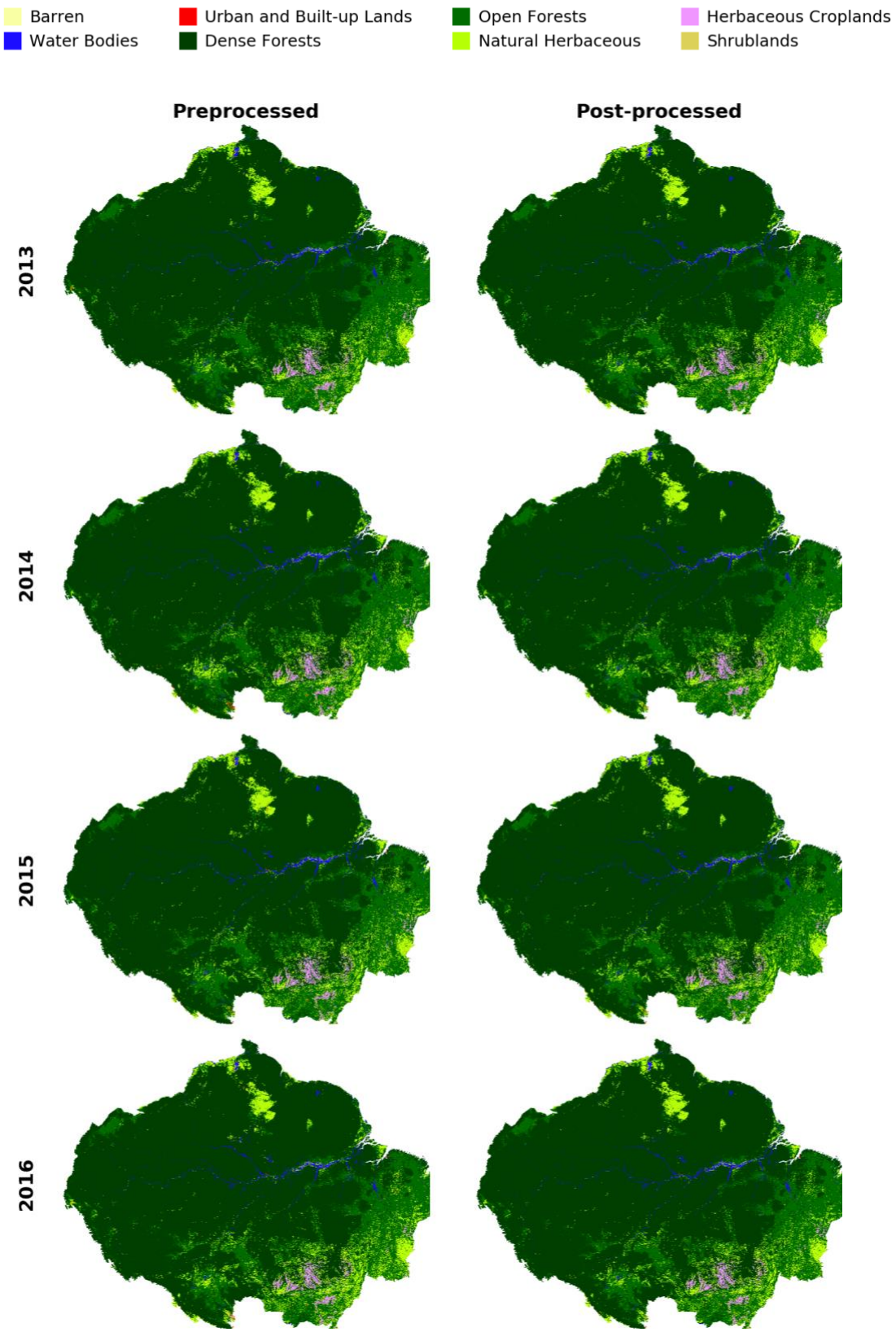


Figure 4 Raw and post-processed 2013-2016 LC maps over the Amazon region.

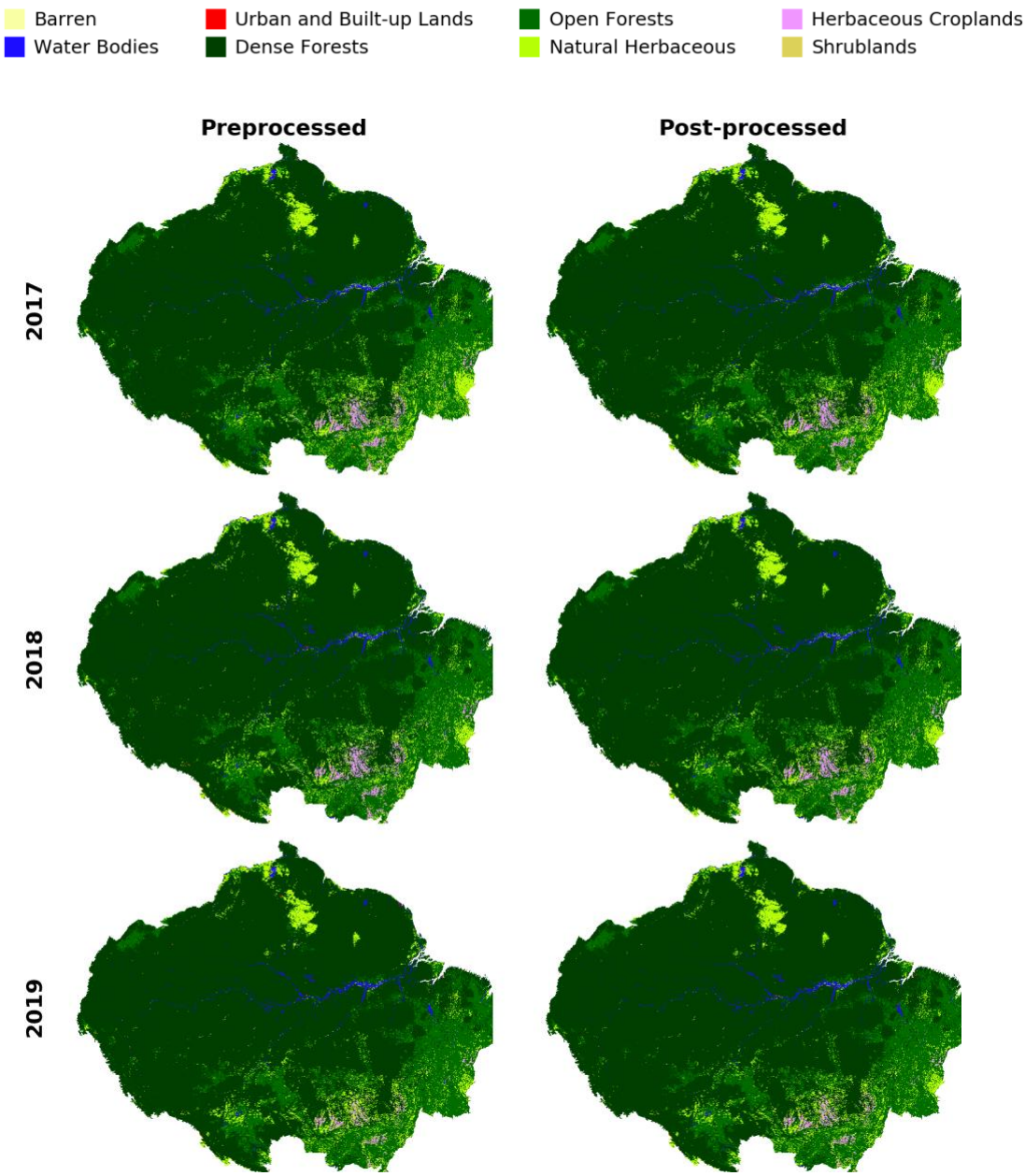


Figure 5 Raw and post-processed 2017-2019 LC maps over the Amazon region.

