MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/08.11.11.21-TDI

DETECTION AND ANALYSIS OF FOREST REGENERATION TRAJECTORIES IN THE LOWER TAPAJÓS REGION

Mariane Souza Reis

Doctorate Thesis of the Graduate Course in Earth System Science, guided by Drs. Maria Isabel Sobral Escada, and Sidnei João Siqueira Sant'Anna, approved in August 25, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/47E2TRB>

INPE
São José dos Campos
2022

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/08.11.11.21-TDI

DETECTION AND ANALYSIS OF FOREST REGENERATION TRAJECTORIES IN THE LOWER TAPAJÓS REGION

Mariane Souza Reis

Doctorate Thesis of the Graduate Course in Earth System Science, guided by Drs. Maria Isabel Sobral Escada, and Sidnei João Siqueira Sant'Anna, approved in August 25, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/47E2TRB>

INPE
São José dos Campos
2022

Cataloging in Publication Data

Reis, Mariane Souza.

R277d Detection and analysis of forest regeneration trajectories in the lower Tapajós region / Mariane Souza Reis. – São José dos Campos : INPE, 2022.

xxxii + 222 p. ; (sid.inpe.br/mtc-m21d/2022/08.11.11.21-TDI)

Thesis (Doctorate in Earth System Science) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2022.

Guiding : Drs. Maria Isabel Sobral Escada, and Sidnei João Siqueira Sant'Anna.

1. Secondary vegetation. 2. Trajectory classification. 3. Amazon. I.Title.

CDU 630*23(811)

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE TESE DE MARIANE SOUZA REIS
BANCA Nº088/2022 REG 121363/2017

No dia 25 de agosto de 2022, as 14h00min, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutora em Ciência do Sistema Terrestre. O trabalho deve incorporar as correções sugeridas pela Banca Examinadora, a critério e sob revisão final do(s) orientador(es).

Novo título: "DETECTION AND ANALYSIS OF FOREST REGENERATION TRAJECTORIES IN THE LOWER TAPAJÓS REGION"

Dr. Silvana Amaral Kampel - Presidente - INPE
Dr. Maria Isabel Sobral Escada - Orientador – INPE
Dr. Sidnei João Siqueira Sant'Anna - Orientador – INPE
Dr. Camilo Daleles Rennó - Membro Interno - INPE
Dr. Luiz Eduardo Oliveira e Cruz de Aragão - Membro Interno – INPE
Dra. Ana Catarina Conte Jakovac – Membro Externo - UFSC
Dr. Daniel E Silva – Membro Externo - WWF Brasil

Documento assinado eletronicamente por **Sidnei João Siqueira SantAnna**, **Tecnologista**, em 30/08/2022, às 14:16 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Silvana Amaral Kampel**, **Pesquisadora**, em 30/08/2022, às 16:35 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Camilo Daleles Rennó**, **Tecnologista**, em 30/08/2022, às 18:42 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Maria Isabel Sobral Escada**, **Tecnologista**, em 31/08/2022, às 13:12 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **ana catarina conte jakovac (E)**, **Usuário Externo**, em 16/09/2022, às 14:37 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel E Silva (E)**, **Usuário Externo**, em 08/10/2022, às 18:16 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Luiz Eduardo Oliveira E Cruz de Aragão**, **Chefe da Divisão de Observação da Terra e Geoinformática**, em 11/10/2022, às 14:32 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.mcti.gov.br/verifica.html, informando o código verificador **9666429** e o código CRC **017D2478**.

Referência: Processo nº 01340.002689/2022-32 SEI nº 9666429

"Though my soul may set in darkness, it will rise in perfect light;

I have loved the stars too truly to be fearful of the night.

What, my boy, you are not weeping? You should save your eyes for sight;

You will need them, mine observer, yet for many another night."

SARAH WILLIAMS
in *"The old astronomer"*, 1868

To the little girl who once dreamed of being a scientist.

ACKNOWLEDGEMENTS

(in Portuguese)

A meus orientadores, Dra. Maria Isabel Sobral Escada e Dr. Sidnei João Siqueira Sant'Anna, pela oportunidade, conversas, amizade, e principalmente paciência durante esse processo.

A Dr. Luciano Vieira Dutra, cuja generosidade foi fundamental tanto na minha formação quanto na proposta e desenvolvimento dos métodos e análises contidos nesse documento, e que foi o primeiro a me dar uma oportunidade no INPE.

Aos membros da banca, Dra. Silvana Amaral Kampel, Dr. Camilo Daleles Rennó, Dr. Luiz Eduardo Oliveira e Cruz de Aragão, Dra. Ana Catarina Conte Jakovac, e Dr. Daniel E Silva, por todas as valiosas sugestões, mas principalmente por terem aceitado participar do processo de avaliação e pela paciência com que lidaram com os contra-tempos.

A todos que me receberam pacientemente em Berlim, em especial: Dr. Patrick Hostert, por ter aceitado participar dessa jornada, Dr. Philippe Rufin, pelas sugestões sempre pertinentes, e Anita Carvalho, conterrânea sempre pronta a me ajudar com os choques culturais e que gentilmente compartilhou suas histórias.

A Angela Harada, Mariana Santos e Dr. Kleber Naccarato, pela disposição, eficiência e todo o auxílio durante essa processo.

A meus colegas de curso, de laboratório e de campo, que sempre me incentivaram e estiveram presentes durante todo o processo com ouvidos amigos, problemas compartilhados e soluções. Em particular, agradeço Tainá O. Assis, pela companhia em todos os prazos e processos, Gabriel Bragion e Juan Doblas, que gentilmente cederam seu tempo e conhecimentos auxiliando na elaboração de códigos, Bruno V. Adorno, Izak F. Justi, Noeli A. P. Moreira e Maria Antônia F. de Oliveira, que jamais se negaram a ajudar na compilação e geração de dados, Luciana P. Trevisan de Souza, pela ajuda com o a diagramação das figuras de linha do tempo, a Renata M. Ribeiro, Sacha Siani, Erick da Silva Santos e Sabrina P. L. P. Corrêa, pelas correções gramaticais de última hora e torcida. Agradeço também Anielli R. de Souza e Danielle S. de Paula, pela gentil e pronta disponibilização dos dados gerados em suas dissertações.

A Janitoni "Louro" Lima e Adão Silva Santos, pelo acolhimento no LBA.

A meus pais, que sempre me apoiaram nessa jornada.

A Osvaldo S. Murahara *sensei*, que me ensinou a sonhar grandes sonhos.

A Daniela Holleben, amiga para todas as horas e situações, que sempre esteve a uma mensagem de distância.

A Akime (*in memoriam*). Você foi uma fortaleza.

A CAPES e DAAD, pela concessão das bolsas de estudos, financiamentos de publicações e trabalhos de campo.

E por fim, a todos os professores, colegas e amigos que participaram da minha formação. Obrigada por viverem esse processo comigo.

ABSTRACT

The analysis of forest regeneration processes, through the classification of remote sensing time series into land use and land cover (LULC) trajectories, provides information to estimate important ecosystem services in the Amazon. However, the current classification methods are either prone to the classification of invalid LULC trajectories and/or demand many processing steps that can hinder the classification of secondary vegetation areas. This study aims to propose a method for the classification and analysis of forest regeneration trajectories based on previous knowledge about land cover dynamics and multi-temporal remote sensing data. This analysis was conducted in a study area in the Lower Tapajós region (Pará state), in four main steps. The first one regards the definition of LULC classes for the study area. We proposed 16 LULC classes described in Land Cover Meta Language (LCML). These classes were grouped into legends with different levels of detail. These were evaluated to determine the most adequate legend to be used in this study. The second step is the proposal of a novel classification algorithm, the Compound Maximum *a Posteriori* (CMAP) classifier. The third step is the classification of 37 years of Landsat images to generate LULC trajectories from 1984 to 2020 in the Lower Tapajós region. Here we compared CMAP to the post-classification comparison approach with and without the use of a post-processing filter to remove invalid transitions. CMAP was able to provide more accurate classifications with only valid transitions. In the fourth step, we used these LULC trajectories to characterize and compare the forest regeneration processes in the Lower Tapajós region. We have found differences in the spatial-temporal pattern of areas concentrating the highest proportions of deforested areas covered by secondary vegetation, the number of regeneration cycles, the persistence of secondary vegetation, and the age of secondary vegetation across the study area. These variations coincide with known shifts in agricultural practices. Riverine areas at the northwest portion of the study area included in agro-extractivist projects present forest regeneration trajectories with a relatively high number of regeneration cycles and persistence of secondary vegetation. They are also amongst the areas with the oldest secondary vegetation observed in the final years of the time series. These results highlight the need for detailed studies of secondary vegetation dynamics in time and space within the Amazon, to improve carbon estimates and enable the proposal of more effective conservation policies, territorial ordering, and sustainable land use practice incentives. Lastly, we identified that up to 11.5% of the deforestation of secondary vegetation in a single year, in this region, occurs in areas regenerating from 11 years or more. These processes can only be fully characterized by the classification of valid LULC trajectories from decades of observations, further evidencing the importance of approaches such as CMAP.

Keywords: Secondary vegetation. Trajectory classification. Amazon.

DETECÇÃO E ANÁLISE DE TRAJETÓRIAS DE REGENERAÇÃO FLORESTAL NA REGIÃO DO BAIXO TAPAJÓS

RESUMO

A análise dos processos de regeneração florestal, por meio da classificação de séries temporais de sensoriamento remoto em trajetórias de uso e cobertura da terra, fornece informações para estimar importantes serviços ecossistêmicos na Amazônia. No entanto, os métodos de classificação atuais são propensos à classificação de trajetórias inválidas e/ou exigem muitas etapas de processamento que podem prejudicar a classificação de áreas de vegetação secundária. Este estudo visou propor um método para a classificação e análise de trajetórias de regeneração florestal com base em conhecimentos prévios acerca da dinâmica da cobertura da terra e dados de sensoriamento remoro multi-temporais. Esta análise foi conduzida em uma área de estudo na região do Baixo Tapajós, no estado do Pará, em quatro etapas principais. A primeira etapa foi a definição de classes de uso e cobertura da terra para a área de estudo. Foram propostas 16 classes de uso e cobertura da terra, descritas em *Land Cover Meta Language* (LCML). Essas classes foram agrupadas em legendas com diferentes níveis de detalhe, que foram avaliadas para determinar a mais adequada para este estudo. O segundo passo foi a proposta de um novo algoritmo de classificação de trajetórias, o *Compound Maximum a Posteriori* (CMAP). O terceiro passo foi a classificação de 37 anos de imagens Landsat para gerar trajetórias de uso e cobertura da terra de 1984 a 2020 na região do Baixo Tapajós. Avaliou-se o CMAP em oposição a abordagens de comparação pós-classificação, com e sem a utilização de um filtro de pós-processamento para remover transições inválidas. O CMAP foi capaz de fornecer classificações mais acuradas e com apenas transições válidas. No quarto passo, as trajetórias calculadas foram utilizadas para caracterizar e comparar os processos de regeneração florestal na área de estudo. Foram encontradas diferenças nos padrões espaço-temporais da concentração das maiores proporções de áreas desmatadas cobertas por vegetação secundária, do número de ciclos de regeneração e da persistência e idade da vegetação secundária em toda a área de estudo. Estas variações coincidem com mudanças conhecidas nas práticas agrícolas. As áreas ribeirinhas na parte noroeste da área de estudo, incluídas em projetos agroextrativistas, apresentaram trajetórias de regeneração florestal com um número relativamente elevado tanto de ciclos de regeneração quanto de persistência da vegetação secundária. Estas também apresentaram vegetação secundária mais antiga nos últimos anos da série temporal. Estes resultados evidenciam a necessidade de estudos detalhados sobre a dinâmica da vegetação secundária no tempo e no espaço dentro da Amazônia, tanto para estimativas de carbono, quanto permitir a proposta de políticas de conservação, ordenamento territorial e incentivos às práticas sustentáveis de uso da terra. Por fim, identificou-se que até 11,5% do desmatamento da vegetação secundária por ano ocorre em áreas em regeneração por 11 anos ou mais na área de estudo. Esses processos só podem ser plenamente caracterizados pela classificação de trajetórias válidas a partir de décadas de observações, evidenciando a importância de classificadores como o CMAP.

Palavras-chave: Vegetação secundária. Classificação de trajetórias. Amazônia.

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ABBREVIATIONS

21S	–	zone 21 south
AA	–	Annual Agriculture
AC	–	Annual Crop
AF	–	Agroforest
AF	–	Aquatic or regularly flooded
AG	–	Agriculture
ALI	–	Advanced Land Imager
ALOS	–	Advanced Land Observing System
ARPA	–	Amazon Protected Areas Program
AS	–	Artificial Surface and Associated Areas
B	–	Blue channel
BA	–	Built-up Area
BAP	–	Best Available Pixel
BFAST	–	Breaks for Additive Season and Trend
BR-010	–	Belém-Brasília highway
BR-163	–	Cuiabá-Santarém highway
BR-230	–	Transamazon highway
BR-364	–	Cuiabá-Porto Velho highway
BS	–	Bare Agricultural Soil
CAPES	–	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CAR	–	Rural Environmental Registry
CCDC	–	Continuous Change Detection and Classification
CLC	–	CORINE Land Cover
CM	–	Cultivated and Managed Terrestrial Area
CMAP	–	Compound Maximum *a Posteriori*
CMDS	–	Classical Multidimensional Scaling
CML	–	Compound Likelihood
CNAL	–	National Council of the Legal Amazon
CONVERT	–	conversion predicate
CORINE	–	COOrdination of INformation on the Environment
CP	–	Clean Pasture
DAAD	–	German Academic Exchange Service
DEGRAD	–	Degradation Mapping in the Brazilian Amazon
DETER	–	Deforestation Detection System in Real-Time
DETEX	–	Detection of Selective Exploitation
DFS	–	Sustainable Forest District
disc.	–	discriminative
DMLD	–	Double Minimum Link Dendrogram
DP	–	Degraded Pasture
DSV	–	Developed Secondary Vegetation

DTW	–	Dynamic Time Warping
Embrapa	–	Brazilian Agricultural Research Corporation
EO-1	–	Earth Observer 1
EROS	–	Earth Resources Observation and Science
ETM+	–	Enhanced Thematic Mapper Plus
EVOLVE	–	evolution predicate
F	–	Forest
FAO	–	Food and Agriculture Organization
FBD	–	Fine Beam Dual
FLONA	–	National Forest
G	–	Green channel
Gebam	–	Executive Group for the Lower Amazon Region
GEE	–	Google Earth Engine
Getat	–	Executive Group for Lands of the Tocantins-Araguaia
GLAD	–	Global Land Analysis and Discovery
GSOM	–	Generalized Self-Organizing Maps
GSTE	–	Grouping by Similarity of Temporal Evolution
HOLD	–	maintenance predicate
I PND	–	I National Development Plan
IA	–	Idle Agricultural Area
IBGE	–	Brazilian Institute of Geography and Statistics
ICM	–	Iterated Conditional Modes
IDL	–	Interface Description Language
IGH	–	Idle Agricultural Area with Green Herbs
II PND	–	II National Development Plan
IL	–	Indigenous Lands
INCRA	–	National Institute for Colonization and Agrarian Reform
INPE	–	National Institute for Space Research
IP	–	Integral Protection
ISS	–	Idle Agricultural Area in a Senescent State
ITTO	–	International Tropical Timber Organization
JAXA	–	Japan Aerospace Exploration Agency
JM	–	Jeffries-Matusita
k-NN	–	k-Nearest Neighbor
L1	–	first legend
L1(10)	–	base legend of 10 classes
L2	–	second legend
L3	–	third legend
L4	–	fourth legend
LA	–	Land
LandTrendR	–	Landsat-based detection of Trends in Disturbance and Recovery
LaSRC	–	Land Surface Reflectance Code

LBA	–	Large-scale Biosphere-Atmosphere experiment in Amazonia
LCCS	–	Land Cover Classification System
LCML	–	Land Cover Meta Language
LEDAPS	–	Landsat Ecosystem Disturbance Adaptive Processing System
LIDAR	–	Light Detection and Ranging
LULC	–	land use and land cover
LULCC	–	land use and land cover change
MA	–	Mature Forest
MapBiomas	–	Annual Mapping of Land Cover and Use in Brazil
MF	–	Modified Forest
MF1	–	Highly Modified Forest
MF2	–	Moderately Modified Forest
MF3	–	Slightly Modified Forest
MG	–	Mining
ML	–	Maximum Likelihood
MODIS	–	Moderate-Resolution Imaging Spectroradiometer
MSI	–	Multispectral Instrument
MSS	–	Multispectral Scanner
NDSI	–	Normalized Difference Snow Index
NIR	–	Near Infrared
NO	–	Not Observed
NSV	–	Natural and Semi-Natural Vegetation
NV	–	Primarily Non-Vegetated Area
OA	–	Overall Accuracy
OLI	–	Operational Land Imager
P.A.	–	Producer's Accuracy
P.S.V.	–	persistence of secondary vegetation
PA	–	Federal Settlement Project
PA	–	Pasture
PA-370	–	Santarém-Curua-Una
PAC	–	Joint Settlements Project
PAE	–	Agro-extractivist Settlement Project
PALSAR	–	Phase Array L-Band Synthetic Aperture Radar
PAS	–	Sustainable Amazon Plan
PDS	–	Sustainable Development Project
PEAEX	–	Agro-extractivist State Settlement Project
PEAS	–	Sustainable Settlement State Project
PIC	–	Integrated Colonization Project
PIN	–	National Integration Plan
Planaveg	–	National Plan for the Recovery of Native Vegetation
PNAP	–	National Strategic Plan for Protected Areas
PNMC	–	National Policy on Climate Change

Polamazônia	–	Programme of Agricultural and Agromineral Poles of the Amazon region
Polonoroeste	–	Northwest Brazil Integrated Development Program
PPA	–	Pluriannual Plan
PPCDAm	–	Action Plan for Prevention and Control of the Legal Amazon Deforestation
PPG7	–	Pilot Program to Conserve the Brazilian Rain Forests
PROBEM	–	Brazilian Molecular Ecology Program for the Sustainable Use of Biodiversity
PRODES	–	Program for Deforestation Monitoring in the Brazilian Legal Amazon
Promanejo	–	Forest Management Support Project
Pronabio	–	National Biological Diversity Program Brazil
PROTERRA	–	Programme for Land Redistribution and Stimulation of Agroindustry in the North and Northeast
Q	–	Quilombola areas
QA	–	Quality Assessment
R	–	Red channel
R.C.D.	–	duration of the forest regeneration cycle
RADAM	–	Radar in the Amazon
RAM	–	Random Access Memory
RD	–	Recent Deforested Area
RECUR	–	recurrence predicate
REIS	–	RapidEye Earth Imaging System
RESEX	–	Extrativist Reserve
RFM	–	Rational Function Model
RMSE	–	Root Mean Square Error
ROCK	–	RObust Clustering using linKs
S1	–	land use and land cover trajectory 1
S2	–	land use and land cover trajectory 2
S3	–	land use and land cover trajectory 3
S4	–	land use and land cover trajectory 4
S5	–	land use and land cover trajectory 5
SAD	–	Deforestation Alert System
SAR	–	Synthetic Aperture Radar
SC	–	Shifting Cultivation
SDNLM	–	Stochastic Distances Nonlocal Means
SEMA	–	Special Secretariat for the Environment
simp.	–	simplified
Sipam	–	Amazonian Protection System
SITS	–	Satellite Image Time Series Analysis
Sivam	–	Amazonian Surveillance System

SLHC	–	Single Link Hierarchical Clustering
SMLD	–	Single Minimum Link Dendrogram
SNUC	–	National System of Conservation Units
SP	–	Shrubby Pasture
SPOT	–	*Satellite Pour l'Observation de la Terre*
SPVEA	–	Superintendency for the Planning of the Economic Valorization of Amazonia
SRTM4	–	Shuttle Radar Topography Mission version 4
SU	–	Sustainable Use
SUDAM	–	Superintendency for the Development of Amazonia
SUFRAMA	–	Superintendence of Manaus Free Trade Zone
SV	–	Secondary Vegetation
SV1	–	Initial Secondary Vegetation
SV2	–	Intermediate Secondary Vegetation
SV3	–	Advanced Secondary Vegetation
SVM	–	Support Vector Machine
SWIR1	–	Short-wave infrared 1
SWIR2	–	Short-wave infrared 2
T.U.B.R.	–	time of use before regeneration
T1	–	Tier 1
TC	–	Permanent Cultures
TC	–	Temporary Cultures
TerraClass	–	Land Use and Land Cover Mapping of Deforested Areas in the Legal Amazon Project
TM	–	Thematic Mapper
TOA	–	Top of Atmosphere
TWDTW	–	Time-Weighted Dynamic Time Warping
U.A.	–	User's Accuracy
UC	–	Conservation Unit
UML	–	Unified Modeling Language
UNEP	–	United Nations Environment Programme
USGS	–	United States Geological Survey
UTM	–	Universal Transverse Mercator
VeRDET	–	Vegetation Regeneration and Disturbance Estimates through Time
VLCE	–	Virtual Land Cover Engine
WFS	–	Web Feature Service
WGS84	–	World Geodetic System 1984
WP	–	Woody Plantation
ZEE	–	Ecological-Economic Zoning

LIST OF SYMBOLS

%	–	percent
$\otimes$	–	Cartesian Product of sets
$\lvert \cdot \rvert$	–	determinant function
$\ldots$	–	omitted values (horizontal)
$\vdots$	–	omitted values (vertical)
$\ddots$	–	omitted values (diagonal)
$\in$	–	pertains
$\cap$	–	union
$\sim$	–	approximately
T	–	transpose function
$\arg_s \max$	–	maximum argument of s
Σ_k	–	covariance matrix of samples from class k
$\omega_t^{k_t}$	–	actual class at time position t
Ω_t	–	set of classes
GB	–	gigabyte
GHz	–	gigahertz
ha	–	hectare
km	–	kilometer
m	–	meter
mm	–	millimeter
μm	–	micrometer
B	–	number of observed variables
B_{ij}	–	Bhattacharyya distance between the classes i and j
C	–	legend based on the conceptual model
e	–	Euler's Number
i	–	index of class
j	–	index of class
JM_{ij}	–	Jeffries–Matusita between the classes i and j
k_t	–	indicator of the class in time t
L	–	the base set of classes
Legend $l(n)$	–	names of legends. In which *Legend* denotes the type of the legend, l denotes the tier level, and n represents the number of classes
ln	–	function of natural logarithm
m	–	mean
m_k	–	mean vector of the class k
n	–	number of classes/samples
n_i	–	number of samples classified as i

n_{ij}	–	number of samples of class j classified as class i
N_s	–	number of valid trajectories
P	–	legends based on on parametrization
$P(a\|b)$	–	probability of a given b
$p_{(lat,\ long,\ t)}$	–	pixel at position (lat, $long$) and time t
$P(s)$	–	*a priori* probability of a trajectory s
$\widehat{p}_{ij}$	–	cell ij of the confusion matrix
s	–	a trajectory
$S(\widehat{O})$	–	desired standard error of the estimated Overall Accuracy
sd	–	standard deviation
S_i	–	a variable calculated for the class i
t	–	an indicator of time
T	–	length of the time sequence
t_y	–	year of observation
U_i	–	User's Accuracy of class i
W_i	–	mapped proportion of the area of class i
$\boldsymbol{X}$	–	a set of images
$\vec{X}$	–	an observation vector

CONTENTS

1 INTRODUCTION [1]

The Amazon basin is considered one of the most important ecological systems on the planet (FOLEY et al., 2007). Besides being a crucial element to the Earth's climate, water, and carbon system (COE et al., 2013; STRAND et al., 2018; SPA, 2021), it holds an unique and diverse biodiversity and human culture (HECKENBERGER et al., 2007; SPA, 2021). The Amazon basin also holds the biggest tropical forest in the world (SAATCHI et al., 2011), which is located mostly in the Brazilian territory. In Brazil, the Amazon biome is commonly referred to simply as the Amazon. This biome occupies almost half of the Brazilian territory and extends over the North, Northeast, and Center-West regions. For legal and planning purposes, the Brazilian government also instituted the Brazilian Legal Amazon. It includes the states of Acre, Amapá, Amazonas, Mato Grosso, Pará, Rondônia, Roraima, Tocantins, and part of the state of Maranhão (BRASIL, 1966; BRASIL, 2007).

The Amazon has been under constant local and regional pressure for its natural resources. This pressure is reinforced by the type of economic development practiced in the region (BECKER, 2007), as well as the current and inefficient control, conservation, and forest management policies. These elements are commonly reflected in land use and land cover (LULC) changes, with particular concern to the deforestation of primary forests (FEARNSIDE, 1990; COE et al., 2013). This type of change has the potential to disrupt carbon storage, water and regional climate patterns (FOLEY et al., 2007), while causing biodiversity loss, greenhouse gas emissions, and surface albedo modification (FEARNSIDE, 2005). It can also lead to the proliferation of tropical diseases (ELLWANGER et al., 2020) and is a threat to indigenous people and other local communities' livelihoods (RORATO et al., 2020; SPA, 2021).

Some of the deforestation impacts can be offset by secondary forest regeneration processes, i.e., the natural growth of secondary vegetation in previously deforested areas (BROWN; LUGO, 1990; CORLETT, 1995), herein treated as forest regeneration, for simplicity. During forest regeneration, carbon is assimilated from the atmosphere and either transformed into biomass or fixed in the soil (RAMANKUTTY et al., 2007). Besides carbon sequestration, forest regeneration processes can provide direct resources to local populations (JUNQUEIRA et al., 2011), as well as important ecosystem services, such as

[1] Excerpts from this chapter are based on Reis et al. (2020b).

biodiversity recovery, soil erosion prevention, soil fertility recovery, hydrological cycle maintenance, and the formation/connection of ecological corridors (CHAZDON et al., 2009; MEYFROIDT; LAMBIN, 2011; JAKOVAC et al., 2021). It is also a fundamental element in traditional agricultural practices, such as shifting cultivation (COSTA, 2016; JAKOVAC et al., 2021). Therefore, it is considered a low-cost and socially effective manner to contribute to the conservation of biodiversity, mitigation and adaptation to climate change, enhancement of food security, and protection of water supplies (CHAZDON et al., 2020).

Nonetheless, there is also a contrasting view about forest regeneration processes in the Amazon, in which the presence of secondary vegetation is seen as the result of failures in agricultural activities in areas that have been deforested for this purpose. Regeneration processes are then associated with the degradation of the land and unsustainable uses of low economic efficiency (COSTA, 2016). This second view sustains discourses against traditional practices of local populations that are important parts of the Amazon economy and culture, and also enables the existence of secondary vegetation patches within human-modified landscapes.

The benefits of forest regeneration processes may vary depending on biophysical elements and management practices. Structural differences in secondary vegetation across the regions within Amazon may be explained by variations in climate (HEINRICH et al., 2021) and soil fertility/type (MORAN et al., 2000). Landscape context and land use history, however, largely influence secondary vegetation structure and successional pathways within the same region (MORAN et al., 2000; JAKOVAC et al., 2021). Allied with different management practices, these characteristics may lead to different forest regeneration processes that result in varied dynamics of biomass accumulation and carbon absorption rates (STEININGER, 2000; AGUIAR et al., 2012).

Details on the spatial-temporal distribution and the intensity of forest regeneration processes in multi-temporal analyses and within different contexts of land management are important information to estimate greenhouse gas emissions and support territorial planning (VIEIRA et al., 2014; PARÁ, 2015), as well as the elaboration of conservation policies and sustainable land use practice incentives. Here, forest regeneration processes were treated as a particular type of LULC trajectory. A trajectory is defined as a sequence of transitions. A transition in LULC is the observation of land use and/or land cover class in the same unit of analysis in two observations, usually successive (MACIEL, 2017). In this context,

land cover refers to the biophysical state of the land surface, including the amount and type of vegetation cover, water, and other materials and structures of natural or anthropic origins. Land use refers to the set of activities carried out on the land and the purpose of manipulating the land cover for activities of human interest (TURNER; MEYER, 1994). Thus, LULC trajectories are defined as the succession of types (i.e. classes) of land use and/or land cover in a given spatial unit of analysis at three or more observed times (MERTENS; LAMBIN, 2000; MENA, 2008). Accordingly, forest regeneration trajectories are LULC trajectories that contain at least one regeneration event. A regeneration event is typically flagged by one or consecutive observations of secondary vegetation within the LULC trajectories, depending of the temporal resolution of the analysis.

A primary method of obtaining LULC trajectories for long, albeit relatively recent, periods is the processing of remote sensing image time series (WULDER et al., 2018). Note that information about land use is not directly derived from remote sensing data. However, land use can be attributed by the analyst according to the knowledge of the study area, in association with the land cover class, and/or using auxiliary information. Here, remote sensing based LULC trajectory studies were divided into the basic steps: 1) problem definition, 2) trajectory classification, and 3) extraction and analysis of information.

Each step brings methodological challenges such as the definition of LULC classes of interest, input data, and classification/analysis methodology. Particularly, the current classification methods are either prone to the classification of invalid LULC trajectories, i.e. those that could never happen in the field, or demand many processing steps that can hinder the classification of secondary vegetation areas. The challenges associated with each step are investigated in this thesis, and may be synthesized in the hypothesis: *incorporating previous knowledge about land cover dynamics and the information of multi-temporal data sets during the classification of LULC trajectories results in consistent and more accurate classification of secondary vegetation areas.*

This hypothesis was tested by classifying and analyzing forest regeneration trajectories in a study area located in the Lower Tapajós region, in the State of Pará, within the Brazilian Amazon. This is an area characterized by different occupation histories and management practices, whose LULC dynamics were influenced by public policies with territorial and institutional expression implemented at varied times (LOBO et al., 2015).

1.1 Research objectives and contributions

This thesis aims to propose a method for the classification and analysis of forest regeneration trajectories based on previous knowledge about land cover dynamics and multi-temporal remote sensing data. To this end, the following specific objectives were established:

a) to propose objective definitions of LULC classes in the studied region, in order to allow for consistent and replicable analyses based on multi-temporal remote sensing data sets;

b) to develop a novel trajectory classifier able to extract only valid land cover trajectories directly from remote sensing time series;

c) to extract LULC trajectories for the study area with annual observations from years 1984 to 2020;

d) to identify, quantify, and characterize forest regeneration trajectories related to different agrarian systems [2] and public policies with territorial expression.

To achieve these objectives, the present document was structured into eight chapters:

a) **Chapter 1:** motivation of the present study, introduction of the hypotheses and research objectives;

b) **Chapter 2:** theoretical background. Review of related studies and analysis of the main methods used to classify and analyze LULC trajectories based on orbital remote sensing data;

c) **Chapter 3:** description of the study area based on geographic and political limits, as well as a brief historic of occupation and overview of public policies;

d) **Chapter 4:** proposal of LULC legends and classification tests over a commonly used medium spatial resolution remotely sensed image from the Landsat family;

[2] An agrarian system is defined as the set of management practices, agrarian structures, and social-economic characteristics of an activity that results in the production/collection of agricultural goods.

e) **Chapter 5:** proposal of the Compound Maximum *a Posteriori* classifier (CMAP);

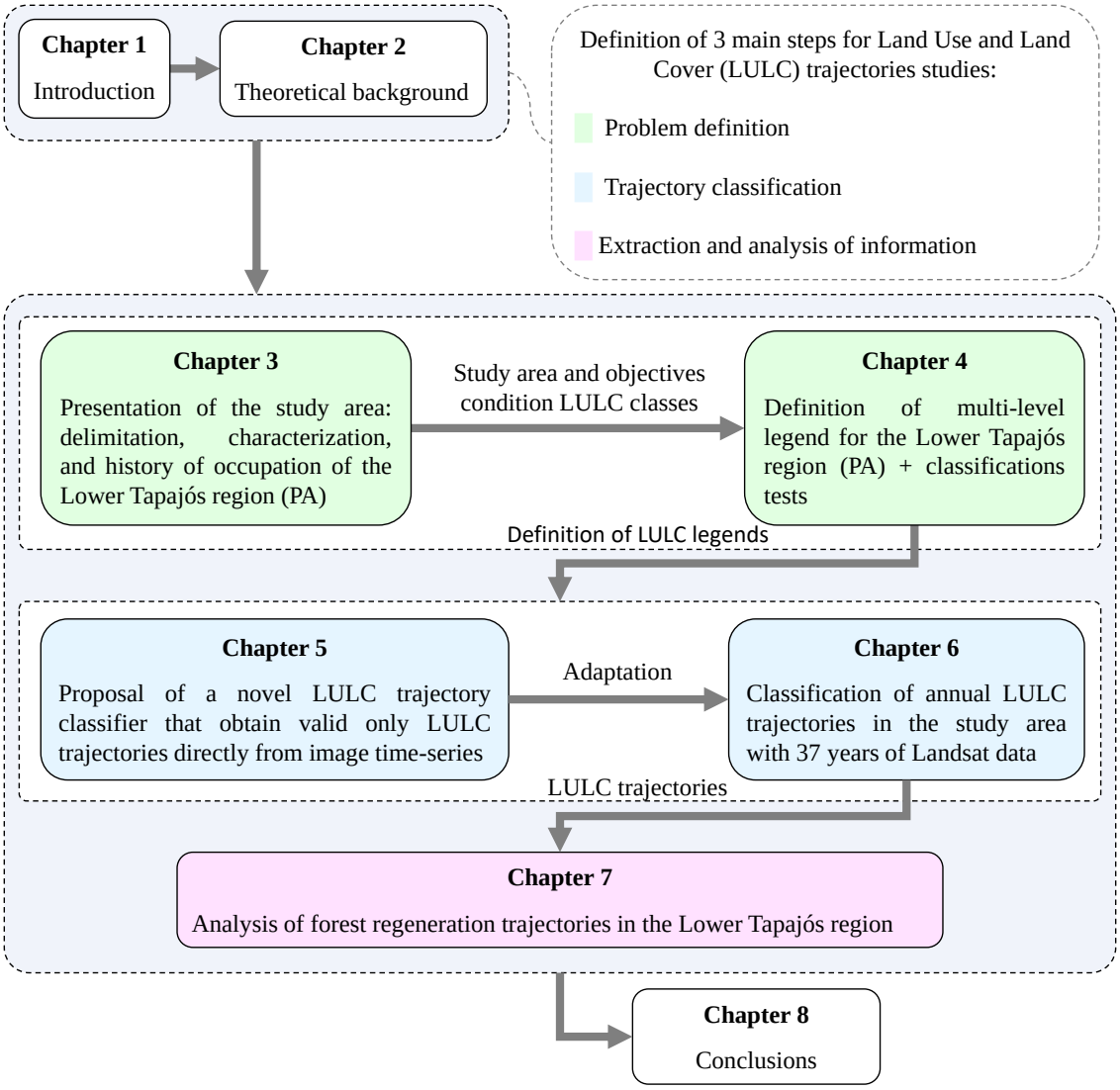
f) **Chapter 6:** use of CMAP to classify 37 years of Landsat data into LULC trajectories;

g) **Chapter 7:** extraction and analysis of forest regeneration trajectories in the Lower Tapajós region;

h) **Chapter 8:** conclusions and perspectives for future studies.

The thesis structure is illustrated in Figure 1.1.

Figure 1.1 - Structure of the present document.



SOURCE: The author.

2 THEORETICAL BACKGROUND[1]

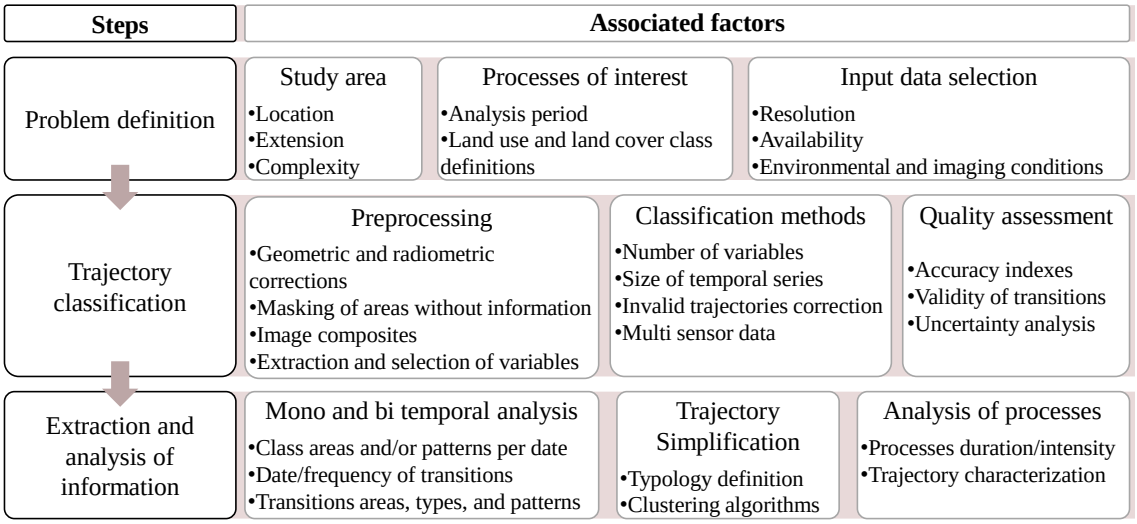
As previously mentioned, LULC trajectories are commonly obtained by processing time series data of orbital remote sensing images. In recent years, new techniques have emerged to analyze remote sensing time series data, as a result of computational and methodological advances, the unprecedented availability of data from different sensors, and improvements in data processing/format standardization (WULDER et al., 2018). Some of these techniques have been reviewed by Banskota et al. (2014), Gómez et al. (2016), and Zhu (2017). The latter studies are comprehensive and more general and, therefore, do not detail important specificities when it comes to analyzing LULC trajectories in forest regeneration studies in the Amazon.

Thus, this chapter aims to identify and describe the main methods used to classify and analyze LULC trajectories based on orbital remote sensing data. The clear definition of objectives, LULC classes, and the spatial-temporal resolution and extension of the study are crucial, given the high degree of complexity in forest regeneration processes. These factors influence both data selection and methods in forest regeneration analyses in the Amazon, as discussed here.

The methods and procedures used to analyze LULC trajectories based on remote sensing data were summarized in three basic steps: 1) problem definition; 2) trajectory classification; and 3) information extraction and analysis. These steps are illustrated in Figure 2.1 and discussed in the context of specific characteristics of forest regeneration studies as follows.

[1]This chapter is an adapted version of Reis et al. (2020b). Sections were expanded to include the theoretical background from the other chapters and the information about available data sets has been updated.

Figure 2.1 - Steps and associated factors for land use and land cover trajectory analyses based on remote sensing data.



SOURCE: The author.

2.1 Problem definition

The study objectives are defined in this step, as well as the location and extent of the studied area, the period of the analysis, the LULC classes, and the required level of accuracy (LU et al., 2014). Both the data and methods to be used in the analysis are selected according to the problem definition. Nonetheless, both the availability (or lack thereof) of remote sensing data and methods to process and extract different types of information often limit and/or shape the objectives of the study. Inadequate spatial and temporal resolution in the used data, difficulties discriminating between different LULC classes, and a lack of available time series data for long periods with dense observations are among the main limitations.

2.1.1 Defining secondary vegetation

There are different expressions used as synonyms for secondary vegetation in the literature. The most common expressions, as well as selected variations in the definition, are presented in Table 2.1.

Table 2.1 - Definitions of secondary vegetation found in the literature.

Expression	**Definition and authors**
Secondary vegetation	Forests formed as a consequence of human impact on forest lands, particularly as a result of abandoned cleared forest lands, generally for agriculture, and excluding planted forests and forests resulting from natural disturbances (BROWN; LUGO, 1990). Clear-cut areas detected by the Program for Deforestation Monitoring in the Brazilian Legal Amazon (PRODES), where vegetation has regenerated after the land has been abandoned (ALMEIDA et al., 2010). "Originally forested areas that had been clear-cut, used for silviculture, agriculture or pasture and then left to regenerate, taking on a forest appearance again" (ALMEIDA et al., 2016).
Secondary forest/ Second-growth forest	"Forests regenerating largely through natural processes after significant human and/or natural disturbance of the original forest vegetation at a single point in time or over an extended period, and displaying a major difference in forest structure and/or canopy species composition with respect to nearby primary forests on similar sites" (CHOKKALINGAM; JONG, 2001). Forests that develop after complete deforestation (CORLETT, 1995; PUTZ; REDFORD, 2010; VIEIRA et al., 2014; WANG et al., 2020). Vegetation that emerges on abandoned farmland or planted estates where the original forest was clear-cut before agricultural use (KAMMESHEIDT, 2002). Forests that occur "when a pixel classified as anthropic cover (e.g., pasture or agriculture) in a given year is replaced in the following year by a pixel of forest cover (excluding mangroves and plantations)" (SILVA JUNIOR et al., 2020).
Capoeira	Secondary vegetation that grows after the clearing of primary forests (IBGE, 2004b). "Areas with secondary vegetation that are either temporarily or permanently removed from agricultural production. (...) Land areas in different stages of natural regeneration after having been radically altered by human activity" (COSTA, 2016). Areas of varying dimensions, in different stages of spontaneous regeneration of forest cover in ecosystems radically altered by human actions (COSTA, 2009). Secondary vegetation in the intermediate stage of succession (SALOMÃO et al., 2012).
Capoeirão	Secondary vegetation in an advanced stage of succession (IBGE, 2004b; SALOMÃO et al., 2012)
Capoeirinha/ Juquira	Secondary vegetation in the initial succession stage (SALOMÃO et al., 2012).

SOURCE: The author.

The presented definitions agree with the idea that secondary vegetation, secondary forests, or *capoeiras* are types of vegetation that regenerate after the original forest cover has suffered some type of disturbance. The main observed differences among the definitions relate to:

a) the type of disturbance: whether the disturbance is of a natural or anthropic origin, or both;

b) the intensity of the disturbance: whether it occurred only in areas where the forest had been clear-cut or in areas with a given level of forest degradation;

c) the land use after disturbance: some definitions incorporate specific types of land use before secondary vegetation;

d) the vegetation structure and composition: whether the resulting vegetation presents similar characteristics to the original forest cover;

e) the regeneration process: whether it is natural or a result of human intervention;

f) the successional stage: in some cases, some expressions refer to different successional stages, whereas these same expressions may also be used in a general way, as is the case with the expression *capoeira*;

g) the characteristics of the data used: certain definitions are associated with the capacity to identify secondary vegetation with remote sensing data, such as those adopted by Almeida et al. (2010) and Silva Junior et al. (2020).

In this document, the expression secondary vegetation was reserved to describe the vegetation that grows without human direct interference in areas with complete removal of the original forest cover, without distinction as to the cause of the removal of this cover or the types of intermediate use. The regeneration of degraded forests, i.e. those with partial loss of original forest cover, was not included in the present discussions.

2.1.2 Study area and processes of interest

Although commonly treated as a single unity, the Amazon presents a heterogeneous spatial pattern associated mainly with different phases of development and land use management history/practices, also affected by biophysical and socio-economical

variations (PERZ; SKOLE, 2003a; MELLO; ARTAXO, 2017). The joint action of varied occupation processes, a not so synergic combination of international market demands, government policies, and agents such as indigenous and traditional populations, rural and urban inhabitants, and small, medium, and large farmers led to many forms of land uses (BECKER, 2005; BECKER, 2007), such as:

a) forest extrativism and traditional fishery;

b) small-scale agriculture/shifting cultivation (mainly by riverine population);

c) selective logging followed or not by forest clear-cut and cattle raising expansion;

d) large-scale cattle raising and agriculture (mainly soy);

e) mineral exploitation;

f) protected areas, in which some of the mentioned activities may be allowed.

Therefore, forest regeneration processes in the Amazon occur in different ecological, social, and political contexts (VIEIRA et al., 2014; CHAZDON et al., 2020). Studies seeking to examine these processes must consider existing different land management systems and regeneration cycles with varying time lengths, as these may affect the definition of the analysis period and LULC classes.

According to Costa (2016), the presence of secondary vegetation in the Brazilian Amazon can be associated with three types of land use decisions:

a) the use of techniques that adopt the regeneration of fallow areas to restore soil fertility, such as the shifting agriculture carried out by riverside and traditional populations;

b) the abandonment of degraded and/or unproductive areas after exhausting the soil and climatic conditions needed for production;

c) the intensification of land use that leads to production in smaller areas, causing secondary vegetation to regenerate on excess lands. In this case, the land surplus can be incorporated into the production process again, thereby leading to the deforestation of the secondary vegetation.

Mello and Alves (2011) mention an additional type of secondary vegetation that results from the abandonment of a recently deforested area that has never been involved in the production process.

According to Wang et al. (2020), secondary vegetation areas are commonly deforested within two years after the start of the regeneration process. Additionally, forest regeneration can involve abrupt or subtle changes in both the short and long-term (NUNES et al., 2020). Riverside regions, for example, may present agricultural activities with fallow and regeneration periods that vary between less than four and more than 15 years (STEININGER, 2000; AFFONSO et al., 2016; JAKOVAC et al., 2017). In the state of Pará, Normative Instruction 08 of October 28, 2015 states different rules for vegetation suppression depending on the age of the secondary vegetation. These rules are stratified as follows: no older than five years, five to 20 years, and older than 20 years (PARÁ, 2015). The categorization between secondary vegetation areas that are less than 20 years old and those that are more than 20 years old is also used to estimate carbon absorption rates (SILVA JUNIOR et al., 2020). To account for abrupt changes, it is interesting to have at least one observation of LULC mapped per year to improve the accuracy of age estimates. Analyses carried out at intervals of more than 20 years are also recommended to study forest regeneration in the Amazon, to allow the full observation of regeneration cycles. Such studies also enable age estimates of older vegetation.

Despite being commonly overlooked, defining LULC classes is a particularly important step in LULC studies given that different class definitions can result in quite different analyses. According to Hansen and Loveland (2012), most large area, medium spatial resolution, land cover monitoring products are focused on forest cover change mapping, with special attention to changes occurring in tropical forests. 'Forest', however, can be a challenging concept to define and the lack of consensus on this concept may have serious consequences in conservation, development, climate, livelihood, biodiversity, and ecosystem services (PUTZ; REDFORD, 2010). Chazdon et al. (2016) point out that a simple change in the definition of the term forest can alter management policies or inventoried forested areas. For instance, both Putz and Redford (2010) and Chazdon et al. (2016) cite examples in which an area nearly devoid of trees would still be considered forested. Furthermore, the 'forest' class sometimes encompasses areas of primary forest, degraded forest, and also secondary vegetation, which is not compatible with forest regeneration studies (PUTZ; REDFORD, 2010).

There are three basic ways to define LULC classes. The first is to arbitrarily define the used classes. The second is to define LULC classes based on the empirical analysis of the data (PEREIRA et al., 2018) and/or image classification results (PANTALEÃO et al., 2012). The third way is to derive classes from a standard classification or legend system. A classification system is a logical framework that retains class names, the criterion for class separation, and the relationship among classes (MCCONNELL; MORAN, 2001). A legend is a subset of classes, which may or may not be obtained from a classification system (MCCONNELL; MORAN, 2001; HEROLD et al., 2006).

The first method, arbitrarily establishing classes, can lead to widely different definitions of classes, data collection protocols, and analytical techniques. For instance, there is a fundamental difference between data used for local and global studies. According to McConnell and Moran (2001), global researchers are often interested in the representation of biophysical aspects and temporal patterns linked to climate change (e.g., carbon dynamic and sequestration rates), usually focusing on major land cover changes across large spatial extents. Local researchers, instead, focus on the characterization of human-induced modifications to land cover at finer scales. Even considering similar scales and objectives, the straightforward comparison between results is sometimes impossible because of slight to moderate differences in the legend definition (AHLQVIST, 2008).

In the second method, the class definition is usually tailored to present accurate classification results giving the used data. Although it can solve accuracy problems, many types of inconveniences may arise. The most obvious one is that studies with similar study areas and objectives can be carried out with different legends because of differences in data type/sensor/date, field data collection, processing methodologies, or the expected quality of the final map. It is the case of the studies carried out by Angelis et al. (2002), Freitas et al. (2008), Dutra et al. (2009), Silva et al. (2011), Pantaleão et al. (2012), and Pereira et al. (2016), for example. These studies were carried out within the same study area in the Lower Tapajós region using similar LULC classes. However, since different classes are merged for analysis, the resultant maps are rarely directly comparable and their usefulness for studies with other types of data is diminished.

The third method, i.e. the use of a unique classification system or standardized legend, has the advantage of allowing direct comparison between class sets (ANDERSON et al., 1976). Among the commonly used classification systems are:

Sistema Básico de Classificação da Cobertura e do Uso da Terra (IBGE, 2013), Anderson et al. (1976), and CORINE (COOrdination of INformation on the Environment) Land Cover (CLC) (EEA, 2019). None of these classification systems has been fully and internationally accepted since they do not apply to every study objective or area of interest (DI GREGORIO; JANSEN, 2005).

In all three methods, any classes defined using subjective criteria can compromise both the reproducibility of a study and eventual comparisons between results. As a universally accepted standard legend is not feasible, one solution to decrease the subjectivity of class definition and allow comparison of results, to a certain degree, is to standardize the used terminology, instead of the categories. This is the main purpose of the Land Cover Meta Language (LCML), developed by the Food and Agriculture Organization (FAO) and the United Nations Environment Programme (UNEP) as an evolution of the Land Cover Classification System (LCCS) (DI GREGORIO; JANSEN, 2005).

The LCML is a Unified Modeling Language (UML) meta-model that proposes the use of common classifiers to describe classes based on quantifiable physiognomic aspects of the land cover. This language was used to describe the classes from LULC studies and projects (UDDIN et al., 2015; ARAFAT et al., 2014; CORD et al., 2010; COUTINHO et al., 2013), as well as some well-known land cover classification systems (HEROLD et al., 2009). In addition, LCML has been proposed as an instrument of harmonization between land cover data sets with different land cover classes (KOSMIDOU et al., 2014) and/or produced considering very different scales (MCCONNELL; MORAN, 2001).

To our knowledge, the most notable effort in using LCML to describe a legend for the Brazilian Amazon was done by the Land Use and Land Cover Mapping of Deforested Areas in the Legal Amazon Project (TerraClass), as described in Coutinho et al. (2013). However, the LCML description was done after the legend definition in a non-systematic way and mainly based on used data, which impacts this legend's usability for local studies and field campaigns. To address this problem, a data-independent LCML formalized legend is proposed in Chapter 4.

2.1.3 Input data selection

The potential to identify the LULC classes in the study scale in a given remote sensing data set is conditioned to the sensor's resolution. The characteristics of the main remote sensors with open archives are compiled in Table 2.2.

Table 2.2 - Characteristics of the main remote sensors with open archives.

Satellite/Sensor	Resolution				Footprint (km)	Operation period
	Spatial (m)	Temporal (days)	Spectral (μm)	Radiometric (bits)		
NOAA(6-19)/AVHRR(2-3)	1100	< 1	4-6 bands (0.58 to 12.50)	10	2400	1979 to 2019
Landsat(1-5)/MSS	68 × 83	16-18	4 bands (0.50 to 1.10) [a]	6-8	185	1972 to 1999
Landsat(4-5)/TM	30 [b]	16	7 bands (0.45 to 12.50)	8	185	1982 to 2013 [c]
Landsat-7/ETM+	30 [b,d]	16	8 bands (0.45 to 12.50)	8	183	from 1999 [e]
Landsat-8/OLI	30 e	16	9 bands (0.43 to 2.30)	12	185	from 2013
Landsat-8/TIRS	100	16	2 bands (10.30 to 12.50)	12	185	from 2013
(Terra/Aqua)/MODIS	250-1000	1-2	36 bands (0.40 to 14.38)	12	2330	from 1999
Terra/ASTER	15-90	16	15 bands (0.52 to 11.65)	8-12	60	from 1999 [f]
CBERS(1-2b)/CCD	20	26	5 bands (0.51 to 0.89)	8	113	1999 to 2010
CBERS(1-2b)/WFI	260	< 5	2 bands (0.63 to 0.89)	8	890	1999 to 2010
CBERS(1-2)/IRMSS	80-160	26	4 bands (0.5 to 12.50)	8	120	1999 to 2008
CBERS-2b/HRC	2.7	130	1 band (0.50 to 0.80)	8	27	2007 to 2010
CBERS-4/PAN	5-10	26-52	4 bands (0.51 to 0.89)	8	60	from 2014
CBERS-4/IRS	40-80	26	4 bands (0.50 to 12.50)	8	120	from 2014
CBERS-(4-04A)/WFI	55-64	< 5	4 bands (0.45 to 0.89)	10	684-866	from 2014
CBERS(4-04A)/MUX	16.5-20	26-31	4 bands (0.45 to 0.89)	8	95-120	from 2014
CBERS-04A/WPM	2-8	31	5 bands (0.45 to 0.90)	10	92	from 2020
Resourcesat(1-2)/LISS III	23.5	24	4 bands (0.52 to 1.70)	7	141	from 2004
Resourcesat(1-2)/AWiFS	56	5	4 bands (0.52 to 1.70)	10	370-810	from 2004
Sentinel-2/MSI	10-60	5	13 bands (0.43 to 2.32)	12	290	from 2015
Amazonia-1/WFI	64	5	4 bands (0.45 to 0.89)	10	850	from 2021
ALOS/PALSAR	10-100 [g]	n.d.	band L (23 cm) [h]	32	30-350 [g]	2006 to 2011
Sentinel-1	5-40 [g]	6-12	band C (5.5 cm) [i]	10	20-410 [g]	from 2014[j]

[a] The sensor carried a thermal band in Landsat 3. [b] 120 m for the thermal band. [c] Significant drops in the number of observations after 2011. [d] 15 m for the panchromatic band. [e] Detection failures from May 2003. [f] Saturation of infrared band from April 2008. [g] Depending on the acquisition mode. [h] 1 to 4 polarizations, depending on the acquisition mode. [i] 1 to 2 polarizations, depending on the acquisition mode. [j] Detection failure in Sentinel-1B from December 2021.

SOURCE: Based on EMBRAPA (2018), ASF (2020), ESA (2020), INPE (2020a, 2020d), NASA (2020a, 2020b), USGS (2020), and Braz and D'Alge (2021).

It should be noted that the coverage and period of operation of a satellite/sensor do not guarantee the availability of data. This can occur because the sensors may not operate regularly in some areas (e.g., the Multispectral Scanner (MSS), which sensed images for the Amazon on a few and sparse dates). Additionally, remote optical sensors are strongly influenced by atmospheric conditions. Given that cloud cover is frequent in the Amazon at certain times of the year, it is possible that an orbital optical image acquired for the desired period and area may not be suitable for analysis. Despite this eventual difficulty of finding cloud-free images for certain dates, the characteristics of data from the Landsat family are appropriate to analyze annual LULC trajectories over long periods. Nonetheless, Sentinel-2 data has great potential for future analyses, given the availability of free images, the resolution characteristics of its sensor, cross-calibration with Landsat data, and the prospects of project continuity (WULDER et al., 2018). Although not fully open, another useful data set with public access is the monthly mosaics of Planet images, made available for the neotropics from 2016 at a 5 m spatial resolution and in real-color composition (`https://www.planet.com/`).

Synthetic Aperture Radar (SAR) data are obtained almost independently of weather conditions and are able to capture information about the land surface in cloud-covered areas (PARADELLA et al., 2005). However, SAR data are more difficult to interpret and process than optical data and are freely available for short periods. Despite the arrival of the first global forest products based on SAR data (SHIMADA et al., 2014), studies based solely on SAR images are still mostly limited to a few dates. With the continuation of the Sentinel-1 project, it is possible that new products based on SAR data, or even multi-sensor approaches, emerge in the future.

2.2 Trajectory classification

The trajectory classification step corresponds to assigning LULC classes to each spatial unit analysis at each time of interest. This step involves three main procedures: 1) preprocessing; 2) classification; and 3) quality assessment.

2.2.1 Preprocessing

Once the remote sensing data is selected, it is necessary to ensure that these data are comparable at the unit of analysis level, which is usually the image pixel. In general, this process involves absolute or relative radiometric and geometric correction of the data (COPPIN et al., 2004; RICHARDS; JIA, 2006; WULDER et al., 2018). It is

also common to mask and/or remove low-quality pixels, or pixels without any information, from the analysis. One way to solve low-quality pixel problems is to compose a new image from selected pixels from radiometrically calibrated images in the period of interest. The selection of the Best Available Pixel (BAP) can be done by combining different factors, such as the low probability of clouds/cloud shadow, distance to clouds, proximity to the date of interest, and quality indexes derived from the images, for example. Different BAP selection techniques have been proposed over the years (HOLBEN, 1986; GRIFFITHS et al., 2013; WHITE et al., 2014).

Besides the pixel digital values, it is possible to conduct the analysis by using different measurements calculated from the images. Examples of these measurements include indexes derived from the mathematical combination of the image channels, sub-pixel fractions (e.g., fractions of vegetation, soil, and shadow obtained from the Linear Model of Spectral Mixture proposed by Shimabukuro and Smith (1995)), texture, transformed images (e.g. principal components, Tasseled Cap), and fused images. There are also specific metrics that consider the multi-temporal characteristics of the data (FRANKLIN et al., 2015), and statistical metrics used to summarize sets of images from a given period, such as the median, quartiles, and standard deviation (RUFIN et al., 2015; SOUZA et al., 2020). According to Lu et al. (2004) variable selection should be based on: 1) the capacity to differentiate between classes of interest, 2) the decreasing data dimensionality to improve classification processes and/or to avoid including noises in the analysis, 3) the limitations in the classification methods, such as whether it is possible to use more than one variable/type of data simultaneously, and 4) whether there is an additional need to standardize data, e.g., if the radiometric calibration of data is not enough to minimize differences on multi-sensor data and/or the effects of factors such as illumination and topography.

Data from the Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager (OLI), and Multispectral Instrument (MSI) sensors onboard Landsat 4-5, Landsat 7, Landsat 8, and Sentinel 2 satellites, respectively, are available with geometric and radiometric corrections in many platforms, such as the United States Geological Survey (USGS) archives (`https://earthexplorer.usgs.gov/`, `https://glovis.usgs.gov/app`). These corrected data are also available in the Google Earth Engine (GEE) platform (GORELICK et al., 2017) with cloud masks. Other algorithms that can be used to preprocess these images are discussed by Zhu (2017), Frantz (2019), and Sanchez et al. (2020).

2.2.2 Classification

LULC trajectory classification can be done either by the successive use of traditional classification methods suitable for single images or change detection in pairs of images successively or using methods focused on the analysis of remote sensing time series images. The first approach appears to be more common in the literature.

The most commonly used method to classify LULC trajectories is post-classification comparison. In this method, images for each date are independently classified and then stacked to create a trajectory. Given that classification is carried out independently for each image, there is no need for a radiometric correction step and it is possible to use multi-sensor data (LU et al., 2004). Although it is necessary to know the land cover on each date, no prior knowledge about the LULC trajectories is needed. One of the main criticisms of this method is that the quality of the LULC trajectories directly depends on the classification results for each date (FULLER et al., 2003; TEWKESBURY et al., 2015). Classification errors can result in invalid trajectories, i.e., LULC trajectories that present inconsistent transitions (AZEREDO et al., 2016; REIS et al., 2020a).

Another common LULC trajectory classification method is to classify the data from a given date and use it as the base classification to be updated on the next date by using binary bi-temporal change detection techniques (COMBER et al., 2004; XIAN; HOMER, 2010; HUANG et al., 2017). This method may involve several steps, which can demand a lot of time and effort from the analyst when it comes to studies with a large number of observations. Additionally, classification errors are cumulative.

It is also possible to directly analyze radiometrically calibrated image time series to detect changes in land cover and then classify the resulting time segments of non-changes into LULC classes. The Continuous Change Detection and Classification (CCDC) algorithm, proposed by Zhu and Woodcock (2014), follows this logic. As such, the CCDC is more suitable for analyses with dense time series data, i.e. those with several observations for each interval of interest. This is not the case with historical analyses in the Amazon, in which few (and sometimes none) cloud-free images are found per year. Different algorithms focused on time segmentation of remote sensing time series data are currently available on GEE (`https://developers.google.com/earth-engine/api_docs`). Among them, the Landsat-based detection of Trends in Disturbance and Recovery (LandTrendR) (KENNEDY et al., 2010) and the Vegetation Regeneration

and Disturbance Estimates through Time (VeRDET) (HUGHES et al., 2017) are particularly interesting to analyze forest regeneration in the Amazon. Both algorithms use image composite-based variables as input data and classify the time segments as *disturbance*, *stable*, or *regeneration* based on the variation of the analyzed values. However, they do not directly provide LULC trajectories.

The classification of the temporal segments of non-change can be achieved by using supervised classifiers that are capable of handling high-dimensional data, such as Random Forest and the Support Vector Machine (SVM), those based on neural networks, or even algorithms based on distances between time series. The Dynamic Time Warping (DTW) is among the most well-known algorithms of the latter type. According to Maus et al. (2016), the DTW distorts the time to adjust the two series to be compared. However, the time elapsed between observations is an important attribute in remote sensing analyses. To solve this problem, Maus et al. (2016) proposed a time-weighted extension of the DTW called Time-Weighted Dynamic Time Warping (TWDTW). The TWDTW shows great potential to discriminate between classes of pasture, agriculture, and forests, for example, due to the characteristic variation of the spectral response of the targets of these classes throughout the year. However, TWDTW does not accurately distinguish between primary forest and secondary vegetation areas, given that these classes usually present very similar behavior throughout the time series. TWDTW is currently available in the dtwSat package (MAUS et al., 2019), in R. Other machine learning methods adapted for time series are available in the SITS (Satellite Image Time Series Analysis) package (E-SENSING, 2019), also developed in the R language. Additionally, SITS unites a set of native tools to classify time series of remote sensing images with a large number of observations into LULC trajectories. This package includes tools to import remote sensing time series directly from web services, and to view, group, and filter these series, as well as to validate the classification results. For sparse time series, i.e., those with one image per date of interest, the classifiers contained in SITS tend to behave like traditional supervised classifiers. Currently SITS is not equipped to deal with invalid transitions.