Appendix VI: LC distribution, stable and unstable sequences

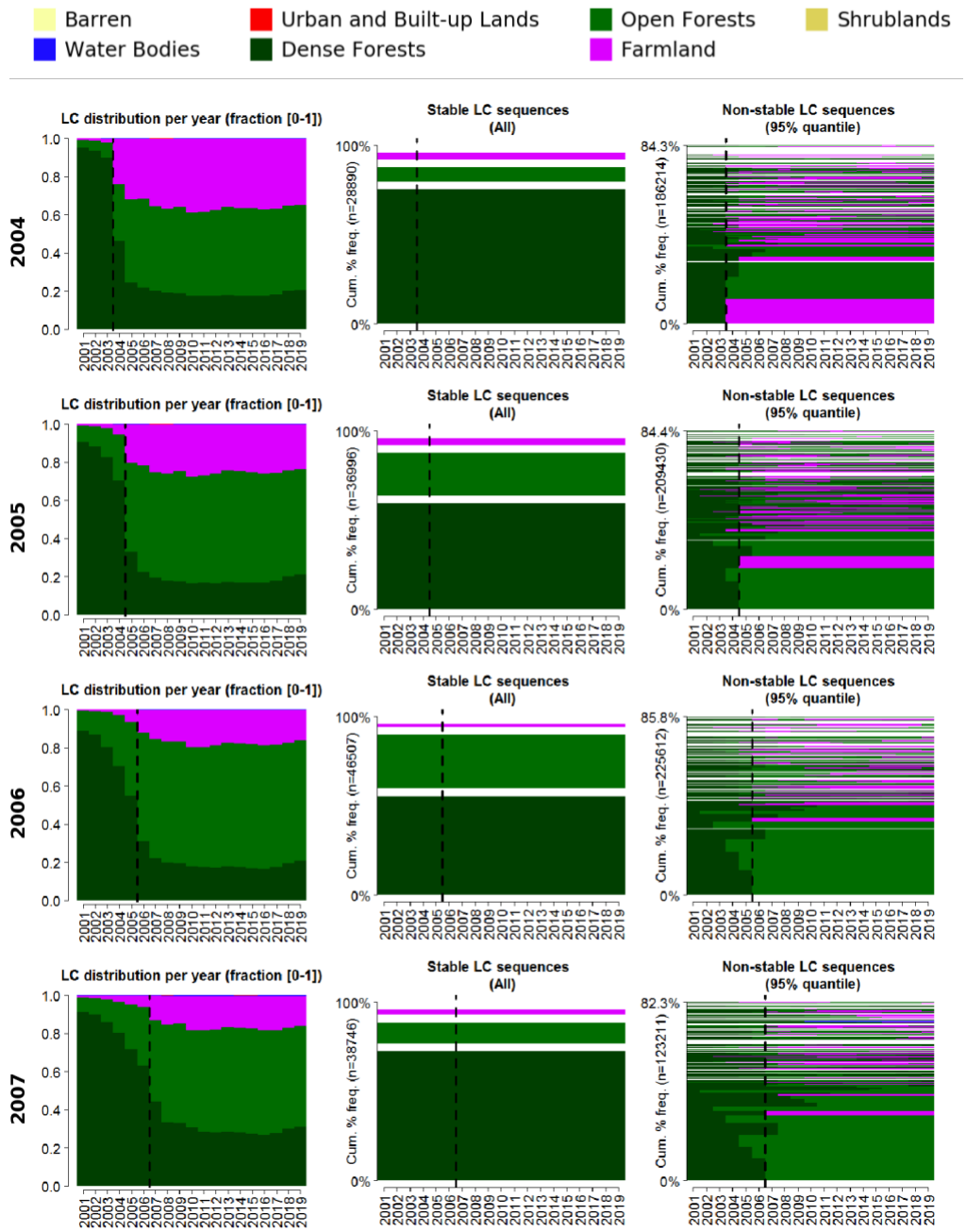


Figure 1 LC distribution, stable and unstable sequences over the period 2001-2019 for deforested areas located across the Amazon humid forest. The target deforested year is highlighted by a dashed vertical line and annotated in the Y-axis of each row.

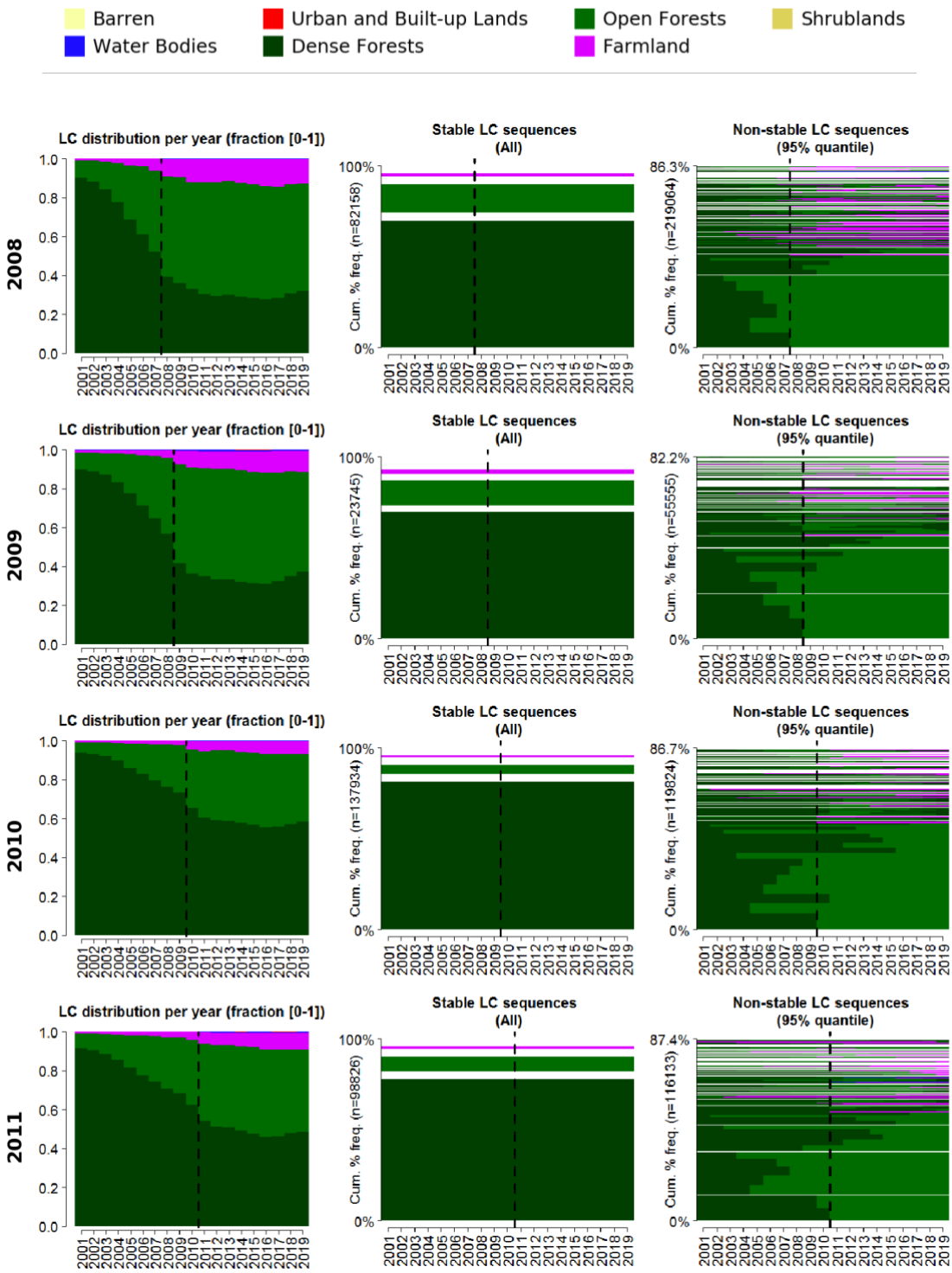


Figure 2 LC distribution, stable and unstable sequences over the period 2001-2019 for deforested areas located across the Amazon humid forest. The target deforested year is highlighted by a dashed vertical line and annotated in the Y-axis of each row.

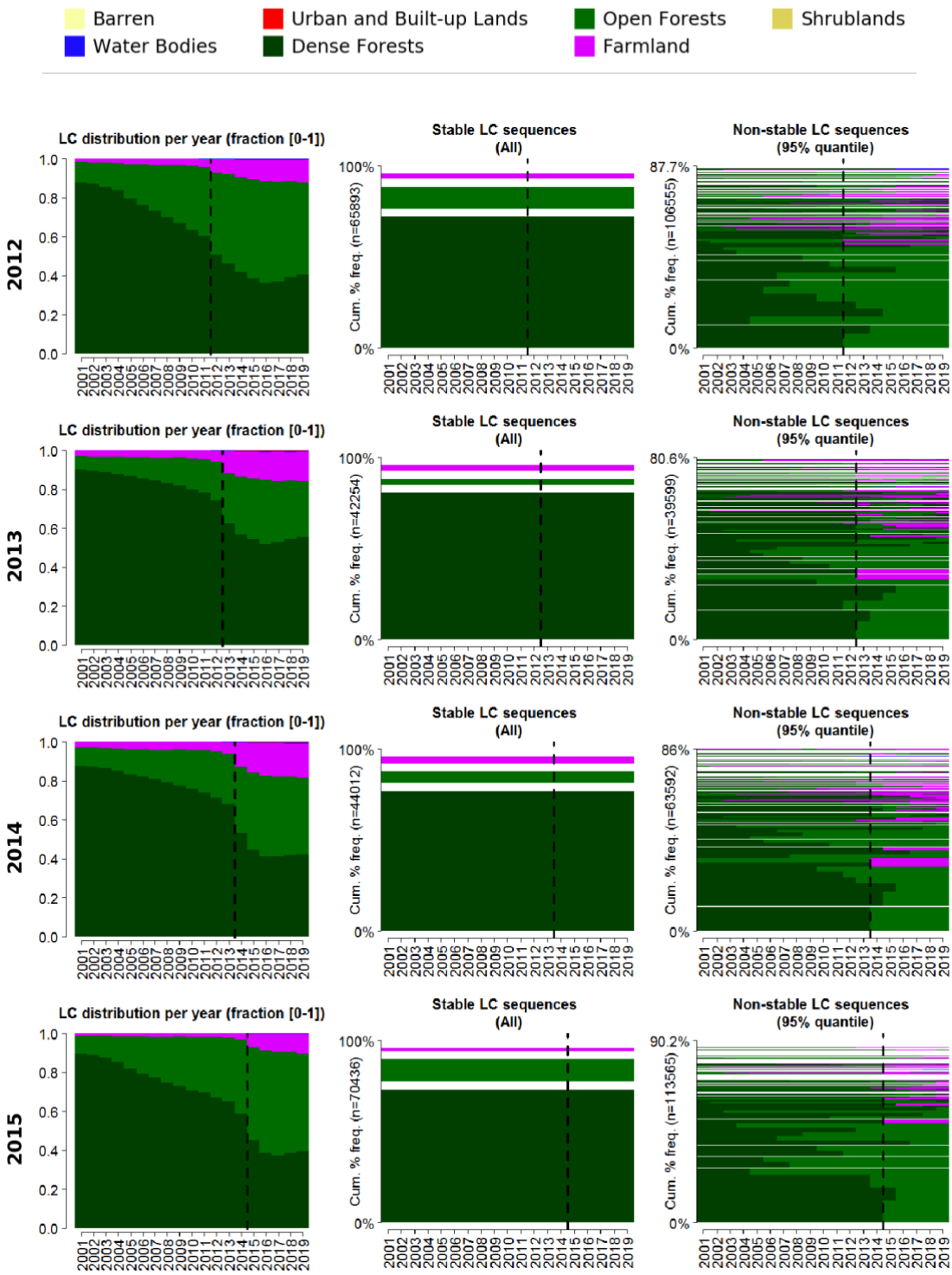


Figure 3 LC distribution, stable and unstable sequences over the period 2001-2019 for deforested areas located across the Amazon humid forest. The target deforested year is highlighted by a dashed vertical line and annotated in the Y-axis of each row.