It is common to treat invalid transitions in multitemporal sets of land cover classifications as classification errors that should be masked or corrected in post-classification steps. Ways to correct these transitions include manual editing and the use of temporal filters (GRIFFITHS et al., 2018; SOUZA et al., 2020). A relatively simple way to prevent the occurrence of invalid transitions in the classification process is to use masks, as demonstrated by the methodology

applied in the Program for Deforestation Monitoring in the Brazilian Legal Amazon (PRODES). Areas previously deforested or that were not originally covered by forests are removed from the annual analysis in PRODES, so only deforestation of primary forests is quantified (CÂMARA et al., 2006) and invalid transitions, like a deforested area turning into a primary forest, are not mapped.

It is also possible to directly classify time series images into LULC trajectories by using supervised classifiers and labeled samples of the trajectories of interest (ZHU, 2017; ANJOS et al., 2015). The selection of samples is usually carried out in a section of the time series of images and used to define the rules to identify the expected trajectories throughout the time series. According to Zhu (2017), this method has a high computational cost and requires prior knowledge about the types of trajectories to be detected.

In addition to these methods, algorithms based on Markov processes have recently been proposed to classify LULC trajectories. These algorithms incorporate both the observed pixel values and information about the probability of joint occurrence between LULC classes ordered over time. One example of proposal includes the Virtual Land Cover Engine (VLCE) (HERMOSILLA et al., 2018). The VLCE is a set of algorithms used to incorporate the temporal dimension in an extension of the well-known spatial context model Iterated Conditional Modes (ICM). This extension allows the incorporation of knowledge on the probabilities of transitions to correct LULC classifications obtained by Random Forest. It is an iterative algorithm that requires several classification steps, including the calculation of probabilities of successions of different classes before and after change events mapped externally to VLCE.

2.2.3 Quality assessment

This step still presents major implementation challenges. It usually depends on the existence of reference data collected in the field, using images with higher spatial resolutions, and/or based on auxiliary data. These types of data can be difficult to acquire when it comes to multi-temporal studies involving several observations because they depend on the systematic collection of data for the same study area over time (FOODY, 2010; LU et al., 2014). In general, quality assessments of LULC trajectories are based on the calculation of accuracy indices derived from contingency matrices, also known as confusion or error matrices, being common to independently calculate the accuracy of each LULC classification for each date as a proxy to assess the accuracy of the LULC trajectory. Given the lack of reference

data for some dates, particularly the oldest dates, the accuracy of one classification is sometimes considered to be the accuracy of the other classifications obtained with the same classes, data, and methods, which may not be correct. According to Sexton et al. (2013), due to the intrinsic variations of each date either in the land cover or the images, classification errors are not constant across years, even when the same techniques and data are used. To calculate accuracy indices for all dates, some studies have focused on collecting reference samples from the same remote sensing images used in the classification. In such cases, it should be noted that the calculated accuracy indices are likely to be overestimated. Nonetheless, different tools can facilitate the selection process of reference samples, including those specially developed to observe time series data. Some of these tools have been analyzed by Jakimow et al. (2020).

Regardless of the selected tool, a crucial step in analyzing the quality of LULC trajectories is to define which feature of the trajectories is to be evaluated. In some cases, the correct classification of a certain set of transitions is sufficient, regardless of the exact moment that specific classes/changes were observed. Other studies may demand the correct observation of each class for each time. Specifically, attention should be paid to the misclassification of areas containing classes under transition. Even when carrying out an analysis of a LULC classification on a single date, the classes used could represent stages in a gradient, with unclear limits between them (POWELL et al., 2004). Powell et al. (2004), for example, analyzed error sources in LULC classifications based on Landsat 5/TM images in an area in the state of Rondônia, Brazil. According to the authors, the dominant classes in the study area were in a gradient between pasture and secondary vegetation. Reference samples were obtained from high-resolution images and based on the visual analysis of five trained interpreters. These interpreters disagreed in almost 30% of the samples, mainly in those located along borders between targets, in mixed pixels (pixels covering more than one class), in transition areas (classes overlapping the gradient), or due to geometric correction problems (POWELL et al., 2004). Thus, the accuracy of the reference samples must be evaluated and considered in this type of analysis, as demonstrated by Foody (2010) and Olofsson et al. (2014).

In studies that involve many dates, however, this type of problem should not heavily affect the classification of images sensed before and after the transition date(s). If the exact moment when an area has changed from one class to the next on the gradient is not important, the obtained trajectory is similar, regardless of which of the two classes is assigned to the area under transition. Therefore, misclassifications

in certain dates do not necessarily imply errors in LULC trajectories. Since forest regeneration is a continuous process, it is natural that the classes involved in the analysis represent a gradient. Fallow agriculture/shrubby pasture areas can be considered transition classes between agriculture/pasture and secondary vegetation, for example. On the other hand, secondary vegetation classes can also be divided into different development stages, and incorrect classification of a stage on a given date may not affect the trajectory analysis. Nonetheless, possible problems regarding class definition (when an area effectively becomes secondary vegetation and at what stage of development) and classification (capacity for discrimination among these classes) can be particularly troublesome in some forest regeneration studies. This characteristic hinders the establishment of the exact time of vegetation formation, even in studies with observations for all dates of interest. Therefore, analyses that seek to quantify secondary vegetation areas on each date or calculate the age of secondary vegetation demand greater care.

Other methods to evaluate trajectory quality involve analyzing LULC trajectories based on logical rules regarding the probability of transitions, without using reference samples (LIU; ZHOU, 2004; AZEREDO et al., 2016). This type of analysis is particularly useful for the identification of invalid transitions as classification errors. These errors can be represented by a transition validity map (REIS et al., 2020a).

It is also possible to assess the quality of the classifications with uncertainty analyses (REIS et al., 2017a) despite these not being a direct indicator of accuracy. An example of this type of analysis is to vary the LULC classification approach (classifier, parameters, set of training samples) and account for variations in the LULC classification/trajectory results (i.e., identify pixels that lead to unstable classifications). Considering supervised classifiers, it is common for algorithms to assign classes to each pixel/object as a function of the lower or higher values of some measurements calculated for each class, such as probability or distance. For results obtained using these classifiers, Wulder et al. (2018) also mention the possibility of generating quality maps with the differences between the highest and second-highest values of these measures. In these analyses, smaller differences indicate greater uncertainty in the classification. Other examples of uncertainty indices are presented and discussed in Gonçalves et al. (2005).

2.3 Extraction and analysis of information

LULC trajectories can be used, for example, to identify the date of LULC changes (ZHU et al., 2016); to quantify the area of each

class in each date, as well as the area of specific changes/transitions (YUAN et al., 2005; HERMOSILLA et al., 2015; FRANKLIN et al., 2015); and to analyze the occurrence and spatial patterns of trajectories (MENA, 2008; MÜLLER-HANSEN et al., 2017). The visual interpretation of results and calculation of indexes from the LULC classification from one or two consecutive dates are among the usual methods. Some of the commonly used indexes are: fragmentation indexes, the percentage/area occupied by each LULC and/or change class, the number and shape of features of a given LULC/change class, the distance between features, the area-perimeter ratio of features, and Shannon's diversity index, among others (GILLANDERS et al., 2008). It should be noted, however, that a comparison of information derived for each date or bi-temporal changes do not configure into an analysis of trajectories *per se*. Despite the potential to generate important results, much of the temporal information contained in the trajectories is ignored by the analyst due to this type of simplification.

The study conducted by Carvalho et al. (2019) exemplifies how to use information derived for each date or pair of dates when carrying out a forest regeneration analysis. The authors evaluated changes in the historical patterns of secondary vegetation accumulation in the state of Pará between 2004 and 2014 using data from TerraClass (TerraClass data are described in Section 2.4). According to the authors, the conversion rate of secondary vegetation to clean pastures/mechanized agriculture increased after 2010, which corresponds to a period with a decrease in the rate of primary forest deforestation. For the same period, the authors identified decreased areas of classes with supposed greater regeneration potential: occupation mosaics and regenerating pastures. Thus, the authors point to a possible tendency for secondary vegetation areas to decrease in the years following the analysis. Additionally, the authors found differences in the spatial concentration of secondary vegetation in the northern and southern regions of the state of Pará. These regions present different historical rates of occupation and deforestation of primary forests.

Information on the age of the secondary vegetation (NUNES et al., 2020; SILVA JUNIOR et al., 2020) and the duration and/or intensity of regeneration processes can also be extracted from LULC trajectories. For example, Müller et al. (2016) and Jakovac et al. (2017) extracted the number and duration of forest regeneration cycles from more than 29 years of Landsat time series data. Müller et al. (2016) first detected deforested areas and then classified secondary vegetation areas along the Cuiaba-Santarém highway between the states of Pará and Mato Grosso. The authors observed differences in the duration and frequency

of regeneration cycles, depending on the types of predominant land use and the proximity of forests. Jakovac et al. (2017) segmented and classified an image time series from the municipalities of Tefé and Alvarães, in the state of Amazonas, using the Breaks for Additive Season and Trend (BFAST) (VERBESSELT et al., 2010) and Random Forest algorithms. The authors observed an average decrease in the regeneration cycle duration between the periods 1987-2000 and 2001-2014 in areas with shifting agriculture, which indicates that an agricultural intensification process occurred in the region (JAKOVAC et al., 2017).

Nevertheless, one study can analyze information on each date, pairs of dates, and trajectories jointly. Wang et al. (2020) used TerraClass data for the entire Brazilian Legal Amazon between 2000 and 2014 to analyze the deforestation of secondary vegetation. Their analysis consisted of comparing consecutive data pairs, which were stratified according to the age categories of secondary vegetation determined by stacking TerraClass data. The authors observed two distinct phases of secondary vegetation deforestation: 1) between 2000 and 2008, and 2) between 2008 and 2014. In the first period, they observed decreasing secondary vegetation/primary forest deforestation rates. In the second period, they observed increasing deforestation of secondary vegetation along with decreased deforestation of primary forest. The authors suggest that an increased pressure on forest systems in the second period had been absorbed by areas that were regenerating. The authors also concluded that 91% of secondary vegetation areas had been converted to pasture at some point in time, regardless of the age of that vegetation.

Another common question in LULC trajectory studies is what are the main types of existing trajectories. LULC trajectories can be typified by the duration/intensity of the observed processes or by their composition, i.e., the ordered sequence of LULC classes. Two main composition-based methods to define LULC trajectory types were identified. The first method consists of exhaustively defining the typologies based on the observed LULC transitions (PINHEIRO et al., 2016; CORSINI, 2018; ASSIS et al., 2020). The second method is the use of automatic clustering algorithms to identify groups of pixels with the same spatio-temporal pattern (AZEREDO, 2017).

To cite an example of the first composition-based method, Corsini (2018) observed differences in aboveground biomass patterns and how these had been impacted by fire and deficit of water in defined types of forest regeneration trajectories in the Brazilian Legal Amazon. These trajectories corresponded to certain class sequences

in TerraClass data for the years 2004, 2008, and 2010. This type of analysis is feasible for a few dates since an increase in the number of analyzed dates generally leads to an even bigger increase in the number of possible trajectories. Given that forest regeneration processes can involve diverse LULC trajectories, identifying all possible typologies *a priori* can be a very complex and expensive task. It can also result in subjective and/or synthetic analyses that focus on only a few previously identified processes.

An example of the second method is the clustering algorithm Grouping by Similarity of Temporal Evolution (GSTE) proposed by Azeredo (2017). The GSTE combines the traditional DTW, Classical Multidimensional Scaling (CMDS), and K-Means Clustering algorithms to analyze how a trajectory behaves over time. According to the author, the GSTE clusters portray characteristics such as the time interval between different processes, the duration of each process, and the classes observed at each given point in time. However, whereas LULC trajectories are generally presented in categories, the current version of GSTE is only applicable to numerical data. LULC classes need to be converted to values associated with different intensities of a single process to carry out a GTSE analysis, which may not be directly feasible when multiple processes are involved. Nonetheless, different clustering methods for categorical data have been proposed in the literature and should be evaluated regarding their applicability for LULC trajectory analyses, such as K-Modes (HUANG, 1998), RObust Clustering using linKs (ROCK) (GUHA et al., 2000), and Generalized Self-Organizing Maps (GSOM) (HSU, 2006).

Furthermore, new techniques to analyze LULC trajectories have been proposed by Azeredo et al. (2016) and Maciel et al. (2019). Azeredo et al. (2016) adapted some patterns from the mobile objects literature to study LULC trajectories. These patterns have been used in forest degradation analyses. Although these patterns have not yet been applied to forest regeneration studies, they could be used to identify:

a) trajectories that converge to specific classes that denote forest regeneration processes;

b) trajectories with a given set of LULC classes in specific time windows. This enables secondary vegetation to be identified in certain time windows rather than on individual dates, thus decreasing quantification problems in areas under transition;

c) trajectories that present similar LULC changes, either in the same or

different periods. This is useful to identify the deforestation of secondary vegetation for certain types of use;

d) trajectories that present similar process intensities, such as similar fallow/regeneration times;

e) trajectories with inconsistent/invalid transitions.

Maciel et al. (2019) formalized a space-time calculation logic called 'LUC Calculus'. This tool can be used to investigate and characterize LULC trajectories according to four main predicates: conversion (CONVERT), recurrence (RECUR), evolution (EVOLVE), and maintenance (HOLD). These predicates, adapted to the context of forest regeneration analyses, can be interpreted as:

a) CONVERT denotes the change between classes for two consecutive times of observation. Usage example: detection of all secondary vegetation areas that have been converted for agricultural use;

b) RECUR denotes the observation of a recurring class after the conversion of that class to another. Usage example: to correct invalid transitions and distinguish secondary vegetation and primary forest areas by the historical analysis of the trajectory;

c) EVOLVE denotes events with one or more different classes interspersed between specific LULC changes. Usage example: to identify deforested primary forests that, after being abandoned, have regenerated to secondary vegetation. These areas can be identified regardless of their intermediate use;

d) HOLD denotes a class that has remained the same throughout all observations over a determined period. Usage example: to identify areas of persistent secondary vegetation.

2.4 Available map products

Different programs that monitor tropical forests on either local or global scales have generated LULC change data for the Amazon that can be used as auxiliary data in LULC trajectory analyses. Among these are programs focused on deforestation and/or degradation of primary forests, such as PRODES (INPE, 2020c), both the Deforestation Detection System in Real-Time (DETER) and its incorporated or

derivative programs, such as the Forest Degradation Mapping in the Brazilian Amazon (DEGRAD) and intense DETER (INPE, 2020b), the Deforestation Alert System (SAD) of the *Instituto Homem e Meio Ambiente da Amazônia* (IMAZON, 2020), and data from the Global Land Analysis and Discovery (GLAD) laboratory, such as the Global Forest Change data set (HANSEN et al., 2013). It is also possible to use natural resource maps, such as the pioneers RADAM (Radar in the Amazon) and RADAMBRASIL (ROESSEL; GODOY, 1974), as well as multi-temporal LULC classifications, such as the *Monitoramento de Uso e Cobertura da Terra do Brasil* (IBGE, 2015, 2017b, 2020). However, projects that map deforested areas into LULC classes, among which secondary vegetation necessarily needs to be discriminated, are of particular interest to forest regeneration studies. Examples of such projects in Amazon include TerraClass and the Annual Mapping of Land Cover and Use in Brazil (MapBiomas). Their main characteristics are summarized in this section.

TerraClass is developed by the National Institute for Space Research (INPE) in partnership with the Brazilian Agricultural Research Corporation (Embrapa). In this project, deforested areas mapped by PRODES in the Brazilian Legal Amazon are classified into land cover classes of interest. There are available classifications for years 2004, 2008, 2010, 2012, and 2014, and the project to generate the ones for years 2018, 2020, 2021, and 2022 (EMBRAPA; INPE, 2021). Unpublished, there are also classifications for 1991 and 2000. The TerraClass project is based on classifying images from the Landsat family (or those with similar resolutions) and data from the Moderate-Resolution Imaging Spectroradiometer (MODIS). Different image processing techniques and stages of visual analysis, manual editing, and auditing are used in this project. The results are freely available with two legend versions, both with 30 m spatial resolutions, and with three classes inherited from PRODES data (Hydrography, Non-forest, and Forest). In the first version, 13 land cover classes are mapped by TerraClass: Deforestation of the year, Annual agriculture, Pasture with exposed soil, Clean pasture, Shrubby pasture, Regeneration with pasture, Secondary vegetation, Reforestation (except for the 2008 data), Urban area, Mining, Mosaic of occupations, Others, and Unobserved area (ALMEIDA et al., 2016). Almeida et al. (2016) calculated a Global Accuracy of 76.6% and a Kappa index of 0.67 for the classifications in the states of Pará and Mato Grosso in 2008, with ground samples of the classes Annual agriculture, Mosaic of occupations, Clean pasture, Shrubby pasture, Regeneration with pasture, and Secondary vegetation. The authors observed the greatest misclassification rate among pasture classes. After merging the Clean pasture, Shrubby pasture, and Regeneration with pasture classes, the Global

Accuracy and Kappa values increased to 89.7% and 0.79, respectively. In the second legend, the classes are: Natural secondary forest vegetation, Deforestation of the year, Perennial agricultural culture, Semi-perennial agricultural culture, Temporary agricultural culture, Shrub cultivated pasture, Herb cultivated pasture, Silviculture, Urbanized area, Mining, Unobserved, and Others.

Since TerraClass masks deforested areas detected by PRODES, of internationally recognized reliability, secondary vegetation areas can only be identified in previously deforested areas (ALMEIDA et al., 2016). This methodological aspect reduces the misclassification between forested and secondary vegetation areas considerably. However, it also means that areas of secondary vegetation can only be observed if the deforested area has been previously detected by PRODES, which in turn only detect deforested areas larger than 6.25 ha. TerraClass also does not map important classes to the characterization of forest regeneration cycles: small-scale agriculture/shifting cultivation. The use of TerraClass data for studies focused on this class usually encompasses a refinement step, in which these areas are mapped using high or very-high resolution images and aggregated to TerraClass products (SOUZA et al., 2019). Furthermore, TerraClass data are not annual and do not encompass older dates. As a result, older regeneration cycles would not be entirely observed if one were to only use this data set. Moreover, TerraClass data have misalignment problems inherited from PRODES (INPE, 2017; WANG et al., 2020). These problems must be corrected in order to carry out pixel-level analyses. Therefore, TerraClass data are indicated either to characterize short trajectories, as demonstrated in Corsini (2018) and Wang et al. (2020) and detailed in Section 2.3, or to be used as an auxiliary data for the LULC trajectory classification process, such as a classification mask, base classification to be updated, and/or for collection of reference samples.

The MapBiomas project is produced by a collaborative network of institutions. This project aims to produce annual LULC classifications by using automatic techniques to classify Landsat images. Data sets usually cover the entire Brazilian territory, plus other special areas depending on the collection. In general, mosaics constructed from Landsat image statistical variables (median, amplitude, and standard deviation, etc.) for each year are classified independently. Invalid transitions are then corrected in the post-classification process with spatial and/or temporal filters based on pre-established rules (SOUZA et al., 2020). The images, LULC classes, accuracy, and methodological details vary depending on the MapBiomas collection. The Global Accuracy values are estimated from samples collected from visual interpretation of

the images. To date (August 2022), seven collections, accounting for ten different data sets, have been published. There is no distinction between primary, degraded, or secondary vegetation classes in collections 3.x to 5 (MAPBIOMAS, 2022b). From collection 6, areas of growth of secondary vegetation are estimated based on the identification of distinct trajectory patterns within the LULC maps. To date, there is no information available about the accuracy of the secondary vegetation classes. Also, this methodology does not allow the identification of secondary vegetation in areas without evidence of previous anthropic use, or fallow cycles with less than three to seven years (MAPBIOMAS, 2022a).

Collections 2.x are the last ones that separated the primary forests and secondary vegetation classes within the classification step. These contain annual data between 2000 and 2016, with Global Accuracy indexes for each year that vary at around 80% for the Amazon biome. Neves et al. (2020) harmonized the legends of the TerraClass (first legend) and MapBiomas collection 2 maps and then compared the classification results between the overlapping areas of the 2004, 2008, 2010, 2012, and 2014 maps. Considering all years in a joint analysis, the authors found that only 3.56% of the areas classified as secondary vegetation in TerraClass products were also classified as such in the MapBiomas 2 products. The remaining areas of secondary vegetation in TerraClass results were mostly classified as Forest (80.34%) and Pasture (13.96%) by MapBiomas. Despite these differences, the agreement among both classifications is around 87% of the pixels. This happens because the predominant class in both maps is forest, which can influence the values of Global Accuracy presented by the MapBiomas data.

There are also studies aimed to reclassify forests that occur in areas with evidence of deforestation as secondary vegetation in other MapBiomas collections (NUNES et al., 2020; SILVA JUNIOR et al., 2020). In their case, evidence of deforestation means the previous observation of a class other than forest. Nunes et al. (2020) used this approach to separate primary forest areas from secondary vegetation areas in the MapBiomas data from collection 3.1, for the Amazon biome. From these data, the authors grouped the pixels of secondary vegetation by estimated age and calculated both the area covered by secondary vegetation in the biome and the area of secondary vegetation deforestation. They also presented carbon capture estimates. It must be highlighted that this approach is unable to identify secondary vegetation areas if there is no evidence of deforestation throughout the time series. Therefore, Nunes et al. (2020) found smaller proportions of deforested areas occupied by secondary vegetation than

TerraClass for corresponding periods. Using a similar approach, Silva Junior et al. (2020) created a database containing increment, extension, and age of secondary vegetation estimates for the years 1986 to 2018 using data from the MapBiomas collection 4.1. These results were also validated by TerraClass data through a method that compares the proportion of the class in a given regular cell-grid. The authors concluded that although there is statistical evidence that the average proportion of secondary vegetation found by processing MapBiomas data is smaller than that mapped by TerraClass (via Mann-Whitney test and p-values < 0.001) the results of the two data sets are comparable.

According to Maurano and Escada (2019), variations in each collection of the historical classifications of MapBiomas limit the applicability of this data when it comes to operational and continuous area estimates. In forest regeneration studies, these variations cause the observation of different trajectories, from which different rates of atmospheric carbon absorption can be calculated, for example. The inability to identify secondary vegetation areas without evidence of deforestation events can also be detrimental to analyses in regions dominated by older secondary vegetation, such as in the northeastern part of Pará state (CAPANEMA et al., 2019). MapBiomas data are suitable for preliminary analyses over large areas, which should, subsequently, be further assessed and possibly refined for more detailed or local forest regeneration studies.

2.5 Chapter conclusions

In this chapter, the main data and methods used to classify and analyze LULC trajectories were reviewed. They were also discussed in the context of their applicability in forest regeneration studies in the Brazilian Amazon. In general, one can either use 1) methods suitable for single images or pairs of images successively or 2) tools focused on the analysis of remote sensing time series images and LULC trajectories.

The studies and remote sensing products reviewed in this chapter are based mainly on Landsat images. These images are freely available for use with atmospheric and geometric corrections. Nonetheless, cloud cover continues to present scientific challenges in the Amazon Biome, mainly when it comes to carrying out historical analyses. Given that the combination of images from multiple sensors constitutes a possible alternative to increase the number of time series observations, methods that combine different types of data are important for LULC trajectories in the Amazon. Therefore, methods capable of dealing with sparse time series, multi-sensor

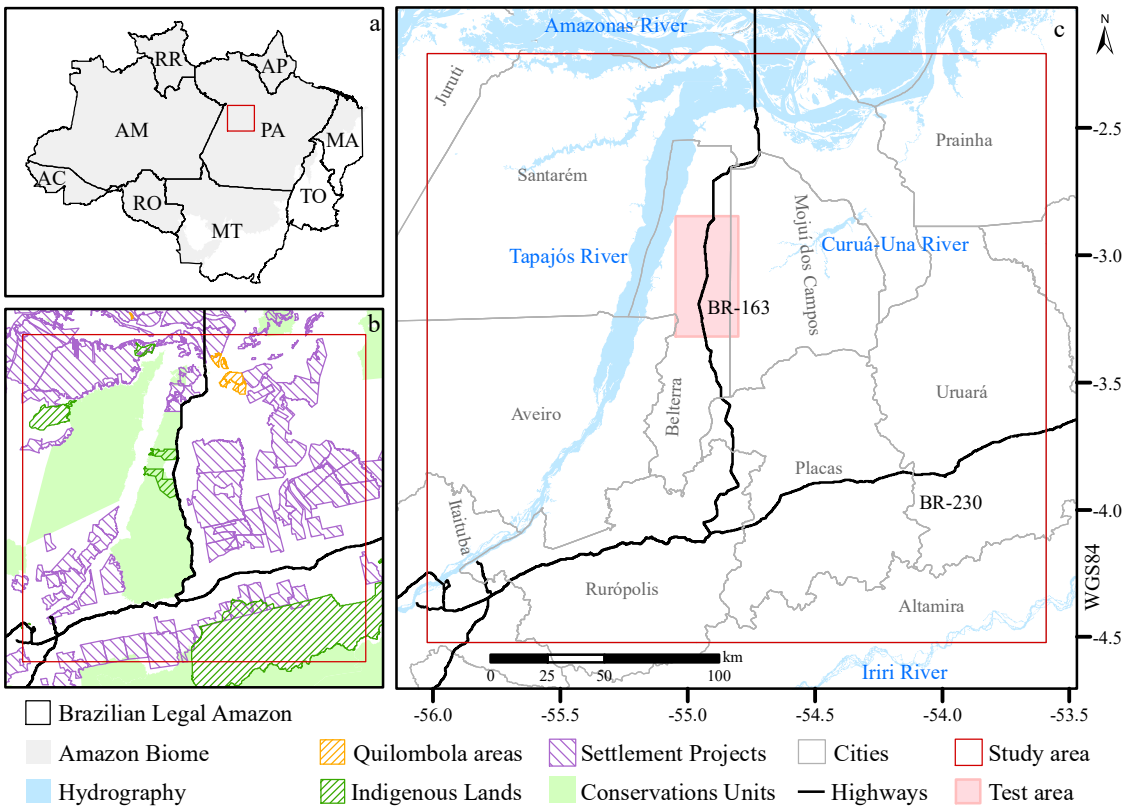
data, and different aggregations of a given legend are of particular interest to forest regeneration studies. Another recurring problem among trajectory classification techniques is that several steps are needed, including a post-classification step to correct invalid trajectories.

Despite the diversity of proposed techniques and tools to analyze LULC trajectories, quality assessment methods are mostly based on calculating accuracy indices for each observed time. Similarly, most studies are still based on mono or bi-temporal techniques. One must pay attention to possible overestimations of accuracy values in these studies. This is particularly important when quantifying LULC class/change areas for each date. In this sense, one of the major challenges when it comes to analyzing LULC trajectories is developing methodologies to estimate their accuracy and allow the adequate information extraction. The analysis of long LULC trajectories is not a trivial task, with many studies opting instead to compare single points in time. The notable exceptions are those that analyze multi-temporal attributes such as duration and number of regeneration cycles. Therefore, there is a pressing need to evaluate LULC trajectories analysis tools that permit the analysis of trajectories with a greater number of dates and classes of interest since such tools could enable regeneration processes hitherto not discriminated to be identified.

3 STUDY AREA

This study focus on an area located in the Lower Tapajós region, in the State of Pará, within the Brazilian Amazon. This area is illustrated in Figure 3.1.

Figure 3.1 - Study area.



a) localization in relation to the Brazilian Legal Amazon; b) different types of land occupation within the study area; c) localization of the study and test area in relation to municipal limits, major highways, and main rivers.

SOURCE: Based on INCRA (s.d.), INPE (2019), IBGE (2019), MMA (2020), FUNAI (2021), and MInfra (2021).

The study area has approximately 69 thousand km^2, from which nearly 13 thousand km^2 corresponds to Conservation Units, 6.2 thousand km^2 corresponds to Indigenous Lands, 265.5 km^2 corresponds to Quilombola areas, and 32.5 thousand km^2 corresponds to Settlement Projects instituted by the National Institute for Colonization and Agrarian Reform (INCRA) and other areas influenced by benefits from the agrarian program (including the area occupied by the National Forests

and the Extrativist Reserves). This region is also inserted in the area of influence of the BR-163 Ecological-Economic Zoning (ZEE) and the BR-163 Sustainable Forest District (DFS). As such, this region has been the target of different types of studies, with history of occupation and public policies generally well documented.

The region presents a humid tropical climate, with average annual rainfall around 1820 mm. The higher precipitations occur from January to May, with a marked dry period between August and November. The annual average temperature is around 25°C, varying less than 5°C throughout the year. The region is inserted into two morpho-structural units, the Lower Plateau of the Amazon (*Planalto Rebaixado da Amazônia*), with an average altitude of around 100 m above sea level, and the Tapajó-Xingu Plateau (*Planalto Tapajós-Xingu*), where the altitude ranges between 120 and 170 m (RADAMBRASIL, 1976; IBAMA, 2004). Altitudes vary between 50 and 600 m with areas of low or strong slope, depending on the morphostructural unit (EMBRAPA, 2010a; FURTADO; PONTE, 2013). The area is drained mainly by the Tapajós River, navigable from Santarém to Itaituba throughout the year. Stretches of the Amazon and Iriri rivers are also present in the delimited area. The predominant typology of original vegetation formation in the study area is the Dense Ombrophilous Forest (IBGE, 2004a; EMBRAPA, 2010a). The dominant soils in the study area are argisols and latosols, with good physical properties and low natural fertility (EMBRAPA, 2010b). Areas of *terra roxa* are also found, mainly near the Transamazon highway (BR-230). Considering adequate management practices, this region presents a mosaic of areas considered good for agriculture and those good for livestock (EMBRAPA, 2016).

Information about the land cover elements was collected on the field for a subset of this region depicted in Figure 3.1 as 'test area'. This area was revisited for five years between 2009 and 2017. These data are detailed in Appendix A.

3.1 Occupation history and public policies that influenced the study area

This section presents a brief historic of occupation and an overview of public policies that influenced the study area. These were discussed regarding the Amazon's three major phases of development, proposed by Becker (2007):

a) 1616 - 1930: period of territorial formation;

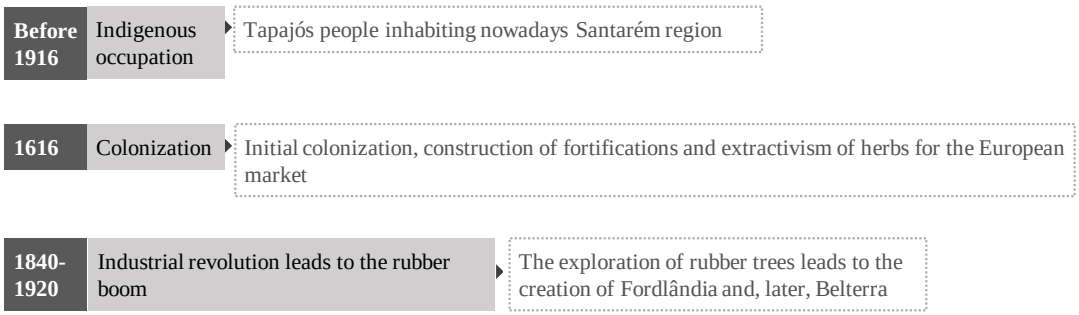
b) 1930 - 1985: period characterized by the importance attributed to regional

planning and the role of the State in the configuration of space;

c) 1985 - nowadays: period characterized by the implementation of policies aimed either at the economic appreciation or at the ecological preservation of the region.

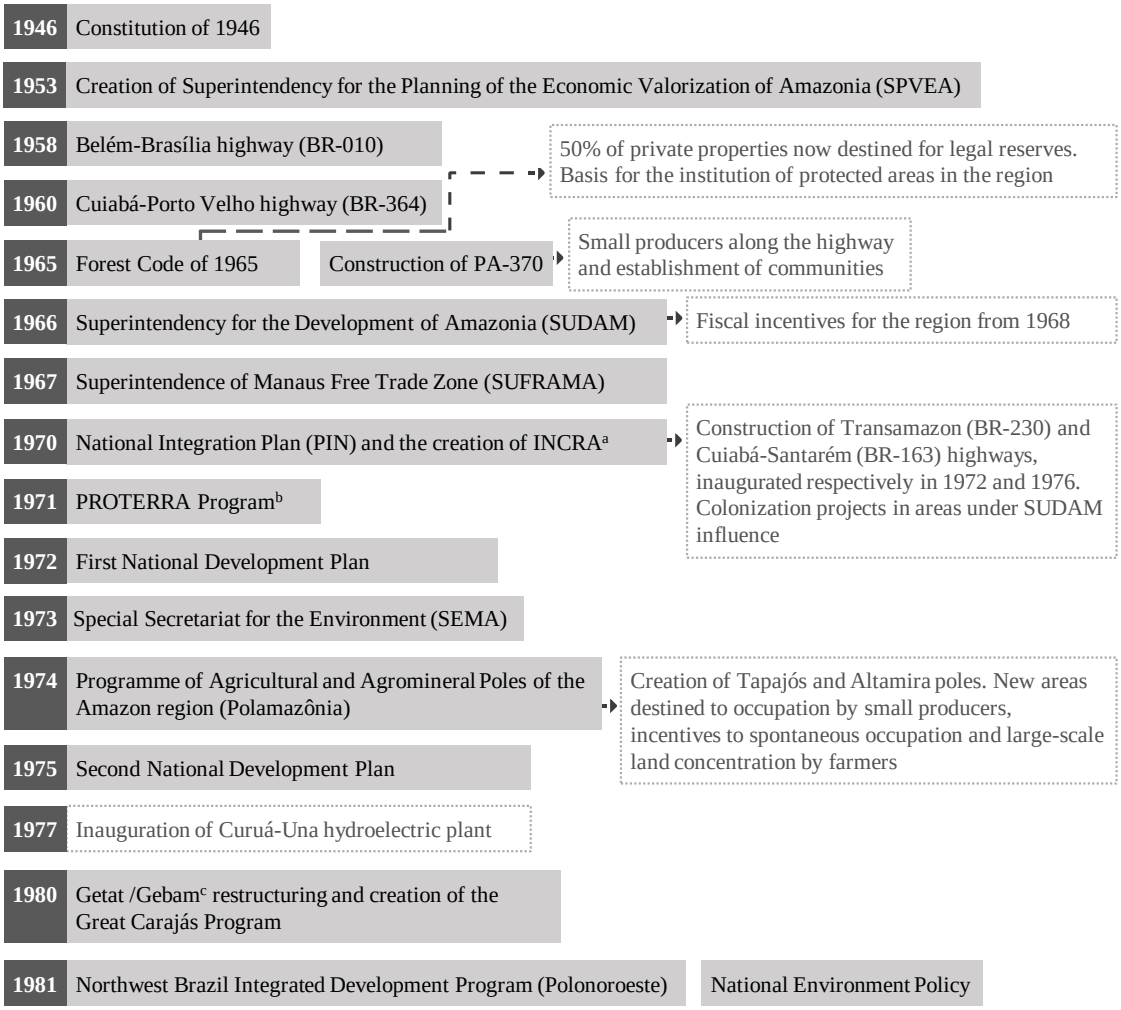
Although the periods before the 1980s are not the object of the study proposed in this thesis, their influences on social organization and impacts on the land use in the region persist to the present day. The main events that occurred in the Brazilian Amazon in these three periods are illustrated in Figures 3.2 to 3.4, as a basis to contextualize the main events that occurred within the study area to the national ones. Note that the period from 1616 to 1930 was only briefly discussed here. Details on these events are detailed in length in studies like Santos (1980) and Weinstein (1993).

Figure 3.2 - Main events, projects, and government policies in the Amazon (up to 1930).



Impacts within the study area are depicted in the dotted boxes.
SOURCE: Based on Becker (2007) and Becker and Stenner (2008).

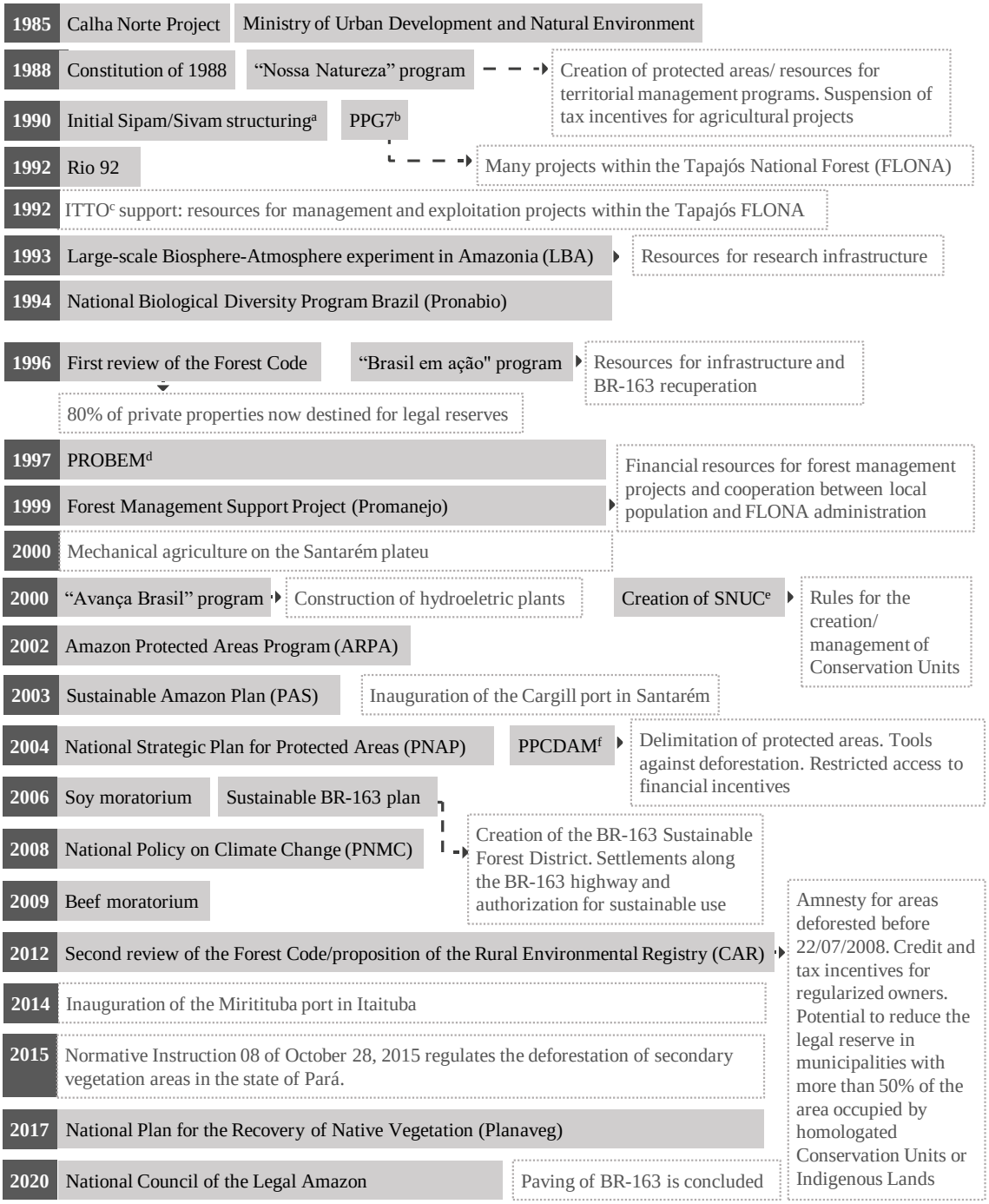
Figure 3.3 - Main events, projects, and government policies in the Amazon/study area (1930 - 1985).



Impacts within the study area are depicted in the dotted boxes. [a] National Institute for Colonization and Agrarian Reform, [b] Programme for Land Redistribution and Stimulation of Agroindustry in the North and Northeast, [c] Executive Group for Lands of the Tocantins-Araguaia (Getat)/ Executive Group for the Lower Amazon Region (Gebam).

SOURCE: Based on Becker (1990), Mello (2006), Becker and Stenner (2008), and Santos (2020).

Figure 3.4 - Main events, projects, and government policies in the Amazon/study area (1985 - nowadays).



Impacts within the study area are depicted in the dotted boxes. [a] Amazonian Protection System (Sipam)/Amazonian Surveillance System (Sivam), [b] Pilot Program to Conserve the Brazilian Rain Forests, [c] International Tropical Timber Organization, [d] Brazilian Molecular Ecology Program for the Sustainable Use of Biodiversity, [e] National System of Conservation Units, [f] Action Plan for Prevention and Control of the Legal Amazon Deforestation.

SOURCE: Based on Becker (1990), Mello (2006), Becker and Stenner (2008), and MMA (s.d.)

The history of occupation of the study area dates from the beginning of the 17^{th} century, with the Portuguese expeditions, followed by the foundation of villages by the Jesuits (BECKER, 2007; BECKER; STENNER, 2008). Santarém, for instance, originated from one of such villages (IBGE, 2017a). This occupation was characterized by the construction of fortifications and the extraction of spices with high value in the European market. Before that date, there were already indigenous civilizations inhabiting the floodplains of the Tapajós river, with groups probably composed of extractivists, fishermen, and manioc farmers (BECKER; STENNER, 2008).

The effective occupation of the region, accompanied by significant economic and demographic growth, took place with the Industrial Revolution and the consequent expansion of the rubber trade and rubber extraction, between 1840 and 1920 (SANTOS, 1980; WEINSTEIN, 1993). Due to rubber extraction, almost the entire southwestern region of Pará was explored, which led to the implementation of Fordlândia, where Aveiro is located today, and then the project that structured the Belterra region, both by Henry Ford (BECKER; STENNER, 2008; IBGE, 2017a). The exploration of rubber also fostered the occupation along the river banks by communities formed by rubber tappers coming from the Northeast region of the country (MMA, 2019). On a much smaller scale, *Pau Rosa* (*Aniba rosaedora Duke*) was explored in the region of Santarém following this period (SANTOS, 2020). This type of exploration led to the aperture of the so-called Paxiúba road, which attracted settlers that founded a few communities in the region.

The second expressive migratory event to the region occurred in the context of the I National Development Plan (I PND), between 1970 and 1972, and the creation of the National Integration Program (PIN). The PIN aimed to expand the road network, through the immediate construction of the Transamazon (BR-230) and Cuiabá-Santarém (BR-163) highways. Furthermore, areas up to 10 km to the left and right of these highways would be destined for colonization and the agrarian reform (BRASIL, 1970), and distributed by the newly created INCRA. In 1971, the Decree-Law N^o 1.164/1971 (BRASIL, 1971) declares the vacant lands located in up to 100 km on each side of the highways in the Legal Amazon 'indispensable to the national security and development'. In the same year, an area of around 64 mil km^2 along the Transamazon highway, between Altamira and Itaituba, was declared as of social interest, and privately owned lands were expropriated. These events led to the official colonization of these areas and included, besides the implementation of the highways and settlements, the opening of neighboring roads, demarcation of

rural lots, and construction of the first urban nuclei (MIRANDA, 1990). These actions brought families from various regions of Brazil to these colonization areas, especially from areas of tension over land disputes in the Northeast of Brazil (MIRANDA, 1990).

According to Mello (2006), as of 1973, INCRA began to favor the occupation of lots of up to 3 thousand hectares, which could be obtained through registration that required a cultivation plan and technical management. With the requirement of technical bases that the local population could not absorb or reproduce, this process would have led to the concentration of large extensions of land by a few owners. From 1974, the official character of land demarcation and settlements was diminished and the occupation movement in the Lower Tapajós region become spontaneous (MIRANDA, 1990).

Private initiatives are also privileged in the context of the II National Development Plan (II PND), between 1975 and 1979. The Superintendence for the Development of the Amazon (*Superintendência do Desenvolvimento da Amazônia* - SUDAM) directed fiscal incentives mainly to agricultural and industrial projects. The government also created, in 1974, the Programme of Agricultural and Agromineral Poles of the Amazon Region (*Programa de Polos Agropecuários e Agrominerais da Amazônia* - Polamazônia), with two poles in the Lower Tapajós region: the Tapajós and Altamira poles. This program aimed to concentrate resources to stimulate migratory flows, increase livestock herds, and improve the urban infrastructure of the Amazon (BECKER, 1990). This period is marked by the occupation of priority areas, as well as areas outside the colonization projects, by small producers, concomitant to the land concentration by large-scale farmers, with expulsion/expropriation of the settlers (SCHMINK; WOOD, 1992).

The concern about the undesirable effects of economic development on the quality of the environment was growing globally since the late 1960s. Although the more intense conservationist pressure over the Amazon region started in the 1980s, we can highlight several measures with Amazon-wide impacts from the 1970s, such as:

a) the institutionalization of the Special Secretariat for the Environment (*Secretaria Especial de Meio Ambiente* - SEMA), in 1973;

b) the 1965 Forest Code (BRASIL, 1965), which instituted the Biological Reserves, Areas of Permanent Preservation, and National Forests (FLONAS). It also determined that 50% of the private properties of Amazon were destined for Legal Reserves;

c) the Constitution of 1988, which incorporated instruments and advances for the preservation of Brazilian ecosystems. The Brazilian Amazon Rainforest was declared a National Patrimony. Furthermore, indigenous peoples were granted rights of ownership over their lands and respect for their cultures, which contrasts with the previous assimilationist vision;

d) the *Nossa Natureza* program, in 1988. This program resulted in the suspension of fiscal credit incentives for agricultural and cattle-raising projects in the Legal Amazon; the institutionalization of financial resources for the environmental area; the creation of Conservation Units (UC); the creation/restructuring of programs to control anthropic actions in the Legal Amazon; and the proposition of the Legal Amazon ZEE as a tool for territorial ordering (MELLO, 2006);

e) the Pilot Program for the Protection of Brazilian Tropical Forests (PPG7), in 1990. It has helped to spread principles, techniques, and alternative productive systems by funding and implementing sustainable forest management projects (MELLO, 2006). For the Lower Tapajós Region, the PPG7 program enabled the creation of protected areas, the allocation of resources for territorial management programs, several studies, and the involvement of the community in environmental management. Resources and studies are continued within the region in the derivative Forest Management Support Project (*Projeto de Apoio ao Manejo Florestal* - Promanejo), with expressive results within the Tapajós National Forest: facilitation of cooperation between the local population and the Tapajós National Forest managers, as well as actions aimed at strengthening the management of the unit, which turns into a reference for other National Forests in the Amazon;

f) the Large-Scale Biosphere-Atmosphere Experiment in the Amazon (LBA), in 1993. This experiment aimed to understand the Amazon ecosystems and their relations with the regional and global climate. It fostered many studies in the Lower Tapajós region, as well as the installation of instrumented research sites in the region, from which we highlight the ones in the Tapajós National Forest (BATISTELLA et al., 2007).

In 1996, the first revision of the Forest Code determined that 80% of private properties in the Amazon should be allocated as legal reserves. This year also marks the resumption of the Union's territorial planning in Amazon (BECKER, 2007), by

the *Brasil em Ação* Program, followed by the *Avança Brasil* program, respectively the Pluriannual Plan (PPA) of 1996-1999 and 2000-2003. Of particular expression in the Lower Tapajós region, Thery (2005) highlights the rehabilitation of the Cuibá-Santarém highway in the first program. These also enabled the creation of fiscal funds for the Amazon region that privileged agricultural activities focused on grains and cattle (MELLO, 2006). It is in the context of these PPAs that a new immigration process is initiated in the Lower Tapajós region, focused on the exploration of mechanized agriculture (SANTOS, 2020; MMA, 2019).

The 2000s are, once again, characterized by the institution of different Amazon-wide programs focused on the preservation, inspection, and/or sustainable development of the Amazon region. Among these, the main one is the Action Plan for Prevention and Control of Deforestation in the Legal Amazon (*Plano de Ação para Prevenção e Controle do Desmatamento na Amazônia Legal* - PPCDAm), initiated in 2004. This plan enabled the creation of Conservation Units and Indigenous Lands, as well as the development and/or improvement of forest monitoring systems, such as PRODES, DETER, DEGRAD, Detection of Selective Exploitation (DETEX), and TerraClass. It also promoted incentives for policies created in the 2010s, such as the Rural Environmental Registry (*Cadastro Ambiental Rural* - CAR); the Soy Moratorium; and the Legal Amazon Macro-Zoning (MMA, 2018; MELLO; ARTAXO, 2017). It also led to the creation of the Plan for Sustainable Regional Development for the Area of Influence of the BR-163 Highway (BR-163 Project), in 2006, which in turn led to (MMA, s.d.):

a) the implementation of settlements along the Santarém-Cuiabá highway;

b) the creation of the BR-163 DFS, a geo-economic and social complex that aimed to prioritize the implementation of public policies for the sustainable use of natural resources;

c) the allocation of resources for National Forests;

d) the concessions for the sustainable use of natural resources;

e) the implementation of the Law for the Management of Public Forests.

The 2010s, on the other hand, were characterized by the loss of significance in environmental laws and agencies in the Amazon. The most expressive one may be the second revision of the Forest Code, in 2012. Among the main changes in

the Amazon biome, we cite: amnesty of areas deforested before July 22, 2008; continuity of agrosilvopastoral activities and tourism in Permanent Preservation Areas consolidated before July 22, 2008; credit and tax incentives for landowners registered in the CAR; and the potential reduction of the Legal Reserve in municipalities with more than 50% of the area occupied by approved Conservation Units/Indigenous Lands.

Contradictorily, it is in this scenario that Brazil committed to restore and reforest 12 Mha by 2030, as part of their Nationally Determined Contributions in the Paris Agreement (BRASIL, 2016). To achieve this, and also to restore international credibility, the Brazilian Government created the National Plan for the Recovery of Native Vegetation (*Plano Nacional de Recuperação da Vegetação Nativa* - PLANAVEG), in 2017, and re-establishment of the National Council of the Legal Amazon (*Conselho Nacional da Amazônia Legal* - CNAL), in 2020, although with a different function, as the PPCDAm is currently inactive, and the CNAL operates without the participation of civil institutions.

In a nutshell, these many occupation events and programs led to varied realities within the Lower Tapajós region. Many municipalities and communities were established at the time of the rubber boom and/or as a response to government incentives for agricultural expansion and occupation of the region. Others originated from indigenous villages, such as Mayatapu, Tupinambá, and Kumaruara. These differentiated occupation histories, as well as political limits imposed by the State, led to differences in political opinions, access to resources, and mainly, in land management practices (AFFONSO et al., 2016). As a result, the region holds areas of extensive cattle raising, mechanized agriculture, as well as areas of shifting cultivation, and latex/wood extraction.

4 TOWARDS A REPRODUCIBLE LAND USE AND LAND COVER HIERARCHICAL CLASS LEGEND FOR THE LOWER TAPAJÓS REGION[1]

As previously discussed in Section 2.1.2, whereas a universal legend for the Brazilian Amazon is not feasible, having a standard legend for a given region could benefit a group of researchers focused on long-term studies, regarding the possibility to maintain class definition along the time or in studies conducted by different teams.

In this chapter, we propose a set of Land Cover Meta Language (LCML) formalized hierarchical LULC legends for the Lower Tapajós region. These legends were designed to be:

a) objective/reproducible: the proposed legends are formalized in LCML using quantifiable and easily recognizable physiognomic characteristics of land cover classes;

b) data-independent: classes in the most detailed level were defined considering field data and experience, in opposition to class separability in a given data set;

c) constructed around a common objective for the study area: legends are derived from a conceptual model that was based on commonly used legends and classification systems in the Amazon (ANDERSON et al., 1976; DI GREGORIO; JANSEN, 2005; IBGE, 2013);

d) open to the inclusion of necessary not predicted classes: new classes can be defined upon the use of the selected LCML classifiers;

e) hierarchical: different legends tiers/levels of detail are proposed to allow for studies in varied scales and, to a certain degree, to accommodate different types of data.

As a secondary objective of this chapter, we also present the comparison of classification results of a representative Landsat 5/TM image from 2010 using both the proposed legends as legends derived from the class separability of classes in this image. This analysis aims to provide results to better understand the usefulness of the proposed legends for the classification of LULC trajectories in the study area.

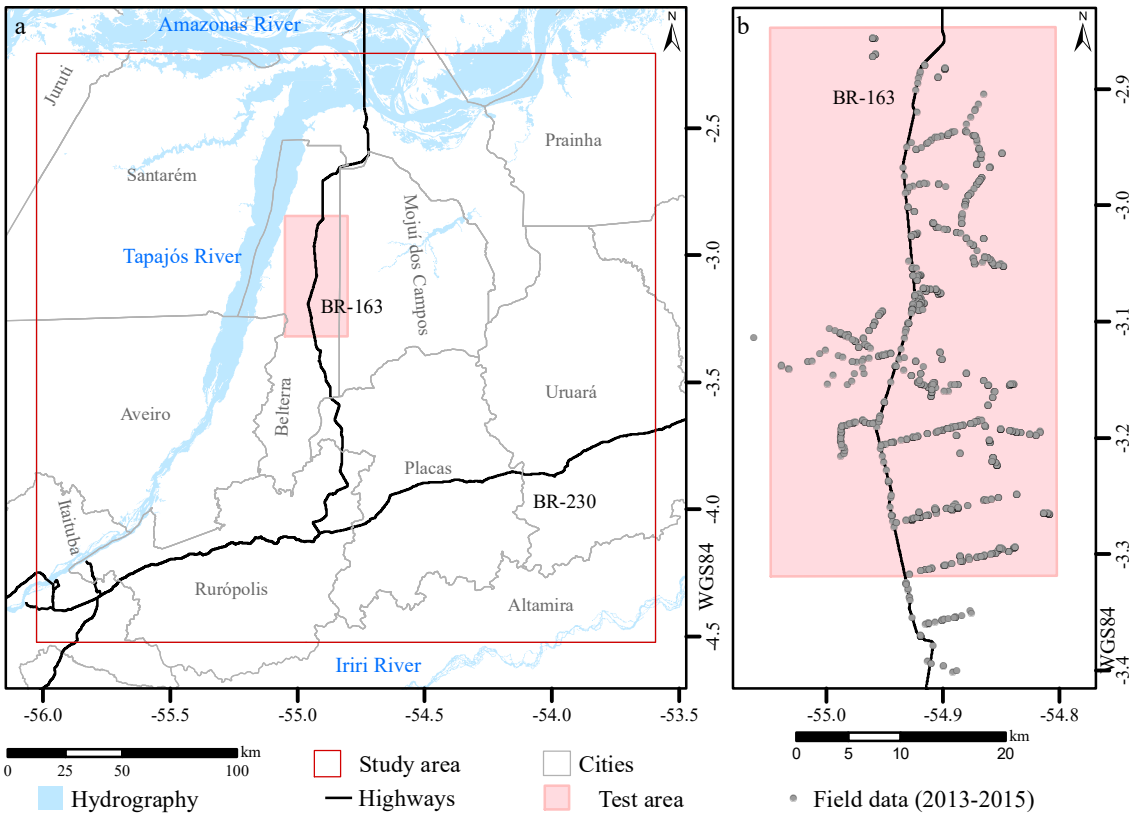
[1]This chapter is a summarized version of Reis et al. (2018).

This chapter is divided as follows. Firstly, we detail the used field data and remote sensing image in Section 4.1. The methods used to define the proposed LULC classes, as well as the image classification steps, are presented in Section 4.2. The proposed legends and classification results are presented in Section 4.3. The applicability of these products is discussed in Section 4.4. The main findings and their relation to the present thesis are detailed in Section 4.5.

4.1 Field data and remote sensing image

Field data were collected along km 60 to 120 of the Cuiabá–Santarém highway (BR-163) and adjacent areas. This area is illustrated in Figure 4.1.

Figure 4.1 - Test area of Chapter 4 in relation to the main study area.



a) area of image coverage (test area); b) zoom to show field data collection points. Due to variations in data collection, it is possible to have more than one point per terrain feature, whereas one point can represent the location of more than one feature/type of land cover.

SOURCE: Based on IBGE (2019) and MInfra (2021).

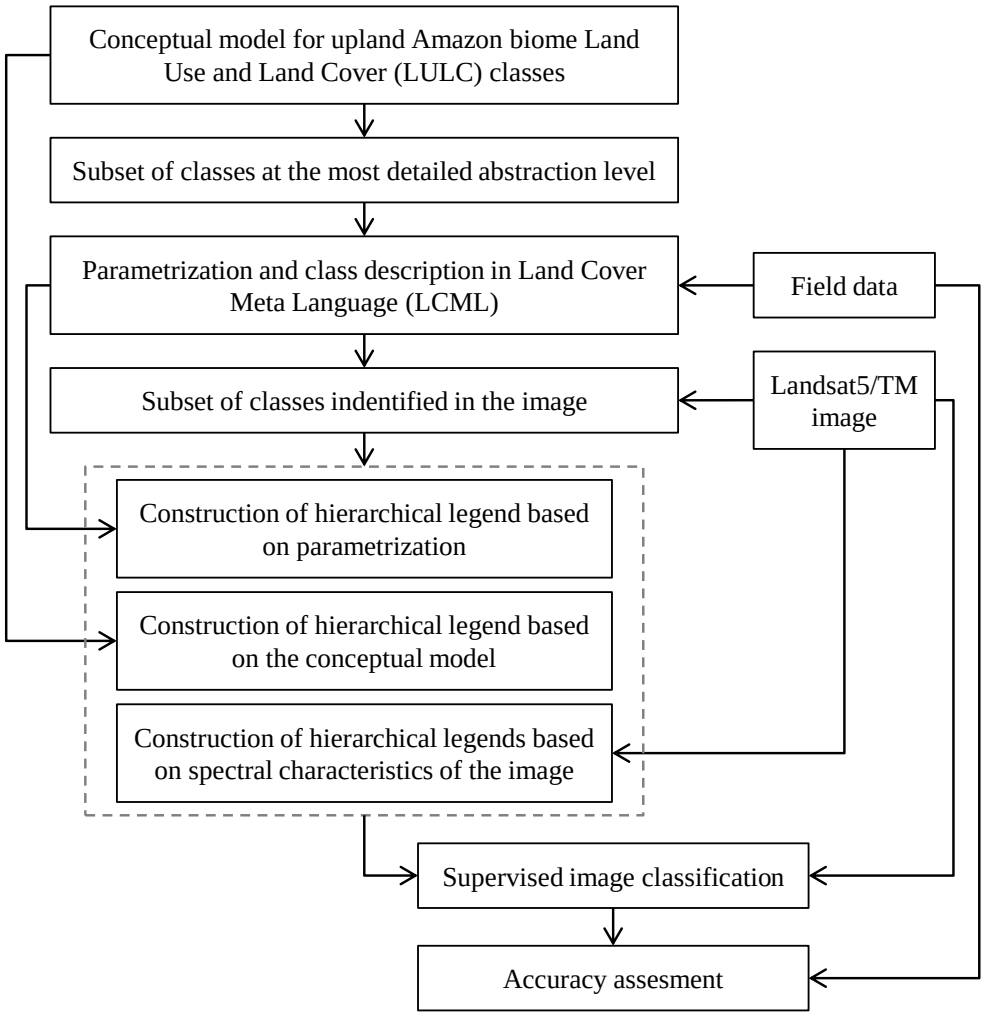
The data consist of geo-located photographs and land cover descriptions, collected in September 2009 and 2010, August 2013, and March 2015. Data from 2013 and 2015 also contain estimations of the proportion and the height (when applicable) of land cover elements. These estimations were made in the field and mainly in a visual way, based on the observation of at least two analysts, and considering the majority of the features identified in the field and on high-resolution remote sensing images, such as those captured by the *Satellite Pour l'Observation de la Terre* (SPOT) 5, SPOT 6 and RapidEye sensors. For 2013 field data, trees were randomly measured using a TruPulse 200B Rangefinder instrument. Major land cover classes in the area were previously identified using TerraClass data for the years 2008 and 2010, PRODES data for the years 2012 and 2014, and also high-resolution images covering the test area. When possible, information about the area's history was acquired from local inhabitants.

The remote sensing data used in this study corresponds to a Landsat 5/TM image downloaded from the INPE catalog (`http://www.dgi.inpe.br/CDSR/`). This image was acquired over the test area on June 29, 2010, path/row 227/62, with 8 bits, 7 spectral channels (0.45–0.52 μm; 0.52–0.60 μm; 0.63–0.69 μm; 0.76–0.90 μm; 1.55–1.75 μm; 10.4–12.5 μm; and 2.08–2.35 μm) and 30 m of spatial resolution (120 m on the thermal band). This image was registered to an orthorectified Landsat 5/TM image from the GeoCover project, and then radiometrically corrected using the method described in Green et al. (2002).

4.2 Methods

The proposed methodology is illustrated in Figure 4.2. In the first step, we define a conceptual LULC class model for the upland Amazon Biome. This is a model that holds class names, relationships, and qualitative definitions. This model was organized hierarchically, meaning that the classes of an upper level can be detailed in subclasses until a given abstraction level. From the classes at the most detailed abstract level, the ones identified in the Lower Tapajós region were formalized in LCML using the FAO Land Cover Classification System 3 software, version 1.8.0 (DI GREGORIO et al., 2015). From this subset of classes, we selected the ones present in the Landsat 5/TM image, which were grouped based on different criteria to form four hierarchical legends. Labeled samples for the classes in these legends were collected and used to classify the Landsat 5/TM image. Classification results were then compared. These steps are detailed in the following sections.

Figure 4.2 - Main steps of the proposed methodology.



SOURCE: The author.

4.2.1 Proposal of a conceptual class model for upland Amazon Biome

The LULC classes expected to be found in the upland Amazon Biome were organized in a conceptual model. This model was proposed based on the analysis of well-known classification systems and legends used in Brazilian Amazon studies, as well as previous field and an ad hoc LULC analysis experience. At this stage, nomenclature and definitions are similar to those used by the LCCS version 2 dichotomous phase (DI GREGORIO; JANSEN, 2005) and the first legend from the TerraClass project (ALMEIDA et al., 2016).