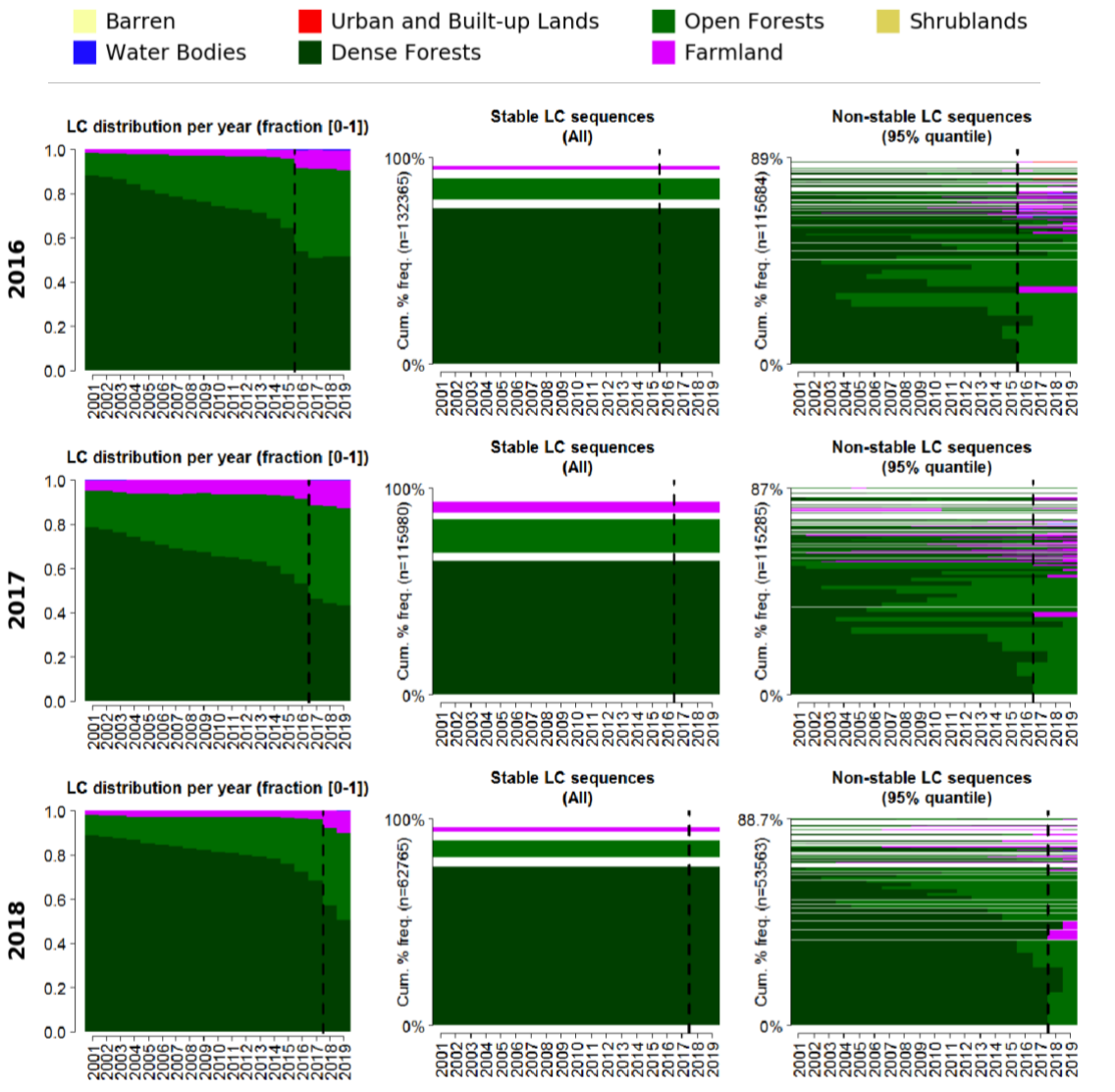


Figure 4 LC distribution, stable and unstable sequences over the period 2001-2019 for deforested areas located across the Amazon humid forest. The target deforested year is highlighted by a dashed vertical line and annotated in the Y-axis of each row.

Appendix VII: Solutions to the minimum period and number of clusters

Best solution - Minimum Period = 12 / Number of clusters = 6

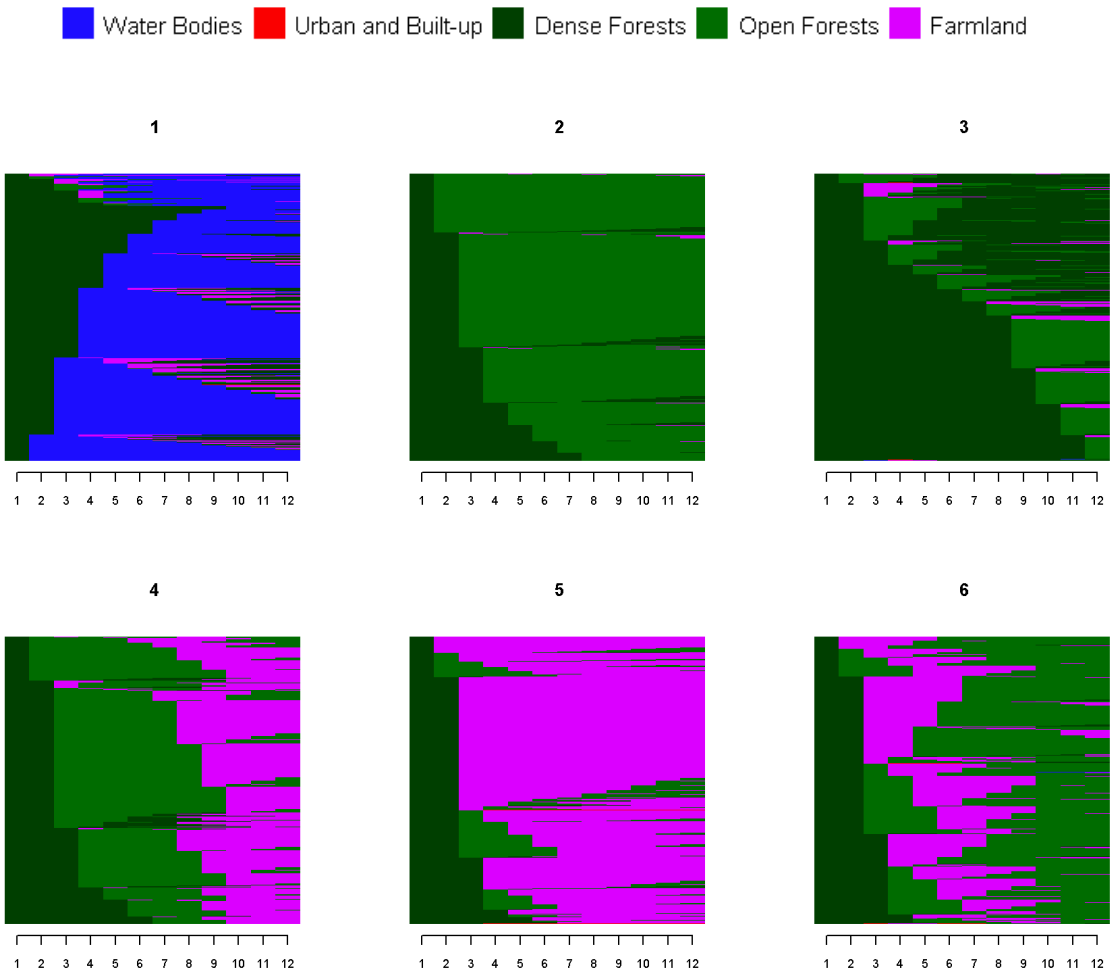


Figure 1a Sequence index plots showing post-loss LC sequences according to the number of clusters and minimum period of the best solution.