4.2.2 Parametrization of LULC classes and formalization to LCML

Following the proposal of LCML, the LULC classes identified in the Lower Tapajós region were characterized by the physiognomy and structure of elements of the cover for each class. LCML classifiers were selected to be descriptive of the classes, while also being relatively easy to understand and identify in the field. The LULC classes were described by the minimum and maximum thresholds of the mean height and the cover proportion of the following five elements of land cover structure:

a) **Soil:** exposed soil. This feature height equals zero. Therefore, only the proportion in relation to other features in the same *stratum* is considered;

b) **Litter:** organic debris. It occurs in the same *stratum* as 'Soil' and a height equal to zero was considered;

c) **Herbaceous vegetation:** plants that have no persistent woody stem above ground. Height may vary and it can be in the same *stratum* as 'Soil' and 'Litter' or not, depending on height and structure. The height usually varies up to 2 m;

d) **Shrubs:** plants that have persistent woody stem above ground and height smaller than 5 m;

e) **Trees:** plants with an elongated woody stem and higher than 5 m. In the presence of emergent trees, two *strata* may be composed of trees.

These elements can be distributed in different strata, so it is possible, and even probable, that the sum of the cover proportion of all elements surpasses 100%, as established in Di Gregorio and Jansen (2005). The definition of thresholds for each classifier/element was based on field data, complemented by the ad hoc knowledge of researchers experienced in field data collection.

4.2.3 Feature selection and definition of hierarchical legends

We were able to define 16 LULC classes in the study area in the most detailed level of abstraction (Section 4.3.2), from which ten were identified in the Landsat 5/TM image (Section 4.3.3). These ten classes were herein organized in the legend level referred to as L1(10). We collected labeled samples from the Landsat 5/TM image for each class, based on field data from 2010. Note that, although the field data from 2010 does not provide information regarding the mean height or cover proportion

of the land cover elements, the classes correspond to the LCML formalized ones, allowing for the present analysis. The labeled samples were then divided into the training and test data set. The training data set was used 1) to select the bands of the Landsat image to be used; 2) to calculate the data-based legends, and 3) to train the supervised algorithm used to classify the selected bands. The test samples were used to evaluate the classification results.

The feature selection process was executed over the bands 1–5 and 7 of Landsat 5/TM image, based on the minimum Jeffries–Matusita (JM) distance between all possible pairings within the ten classes set. JM distance is given by (SCHOWENGERDT, 2006):

$$JM_{ij} = \sqrt{2(1 - e^{-B_{ij}})}, \tag{4.1}$$

in which B_{ij} is the Bhattacharyya distance between the classes i and j. For Gaussian distributions, it is given by:

$$B_{ij} = \frac{1}{8}(m_i - m_j)^{\intercal} \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (m_i - m_j) + \frac{1}{2} ln \frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| \, |\Sigma_j|}}, \tag{4.2}$$

$\intercal$ is the transverse operation, m_k is the mean value of samples from class k and Σ_k is the covariance matrix of these samples.

All possible combinations from 1 to 6 bands were tested. We selected the set of features with three bands that presented the highest minimum JM distance, as the difference to the best-ranked set with four features was small (0.44 to 0.47). Bands 3, 4, and 5 were selected for analysis.

The data-based hierarchical legends were calculated by the approaches here defined as the 'Single Minimum Link Dendrogram' (SMLD) and the 'Double Minimum Link Dendrogram' (DMLD). The SMLD is based on the Single Link Hierarchical Clustering (SLHC) algorithm, used in Negri (2009) for similar purposes. Consider an image to be classified into n classes and a set of labeled pixels for each class. From these samples, it is possible to determine a representative probability density function for each class and to calculate a dissimilarity measure between pairs of these functions. In the first iteration of SMLD, the pair of classes that presents the lowest dissimilarity value is aggregated, generating a set of $n - 1$ classes. For the second iteration, dissimilarity values are recalculated considering this new set of classes. Again, the pair with the minimum value is aggregated. This process is iteratively

made until some established stop criterion is achieved. Here, we aggregated classes until only two remained. Results are then organized as a dendrogram, in which different sets of classes are determined by a given threshold, herein the minimum dissimilarity value between the pairs of classes in that set. The DMLD works similarly to SMLD. The main difference between the two is that DMLD aggregates two pairs of classes with the lowest dissimilarity values instead of one pair of classes at each iteration. We used the JM distance as the dissimilarity measure both in SMLD and DMLD.

Classes in L1(10) were also organized in two other hierarchical legends, herein the theoretical legends. In the first one, classes are merged considering the hierarchy proposed in the conceptual class model. In the second one, classes were merged considering the LCML parametrization, by grouping classes with proximate values of the land cover class elements.

4.2.4 Image classification

To illustrate the potential and shortcomings of the defined classes/legends for land cover classification studies, we classified the Landsat 5/TM image described in Section 4.1 using supervised algorithms and a Monte Carlo approach, based on Reis et al. (2017b). Bands 3, 4, and 5 of the Landsat 5/TM image were classified using all the proposed legends. Firstly, training samples were merged according to the legend and then used to train three supervised algorithms implemented in the R software: the Tree Decision-based algorithm J48 within RWeka package (HORNIK et al., 2009; WITTEN; FRANK, 2005), k-Nearest Neighbor (k-NN) (as implemented in RWeka package), and Maximum Likelihood (ML), based on the Gaussian distribution and implemented in Rasclass package (RASCLASS..., 2016). For classification using the J48 algorithm, we adopted the default pruning confidence threshold (0.25), seed (1), and number of folds (3). We varied the minimum number of instances from 2 to 10, in steps of 1. For k-NN, we varied the number of neighbors from 1 to 20, in steps of 1.

Considering each legend and algorithm separately, 100 pixels of each class were randomly selected from the training set and used to train the classifier and obtain one classification. This process was repeated 100 times, creating 100 classifications for each legend and algorithm. Each pixel in the final classification was assigned the most frequent (mode) label in the 100 classifications. This process was done to supposedly obtain image classifications more independent from the sample collection and to avoid problems of spatial correlation.

We observed that similar results were obtained using all classifiers, although the best configuration for J48 or k-NN varied depending on the legend level. For simplicity, we selected the ML results to be presented in this chapter. This classifier was selected because it presented relatively good classification results without the need for parameter tuning.

4.2.5 Accuracy assessment and comparison of results

For the accuracy assessment of the final classifications, 50 pixels for each class from the test set were randomly selected. With these samples, a confusion matrix was constructed, and the Overall Accuracy and Kappa Index (CONGALTON, 1991; CONGALTON; GREEN, 2008) were calculated, as well as the Producer's and User's Accuracy for each class. This process was repeated 1,000 times, resulting in 1,000 confusion matrices for each classification and 1,000 values for each index. We analyzed the average confusion matrix (each cell in this matrix is the average value of the 1000 values of this cell in the matrices replication), as well as the average value of the resulting indexes. In the last analysis, the indexes' average values were compared using an unpaired t-test with 1% of significance level [2].

4.3 Results

The proposed conceptual class model for the Brazilian Amazon Biome is presented in Section 4.3.1. The parametrization and LCML formalization of the classes identified in the Lower Tapajós region are described in Section 4.3.2. The derived hierarchical legends are presented in Section 4.3.3, followed by the image classification results, illustrated in Section 4.3.4.

4.3.1 The proposed conceptual class model for the Brazilian Amazon Biome

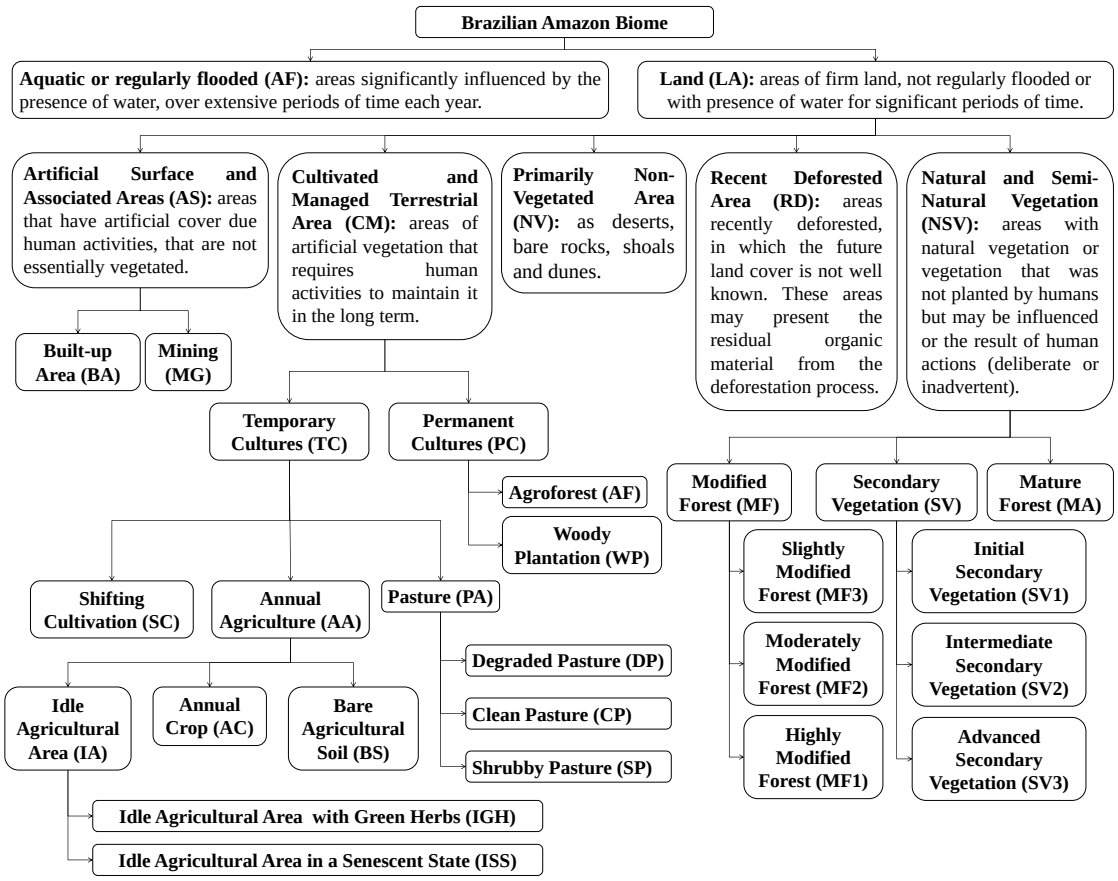
The conceptual class model is illustrated in Figure 4.3, with some of the major LULC classes described. More detailed classes and the criteria for subdivision are explained as follows.

As we choose to focus on the upland classes, only on the classes on the *Land* category were further developed into sub-classes. The *Land* category is further divided into five classes: *Artificial Surface and Associated Areas* (AS), *Cultivated and Managed*

[2]Note that we opted for unpaired tests in this analysis because we did not use the same seed for the random selection of test samples in different images. Future studies can benefit from redesigning this feature to allow for more robust paired tests.

Terrestrial Area (CM), *Primarily Non-Vegetated Area (NV)*, *Recent Deforested Area (RD)*, and *Natural and Semi-Natural Vegetation* (NSV). Although these classes are defined in Figure 4.3, some considerations must be made about the *Natural and Semi-Natural Vegetation* class.

Figure 4.3 - Proposed conceptual class model for the Brazilian Amazon biome.



SOURCE: The author.

Generally, it is very difficult to determine a single and well-accepted definition of 'forest'. Although it is commonly accepted that a closed forest may be a class in which more than 70% of the area is covered by three crowns (DI GREGORIO; JANSEN, 2005; PUTZ; REDFORD, 2010; COUTINHO et al., 2013), it can be very difficult to define exactly what a pristine forest is in the Amazon biome. As we cannot guarantee that past modifications were not made, we adopted the term *Mature Forest* to describe well-structured, climax forests, with small to

no evidence of alteration. If a forest is modified to the point of losing its original characteristics without suffering deforestation event, it may be classified as *Modified Forest*. This class covers both forests being degraded by factors as logging, fire, invasion by exotic species or other natural or anthropic activities, as well as forests being managed by silviculture practices (except forest plantations) employed to favor economically valuable species.

Concerning originally covered areas of forest, there is also any kind of semi-natural vegetation (not planted by humans, but resultant of human activities) that has grown in areas in which the original vegetation was completely removed. These areas were classified as *Secondary Vegetation*. These classes can be further divided regarding the intensity of modification or the stage of development. Although these subclasses occur in continuum gradients, it is feasible to state one class for the beginning of the process, one for the final stage, and an intermediary one. For *Secondary Vegetation*, these stages correspond to *Initial Secondary Vegetation*, *Intermediate Secondary Vegetation*, and *Advanced Secondary Vegetation*. Similar categorization was used in studies like Lu et al. (2003), Vieira et al. (2003), and Salomão et al. (2012). In an analog way, *Modified Forests* also may be divided into *Slightly Modified Forest*, *Moderately Modified Forest*, and *Highly Modified Forest*.

The *Cultivated and Managed Terrestrial Area* class was divided into different subclasses considering the differences in management and structure. Firstly, they were divided into *Permanent Culture* and *Temporary Culture*. The first one encompasses classes that are dominated by shrubs and trees and do not need regular replanting. The second class refers to areas composed mainly of herbaceous cultures which present some temporal sequencing depending behavior within one year. *Permanent Culture* can be further divided into *Woody Plantation* and *Agroforest*. Agricultural areas of permanent, arboreal, cultures were included in *Woody Plantation*, as well as reforested areas not destined for agriculture. *Agroforest* represents areas in which husbandry and cropping are practiced among natural, semi-natural, or domesticated trees.

The *Temporary Culture* class was subdivided into *Annual Agriculture*, *Pasture*, and *Shifting Cultivation*. *Annual Agriculture* denotes areas used to cultivate annual crops. This class may be divided into *Bare Agricultural Soil*, *Annual Crop*, and *Idle Agricultural Area*. As the names suggest, *Bare Agricultural Soil* represents areas of bare soil, recently harvested or prepared for planting. *Annual Crop* denotes areas with cultivated cultures at the time of analysis, in any stage of development.

Idle Agricultural Area represents areas used to describe fallow annual agriculture areas, not being used for cultivation at the time of analysis. This last class may be split into other two classes, with respect to the state of the vegetation covering the area at the time being analyzed: *Idle Agricultural Area in a Senescent State*, in which haulm or invasive plants are expected to be present and, in their majority, senescent; and *Idle Agricultural Areas with Green Herbs*, in which invasive plants must be, in their majority, still green. Though *Annual Agriculture* classes represent mainly mechanized, extensive areas, some small low-productivity plantations are also expected. These areas differ from *Shifting Cultivation* mainly regarding land management. In *Shifting Cultivation*, cultivated fields are typically managed for two or three years before being left to fallow, and fallow lands are lands unused for three to fifteen years. It is commonly practiced in small areas and has a mosaic and dynamic footprint consisting of cultivated fields and fallow lands.

The *Pasture* class represents areas mainly covered by grass. Mainly because of management practices, the structure of this class may vary over time. This class was divided into: *Clean Pasture*, *Shrubby Pasture*, and *Degraded Pasture*. *Clean Pasture* denotes pasture areas, in which shrubby invasive plants can be found only in small quantities. *Shrubby Pasture* describes areas with a significant presence of shrubby invasive plants. *Degraded Pasture* denotes pasture areas with reduced productivity and/or with pasture grass with decreased vigor. If needed for further characterization, these classes can be also classified by the presence or not of original remaining trees or cultivated palms, as illustrated in Figure 4.4.

Figure 4.4 - Examples of pasture classes.

(a) CP with Palms (b) SP without Palms (c) SP with Palms

In which CP = Clean Pasture and SP = Shrubby Pasture.

SOURCE: The author.

Regarding the class *Artificial Surfaces and Associated Areas*, the main two subclasses found in Amazon were *Built-up Area* and *Mining*. *Built-up Area* is defined as an area that was constructed by humans. For field characterization or further detailed studies, it may be of interest to classify this area in *Impervious Surface* (roofs, paved streets, and/or roads) and *Non-Impervious Surface* (gardens, parks, non-paved roads, among others). It may also be interesting to typify these elements regarding the type of material, and size, among other characteristics. *Mining* represents areas used for any type of mineral extraction.

It is possible to find areas in which the LULC class is being changed at the time of analysis, like an area of secondary vegetation being cleaned for pasture purposes, for instance. Since this is an area that does not belong to a specific LULC class, they were not defined in the conceptual model. In this sense, studies should be free to add additional classes such as *Transition area* or *Areas without information*, even if those cannot be properly defined using the chosen LCML classifiers.

4.3.2 Parametrization of LULC classes and LCML translation

From the classes defined in the previous subsection, it was possible to identify 13 land cover classes in the area with field data: Bare Agricultural Soil (BS), Idle Agricultural Area in a Senescent State (ISS), Idle Agricultural Area with Green Herbs (IGH), Annual Crop (AC), Shifting Cultivation (SC), Clean Pasture (CP), Shrubby Pasture (SP), Woody Plantation (WP), Initial Secondary Vegetation (SV1), Intermediate Secondary Vegetation (SV2), Advanced Secondary Vegetation (SV3), Slightly Modified Forest (MF3), and Mature Forest (MA). From these and the previous knowledge of the region, it was also possible to extrapolate the characteristics of other three classes: Highly Modified Forest (MF1), Moderately Modified Forest (MF2), and Recent Deforested Area (RD). The proportion and the mean height thresholds of land cover elements are defined in Table 4.1. In this table, the mean height of 'Soil' and 'Litter' is not specified because it was previously defined as zero. The mean height of 'Herbaceous' was not decisive to define any of the classes, so it was also suppressed from Table 4.1.

Table 4.1 - Thresholds of land cover elements: cover proportion (%) and mean height (m).

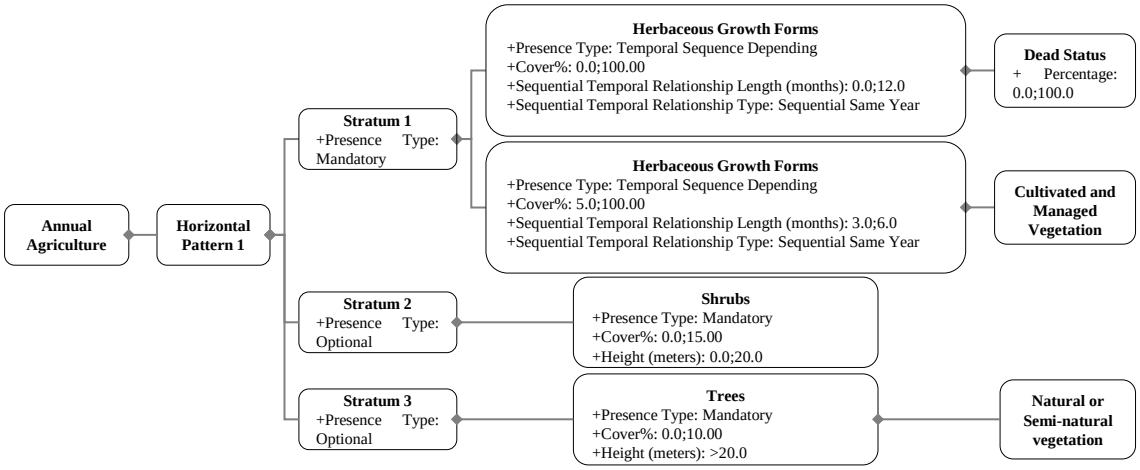
Land Cover Classes	Proportion										Mean Height			
	Herbaceous		Shrubs		Trees		Litter		Soil		Shrubs		Trees	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Bare Agricultural Soil	0	10	0	5	0	10 [a]	0	20	80	100	*	*	*	*
Idle Agricultural Area in a Senescent State	0	10	0	20	0	10 [a]	20	100	0	80	*	*	*	*
Idle Agricultural Area with Green Herbs	20	100	0	20	0	10 [a]	0	20	0	80	*	*	*	*
Annual Crop	5	100	0	5	0	10 [a]	0	20	0	95	*	*	*	*
Recent Deforested Area	0	15	0	15	30 [a,b]	100 [a,b]	0	70	0	70	*	*	*	*
Shifting Cultivation	30	80	0	40	0	0	0	40	0	40	*	*	*	*
Clean Pasture	60	100	0	15	0	10 [a]	*	*	0	40	*	*	*	*
Shrubby Pasture	20	85	15	40	0	10 [a]	*	*	*	*	*	*	*	*
Woody Plantation	*	*	*	*	60	100	*	*	*	*	*	*	*	*
Initial Secondary Vegetation	*	*	40	100	0	10 [a]	*	*	0	40	0.5	5	*	* [a]
Intermediate Secondary Vegetation	*	*	0	40	60	100 [a]	*	*	*	*	2	5	5	15 [a]
Advanced Secondary Vegetation	*	*	0	40	60	100 [a]	*	*	*	*	2	5	15	20 [a]
Highly Modified Forest	*	*	*	*	10	40	*	*	*	*	*	*	20	**
Moderately Modified Forest	*	*	*	*	40	80	*	*	*	*	*	*	20	**
Slightly Modified Forest	*	*	*	*	80	90	*	*	*	*	*	*	20	**
Mature Forest	*	*	*	*	90	100	*	*	*	*	*	*	20	**

In which * denotes that the proportion of an element or height does not interfere with the definition of the respective class and ** that the element may reach its maximum height, but this height may vary. [a] 10% of remaining natural trees, higher than 20 m, allowed. [b] Cut trees or burnt standing ones.

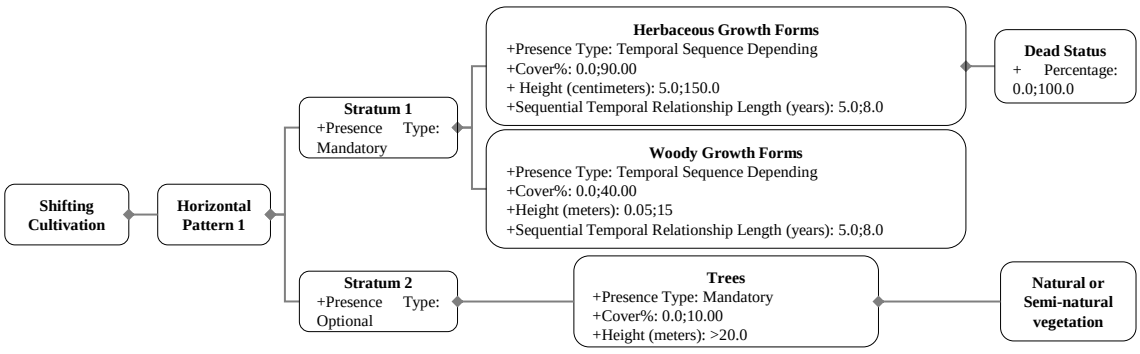
SOURCE: The author.

The thresholds presented in Table 4.1 are intended to help to identify classes in the field in a reproducible way, and not as an absolute descriptor. Furthermore, there are some classes in which these thresholds are superimposed, as happens with *Shifting Cultivation* and *Shrubby Pasture*, *Woody Plantation* and some secondary vegetation/modified forest classes, and also between *Idle Agricultural Area with Green Herbs* and *Clean Pasture*. Although these classes may not be fully distinguishable considering the defined thresholds, they present different spatial patterns, compositions, and/or temporal behaviors, which can be fully described in LCML. Some examples of LCML formalized classes are shown in Figures 4.5 and 4.6.

Figure 4.5 - Examples of LCML modeling of classes with temporal sequencing behavior.



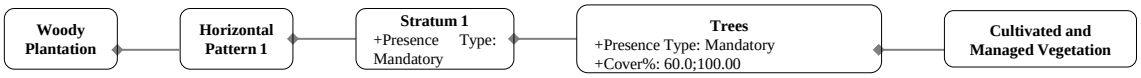
(a) Annual Agriculture



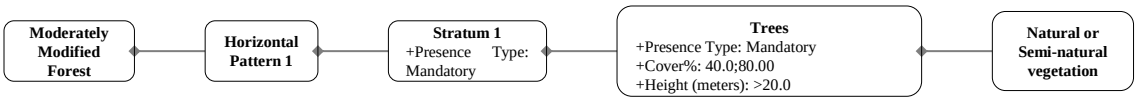
(b) Shifting Cultivation

SOURCE: The author.

Figure 4.6 - Examples of LCML modeling of classes without temporal sequencing behavior.



(a) Woody Plantation



(b) Moderately Modified Forest

SOURCE: The author.

For instance, consider the classes *Idle Agricultural Area with Green Herbs* and *Clean Pasture*, exemplified in Figure 4.7. As can be seen, differentiating these classes may be a confusing task, even in the field. However, besides the difference in use, these classes also present different temporal behaviors. While one pasture area may present *Clean Pasture* and/or *Shrubby Pasture* throughout the year, one annual agricultural area would probably present the classes *Annual Crop*, *Bare Agricultural Soil*, and *Idle Agricultural Area* in succession. Therefore, this time sequencing component of *Annual Agriculture* may be useful to distinguish this class. Spectral variation in images from temporal series is already used in the TerraClass project to classify the corresponding 'Annual Agriculture' class. This difference in temporal behavior for agricultural classes was formalized in LCML, as illustrated in Figure 4.5a. Note that all agricultural classes (BS, IGH, ISS, and AC) were modeled as a single class, *Annual Agriculture*, in LCML.

Figure 4.7 - Examples of areas of Idle Agricultural Area with Green Herbs and Clean Pasture.

(a) Idle Agricultural Area with Green Herbs (b) Clean Pasture

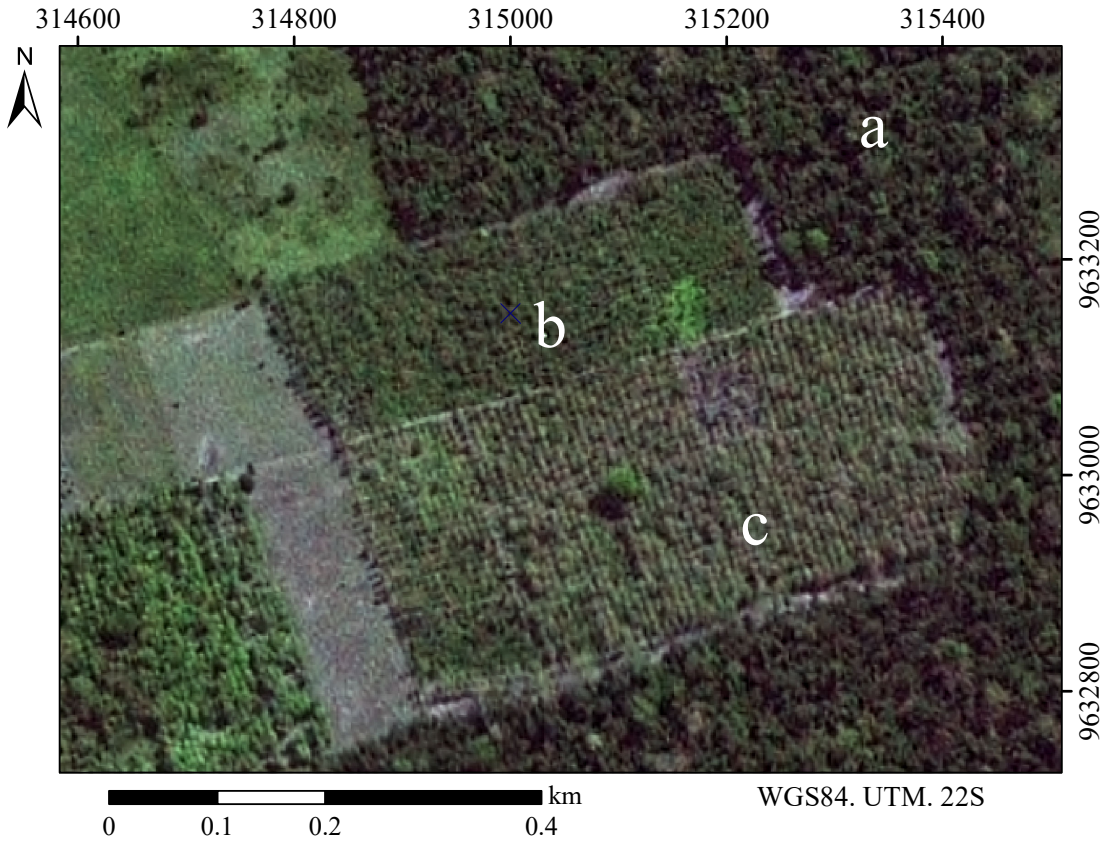
SOURCE: The author.

The classes *Shifting Cultivation* and *Shrubby Pasture*, although also presenting differences in temporal behavior, can be distinguished given their very different patterns. *Shifting Cultivation* areas present no infrastructure for cattle and the species cultivated are generally for consumption and somewhat diversified, while in *Shrubby Pasture* the shrubs are invasive plants. In LCML modeling, this can be differentiable by the inclusion of a 'grazing' classifier for *Gramineae* and 'natural or semi-natural vegetation' for shrubs, both in *Shrubby Pasture*. In remote sensing images, they may be distinguished by the format and size of features, since pastures are usually larger than areas of *Shifting Cultivation*.

The thresholds defined for the *Woody Plantation* class also present some overlapping with the ones in classes *Intermediate Secondary Vegetation*, *Advanced Secondary Vegetation*, *Moderately Modified Forest*, *Slightly Modified Forest*, and *Mature Forest*. However, distribution patterns and, in some cases, the number of species, may be easy to distinguish, even to an inexperienced observer. Since the trees were human-planted in *Woody Plantation*, they may be regularly distributed and present a homogeneous texture or be of specific, easily recognizable, commercial species. In secondary vegetation, modified or mature forests, a regular pattern in the distribution or single species are not expected. To illustrate this point, Figure 4.8 presents a subset of a high-resolution optical remote sensing image from an area in Brasil Novo, also in Pará state. This is a SPOT-6 image, acquired on August 19, 2015, in which the spectral bands (6 m of spatial resolution) were fused with the panchromatic band (1.5 m of spatial resolution). In this figure, it is possible to

verify the aforementioned difference in pattern from human-planted trees (indexes b and c in the figure) to trees not planted by humans in a uniform pattern (index a). This difference may also be noticeable in medium-resolution images because of differences in texture.

Figure 4.8 - A subset of a SPOT-6 image from an area in Brasil Novo, Pará state, Brazil, R(3)G(2)B(1) color composition.



(a) area of mature forest; (b) area of planted cocoa; (c) area of mixed woody plantation.

SOURCE: The author.

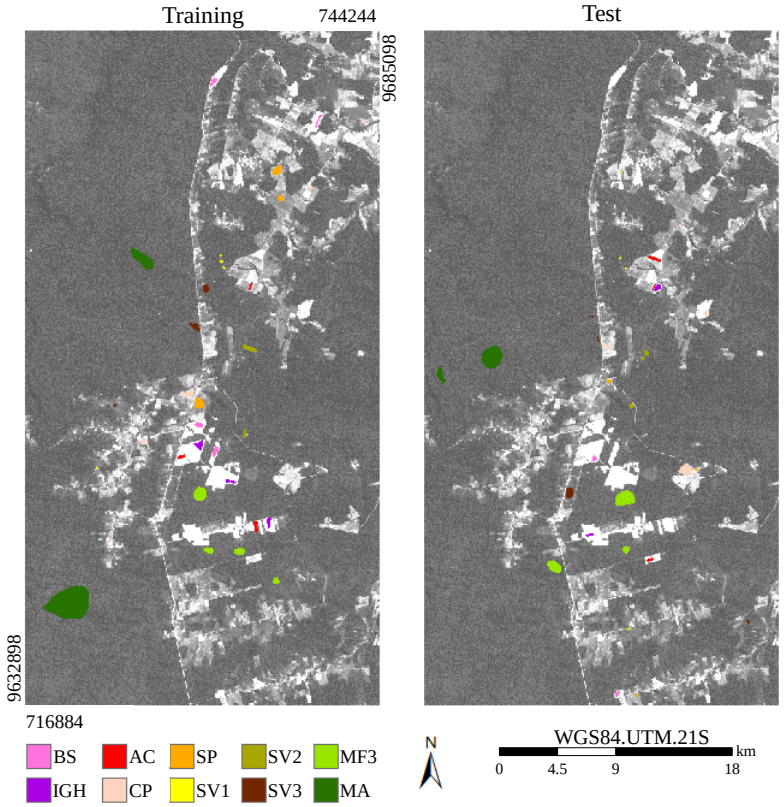
Distribution pattern also seems to be useful to distinguish different types of alteration that lead to modified forests. Some examples of different patterns of modified forests and how these alter the appearance of the classes in remote sensing images are illustrated in studies like Asner et al. (2004) and Pinheiro et al. (2016).

Although the estimation of the cover percentage and average height of vegetation may not be exact in the field, the defined thresholds help to explain the classes to the analyst, helping to set the point at which a class that occurs in a gradient becomes another class. In our experience, they were useful to help to separate *Clean Pasture* from *Shrubby Pasture*, or to better argue the point in which *Shrubby Pasture* turns into an area of *Initial Secondary Vegetation*, as well as distinguishing stages of secondary vegetation or modified forests.

4.3.3 Defined legends

In the 2010 Landsat 5/TM image, ten LULC classes were identified: *Bare Agricultural Soil*, *Idle Agricultural Area with Green Herbs*, *Annual Crop*, *Clean Pasture*, *Shrubby Pasture*, *Initial Secondary Vegetation*, *Intermediate Secondary Vegetation*, *Advanced Secondary Vegetation*, *Slightly Modified Forest*, and *Mature Forest*. Samples of these classes can be seen in Figure 4.9, with band 5 of the Landsat 5/TM image for reference, as well as the number of samples collected for each class (number of pixels considering the image spatial resolution).

Figure 4.9 - Training and test sample sets, superposed on band 5 of the Landsat 5/TM image for reference.



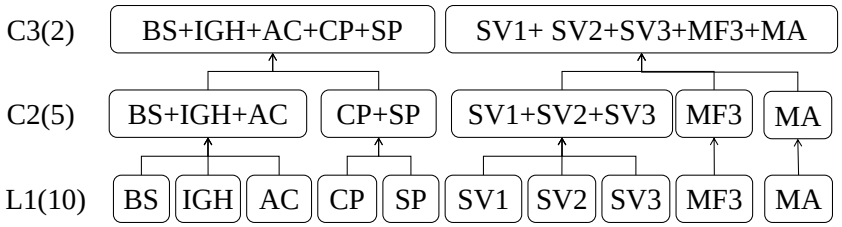
# of samples in each set		
Class	Training	Test
BS	940	256
IGH	694	344
AC	459	407
CP	831	991
SP	1175	224
SV1	158	167
SV2	516	332
SV3	599	609
MF3	2032	2727
MA	8880	2898

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, and MA = Mature Forest.

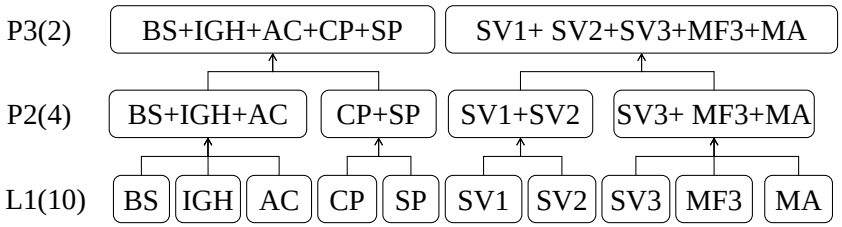
SOURCE: The author.

The ten LULC classes were then organized into three types of legends: based on the conceptual model, based on the parametrization, and based on automatic clustering algorithms. The first two legends are illustrated in Figure 4.10. In this figure, each legend is named as *Legend* $l(n)$, in which *Legend* denotes the type of the legend, either 'C', 'P' or 'L', respectively for the legends based on the conceptual model, based on parametrization, and the base set of classes, l denotes the tier level, and n represents the number of classes in the legend.

Figure 4.10 - Theoretical legends.



(a) Based on the conceptual class model



(b) Based on the LCML parametrization

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, and MA = Mature Forest.

SOURCE: The author.

These legends may be useful for:

a) **L1(10):** the ten identified LULC classes. This is the most detailed legend and may be useful for data field collection and mappings with diverse objectives;

b) **C2(5):** first grouping of the ten LULC classes, following the relationships established by the conceptual class model. This legend has the potential to yield better classification results than L1(10) while maintaining significant classes for decision-makers. It is especially interesting for forest regeneration studies since the secondary vegetation classes are nor merged with other types of classes;

c) **P2(4):** grouping of the ten LULC classes based on similarity in the parametrization. Good classification results in remote sensing data are expected with this legend (REIS et al., 2017b);

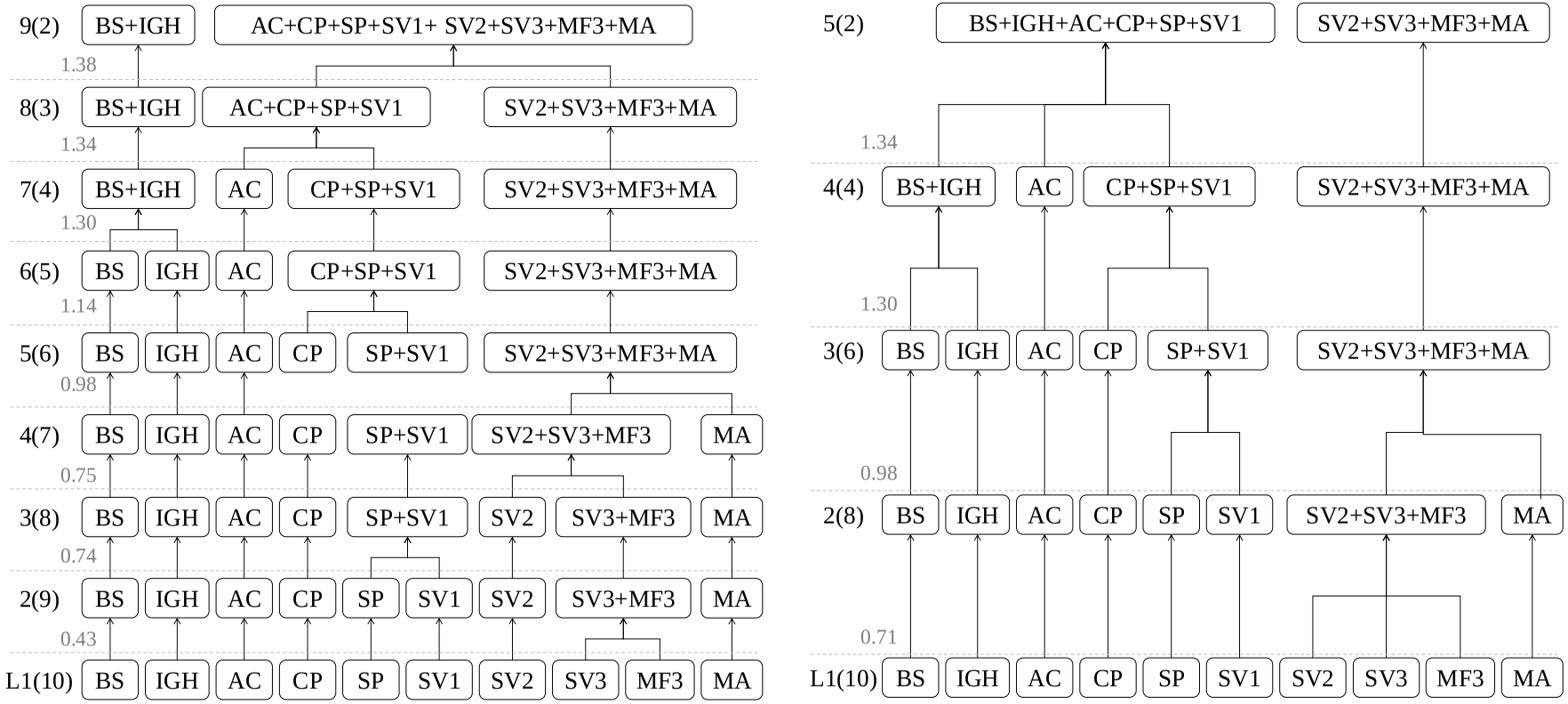
d) **C3(2)/P3(2):** this legend possesses only two classes, which can be

achieved either by following the conceptual class model or the classifiers in LCML and for which the most accurate classification results are expected. This legend may be particularly interesting for deforestation-related studies.

The legends based on the automatic clustering algorithms, the data-based legends, are depicted in Figure 4.11. In these legends, tiers are presented as *Legend_l*(n), in which *Legend* can be either 'SMLD' or 'DMLD'. In Figure 4.11, the minimum dissimilarity value between the pairs of classes in each tier is registered in SMLD, and the second minimum in DMLD. Note that all legends present the same set of ten classes (L1(10)) in the first tier and that SMLD_5(6) and DMLD_3(6) are the same, and so are SMLD_7(4) and DMLD_4(4).

It is possible to notice that, with the obvious exception of L1(10), automatic clustering algorithms did not generate any legend tier equal to one within the theoretical legends. The first classes to be merged on data-based legends are *Advanced Secondary Vegetation* (SV3) and *Slightly Modified Forest* (MF3), with the classes *Shrubby Pasture* (SP) and *Initial Secondary Vegetation* (SV1) usually merged in the sequence. These results indicate the difficulty in separating these classes based only on the spectral characteristics of the Landsat 5/TM image. On the other hand, the classes *Bare Agricultural Soil* (BS), *Idle Agricultural Area with Green Herbs* (IGH), and *Annual Crop* (AC) were merged in the final tiers of data-based legends, due to very different spectral characteristics in the selected features of Landsat 5/TM data. These results mean that, for some classifiers such as Maximum Likelihood, it is important to keep the agricultural subclasses as different classes, whereas the distinction of secondary vegetation and forests may depend on multi-temporal data sets. Furthermore, *Initial Secondary Vegetation* is commonly characterized as a transition class from *Shrubby Pasture* to more developed secondary vegetation classes in LULC studies. This may be exemplified by the existence of a class 'Secondary Vegetation with Pasture', in the first TerraClass legend. Given the spectral similarity of these classes in the selected features of the Landsat 5/TM data, some studies considering LULC trajectory analysis may find it convenient to merge these classes, or at least include suppositions of confusion between them during the interpretation of results.

Figure 4.11 - Legends based on spectral separability of data.



(a) SMLD

(b) DMLD

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, and MA = Mature Forest.

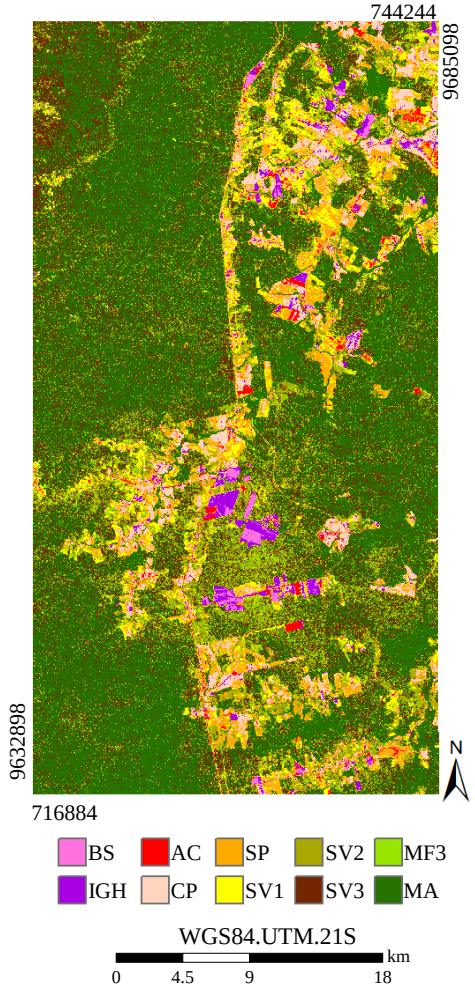
SOURCE: The author.

4.3.4 Image classification

As previously explained, the samples illustrated in Figure 4.9 were grouped in each legend and used to classify bands 3, 4, and 5 of the Landsat 5/TM image using a Monte Carlo approach and ML classifier. We will first focus on the classification results obtained by the legends that are independent of the remote sensing image, i.e., those based on the conceptual class model and parametrization. The classified images, as well as the confusion matrix (in %) and accuracy indexes, are presented in Figures 4.12 to 4.15.

These classifications were obtained with a simple classification algorithm, only one image for only one date, and a pixel-wise approach. Therefore, it was expected that classes formed by the same elements of land cover in similar proportions would present some confusion in the classification results. This is clear in the analysis of results from L1(10) legend, given the high amount of samples of *Idle Agricultural Area with Green Herbs* classified as *Clean Pasture*, or the confusion between *Slightly Modified Forest* and *Intermediate Secondary Vegetation/Advanced Secondary Vegetation*. Classes without superposing thresholds but with successional behavior, such as *Shrubby Pasture/Initial Secondary Vegetation* or *Intermediate Secondary Vegetation/Advanced Secondary Vegetation*, also presented high confusion. Despite the high confusion between some classes, the use of L1 legend resulted in the good classification of classes as *Bare Agricultural Soil*, *Annual Crop*, and, to a lesser degree, *Initial Secondary Vegetation* and *Mature Forest*.

Figure 4.12 - Classification results obtained using the Maximum Likelihood classifier and the L1(10) legend.



Average Confusion Matrix (%)

Classification \ Reference	1	2	3	4	5	6	7	8	9	10
1.BS	**98**	0	0	0	0	0	0	0	0	0
2.IGH	2	**42**	0	6	9	0	0	0	0	0
3.AC	0	5	**100**	7	1	3	0	0	0	0
4.CP	0	53	0	**77**	18	0	0	0	0	0
5.SP	0	0	0	10	**43**	8	2	0	0	0
6.SV1	0	0	0	0	29	**87**	2	1	3	0
7.SV2	0	0	0	0	0	1	**44**	52	29	1
8.SV3	0	0	0	0	0	0	21	**22**	23	19
9.MF3	0	0	0	0	0	1	30	21	**34**	4
10.MA	0	0	0	0	0	0	1	4	11	**76**

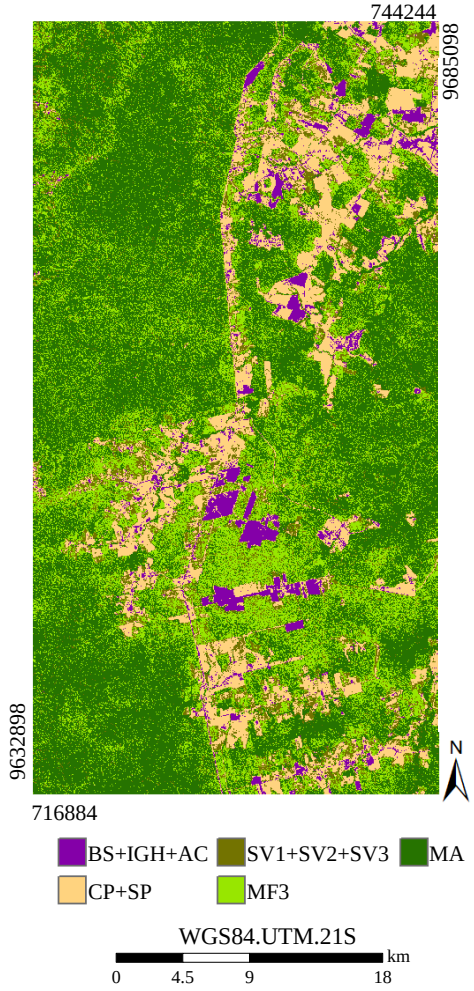
Accuracy Indexes

Class	P.A.	U.A.
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.42 ± 0.07	0.71 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.03
CP	0.78 ± 0.06	0.52 ± 0.07
SP	0.43 ± 0.06	0.68 ± 0.04
SV1	0.87 ± 0.04	0.72 ± 0.04
SV2	0.44 ± 0.06	0.35 ± 0.04
SV3	0.22 ± 0.06	0.26 ± 0.06
MF3	0.34 ± 0.07	0.38 ± 0.06
MA	0.76 ± 0.06	0.83 ± 0.05
Overall Accuracy	0.62 ± 0.02	
Kappa	0.58 ± 0.02	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, P.A. = Producer's Accuracy and U.A. = User's Accuracy. In the average confusion matrix the main diagonal in presented in bold font. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

Figure 4.13 - Classification results obtained using the Maximum Likelihood classifier and the C2(5) legend.



Average Confusion Matrix (%)

		Reference				
		1	2	3	4	5
Classification	1.BS+IGH+AC	**98**	16	1	0	0
	2.CP+SP	2	**84**	11	0	0
	3.SV1+SV2+SV3	0	0	**27**	19	2
	4.MF3	0	0	56	**66**	13
	5.MA	0	0	5	15	**85**

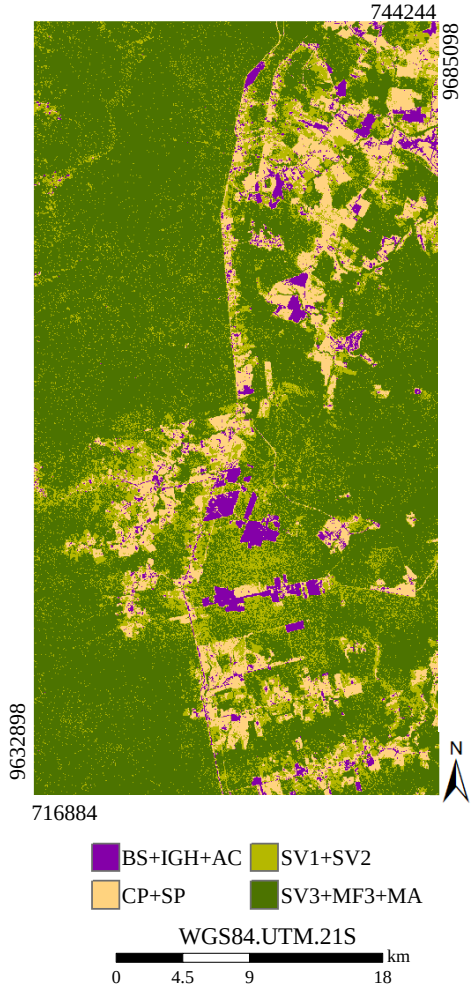
Accuracy Indexes

Class	P.A.	U.A.
BS+IGH+AC	0.99 ± 0.02	0.86 ± 0.04
CP+SP	0.84 ± 0.05	0.87 ± 0.04
SV1+SV2+SV3	0.27 ± 0.06	0.56 ± 0.09
MF3	0.67 ± 0.07	0.49 ± 0.04
MA	0.85 ± 0.05	0.81 ± 0.05
Overall Accuracy	0.72 ± 0.02	
Kappa	0.65 ± 0.03	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, P.A. = Producer's Accuracy and U.A. = User's Accuracy. In the average confusion matrix the main diagonal in presented in bold font. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

Figure 4.14 - Classification results obtained using the Maximum Likelihood classifier and the P2(4) legend.



Average Confusion Matrix (%)

		Reference			
		1	2	3	4
Classification	1.BS+IGH+AC	**98**	16	1	0
	2.CP+SP	2	**83**	11	0
	3.SV1+SV2	0	1	**69**	33
	4.SV3+MF3+MF	0	0	19	**67**

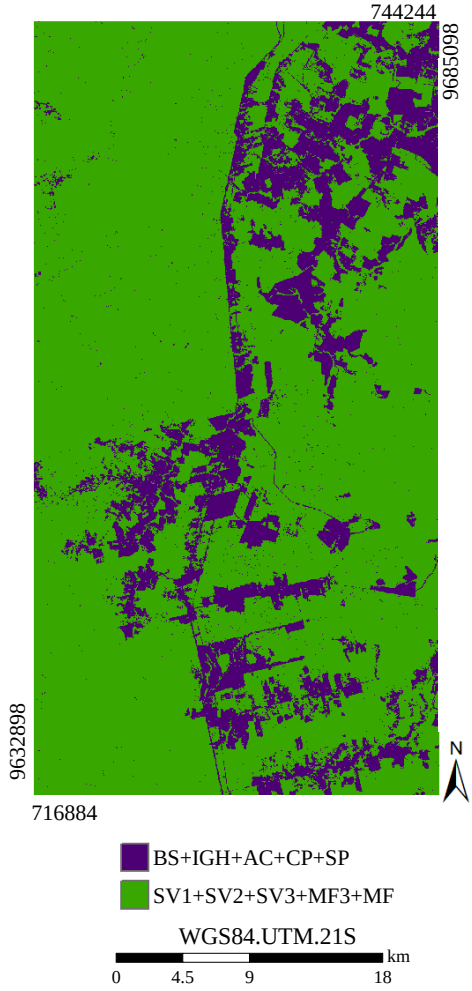
Accuracy Indexes

Class	P.A.	U.A.
BS+IGH+AC	0.98 ± 0.02	0.86 ± 0.04
CP+SP	0.83 ± 0.05	0.87 ± 0.04
SV1+SV2	0.69 ± 0.06	0.67 ± 0.05
SV3+MF3+MF	0.67 ± 0.06	0.78 ± 0.05
Overall Accuracy	0.79 ± 0.03	
Kappa	0.73 ± 0.03	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, P.A. = Producer's Accuracy and U.A. = User's Accuracy. In the average confusion matrix the main diagonal is presented in bold font. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

Figure 4.15 - Classification results obtained using the Maximum Likelihood classifier and the C3(2)/P3(2) legend.



Average Confusion Matrix (%)

		Reference	
		1	2
Classification	1.BS+IGH+AC+CP+SP	**100**	4
	2.SV1+SV2+SV3+MF3+MF	0	**96**

Accuracy Indexes

Class	P.A.	U.A.
BS+IGH+AC+CP+OP	1.00 ± 0.00	0.97 ± 0.02
SV1+SV2+SV3+MF3+MF	0.96 ± 0.03	1.00 ± 0.00
Overall Accuracy	0.98 ± 0.01	
Kappa	0.96 ± 0.03	

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, P.A. = Producer's Accuracy and U.A. = User's Accuracy. In the average confusion matrix the main diagonal in presented in bold font. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

The average global accuracy indices (Overall Accuracy, Kappa) of classifications using the data-based legends are depicted in Table 4.2, along with the standard deviation of values. The values for class-oriented indices (Producer's Accuracy and User's Accuracy) are illustrated in Table 4.3.

Table 4.2 - Global accuracy indices for the classifications obtained using the Maximum Likelihood classifier and the data-based legends.

Legend	**Overall Accuracy**	**Kappa**
SMLD_2(9)	0.70 ± 0.02	0.66 ± 0.02
SMLD_3(8)	0.73 ± 0.02	0.69 ± 0.02
DMLD_2(8)	0.77 ± 0.02	0.74 ± 0.02
SMLD_4(7)	0.82 ± 0.02	0.79 ± 0.02
SMLD_5(6)/DMLD_3(6)	0.83 ± 0.02	0.80 ± 0.02
SMLD_6(5)	0.86 ± 0.02	0.83 ± 0.02
SMLD_7(4)/DMLD_4(4)	0.89 ± 0.02	0.85 ± 0.03
SMLD_8(3)	0.87 ± 0.02	0.81 ± 0.04
SMLD_9(2)	0.92 ± 0.03	0.85 ± 0.05
DMLD_5 (2)	0.99 ± 0.01	0.97 ± 0.02

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, and MA = Mature Forest. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

Table 4.3 - Class-oriented accuracy indices for the classifications obtained using the Maximum Likelihood classifier and the data-based legends.

	P.A.	**U.A.**
SMLD_2(9)		
BS	0.97 ± 0.02	1.00 ± 0.00
IGH	0.42 ± 0.07	0.71 ± 0.07
AC	1.00 ± 0.00	0.86 ± 0.04
CP	0.78 ± 0.05	0.52 ± 0.03
SP	0.43 ± 0.06	0.70 ± 0.07
SV1	0.88 ± 0.04	0.73 ± 0.04
SV2	0.48 ± 0.07	0.55 ± 0.06
SV3+MF3	0.48 ± 0.07	0.43 ± 0.05
MA	0.84 ± 0.05	0.86 ± 0.04

(Continue)

Table 4.3 - Continuation.

	P.A.	**U.A.**
SMLD_3(8)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.78 ± 0.07
AC	1.00 ± 0.00	0.88 ± 0.04
CP	0.79 ± 0.06	0.56 ± 0.04
SP+SV1	0.81 ± 0.06	0.87 ± 0.04
SV2	0.47 ± 0.06	0.56 ± 0.06
SV3+MF3	0.51 ± 0.07	0.43 ± 0.05
MA	0.84± 0.05	0.86 ± 0.04
DMLD_2(8)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.71 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.04
CP	0.79 ± 0.06	0.53 ± 0.04
SP	0.43 ± 0.06	0.70 ± 0.07
SV1	0.87 ± 0.04	0.73 ± 0.04
SV2+SV3+MF3	0.84 ± 0.05	0.85 ± 0.04
MA	0.87 ± 0.05	0.88 ± 0.04
SMLD_4(7)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.74 ± 0.07
AC	1.00 ± 0.00	0.87 ± 0.04
CP	0.78 ± 0.06	0.56 ± 0.04
SP+SV1	0.81 ± 0.05	0.89 ± 0.04
SV2+SV3+MF3	0.86 ± 0.05	0.86 ± 0.04
MA	0.86 ± 0.05	0.88 ± 0.04
SMLD_5(6)/DMLD_3(6)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.43 ± 0.06	0.76 ± 0.07
AC	1.00 ± 0.00	0.88 ± 0.044
CP	0.79 ± 0.05	0.56 ± 0.03
SP+SV1	0.82 ± 0.05	0.88 ± 0.04
SV2+SV3+MF3+MA	0.98 ± 0.02	1.00 ± 0.00

(Continue)

Table 4.3 - Continuation.

	P.A.	**U.A.**
SMLD_6(5)		
BS	0.98 ± 0.02	1.00 ± 0.00
IGH	0.46 ± 0.07	0.87 ± 0.06
AC	1.00 ± 0.00	0.90 ± 0.03
CP+SP+SV1	0.89 ± 0.05	0.63 ± 0.03
SV2+SV3+MF3+MA	0.97 ± 0.02	1.00 ± 0.00
SMLD_7(4)/DMLD_4(4)		
BS+IGH	0.71 ± 0.06	0.90 ± 0.04
AC	1.00 ± 0.00	0.93 ± 0.03
CP+SP+SV1	0.87 ± 0.05	0.75 ± 0.04
SV2+SV3+MF3+MA	0.97 ± 0.02	1.00 ± 0.00
SMLD_8(3)		
BS+IGH	0.68 ± 0.06	0.95 ± 0.04
AC+CP+SP+SV1	0.95 ± 0.03	0.74 ± 0.04
SV2+SV3+MF3+MA	0.98 ± 0.02	0.99 ± 0.01
SMLD_9(2)		
BS+IGH	0.88 ± 0.04	0.96 ± 0.03
AC+CP+SP+SV1+SV2+SV3+MF3+MA	0.97 ± 0.03	0.89 ± 0.03
DMLD_5 (2)		
BS+IGH+AC+CP+SP+SV1	0.99 ± 0.01	0.98 ± 0.02
SV2+SV3+MF3+MA	0.98 ± 0.02	0.99 ± 0.01

In which BS = Bare Agricultural Soil, IGH = Idle Agricultural Area with Green Herbs, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF3 = Slightly Modified Forest, MA = Mature Forest, P.A. = Producer's Accuracy, and U.A. = User's Accuracy. Accuracy indexes are presented as $m \pm sd$ in which m is the average value of the index and sd is the standard deviation.

SOURCE: The author.

Higher accuracy values were obtained using the data-based legends. This result was expected for the used Landsat 5/TM image but it may vary if another type of data is used. Although the usefulness of legends based on remote sensing data depends on the study being carried on, it is interesting to note that the legend SMLD_2(9) has led to higher accuracy values than C2(5), with higher detail. Given that some LULC studies do not distinguish between modified and secondary forests,

the merging of the classes *Slightly Modified Forest* (MF3) and *Advanced Secondary Vegetation* (SV3) can be inconsequential to the objective but necessary to achieve better accuracy values for single-time classifications, which may be corrected using multi-temporal data such PRODES, if available.

Considering regional studies, SMLD_7(4)/ DMLD_4(4) seems to present a good compromise between accuracy and detail, with agricultural, pasture, and developed forest classes separated. The main problem of these legends is the union of *Initial Secondary Vegetation* (SV1) with *Clean Pasture* (CP) and *Shrubby Pasture* (SP), as previously discussed. Considering the legends with the lower level of detail (2 classes), we have very good accuracy values for both L4(2) and DMLD_5(2), but lower values for SMLD_9(2). Although DMLD_5(2) has shown higher accuracy than C3(2)/P3(2), this difference is small. There is the same possible problem of grouping *Initial Secondary Vegetation* to the agricultural classes in DMLD_5(2).

In general, classes that presented the lowest values of mean Producer's and User's Accuracy using one legend have been merged into another class to form the next tier of the legends SMLD and DMLD. Although the high Producer's Accuracy and User's Accuracy values for the classes BS+IGH+AC of both P2(5) and C2(4), the class *Annual Crop* (AC) has been well classified even using ten classes (L1(10)). This good classification was maintained until this class was merged with another in either SMLD or DMLD. The same happened to the *Bare Agricultural Soil* (BS) class. These results suggest that the use of automatic clustering methods may be useful not only to form legends but also to determine which subclasses could remain separated from the class of interest in a study to preserve a higher level of detail for other possible uses of the generated classification.

4.4 Discussion

In this chapter, we proposed a set of hierarchically organized LULC classes for the Brazilian Amazon biome. We also defined 16 detailed LULC classes in LCML, based on field data and experience. As suggested by Di Gregorio and Jansen (2005), these classes may be easily understandable and recognized. Furthermore, very simple classifiers (average height and proportion of cover of five elements of land cover classes) were used. However, since many of these classifiers presented overlapping thresholds between very different classes, it is clear that they can not be configured as absolute descriptors of each class. For instance, analysts may need different class definitions in their studies, or their definition may be based on criteria like above-ground biomass, Net Primary Productivity, forest

stand volume, basal area, average stand height, diameter at breast height, age, and/or many others, as used by studies such Lu et al. (2003), Vieira et al. (2003), Salomão et al. (2012), and Romero-Sanchez and Ponce-Hernandez (2017) to justify the subdivision of secondary vegetation classes. Nonetheless, many of these classifiers are implemented in LCML and the users are free to further describe the classes or even to include new subclasses, without compromising the use of the proposed conceptual model and formalized classes. Moreover, relatively simple methods based on multi-temporal data can differentiate between classes with overlapping thresholds, as is the case of *Idle Agricultural Area with Green Herbs* and *Clean Pasture* (ALMEIDA et al., 2016). There are also methods based on multi-temporal remote sensing images focused on detecting and measuring forest disturbances (TRITSCH et al., 2016; JARRON et al., 2017).