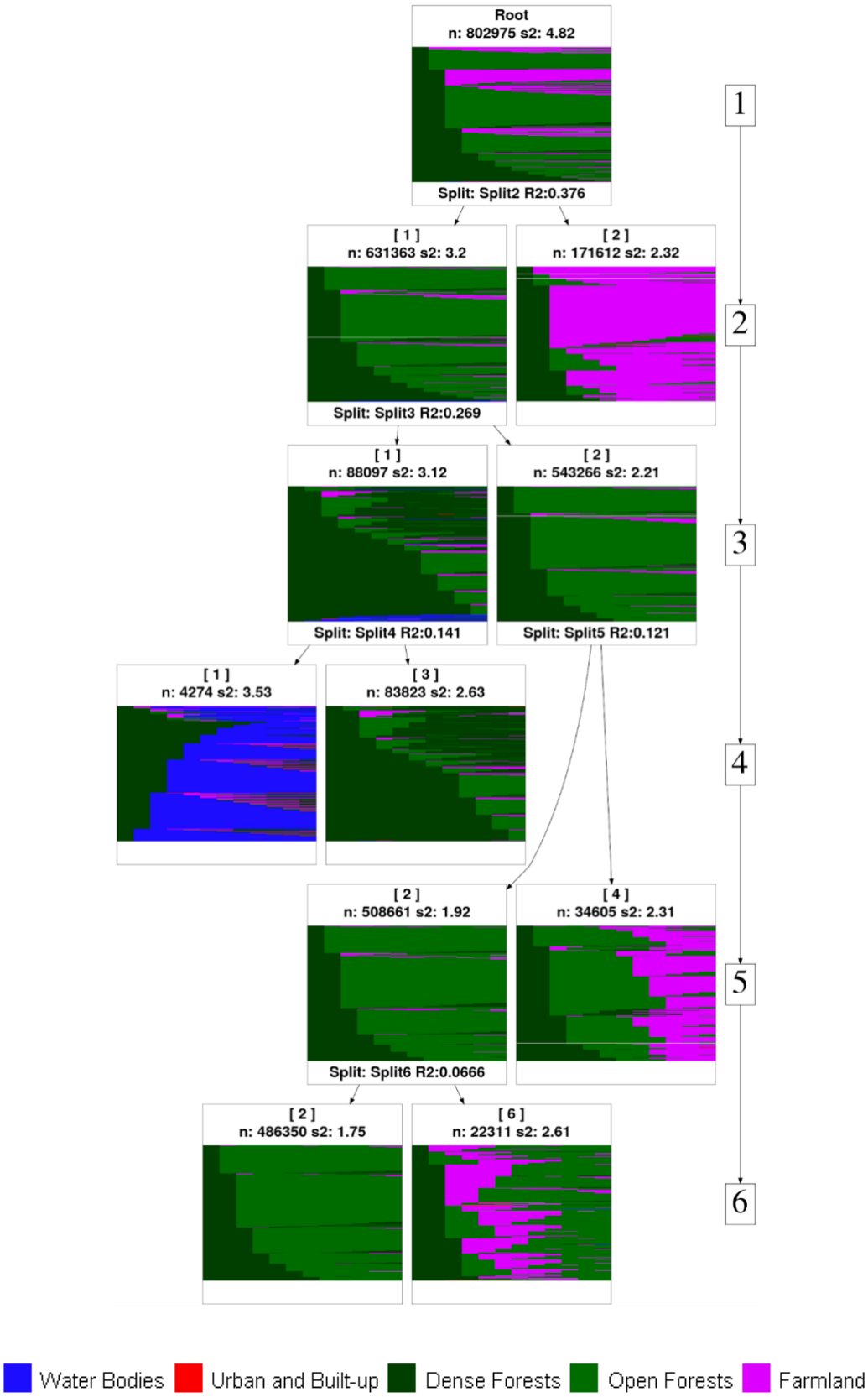


Figure 1b Tree-structure diagram showing how cluster partitions for the best solution (minimum period of 12 and 6 clusters) were generated. The number of sequences per partition is denoted by n.

Alternative solution - Minimum Period = 8 / Number of clusters = 6

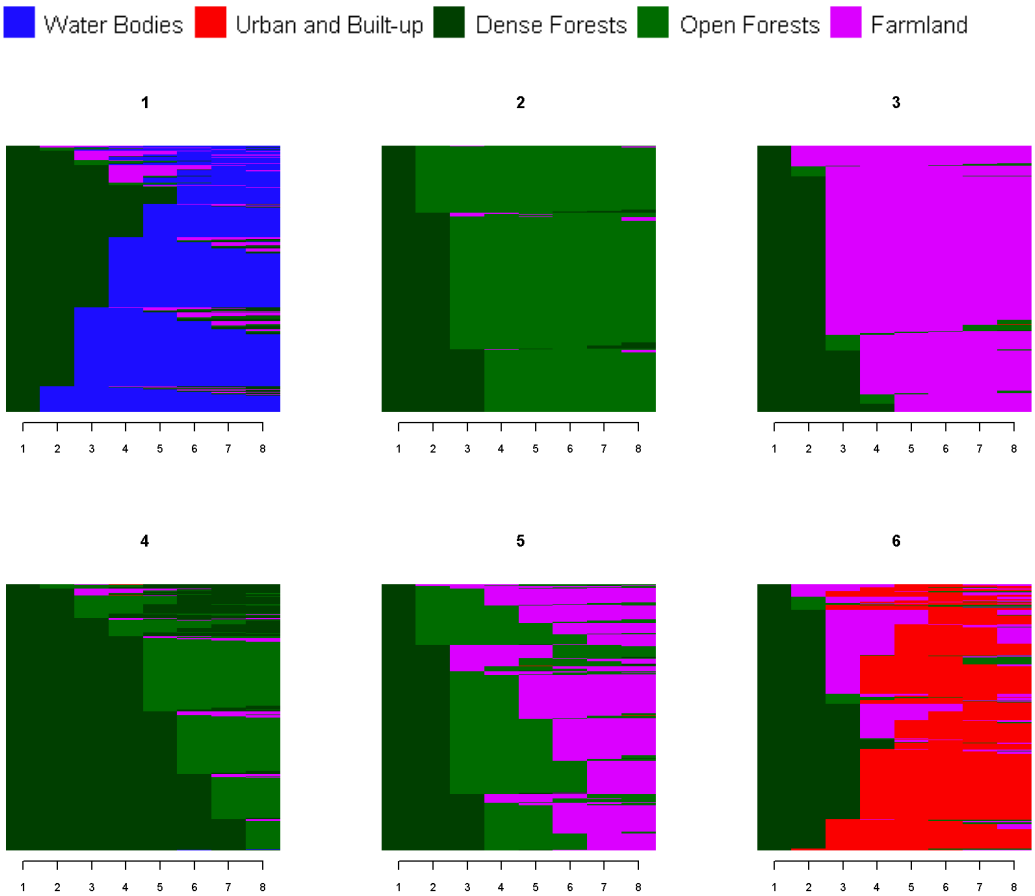


Figure 2a Sequence index plots showing post-loss LC sequences according to the alternative solution (minimum period of 8 and 6 clusters).