Despite this lacking characteristic of the proposed thresholds, it must be highlighted that they are very easily understood by non-specialist analysts, and may be particularly useful for studies that lay on Community-Based Observations (PRATIHAST et al., 2014). Additionally, there is potential to measure them directly from very-high-resolution images (cover proportion) and Light Detection and Ranging (LIDAR) data (average height), which would greatly diminish the subjectivity of studies. Most importantly, the use of the proposed thresholds is useful to delimit classes that occur successively in a given area, which tend to be the most difficult ones to describe. Note that the difference among *Clean Pasture*, *Shrubby Pasture*, *Initial Secondary Vegetation*, *Intermediate Secondary Vegetation*, and *Advanced Secondary Vegetation* is clearly defined, as well as the difference between *Mature Forest* and the *Modified Forest* subclasses. We understand that one can argue that the difference between 15% and 16% of shrubs to qualify an area as *Shrubby Pasture* is arbitrary and it is practically impossible to be determined precisely in the field or even in very-high-resolution images. However, this is not an exclusive problem of the delimitation of LCML thresholds but the very core of the Fuzzy set theory (ZADEH, 1965). Although the necessity of separating the classes in Boolean sets for field data collection or remote sensing analysis can be argued (FISHER, 2010), it is still the most common approach because it is operational and simpler than fuzzy methodologies. Furthermore, to use the Boolean class definition, clear thresholds must be defined. In this sense, some misplacement between very similar classes is expected, since measures contain errors.

Another perceived advantage of the use of LCML formalized legends is the possibility

to compare results obtained with different class definitions, as presented by Reis et al. (2017c). It is important to keep in mind, however, that although LCML formalized legends are easily understandable, they may not be directly comparable and a harmonization process would still be necessary (JANSEN et al., 2008).

Regarding the Landsat 5/TM classification results, the comparison among the average Kappa values demonstrates that less detailed legends result in more accurate classifications. This behavior was expected and had been previously documented (NEGRI, 2009; PANTALEÃO et al., 2012; REIS et al., 2017b). The legends based on data normally lead to the most accurate maps, even though these are not always the most useful ones. However, it is important to highlight once more that the use of different data, be it the image (type, date, or pre-processing), the initial set of classes, or the collection of reference samples would lead to different legends when these are derived using automatic clustering algorithms. Therefore, it is fundamental to verify if the chosen classes are significant for a determined study objective, or if those are just the ones that yield better accuracy figures. Nonetheless, automatic clustering algorithms based on data may be useful to analyze the data being used. They are also able to point subclasses that could be preserved to generate classifications with higher detail levels that could still be generalized to a theoretical legend, not greatly compromising the accuracy.

4.5 Chapter conclusions

This chapter proposes a conceptual class model for the upland Amazon Biome, with the classes in the most abstract level of detail fully described using simple LCML classifiers. We also presented different ways to merge classes to create hierarchical LULC legends, which were tested for supervised classification of a commonly used remote sensing image. These results are useful to guide researchers in the definition of vegetated classes from the upland Brazilian Amazon biome. They are also useful to explain LULC classes to non-specialists with objectivity, mainly in classes happening successively in a given area. Moreover, the use of the defined classes enables reproducibility of data acquisition for multi-temporal studies and can be an important tool for the harmonization of different products and the translation of field data into information applicable to local, regional or global scale studies. We could also perceive that the collection of field data with estimates of mean height and cover proportion of soil, litter, herbaceous vegetation, shrubs, and trees improves class description, both for the use of the proposed class definition as well as to allow the researchers to define LULC classes based on their own thresholds.

5 WEIGHTING TRANSITIONS TO IMPROVE LULC TRAJECTORIES CLASSIFICATION: THE COMPOUND MAXIMUM *A POSTERIORI* ALGORITHM[1]

As discussed in Chapter 2, the existing approaches commonly used for classifying remote sensing time-series in LULC trajectories can present serious drawbacks, such as the possibility to derive invalid trajectories, masking particular types of trajectories, and/or the use of complex data inputs and several classification steps. To overcome these aspects, this chapter presents a novel algorithm named Compound Maximum *a Posteriori* (CMAP) classifier.

CMAP is a generative classifier that incorporates pixel values observed at all points in the time series and the joint distribution of the classes over time. In CMAP, LULC trajectories are considered to be a Markov chain, so the probabilities between transitions can be globally modeled by a transition probability matrix. This matrix can either be estimated from auxiliary data or defined by the analyst. The CMAP classifies the time series of remote sensing images directly into valid-only LULC trajectories with a single step and in one location at a time. These characteristics provide some specific advantages for forest regeneration studies, such as:

a) the possibility to include the information derived from multi-temporal data in a simple way to better separate *Secondary Vegetation* from *Forest* areas, even if the deforestation event is not registered in the data used;

b) given that all dates are classified simultaneously, classification errors do not propagate in the trajectories, as observed with post-classification rule-based methods to eliminate invalid trajectories;

c) CMAP uses a supervised classification approach that requires reference samples of land cover classes in each time as the input, instead of reference samples of transitions/trajectories;

d) CMAP also permits the use of multi-sensor data and different tiers of a hierarchical legend for each date and/or data type. The latter is an important characteristic to analyze areas and/or periods with low data availability;

e) CMAP can be used for sparse time series, as is the case of analysis based on Landsat data acquired before the 2000s.

[1]This chapter is an adapted version of Reis et al. (2020a).

The proposed algorithm is described in Section 5.1. CMAP was then tested in two case studies, in which we compared land cover trajectories obtained by CMAP to those obtained using the traditional Maximum Likelihood (ML) classifier in a post-classification comparison approach. In the first case study, we classified six images from the same sensor, using the same land cover legend. In the second case study, we classified three images from different types of sensors, using different land cover legend levels. The data and methodology employed for these case studies are presented in Section 5.2. The results are presented and discussed in Sections 5.3 and 5.4. Chapter conclusions are drawn in Section 5.5.

5.1 Compound Maximum a Posteriori algorithm

Consider a single geographic position in which one wants to estimate the trajectory $s = \{\omega_1^{k_1}, ..., \omega_t^{k_t}, ..., \omega_T^{k_T}\}$, T is the length of the time sequence and $\omega_t^{k_t} \in \Omega_t = \{\omega_t^{k_1}, ..., \omega_t^{k_t}, ...\omega_t^{K_t}\}$, in which $\omega_t^{k_t}$ is the actual class at time position t of s and k_t is the indicator of the class in the set Ω_t that holds the K_t possible classes on time t. Additionally, $s \in \mathbf{\Omega} = \Omega_1 \otimes ... \otimes \Omega_t \otimes ... \otimes \Omega_T$, in which $\otimes$ is the Cartesian Product of sets. A given observation vector $\vec{X} = \{\vec{x_1}, ..., \vec{x_t}, ..., \vec{x_T}\}$ contains the T temporal observations that can indicate the trajectory composition, like the digital numbers in correspondent pixels in a given set of images $\boldsymbol{X}$, and $\vec{X} \in \boldsymbol{X}$.

A generative method (NG; JORDAN, 2001) for trajectory classification can be formulated as:

$$\hat{s} = \arg_s \max(P(\vec{X}, s), s \in \mathbf{\Omega}, \vec{X} \in \boldsymbol{X}). \tag{5.1}$$

Based on the definition of the conditional probability:

$$\hat{s} = \arg_s \max(P(\vec{X}|s) \times P(s), s \in \mathbf{\Omega}, \vec{X} \in \boldsymbol{X}). \tag{5.2}$$

in which $P(s)$ is the *a priori* probability of a trajectory s. Supposing that the observations are independent in time and that each one depends only on the observed object, we have:

$$P(\vec{X}|s) = P(\vec{x_1}|\omega_1^{k_1}) \times ... \times P(\vec{x_t}|\omega_t^{k_t}) \times ... \times P(\vec{x_T}|\omega_T^{k_T}), \tag{5.3}$$

As an example, for the Gaussian distribution, $P(\vec{x_t}|\omega_t^{k_t})$ is calculated by

(THEODORIDIS; KOUTROUMBAS, 2009):

$$P(\vec{x_t}|\omega_t^{k_t}) = \int (2\pi)^{-B/2}|\Sigma_{k_t}|^{-1/2} \times \exp\{-1/2(\vec{x_t} - m_{k_t})^\intercal \Sigma_{k_t}^{-1}(\vec{x_t} - m_{k_t})\}d\vec{x_t} \quad (5.4)$$

in which B is the number of observed variables (like the channels of the image) of the time t, m_{k_t} and Σ_{k_t} are respectively the mean vector and the covariance matrix of the class $\omega_t^{k_t}$, $|\cdot|$ is the determinant and $\intercal$ is the transpose.

Observe that Equation 5.2 returns the same trajectory as stacking the Maximum Likelihood classifications at each point of time if the *a priori* probabilities of all trajectories are equal. The rule expressed in Equation 5.3, once the *a priori* term is removed, is called Compound Likelihood (CML). Inserting Equation 5.3 into Equation 5.2, we have what we refer to as the Compound Maximum *a Posteriori* (CMAP) estimation of a trajectory.

In Equation 5.2, $P(s)$ is the *a priori* probability of a trajectory s and can be given by:

$$P(s) = P(\omega_1^{k_1} \cap ... \cap \omega_t^{k_t} \cap ... \cap \omega_T^{k_T}). \quad (5.5)$$

Assuming the premise that the *a priori* probability of a class in a given time depends only on the class in the previous time,

$$P(s) = P(\omega_1^{k_1}) \times P(\omega_2^{k_2}|\omega_1^{k_1}) \times ... \times P(\omega_T^{k_T}|\omega_{T-1}^{k_{T-1}}). \quad (5.6)$$

In this case, Equation 5.6 is the basic assumption of Markov chains (THEODORIDIS; KOUTROUMBAS, 2009), adapted here for trajectory classification purposes. The conditional probabilities of Equation 5.6 can be represented by a matrix that tabulates the probability of each transition between legends used on two consecutive dates, or a *transition matrix* for simplicity, as illustrated in Figure 5.1. Given the premise of dependence on a previous time, this matrix will be referred to as the 'forward transition matrix' in this document. Following the concepts of Ng and Jordan (2001), this structure is considered discriminative. In the forward transition matrix, we consider that $P(\omega_t^{k_t}|\omega_{t-1}^{k_{t-1}})$ is known and

$$\sum_{k_t=1}^{K_t} P(\omega_t^{k_t}|\omega_{t-1}^{k_{t-1}}) = 1. \quad (5.7)$$

Equation 5.7 is based on a hard classification approach, in which the summation includes all possible classes and these are mutually exclusive. In the example presented in Figure 5.1, it means that the sum of probabilities in each line of the matrix equals 1.

Figure 5.1 - Discriminative structure of the forward transition matrices.

$t-1$ \ t	ω_t^1	$\dots$	$\omega_t^{k_t}$	$\dots$	$\omega_t^{K_t}$
ω_{t-1}^1	$P(\omega_t^1 \mid \omega_{t-1}^1)$	$\dots$	$P(\omega_t^{k_t} \mid \omega_{t-1}^1)$	$\dots$	$P(\omega_t^{K_t} \mid \omega_{t-1}^1)$
$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$
$\omega_{t-1}^{k_{t-1}}$	$P(\omega_t^1 \mid \omega_{t-1}^{k_{t-1}})$	$\dots$	$P(\omega_t^{k_t} \mid \omega_{t-1}^{k_{t-1}})$	$\dots$	$P(\omega_t^{K_t} \mid \omega_{t-1}^{k_{t-1}})$
$\vdots$	$\vdots$	$\ddots$	$\ddots$	$\ddots$	$\vdots$
$\omega_{t-1}^{K_{t-1}}$	$P(\omega_t^1 \mid \omega_{t-1}^{K_{t-1}})$	$\dots$	$P(\omega_t^{k_t} \mid \omega_{t-1}^{K_{t-1}})$	$\dots$	$P(\omega_t^{K_t} \mid \omega_{t-1}^{K_{t-1}})$

SOURCE: The author.

However, it is also possible to assume that the *a priori* probability of a class in a given time depends only on the class in the posterior time. This may be useful in cases in which the information for more recent datasets is more accurate. In this case

$$P(s) = P(\omega_T^{k_T}) \times P(\omega_{T-1}^{k_{T-1}} | \omega_T^{k_T}) \times ... \times P(\omega_1^1 | \omega_2^{k_2}), \tag{5.8}$$

$$\sum_{k_{t-1}=1}^{K_{t-1}} P(\omega_{t-1}^{k_{t-1}} | \omega_t^{k_t}) = 1, \tag{5.9}$$

and the transition matrix, referred to as the 'backward transition matrix', changes accordingly to depict the values of $P(\omega_{t-1}^{k_{t-1}} | \omega_t^{k_t})$.

Nonetheless, one may be interested in defining the values of $P(s)$ based on less restrictive assumptions than the discriminative transition matrices. Here we propose a simplified model, in which $P(s) = 0$ for invalid trajectories and $P(s) = 1/N_s$ for valid trajectories, in which N_s is the number of valid trajectories.

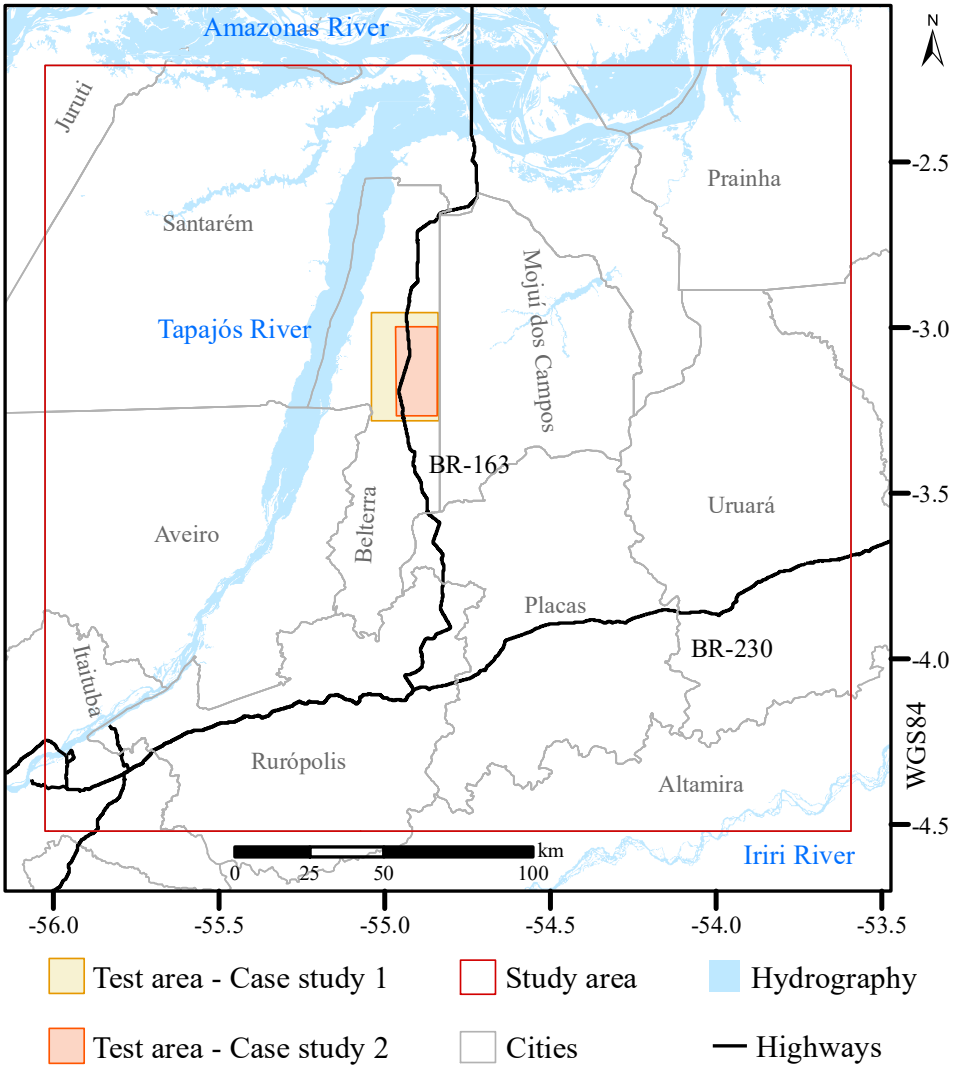
5.2 Case studies - materials and methods

In this section, we present the data and methodology for two applications of CMAP to generate land cover maps. Case Study 1 corresponds to what we considered a usual case of land cover trajectory classification: the classification of 6 images, one for each year between 2005 and 2010, from the commonly used sensor Thematic Mapper (TM) onboard Landsat5, with the same land cover class legend. Case Study 2 corresponds to the classification of images from different sensors and using different levels of a land cover class legend for each date/image. For this case study, we used one image from each of the following satellite/sensors: the Advanced Land Observing System (ALOS)/Phase Array L-Band Synthetic Aperture Radar (PALSAR), Landsat5/TM, and Earth Observer 1 (EO-1)/Advanced Land Imager (ALI), respectively from years 2008, 2010, and 2013. The test area, data, and methods used for each case study are described in the following sections.

5.2.1 Test area

The case studies were conducted in superposed test areas within the main study area. These areas are illustrated in Figure 5.2 along with political and natural limits.

Figure 5.2 - Test areas in relation to main study area.

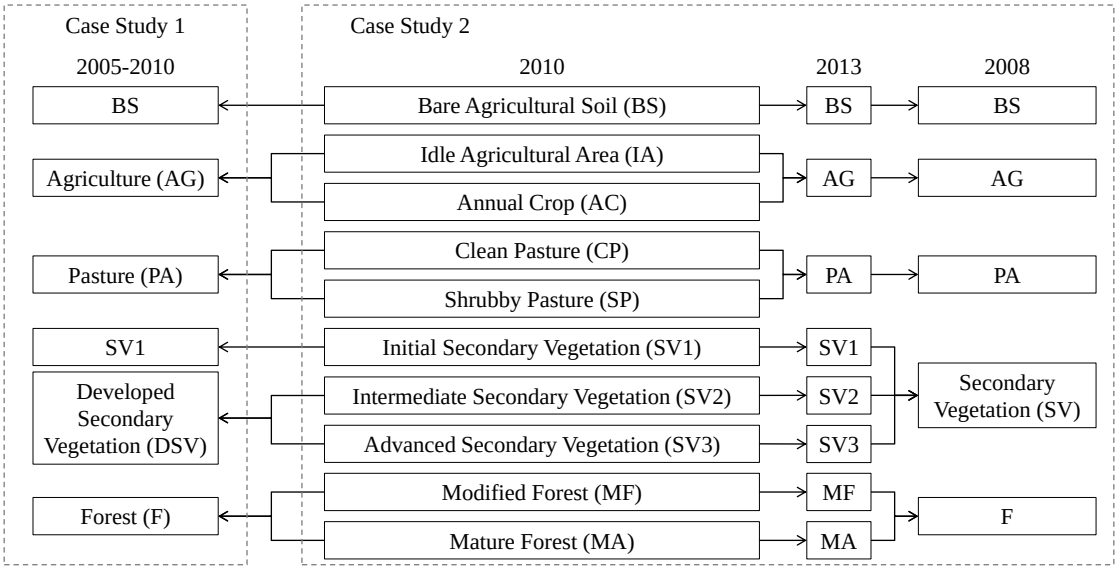


SOURCE: Based on INCRA (s.d.), IBGE (2019), MMA (2020), FUNAI (2021), and MInfra (2021).

5.2.2 Land cover legends and transition matrices

The ten most detailed land cover classes identified in the fields (Chapter 4) were organized in four different legend levels. The defined legends are illustrated in Figure 5.3. The legend with 6 classes was used to classify all dates from Case Study 1. The remaining legend levels, with 5, 10, and 8 classes, were used to classify, respectively, the images from 2008, 2010, and 2013 in Case Study 2.

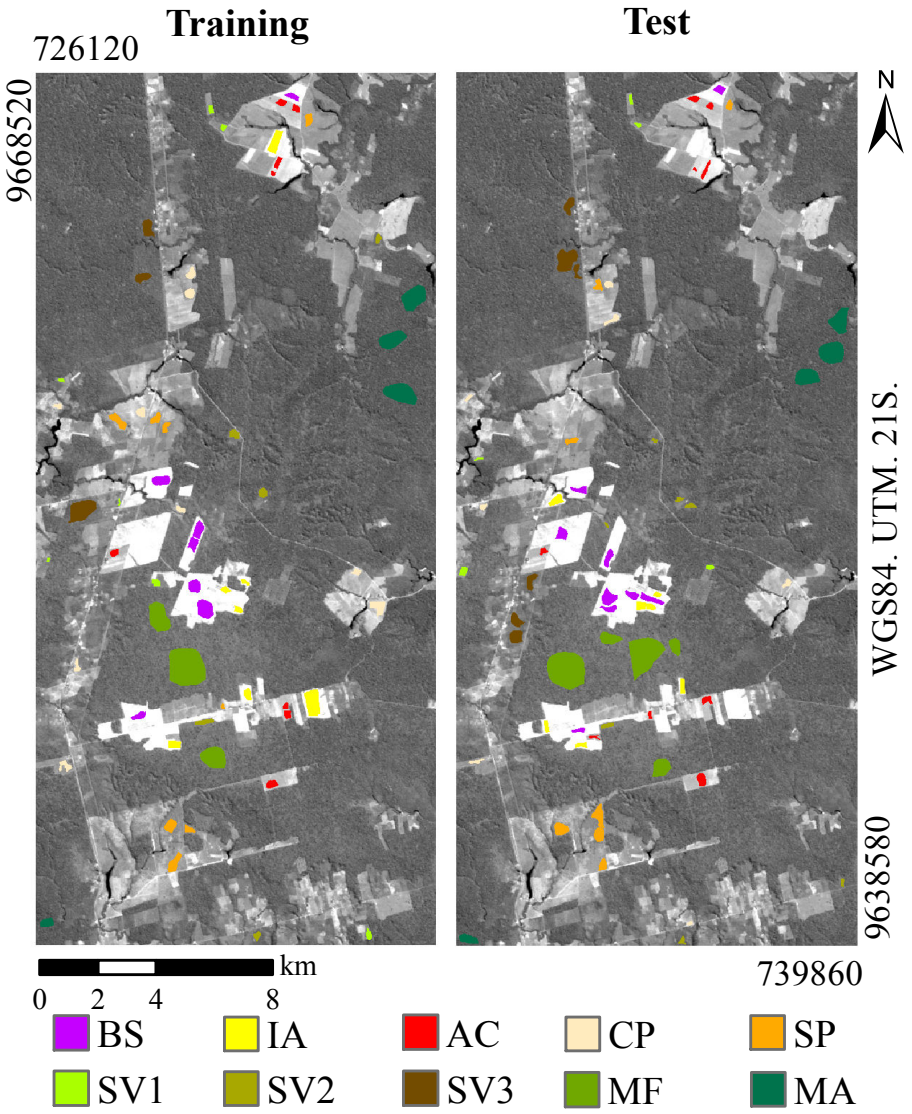
Figure 5.3 - Legends used in each Case Study.



SOURCE: The author.

Labeled samples of each class of a correspondent legend were manually collected over the images, considering the field information, maps from the TerraClass project, and visual interpretation of ancillary images. These samples were randomly divided into training and test sets. We collected at least 147 samples (pixels) for each class/year for the training set and at least 95 samples/class/year for the test sets. The class *Agriculture* was not found in the study area for years 2006 and 2009, so samples of the remaining 5 classes used in Case Study 1 were collected for these years. As an example, the labeled samples collected for 2010 using the most detailed legend level are presented in Figure 5.4. In this figure, the samples are illustrated over the band 5 of the Landsat5/TM image of the same year.

Figure 5.4 - Labeled samples collected for 2010, in the most detailed legend used in Case Study 2, over the band 5 of the Landsat5/TM image acquired on June 26, 2010.



In which BS = Bare Agricultural Soil, IA = Idle Agricultural Area, AC = Annual Crop, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF = Modified Forest, and MA = Mature Forest.

SOURCE: The author.

We used both the simplified and the discriminative approaches for calculating the *a priori* probabilities of trajectories in each case study. For the simplified approach, we considered as valid trajectories those composed only of valid transitions, as illustrated by Figures 5.5 and 5.6. For the discriminative approach, we considered

the backwards transition matrices depicted in Figures 5.7 and 5.8. In these matrices, the sum of each column equals to 1 (Equation 5.9). In Equation 5.8, $P(\omega_T^{k_T}) = 1/K_T$.

Figure 5.5 - Validity of transitions for Case Study 1.

t / $t-1$	BS	AG	PA	SV1	DSV	F
Bare Agricultural Soil (BS)	V	V	V	V	I	I
Agriculture (AG)	V	V	V	V	I	I
Pasture (PA)	V	V	V	V	I	I
Initial S.V. (SV1)	V	V	V	V	V	I
Developed S.V. (DSV)	V	V	V	V	V	I
Forest (F)	V	V	V	V	I	V

I=invalid transition; V=valid transition; S.V.=Secondary Vegetation. Transitions are considered as invalid by definition (any class different from *Forest* to *Forest*) or because of the time lapse between classifications (*Bare Agricultural Soil*/*Agriculture*/*Pasture* to *Developed Secondary Vegetation*.

SOURCE: The author.

Figure 5.6 - Validity of transitions for Case Study 2.

2008 \ 2010	BS	IA	CA	CP	SP	SV1	SV2	SV3	MF	MA
Bare Ag. Soil (BS)	V	V	V	V	V	V	I	I	I	I
Agriculture (AG)	V	V	V	V	V	V	I	I	I	I
Pasture (PA)	V	V	V	V	V	V	I	I	I	I
Secondary Veg. (SV)	V	V	V	V	V	V	V	V	I	I
Forest (F)	V	V	V	V	V	V	I	I	V	V

(a) Transitions from 2008 to 2010.

2010 \ 2013	BS	AG	PA	SV1	SV2	SV3	MF	MA
Bare Ag. Soil (BS)	V	V	V	V	I	I	I	I
Idle Ag. Area (IA)	V	V	V	V	I	I	I	I
Annual Crop (AC)	V	V	V	V	I	I	I	I
Clean Pasture (CP)	V	V	V	V	I	I	I	I
Shrubby Pasture (SP)	V	V	V	V	I	I	I	I
Initial S.V. (SV1)	V	V	V	V	V	I	I	I
Intermediate S.V. (SV2)	V	V	V	V	V	V	I	I
Advanced S.V. (SV3)	V	V	V	V	I	V	I	I
Modified Forest (MF)	V	V	V	V	I	I	V	I
Mature Forest (MA)	V	V	V	V	I	I	V	V

(b) Transitions from 2010 to 2013.

I=invalid transition; V=valid transition; S.V.=Secondary Vegetation; Ag.=Agricultural

SOURCE: The author.

Figure 5.7 - Backward discriminative transition matrix used for Case Study 1.

$t-1$ \ t	BS	AG	PA	SV1	DSV	F
Bare Ag. Soil (BS)	1/6	1/6	1/6	1/6	0	0
Agriculture (AG)	1/6	1/6	1/6	1/6	0	0
Pasture (PA)	1/6	1/6	1/6	1/6	0	0
Initial S.V. (SV1)	1/6	1/6	1/6	1/6	1/2	0
Developed S.V. (DSV)	1/6	1/6	1/6	1/6	1/2	0
Forest (F)	1/6	1/6	1/6	1/6	0	1

S.V.=Secondary Vegetation; Ag.=Agricultural.

SOURCE: The author.

Figure 5.8 - Backward discriminative transition matrices used for Case Study 2.

2008 \ 2010	BS	IA	AC	CP	SP	SV1	SV2	SV3	MF	MA
Bare Ag. Soil (BS)	1/5	1/5	1/5	1/5	1/5	1/5	0	0	0	0
Agriculture (AG)	1/5	1/5	1/5	1/5	1/5	1/5	0	0	0	0
Pasture (PA)	1/5	1/5	1/5	1/5	1/5	1/5	0	0	0	0
Secondary Veg. (SV)	1/5	1/5	1/5	1/5	1/5	1/5	1	1	0	0
Forest (F)	1/5	1/5	1/5	1/5	1/5	1/5	0	0	1	1

(a) Transitions from 2008 to 2010.

2010 \ 2013	BS	AG	PA	SV1	SV2	SV3	MF	MA
Bare Ag. Soil (BS)	1/10	1/10	1/10	1/10	0	0	0	0
Idle Ag. Area (IA)	1/10	1/10	1/10	1/10	0	0	0	0
Annual Crop (AC)	1/10	1/10	1/10	1/10	0	0	0	0
Clean Pasture (CP)	1/10	1/10	1/10	1/10	0	0	0	0
Shrubby Pasture (SP)	1/10	1/10	1/10	1/10	0	0	0	0
Initial S.V. (SV1)	1/10	1/10	1/10	1/10	1/2	0	0	0
Intermediate S.V. (SV2)	1/10	1/10	1/10	1/10	1/2	1/2	0	0
Advanced S.V. (SV3)	1/10	1/10	1/10	1/10	0	1/2	0	0
Modified Forest (MF)	1/10	1/10	1/10	1/10	0	0	1/2	0
Mature Forest (MA)	1/10	1/10	1/10	1/10	0	0	1/2	1

(b) Transitions from 2010 to 2013.

I=invalid transition; V=valid transition; S.V.=Secondary Vegetation; Ag.= Agricultural.

SOURCE: The author.

5.2.3 Remote sensing images

The used images are described in Table 5.1. The Landsat5/TM images were collected from the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center Archive at the L1 processing level and tier 1 terrain correction. These images present co-registration with Root Mean Square Error (RMSE) less than 12 m, considered a suitably geometric correction for time-series analysis. The image from ALOS/PALSAR was acquired from the Japan Aerospace Exploration Agency (JAXA) in Fine Beam Dual (FBD) mode, at 1.1 processing level, in single-look complex format, and in slant range. The image from EO-1/ALI was acquired from the USGS/EROS Center Archive at Level 1 Gst processing level.

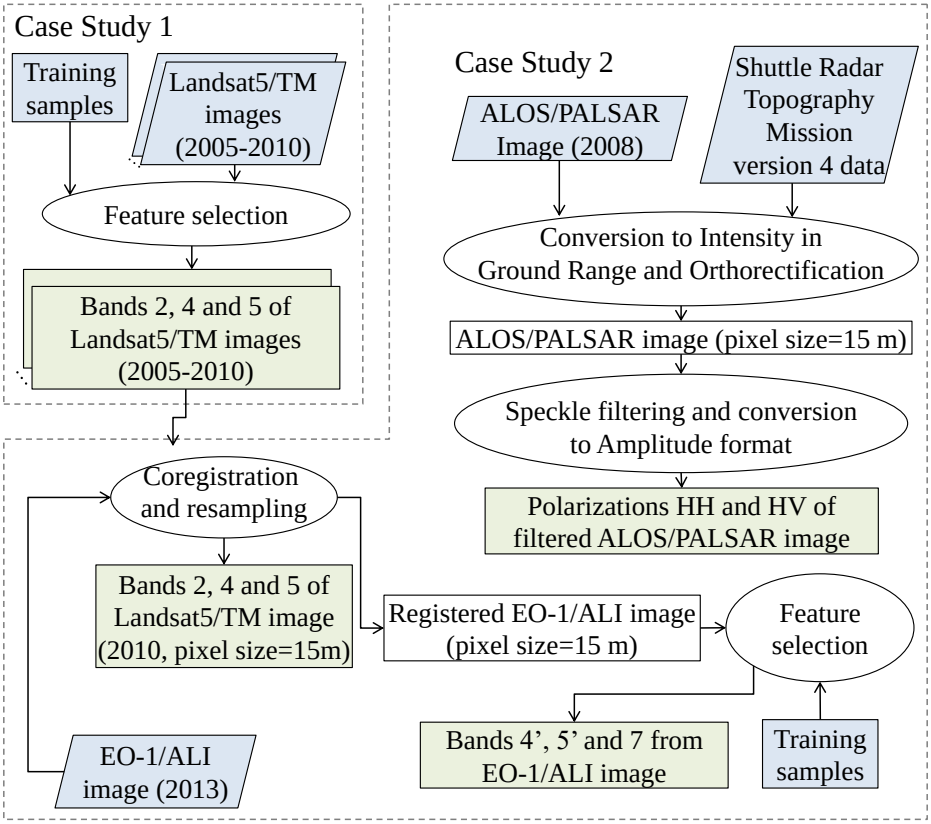
Table 5.1 - Images description.

Characteristics	Landsat5/TM	ALOS/PALSAR	EO-1/ALI
Channels	7 spectral bands: 1 (0.45–0.52 μm); 2 (0.52–0.60 μm); 3 (0.63–0.69 μm); 4 (0.76–0.90 μm);5 (1.55–1.75 μm); 6 (10.4–12.5 μm); 7 (2.08–2.35 μm)	2 polarizations in L band (23 cm): HH and HV	10 spectral bands: Pan (0.48–0.69 μm); 1' (0.433–0.453 μm); 1 (0.45–0.515 μm) 2 (0.525–0.605 μm); 3 (0.63–0.69 μm); 4 (0.775–0.805 μm); 4' (0.845–0.89 μm); 5' (1.2–1.3 μm); 5 (1.55–1.75 μm) 7 (2.08–2.35 μm)
Acquisition date (day/month/year)	01/07/2005; 21/08/2006 21/06/2007; 23/06/2008 12/07/2009; 29/06/2010	15/06/2008	05/10/2013
Spatial resolution	30 m (120 m in band 6)	10 m in range and 4.5 m in azimuth	30 m (10 m in band Pan)
Acquisition Mode	-	Fine Beam Dual	-
Processing level	Level 1	Level 1.1	Level 1Gst
Selected channels	2, 4 and 5	HH and HV (amplitude)	4', 5' and 7
Image size (pixels)	750 × 1200 (Case Study 1) 916 × 1996 (Case Study 2)	916 × 1996	916 × 1996

SOURCE: The author.

All Landsat5/TM images were used for Case Study 1, at the original 30 m pixel size. For Case Study 2, we used the Landsat5/TM image from June 28, 2010, along with the ALOS/PALSAR and the EO-1/ALI images, all with 15 m pixel size, re-sampled using the nearest neighbor approach to preserve the statistical properties of the images. All images were preprocessed in order to be co-registered and to have the same pixel size and geographic projection, as summarized in Figure 5.9 and explained as follows.

Figure 5.9 - Data preprocessing.



Inputs are highlighted in light blue. Final preprocessed images are highlighted in light green.

SOURCE: The author.

Since the Landsat5/TM images were downloaded suitably geometrically corrected, preprocessing of these images comprised a feature selection step for both case studies, done to remove redundant or noisy spectral bands from the analysis. For Case Study 2, there is the additional re-sampling step. Feature selection was based

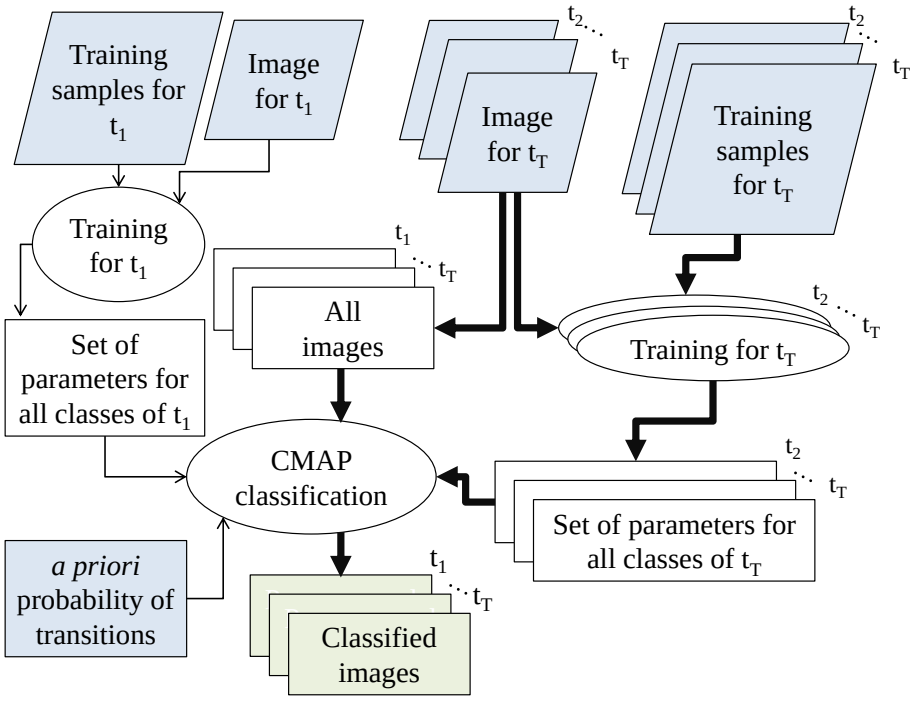
on the minimum Jeffries-Matusita (JM) distance (SCHOWENGERDT, 2006) between all pairs of classes of the 10 class legend used for Landsat5/TM classification in Case Study 2 (Figure 5.3). Bands 2, 4, and 5 were selected for classification. The ALOS/PALSAR image was first geocoded to UTM (zone 21S)/WGS84, converted to ground range, and transformed to intensity in ASF MapReady 3.0 software. It was then orthorectified using Shuttle Radar Topography Mission version 4 (SRTM4) data and the Rational Function Model (RFM) from the PCI Geomatics 13.0 software (with less than 7.5 m of RMSE), and re-sampled to square pixels of 15 m. The orthorectified ALOS/PALSAR image was then filtered by the Stochastic Distances Nonlocal Means (SDNLM) filter (TORRES et al., 2014) and converted to amplitude values, as detailed in Reis et al. (2020c). The parameters used in SDNLM are filtering window=5x5 pixels, patch=3x3 pixels, and confidence level=90%. This data set can be fitted to a Gaussian distribution, considering a Chi-square adherence test with a 5% of significance level (REIS et al., 2020c).

The EO-1/ALI image was geometrically corrected using first-order polynomial transformation and control points collected using the Landsat5/TM image from 2010 as the reference, with less than half a pixel (15 m) of RMSE. Bands 4', 5', and 7 from the EO-1/ALI image were selected for use based on the minimum JM distance between all pairs of classes from the 8 class legend used to classify this image. In Study Case 1, areas under clouds or cloud shadows at any given point were masked from analysis (2485 pixels, less than 0.3% of the image subset). The image subsets used for Study Case 2 were cloud-free.

5.2.4 Image classification and assessment

The land cover trajectory classification process is similar for the two case studies and is illustrated in Figure 5.10. In both cases, CMAP was calculated considering the training samples collected over each image and using Gaussian distribution. We also classified each image using the traditional ML pixel-wise classifier, to verify the impact of including the *a priori* probabilities of each land cover trajectory. We implemented CMAP and ML classifiers in R language and processed the classifications for both case studies in a computer with Windows 7, Intel® Core™ i7-4770 (3.40 GHz) Processor, and 8 GB of Random Access Memory (RAM). Processing time for each set of classifications using a non-optimized version of the code varied between 12 and 17 minutes. We also independently implemented a version of the algorithm in Interface Description Language (IDL) which we used to verify the correctness of results acquired using the R version.

Figure 5.10 - General methodology for Compound Maximum *a Posteriori* (CMAP) classification.



Thicker connection lines indicate a high number of outputs. Inputs are highlighted in light blue and outputs are highlighted in light green.

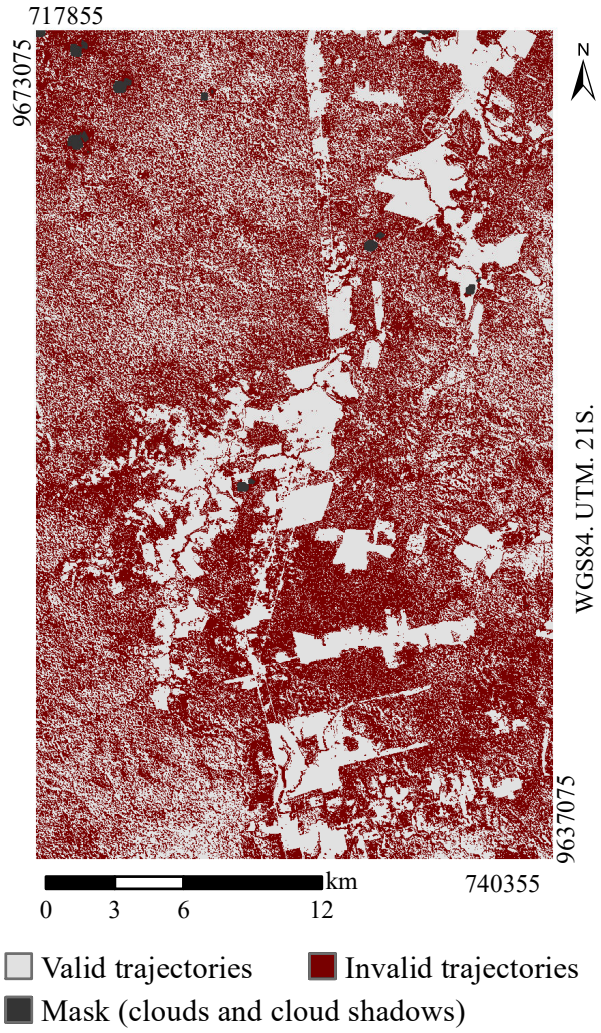
SOURCE: The author.

Classified images were evaluated using confusion matrices and global and class-focused accuracy indexes. Both the confusion matrices and the derived accuracy indexes are calculated based on the Monte Carlo approach and evaluated by the average and standard deviation values. For each classified image, we randomly selected 50 samples per land cover class from the test set and constructed one confusion matrix, from which the Overall Accuracy (OA), Global Kappa, Producer's Accuracy (P.A.), and User's Accuracy (U.A.) indexes were calculated. This process was repeated 100 times for each land cover classified image, with varying sample selection, resulting in 100 values for each index, which were compared using the Wilcoxon unpaired test at 1% significance level. Additionally, we also verified the number and placement of disagreements in pixel classification considering the ML and CMAP classifiers, as well as the number of invalid transitions obtained when stacking ML classifications. We also checked the eventual classification of invalid trajectories from CMAP results, to further validate the algorithm.

5.3 Case Study 1 - results and discussion

As previously mentioned, a common way to obtain land cover trajectories would be stacking land cover classifications obtained by the independent use of supervised classifiers, such as ML. The stacking of ML classifications in Case Study 1 resulted in 54.9% of the observed pixels being classified as an invalid trajectory, that is, a land cover trajectory that contains at least one impossible transition resulting from obvious classification errors. The invalid trajectories obtained using ML are illustrated in Figure 5.11. In CMAP, these trajectories are non-existent.

Figure 5.11 - Validity of land cover trajectories obtained by stacking ML classifications of Case Study 1.



SOURCE: The author.

The number of pixels in which the traditional ML classification and the CMAP ones disagreed is illustrated in Table 5.2. As can be seen, the use of the discriminative transition matrix to calculate the *a priori* probabilities of trajectories resulted in more pixels with disagreement than the use of the simplified approach. This was expected since the simplified approach effectively produces results similar to concatenating ML classifications but only considering possible transitions, whereas the discriminative matrix has the potential to change the classification at each time by weighting possible transitions differently.

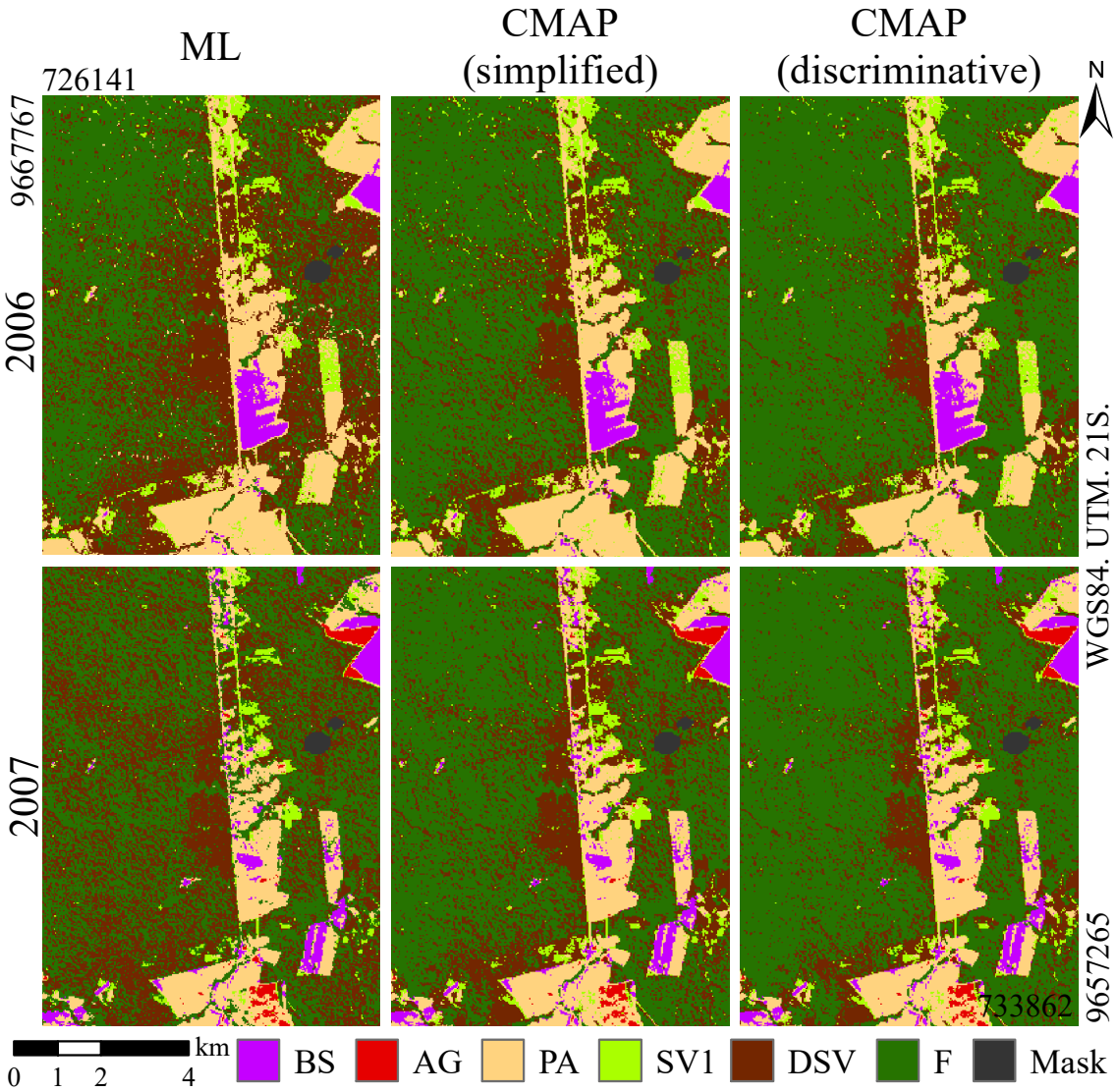
Table 5.2 - Per year and trajectory percentage of disagreement between CMAP and ML classifications in Case Study 1, disregarding masked areas.

Date	CMAP (simplified)	CMAP (discriminative)
2005	17.2	19.3
2006	10.8	11.7
2007	31.4	34.5
2008	19.9	21.3
2009	17.5	20.3
2010	16.5	19.3
Trajectory	54.9	56.0

SOURCE: The author.

A subset of the classifications of years 2006 and 2007, respectively the ones in which fewer and more disagreements occurred, are presented in Figure 5.12. Notice that in ML classifications there are many pixels misclassified as *Developed Secondary Vegetation* scattered among those classified as *Forest*. In CMAP results, pixels classified as *Developed Secondary Vegetation* are concentrated in features located in previously deforested areas, although the information about the actual event of deforestation was not included in the analysis. As can be seen, just by avoiding invalid or improbable transitions, the use of CMAP reduces the classification noise, without considering a contextual approach.

Figure 5.12 - Subset of classifications of Case Study 1.



In which BS = Bare Agricultural Soil, AG = Agriculture, PA = Pasture, SV1 = Initial Secondary Vegetation, DSV = Developed Secondary Vegetation, and F = Forest.

SOURCE: The author.

As a secondary effect of only classifying valid trajectories, CMAP also provides improved land cover classification for each date. The complete set of accuracy indexes for the classifications of Case Study 1 is presented in Table 5.3. In this table, the highest value for each index, as well as those considered statistically similar using the Wilcoxon unpaired test at 1% of significance level, are highlighted in bold font.

The use of CMAP led to higher values of Global Kappa and Overall Accuracy indexes

than the traditional ML classifier for 5 of the 6 classified images of Case Study 1. Only for 2006, ML and CMAP with the simplified approach for the trajectory probability calculation obtained statistically similar overall indexes. Regarding the approach used for calculating the trajectory probability, the simplified one achieved similar overall indexes to the one based on the discriminative matrix for 5 of the 6 images. The average Overall Accuracy values differed in less than 0.02 for all CMAP classifications.

Improvements in overall classification indexes occur mainly due to the improvement of the classification of the classes *Initial Secondary Vegetation*, *Developed Secondary Vegetation*, and *Forest*, as expected given those are the classes involved in the majority of the defined invalid transitions. The decrease in the misclassification of these classes can be assessed in the average confusion matrices of the classifications for 2007 presented in Table 5.4. In this sense, we also perceive a significant improvement of the User's Accuracy of the classes *Developed Secondary Vegetation* and *Forest* in CMAP classifications, compared to the ML-based ones. It happens because the transitions that result in these classes are limited in CMAP, so the number of pixels of other classes misclassified as *Developed Secondary Vegetation* and *Forest* tends to reduce. As expected, classes not limited in CMAP in this case study (*Bare Agricultural Soil*, *Agriculture*, and *Pasture*) are very similar in all the analyzed classifications of each date, as indicated by statistically similar Producer's and User's Accuracy average values for the majority of classifications.

Table 5.3 - Average values of the accuracy indexes for the classifications in Case Study 1.

	2005						2006					
	ML		CMAP (simp.)		CMAP (disc.)		ML		CMAP (simp.)		(disc.)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A	U.A.	P.A
BS	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
AG	**0.90**	**0.44**	**0.90**	**0.44**	**0.90**	**0.44**	-	-	-	-	-	-
PA	**0.68**	**1.00**	**0.68**	**1.00**	**0.68**	**1.00**	**1.00**	**1.00**	**0.99**	**1.00**	**1.00**	**1.00**
SV1	0.79	0.47	**0.84**	**0.59**	**0.83**	**0.58**	**0.99**	0.96	0.99	0.99	**1.00**	**1.00**
DSV	0.68	0.88	0.70	**0.96**	**0.75**	0.91	0.94	**0.96**	0.94	0.93	**0.98**	0.87
F	0.61	0.69	**0.97**	0.94	0.85	**0.97**	**0.95**	0.96	0.93	0.94	0.88	**0.98**
Kappa	0.69±0.02		**0.79±0.02**		**0.78±0.02**		**0.97±0.01**		**0.97±0.01**		0.96±0.01	
OA	0.75±0.02		**0.82±0.02**		**0.81±0.02**		**0.97±0.01**		**0.97±0.01**		0.96±0.01	
	2007						2008					
	ML		CMAP (simp.)		CMAP (disc.)		ML		CMAP (simp.)		(disc.)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A	U.A.	P.A
BS	**0.98**	**0.99**	**0.98**	**0.99**	**0.98**	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
AG	**0.98**	**1.00**	**0.99**	**0.99**	**0.99**	**1.00**	**0.99**	**0.95**	**0.99**	**0.95**	**0.98**	**0.96**
PA	**0.90**	**0.98**	0.87	**0.98**	0.87	**0.98**	**0.66**	**0.93**	**0.67**	**0.94**	**0.66**	**0.93**
SV1	**0.95**	**0.82**	0.94	**0.82**	**0.96**	**0.82**	**0.88**	**0.55**	**0.86**	**0.56**	**0.87**	**0.55**
DSV	0.71	0.93	0.87	**0.97**	**0.89**	0.90	0.81	0.85	0.91	**0.93**	**0.96**	0.84
F	0.84	0.60	**0.98**	0.84	0.91	**0.89**	0.84	0.79	**0.92**	0.88	0.85	**0.95**
Kappa	0.86±0.02		**0.92±0.02**		**0.92±0.01**		0.82±0.02		**0.85±0.02**		**0.85±0.02**	
OA	0.89 ±0.02		**0.93±0.01**		**0.93±0.01**		0.85±0.02		**0.88±0.02**		**0.87±0.01**	
	2009						2010					
	ML		CMAP (simp.)		CMAP (disc.)		ML		CMAP (simp.)		(disc.)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.	U.A.	P.A	U.A.	P.A
BS	**1.00**	**0.98**	**1.00**	**0.98**	**1.00**	**0.98**	**0.93**	**0.99**	**0.93**	**0.99**	**0.93**	**0.99**
AG	-	-	-	-	-	-	**0.97**	**0.80**	**0.98**	**0.80**	**0.97**	**0.80**
PA	**0.64**	**0.92**	**0.65**	**0.92**	**0.64**	**0.92**	**0.75**	**0.79**	**0.75**	**0.80**	**0.76**	**0.79**
SV1	0.77	**0.51**	0.75	**0.51**	**0.80**	**0.51**	**0.67**	**0.76**	**0.67**	**0.76**	**0.68**	**0.77**
DSV	0.81	0.69	0.92	**0.90**	**0.95**	0.83	0.77	0.80	0.81	**0.84**	**0.88**	0.77
F	0.77	0.82	**0.97**	0.90	0.87	**0.95**	0.86	0.76	**0.91**	0.80	0.85	**0.89**
Kappa	0.73±0.03		**0.80±0.02**		**0.79±0.02**		0.78±0.03		**0.80±0.02**		**0.80±0.02**	
OA	0.78±0.02		**0.84±0.02**		**0.84±0.02**		0.82±0.02		**0.83±0.02**		**0.83±0.02**	

simp.=simplified, disc.=discriminative, P.A.=Producer's Accuracy, U.A.=User's Accuracy, OA=Overall Accuracy. Kappa's and OA's values are presented as $m \pm sd$ in which m is the average value and sd is the standard deviation of values. The highest average value, as well as those statistically similar at 1% of significance level, are highlighted in bold font.

SOURCE: The author.

Table 5.4 - Average confusion matrix (%) of the classifications of 2007 in Case Study 1.

		Reference					
		BS	AG	PA	SV1	DSV	F
ML	Bare Agricultural Soil (BS)	99.3	0.0	1.2	0.7	0.0	0.0
	Agriculture (AG)	0.0	99.5	0.0	1.6	0.0	0.0
	Pasture (PA)	0.7	0.0	98.0	10.5	0.0	0.0
	Initial S.V. (SV1)	0.0	0.5	0.9	81.6	0.8	2.0
	Developed S.V. (DSV)	0.0	0.0	0.0	0.7	93.1	37.7
	Forest (F)	0.0	0.0	0.0	5.0	6.2	60.3
CMAP (simplified)		BS	AG	PA	SV1	DSV	F
	Bare Agricultural Soil (BS)	99.3	0.0	1.4	0.9	0.0	0.0
	Agriculture (AG)	0.0	99.4	0.0	1.4	0.0	0.0
	Pasture (PA)	0.7	0.0	97.8	13.8	0.0	0.1
	Initial S.V. (SV1)	0.0	0.6	0.8	82.4	0.8	2.8
	Developed S.V. (DSV)	0.0	0.0	0.0	1.6	97.0	13.4
	Forest (F)	0.0	0.0	0.0	0.0	2.2	83.8
CMAP (disc.)		BS	AG	PA	SV1	DSV	F
	Bare Agricultural Soil (BS)	99.2	0.0	0.9	0.8	0.0	0.0
	Agriculture (AG)	0.0	99.6	0.0	1.1	0.0	0.0
	Pasture (PA)	0.8	0.0	98.1	13.9	0.0	0.0
	Initial S.V. (SV1)	0.0	0.4	1.0	82.0	0.6	2.0
	Developed S.V. (DSV)	0.0	0.0	0.0	2.2	90.4	8.6
	Forest (F)	0.0	0.0	0.0	0.0	9.0	89.4

S.V.= Secondary Vegetation.

SOURCE: The author.

5.4 Case Study 2 - results and discussion

The validity of land cover trajectories obtained by stacking the ML classifications of Case Study 2 is illustrated in Figure 5.13. The percentage of pixels in which the traditional ML classification and the CMAP ones disagreed is presented in Table 5.5. The complete set of accuracy indexes for the classifications of Case Study 2 is presented in Table 5.6. The highest value for each index, as well as those statistically similar using the Wilcoxon unpaired test at 1% of significance level, are highlighted in bold font.

Figure 5.13 - Validity of land cover trajectories obtained by stacking ML classifications of Case Study 2.

SOURCE: The author.

Table 5.5 - Per year and trajectory percentage of disagreement between CMAP and ML classifications in Case Study 2.

Date	CMAP (simplified)	CMAP (discriminative)
2008	30.4	31.0
2010	20.8	22.8
2013	12.9	15.4
Trajectory	50.1	52.9

SOURCE: The author.

Approximately half of the pixels (50.1%) in the used images would be classified as an invalid land cover trajectory by the stacking of the ML classifications in Case Study 2. Because of its formulation, all land cover trajectories obtained by CMAP are valid. Similarly to Case Study 1, we also obtained more accurate land cover classification using CMAP rather than using ML, with classes involved in invalid transitions being the most changed.

Table 5.6 - Average values of the accuracy indexes for the classifications in Case Study 2.

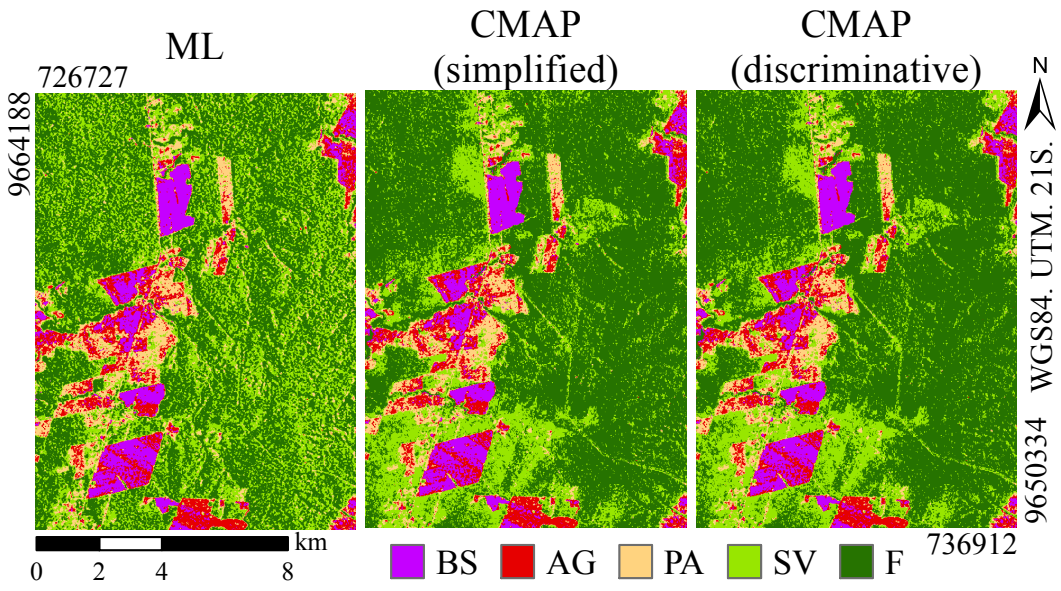
	2008					
	ML		CMAP (simplified)		CMAP (discriminative)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
BS	0.62	**0.83**	**0.64**	**0.84**	**0.63**	**0.83**
AG	**0.53**	**0.57**	**0.54**	**0.58**	**0.54**	**0.57**
PA	0.72	**0.42**	**0.78**	**0.42**	**0.77**	**0.40**
SV	0.49	0.42	**0.72**	**0.78**	**0.73**	**0.79**
F	0.54	0.63	**0.80**	**0.78**	**0.80**	**0.80**
Kappa	0.47±0.04		**0.60±0.03**		**0.60±0.03**	
OA	0.58±0.03		**0.68±0.03**		**0.68±0.03**	
	2010					
	ML		CMAP (simplified)		CMAP (discriminative)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
BS	**0.92**	**0.89**	**0.91**	**0.89**	**0.92**	**0.89**
IA	**0.89**	**0.89**	**0.89**	**0.88**	**0.89**	**0.89**
AC	**0.82**	**0.93**	**0.82**	**0.93**	**0.83**	**0.93**
CP	**0.93**	**0.65**	**0.93**	**0.65**	**0.93**	**0.65**
SP	0.79	**0.93**	0.79	**0.93**	**0.81**	**0.93**
SV1	0.86	**0.83**	0.85	**0.85**	**0.90**	0.73
SV2	**0.55**	0.50	0.47	0.45	0.48	**0.53**
SV3	0.50	0.54	**0.71**	**0.57**	**0.73**	**0.55**
MF	**0.54**	**0.48**	0.51	**0.49**	0.48	**0.49**
MA	**0.78**	0.88	0.75	**0.96**	0.72	**0.96**
Kappa	**0.73±0.02**		**0.73±0.02**		**0.73±0.02**	
OA	**0.75±0.02**		**0.76±0.02**		**0.76±0.02**	
	2013					
	ML		CMAP (simplified)		CMAP (discriminative)	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
BS	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
AG	**0.88**	**1.00**	**0.87**	**1.00**	**0.88**	**1.00**
PA	**1.00**	**0.75**	**1.00**	**0.74**	**0.99**	**0.76**
SV1	**0.85**	0.83	0.84	**0.91**	**0.86**	0.74
SV2	0.60	0.42	**0.76**	**0.49**	0.61	**0.51**
SV3	**0.59**	0.47	**0.61**	**0.55**	**0.61**	0.52
MF	0.40	0.65	0.45	**0.68**	**0.46**	**0.67**
MA	0.73	0.74	**0.80**	0.77	0.75	**0.84**
Kappa	0.69±0.02		**0.73±0.02**		0.72±0.02	
OA	0.73±0.02		**0.77±0.02**		0.75±0.02	

P.A.=Producer'a Accuracy, U.A.=User's Accuracy, OA=Overall Accuracy. Kappa's and OA's values are presented as $m \pm sd$ in which m is the average value and sd is the standard deviation of values. The highest average value, as well as those statistically similar at 1% of significance level, are highlighted in bold font.