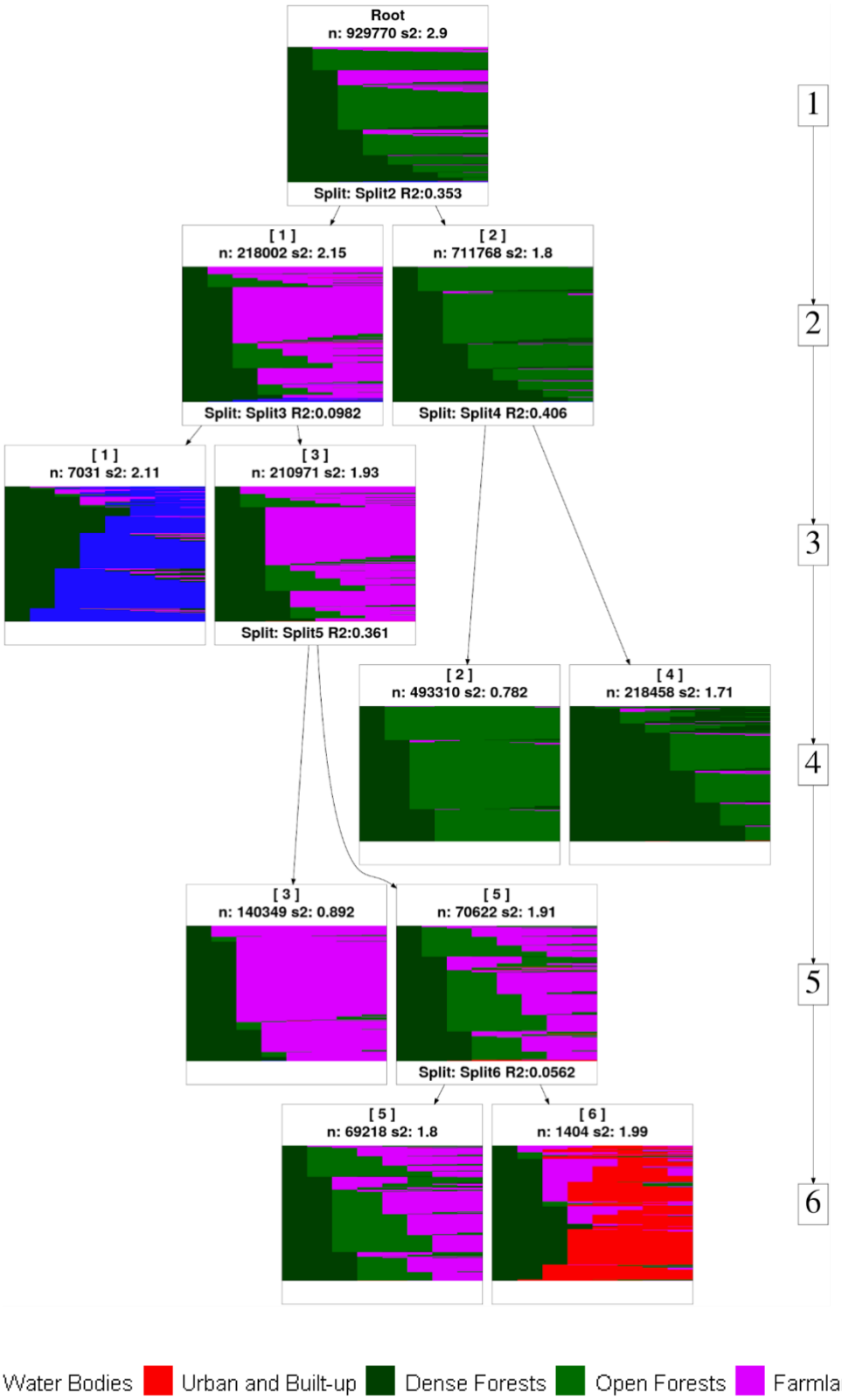


Figure 2b Tree-structure diagram showing how cluster partitions for the best solution (minimum period of 8 and 6 clusters) were generated. The number of sequences per partition is denoted by n.

Alternative solution - Minimum Period = 14 / Number of clusters = 6

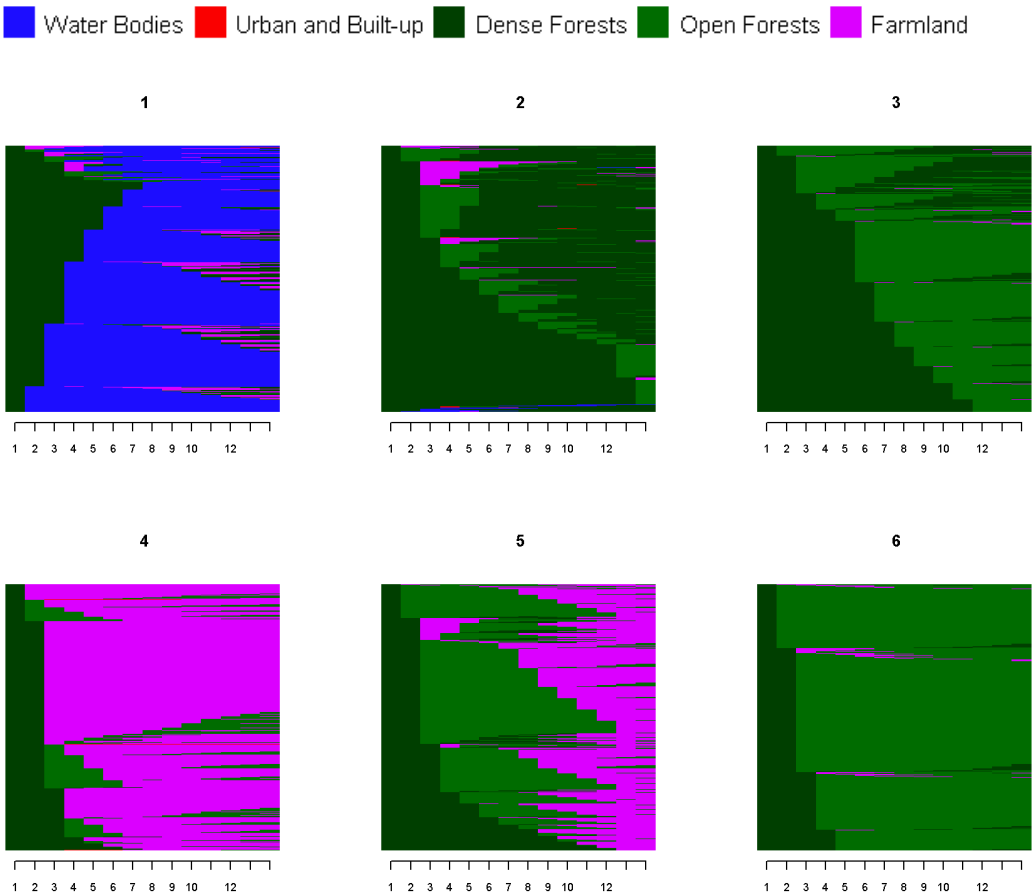


Figure 3a Sequence index plots showing post-loss LC sequences according to the alternative solution (minimum period of 14 and 6 clusters).