SOURCE: The author.

The improvement in classification results is particularly expressive for the classification of the ALOS/PALSAR image of 2008 (gain of 30.4% in the average Kappa value and more than 30% of changed pixels in relation to ML classification) mainly due to the contribution of optical data in other dates to decrease the confusion between *Secondary Vegetation* and *Forest*, as illustrated in Figure 5.14 and Table 5.7. As is widely known, optical data may be greatly affected by cloud cover whereas SAR data can be acquired almost independently of atmospheric conditions (JENSEN, 2007). Therefore, in areas frequently covered by clouds, such as tropical forests, it is probable that only SAR data may be available during certain periods of the year. This type of data usually delivers poorer results for land cover classification than optical data in tropical forests, mainly due to the confusion between primary and secondary forests (PEREIRA et al., 2016; REIS et al., 2020c), and more refined data processing and classification methods are necessary (LI et al., 2012). However, CMAP allows the use of information from optical data available on other dates to better classify the SAR image on the date of interest, even allowing the use of different legends to meet accuracy requirements.

Figure 5.14 - Subset of classifications of Case Study 2, obtained using ALOS/PALSAR data of 2008.



In which BS = Bare Agricultural Soil, AG = Agriculture, PA = Pasture, SV = Secondary Vegetation, and F = Forest.

SOURCE: The author.

Table 5.7 - Average confusion matrix (%) of the classifications of ALOS/PALSAR data in Case Study 2.

		Reference				
		BS	AG	PA	SV	F
ML	Bare Agricultural Soil (BS)	82.8	32.6	17.4	0.0	0.0
	Agriculture (AG)	16.5	56.9	33.2	0.2	0.0
	Pasture (PA)	0.7	9.4	41.6	4.6	1.8
	Secondary Vegetation (SV)	0.0	1.0	7.5	41.6	35.4
	Forest (F)	0.0	0.0	0.3	53.6	62.8
CMAP (simp.)		BS	AG	PA	SV	F
	Bare Agricultural Soil (BS)	83.5	31.8	16.1	0.0	0.0
	Agriculture (AG)	15.6	57.9	33.5	0.2	0.0
	Pasture (PA)	0.8	9.1	42.0	2.1	0.0
	Secondary Vegetation (SV)	0.0	1.1	8.2	77.8	21.6
	Forest (F)	0.0	0.0	0.4	19.9	78.4
CMAP (disc.)		BS	AG	PA	SV	F
	Bare Agricultural Soil (BS)	83.4	31.2	17.8	0.0	0.0
	Agriculture (AG)	15.8	57.2	34.1	0.1	0.0
	Pasture (PA)	0.8	10.1	39.9	1.3	0.0
	Secondary Vegetation (SV)	0.0	1.3	8.1	79.1	20.4
	Forest (F)	0.0	0.1	0.1	19.4	79.6

SOURCE: The author.

5.5 Chapter conclusions

This study presented the Compound Maximum *a Posteriori* (CMAP) classifier and two applications for land cover trajectories classification. For both applications, CMAP was capable of incorporating the knowledge of land cover dynamics to classify multi-temporal sets of remote sensing images directly into valid land cover trajectories. These results were compared to those obtained using the post-classification comparison approach and the traditional Maximum Likelihood (ML) classifier.

For the presented case studies, at least 50% of the land cover trajectories generated by post-classification comparison and ML were considered invalid. By its own design, CMAP does not return invalid transitions/trajectories as classification results. In a way, CMAP expands the ML classifier by adding a multi-temporal *a priori* probability of trajectories and excluding invalid transitions from the analysis. The main difference between CMAP and ML is that in CMAP the land cover trajectory is classified as a whole, so the decision about the attributed land cover classes is done after all probabilities of each time are calculated. In ML, decisions are made

independently for each date. It means that, for each image being analyzed, CMAP is able to navigate among the class likelihoods in order to choose the highest one that results in a valid transition and not simply the highest one, which is not viable in hard algorithms like the traditional ML classifier. In this sense, we verified that not only the use of CMAP leads to better accuracy indexes, but it also reduces the classification noise for individual dates.

CMAP is a fairly direct way of introducing previous knowledge into the classification process. The only additional input in relation to the traditional ML classifier is the calculation of the *a priori* probability of trajectories, which may be as simple as defining which transitions are valid and which are not. We proposed two ways to calculate these probabilities, namely the simplified approach and the one based on discriminative transition matrices. Although both approaches resulted in land cover classification with similar overall accuracy indexes, the discriminative approach may be more suitable for difficult classification situations, because of the possibility of weighting each transition differently. Another interesting feature of CMAP is that it allows for different legends to be used for each image, which may be interesting in cases of low class-separability for specific data sets, as illustrated by the classification of one SAR image.

A few limitations were identified in the way we applied the CMAP classifier in the present study. The first one is the use of land cover training samples collected for each image to be classified which may be very expensive for studies with a high number of images. However, parallel studies show good results in using the spectral information from one Landsat5/TM image to train an ML classifier that will be used to classify a different calibrated image from the same sensor (REIS et al., 2019). This process is known as signature extension, spectral extensibility, generalization of training samples (WOODCOCK et al., 2001), or transfer learning. Thus it is also expected that using transferred training samples for use in CMAP will lead to accurate classification results. The second perceived limitation regards areas under clouds/cloud shadows that were masked from analysis in this study. Although these areas were small, greater areas are expected when using a higher number of optical images, even considering image composition techniques. How to properly integrate areas under clouds/shadows in CMAP and the impacts of using transferred training samples to classify image composites in CMAP will be investigated in the following chapter.

Additionally, we focused on simple assumptions about the land cover transitions

in the present study, such as if these were possible or impossible given the class definition and time gap. Although we have found promising results using CMAP this way, future studies could benefit from more rigorous methods of calculating the used transition matrices, as well as analysis in different classification scenarios, i.e. study area, classes of interest, number of images, and type of data. In this sense, it is a relatively straightforward process to substitute the statistical distribution used within CMAP, which could be more adequate for non-Gaussian data sets. Furthermore, there are many opportunities in adjusting the theoretical basis of CMAP to allow the use of other base classifiers, such as using Random Forest to derive the probabilities for each observation (DUTRA et al., 2022), which in turn would allow for the use of intra-annual times series, as done within SITS package, within the current inter-annual analysis.

6 CLASSIFYING ANNUAL LULC TRAJECTORIES WITH 37 YEARS OF LANDSAT DATA

The present chapter aims to investigate the classification of valid only LULC trajectories for longer remote sensing time series. Particularly, we investigate the challenges associated with using CMAP to classify annual LULC trajectories from 1984 to 2020, to be latter used for the analysis of the forest regeneration trajectories. This classification task is more complex than the ones presented in the previous chapter because it involves a bigger area and 37 years of observations. These characteristics led to problems:

a) related to specific constraints of CMAP regarding the length of the time series and number of possible trajectories;

b) related to areas without observation due to the presence of clouds/cloud shadows or with detection failure;

c) related to the difficulty in collecting labeled samples for the complete time series.

Regarding the last problem, it is possible to solve this problem by using the spectral information from known objects in a given set of images to train a supervised classifier that will be used to identify features pertaining to the same class in another image, distant in either space or time. This process is known as signature extension, spectral extensibility, and generalization of training samples(WOODCOCK et al., 2001). According to Olthof et al. (2005), the performance of such methods is dependent on radiometric consistency in the set of images, which may be affected by factors such as atmospheric/imaging conditions and phenology. These can usually be minimized by choosing images at proximate times of the year (COPPIN et al., 2004).

We propose and evaluate solutions for these problems in this chapter. We use combinations of legends and class transition/trajectory *a priori* probabilities to reduce the number of calculated trajectories and computational costs of CMAP. We also propose methods to deal with times without observations during the classification process. Another differential in this chapter is the analysis of the impacts of training supervised classifiers using transferred samples, i.e. labeled samples, and correspondent spectral information, of the classes of interest that may have been collected in different images than the ones being classified.

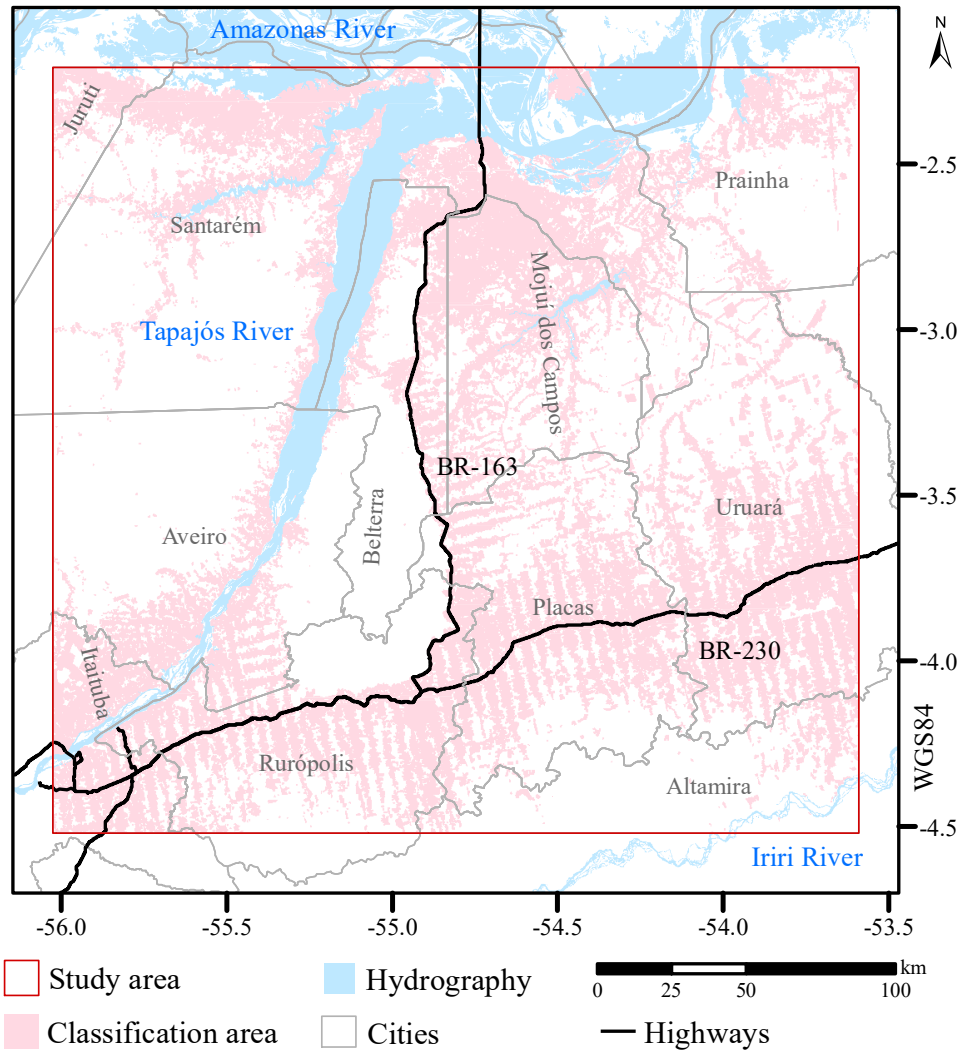
This chapter is structured as follows: the study area, LULC classes, imagery, and used auxiliary data are detailed in Section 6.1. The methods used to classify the LULC trajectories are detailed in Section 6.2. Classification results are presented in Section 6.3 and discussed in Section 6.4. These results were evaluated regarding the accuracy of classifications, and improvements of CMAP over the traditional ML classifier and post-processing filters to solve problems with invalid transitions. The main conclusions of this experiment and their fit within the present thesis are then summarized in Section 6.5.

6.1 Materials

6.1.1 Study area

This chapter focus on the main study area of the present thesis, as illustrated in Figure 6.1. In this study, we only classified areas that are 150 m apart from areas that have been deforested up to 2020 ('classification area' in Figure 6.1). This decision was made to avoid the costs of classifying areas in which forest regeneration trajectories can not occur, by definition. Previously deforested areas were extracted from the PRODES 2020 consolidated data. The buffer of 150 m (or 5 pixels) was empirically selected to account for historic misalignment problems within the PRODES data (INPE, 2017; WANG et al., 2020).

Figure 6.1 - Study area.



SOURCE: Based on IBGE (2019) and MInfra (2021).

6.1.2 Imagery

Landsat data from the sensors TM, ETM+, and OLI, respectively onboard satellites Landsat 4-5, Landsat 7, and Landsat 8 were selected for analysis. These were obtained from Collection 1, ready for use in the GEE platform. Only data in surface reflectance within Tier 1 (T1) correction level were used. The images on Tier 1 have been radiometrically and geometrically corrected. According to USGS (2022), data from different Landsat sensors are inter-calibrated within Tier 1, and the georegistration of scenes is consistent with up to 12 m of Root Mean Square Error

(RMSE) between pairs of images. Thus, this data set is well-suited for pixel-level time series analysis. Atmosphere correction is done using one of two algorithms: the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS), for TM and ETM+ images, or the Land Surface Reflectance Code (LaSRC), for OLI images. Additionally, the images also present a Quality Assessment (QA) band, which flags pixels with observations problems derived from detection failures, atmospheric conditions, and other adverse conditions of the land surface that may compromise the radiometric integrity of the pixel. Cloud, shadow, water, and snow masks within the QA band were produced using CFMASK (USGS, 2022).

Two sets of data were used in this study. The first one encompasses the images collected between June and November of each year, which is a period of lower cloud cover in the study area. This set was used mainly for visual interpretation. The second set of data is composed of images with less than 50% of cloud cover over land, as computed within the QA band, and acquired between August and November of each year. This second time-window corresponds to the dry period in the region (RADAMBRASIL, 1976). This data set was used for classification purposes, and thus selected both to preserve the spectral correspondence between images, as well as to improve the separation of emerging secondary vegetation from pastures and other cultivated areas (RUFIN et al., 2015; MÜLLER et al., 2016). The correspondent images in Top of Atmosphere (TOA) format were also analyzed within the GEE platform, due to methodology constraints of preprocessing algorithms (Section 6.2.1).

6.1.3 Field data and auxiliary information

Field data from 2010 to 2017, detailed in Chapter 3, were used to construct a visual interpretation key for the adopted LULC classes and Landsat images. Other used auxiliary data encompass:

a) the TerraClass classifications (second legend) for the years 1991, 2000, 2004, 2008, 2010, 2012, and 2014;

b) PRODES data from the year 2020;

c) the MapBiomas Collection 5 dataset.

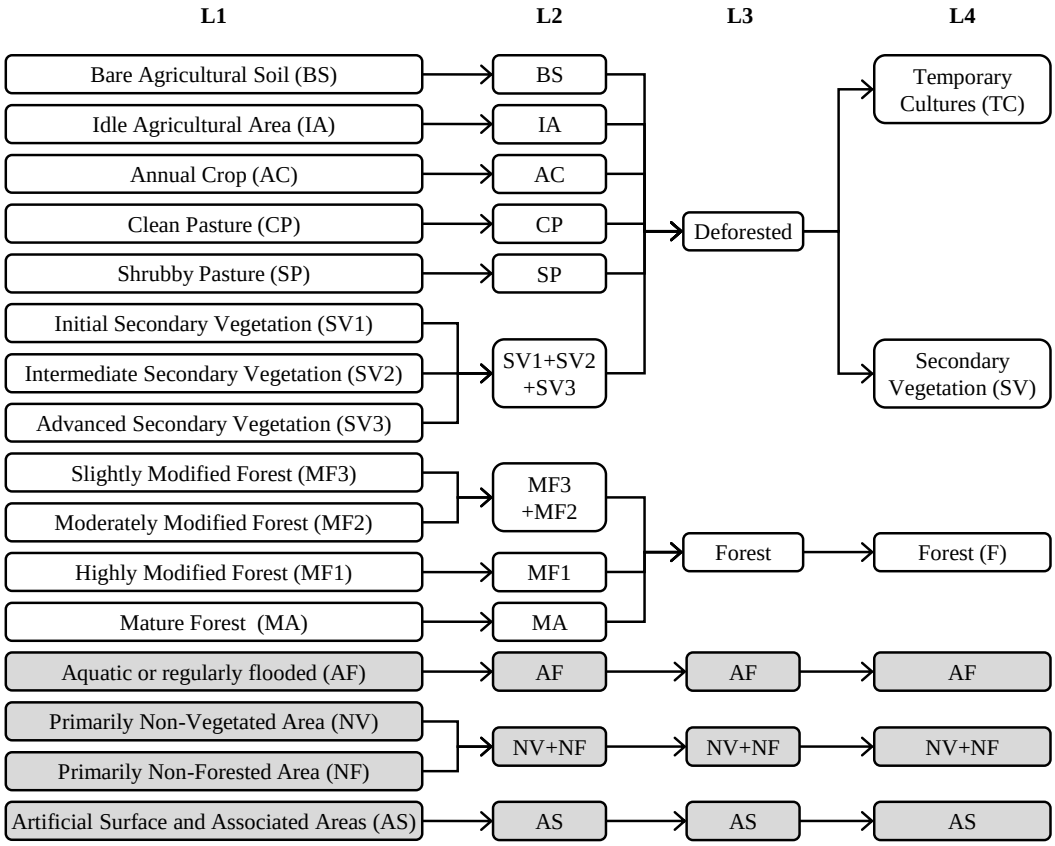
All products were converted to WGS84 reference system. A co-registration step was not executed. The monthly mosaics of Planet images (`https://www.planet.com/`, accessed by the Web Feature Service (WFS) within QGIS software) and the

localities data set from the Brazilian Institute of Geography and Statistics (IBGE) (IBGE, 2019) were also used.

6.1.4 Legend definition

The classification process used in this study previews the existence of different legend levels. These legends largely correspond to aggregations of the classes defined in Chapter 4 and are illustrated in Figure 6.2. The L1 legend represents the most abstract detail level considered for the study area. The L2 legend was defined as the most detailed level in which we could separate the classes based on the visual interpretation of the Landsat mosaics described in Section 6.2.1. The L3 legend is the least detailed one, used to optimize the classification process within CMAP. Lastly, the L4 legend encompasses the final classes of interest for the study area. Not depicted in this figure but present due to the nature of the digital classification process, we also have the class *Not Observed*, which represents pixels not observed due to cloud cover, cloud shadow, or detection failures.

Figure 6.2 - Legends used during the classification process.



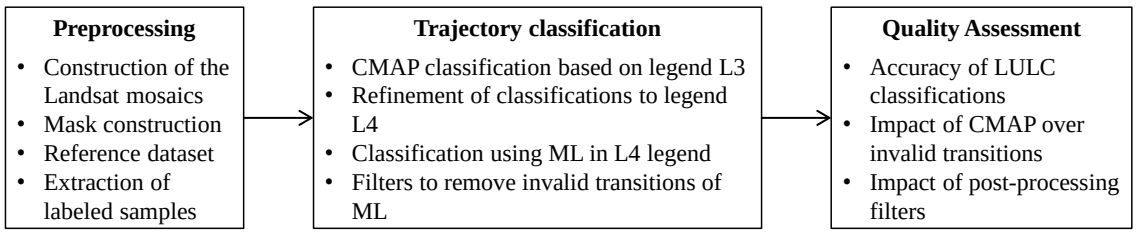
Classes identified but masked from analysis (Appendix B) are highlighted in light-gray.

SOURCE: The author.

6.2 Methods

The classification of the Landsat time-series into LULC trajectories follows the basic steps previously discussed in Chapter 2. These steps were adapted to the present classification problem, as summarized in Figure 6.3, and detailed in the following sections.

Figure 6.3 - Summary of classification steps.



SOURCE: The author.

6.2.1 Preprocessing

First, we derived the image composites used for visual interpretation and classification. These will be referred to as 'Landsat mosaics'. For both sets, we considered the QA band as the indicative of adequate pixels to compose the Landsat mosaic. Pixels flagged as detection failures, clouds, cloud shadows, snow, or those with saturated/invalid values (such as reflectance lower than zero) were removed from the analysis. For the set of mosaics used for visual interpretation (images from June to November of each year), we generated both yearly and monthly mosaics by extracting the median value of the remaining pixels in the Surface Reflectance images. Images of all sensors were processed jointly. The median was selected because it provides comparable mosaics derived from sets of images that may have been acquired with different frequencies in each year, and also presented better classification results and a lower frequency of remaining areas of cloud/cloud shadow when compared to tested BAP techniques for the study area (unpublished results).

Small but frequent and scattered areas of clouds/cloud shadows are not detected by the QA band and remain in the mosaics even after the use of the median. The use of images with less than 50% of cloud cover helps but is not enough to solve this problem. To mask residual clouds/cloud shadows, we opted to use two additional criteria: the simple cloud score for Landsat data, and the distance to a pixel flagged as cloud/cloud shadow. The simple cloud score or Landsat data is a native algorithm in GEE that computes the cloud-likelihood score of a pixel, with likelihoods varying from zero to 100, based on a combination of brightness, temperature, and the Normalized Difference Snow Index (NDSI) from the images in TOA format. Although it is not considered a robust cloud detector, it is a useful

index to compare different sets of images (GEE, 2022). In this study, areas with likelihoods of being a cloud higher than 20 were considered as clouds. Following this process, we also masked any pixel that is within a Chebyshev distance (also known as the chessboard distance) of 2 pixels from any pixel flagged as cloud/cloud shadow. These further masked images were used to compute the second set of mosaics, with images from August to November. Only yearly mosaics were generated, also using the median value of Surface Reflectance images from each year. As this set was designed as the classification input, images from the OLI sensor were processed separately from TM or ETM+, due to small differences in spectral resolution.

Following the construction of the Landsat mosaics, we also masked areas of no interest from the classification process. These are areas in which forest regeneration events either can not occur or would represent very specific and diverse regeneration dynamics that are outside of the scope of this thesis. These areas correspond to the classes *Aquatic or regularly flooded*, *Primarily Non- Vegetated Area*, *Artificial Surface and Associated Areas*, as defined in Chapter 4, and the not previously defined class *Primarily Non-Forested Areas*. The latter class corresponds to areas with original vegetation different than forests usually expected within the Amazon Biome, such as savannas and other types of natural vegetation. The masking process was done based on the Landsat mosaics from June to November and the visual interpretation of the data, as detailed in Appendix B.

The last preprocessing step was the collection of labeled samples. The most straightforward proposal for the collection of samples in the Study Area is the use of areas of agreement between MapBiomas and TerraClass data sets of the same year. Besides supposedly returning classifications with higher accuracy, looking only into agreement areas between the data sets also has the advantage of discarding the misaligned portions of TerraClass features, as the Collection 5 of MapBiomas was also constructed over the Landsat Collection 1 within GEE (SOUZA et al., 2020). For this analysis, it was necessary to first establish the correspondence between the legends adopted for the present analysis and in each project. This harmonization is detailed in Appendix B.

However, the samples obtained in areas of agreement between the MapBiomas and TerraClass datasets could not be promptly used in the current classification scheme. CMAP inherited many characteristics of the Maximum Likelihood (ML) classifier. One of them is the expectation of labeled samples of uni-modal classes for training. As such, the direct use of the agreement between the TerraClass and MapBiomas

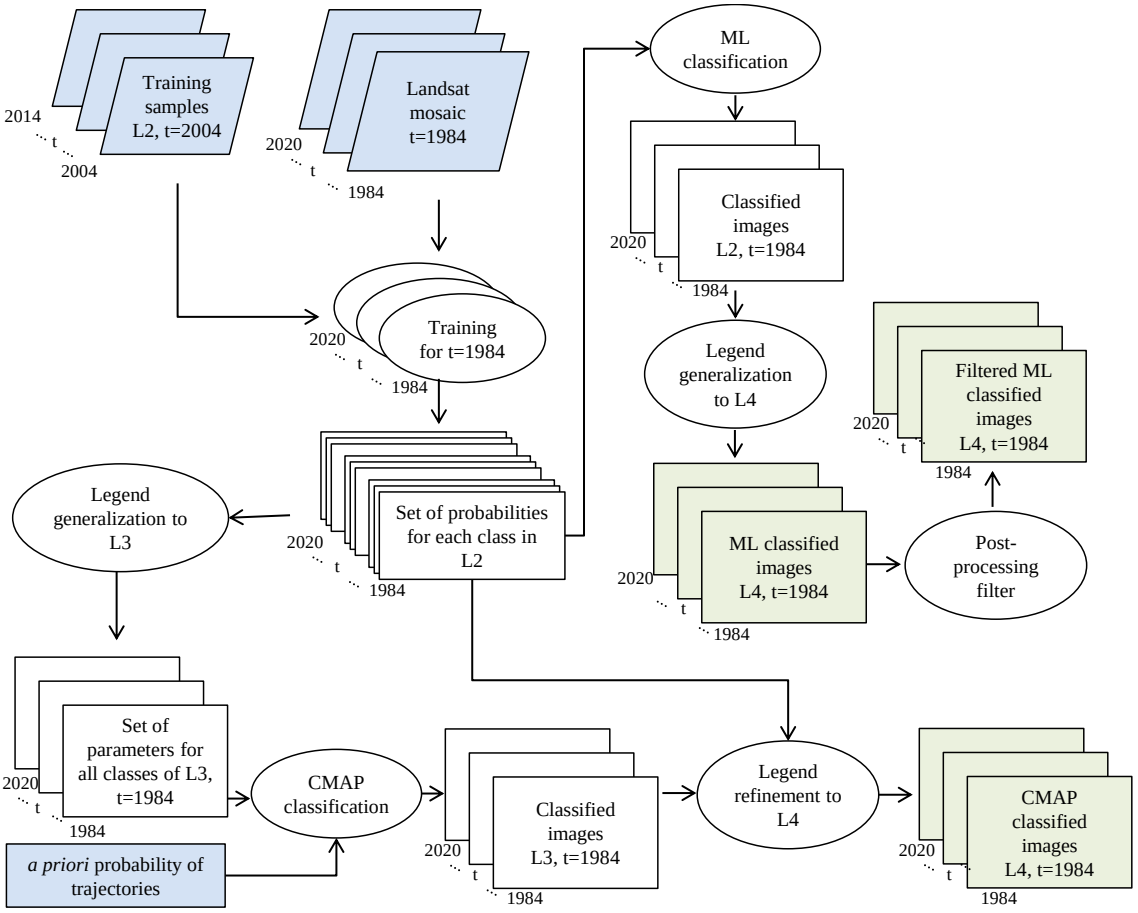
data, which mixes classes with high spectral variability, can hinder the classification process in CMAP. To minimize this problem, a new set of labeled samples was collected directly over the Landsat mosaics from August to November, for the years 2004, 2008, 2010 (TM+ETM+ mosaics), and 2014 (OLI mosaic). These samples were used to train the CMAP classifier, as explained in Section 6.2.2. The data from the remaining years were reserved to test the classification results and did not pass through a refinement phase. The steps for labeled sample collection are also detailed in Appendix B. As a result of this process, we collected thousands of labeled samples for the classes in legend L2, for the years 1993, 2004, 2008, 2010, 2014, and 2016 (Appendix B).

6.2.2 Trajectory classification

The classification of LULC trajectories was based on different legend levels and additional steps to deal with *Not Observed* areas. The classification process is illustrated in Figure 6.4 and explained as follows. Note that the treatment of *Not Observed* areas, not explicit in this figure, was done concomitantly with the classification step. It will be explained separately for clarity.

First, we randomly selected 5,000 samples per class/year from the manually collected set to train the classifier and extracted the pixel values of the Landsat Mosaics of the correspondent year. Only bands from the Near-Infrared (band 4 for TM and ETM+, band 5 for OLI) and Short-wave Infrared (bands 5 and 7 for TM and ETM+, bands 6 and 7 for OLI) were considered for classification. This selection aims to minimize residual differences between the images in the visible bands (WULDER et al., 2018).

Figure 6.4 - General classification flowchart.



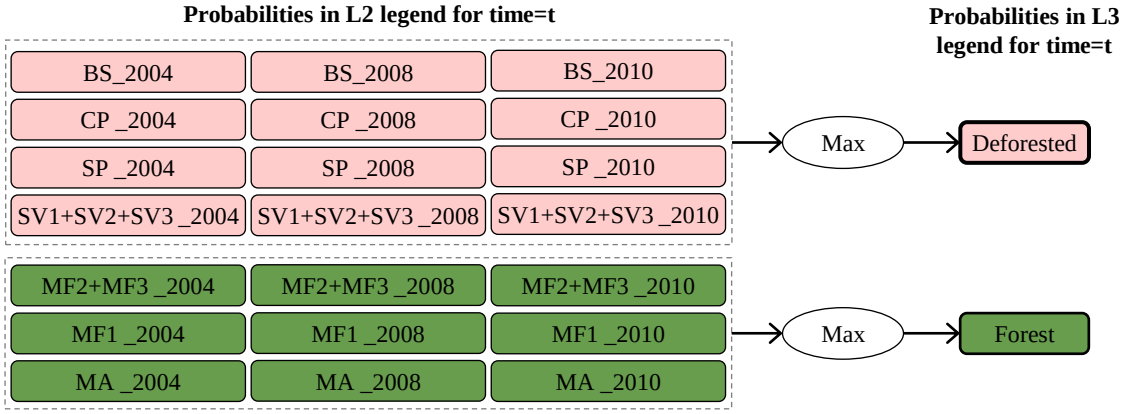
Input data are highlighted in light blue. Outputs are highlighted in light green.
SOURCE: The author.

We independently calculated one mean vector and one covariance matrix for each class/year, based on each set of labeled samples/images. Each pair of mean vector and covariance matrix was used to calculate one value of $P(\vec{x_t}|\omega_t^{k_t})$/pixel. For TM/ETM+ mosaics, we used labeled samples collected for years 1993 (1 class), 2004 (6 classes), 2008 (6 classes), 2010 (6 classes) and 2016 (1 class) (Appendix B). Therefore, we have 20 values of $P(\vec{x_t}|\omega_t^{k_t})$/pixel per year for TM/ETM+ mosaics. OLI mosaics were trained using samples from 2014 (6 classes) and 2016 (1 class), resulting in seven values of $P(\vec{x_t}|\omega_t^{k_t})$/pixel per year. This arrangement present two main problems within CMAP: 1) the number of possible trajectories for long time series, and 2) calculated values for areas with clouds/cloud shadows.

Given the nature of CMAP, classifying 20 LULC classes in 37 years, as is the case

of the TM/ETM+ Mosaics, would mean the calculation of 20^{37} trajectories. To avoid the time and computational costs associated to such a task, we reduced the input legend for CMAP from L2 to L3. For a given year, the probabilities of classes *Deforested* and *Forest* were set as the maximum values of the probabilities calculated for correspondent classes in that same year, based on samples from any date. This process is illustrated in Figure 6.5.

Figure 6.5 - Example for the generalization of the probabilities calculated for an TM/ETM+ mosaic from L2 to L3 legend.



In which BS = Bare Agricultural Soil, CP = Clean Pasture, SP = Shrubby Pasture, SV1 = Initial Secondary Vegetation, SV2 = Intermediate Secondary Vegetation, SV3 = Advanced Secondary Vegetation, MF1 = Highly Modified Forest, MF2 = Moderately Modified Forest, MF3 = Slightly Modified Forest, and MA = Mature Forest. Idle Agricultural Area and Annual Crop were omitted from the legend because no labeled samples were found for these classes. Probabilities samples in L3 legend are depicted as $[class]\_[year\ of\ transferred\ samples]$.

SOURCE: The author.

CMAP uses the *a priori* probabilities of the trajectories ($P(s)$) as input. We opted for the simplified approach proposed in Chapter 5, in which $P(s) = 0$ for invalid trajectories and $P(s) = 1/N_s$ for valid trajectories. Here, N_s is the number of valid trajectories. A valid trajectory is one that only has valid transitions, as determined by the matrix in Figure 6.6. In L3, only the classes *Deforested* and *Forest* participate of the classification process. Because a *Deforested* area can never change to a *Forest* area, only 38 trajectories are valid for the study area, i.e. 1 considering 37 observations of *Forest* and 37 considering the first observation of *Deforested* in each year from the analyzed period.

Figure 6.6 - Validity of transitions for L3 legend.

$t-1$ \ t	Deforested	Forest
Deforested	Valid	Invalid
Forest	Valid	Valid

SOURCE: The author.

So far, we have completed the classification process for the L3 legend. The next step is the refinement of these classifications to the L4 legend. Here, we reclassified *Deforested* areas as *Secondary Vegetation* or *Temporary Cultures* by the analysis of the maximum value of the probabilities calculated using the manually collected samples in the L2 legend. Note that *Temporary Cultures* is obtained if the maximum probability corresponds to classes *Clean Pasture*, *Shrubby Pasture*, or *Bare Agricultural Soil*. This classification scheme returns the same results as using CMAP with the three classes in L4 in 37 years and considering that only a class different from *Forest* becoming *Forest* as an invalid transition in the simplified approach of determining $P(s)$. This happens because both CMAP and the methods used to generalize and refine legends focus on the maximum value of probabilities. The use of the simplified approach to define $P(s)$ leads to constant values for all valid trajectories that do not influence the calculation of the maximum.

For comparison purposes, we also derived the ML classifications using the same input dataset. These classifications were obtained by assigning to each pixel the class that returns the maximum probability value in the sets of each year generated considering the L2 legends, and then merging the correspondent classes to generalize the results into the L4 legend. We also generated a set named 'Filtered ML', using the same process to separate these classes within the MapBiomas dataset: all classes labeled as *Secondary Vegetation* or *Forest* in the ML classifications were merged and re-separated based on the existence of areas classified as *Temporary Cultures* in previous observations.

As previously mentioned, *Not Observed* areas were treated concomitantly with the classification process. It was done in three stages:

a) for the years 2014 to 2020, in which we have two types of mosaics, we used the probabilities calculated in the ETM+ Mosaics to fill *Not Observed* areas in probability sets from the OLI Mosaics. These modified probability

sets were the ones used for classification (CMAP and ML);

b) remaining areas without observation, in any year (1984-2020) then received a value of $P(\vec{x_t}|\omega_t^{k_t}) = c$ for every class, in which c is a constant value that does not impact the maximum calculation of CMAP. This way, CMAP is not affected by pixels without information but is still capable to use that time to ponder the *a priori* probability of the possible transitions;

c) areas without observation in both types of mosaics are relabeled as *Not Observed* (NO) in the final classification (CMAP, ML, and Filtered ML).

6.2.3 Quality assessment

Given the lack of available reference data for the study area in the whole period of analysis, we used the areas of agreement between TerraClass and MapBiomas classifications (generalized to L4) from 1991, 2000, and 2012 to evaluate the three classifications (CMAP, ML, and Filtered ML) obtained for each correspondent year. The classification accuracy was calculated based on the insertion of Olofsson et al. (2014) recommendations within a Monte Carlo approach. Firstly, we calculated the sample size (n) for each classified image (OLOFSSON et al., 2014):

$$n \approx \left(\frac{\sum W_i S_i}{S(\widehat{O})} \right)^2 \tag{6.1}$$

in which $S(\widehat{O})$ is the desired standard error of the estimated Overall Accuracy, set to 0.01 in this study, W_i is the mapped proportion of the area of class i, derived from the agreement between TerraClass and MapBiomas, and $S_i = \sqrt{U_i(1-U_i)}$, being U_i the User's Accuracy of class i, fixed as 0.65 for *Temporary Cultures*, 0.70 for *Secondary Vegetation*, and 0.60 for *Forest*. These values correspond to an approximation of the lowest User's Accuracy calculated for correspondent classes in Chapter 5 - Case Study 1. Values of n varied between 2843 to 3845 for each year. Based on these values, we fixed n=4500 samples, so we would be able to allocate 1500 samples/per class, stratified from the classifications in areas in which we had labeled test samples.

From these samples, we estimate each cell of the confusion matrix ($\widehat{p}_{ij}$) as:

$$\widehat{p}_{ij} = W_i \frac{n_{ij}}{n_i}, \tag{6.2}$$

in which W_i was set as the proportion of class i outside the mask in the the classified image being analyzed, n_{ij} is the number of samples of class j classified as class i, and n_i is the number of samples classified as i. The values of Overall Accuracy, User's Accuracy, and Producer's Accuracy were calculated from this matrix. This process was repeated 1000 times for each classification/year, and calculated values for each accuracy index were compared using the Wilcoxon unpaired test at 1% significance level [1].

Similarly to the analysis executed in Chapter 5, we verified the number of disagreements between pairs of classifications of the same year. We also computed the number of invalid transitions obtained when stacking LULC classifications for each year.

6.3 Results

The preprocessing and classification results are presented in this section. The Landsat Mosaics are illustrated in Section 6.3.1, along with a brief discussion about the adopted preprocessing methods. In Section 6.3.2, we compare the obtained LULC classifications regarding their impacts on invalid trajectories, accuracy, and mapping results.

6.3.1 Landsat mosaics

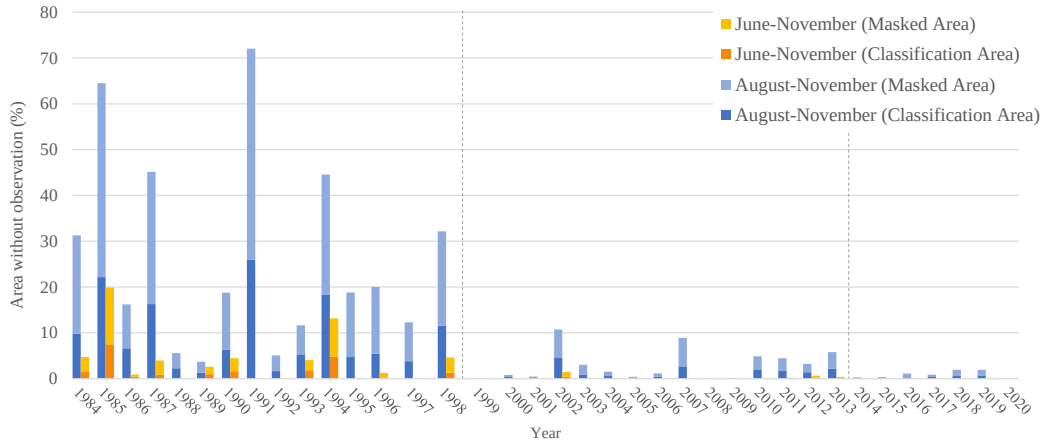
The proportion of *Not Observed* areas due to clouds, shadows, and/or detection failure from the Landsat Mosaics are illustrated in Figure 6.7. This figure regards the whole study area, with results stratified considering the classification mask[2]. For years 2014 to 2020, we filled *Not Observed* from the OLI Mosaics with observed areas in the ETM+ mosaics prior to calculation. As can be seen in Figure 6.7, *Not Observed* pixels are common up to 1998. Lower and less frequent percentages of this type of pixel are found after the launch of the Landsat 7 satellite, carrying the ETM+ sensor, in 1999, followed by even lower values after 2014, with the addition of OLI images. Mosaics from August-November present a higher proportion of *Not Observed* pixels than the ones from June-November, as expected. In the study area, June-July are, historically, the months with a higher percentage of clear observations. Thus their removal caused a significant decrease in the chances of finding a suitable pixel for analysis. The additional filters applied in the August-November mosaics further

[1] Note that although all the sampling was done over the same original set of test samples, each random selection is completely independent from the other. This setup was chosen to allow the stratification of samples selection based on each classification

[2] The proportional values of *Not Observed* areas calculated considering only the classification area and the August-November Mosaics are presented in Section 6.3.2.

decreased the number of available observations. Figure 6.8 shows a subset of the mosaics of 1991, to exemplify some effects of reducing the time window of images from June-November to August-November, as well as the applied filters.

Figure 6.7 - Proportion of areas without observations due to clouds, shadows and/or detection failure in the two sets of Landsat mosaics.



Values are given in percentage (%) relative of the total number of pixels in each mosaic (8581 × 9021 = 77409201 pixels) and stratified in relation to coverage of the mask. The dashed lines indicate the inclusion of Landsat 7 and 8 images, respectively from 1999 and 2014.

SOURCE: The author.

As exemplified by Figure 6.8:

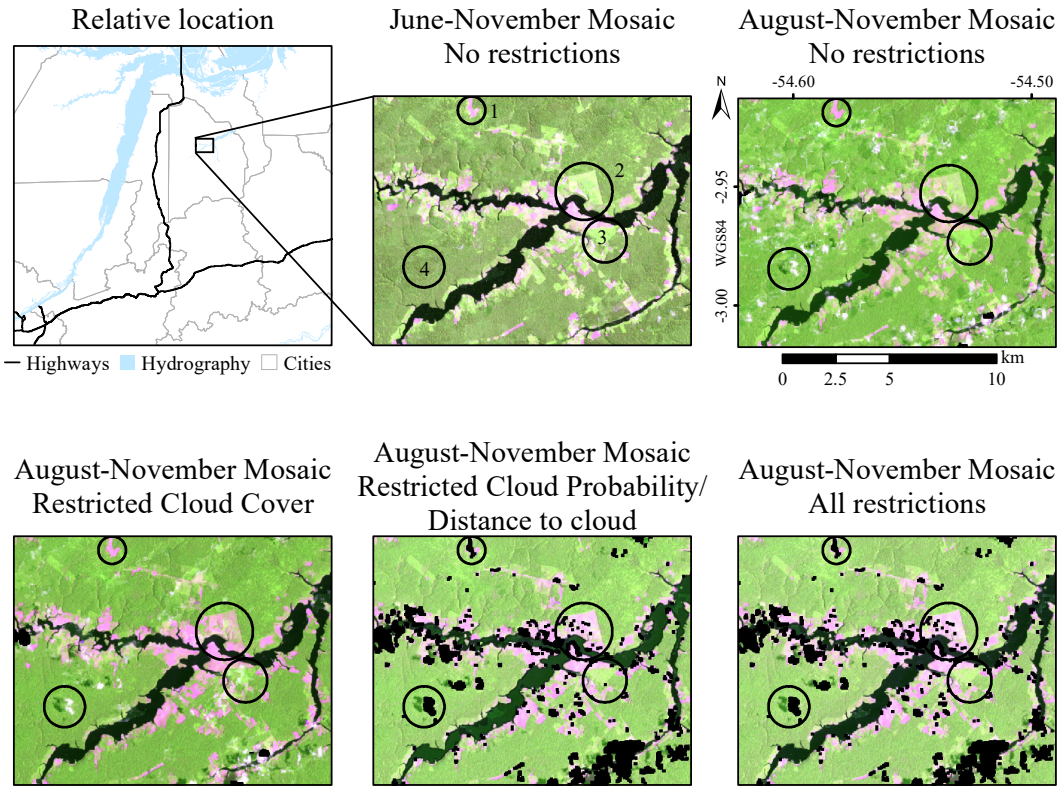
a) CFmask, with the parameters applied by the USGS, does not identify all areas of clouds/clouds shadows. The use of the median to mosaic the Landsat time-series of each year tends to choose a suitable observation, as clouds and cloud-shadows usually fall far from the median spectral observation, but only if a clear observation is available within the data set. For instance, the area highlighted by a black circle with Index 4 is covered by clouds only in the mosaics from August to November. In this area, it is also possible to see that the filter of cloud probability is useful to mask clouds missed by Fmask, but not in masking cloud-shadows.

b) Some areas of saturated soil exposure, such as mining areas and bare agricultural soil, were also masked by the cloud probability filter. Examples

of this process can be seen in the areas highlighted with Indexes 1 and 2. We found the loss of these pixels an acceptable compromise since unchecked clouds within the Landsat mosaics tend to cause the false detection of a deforestation event within CMAP, which in turn would lead to the classification of secondary vegetation in forested areas.

c) As previously mentioned, mosaics constructed using the narrow window of August to November present a significant loss of observations. Nonetheless, prioritizing observations with similar spectral behavior across time is of utmost importance in classification processes that use transferred samples. For instance, we can see an area of *Shrubby Pasture* highlighted by the black circle and Index 2. In the June-November Mosaic, this area presents similar spectral behavior to areas of *Secondary Vegetation*, as such the one highlighted by Index 3. In the August-November Mosaics, we can see a higher soil exposure in the *Shrubby Pasture*, while the *Secondary Vegetation* appears to present a more constant spectral behavior.

Figure 6.8 - Examples of the effect of the restrictions in the Landsat Mosaics, using images obtained from Landsat/TM images from 1991.



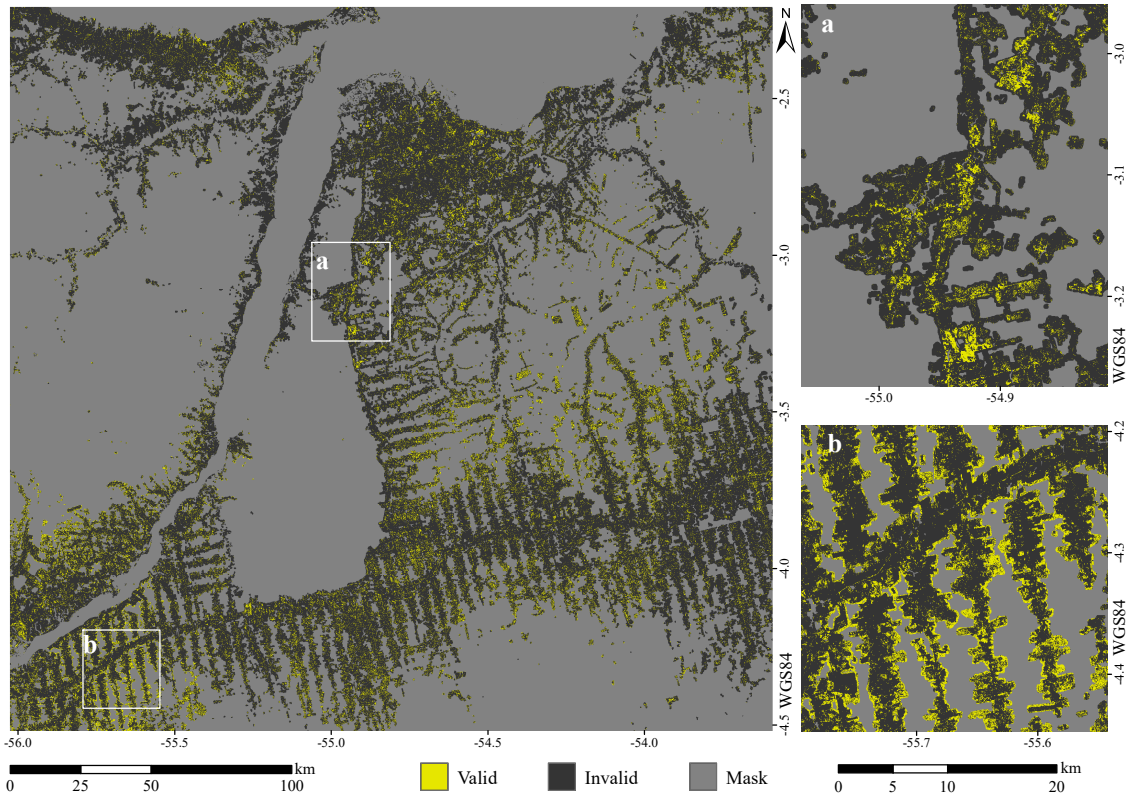
Color composition 5(R)4(G)B(7) with 2% clip contrast. Highlights depicting 1) area of exposed soil masked as cloud due the cloud probability filter; 2) area with higher soil exposure in August-November mosaics; 3) area with seemingly constant vegetation cover in all mosaics; 4) cloud masked by the cloud probability filter. Note that the highlighted areas are only identified on the June-November Mosaic.

SOURCE: The author.

6.3.2 Classifications

Similarly to the results of Chapter 5, the stacking of ML classifications leads to a high amount of transitions that would be impossible in the real world. Here, it resulted in nearly 89% of the pixels within the classification area labeled as an invalid trajectory. The distribution of these pixels within the ML classifications is illustrated in Figure 6.9. CMAP and Filtered ML do not return invalid trajectories.

Figure 6.9 - Validity of land cover trajectories obtained by stacking ML classifications from 1984 to 2020.

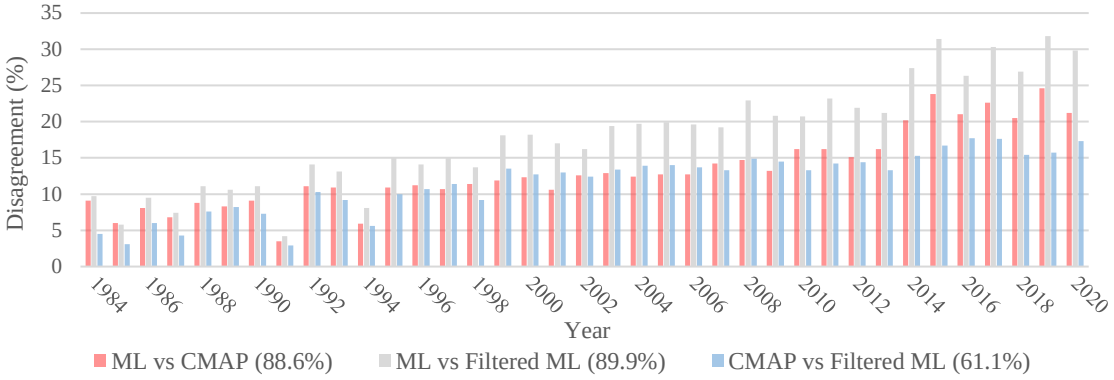


SOURCE: The author.

An invalid trajectory is flagged when a pixel is classified as *Forest* if it has been classified as *Temporary Cultures* or *Secondary Vegetation* at a previous time. In Figure 6.9, it is possible to see that the pattern of distribution of invalid trajectories varies within the study area. Figure 6.9.a exemplifies an area in which the invalid trajectories occur near the boundaries with the masks, whereas in Figure 6.9.b they occur within the center of the features. In both cases, these are areas where ML tends to misclassify *Forest* areas as *Secondary Vegetation* and vice versa.

In this study, CMAP and ML LULC classifications only differ in the case that ML mapped an invalid trajectory. The same consideration is not applicable to disagreements between Filtered ML and ML/CMAP. The disagreement between the pairs of classification in each year and regarding the trajectory is depicted in Figure 6.10.

Figure 6.10 - Per year disagreement between pairs of classifications of the same year, disregarding masked areas.



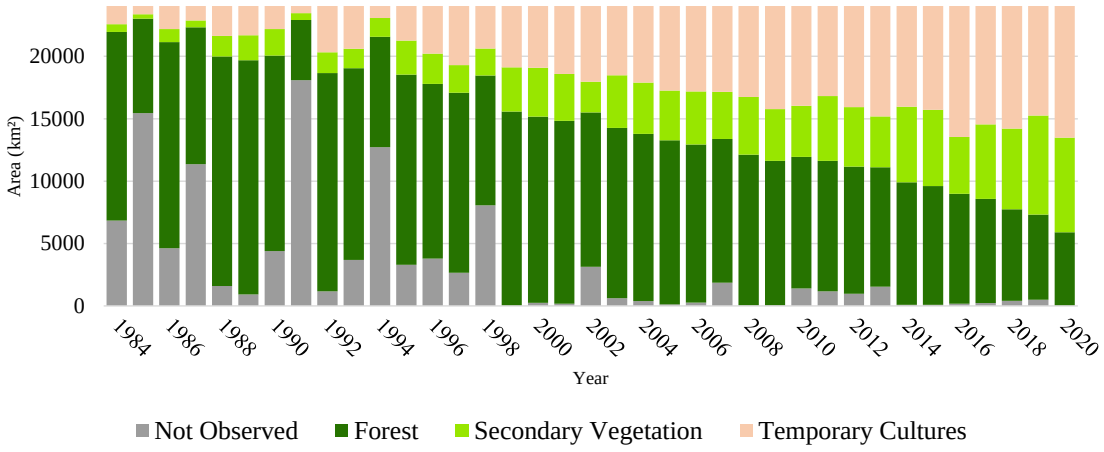
The total disagreement between trajectories is given in parentheses. Disagreements between trajectories are flagged if compared trajectories differ in at least one observation.

SOURCE: The author.

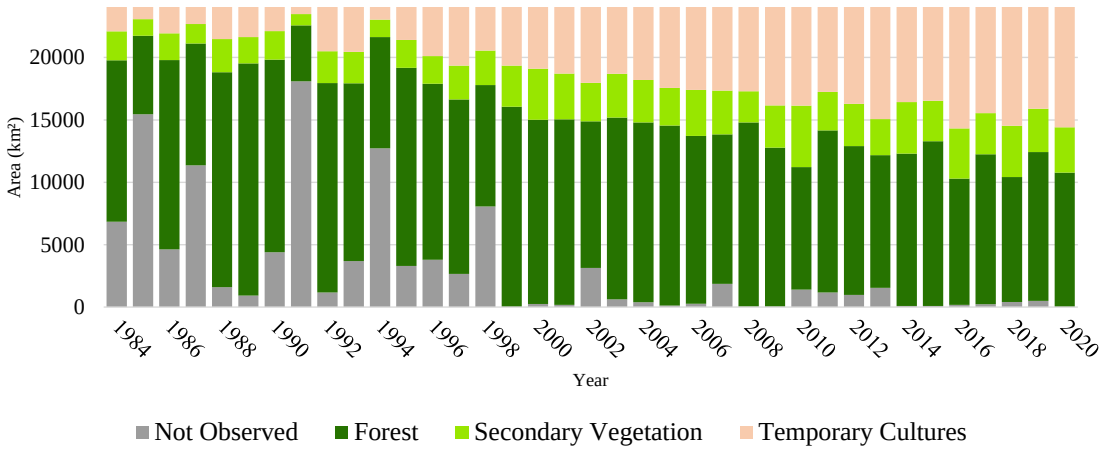
The disagreements between classifications tend to increase with more recent Landsat mosaics being classified. This behavior is more evident when comparing ML and Filtered ML classifications. This result was expected, as Filtered ML highly depends on the previous classifications, with changes in relation to ML classifications cumulative in nature. Trajectories classified by ML and Filtered ML present the highest rate of disagreement, which indicates that CMAP is a well-balanced solution to avoid invalid trajectories. The same behavior can be seen when analyzing the individual LULC classifications of each year.

The area mapped as each class per year and classification approach is presented in Figure 6.11. Here, we only consider the classification area. In this figure, it is also possible to see the amount of *Not Observed* areas in each year. The corresponding values obtained using the filtered MapBiomas collection are shown in Figure 6.12.

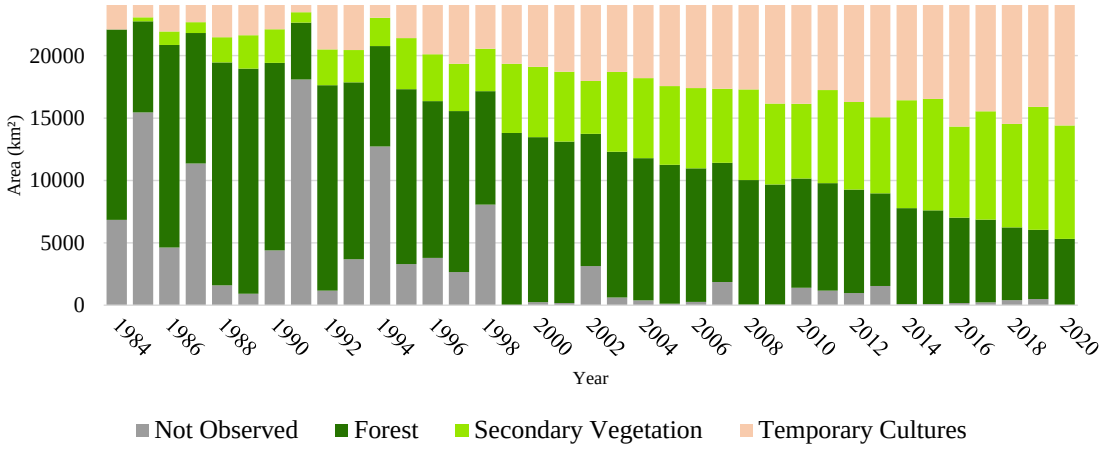
Figure 6.11 - Area mapped of each class per year and classification approach.



(a) CMAP



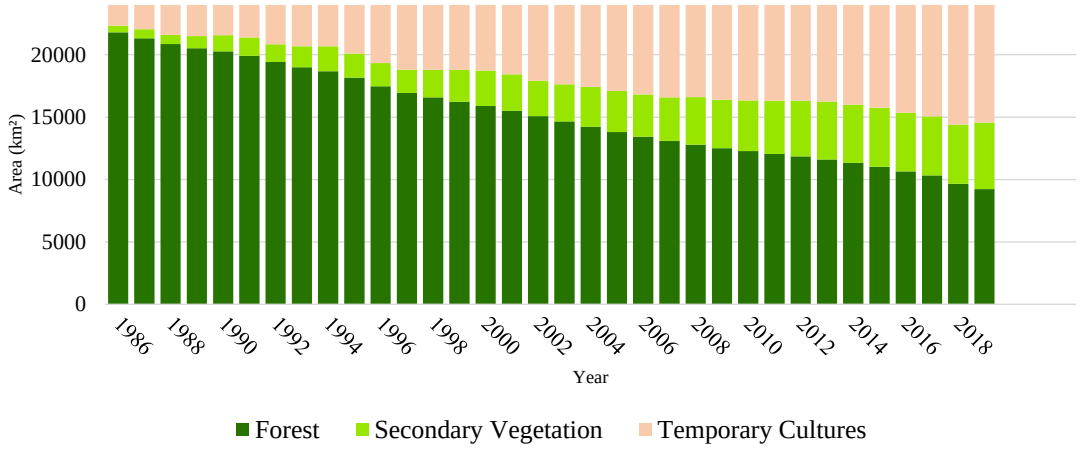
(b) ML



(c) Filtered ML

SOURCE: The author.

Figure 6.12 - Area mapped of each class per year within MapBiomas Collection 5 data set.



Areas of grassland, urban areas, and water were too small to appear in this figure and were removed from the legend.

SOURCE: The author.

In all data sets, we see a tendency for increase with time of the areas of *Temporary Cultures* and *Secondary Vegetation*. Nonetheless, in ML classifications the proportion of *Secondary Vegetation* areas appears to be smaller than the other data sets, mainly from 2014 to 2020, when the least amount of *Not Observed* areas makes this comparison more straightforward. The Filtered ML classifications deliver the highest amount of *Secondary Vegetation*. This result was expected, given that misclassifications of *Forests* in ML tend to increase the amount of *Secondary Vegetation* in Filtered ML.

Additionally, CMAP has presented the smaller absolute difference in area with Filtered MapBiomas data for 18 of the analyzed years, followed by ML in 14 years. The Filtered ML classifications only presented the lower difference in areas for two years. As we applied the same post-classification filter in both Filtered ML and Filtered MapBiomas datasets, both are prone to cumulative errors for the classification of *Secondary Vegetation*. In this sense, the fact that Filtered ML presented the most divergent estimate in areas to MapBiomas classifications further evidence the importance of accurate classifications before the use of a post-processing filter to eliminate invalid transitions.

The accuracy indexes calculated for classifications of the years 1991, 2000, and

2012 are presented in Table 6.1. The corresponding average confusion matrices are presented in Table 6.2. As can be seen, CMAP provided the highest values of Overall Accuracy in the three assessed years. CMAP also presented the highest values of User's Accuracy for the class *Secondary Vegetation* for all cases, even though these values are still relatively low.

Table 6.1 - Average values of the accuracy indexes for the classifications of 1991, 2000, and 2012.

1991	CMAP		ML		Filtered ML	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
Temporary Cultures	**0.88**	0.67	0.85	**0.69**	0.85	0.58
Secondary Vegetation	**0.40**	0.62	0.20	0.59	0.31	**0.72**
Forest	0.97	**0.97**	0.95	0.86	**0.98**	0.93
Overall Accuracy	**0.91**		0.83		0.87	
2000	CMAP		ML		Filtered ML	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
Temporary Cultures	**0.92**	0.74	0.91	**0.75**	0.91	0.72
Secondary Vegetation	**0.42**	0.73	0.26	0.55	0.33	**0.83**
Forest	0.96	**0.93**	0.92	0.86	**0.98**	0.84
Overall Accuracy	**0.86**		0.81		0.81	
2012	CMAP		ML		Filtered ML	
	U.A.	P.A.	U.A.	P.A.	U.A.	P.A.
Temporary Cultures	**0.84**	**0.94**	0.83	0.91	0.83	0.89
Secondary Vegetation	**0.63**	0.71	0.49	0.43	0.51	**0.82**
Forest	0.94	**0.83**	0.83	0.82	**0.98**	0.71
Overall Accuracy	**0.84**		0.78		0.79	

P.A.=Producer's Accuracy and U.A.=User's Accuracy. The highest average value is highlighted in bold font. The unpaired Wilcoxon test did not return similar values to the highest one at 1% significance level.

SOURCE: The author.

Table 6.2 - Average confusion matrix (%) for the classifications of 1991, 2000, and 2012.