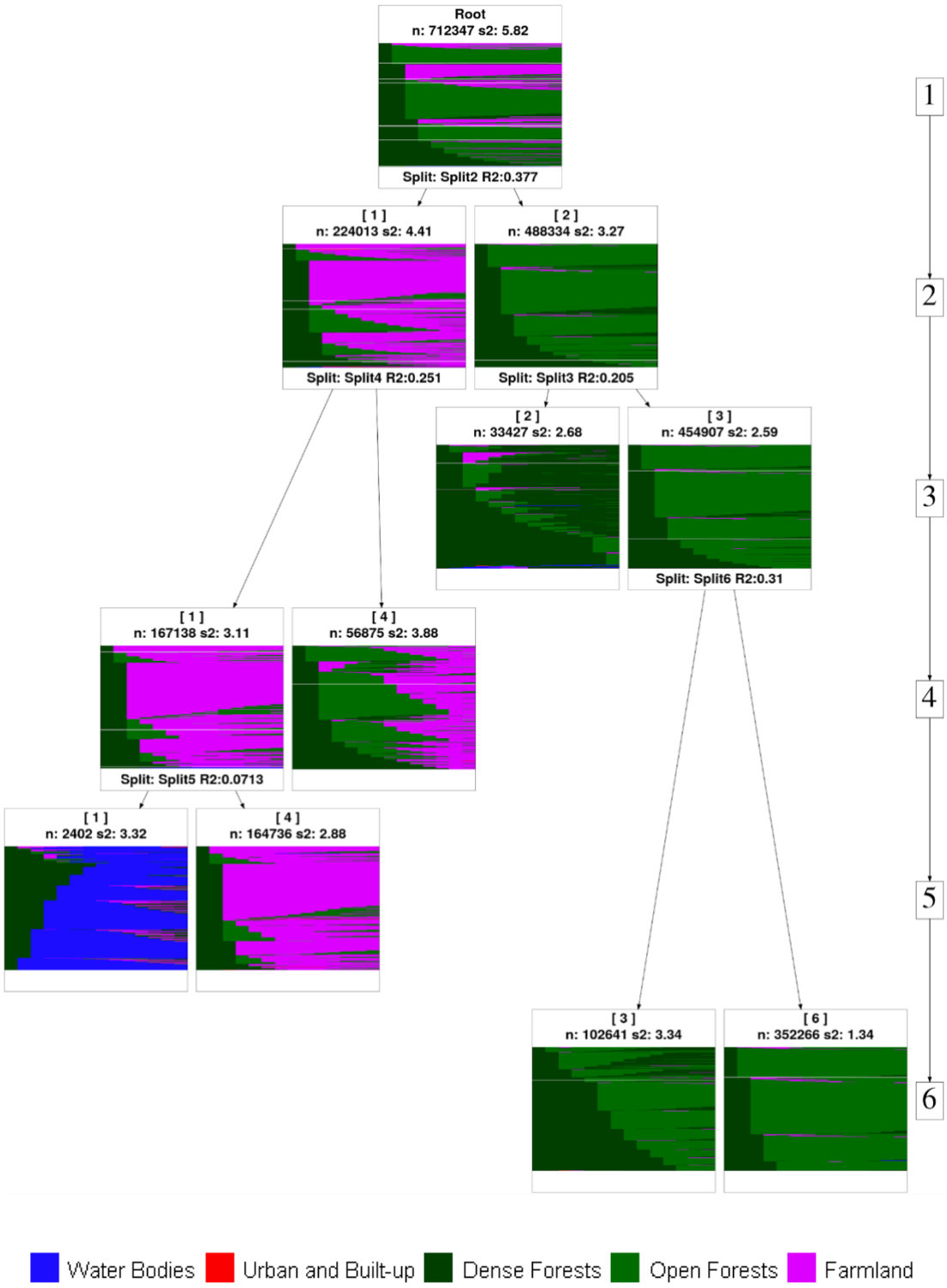


Figure 3b Tree-structure diagram showing how cluster partitions for the best solution (minimum period of 8 and 6 clusters) were generated. The number of sequences per partition is denoted by n.

Alternative solution - Minimum Period = 16 / Number of clusters = 6

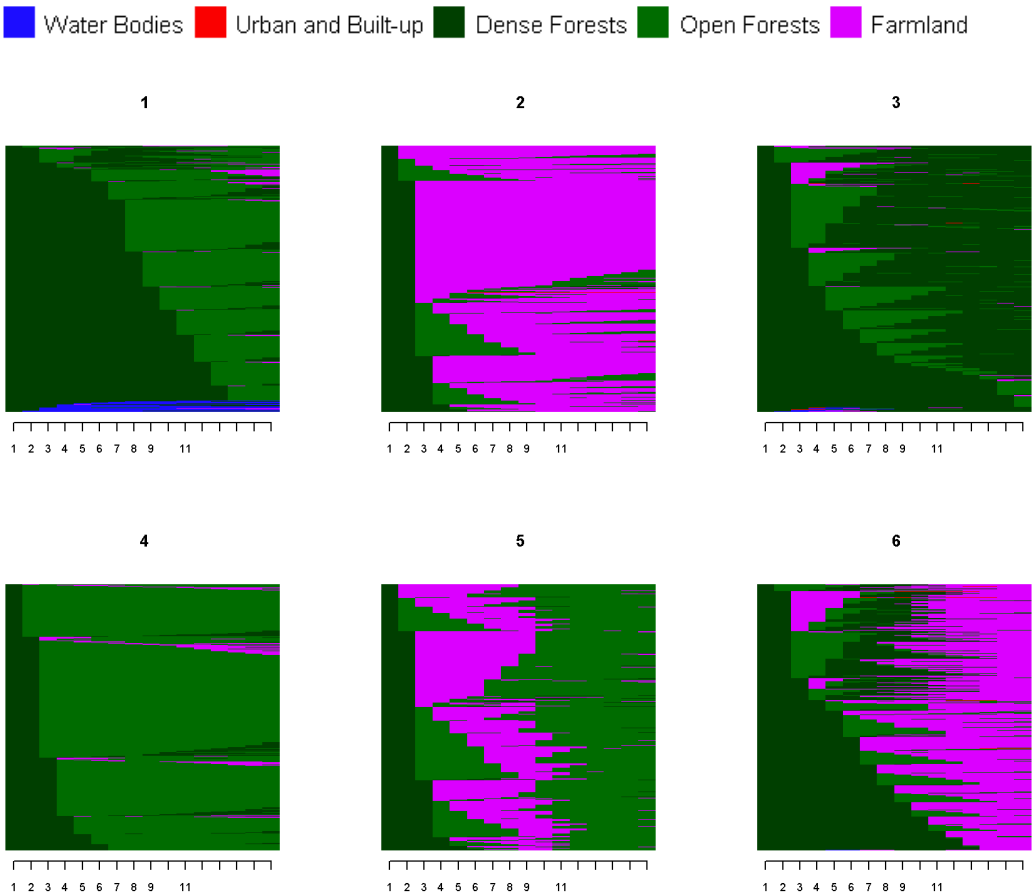


Figure 4a Sequence index plots showing post-loss LC sequences according to the alternative solution (minimum period of 14 and 6 clusters).

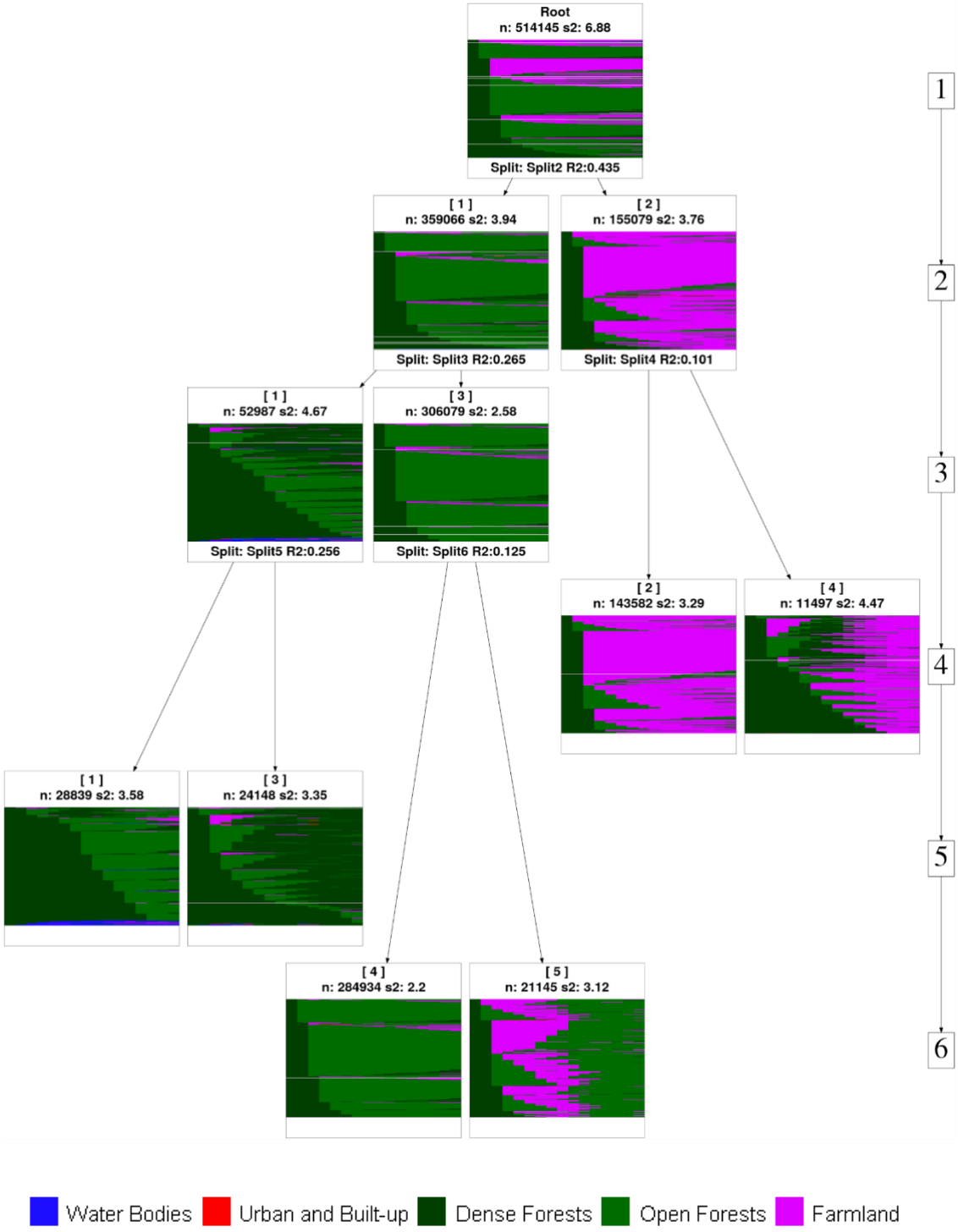


Figure 4b Tree-structure diagram showing how cluster partitions for the best solution (minimum period of 8 and 6 clusters) were generated. The number of sequences per partition is denoted by n.

Appendix VIII: Statistical analysis

Table 1 Median (MD) and interquartile range (IQR) computed by post-loss LC trajectory group across the internal and external variables.

a) Internal variables

Post-loss LC group	Variable and stats									
	Prox. to Water (km)		Prox. to Urban/Built-up (km)		Prox. to Dense Forests (km)		Prox. to Open Forests (km)		Prox. to Farmland (km)	
	MD	*IQR*	*MD*	*IQR*	*MD*	*IQR*	*MD*	*IQR*	*MD*	*IQR*
Water Persistent	0.0	0.5	99.5	137.5	0.5	0.5	3.2	3.5	2.1	4.6
Open Forest Persistent	74.0	75.1	286.2	280.2	0.5	0.7	0.0	0.0	3.9	10.2
Dense Forest Persistent	77.5	79.5	267.2	258.4	0.2	0.5	0.2	1.2	2.6	7.0
Open Forest-Farmland	80.9	66.3	349.5	253.5	0.9	0.9	0.2	0.7	0.0	1.4
Farmland Persistent	81.6	67.0	380.3	172.8	0.9	0.7	0.5	1.2	0.0	0.7
Farmland-Open Forest	85.6	71.9	328.6	197.6	0.9	0.9	0.0	0.2	2.1	6.3

b) External variables

Post-loss LC group	Variable and stats											
	Inaccess. (days)		Elevation (m)		Slope (angle)		Precip. (mm)		Prox. to conser. Pas (km)		Prox. to exploit. PAs (km)	
	MD	*IQR*	*MD*	*IQR*	*MD*	*IQR*	*MD*	*IQR*	*MD*	*IQR*	*MD*	*IQR*
Water Persistent	0.3	0.4	161	120	0.7	0.8	1813	186	104.4	47.8	104.4	47.8
Open Forest Persistent	0.2	0.2	237	142	1.6	1.7	1992	324	107.6	118.0	107.6	118.0
Dense Forest Persistent	0.2	0.2	242	180	1.3	1.5	1979	381	103.9	117.6	103.9	117.6
Open Forest-Farmland	0.2	0.2	300	175	1.2	1.2	1945	319	100.9	117.6	100.9	117.6
Farmland Persistent	0.2	0.2	324	144	1.1	1.1	1948	243	112.9	116.0	112.9	116.0
Farmland-Open Forest	0.2	0.2	285	141	1.4	1.4	1934	261	129.9	106.0	129.9	106.0

Appendix IX: Socialisation

Conference proceedings

COCA-CASTRO, A., M. RUSSWURM, L. REYMONDIN & M. MULLIGAN. 2019. Sequential Recurrent Encoders For Land Cover Mapping In The Brazilian Amazon Using MODIS Imagery And Auxiliary Datasets. In: 23rd IEEE Geoscience and Remote Sensing Society (IGARSS) conference - Poster - Session: WEP2.PT Big Data and Machine Learning - Machine Learning for Land Application, Yokohama, Japan. https://doi.org/10.1109/IGARSS.2019.8899114

COCA-CASTRO, A., M. RUSSWURM & M. MULLIGAN. 2019. An Exploration of Convolutional Recurrent Networks for Large-Scale Land Cover Prediction Using Modis Archives. In: Proceedings of 2019 Big Data from Space (BiDS'19) - Poster - Session: Machine Learning and Artificial Intelligence, Munich, Germany.

Conference abstracts

COCA-CASTRO, A., M. RUSSWURM & M. MULLIGAN. 2020. Multi-temporal mapping of pantropical post-loss land cover using dense earth-observation time series and global pre-existing maps. In: EGU2020: Sharing Geoscience Online - Poster presentation -Session: Tropical landscapes and peatlands: Biogeochemistry, ecohydrology and land use impacts, Vienna, Austria

COCA-CASTRO, A., M. RUSSWURM & M. MULLIGAN. 2018. Towards multi-temporal data-driven models for extracting and learning information from remote sensing time series and existing ancillary data for land cover classification. In: EGU General Assembly 2018 - Oral presentation -Session: Information extraction from satellite observations using data-driven methods, Vienna, Austria

COCA-CASTRO, A., L. REYMONDIN, J. REBETEZ, H. F. SATIZABAL-MEJIA, A. PEREZ-URIBE, M. MULLIGAN, T. SMITH & H. HYMAN. 2017. Big earth observation data analytics for modelling pan-tropical land-use change trajectories for newly deforested areas. In: EGU General Assembly 2017- PICO session: IE3.4/ESSI2.12 Media Monitoring the Sustainable Development Goals with the huge Remote Sensing archives, Vienna, Austria.

Lightning talks

COCA-CASTRO, A. 2018. Advanced classification of remote sensing imagery using GEE and Tensorflow. In: 2018 Google Earth Engine User Summit. Dublin, Republic of Ireland. Online: https://www.youtube.com/watch?v=oK7E4oJuZmA

Repositories

COCA CASTRO, A. 2019. acocac/MTLCC-MODIS-GCP: First public pre-release (Version 1.0). MTLCC-MODIS-GCP contains scripts to download/pre-process MODIS data to train a Multitemporal Land Cover Classification (MTLCC) Network using the Google Cloud Infrastructure. In: Zenodo, http://doi.org/10.5281/zenodo.3604771.

Others

Author & co-author of the following short blog posts published in the blog of the Earth Observation and Environmental Sensing (EOES) Hub at King's College London, https://blogs.kcl.ac.uk/eoes/.

- 2018 Earth Engine User Conference Report;
- Data-driven revolution in Earth Observation (EO) research;
- Monitoring the Sustainable Development Goals with the huge Remote Sensing archives.