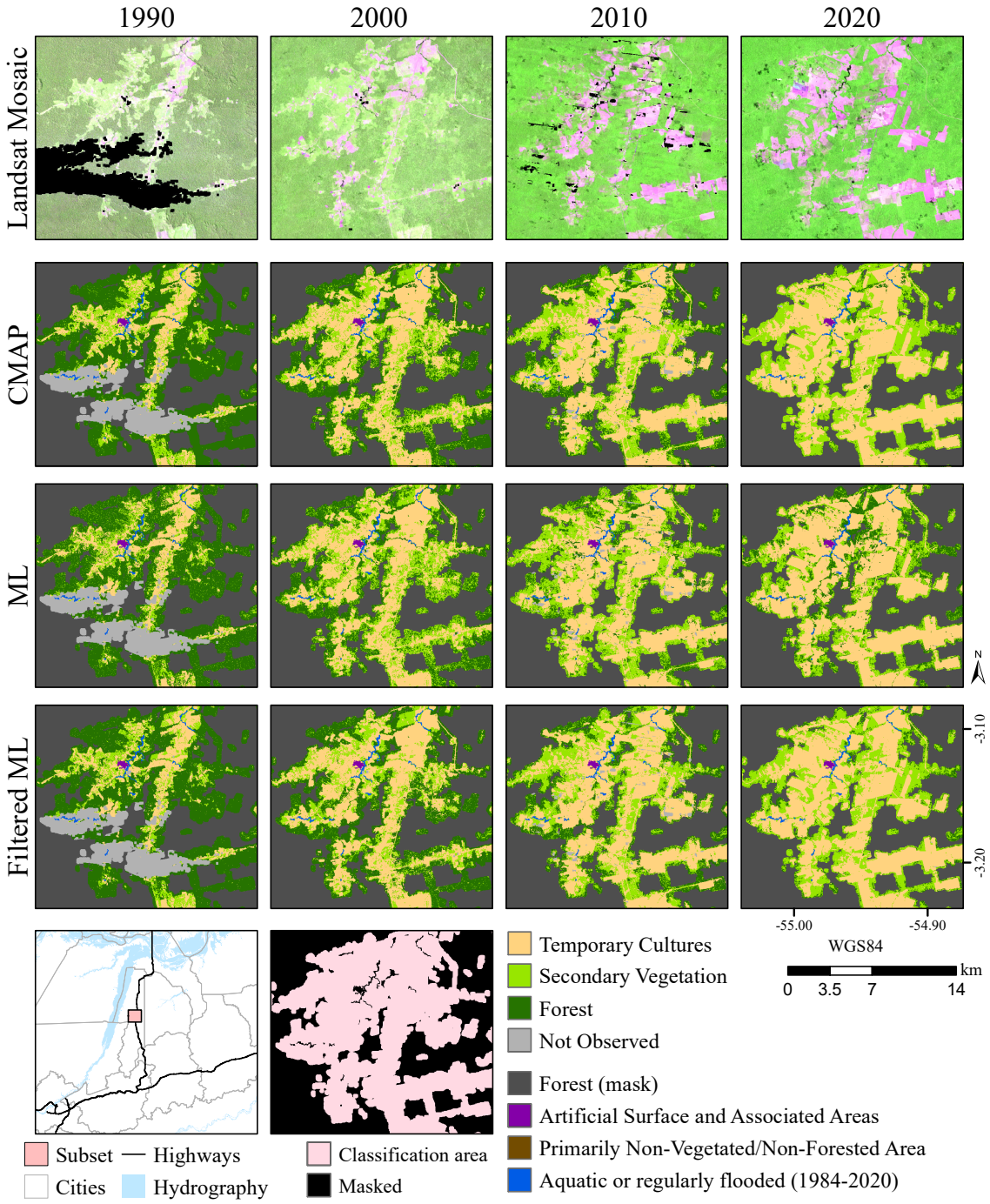
1991	CMAP			ML			Filtered ML		
	TC	SV	F	TC	SV	F	TC	SV	F
Temporary Cultures (TC)	8.3	0.6	0.5	7.6	0.5	0.9	7.6	0.5	0.9
Secondary Vegetation (SV)	3.2	3.7	2.3	1.7	3.0	10.5	5.2	4.4	4.4
Forest (F)	1.0	1.7	78.7	1.8	1.7	72.4	0.4	1.3	75.3
2000	CMAP			ML			Filtered ML		
	TC	SV	F	TC	SV	F	TC	SV	F
Temporary Cultures (TC)	19.0	0.8	0.9	18.7	0.8	1.1	18.8	0.8	1.1
Secondary Vegetation (SV)	5.6	6.9	4.0	4.3	4.5	8.4	6.9	7.8	9.0
Forest (F)	1.1	1.7	60.0	1.9	2.8	57.4	0.3	0.8	54.5
2012	CMAP			ML			Filtered ML		
	TC	SV	F	TC	SV	F	TC	SV	F
Temporary Cultures (TC)	29.4	3.4	2.3	27.8	2.8	2.9	27.8	2.8	2.9
Secondary Vegetation (SV)	1.3	13.1	6.3	0.8	7.2	6.7	3.3	15.4	11.7
Forest (F)	0.6	2.0	41.7	2.0	6.9	42.9	0.1	0.6	35.4

Reference and classifications are depicted in the columns and rows, respectively.

SOURCE: The author.

Figures 6.13 to 6.15 illustrate subsets of the obtained classifications for years 1990, 2000, 2010, and 2020, along with the correspondent Landsat mosaics from August to November. These figures exemplify a pattern observed in the whole study area: whereas the ML classifier tends to classify areas previously deforested as *Forest*, CMAP and ML Filtered tend to attribute the class *Secondary Vegetation* in forested areas that, although suffered some form of degradation, did not have the forest cover completely removed. This event is clearer in Figure 6.13, in which it was expected to see small areas classified as *Forest* along the border of the *Forest* Mask. However, these areas have been burnt sporadically over the years, which lead to the misclassification of these areas at different points in time. Areas less severely affected by fire are depicted in Figures 6.14 and 6.15. In these cases, CMAP tends to be more robust in the classification of *Forest* than Filtered ML.

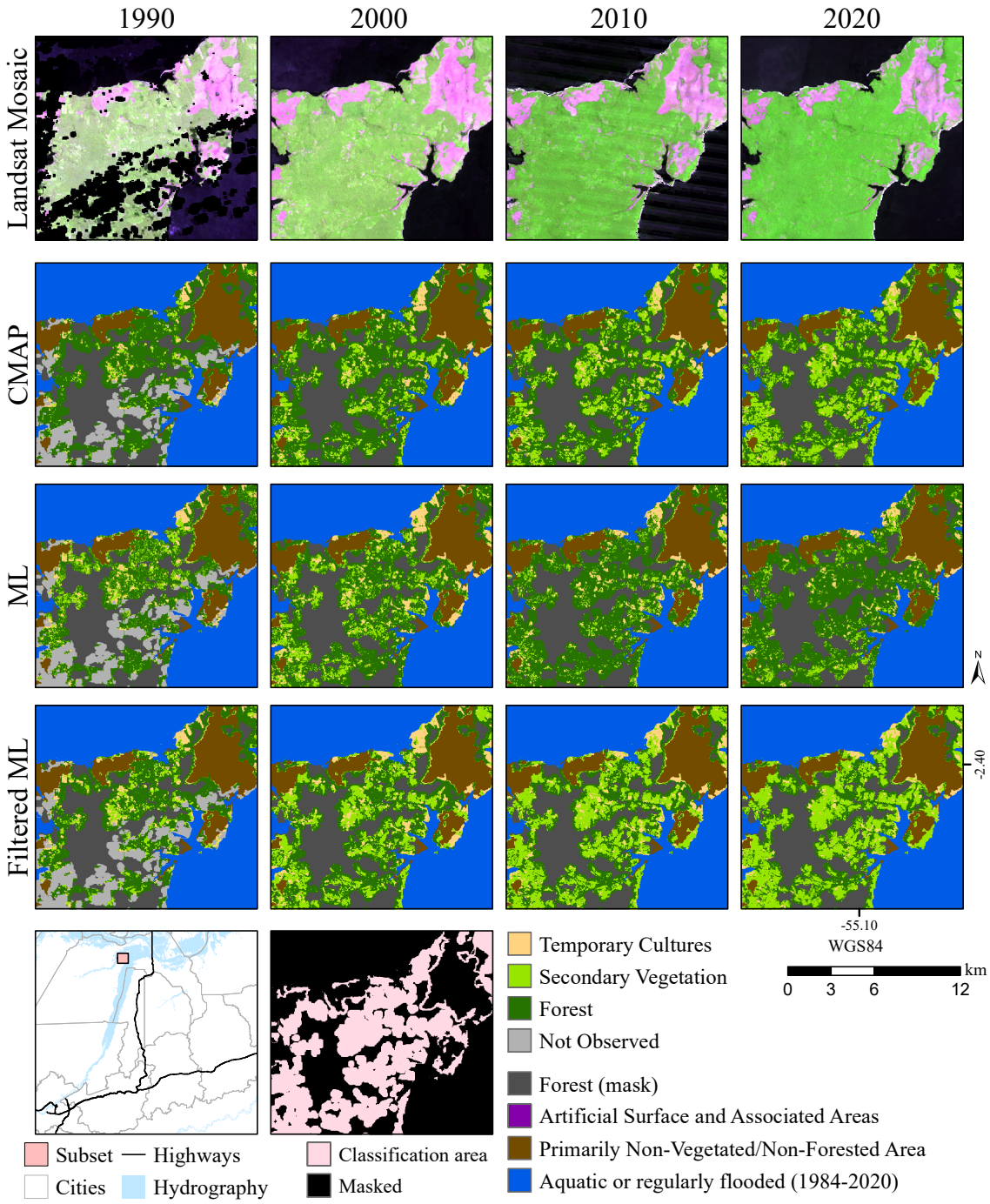
Figure 6.13 - Subset of the classified images. Example in the Cuiabá-Santarém highway (BR-163).



Landsat mosaics are illustrated in color composition R(Swir1)G(NIR)B(Swir2) with 2% clip contrast. Data from 2020 are composites from data derived from Landsat 8 with *Not Observed* areas filled by the correspondent data derived from Landsat 7.

SOURCE: The author.

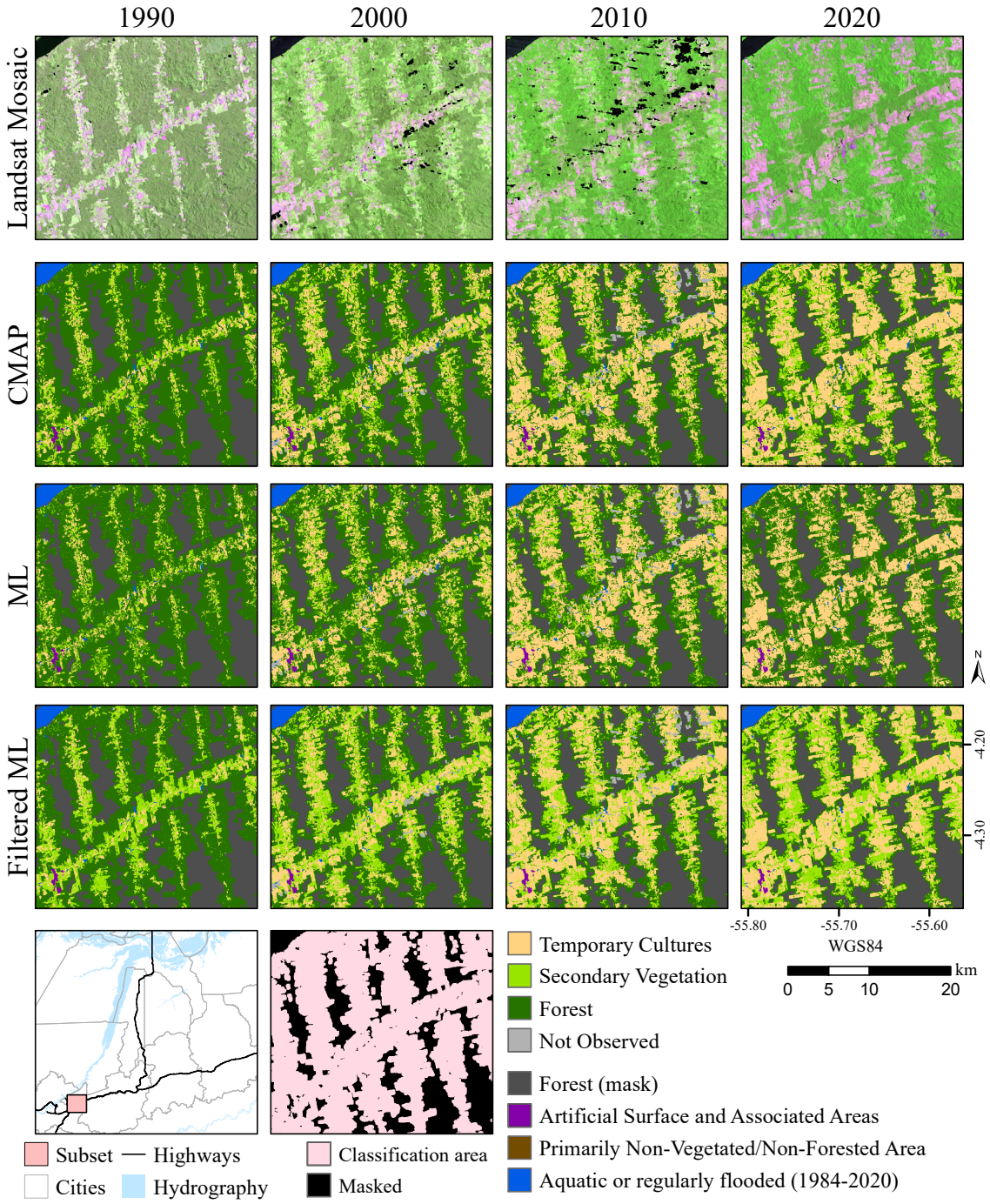
Figure 6.14 - Subset of the classified images. Riverside example.



Landsat mosaics are illustrated in color composition R(Swir1)G(NIR)B(Swir2) with 2% clip contrast. Data from 2020 are composites from data derived from Landsat 8 with *Not Observed* areas filled by the correspondent data derived from Landsat 7.

SOURCE: The author.

Figure 6.15 - Subset of the classified images. Example in the Transamazon highway (BR-230).



Landsat mosaics are illustrated in color composition R(Swir1)G(NIR)B(Swir2) with 2% clip contrast. Data from 2020 are composites from data derived from Landsat 8 with *Not Observed* areas filled by the correspondent data derived from Landsat 7.

SOURCE: The author.

6.4 Discussion

Both ML and CMAP classifiers, so far, uses 1) one observation per analyzed time and 2) labeled samples for each LULC class as input. As highlighted in this chapter, the generation of both inputs involve several methodological characteristics. Reducing the Landsat observations within one year into one annual mosaic is a relatively easy process, given the recent advances in data format standardization (WULDER et al., 2018) and cloud-storage and processing (GORELICK et al., 2017). Nonetheless, there are many decisions to be taken during this step, molded by the premisses and limitations of the present study, that can significantly impact the results, as discussed as follows.

Firstly, reference data for all years of interest are not prompt available for the study areas and the 37 years of interest. To circumvent this problem, we opted to use transferred labeled samples to train ML and CMAP classifiers. The use of such samples assumes very similar behavior of the classes of interest in all observations, which in turn precludes the use of uncalibrated time series, the use of images from different seasons/points of the agricultural calendar, and the use of multi-sensor data sets. As seen in this study, these constraints can lead to large areas without observations for given years, mainly those in the 1980s and 1990s. In this sense, the use of larger time-windows or multi-sensor data to fill missing observations would demand correspondent reference data to train the classifiers accordingly. The launch of TerraClass data sets for years 2018-2022, as well as continued improvements done to the MapBiomas Collections, may improve the collection of more recent samples, which could also enable the use of multi-sensor data in studies focused on more recent images.

The difference in data availability over the years precludes the use of metrics that consider the multi-temporal characteristics of the data and/or more refined statistical metrics such as the quartiles of dry *vs* wet periods, among others (RUFIN et al., 2015; SOUZA et al., 2020). In this study, we also opted to remove the visible bands from the analysis, as calibration problems were causing several classification errors along the time series. As such, the current analysis was based only on the spectral response of the NIR, SWIR1, and SWIR2 bands, which can have led to an overall loss of spectral information. Furthermore, as shown in Reis et al. (2019), although the Surface Reflectance data from Landsat Collection 1 is suitable for classification processes using ML and transferred samples, the use of relative calibration methods can enhance the classification accuracy of *Forest* and

Secondary Vegetation areas. However, such methods demand the existence of labeled samples either for those specific classes or in areas of No-change between all images, which is the main limitation of the present study. Nonetheless, the USGS recently launched the Landsat Data Collection 2 (USGS, 2022), with improved methods for radiometric calibration, which can solve some of the identified problems. New tests using this collection and transferred samples are necessary to assess the suitability of this data set.

Besides the highlighted constrains, the lack of reference samples for the whole study area and/or in all years of interest is also the cause of the main limitation in the analysis of the results presented in this chapter: the current inability to fully assess the classifications' accuracy. As classification errors and the quality of reference samples varies across the years, even in the same study area and using the same data and methods (POWELL et al., 2004; SEXTON et al., 2013), the accuracy estimated for the years 1991, 2000, and 2012 cannot be generalized for the whole set of classifications. Even for these years, the estimates must be interpreted carefully. The used reference samples were obtained from other classifications, and as such may be biased for this type of analysis. For instance, we do not have samples for areas more difficult to classify in which the reference data sets tend to disagree, or from persistent areas of *Secondary Vegetation*, i.e. those in which a deforestation event was not flagged, since these areas will not appear as an agreement between TerraClass and MapBiomas data. In 1991, we have a further aggravating factor: around 75% of the classification area was not classified due to cloud cover. Thus, the accuracy values herein presented for this year correspond to a relatively small subset of the study area. A rigorous accuracy analysis to fully assess the presented classifications, however, would demand independent and experienced image interpreters, the standardization of the datasets and interpretation keys for analysis, and methodological improvements over the good practices suggested by Olofsson et al. (2014), that unfortunately were out of the resources available within this thesis, but should be considered in future projects.

Although the presented accuracy indexes cannot be interpreted as an absolute assessment of accuracy, they present valuable information to allow us to compare the results from each classification approach. As demonstrated by the results, CMAP was able to provide higher values of Overall Accuracy and User's Accuracy for *Secondary Vegetation* for all assessed years, when compared to independently classifying each image with the ML classifier, or the use of post-classification filters. Here, we only compared classification approaches with similar scenarios, i.e.

using ML as the base classifier and the same inputs (Landsat Mosaics, legend, and training samples). Although accurate results can be obtained with the use of different inputs/base classifiers and the use of the post-classification filters (MÜLLER et al., 2016; SOUZA et al., 2020), varying the classifier without varying the inputs do not appear to improve results, as observed by preliminary classification tests (REIS et al., 2017a). In those tests, we classified a Landsat5/TM image within our study area, with varied legend levels correspondent to those presented in Chapter 4 and the supervised classifiers ML, Support Vector Machine (SVM), J48 Decision Tree, and k-Nearest Neighbor (k-NN). This is a robust experiment, in which we varied the collection of labeled samples and tuning of classifiers to properly assess the differences between the classifiers. This analysis showed small differences in the Overall Accuracy from the best results for each classifier, with ML and SVM being the classifiers less affected by variations in the labeled samples. Therefore, we believe that changing the base classifier would only provide significantly improved classification results if it allowed for the inclusion of different data sets as input, which in turn will only be feasible with the collection of an improved set of reference data, and for more recent years.

Furthermore, we used a very simple way to calculate the *a priori* probability of each trajectory, which effectively only prevents the classification of invalid transitions. It is possible that weighting the transitions differently would allow us to limit deforestation events, which in turn could help to avoid the classification of *Secondary Vegetation* in areas not completely deforested. In this sense, the use of more detailed legends could also be beneficial. However, CMAP has a crucial limitation in its present form, which is the need to exhaustively calculate all trajectories derived from all possible combinations of classes in all analyzed years. Possible solutions for this problem involve either the decreasing of the analyzed years, for example by fixing the results on given years to allow the classifications of each temporal subset to run out of sync, or by the use of genetic algorithms. The theoretical basis for the first approach is currently in development.

6.5 Chapter conclusions

We used three approaches of trajectory classification in this study: 1) using the Maximum Likelihood (ML) classifier to classify each Landsat Mosaic independently and then stacking these classifications to create the LULC trajectories; 2) using a post-classification filter to reclassify areas of *Forest* and *Secondary Vegetation* within the ML classifications, to eliminate invalid transitions; and 3) using CMAP

to directly classify the Landsat Mosaics from 1984 to 2020 into LULC trajectories.

All approaches were trained using transferred samples from a more detailed legend than the one considered in the final classification result. As expected, independently classifying each time with the ML classifier led to a huge amount of invalid trajectories. Applying the post-classification filter removed invalid transitions, but the errors are cumulative within the classifications. Among the three approaches, CMAP was able to avoid invalid transitions while using the observation information to auto-regulate. This characteristic led to the classifications of LULC trajectories with less noise and more consistent with what we expected for the study area. As such, CMAP provided classifications with higher calculated User's Accuracy for the class of interest and enabled the classification of persistent areas of *Secondary Vegetation*, which will be analyzed in the following chapter.

7 FOREST REGENERATION TRAJECTORIES ANALYSIS

In the present chapter, we aim to identify, quantify, and characterize forest regeneration trajectories related to different land use historic, agrarian systems, and public policies with territorial expression. This analysis was executed based on two different approaches:

a) by the analysis of the spatial distribution/concentration of secondary vegetation along time; and

b) as forest regeneration trajectories, i.e. a LULC trajectory with at least one observed forest regeneration event. We assumed the need for at least two consecutive observations of secondary vegetation in annual LULC trajectories to characterize a regeneration event. This definition excludes ephemeral secondary vegetation in the first stage of development that is soon deforested. We assumed that these are either part of the management practices or noise derived from the classification process, so their computation as regeneration events would artificially inflate the analysis unnecessarily.

The first approach aims to verify if and how the concentration of secondary vegetation, as well as the frequency and location of events of secondary vegetation deforestation/growth, vary in space and time in the study area, and in relation to the deforestation of primary forest. The second approach, based on forest regeneration trajectories, aims to evaluate other regeneration processes' attributes, herein defined as:

a) Number and duration of regeneration cycles: a regeneration cycle begins with a deforestation event, of primary forest or secondary vegetation, pass through one regeneration event, and ends with the deforestation of the secondary vegetation. Note that one forest regeneration trajectory, that per definition is as long as the analysis period (1984 - 2020), can encompass more than one forest regeneration cycle [1];

b) Persistence of secondary vegetation: within a regeneration cycle, the persistence of secondary vegetation is defined as the duration of the

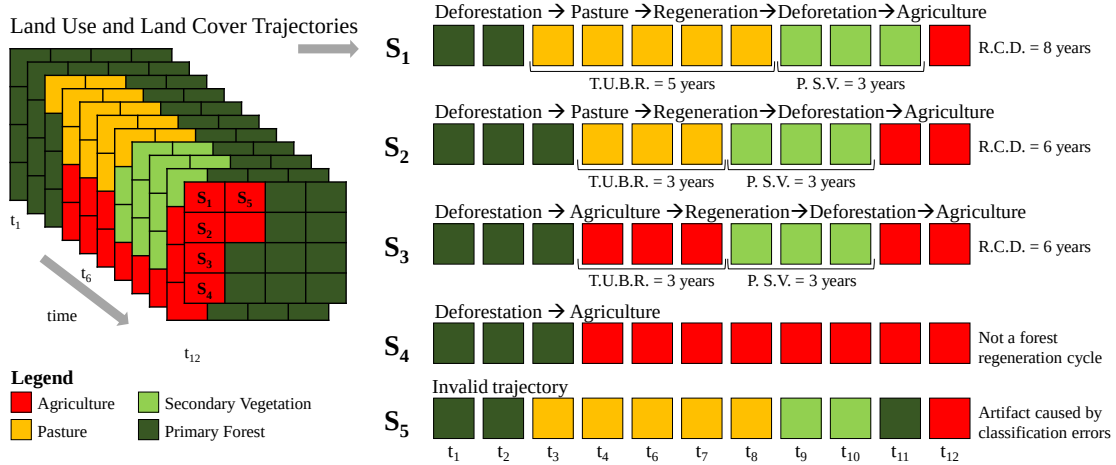
[1]Here we assumed that the time and the type of use before the regeneration event are part of the regeneration cycle since it can impact the regeneration process.

regeneration event. It can also be interpreted as the age of the secondary vegetation being deforested at the end of a regeneration cycle;

c) Time of use before regeneration: duration of LULC sequence, within a regeneration cycle, detected before the regeneration event.

Some examples of LULC trajectories are exemplified in Figure 7.1. In this figure, three LULC trajectories characterize forest regeneration trajectories (S1, S2, and S3). The three chosen attributes are illustrated for these trajectories. A LULC trajectory that does not characterize a forest regeneration trajectory (S4) and an example of an invalid trajectory (S5) were also illustrated. Note that some exemplified trajectories present the same sequence of events with different durations, and therefore, different values for each attribute (S1 and S2) whereas others present different sequences of events, which may impact the adopted typology and interpretation of results, but identical attribute values (S2 and S3). This is a known problem within this type of study (OVIEDO; DOBLAS, 2022) and may result in homogenized analyses.

Figure 7.1 - Examples of Land Use and Land Cover trajectories.



In which: t_y = year of observation; T.U.B.R. = time of use before regeneration; P.S.V. = persistence of secondary vegetation; R.C.D. = duration of the forest regeneration cycle. One forest regeneration cycle begins with a deforestation event (of primary forest or secondary vegetation) and ends with the deforestation of the secondary vegetation. Note that one forest regeneration trajectory can encompass one to several forest regeneration cycles.

SOURCE: Reis et al. (2020b)

In this chapter, we aim to identify if differences in forest regeneration attributes can be explained by variations in time of deforestation, agrarian systems, and/or the influence of public policies with territorial expression. To execute this analysis, we further processed the LULC trajectories classified by CMAP and presented in Chapter 6. Adequately extracting forest regeneration trajectories and their attributes relies on the assumption that we have information about the labeled classes in a given spatial unit of analysis at all times of interest within the LULC trajectory. This assumption is necessary to not dismiss or artificially divide forest regeneration events due to missing information about the land cover at given times. As previously discussed, LULC trajectories obtained in Chapter 6 present several pixels with missing information. Therefore, the first step to derive the forest regeneration trajectories was to assign classes to these pixels. This was done with the use of post-processing filters, as described in Section 7.2.1. Following this, we derived the forest regeneration trajectories and extracted the attributes of interest, as illustrated in Section 7.2.2. Results were analyzed regarding their variation in time and space, and also using data sets that indicate differences in public policies with territorial expression, as well as those representing the agrarian systems (Section 7.1). These results are described in Section 7.3 and discussed in Section 7.4. The chapter conclusions are delineated in Section 7.5.

7.1 Materials

The present analysis was conducted over the LULC trajectories classified by CMAP (Chapter 6). This data set contains one annual LULC classification for each year between 1984 and 2020, in three LULC classes of interest, previously defined on Chapter 4:

a) *Forest*: well-structured, climax forests, with no evidence of clear-cut. Alterations in structure due to logging, fire, invasion by exotic species, or other natural/anthropic activities are allowed, as well as forests being managed by silviculture practices employed to favor economically valuable species;

b) *Secondary Vegetation*: any kind of semi-natural vegetation (not planted by humans, but resultant of human activities) that has grown in areas in which the original vegetation was completely removed;

c) *Temporary Cultures*: areas composed mainly of herbaceous cultures, which present some temporal sequencing depending behavior within one year.

This class includes areas of Pasture and Annual Agriculture;

plus areas without information (*Not Observed*) and other masked areas:

a) *Aquatic or regularly flooded*: areas significantly influenced by the presence of water over extensive periods during the year;

b) *Primarily Non-Vegetated/ Non-Forested Area*: areas not originally covered by forests;

c) *Artificial Surface and Associated Areas*: areas that have artificial, usually not vegetated, cover due to human activities.

Other two data sets were used to stratify the analysis, one representing the different agrarian systems in the region, and one delineating the areas under the influence of public policies with territorial expression. These are detailed as follows.

7.1.1 Agrarian systems

Here, an agrarian system is defined as the set of management practices, agrarian structures, and social-economic characteristics of an activity that results in the production/collection of agricultural goods. We selected the LULC typologies map obtained by Souza (2017) to identify areas with different agrarian systems. This map was generated based on the analysis of patterns of land cover mosaics within a regular grid of 8 km× 8 km in a study area within the Pará state, using a refined version of the TerraClass map for year 2012 (first legend) and data mining techniques. The refined version of TerraClass consists of the identification of areas of *Small-Scale Agriculture* within areas classified as *Occupation Mosaics* and *Secondary Vegetation* in the TerraClass data set. This identification was done visually, based on RapidEye Earth Imaging System (REIS)/RapidEye-2 images from 2011, 2012, and 2014. The author classified the land cover patterns into six categories of interest (SOUZA, 2017):

a) *Intensive Annual Crop:* areas used for mechanized large-scale annual agriculture. The land management in these areas includes the use of fertilizers and pesticides;

b) *Large-scale Livestock:* areas used for the creation of livestock. Associated land management practices can involve techniques like pasture rotation. The infrastructure (pen, paddocks, wire fence, bovine trough, and farmhouse) are usually in good condition;

c) *Transitory System:* systems in transition to more intensive uses. Land use is generally oriented to herding livestock or the production of cocoa, banana, cassava, and/or pepper;

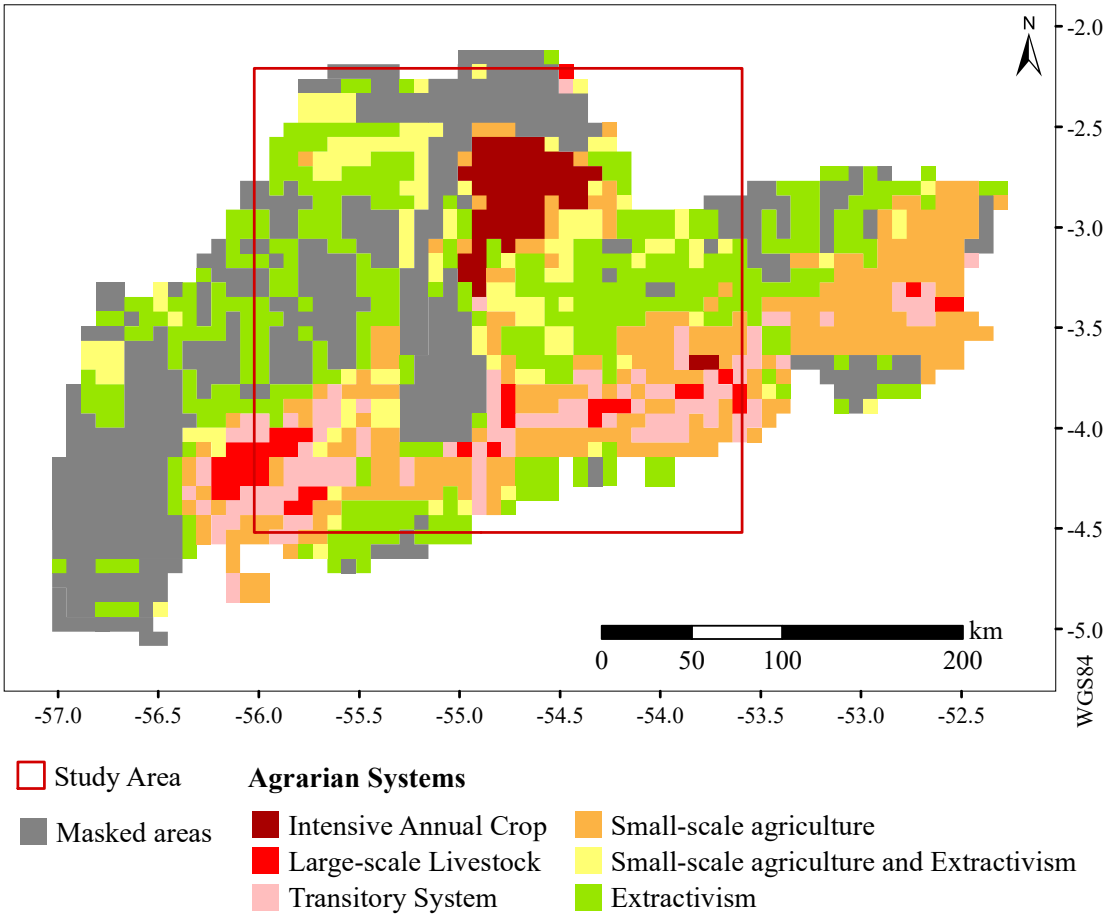
d) *Small-scale agriculture:* land use systems based on the growth of crops such as rice, beans, corn, and cassava. These can be associated to herding livestock and fruits production;

e) *Small-scale agriculture and Extractivism:* areas with the growth of small-scale crops such as rice, beans, corn, cassava, vegetables, and fruit. It is also possible to observe activities based on the non-timber extraction of forest resources and fishing;

f) *Extractivism:* areas characterized by the predominance of activities oriented to non-timber extraction of forest resources, with the occasional existence of small areas of cassava crops for flour production.

This data set is illustrated in Figure 7.2.

Figure 7.2 - Map of agrarian systems, with the identification of the study area limits.

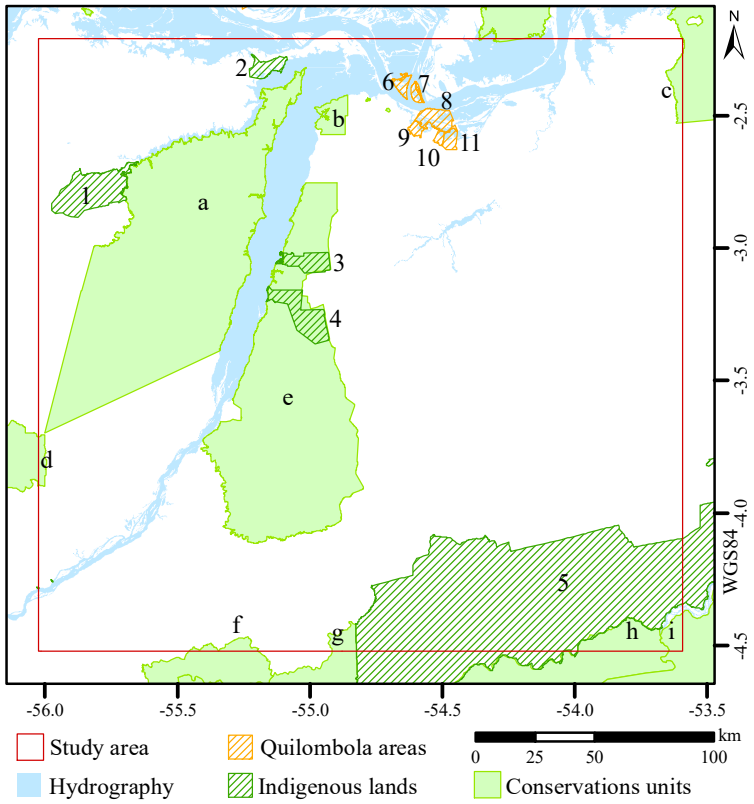


SOURCE: Modified from Souza (2017).

7.1.2 Areas under the influence of different institutions and public policies

We further evaluate differences in forest regeneration trajectories located within areas delimited as the result of public policies and under the influence of different sets of land use rules. These areas correspond to the limits of Conservation Units (MMA, 2020), Indigenous Lands (FUNAI, 2021), Quilombola areas, and INCRA's Settlement Projects (INCRA, s.d.). These limits are detailed in Figures 7.3 and 7.4.

Figure 7.3 - Details about Conservation Units, Indigenous Lands, and Quilombola areas within the study area.

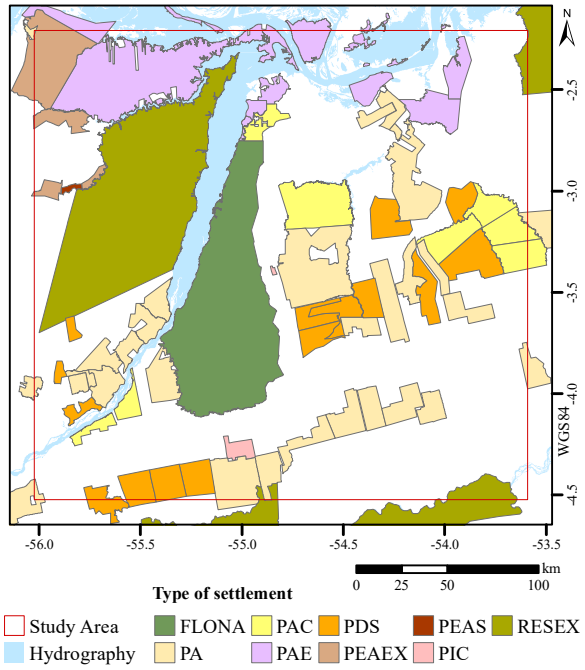


Index	Type	Name	Creation year
a	SU	Tapajós-Arapiuns Extrativist Reserve	1998
b	SU	Alter do Chão Environmental Protection Area	2003
c	SU	Renascer Extrativist Reserve	2009
d	IP	Amazônia National Park	1974
e	SU	Tapajós National Forest	1974
f	SU	Trairão National Forest	2006
g	SU	Riozinho do Anfrísio Extrativist Reserve	2004
h	SU	Rio Iriri Extrativist Reserve	2006
i	IP	Estação Ecológica da Terra do Meio	2005
1	IL	Maró	2008 (Identified)
2	IL	Cobra Grande	2008 (Identified)
3	IL	Bragança-Marituba	2003/2016 (Identified/Declared)
4	IL	Munduruku-Taquara	2003/2016 (Identified/Declared)
5	IL	Cachoeira Seca	1985/2016 (Identified/Approved)
6	Q	Arapemã	2003 (Initial process)
7	Q	Saracura	2003 (Initial process)
8	Q	Maria Valentina	2007 (Initial process)
9	Q	Bom Jardim	2003 (Initial process)
10	Q	Murumuru	2003 (Initial process)
11	Q	Tiningu	2003 (Initial process)

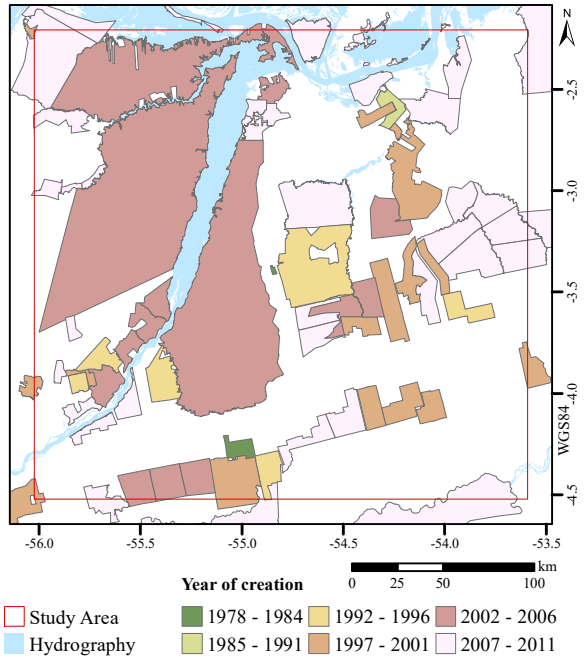
In which IL = Indigenous Lands, and Q = Quilombola areas. Conservation Units are categorized in: SU = Sustainable Use and IP = Integral Protection.

SOURCE: Based on INCRA (s.d.), IBGE (2019), MMA (2020), ISA (2020), FUNAI (2021), and CPISP (2022).

Figure 7.4 - Settlement Projects within the study area, as well as other areas benefited from agrarian reform projects.



(a) Type



(b) Year of creation

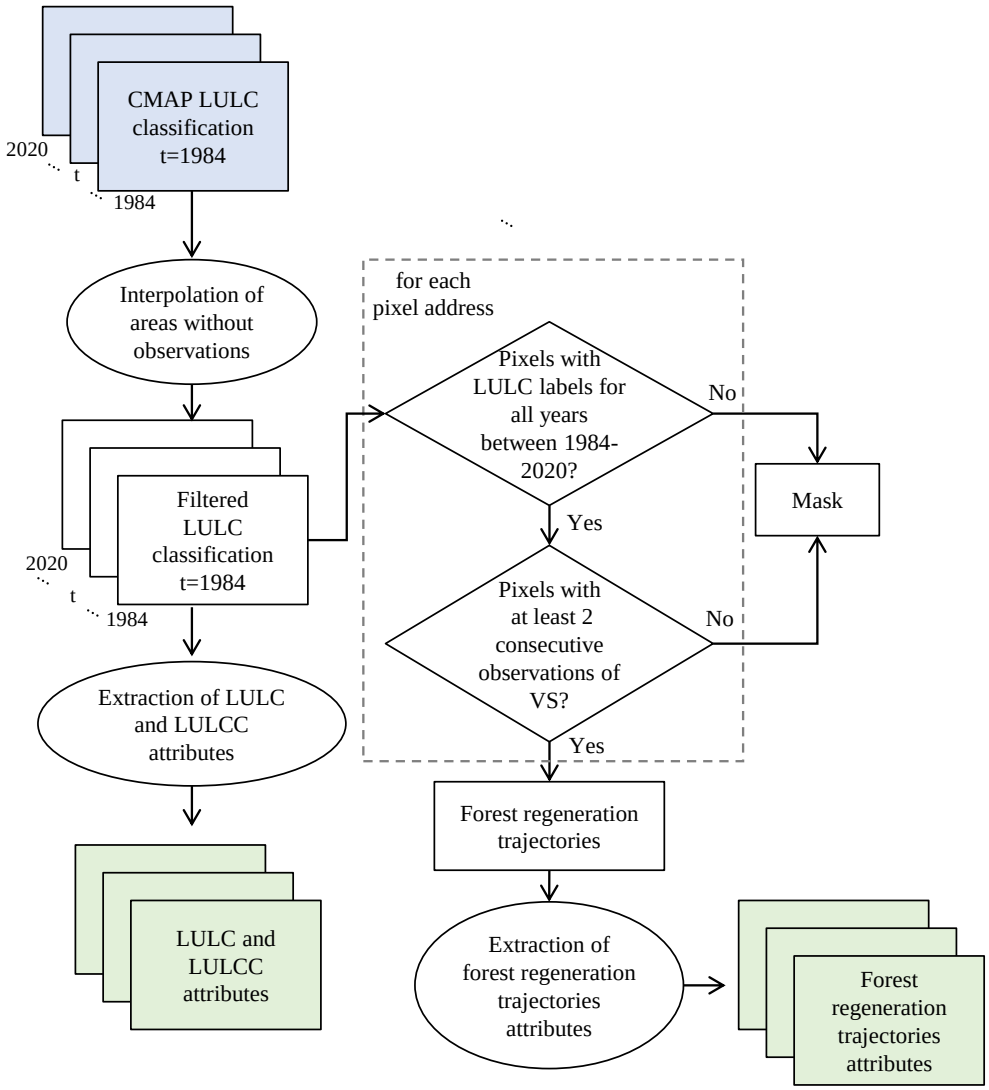
In which PA = Federal Settlement Project, PAC = Joint Settlements Project, PAE = Agro-extractivist Settlement Project, PDS = Sustainable Development Project, PEAEX = Agro-extractivist State Settlement Project, PEAS = Sustainable Settlement State Project, PIC = Integrated Colonization Project, FLONA = National Forest, and RESEX = Extrativist Reserve.

SOURCE: Based on INCRA (s.d.) and IBGE (2019).

7.2 Methods

The methodology is divided into four main steps: 1) the interpolation of areas without observations; 2) the extraction of forested regeneration trajectories and their attributes; 3) the extraction of spatio-temporal patterns of secondary vegetation (LULC and LULC change attributes); and 4) the visualization and analysis of the sets of attributes. These steps are summarized in Figure 7.5 and detailed as follows.

Figure 7.5 - Methodological flowchart for the extraction and analysis of Land Use and Land Cover (LULC), LULC change (LULCC), and forest regeneration trajectories.



Input data are highlighted in light blue. Outputs are highlighted in light green.
SOURCE: The author.

7.2.1 Interpolation of areas without observations

We propose the use of four sequential filters for the interpolation of areas without observations, i.e. those labeled as *Not Observed* in the CMAP classifications. These filters are:

a) Filter 1: be $p_{(lat,\ long,\ t)}$ the pixel at position $(lat,\ long)$ and time t labeled as *Not Observed*. This pixel will be relabeled as *Forest* if any $p_{(lat,\ long,\ t_0)}$ with the time $t_0 >$ time t is labeled as *Forest*. This filter is based on the classification premiss of CMAP in which transitions from any class to *Forest* are invalid;

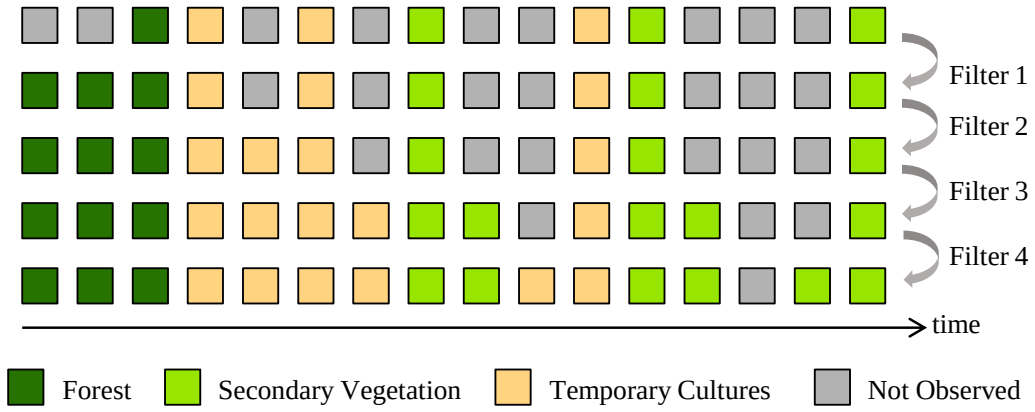
b) Filter 2: relabel the pixel $p_{(lat,\ long,\ t)}$ as a given class ω if $p_{(lat,\ long,\ t-1)}$ and $p_{(lat,\ long,\ t+1)}$ are also labeled as ω;

c) Filter 3: relabel the pixel $p_{(lat,\ long,\ t)}$ as the class of $p_{(lat,\ long,\ t-1)}$. The use of this filter turns the previous one redundant. We decided to use both filters, in order to assess the necessity of interpolating between different classes;

d) Filter 4: relabel the pixel $p_{(lat,\ long,\ t)}$ as the class of $p_{(lat,\ long,\ t+1)}$. As this filter is used after Filter 3, it will only affect pixels originally with two or more consecutive times without observations.

Filters 2 to 4 are relatively conservatory filters, that assume that no changes have occurred in LULC when compared to the nearest observation in time. One example of the effects of each filter over a LULC trajectory is exemplified in Figure 7.6. We applied each filter sequentially over the LULC trajectories and evaluated their impacts based on the remaining pixels containing times without observation. This analysis was conducted by calculating the maximum sequence of times without observations for each pixel before and after the application of each filter. Pixels that remain without observations after the application of the four filters are removed from the analysis.

Figure 7.6 - Effects of each filter for times without information (*Not Observed* class) within the LULC trajectories.



SOURCE: The author.

7.2.2 Extraction of forest regeneration trajectories and attributes

First, we delimit areas with regenerations events, herein defined as areas with two consecutive observations of the class *Secondary Vegetation*. This step typifies the areas of *Secondary Vegetation* as 'a single observation' and 'multiple observations' (two or more). LULC trajectories with at least one regeneration event were flagged as forest regeneration trajectories. From these trajectories and considering the pixel as the unit of analysis, we derived the following attributes:

a) *Age of Secondary Vegetation*: we counted the number of consecutive observations of *Secondary Vegetation* for pixels labeled as this class in a given year and the previous ones. Single observations of *Secondary Vegetation* receive an age equal to one year. This attribute computes observations of *Secondary Vegetation* in all years of the analysis but we cannot extrapolate the values. This means, for instance, that for the year 1985, an *Age of Secondary Vegetation* equal to two means that the secondary vegetation is two years old or older. Similarly, a LULC trajectory with all observations labeled as secondary vegetation will present age values interpreted as 37 years old or older in 2020. This is the only attribute that accounts for possible LULC changes before the time of analysis;

b) *Number of Regeneration Cycles*: number of sets of consecutive occurrences of *Secondary Vegetation with multiple observations*;

c) *Persistence of Secondary Vegetation*: number of elements in each set of consecutive occurrences of *Secondary Vegetation with multiple observations*. As each forest regeneration trajectory can present multiple values of this attribute, results were registered as the mean, median, maximum, and minimum values, as well as the standard deviation of values. We only computed cases in which we observed both the start and the end of the regeneration event. In other words, regeneration events with the first observation of *Secondary Vegetation* in 1984, or the last one in 2020, were not accounted for in the results. This process was done to avoid computing underestimated values. Regeneration processes starting before 1985 or ending after 2020 will not be accounted for within this attribute and may be visualized within *Age of Secondary Vegetation*;

d) *Time of use before Regeneration*: number of consecutive occurrences of *Temporary Cultures*, interspersed or not by *Secondary Vegetation with a single observation*, observed before the first occurrence of *Secondary vegetation with multiple observations* in a regeneration event. This attribute is set to zero if the last observation before a regeneration event is of class *Forest*. Because of the possibility of multiple values within a forest regeneration trajectory, results for this attribute were also presented as the mean, median, maximum, minimum, and standard deviation of values. Similarly to the previous attribute, we did not compute use events with the first or last observation of *Secondary Vegetation with a single observation*/*Temporary Cultures* in 1984 and/or 2020.

7.2.3 Extraction of other LULC and LULC changes attributes

We extracted one set of attributes related to the LULC information of each year, namely the *Relative area of Secondary Vegetation*. This attribute was calculated considering a given polygon as the unit of analysis, and corresponds to the ratio between the area mapped as *Secondary Vegetation* and the area of accumulated deforestation of a given year. The accumulated deforestation is given by the union of areas labeled as *Temporary Cultures* and *Secondary Vegetation* in a given year. We considered four sets of polygons: 1) the limit of the study area, used when presenting a single value, 2) the agrarian systems defined by Souza (2017), 3) the limits of areas influences by public policies with territorial expression, and 4) a regular grid, used to analyze the spatial distribution of data without bias from the other polygons. For this last analysis, it was necessary to first establish the size of

the regular grid. We tested different square cells sizes ranging from 1 km×1 km to 15 km×15 km, in increments of 1 km at each side, and selected the one that presented the best compromise between context, i.e. that captured both areas of *Secondary Vegetation*and surrounding deforested areas without homogenizing either class, and resolution, i.e. in which areas with very different landscape patterns were kept separated. This analysis was made visually and assisted by the data sets depicting the agrarian systems and the limits of areas affected by public policies and institutions (Section 7.1). Examples of the tested cell sizes are illustrated in Figure 7.7. We selected the cells with size 5 km×5 km for further analyses.

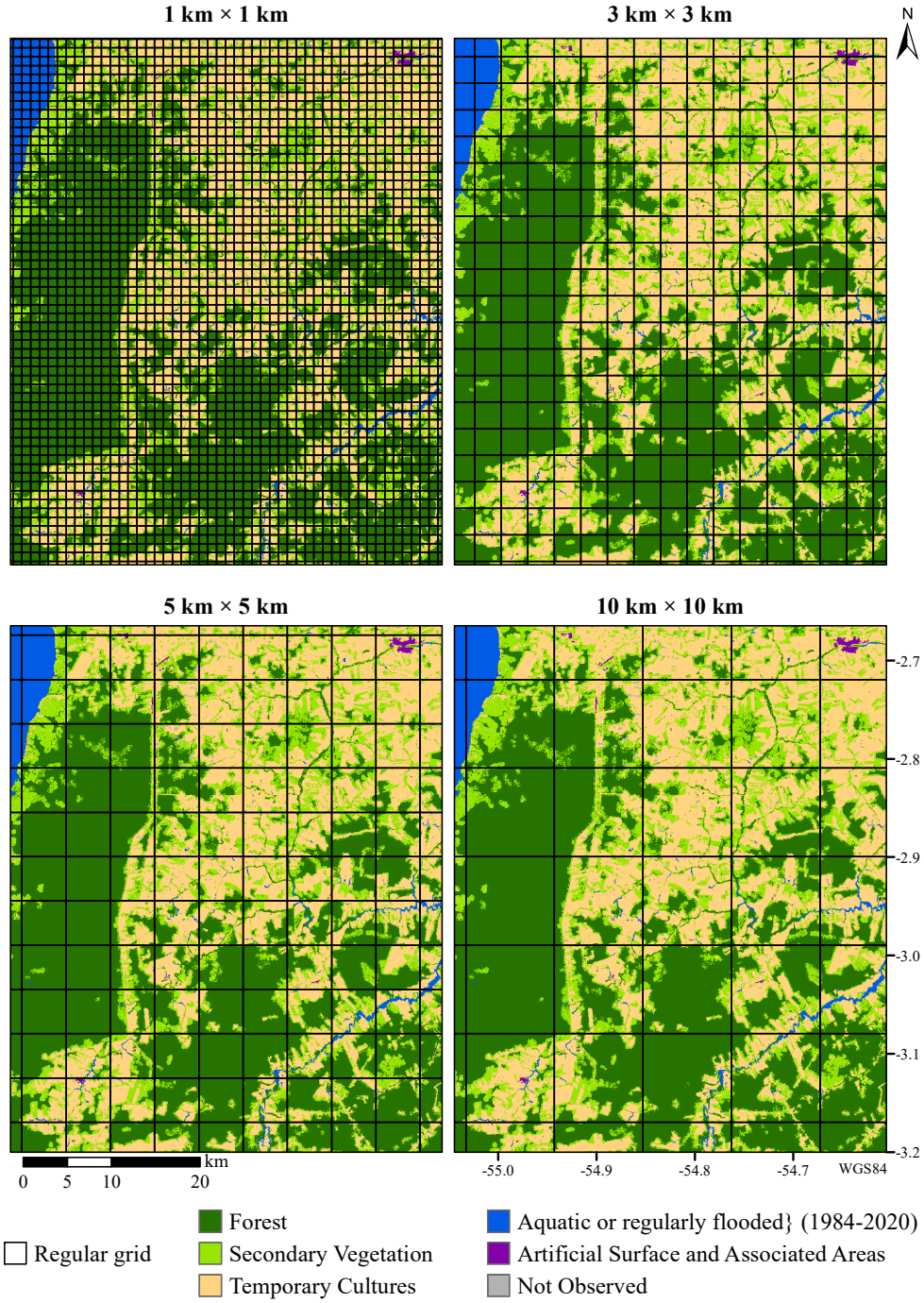
We also derived other three sets of attributes based on the transitions occurring between two consecutive years but characterized considering information derived from the trajectories. These are herein referred to as 'LULCC attributes' and were extracted considering the pixel as the analysis unit. These are:

a) *Deforestation of Forest*: pixels labeled as *Forest* in the previous year and either as *Temporary Cultures* or *Secondary Vegetation* in the analyzed year. These pixels were further characterized by counting the number of observations of *Forest* before the deforestation event;

b) *Deforestation of Secondary Vegetation*: pixels labeled as *Secondary Vegetation* in the previous year and as *Temporary Cultures* in the analyzed year. These pixels were further characterized by the *Age of Secondary Vegetation* before the deforestation event;

c) *Growth of Secondary Vegetation*: pixels labeled as *Secondary Vegetation* in the analyzed year and either *Forest* or *Temporary Cultures* in the previous year. These pixels were further characterized by the *Time of use before Regeneration*.

Figure 7.7 - Variation in cell size in regular grids.



Regular grid with different cell sizes superimposed over the classification of 2020. Note that cell sizes lower than 5 km × 5 km tend to homogenize larger areas of *Temporary Cultures*, whereas higher cell sizes tend to merge areas with very different landscape patterns.

SOURCE: The author.

7.2.4 Visualization and analysis of the attributes

We extracted the median, maximum, minimum, mode, and standard deviation of each attribute to the regular grid. These values were compared and interpreted based on the historic of the study area and known differences in land management. In this analysis, we masked cells with less than 100 observed forest regeneration events. We also calculated the median *Age of Secondary Vegetation* and *Relative area of Secondary Vegetation* for the polygons depicting the agrarian systems and areas influenced by public policies with territorial expression, respectively considering the values calculated for the years 2012 (date of the agrarian systems map) and 2020. These attributes were selected because they can be observed at specific points in time for which the used limits are considered valid.

7.3 Results

The results are presented in this section. Four main aspects were analyzed: 1) the impacts of the post-processing filters in areas without observations (Section 7.3.1); 2) the spatial distribution and temporal dynamics of the LULC and LULCC attributes (Section 7.3.2); 3) spatial patterns of the forest regeneration trajectories attributes (Section 7.3.3); and 4) how a subset of these attributes vary among different agrarian systems and areas under the influence of different public policies and institutions (Section 7.3.4).

7.3.1 Post-processing filters impact

The number of pixels without observations in the LULC trajectory, before and after the application of each filter, is presented in Table 7.1. These quantities are stratified by the value of the maximum sequence of times without observations.

As can be seen, the LULC trajectories presented up to 17 sequential times without observations, although these cases are very rare (2 cases in 26,681,059 classified pixels), with 95.3% of the classified pixels presenting a maximum of three or fewer sequential times without observations. Nonetheless, only 2.4% of the classified pixels presented observations for all the times within the LULC trajectories. The use of the first filter increases the amount of fully observed trajectories up to 61.5%. The use of the remaining filters further increased this number to 97.3% of the classified pixels, which we found a good compromise. The 610,438 pixels with at least one remaining time without observations were masked from the following analysis. This represents 0.79% of the study area (including masked areas) without information, located mainly in the Santarém Plateau and in the Uruará portion of the Transamazon

highway.

Table 7.1 - Maximum number of sequential times without observation in the LULC trajectories. The cumulative percentage is depicted between the parentheses.

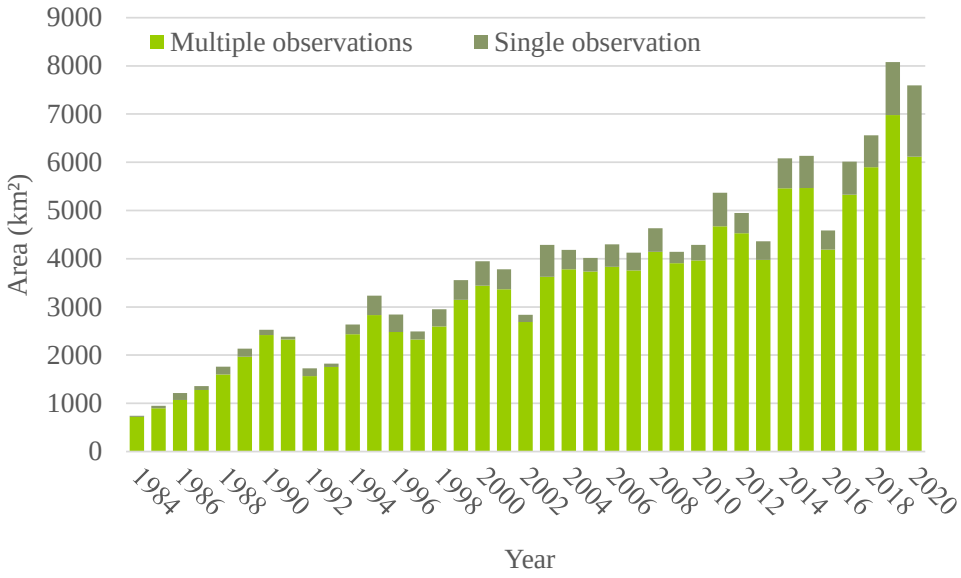
#	Unfiltered		Filter 1		Filter 2		Filter 3		Filter 4	
0	639299	(2.4)	16416046	(61.5)	18859892	(70.7)	22718070	(85.1)	26070621	(97.7)
1	12643739	(49.8)	6601153	(86.3)	4157307	(86.3)	2534144	(94.6)	413234	(99.3)
2	8713528	(82.4)	2527426	(95.7)	2527426	(95.7)	1141526	(98.9)	42744	(99.4)
3	3421210	(95.3)	868104	(99.0)	868104	(99.0)	109711	(99.3)	148604	(99.9)
4	1096714	(99.4)	233371	(99.9)	233371	(99.9)	166750	(99.9)	3577	(99.9)
5	127653	(99.9)	26818	(99.9)	26818	(99.9)	7738	(99.9)	1070	(99.9)
6	24291	(99.9)	5165	(99.9)	5165	(99.9)	985	(99.9)	427	(99.9)
7	6571	(99.9)	982	(99.9)	982	(99.9)	1070	(99.9)	485	(99.9)
8	5355	(99.9)	1260	(99.9)	1260	(99.9)	614	(99.9)	100	(99.9)
9	1518	(99.9)	302	(99.9)	302	(99.9)	173	(99.9)	112	(99.9)
10	710	(99.9)	154	(99.9)	154	(99.9)	81	(99.9)	33	(99.9)
11	217	(99.9)	113	(99.9)	113	(99.9)	144	(99.9)	4	(99.9)
12	152	(99.9)	112	(99.9)	112	(99.9)	1	(99.9)	3	(99.9)
13	6	(99.9)	1	(99.9)	1	(99.9)	4	(99.9)	16	(99.9)
14	17	(99.9)	7	(99.9)	7	(99.9)	6	(99.9)	25	(99.9)
15	55	(99.9)	28	(99.9)	28	(99.9)	38	(99.9)	3	(99.9)
16	22	(99.9)	16	(99.9)	16	(99.9)	3	(99.9)	1	(100.0)
17	2	(100.0)	1	(100.0)	1	(100.0)	1	(100.0)	0	(100.0)

SOURCE: The author.

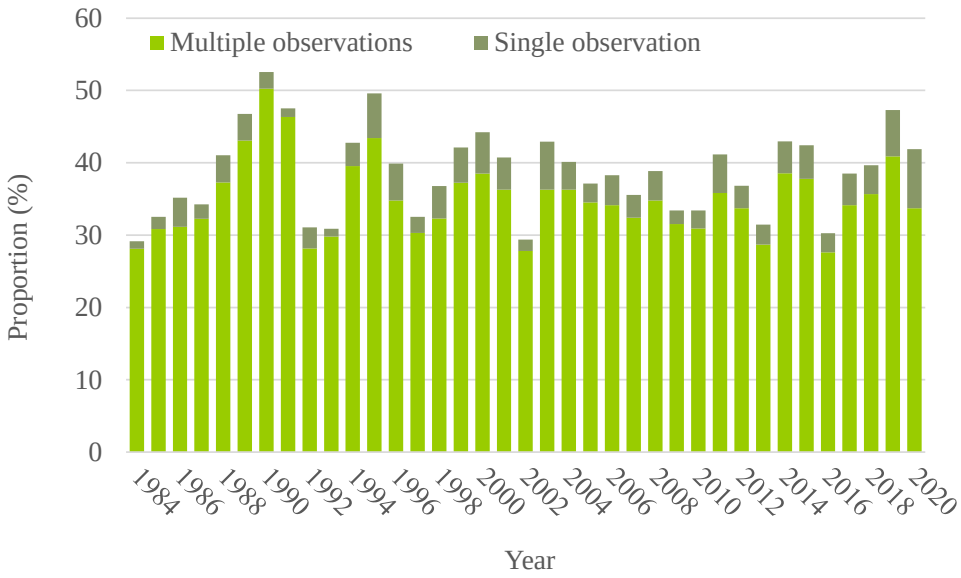
7.3.2 Spatial distribution and temporal dynamics of Secondary Vegetation

Overall, 21.3% of the study area presented at least one observation of *Secondary Vegetation* between 1984 and 2020. At least 84% of the observed *Secondary Vegetation* areas in each year are part of a regeneration event (i.e. two or more consecutive observations of *Secondary Vegetation*). The absolute area (km^2) of *Secondary Vegetation* per year is presented in Figure 7.8.a. Figure 7.8.b presents the *Relative area of Secondary Vegetation* per year.

Figure 7.8 - Area of *Secondary Vegetation* per year.



(a) Absolute



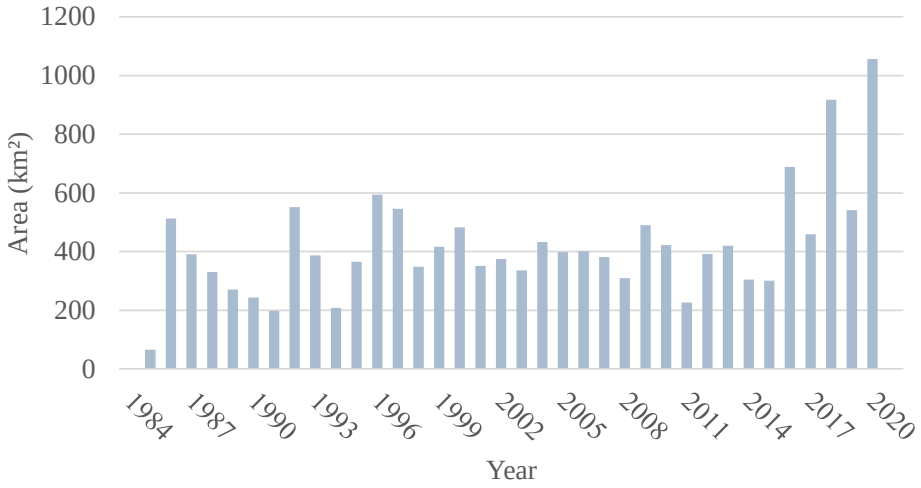
(b) Relative

SOURCE: The author.

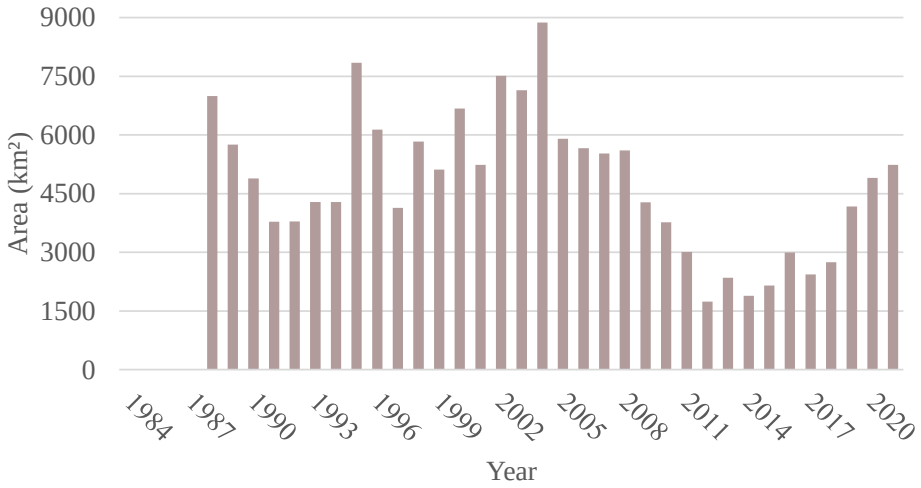
As can be seen in Figure 7.8.a, the area of *Secondary Vegetation* in the study area increases with time. However, this increase in area mostly follows the increase in deforested areas, as observed by the smaller fluctuations in the *Relative area of Secondary Vegetation*. We also see a marked increase in the proportion of *Secondary*

Vegetation with a single observation between 1995-2004 and 2011-2019, with mean proportions equal to 11.0 and 10.6%, respectively, against mean values of 6.4 and 8.4% in 1984-1994 and 2005-2010. The periods between 1995-2004 and 2011-2019 are also characterized by the rise of the deforestation rates of *Forest* in the Pará state, as illustrated in Figure 7.9.

Figure 7.9 - Deforestation of *Forest* per year in the Study Area and the Pará state.



(a) Study Area



(b) Pará

Pará deforestation rates were obtained from INPE (2022).

SOURCE: The author.

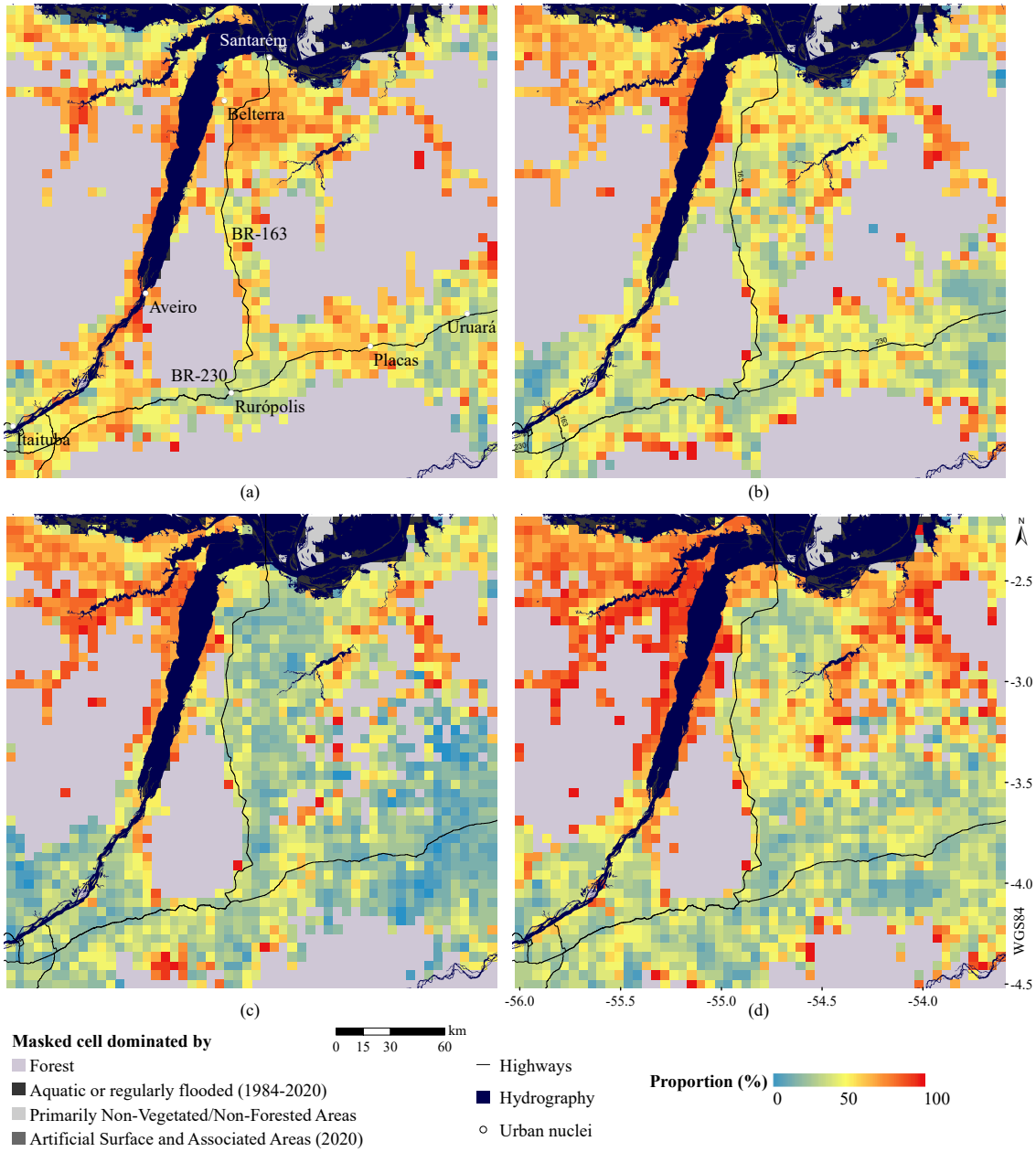
The proportion of *Secondary Vegetation* within deforested areas has been relatively constant over the years, with noticeable increases in 1990, 1995, and 2019. However, this proportion does not present a constant spatial pattern over the years, as illustrated by Figure 7.10 for the years 1990, 2000, 2010, and 2020.

In 1990, the higher values of *Relative area of Secondary Vegetation* are located in the Santarém Plateau and in the deforested stripes near the rivers. Newly deforested areas near the rivers or distant from the highways also tend to present high values of this attribute in the following years. In contrast, the Santarém Plateau presents lower values in 2000 and 2010, with a slightly increase in 2020. Lower values of *Relative area of Secondary Vegetation* are usually seen in newly deforested areas and along the Transamazon highway (BR-230), with a decrease in values near the Itaituba and Placas urban nuclei from 2000.

The areas of deforestation events, both of *Secondary Vegetation* and *Forest* are illustrated in Figure 7.11. The correspondent values normalized by the accumulated deforested area up to the evaluated year are also presented in this figure.

Overall, the deforestation rate of *Secondary Vegetation* increases over the years, following the rise in deforested areas. The *Deforestation of Forest* was higher than the one of *Secondary Vegetation* from 1985 to 1988, with an inverted trend from 1989 to 2020. Deforested *Secondary Vegetation* is mostly younger than five years, which corroborates the field data collected by Affonso et al. (2016) in the region. The deforestation of *Secondary Vegetation* older than two years only surpasses the deforestation of *Forest* for years 2002, 2013, and 2016. The deforestation of *Secondary Vegetation* older than 10 years did not surpass 300 km^2 in the study area, with the highest area computed in 2016 (275 km^2). This is also the year with the highest computed proportion of deforestation of *Secondary Vegetation* older than 10 years when compared to the total of deforested areas of *Secondary Vegetation*, equal to 11.4%.

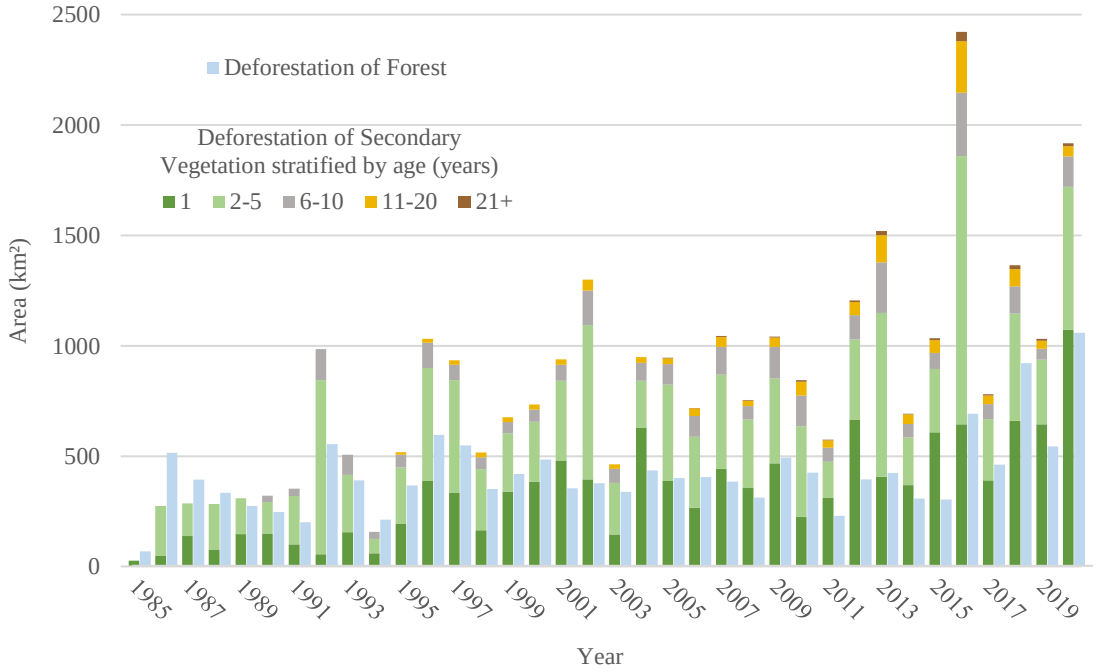
Figure 7.10 - *Relative area of Secondary Vegetation* for years 1990, 2000, 2010, and 2020.



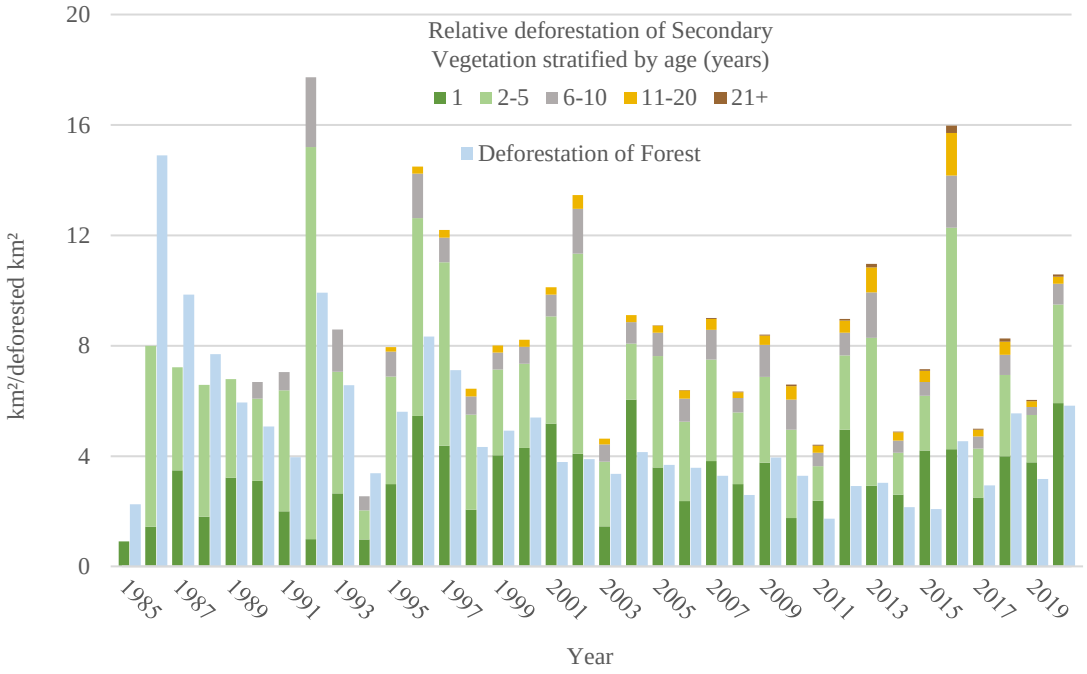
(a) 1990, (b)2000, (c) 2010, (d) 2020. Regular grid of 5 km × 5 km. Cells with 99% or more of coverage corresponding to areas not prone to the growth of *Secondary Vegetation* (i.e. corresponding to the classes *Forest*, *Aquatic or regularly flooded* (1984-2020), *Primarily Non-Vegetated/Non-Forested Areas*, *Artificial Surface and Associated Areas* (2020)) are masked from analysis. Hydrography, highways, and urban nuclei location from IBGE (2019).

SOURCE: The author.

Figure 7.11 - Area of deforestation events detected per year.



(a) Absolute area



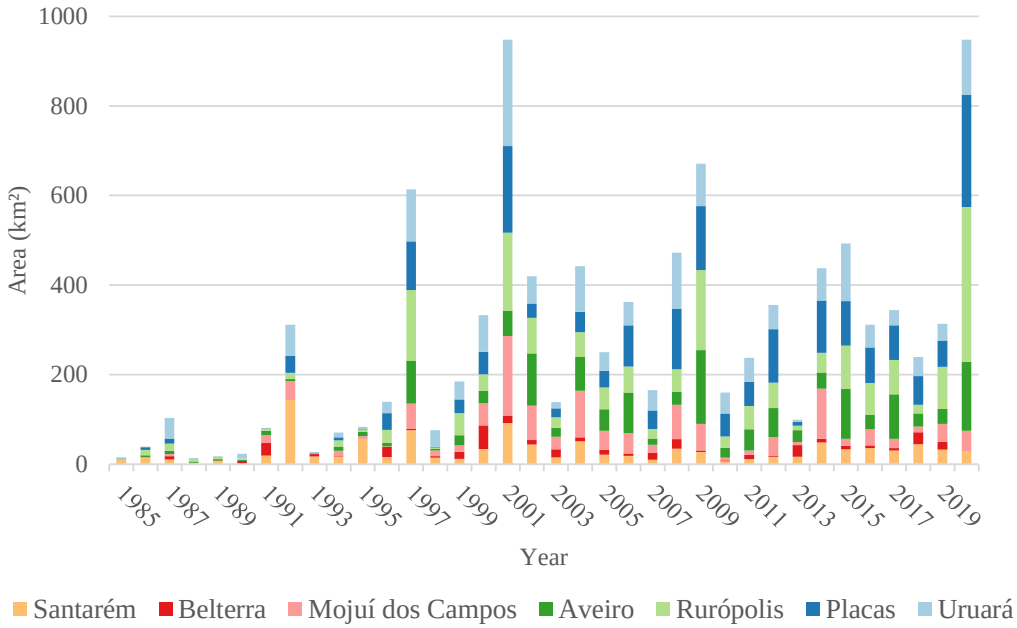
(b) Normalized by the accumulated deforestation

The age category describes the number of observations from 1984. As such, it is undervalued at the higher category in each year.

SOURCE: The author.

There is a relative increase in the deforestation of *Secondary Vegetation* in years 1992, 1996-1997, 2002, and 2016. The increase in *Deforestation of Secondary Vegetation* in 1992 and 1997 coincides with the peaks of burned areas in the region mapped by MapBiomas-Fire, illustrated in Figure 7.12. The spatial distribution of the *Deforestation of Secondary Vegetation* (1-5 years) for 1992, 1996, 2002, and 2016 is illustrated in Figure 7.13.

Figure 7.12 - Burned area per municipality and year.

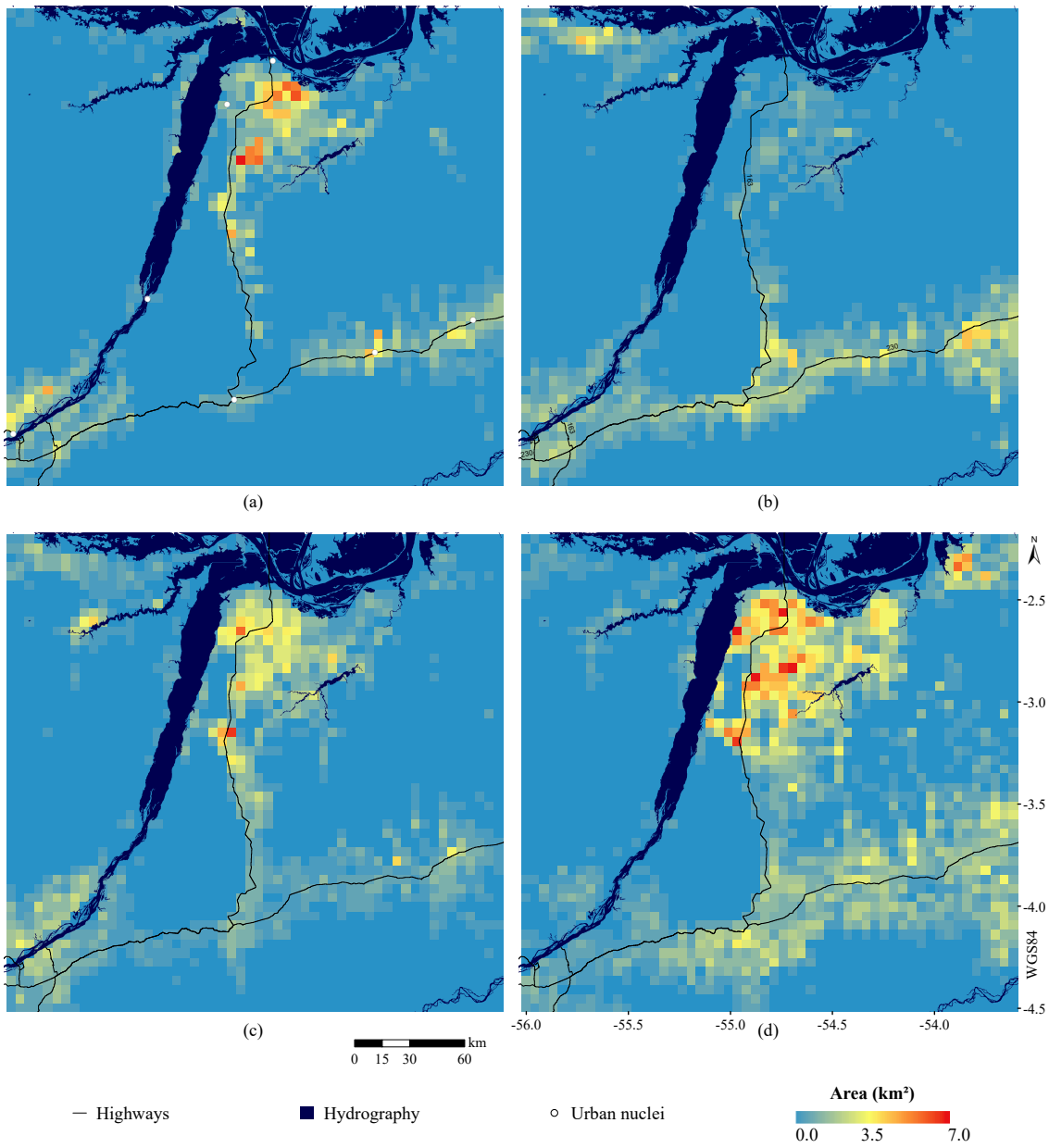


Results covering the cities of Aveiro, Belterra, Placas, Santarém, Rorópolis, and Uruará, all in Pará state.

SOURCE: Modified from MapBiomas (2022c).

In 1992, most burned areas were located in Santarém, whereas most observations of *Deforestation of Secondary Vegetation* are located in Santarém and Mojuí dos Campos. In this year, the *Deforestation of Secondary Vegetation* and *Deforestation of Forest* occurs concentrated in similar areas.

Figure 7.13 - Spatial distribution of the *Deforestation of Secondary Vegetation* mapped in selected years.



(a) 1990; (b) 1996; (c) 2002; (d) 2016.
SOURCE: The author.

In 1997, most burned areas were located in Rurópolis, Placas, and Uruará, which are also areas with higher rates of *Deforestation of Secondary Vegetation* (1-5 years) in 1996. These areas are not so evident in 1997. We notice less concentrated events of *Deforestation of Forest* in the Santarém Plateau in these years, although

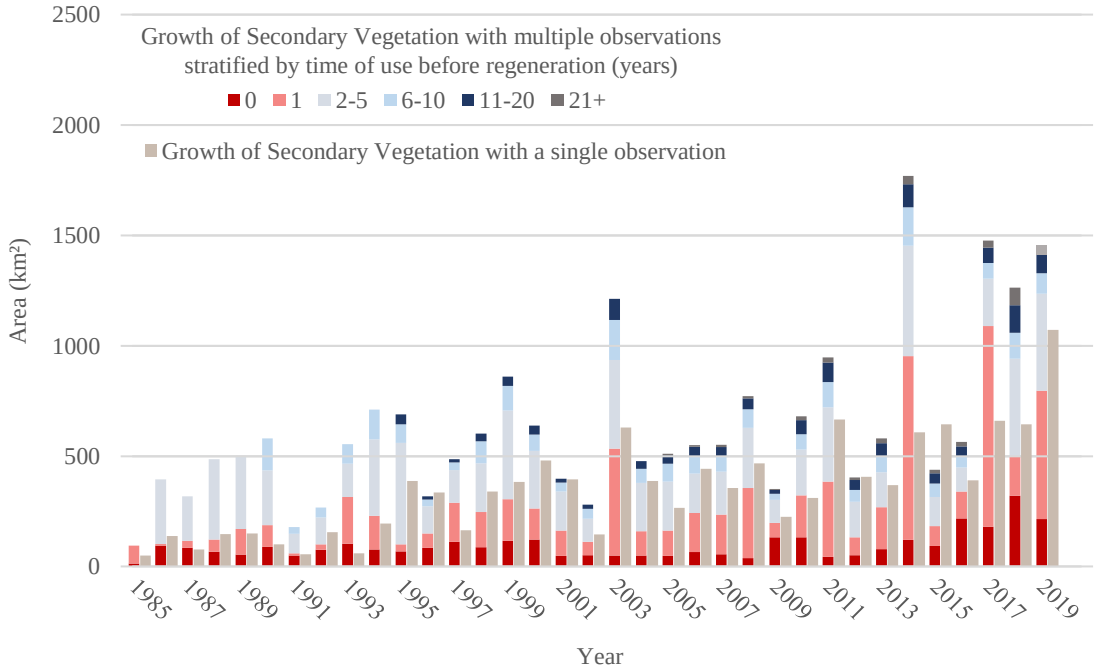
the *Deforestation of Secondary Vegetation* is still present. Also interesting is the concentration of events of *Deforestation of Secondary Vegetation* in the riverine areas along the Amazon river, near the Juruti portion of the study area. This is an area with historically very few registered burned areas (MAPBIOMAS, 2022c) and that presented concomitant events of *Deforestation of Forest*. This last pattern was not observed in the other years with *Deforestation of Secondary Vegetation* peaks.

The high amount of observed events of *Deforestation of Secondary Vegetation* in 2002 can be related to the entry of soy in the Santarém Plateau, which advanced into areas occupied by secondary vegetation, small-scale agriculture, pastures, and also forests (COELHO et al., 2021). In this sense, the focus of deforestation of *Secondary Vegetation* younger than five years occurred mainly in the Santarém Plateau and along the Belterra portion of the Cuiabá-Santarém highway in 2002. The Santarém Plateau also concentrates the observations of *Deforestation of Secondary Vegetation* older than five years. Observations of *Deforestation of Forest*, on the other hand, appear to be scattered along the study area.

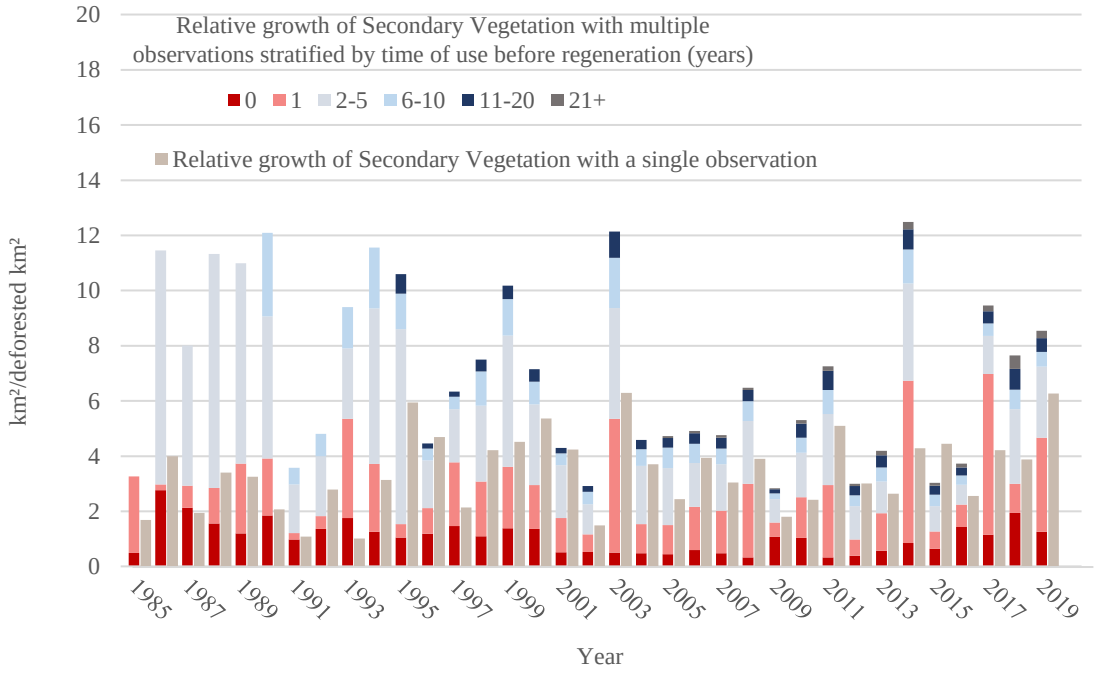
In 2016, we observed areas of *Deforestation of Secondary Vegetation* younger than five years mostly on the Satantarém Plateau and areas along the Transamazon highway, with a noticeable smaller amount of deforestation at the West portion of the study area, both in areas with younger and older *Secondary Vegetation*. The patterns of deforestation of older *Secondary Vegetation* are concentrated in the Santarém Plateau, whereas the deforestation of *Forest* is located in more distant areas along the Transamazon highway.

Similarly to the deforestation trend, the areas of *Growth of Secondary Vegetation* also increase over the years. The majority of the growths occur in areas being used for less than six years, as illustrated in Figure 7.11. This figure illustrates both the absolute as well as the normalized areas of *Growth of Secondary Vegetation*. The spatial pattern of this attribute in selected years is shown in Figure 7.15. We did not observe spatial differences in the pattern of *Growth of Secondary Vegetation* with a single observation or multiple observations, so separate maps were omitted from this document.

Figure 7.14 - Area of events of *Secondary Vegetation* growth detected per year .



(a) Absolute area

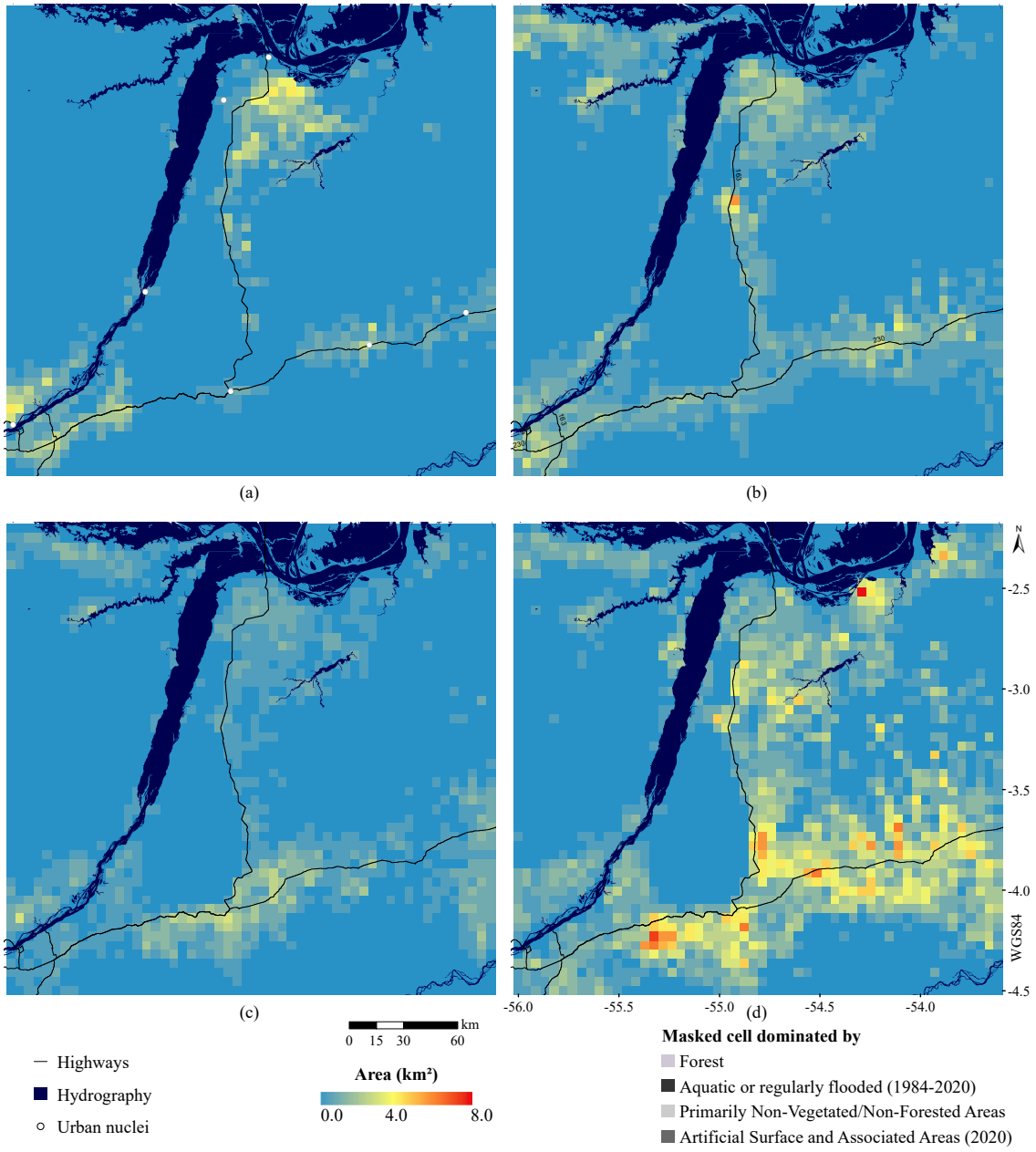


(b) Normalized by the accumulated deforestation

The time of use category describes the number of observations from 1984. As such, it is undervalued at the higher category for each year.

SOURCE: The author.

Figure 7.15 - Spatial distribution growth of *Secondary Vegetation* for years 1990, 2000, 2010, and 2019.



(a) 1990; (b) 2000; (c) 2010; (d) 2019.

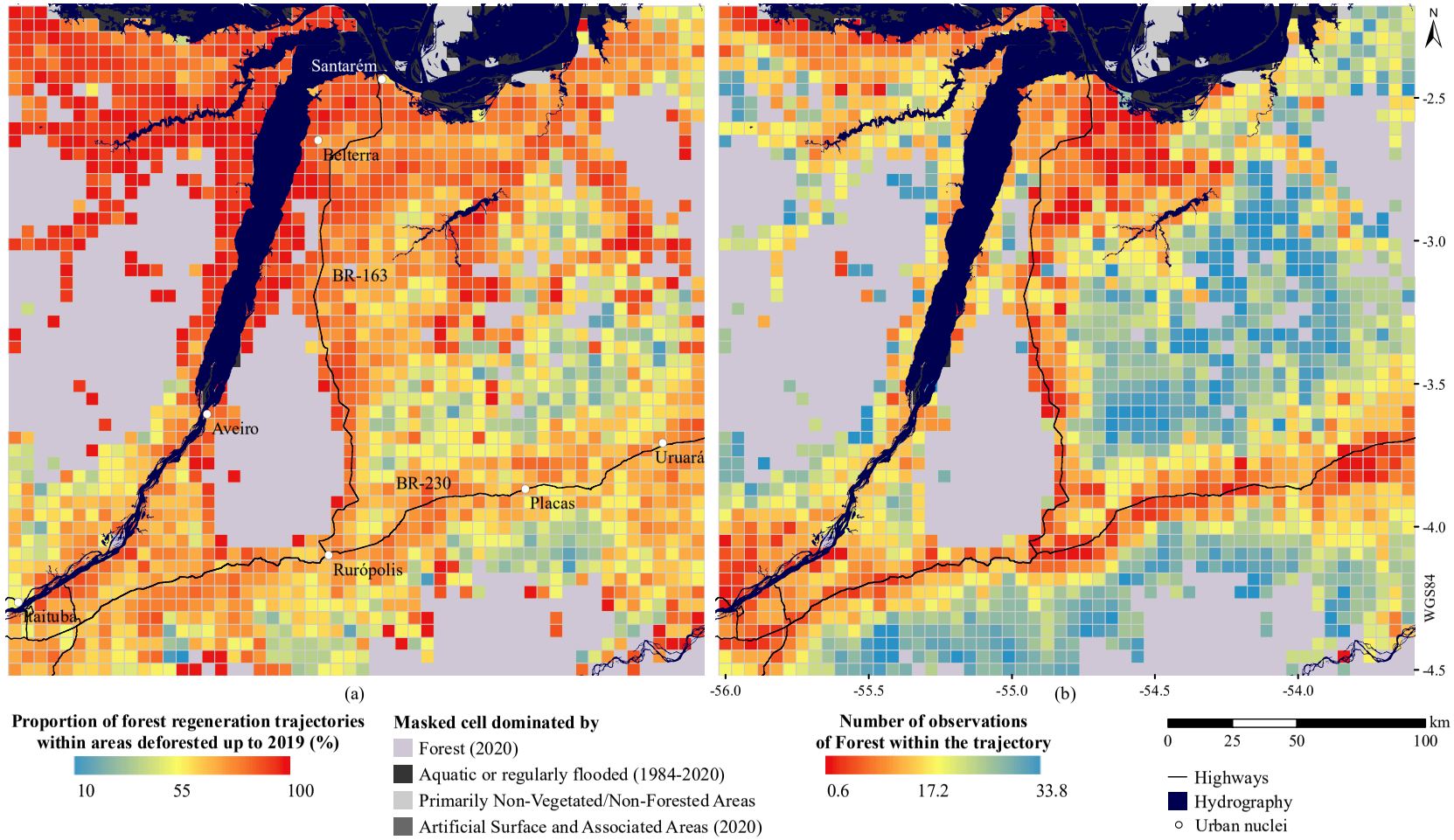
SOURCE: The author.

Up to 320.5 km^2/year (2018) of *Growth of Secondary Vegetation* occurred in areas without previous agricultural use (Figure 7.14.a). Costa (2016), for instance, suggests that such type of growth could be related to the land market. In the study area, we observed higher proportions of *Growth of Secondary Vegetation* in areas not previously used before the year 2000, and once again after 2016. Furthermore, we highlight the decrease in the normalized *Growth of Secondary Vegetation* between 2004 and 2013 (Figure 7.14.b). This is a period of reduced/stable *Deforestation of Forest* in the Pará state/study area. In this sense, it is possible that the actions derived from the PPCDAm, instead of affecting the deforestation rates in the study area, as happened with the Amazon biome and the Pará state, had instead inhibited the abandonment of previously deforested areas. This is also a period in which many areas within the Santarém Plateau have been converted from small-scale agriculture to large-scale agriculture (mainly soy) which also explains the decrease in the area of *Growth of Secondary Vegetation*.

7.3.3 Spatial patterns of the forest regeneration trajectories attributes

By definition, forest regeneration events can only occur in areas deforested up to 2019 in this study. After removing the areas with at least one missing observation between 1984 and 2020, deforested areas in 2019 added up to 16,523 km^2, from which ∼11,697 km^2 (70.8%) presented at least one regeneration event, and therefore, correspond to forest regeneration trajectories. The spatial distribution of these trajectories are illustrated in Figure 7.16, as the proportion of pixels labeled as forest regeneration trajectories from the total number of deforested pixels up to 2019 in each 5 km × 5 km cell.

Figure 7.16 - Distribution of forest regeneration trajectories and time of first deforestation.



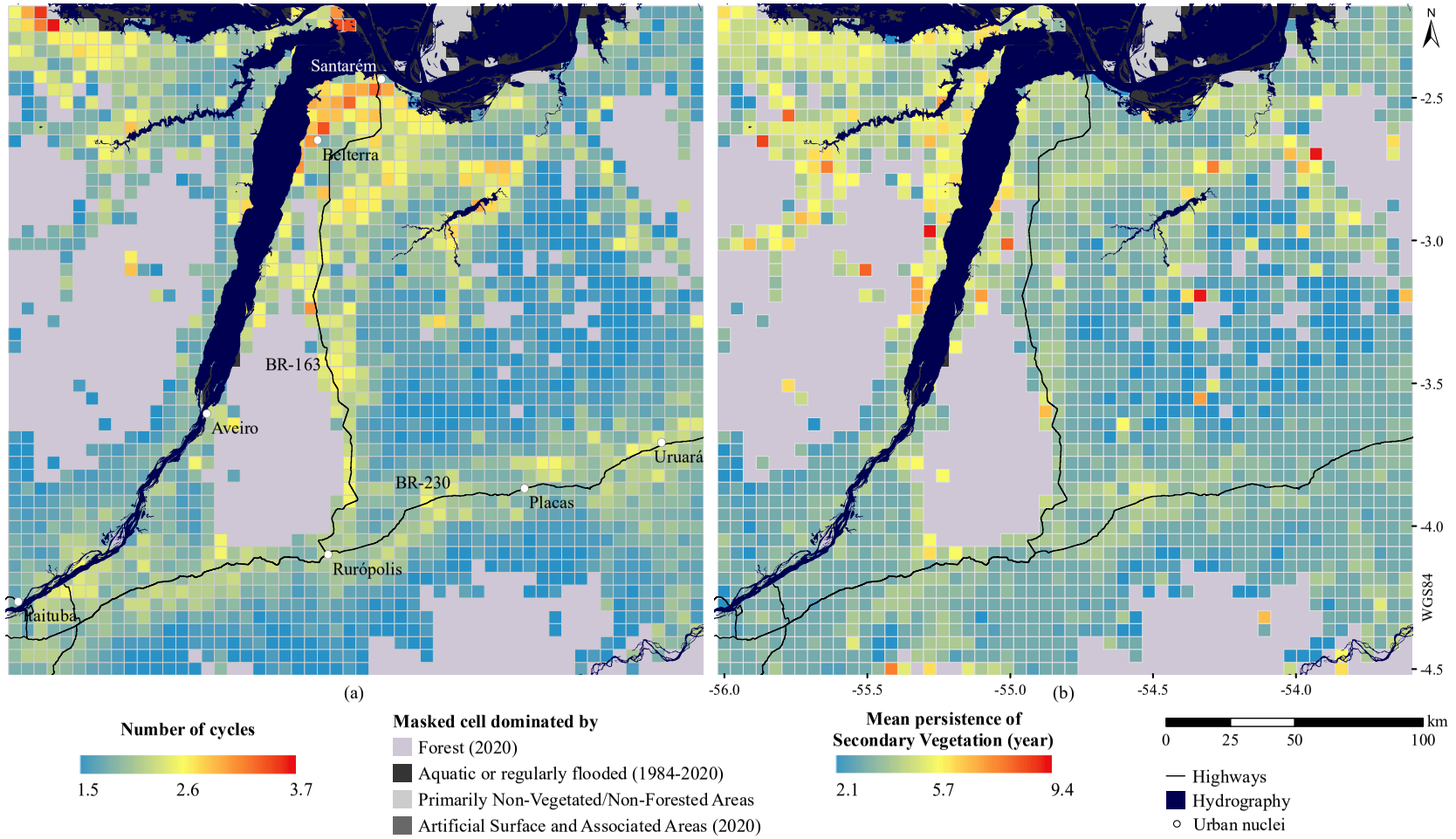
(a) Proportion of forest regeneration trajectories in relation to areas deforested up to 2019 and (b) median number of observations of *Forest* between 1984-2020. In (b), a higher number of observations indicate more recent first-deforestation events. Hydrography, highways, and urban nuclei location from IBGE (2019).

SOURCE: The author.

The highest proportion of forest regeneration trajectories was observed near the rivers, the Santarém Plateau, and areas north of the Tapajós-Arapiuns Extrativist Reserve, as well as areas near the *Forest* matrix. The first two are areas historically used for small-scale agriculture. They are also areas of older occupation, as denoted by Figure 7.16.b. In this figure, the number of observations of the class *Forest* within the trajectory is used as a proxy for the deforestation time. As such, a lower number of observations indicates older deforestation events. The other two areas can be considered as areas of relatively limited access that are also the nearest to the unperturbed *Forest* matrix, which can foster abandonment rates and availability of seeds (JAKOVAC et al., 2021). In contrast, areas of older occupation along the Transamazon highway, usually occupied by pasture since the deforestation of *Forest*, present a relatively lower proportion of forest regeneration trajectories. As expected, the newly deforested areas at the East of the study area present the lowest proportion of forest regeneration trajectories.

We identified a total of 18,248,581 pixels depicting regeneration events in the study area, with duration ranging from two (the majority of events - 37.7%) to 35 years (very rare - 0.0005%). Around 75% of the regeneration events persisted for up to five years. Events that persisted for 20 years or more account for around 1% of the total. The *Number of Regeneration Cycles* varied from one (46.2% of the forest regeneration trajectories) to 10 (only four registered trajectories) in the study area and during the period from 1984 to 2020. Around 96% of the forest regeneration trajectories presented four regeneration cycles or less. These attributes are illustrated in Figure 7.17.

Figure 7.17 - Number of cycles and mean Persistence of *Secondary Vegetation* within the forest regeneration trajectories.



(a) Number of Regeneration Cycles and (b) mean Persistence of Secondary Vegetation (considering only complete observations). The grid cells are filled with the median value. Hydrography, highways, and urban nuclei location from IBGE (2019).

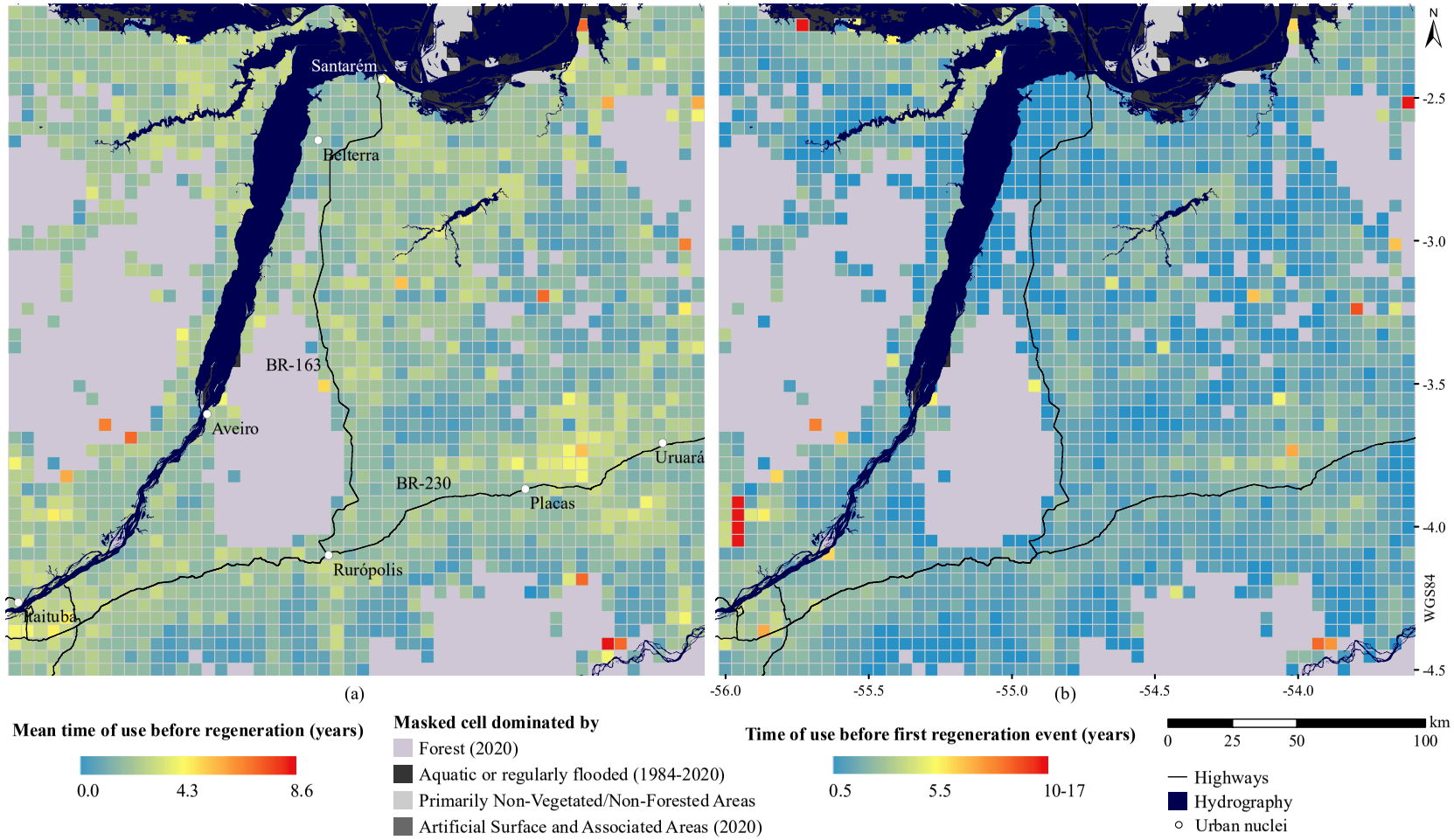
SOURCE: The author.

Interestingly, areas with higher numbers of forest regeneration cycles also present higher mean *Persistence of Secondary Vegetation*. This is particularly evident in areas north of the Tapajós-Arapiuns Extrativist Reserve, and west of the Cuiabá-Satarém highway, which are respectively areas of Agro-extractivist Settlement Projects and Sustainable Conservation Units [2], with agrarian systems based on small-scale agriculture and/or extractivism. This result can be explained by the role the *Secondary Vegetation* has within Shifting Cultivation practices. In those, cycles of regeneration and deforestation of *Secondary Vegetation* are used to restore soil fertility. In this sense, these areas tend to present multiple cycles of forest regeneration by design, with regular persistence of *Secondary Vegetation*. As expected, there are also high values of *Persistence of Secondary Vegetation* in the Tapajós-Arapiuns Extrativist Reserve and within the Tapajós National Forest, expected due to their status as Conservation Units.

Accordingly, the mean *Time of use before Regeneration* of the trajectories is illustrated in Figure 7.18. Along with this attribute, we also present the *Time of use before Regeneration* calculated after the deforestation of the class Forest. As can be seen in this figure, areas near Santarém, along the Cuiabá-Santarém highway, as well as those near the Curuá-Una river, and the ones in the Jutiti municipality tend to present higher values of *Time of use before Regeneration*.

[2]Conservation Units of Sustainable Use admit the presence of residents in their interior and some agrarian practices.

Figure 7.18 - Time of use before forest regeneration events.



(a) Mean time of use before regeneration and (b) time of use before first regeneration event (considering only trajectories with deforestation of *Forest*). The grid cells are filled with the median value.

SOURCE: The author.

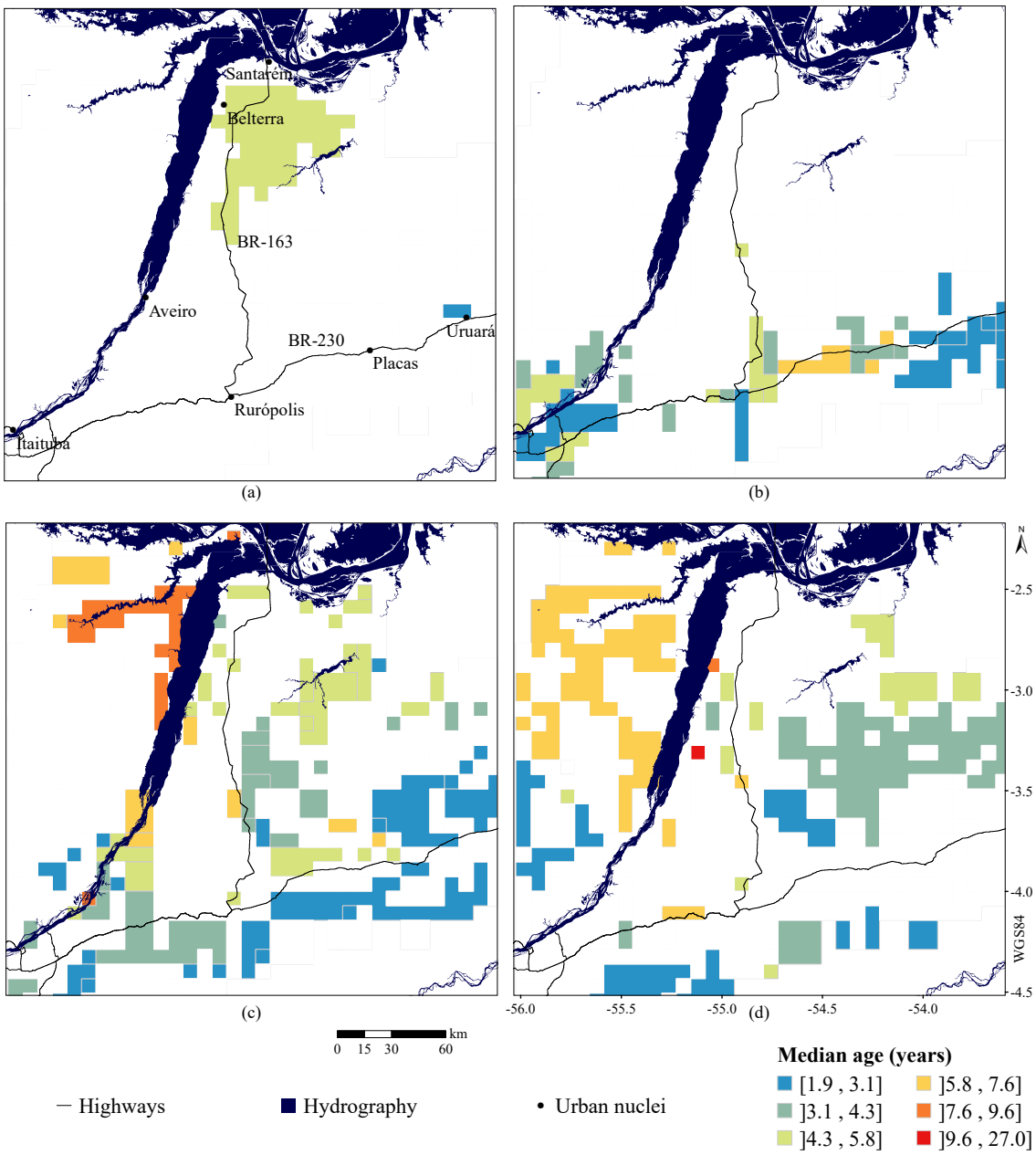
7.3.4 Age variation of *Secondary Vegetation* among different agrarian systems and areas under the influence of different public policies

The median *Age of Secondary Vegetation* in 2012, extracted considering different agrarian systems, is presented in Figure 7.19. The *Relative area of Secondary Vegetation* in the same year and considering the same unity of analysis is illustrated in Figure 7.20. In these figures, areas used with similar practices and with similar results are presented jointly. This is the case for *Large-scale Livestock* and *Transitory Systems*, as well as *Small-scale Agriculture* and *Small-scale Agriculture and Extractivism*.

Overall, the highest values of median *Age of the Secondary Vegetation* were found in areas used for *Extractivism* inside the Tapajós National Forest, and *Small-scale Agriculture and Extractivism* in riverine areas west of the Tapajós river. The area of *Intensive Annual Crop* within the Santarém Plateau presented a median age of 4.7 years (mean age of 6.1 years, with a standard deviation of 5.4 years), which was expected given the known remains of *Small-scale Agriculture* within the *Intensive Annual Crop* areas in the region (COELHO et al., 2021; PAULA et al., 2022). The results, however, reveal that the *Age of Secondary Vegetation* varies within areas with supposedly the same type of agrarian system. Areas used for *Extractivism*, for instance, present both the lowest and highest values of this attribute for 2012.

In general, the median *Age of Secondary Vegetation* tends to vary differently in relation to the *Relative area of Secondary Vegetation* and the median number of observations of *Forest* before the deforestation event. In areas dominated by *Small-scale Agriculture* we observed that higher median values of *Age of Secondary Vegetation* occur within areas with higher *Relative area of Secondary Vegetation*, possibly meaning that the vegetation is left to regenerate for more time because more areas are being used for Shifting Cultivation. In contrast, *Intensive Livestock* and *Transitory System* tend to present the inverse relationship, which may occur due to *Secondary Vegetation* areas that are soon deforested for pasture maintenance. Furthermore, areas with older deforestation tend to present older *Secondary Vegetation*, with a marked exception of *Extractivism* systems. These are less dependent on converting the *Forest* into other types of land cover, presenting a lower area of *Secondary Vegetation* and different patterns than the other agrarian systems by default.

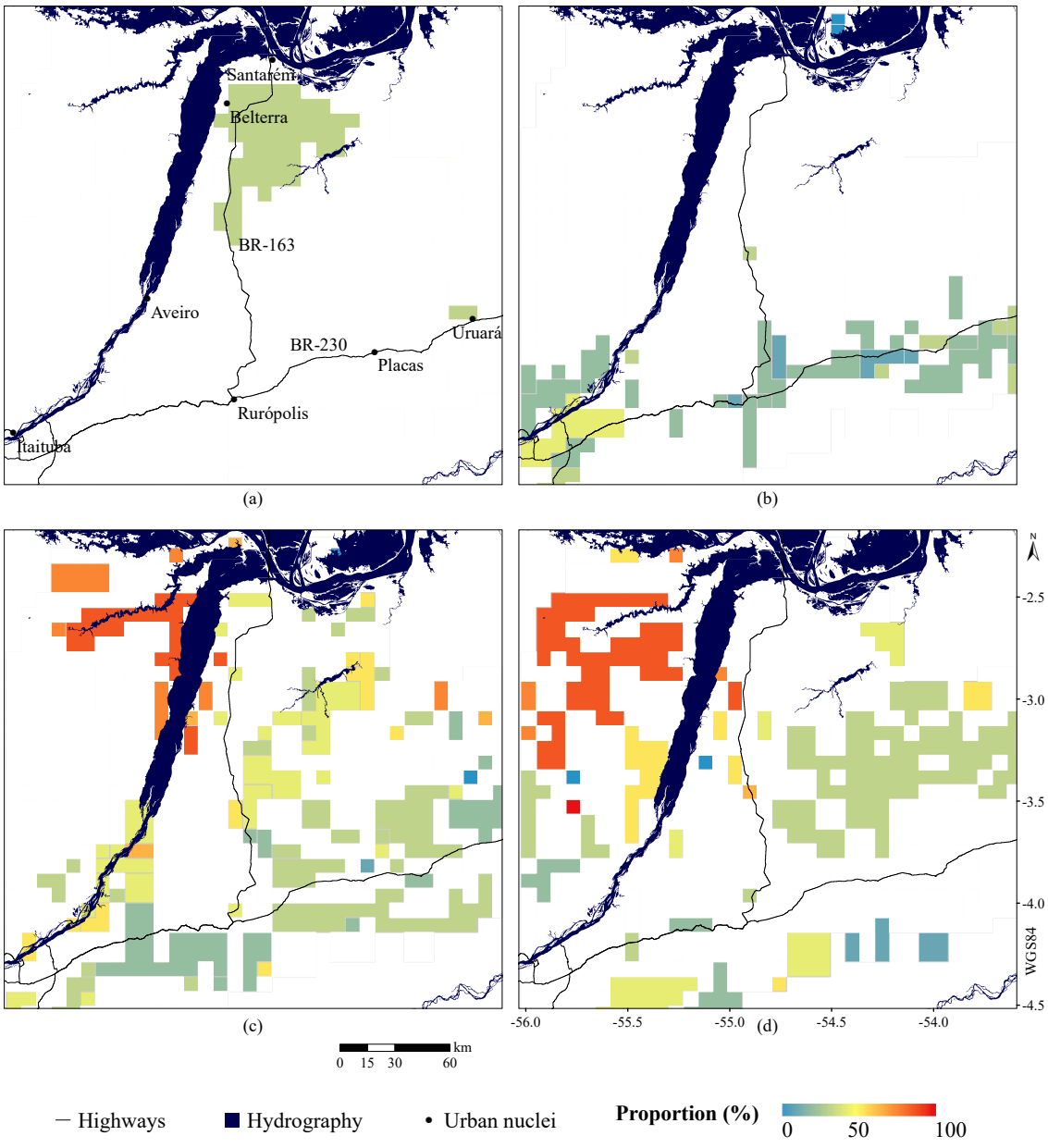
Figure 7.19 - Median Age of Secondary Vegetation in 2012 stratified by types of agrarian systems.



(a) Intensive Annual Crop, (b) Large-scale Livestock+Transitory System, (c) Small-scale Agriculture+Small-scale Agriculture and Extractivism, (d) Extractivism. Hydrography, highways, and urban nuclei location from IBGE (2019). Types of agrarian systems from Souza (2017).

SOURCE: The author.

Figure 7.20 - Relative area of Secondary Vegetation in 2012 stratified by types of agrarian systems.

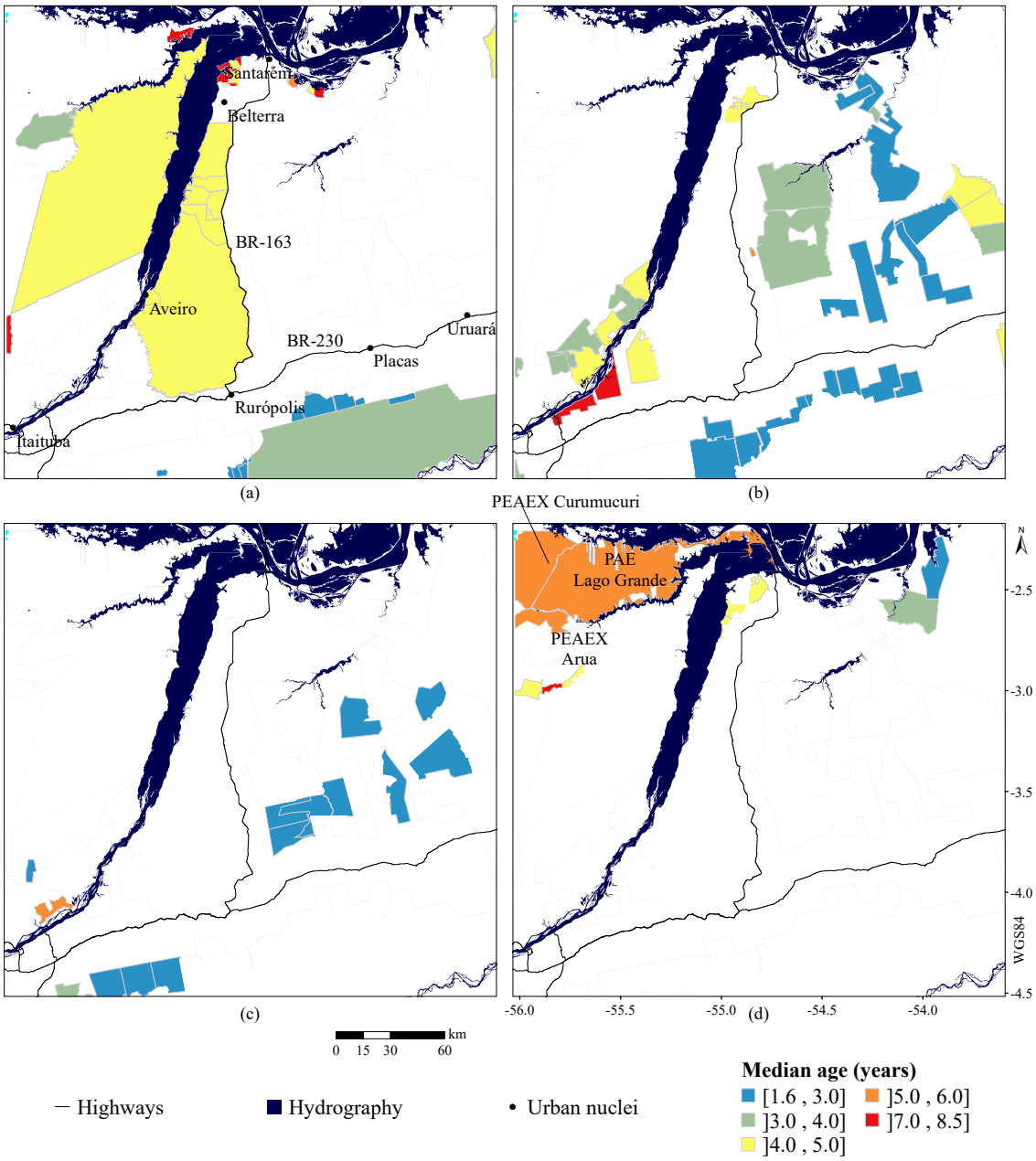


(a) Intensive Annual Crop, (b) Large-scale Livestock+Transitory System, (c) Small-scale Agriculture+Small-scale Agriculture and Extractivism, (d) Extractivism. Hydrography and urban nuclei location from IBGE (2019). Highways from MInfra (2021). Types of agrarian systems from Souza (2017).

SOURCE: The author.

The superposition of limits among Conservation Units, Indigenous Lands, and Settlement Projects are characteristic of the study area. Therefore, for the analysis of areas influenced by public policies with spatial expression, we considered the subsets of areas with the presence of more than one institution as a different polygon. The median *Age of Secondary Vegetation* in 2020, extracted considering these polygons, is presented in Figure 7.21. The *Relative area of Secondary Vegetation* is depicted in Figure 7.22. Areas without *Secondary Vegetation* in 2020, or with 99% or more of the area occupied by areas not prone to the growth of *Secondary Vegetation*, were removed from the analysis. As can be seen in this figure, Settlement Projects along the Transamazon highway and those located in nearly deforested areas tend to present younger *Secondary Vegetation* in 2020. Relatively high values of mean *Age of Secondary Vegetation* were observed in the Conservation Units and Quilombola Areas, as well as Settlement Projects focused on extractivism, such as the Agro-extractivist Settlement Projects (PAEs) Lago Grande and Salé and the Agro-extractivist State Settlement Project (PEAEX) Curumucuri. Areas within Federal Settlement Projects (PAs), Joint Settlements Projects (PACs), and Integrated Colonization Projects (PICs) presented more varied values.

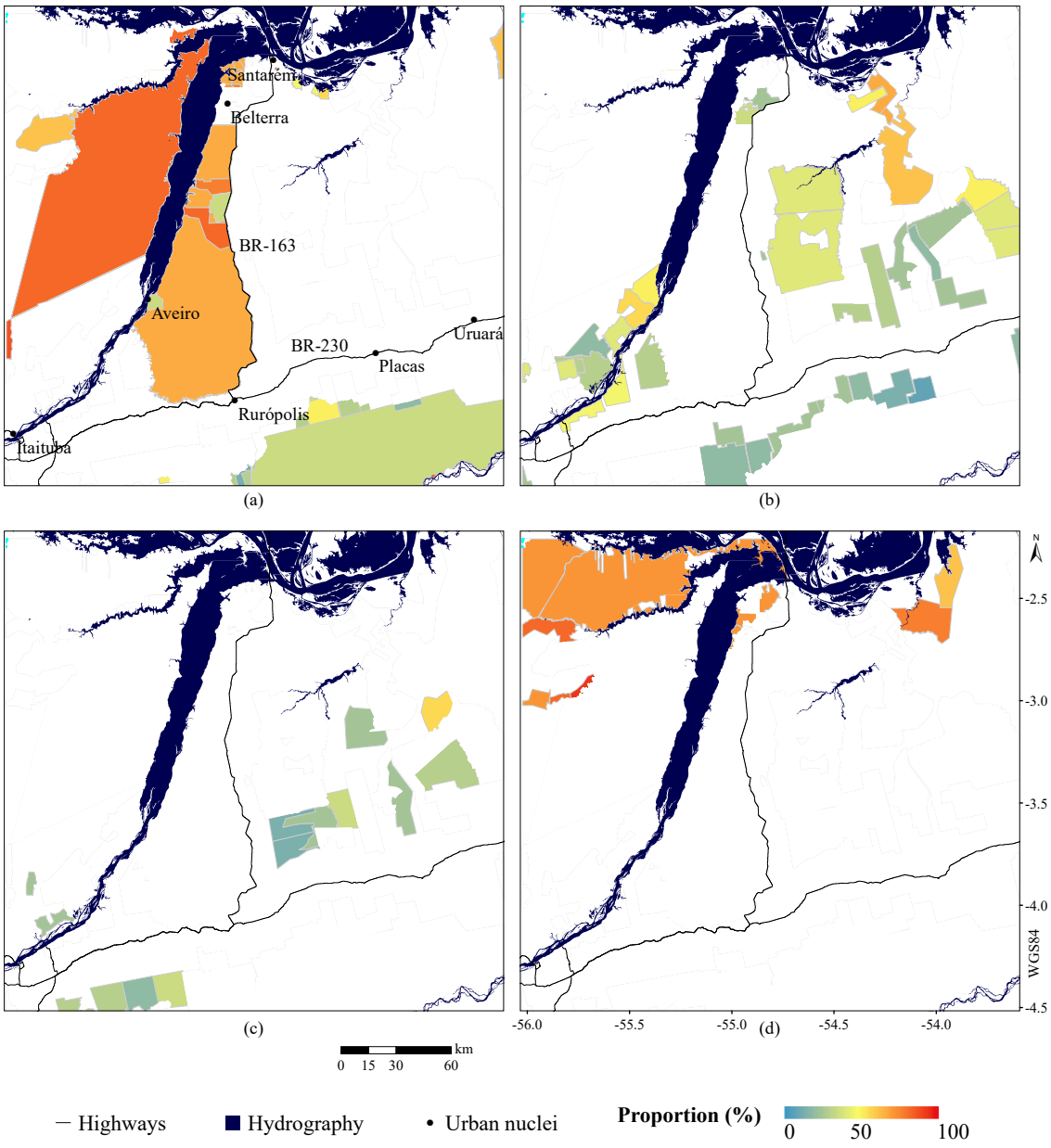
Figure 7.21 - Median Age of Secondary Vegetation in 2020 stratified by types of areas delimited by public policies with spatial expression.



(a) Conservation Units+Indigenous Lands+Quilombola Areas +superpositions to Settlement Projects (b) Federal Settlement Project (PA)+Joint Settlements Project (PAC)+Integrated Colonization Project (PIC), (c) Sustainable Development Project (PDS), (d) Agro-extractivist Settlement Project (PAE)+Agro-extractivist State Settlement Project (PEAEX)+Sustainable Settlement State Project (PEAS). Hydrography and urban nuclei location from IBGE (2019). Highways from MInfra (2021). Delimited areas from INCRA (s.d.), FUNAI (2021), and MMA (2020).

SOURCE: The author.

Figure 7.22 - Relative area of Secondary Vegetation in 2020 stratified by types of areas delimited by public policies with spatial expression.



(a) Conservation Units+Indigenous Lands+Quilombola Areas +superpositions to Settlement Projects (b) Federal Settlement Project (PA)+Joint Settlements Project (PAC)+Integrated Colonization Project (PIC), (c) Sustainable Development Project (PDS), (d) Agro-extractivist Settlement Project (PAE)+Agro-extractivist State Settlement Project (PEAEX)+Sustainable Settlement State Project (PEAS).

SOURCE: The author.

7.4 Discussion

Although the area of *Secondary Vegetation* is increasing with time, its proportion to the deforested area fluctuates, with marked periods of increment and reduction. The *Deforestation of Secondary Vegetation* generally surpasses the *Deforestation of Forest* from 1989 onwards, with the exception of year 1994, in which the *Deforestation of Forest* was slightly higher. Furthermore, we have found differences in the *Relative area of Secondary Vegetation* and forest regeneration trajectories attributes for the Lowe Tapajós region that can be related to known differences in agricultural practices in the region.

For instance, the areas concentrating the highest *Relative area of Secondary Vegetation* values vary over the years within the study area. These variations coincide with known shifts in agricultural practices in the region, such as the soy entry in the Santarém Plateau. This is a region historically used for small-scale agriculture, with many areas being converted to large-scale agriculture from 2000 to 2006 (SANTOS, 2020; COELHO et al., 2021; PAULA et al., 2022). However, small-scale agriculture areas are still present up to 2020 (PAULA et al., 2022). According to Coelho et al. (2021), the area used for large-scale agriculture in the Santarém Plateau remained relatively stable from 2007, when compared to other regions within the Amazon. In the last evaluated years, this region presented areas interspersed with low and relatively high concentrations of *Secondary Vegetation*. We observed a high concentration of *Secondary Vegetation* in this area in the 1990s, followed by a marked decrease in concentration from 2000 to 2010.

We also noticed shifts in the *Relative area of Secondary Vegetation* in riverine areas, which are areas with the presence of small-scale agriculture well documented (AFFONSO et al., 2016; SOUZA, 2017). Interestingly, these tended to present lower values in the initial years of the analysis. From the 2000s we could identify an increase in the *Relative area of Secondary Vegetation*, mainly in newly deforested areas. Riverine areas and those at the edge of dense *Forest* matrices also present the higher *Relative area of Secondary Vegetation* and proportion of deforested pixels with at least one forest regeneration event from 1984 to 2020. These results are aligned with the ones obtained by Perz and Skole (2003b), and more recently, Oviedo and Doblas (2022), in which the authors assess that traditional activities developed in the context of small-scale agriculture have positive effects on the extent of secondary vegetation in the Amazon as a whole. In general, areas with the presence of small-scale agriculture also presented a higher number of regeneration cycles, as

expected due to the adopted management practices. This characteristic was observed in the Santarém Plateau, areas along the Cuiabá-Santarém highway, and also riverine areas.

In contrast, the areas along the Transamazon highway, historically deforested for the implementation of pastures for extensive cattle raising, did not exhibit similar changes in the *Relative area of Secondary Vegetation* over time. These also tend to present lower values of *Relative area of Secondary Vegetation* for all analyzed years, which is in accordance with the low concentration of secondary vegetation found in areas used for livestock in the Brazilian legal Amazon found by Almeida et al. (2010). In the Lower Tapajós region, these are areas with relatively old deforestation of *Forest* but a low concentration of forest regeneration events. This does not mean, however, that these are not dynamic areas regarding the deforestation and growth of *Secondary Vegetation*, although this pattern varies from year to year. These results indicate that areas along the Transamazon highway can be left to regenerate briefly but are again deforested/managed to keep their use.

Around 96% of the forest regeneration events identified in the Lower Tapajós region presented four regeneration cycles or less. This number of cycles is similar to the one found in areas of shifting cultivation in Tefé and Alvarães, in the state of Amazonas, by Jakovac et al. (2017), or in pasture-dominated areas along the Cuibá-Santarém highway, in the area near Novo Progresso, in the Southwest of the Pará state (MÜLLER et al., 2016). The majority of regeneration events observed in the area happened in areas being used for five years or less. Additionally, around 75% of the fully observed regeneration events persisted for up to five years. Per year, up to 91% of the deforested *Secondary Vegetation* has been growing for five years or less. Although the proportion of the deforested *Secondary Vegetation* with five years or less has oscillated during the time series, they have kept somewhat constant from 1990 (the time for which we have full observations for this age) with mean values around 86% with a standard deviation of 3%. The exception is the period between 2010 and 2016, in which this proportion drops to ∼73% with standard deviation of 5%. The permanence of *Secondary Vegetation* in the Lower Tapajós region corroborates the estimates calculated by Almeida (2009), in which half of the total area of secondary vegetation in the Brazilian Legal Amazon usually takes nearly five years to be deforested.

Age and *Relative area of Secondary Vegetation* varied within the same type of agrarian system or in areas affected by similar public policies with territorial

expression in the Lower Tapajós region. Although related, time since the deforestation of *Forest* could not explain the observed differences. Nonetheless, riverine areas in the northwest portion of the study area, and those neighboring the Tapajós-Arapiuns Extrativist Reserve, present forest regeneration trajectories with a relatively high *Number of Regeneration Cycles*, while also presenting the highest medium values of *Persistence of Secondary Vegetation*. These are areas included within Agro-extractivist Settlement Projects (PAE) and Agro-extractivist State Settlement Projects (PEAEX), more specifically the PAE Lago Grande, PEAEX Arua, and PEAEX Curumucuri. They also presented the older *Secondary Vegetation* in the final years of the time series. These results complement the analysis carried out by Oviedo and Doblas (2022), in which the authors conclude that the presence of traditional occupation in protected areas tends to enable forest regeneration processes, while also acting as a buffer for the deforestation of primary forest. In addition, the secondary vegetation that grows in areas of shifting cultivation tends to accumulate biomass faster than those growing in abandoned pasture areas (WANDELLI; FEARNSIDE, 2015). Therefore, public policies oriented toward the regeneration of forests could highly benefit from the inclusion of the management practices of the traditional populations.

7.5 Chapter conclusions

In this chapter, we have characterized both the spatial patterns and temporal dynamic of secondary vegetation, as well as looked into differences in forest regeneration trajectories in the Lower Tapajós region, in the Pará state. We have found that the Santarém Plateau concentrates forest regeneration trajectories with a relatively high number of regeneration cycles, probably due to the historic use for small-scale agriculture that persists even with the entry of soy in the region. Areas occupied by agro-extractivist projects, both federal and from the state, at the West portion of the Tapajós river also tend to concentrate trajectories with a relatively high number of regeneration cycles, but also with high mean persistence of the secondary vegetation. In 2020, this second region also presents a high proportion of deforested areas occupied by secondary vegetation, which is older than the vegetation found in the Santarém Plateau, along the highways, in newly deforested areas, and in protected areas such as Indigenous Lands and Conservation Units within the region.

Furthermore, we found differences in patterns of deforestation and growth of secondary vegetation over the years. These differences were found regarding the age of the deforested secondary vegetation, the time of use before the regeneration event,

the total amount of regenerated area, and regarding spatial patterns of these events. Although these results can not be generalized to regional scales, it highlights the importance of analysis based on LULC trajectories with at least annual observations, as snapshots in higher intervals may fail to capture these fluctuations.

Additionally, our results demonstrate that the concentration and age of secondary vegetation vary within areas with similar agrarian systems, and presumed similar management techniques, even on the local scale. The same was observed for areas affected by similar types of public polices. Furthermore, we were able to identify differences regarding the age of the secondary vegetation to the deforestation time, and/or proportion of deforested area occupied by secondary vegetation in these areas. These results should be further investigated within land use and land cover change models and rigorous statistic analyses. Nonetheless, they highlight the need for detailed studies of secondary vegetation dynamics in time and space within the Amazon, to improve carbon estimates and enable the elaboration of more effective conservation policies, territorial ordering, and sustainable land use practice incentives.

8 CONCLUSIONS

In this study, we proposed methods for the classification and analysis of forest regeneration trajectories based on previous knowledge about land cover dynamics and multi-temporal remote sensing data. Here, forest regeneration trajectories were treated as a particular type of LULC trajectory, obtained from the processing of remote sensing image time series. Our literature review led us to segment remote sensing based LULC trajectory studies into the basic steps used to orient this research: 1) problem definition, 2) trajectory classification, and 3) extraction and analysis of information. Applying these steps to our particular problem, we elaborated four main specific objectives within this thesis, explored in Chapters 4 to 7. Meeting these objectives presented particular problems derived both from the inherent complexity of the problem and methodological constrains, thus demanding the creation and adaptation of tools. In this final chapter, we aim to highlight the proposed solutions and observed limitations for the problems encountered at meeting each of the original specific objectives, as well as to synthesize the main findings, contributions, and perspective for future studies. We then discuss how these relate to the main hypothesis of this thesis.

The first specific objective was *to propose objective definitions of LULC classes in the studied region, in order to allow for consistent and replicable analyses based on multi-temporal remote sensing data sets*. LULC trajectories studies demand particular attention to the definition of LULC classes and the selection of adequate remote sensing data and legend. It is a well know problem that LULC classes defined by subjective criteria can diminish the significance of a study, hindering the reproducibility and the comparison of results with other studies (Section 2.1.2). Our adopted solution to decrease the subjectivity of class definition and allow comparison of results was to identify and describe the LULC classes in the Lower Tapajós region using the terminology proposed by the Land Cover Meta-Language (LCML). We selected quantifiable and easily recognizable physiognomic characteristics of land cover classes measured at the field level in the study area. We then described 16 LULC classes in LCML, which were grouped into legends with different levels of detail. We then classified a Landsat5/TM image using supervised classifications approaches and each of these legends, to evaluate their potential and constraints for studies based on Landsat-like data. These experiments were described in Chapter 4.

Our results indicate that our proposed thresholds (height and cover proportion of soil, litter, herbaceous vegetation, shrubs, and trees) are not enough to properly

describe the LULC classes in the Lower Tapajós, since multi-temporal knowledge about the classes is also necessary. This finding is particularly interesting for remote sensing analysts alike because they highlight the impossibility of accurately defining some land cover classes with a single observation in time, even with very high-resolution images. Nonetheless, the proposed thresholds are useful to delimit classes that happen successively and to help to standardize field campaigns.

The classification results revealed that classes formed by the same elements of land cover with similar thresholds present high confusion. Additionally, classifications obtained using legends based on the spectral characteristics of the remote sensing image being analyzed, although more accurate, are not always useful. We also observed that diminishing the number of classes in an analysis may be an important tool to improve accuracy.

Regarding the context of the present thesis, this chapter presents important elements for the construction of the legends used in the following chapters. First, it allows for the objective definition of the most detailed set of LULC classes that can be identified from the collected field data. From the joint analysis of this data, we further developed an interpretation key for the used classes in the Lower Tapajós region, which was used for the collection of labeled samples. Our results also show us the need for novel classification strategies, since we can only separate secondary vegetation from other forested classes considering multi-temporal and inter-annual information. This finding ultimately led to the proposal of the next specific objective. Lastly, our results further provided insights about the needed temporal window for the following analysis and legends used in Chapters 5 and 6.

There are aspects of this study that may be further explored. For instance, we did not develop the legend for aquatic or regularly flooded areas. It may also be interesting to estimate the cover proportion of elements from images obtained by unmanned aerial vehicles (UAVs), which can be useful for field data collection and to further reduce the subjectivity of estimates and/or enable different teams of researchers to define different legends based on collected data.

The second specific objective was *to develop a novel trajectory classifier able to extract only valid land cover trajectories directly from remote sensing time series*. In Chapter 5, we proposed a novel trajectory classifier named Compound Maximum *a Posteriori* (CMAP) classifier. CMAP incorporates the multi-temporal data sets to weight the classification of LULC trajectories, using a global generative classification approach. CMAP was tested in two case studies, in which we compared CMAP

results with those obtained by stacking classifications from the Maximum Likelihood (ML) classifier, based on the same class definition, images, and training samples. Our results show that the use of ML and post-classification comparison resulted in invalid land cover trajectories in more than 50% of the test area in both case studies, even though our experiments did not consider more than 6 dates. CMAP was able to classify only valid trajectories with improved accuracy and reduction of the classification noise.

In the context of this thesis, CMAP was crucial to include multi-temporal information to enable the discrimination between secondary vegetation and other forested classes. CMAP has also been already used in classification problems within the Atlantic Forest in parallel studies (MACIEL et al., 2021). Furthermore, our literature review (Chapter 2) highlights the need for LULC trajectory classification methods adequate for the processing of sparse time series, multi-sensor data, and different aggregations of a given legend. Another recurring problem among trajectory classification techniques is that several steps are needed, including a post-classification step to correct invalid trajectories. The CMAP classifier stands out for classifying image time series directly on valid-only LULC trajectories in a single step using multi-sensor data at different tiers of a hierarchical legend. Currently, the computational cost of CMAP is high for large areas and/or long time series, which can limit its operational use. CMAP also inherited many characteristics of the Maximum Likelihood (ML) classifier, such as the expectation of training samples of uni-modal classes. Another problem specific to our implementation is the expectation of Gaussian data. In this sense, improvements to the performance of CMAP and the incorporation of new base classifiers to the algorithm, as well as new methods for calculating the probability of transitions within the classifier, are already in motion (DUTRA et al., 2022), and may further settle CMAP as an important tool for the classification of time series.

Our third specific objective was *to extract LULC trajectories for the study area with annual observations from years 1984 to 2020*. This objective imposed varied challenges derived from the length of this time series. Therefore, in Chapter 6, we proposed solutions for a few of the constraints identified for CMAP and classify the LULC trajectories for the Lower Tapajós region using 37 years of Landsat data and in three classes of interest: *Forest*, *Secondary Vegetation*, and *Temporary Cultures*. We also showed that CMAP is more appropriate for this task than the use of post-classification comparison even with the use of post-classification filters, which is the methodology adopted by MapBiomas so far. From a methodological point of

view, we also presented solutions for dealing with the cloud problem within LULC trajectories classifications and the scarcity of training samples for multi-temporal supervised classifications. The latter was done by the use of transferred samples, i.e labeled samples and correspondent spectral information collected from a subset of the images being classified.

This research presented many drawbacks that remained unsolved within this thesis:

a) the use of transferred samples assumes very similar behavior of the classes of interest in all observations. This aspect limits the adequate time window for image acquisition and uncalibrated/multi-sensor data sets, which in turn led to the use of few spectral bands of Landsat data and a huge amount of areas without observation in the time series. We expect that advancements in data calibration may reduce this problem for historic analysis;

b) the difference in data availability over the years precludes the use of multi-temporal metrics that could offer solutions to the classification of classes such as Agriculture and Pasture for the oldest images in the time series;

c) as the main limitation within this analysis, we highlight the lack of reference samples for the whole study area and/or in all years of interest. As such, we were unable to fully assess the accuracy of the classifications. Even for the years in which we had labeled samples, these were derived from other classification products and may be biased for this type of analysis. For instance, we do not have samples for areas in which the reference classifications differed, along feature borders, or from persistent areas of secondary vegetation.

Nonetheless, this study indicates important paths for advancement in LULC trajectories classification in operational contexts, which will greatly benefit from parallel studies using the CMAP classification approach with different base classifiers (DUTRA et al., 2022) and/or theoretical advancements in pattern recognition to allow CMAP to consider information from external classification to delimit different sets of possible trajectories depending of the geographic position (unpublished). These advancements would allow for the use of different inputs and the rupture of the classification process in one or more steps. These could, in turn, be used for LULC

classification conditioned to TerraClass or PRODES data sets, for example, to fill or expand the classification time series of these projects.

Our last specific objective was *to identify, quantify, and characterize forest regeneration trajectories related to different agrarian systems and public policies with territorial expression*. We used the LULC trajectories classified in Chapter 6 for the analysis presented in Chapter 7. Our results reveal that:

a) the spatial pattern of areas concentrating the highest proportions of deforested areas covered by secondary vegetation varies over the years. These variations coincide with known shifts in agricultural practices, such as the soy entry in the region;

b) although the absolute area of secondary vegetation is increasing with time, its proportion to the deforested area fluctuates, with marked periods of increment and reduction;

c) around 96% of the forest regeneration events identified in the Lower Tapajós region presented four regeneration cycles or less. The higher values are concentrated in areas with the presence of small-scale agriculture well documented by field information;

d) around 75% of the fully observed regeneration events persisted no more than five years;

e) from 2010 to 2016, there is a slight shift indicating a higher proportion of deforestation of older secondary vegetation. This period coincides with policies to combat the deforestation of *Forests* and precedes the Normative Instruction 08 of October 28, 2015 (PARÁ, 2015). We suggest that this shift could be related to the preference to deforest older *Secondary Vegetation* instead of *Forests*;

f) the majority of regeneration events happened in areas being previously used for five years or less;

g) age and proportion of secondary vegetation varied within the same type of agrarian system or in areas affected by similar public policies with territorial expression in the Lower Tapajós region;

h) riverine areas at the northwest portion of the study area included in agro-extractivist projects present forest regeneration trajectories with a

relatively high number of regeneration cycles and persistence of secondary vegetation, as well as the older secondary vegetation observed at the final years of the time series.

This study presents a limitation common to all analysis based on the classification of remote sensing data: errors in the classification reflect as errors in the analysis. Within the resources of this thesis, we could not fully assess the classification errors neither for the classifications of each year nor for the calculated trajectories. Furthermore, we needed to interpolate many areas without observations, which could further influence the observed values. Another common problem of analyses based on remote sensing images is that our interpretation of the results is based solely on changes in the spectral response of the images. As such, we cannot fully affirm if the changes we observed are caused by land decisions affected by specific events. As such, these results should be further investigated within land use and land cover change models and rigorous statistic analyses. Furthermore, these should be further cross-examined with data from past field campaigns, and new field information may be needed for us to fully understand some of the observed forest regeneration dynamics.

The results obtained in Chapter 6 highlight differences in forest regeneration dynamics even at the local scale. As different forest regeneration processes may impact the dynamics of biomass accumulation and carbon absorption rates, there is a pressing need to better characterize forest regeneration events in the Brazilian Amazon to improve the precision and accuracy of carbon emission estimates. From a socio-economic point of view, such studies can reveal how different forms of occupation produce different landscapes and forest regeneration dynamics, which are important to consider when it comes to elaborating conservation policies, territorial ordering, and sustainable land use practice incentives. In particular, our results point out that areas used for small-scale agriculture within agro-extractivist settlements tend to accumulate older secondary vegetation in the Lower Tapajós region. So far, these are areas for which we have LULC maps that are sparse in time and space since classes such as Shifting Cultivation have been invisible to Amazon-wide LULC classifications. These are areas that present important economical roles within the Amazonian population, usually hidden by inadequate economical indexes. Although we cannot generalize our results to regional scales, they indicate the need to better characterize these areas in an Amazon-wide context to understand their roles as providers of ecosystem services and to create protection mechanisms.

From the obtained results, we are able to go back to the hypothesis that structured this study: *incorporating previous knowledge about land cover dynamics and the information of multi-temporal data sets during the classification of LULC trajectories results in consistent and more accurate classification of secondary vegetation areas.* This hypothesis was proven true by the proposal and evaluation of the Compound Maximum *a Posteriori* (CMAP) classifier, in Chapters 5 and 6, and the later use of the classified LULC trajectories to identify differences in the number of regeneration cycles, the persistence of secondary vegetation, and in patterns of deforestation and growth of secondary vegetation in riverine areas, areas historically occupied by small-scale agriculture, and areas deforested for the establishment of pastures in the Lower Tapajós region. Lastly, we identified that up to 11.5% of the deforestation of secondary vegetation in a single year, in this region, occurs in areas regenerating from 11 years or more. These processes can only be fully characterized by the classification of valid LULC trajectories with 22+ years of observations, further evidencing the importance of approaches such as CMAP.

REFERENCES

AFFONSO, A. G.; ESCADA, M. I. S.; AMARAL, S.; SOUZA, A. R. d.; SIQUEIRA, J. M. d.; TORRES, N. C.; CAMILOTTI, V. L.; DAL'ASTA, A. P.; COSTA, L. C. O.; SOARES, F. d. R. **As comunidades ribeirinhas do Baixo Tapajós (PA): infraestrutura, mobilidade, serviços sócio ambientais e conectividade**. São José dos Campos, 2016. 150 p. Available from: <http://urlib.net/rep/8JMKD3MGP3W34P/3M7C69L>. Access in: 19 Mar. 2021. 12, 42, 154, 174

AGUIAR, A. P. D.; OMETTO, J. P.; NOBRE, C.; LAPOLA, D. M.; ALMEIDA, C.; VIEIRA, I. C.; SOARES, J. V.; ALVALA, R.; SAATCHI, S.; VALERIANO, D.; CASTILLA-RUBIO, J. C. Modeling the spatial and temporal heterogeneity of deforestation-driven carbon emissions: the INPE-EM framework applied to the Brazilian Amazon. **Global Change Biology**, v. 18, n. 11, p. 3346–3366, 2012. 2

AHLQVIST, O. Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: a study of 1992 and 2001 U.S. National Land Cover Database changes. **Remote Sensing of Environment**, v. 112, n. 3, p. 1226 – 1241, 2008. ISSN 0034-4257. 13

ALASKA SATELLITE FACILITY (ASF). **ALOS PALSAR**. 2020. Available from: <https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/>. Access in: 18 Dec. 2020. 15

ALMEIDA, C. A. **Estimativa da área e do tempo de permanência da vegetação secundária na Amazônia Legal por meio de imagens LANDSAT/TM"**. 2009. 129 p. Thesis (Master in Remote Sensing) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2009. 175

ALMEIDA, C. A.; VALERIANO, D. M.; ESCADA, M. I. S.; RENNÓ, C. D. Estimativa de área de vegetação secundária na Amazônia Legal Brasileira. **Acta Amazonica**, v. 40, n. 2, p. 289–302, 2010. 9, 10, 175

ALMEIDA, C. A. d.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; ADAMI, M.; VENTURIERI, A.; DINIZ, C. G.; DESSAY, N.; DURIEUX, L.; GOMES, A. R. 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**, v. 46, n. 3, p. 291–302, 2016. 9, 27, 28, 46, 74

ANDERSON, J. R.; HARDY, E. E.; ROACH, J. T.; WITMER, R. E. **A land use and land cover classification system for use with remote sensor data**. Washington, 1976. 28 p. Available from: <https://pubs.usgs.gov/pp/0964/report.pdf>. Access in: 19 Mar. 2021. 13, 14, 43

ANGELIS, C. F.; FREITAS, C. C.; VALERIANO, D. M.; DUTRA, L. V. Multitemporal analysis of land use/land cover JERS-1 backscatter in the Brazilian tropical rainforest. **International Journal of Remote Sensing**, v. 23, n. 7, p. 1231–1240, 2002. 13

ANJOS, D. S. d.; LU, D.; DUTRA, L. V.; SANT'ANNA, S. J. S. Change detection techniques using multisensor data. In: THENKABAIL, P. S. (Ed.). **Remotely sensed data characterization, classification, and accuracies**. London: CRC Press, 2015. v. 1, p. 375–395. ISBN 9781482217865. 20

ARAFAT, S. M.; SALEH, N. S.; ABOELGHAR, M.; ELSHRKAWY, M. Mapping of North Sinai land cover according to FAO-LCCS. **The Egyptian Journal of Remote Sensing and Space Science**, v. 17, n. 1, p. 29 – 39, 2014. 14

ASNER, G. P.; KELLER, M.; SILVA, J. N. Spatial and temporal dynamics of forest canopy gaps following selective logging in the eastern Amazon. **Global Change Biology**, v. 10, n. 5, p. 765–783, 2004. 59

ASSIS, T. O.; AGUIAR, A. P. D.; RANDOW, C. von; GOMES, D. M. P.; KURY, J. N.; OMETTO, J. P.; NOBRE, C. A. CO2 emissions from forest degradation in Brazilian Amazon. **Environmental Research Letters**, v. 15, n. 10, p. 104035, 2020. 24

AZEREDO, M. **Mineração e análise de trajetórias de mudança de cobertura da terra: explorando padrões comportamentais no contexto da degradação florestal**. 2017. 150 p. Thesis (PhD in Applied Computing) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017. Available from: <http://urlib.net/rep/8JMKD3MGP3W34P/3NGPJ7H>. Access in: 24 May 2018. 24, 25

AZEREDO, M.; MONTEIRO, A. M. V.; ESCADA, M. I. S.; FERREIRA, K. R.; VINHAS, L.; PINHEIRO, T. F. Mineração de trajetórias de mudança de cobertura da terra em estudos de degradação florestal. **Revista Brasileira de Cartografia**, v. 68, n. 4, p. 717–731, 2016. 18, 22, 25

BANSKOTA, A.; KAYASTHA, N.; FALKOWSKI, M. J.; WULDER, M. A.; FROESE, R. E.; WHITE, J. C. Forest monitoring using Landsat time series data: a review. **Canadian Journal of Remote Sensing**, v. 40, n. 5, p. 362–384, 2014. 7

BATISTELLA, M.; ALVES, D.; ARTAXO, P.; BUSTAMANTE, M.; KELLER, M.; AO, F. L.; MARENGO, J. A.; MARTINELLI, L.; NOBRE, C. A. **Plano científico - LBA2**. [S.l.], 2007. Available from: <http://lba2.inpa.gov.br/media/arquivos_lba/pdfs/LBA_2_Plano_Cientifico_Portugues_Ago_2007.pdf>. Access in: 25 Feb 2019. 40

BECKER, B. K. **Amazônia**. São Paulo: Ática, 1990. 112 p. 36, 37, 39

_____. Geopolítica da Amazônia. **Estudos Avançados**, v. 19, p. 71 – 86, 2005. 11

_____. **Amazônia**: geopolítica na virada do III milênio. 2. ed. Rio de Janeiro: Garamond, 2007. 172 p. 1, 11, 34, 35, 38, 40

BECKER, B. K.; STENNER, C. **Um futuro para a Amazônia**. São Paulo: Oficina de Textos, 2008. 150 p. 35, 36, 37, 38

BIG EARTH OBSERVATION DATA ANALYTICS FOR LAND USE AND LAND COVER CHANGE INFORMATION (E-SENSING). **SITS - Satellite Image Time Series Analysis**. 2019. Available from: <https://github.com/e-sensing/sits/blob/master/README.md>. Access in: 05 May 2019. 19

BRASIL. PRESIDÊNCIA DA REPÚBLICA. Lei no 4.771, de 15.09.1965. Institui o novo Código Florestal. **Jus Brasil, Legislação**, 1965. Available from: <http://www.planalto.gov.br/ccivil_03/Leis/L6938.htm>. Access in: 19 Feb. 2019. 39

_____. Lei no 5173, de 27.10.1966. Dispõe sobre o Plano de Valorização Econômica da Amazônia; extingue a Superintendência do Plano de Valorização Econômica da Amazônia (SPVEA), cria a Superintendência do desenvolvimento da Amazônia (SUDAM), e dá outras providências. **Diário Oficial da União**, 1966. Available from: <http://www.planalto.gov.br/ccivil_03/leis/L5173compilado.htm>. Access in: 19 Mar. 2021. 1

_____. Decreto-lei no 1.106, de 16.06.1970. Cria o Programa de Integração Nacional, altera a legislação do imposto de renda das pessoas jurídicas na parte referente a incentivos fiscais e dá outras providências. **Jus Brasil, Legislação**, 1970. Available from: <http://www.planalto.gov.br/ccivil_03/decreto-lei/1965-1988/Del1106.htm>. Access in: 18 Jan. 2019. 38

_____. Decreto-lei no 1.164, de 01.04.1971. Declara indispensáveis à segurança e ao desenvolvimento nacionais terras devolutas situadas na faixa de cem quilômetros de largura em cada lado do eixo de rodovias na Amazônia legal, e dá outras providências. **Jus Brasil, Legislação**, 1971. Available from: <http://www.planalto.gov.br/ccivil_03/decreto-lei/Del1164.htm>. Access in: 19 Jan. 2019. 38

_____. Lei complementar no 124, de 03.01.2007. Institui, na forma do art. 43 da Constituição Federal, a Superintendência do Desenvolvimento da Amazônia - SUDAM; estabelece sua composição, natureza jurídica, objetivos, área de competência e instrumentos de ação; dispõe sobre o Fundo de Desenvolvimento da Amazônia - FDA; altera a Medida Provisória no 2.157-5, de 24 de agosto de 2001; revoga a Lei Complementar no 67, de 13 de junho de 1991; e dá outras providências. **Diário Oficial da União**, 2007. Available from: <http://www.planalto.gov.br/ccivil_03/leis/lcp/Lcp124.htm>. Access in: 19 Mar. 2021. 1

_____. **Intended Nationally Determined Contribution**: towards achieving the objective of the united nations framework convention on climate change. 2016. Available from: <https://www4.unfccc.int/sites/NDCStaging/Pages/Party.aspx?party=BRA>. Access in: 22 Sept. 2021. 42

BRASIL. MINISTÉRIO DA INFRAESTRUTURA (MINFRA). **Mapas e bases dos modos de transportes**. 2021. Available from: <https://www.gov.br/infraestrutura/pt-br/assuntos/dados-de-transportes/bit/bitmodosmapas>. Access in: 25 Feb. 2022. 33, 44, 82, 107, 170, 172

BRASIL. MINISTÉRIO DO MEIO AMBIENTE (MMA). **Plano de Ação para a Prevenção e Controle do Desmatamento e das Queimadas no Cerrado (PPCerrado) e Plano de Ação para a Prevenção e Controle do Desmatamento na Amazônia Legal (PPCDAm)**: Fase 2016-2020. volumes I, II e III. Brasília, 2018. Available from: <http://combateaodesmatamento.mma.gov.br/images/Doc_ComissaoExecutiva/Livro-PPCDam-e-PPCerrado_20JUN2018.pdf>. Access in: 03 Mar. 2019. 41

_____. **Plano de Manejo da Floresta Nacional do Tapajós**: Volume 1 - diagnóstico. [S.l.], 2019. Available from: <https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/amazonia/lista-de-ucs/flona-do-tapajos/arquivos/plano_de_manejo_flona_do_tapajos_2019_vol1.pdf>. Access in: 26 June 2022. 38, 41

_____. **Downloads de dados geográficos**. 2020. Available from: <http://mapas.mma.gov.br/i3geo/datadownload.htm>. Access in: 25 Feb. 2022. 33, 82, 140, 141, 172

_____. **Projeto BR–163**. s.d. Available from: <http://www.mma.gov.br/florestas/projeto-br-163>. Access in: 3 Mar. 2019. 37, 41

BRAZ, B. C.; D'ALGE, J. C. **AMAZONIA 1**: descritivo da missão e do satélite. [S.l.], 2021. Available from: <http://www.inpe.br/amazonia1/sobre_satelite/>. Access in: 28 Feb. 2022. 15

BROWN, S.; LUGO, A. E. Tropical secondary forests. **Journal of Tropical Ecology**, v. 6, n. 1, p. 1–32, 1990. 1, 9

CÂMARA, G.; VALERIANO, D. M.; SOARES, J. V. **Metodologia para o cálculo da taxa anual de desmatamento na Amazônia Legal**. São José dos Campos: INPE, 2006. 20

CAPANEMA, V. P.; SANCHES, I. D.; ESCADA, M. I. S. Comparação entre os produtos temáticos de uso e cobertura da terra do terraclass amazônia e mapbiomas: teste de aderência entre classes. In: GHERARDI, D. F. M.; SANCHES, I. D.; ARAGÃO, L. E. O. e. C. d. (Ed.). **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2019. p. 724–727. ISBN 978-85-17-00097-3. Available from: <http://urlib.net/rep/8JMKD3MGP6W34M/3U67378>. Access in: 19 Mar. 2021. 30

CARVALHO, R.; ADAMI, M.; AMARAL, S.; BEZERRA, F. G.; AGUIAR, A. P. D. Changes in secondary vegetation dynamics in a context of decreasing deforestation rates in Pará, Brazilian Amazon. **Applied Geography**, v. 106, p. 40–49, 2019. 23

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. When is a forest a forest? Forest concepts and definitions in the era of forest and landscape restoration. **Ambio**, p. 1–13, 2016. 12

CHAZDON, R. L.; LINDENMAYER, D.; GUARIGUATA, M. R.; CROUZEILLES, R.; BENAYAS, J. M. R.; CHAVERO, E. L. Fostering natural forest regeneration on former agricultural land through economic and policy interventions. **Environmental Research Letters**, v. 15, n. 4, p. 043002, 2020. 2, 11

CHAZDON, R. L.; PERES, C. A.; DENT, D.; SHEIL, D.; LUGO, A. E.; LAMB, D.; STORK, N. E.; MILLER, S. E. The potential for species conservation in tropical secondary forests. **Conservation Biology**, v. 23, n. 6, p. 1406–1417, 2009. 2

CHOKKALINGAM, U.; JONG, W. D. Secondary forest: a working definition and typology. **The International Forestry Review**, v. 3, n. 1, p. 19–26, 2001. 9

COE, M. T.; MARTHEWS, T. R.; COSTA, M. H.; GALBRAITH, D. R.; GREENGLASS, N. L.; IMBUZEIRO, H. M. A.; LEVINE, N. M.; MALHI, Y.; MOORCROFT, P. R.; MUZA, M. N.; POWELL, T. L.; SALESKA, S. R.; SOLORZANO, L. A.; WANG, J. Deforestation and climate feedbacks threaten the ecological integrity of south–southeastern Amazonia. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 368, n. 1619, p. 20120155, 2013. 1

COELHO, A.; AGUIAR, A. P. D.; TOLEDO, P. M. d.; ARAÚJO, R.; CANTO, O. d.; FOLHES, R.; ADAMI, M. Rural landscapes and agrarian spaces under soybean expansion dynamics: a case study of the santarém region, brazilian amazonia. **Regional Environmental Change**, v. 21, n. 4, p. e100, 2021. 159, 168, 174

COMBER, A. J.; LAW, A. N. R.; LISHMAN, J. R. Application of knowledge for automated land cover change monitoring. **International Journal of Remote Sensing**, v. 25, n. 16, p. 3177–3192, 2004. 18

COMISSÃO PRÓ-ÍNDIO DE SÃO PAULO (CPISP). **Observatório Terras Quilombolas**. 2022. Available from: <https://cpisp.org.br/direitosquilombolas/observatorio-terras-quilombolas/>. Access in: 07 Mar. 2022. 141

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing of Environment**, v. 37, n. 1, p. 35 – 46, 1991. ISSN 0034-4257. 50

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. [S.l.]: CRC Press, 2008. 50

COPPIN, P.; JONCKHEERE, I.; NACKAERTS, K.; MUYS, B.; LAMBIN, E. Digital change detection methods in ecosystem monitoring: a review. **International Journal of Remote Sensing**, v. 25, n. 9, p. 1565–1596, 2004. 16, 105

CORD, A.; CONRAD, C.; SCHMIDT, M.; DECH, S. Standardized FAO-LCCS land cover mapping in heterogeneous tree savannas of West Africa. **Journal of Arid Environments**, v. 74, n. 9, p. 1083 – 1091, 2010. 14

CORLETT, R. Tropical secondary forests. **Progress in Physical Geography**, v. 19, n. 2, p. 159–172, 1995. 1, 9

CORSINI, C. R. **Análise multitemporal das mudanças de biomassa da vegetação secundária na Amazônia brasileira**. 2018. 106 p. Thesis (PhD in Earth System Science) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2018. Available from: <http://urlib.net/rep/8JMKD3MGP3W34R/3R44PJP>. Access in: 12 May 2019. 24, 28

COSTA, F. A. Dinâmica agrária e balanço de carbono na Amazônia. **Revista EconomiA**, v. 10, n. 1, p. 117–151, 2009. 9

_____. Contributions of fallow lands in the brazilian amazon to CO2 balance, deforestation and the agrarian economy: inequalities among competing land use trajectories. **Elementa Science of the Anthropocene**, v. 4, 2016. 2, 9, 11, 162

COUTINHO, A.; ALMEIDA, C.; VENTURIERI, A.; ESQUERDO, J.; SILVA, M. **Uso e cobertura da terra nas áreas desflorestadas da Amazônia Legal**: TerraClass, 2008. Brasília,Belém, 2013. Available from: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/87809/1/TerraClass-completo-baixa-pdf.pdf>. Access in: 21 July 2020. 14, 51

DEFINIENS. **eCognition developer 8 user guide**. Munich, 2009. 236 p. 219

DI GREGORIO, A.; JANSEN, L. J. M. **Land Cover Classificiation System**: LCCS classification concepts and user manual for software version 2.0. Rome, Italy, 2005. 14, 43, 46, 47, 51, 73

DI GREGORIO, A.; O'BRIEN, D.; BEJAR, R.; MAFFEI, S.; BELLANCINI, A.; MARCHI, D. D.; LEONARDI, U. **FAO Land Cover Classification System 3 (Version 1.8.0)**. 2015. Available from: <www.geovis.net/Downloads.htm>. 45

DUTRA, L. V.; NEGRI, R. G.; LIMA, J.; SANT'ANNA, S. J. S.; PEREIRA, L. O. **Informações acerca de trabalhos de campo.** 2021. [Personal communication].Messages exchanged via social media on 30 Ago. 2021. 218

DUTRA, L. V.; RENNÓ, C. D.; GAMBA, P. **A generative method for simultaneous classification of multi-source time series of Remote Sensing data using Decision Tree Classifiers**. 2022. Cooperation project in progress. 104, 181, 182

DUTRA, L. V.; SCOFIELD, G. B.; NETA, S. R. A.; NEGRI, R. G.; FREITAS, C. C.; ANDRADE, D. Land cover classification in amazon using alos palsar full polarimetric data. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. p. 7259–7264. Available from: <http://urlib.net/dpi.inpe.br/sbsr\spacefactor\@m{}80/2008/11.17.18.24>. Access in: 19 Oct. 2015. 13

ELLWANGER, J. H.; KULMANN-LEAL, B.; KAMINSKI, V. L.; VALVERDE-VILLEGAS, J. M.; VEIGA, A. B. G.; SPILKI, F. R.; FEARNSIDE, P. M.; CAESAR, L.; GIATTI, L. L.; WALLAU, G. L.; ALMEIDA, S. E.; BORBA, M. R.; HORA, V. P.; CHIES, J. A. B. Beyond diversity loss and climate change: impacts of Amazon deforestation on infectious diseases and public health. **Anais da Academia Brasileira de Ciências**, v. 92, 2020. 1

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). **Zoneamento Ecológico-Econômico da Zona Oeste do Estado do Pará**: Diagnóstico socioambiental. Volume 1. [S.l.], 2010a. Available from: <http://www.amazonia.cnptia.embrapa.br/publicacoes_para.html>. Access in: 17 Jan. 2019. 34

_____. **Zoneamento Ecológico-Econômico da Zona Oeste do Estado do Pará**: Gestão territorial - diretrizes de uso e ocupação. Volume 2. [S.l.], 2010b. Available from: <http://www.amazonia.cnptia.embrapa.br/publicacoes_para.html>. Access in: 17 Jan. 2019. 34

_____. **Mapas de solos e de aptidão agrícola das áreas alteradas do estado do Pará**. 2016. Available from: <http://www.amazonia.cnptia.embrapa.br/publicacoes_estados/Para/Aptidao_Solos/Mapas_de_solos_e_aptidao_agricola_das_areas_alteradas_do_Para.pdf>. Access in: 17 Jan. 2019. 34

_____. **Satélites de monitoramento**. 2018. Available from: <https://www.embrapa.br/satelites-de-monitoramento>. Access in: 18 Dec. 2020. 15

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA); INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Nota técnica**: estimativa da área de vegetação secundária na Amazônia Legal (2020). 2021. Available from: <https://www.gov.br/inpe/pt-br/assuntos/ultimas-noticias/>. Access in: 20 Jan. 2020. 27

EUROPEAN ENVIRONMENT AGENCY (EEA). **Updated CLC illustrated nomenclature guidelines**. 2019. Available from: <https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf>. Access in: 29 Sept. 2020. 14

EUROPEAN SPACE AGENCY (ESA). **Sentinel online**. 2020. Available from: <https://sentinel.esa.int/web/sentinel/home>. Access in: 18 Dec. 2020. 15

FEARNSIDE, P. M. Environmental destruction in the Brazilian Amazon. In: GOODMAN, D.; HALL, A. (Ed.). **The future of Amazonia: destruction or sustainable development?** London: Palgrave Macmillan UK, 1990. p. 179–225. 1

_____. Deforestation in Brazilian Amazonia: history, rates, and consequences. **Conservation Biology**, v. 19, n. 3, p. 680–688, 2005. 1

FISHER, P. F. Remote sensing of land cover classes as type 2 fuzzy sets. **Remote Sensing of Environment**, v. 114, n. 2, p. 309 – 321, 2010. ISSN 0034-4257. 74

FOLEY, J. A.; ASNER, G. P.; COSTA, M. H.; COE, M. T.; DEFRIES, R.; GIBBS, H. K.; HOWARD, E. A.; OLSON, S.; PATZ, J.; RAMANKUTTY, N.; SNYDER, P. Amazonia revealed: forest degradation and loss of ecosystem goods and services in the amazon basin. **Frontiers in Ecology and the Environment**, v. 5, n. 1, p. 25–32, 2007. 1

FOODY, G. M. Assessing the accuracy of land cover change with imperfect ground reference data. **Remote Sensing of Environment**, Elsevier, v. 114, n. 10, p. 2271–2285, 2010. 20, 21

FRANKLIN, S. E.; AHMED, O. S.; WULDER, M. A.; WHITE, J. C.; HERMOSILLA, T.; COOPS, N. C. Large area mapping of annual land cover dynamics using multitemporal change detection and classification of Landsat time series data. **Canadian Journal of Remote Sensing**, v. 41, n. 4, p. 293–314, 2015. 17, 23

FRANTZ, D. FORCE–Landsat+ Sentinel-2 analysis ready data and beyond. **Remote Sensing**, v. 11, n. 9, p. 1124, 2019. 17

FREITAS, C. C.; SOLER, L. S.; SANT´ANNA, S. J. S.; DUTRA, L. V.; SANTOS, J. R.; MURA, J. C.; CORREIA, A. H. Land use and land cover mapping in Brazilian Amazon using polarimetric airborne P-band SAR data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 46, n. 10, p. 2956–2970, 2008. 13

FULLER, R.; SMITH, G.; DEVEREUX, B. The characterisation and measurement of land cover change through remote sensing: problems in operational applications? **International Journal of Applied Earth Observation and Geoinformation**, v. 4, n. 3, p. 243 – 253, 2003. 18

FUNDAÇÃO NACIONAL DO ÍNDIO (FUNAI). **Geoprocessamento e mapas**. 2021. Available from: <https://www.gov.br/funai/pt-br/atuacao/terras-indigenas/geoprocessamento-e-mapas>. Access in: 25 Feb. 2022. 33, 82, 140, 141, 172

FURTADO, A. M. M.; PONTE, F. C. Mapeamento de unidades de relevo do estado do Pará. **Revista GeoAmazônia**, v. 2, n. 2, p. 56–67, 2013. 34

GILLANDERS, S. N.; COOPS, N. C.; WULDER, M. A.; GERGEL, S. E.; NELSON, T. Multitemporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends. **Progress in Physical Geography**, v. 32, n. 5, p. 503–528, 2008. 23

GÓMEZ, C.; WHITE, J. C.; WULDER, M. A. Optical remotely sensed time series data for land cover classification: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 116, p. 55–72, 2016. 7

GONÇALVES, L.; FONTE, C. C.; JÚLIO, E. N. B. S.; CAETANO, M. Evaluation of remote sensing images classifiers with uncertainty measures. In: _____. **Spatial data quality from process to decisions**. Boca Raton: CRC Press, 2005. p. 163–177. 22

GOOGLE EARTH ENGINE (GEE). **ee.Algorithms.Landsat.simpleCloudScore**. 2022. Available from: <https://developers.google.com/earth-engine/apidocs/ee-algorithms-landsat-simplecloudscore>. Access in: 10 Mar. 2022. 112

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google earth engine: planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18–27, 2017. 17, 131

GREEN, G.; SCHWEIK, C.; HANSON, M. **Radiometric calibration of Landsat Multi-Spectral Scanner and Thematic Mapper images: guidelines for the Global Change Community**. [S.l.], 2002. 45

GRIFFITHS, P.; JAKIMOW, B.; HOSTERT, P. Reconstructing long term annual deforestation dynamics in Pará and Mato Grosso using the Landsat archive. **Remote Sensing of Environment**, v. 216, p. 497–513, 2018. 19

GRIFFITHS, P.; LINDEN, S. van der; KUEMMERLE, T.; HOSTERT, P. A pixel-based landsat compositing algorithm for large area land cover mapping. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 6, n. 5, p. 2088–2101, 2013. 17

GUHA, S.; RASTOGI, R.; SHIM, K. Rock: a robust clustering algorithm for categorical attributes. **Information Systems**, v. 25, n. 5, p. 345–366, 2000. 25

HANSEN, M. C.; LOVELAND, T. R. A review of large area monitoring of land cover change using Landsat data. **Remote Sensing of Environment**, v. 122, p. 66 – 74, 2012. Landsat Legacy Special Issue. 12

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. High-resolution global maps of 21st-century forest cover change. **Science**, v. 342, n. 6160, p. 850–853, 2013. 27

HECKENBERGER, M. J.; RUSSELL, J. C.; TONEY, J. R.; SCHMIDT, M. J. The legacy of cultural landscapes in the brazilian amazon: implications for biodiversity. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 362, n. 1478, p. 197–208, 2007. 1

HEINRICH, V. H. A.; SILVA, R. D. d.; CASSOL, H. L. G.; ROSAN, T. M.; ALMEIDA, C. T. d.; JÚNIOR, C. H. L. S.; CAMPANHARO, W. A.; HOUSE, J. I.; STICH, S.; HALES, T. C.; ADAMI, M.; ANDERSON, L. O.; ARAGÃO, L. E. O. e. C. d. Large carbon sink potential of secondary forests in the brazilian amazon to mitigate climate change. **Nature Communications**, v. 12, n. 1, p. e1785, 2021. 2

HERMOSILLA, T.; WULDER, M. A.; WHITE, J. C.; COOPS, N. C.; HOBART, G. W. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. **Remote Sensing of Environment**, v. 170, p. 121 – 132, 2015. 23

_____. Disturbance-informed annual land cover classification maps of Canada's forested ecosystems for a 29-year Landsat time series. **Canadian Journal of Remote Sensing**, v. 44, n. 1, p. 67–87, 2018. 20

HEROLD, M.; HUBALD, R.; DI GREGORIO, A. **Translating and evaluating land cover legends using the UN Land Cover Classfication System (LCCS). GOFC-GOLD Report No. 43**. Jena, Germany, 2009. 14

HEROLD, M.; WOODCOCK, C. E.; GREGORIO, A. D.; MAYAUX, P.; BELWARD, A. S.; LATHAM, J.; SCHMULLIUS, C. C. A joint initiative for harmonization and validation of land cover datasets. **IEEE Transactions on Geoscience and Remote Sensing**, v. 44, n. 7, p. 1719–1727, 2006. 13

HOLBEN, B. N. Characteristics of maximum-value composite images from temporal avhrr data. **International Journal of Remote Sensing**, v. 7, n. 11, p. 1417–1434, 1986. 17

HORNIK, K.; BUCHTA, C.; ZEILEIS, A. Open-source machine learning: R meets Weka. **Computational Statistics**, v. 24, n. 2, p. 225–232, 2009. 49

HSU, C.-C. Generalizing self-organizing map for categorical data. **IEEE transactions on Neural Networks**, v. 17, n. 2, p. 294–304, 2006. 25

HUANG, S.; RAMIREZ, C.; KENNEDY, K.; MALLORY, J.; WANG, J.; CHU, C. Updating land cover automatically based on change detection using satellite images: case study of national forests in Southern California. **GIScience & Remote Sensing**, v. 54, n. 4, p. 495–514, 2017. 18

HUANG, Z. Extensions to the k-means algorithm for clustering large data sets with categorical values. **Data Mining and Knowledge Discovery**, v. 2, n. 3, p. 283–304, 1998. 25

HUGHES, M. J.; KAYLOR, S. D.; HAYES, D. J. Patch-based forest change detection from Landsat time series. **Forests**, v. 8, n. 5, p. 166, 2017. 19

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Mapa de vegetação do Brasil**. 2004a. Available from: <https://mapas.ibge.gov.br/tematicos/vegetacao>. Access in: 19 Jan. 2019. 34

_____. **Vocabulário básico de recursos naturais e meio ambiente**. 2. ed. Rio de Janeiro, 2004b. Available from: <https://ww2.ibge.gov.br/home/presidencia/noticias/vocabulario.pdf>. Access in: 6 Mar. 2019. 9

_____. **Manual técnico de uso da terra**. 3. ed. Rio de Janeiro, 2013. 171p. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv81615.pdf>. Access in: 29 Aug. 2020. 14, 43

_____. **Mudanças na cobertura e uso da terra 2000-2010-2012**. Rio de Janeiro, 2015. 48p. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv94724.pdf>. Access in: 21 July 2020. 27

_____. **Cidades**. 2017a. Available from: <https://cidades.ibge.gov.br/>. Access in: 17 Jan. 2019. 38

_____. **Monitoramento da cobertura e uso da terra do Brasil 2000-2010-2012-2014**. Rio de Janeiro, 2017b. 35p. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101469.pdf>. Access in: 02 Oct. 2020. 27

_____. **Downloads**. 2019. Available from: <https://www.ibge.gov.br/geociencias/downloads-geociencias.html>. Access in: 25 Feb. 2022. 33, 44, 82, 107, 109, 141, 142, 155, 163, 165, 169, 170, 172, 219

_____. **Monitoramento da cobertura e uso da terra do Brasil**: 2016-2018. Rio de Janeiro, 2020. 26p. Available from: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101703.pdf>. Access in: 27 July 2020. 27

INSTITUTO BRASILEIRO DO MEIO AMBIENTE E DOS RECURSOS NATURAIS RENOVÁVEIS (IBAMA). **Floresta Nacional do Tapajós plano de manejo: volume I - informações gerais**. Brazil, 2004. 76p,. 34

INSTITUTO HOMEM E MEIO AMBIENTE DA AMAZÔNIA (IMAZON). **Monitoramento da Amazônia**. 2020. Available from: <https://imazon.org.br/>. Access in: 02 Oct. 2020. 27

INSTITUTO NACIONAL DE COLONIZAÇÃO E REFORMA AGRÁRIA (INCRA). **Acervo fundiário**: download shapefile. s.d. Available from: <https://certificacao.incra.gov.br/csv_shp/export_shp.py>. Access in: 25 Feb. 2022. 33, 82, 140, 141, 142, 172

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **Nota técnica - Correção da máscara do PRODES**. 2017. Available from: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes/pdfs/nt_deslocamentomascara.pdf>. Access in: 02 Oct. 2020. 28, 106

_____. **TerraBrasilis catalogue**: Limites dos Estados da Amazônia Legal. 2019. Available from: <http://terrabrasilis.dpi.inpe.br/downloads/>. Access in: 25 Feb. 2022. 33

_____. **CBERS: Satélite Sino-Brasileiro de Recursos Terrestres**. 2020. Available from: <http://www.cbers.inpe.br/>. Access in: 18 Dec. 2020. 15

_____. **DETER e DETER intenso**. 2020. Available from: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/deter>. Access in: 02 Oct. 2020. 27

_____. **Monitoramento do desmatamento da floresta amazônica brasileira por satélite**. 2020. Available from: <http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes>. Access in: 02 Oct. 2020. 26

_____. **Satélites**. 2020. Available from: <http://www.dgi.inpe.br/documentacao/satelites>. Access in: 02 Oct. 2020. 15

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). TERRABRASILIS. **PRODES (Desmatamentos)**: taxas de desmatamento - Amazônia Legal - estados. 2022. Available from: <http://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/legal_amazon/rates>. Access in: 04 Aug. 2022. 153

INSTITUTO SOCIOAMBIENTAL (ISA). **Terras indígenas do Brasil**. 2020. Available from: <https://terrasindigenas.org.br/>. Access in: 07 Mar. 2022. 141

JAKIMOW, B.; LINDEN, S. van der; THIEL, F.; FRANTZ, D.; HOSTERT, P. Visualizing and labeling dense multi-sensor earth observation time series: the EO Time Series Viewer. **Environmental Modelling & Software**, v. 125, p. 104631, 2020. 21

JAKOVAC, C. C.; DUTRIEUX, L. P.; SITI, L.; PEÑA-CLAROS, M.; BONGERS, F. Spatial and temporal dynamics of shifting cultivation in the middle-amazonas river: expansion and intensification. **PloS One**, v. 12, n. 7, p. e0181092, 2017. 12, 23, 24, 175

JAKOVAC, C. C.; JUNQUEIRA, A. B.; CROUZEILLES, R.; PEñA-CLAROS, M.; MESQUITA, R. C. G.; BONGERS, F. The role of land-use history in driving successional pathways and its implications for the restoration of tropical forests. **Biological Reviews**, v. 96, n. 4, p. 1114–1134, 2021. 2, 164

JANSEN, L. J.; GROOM, G.; CARRAI, G. Land-cover harmonisation and semantic similarity: some methodological issues. **Journal of Land Use Science**, v. 3, n. 2-3, p. 131–160, 2008. 75

JARRON, L. R.; HERMOSILLA, T.; COOPS, N. C.; WULDER, M. A.; WHITE, J. C.; HOBART, G. W.; LECKIE, D. G. Differentiation of alternate harvesting practices using annual time series of Landsat data. **Forests**, v. 8, n. 1, 2017. 74

JENSEN, J. R. **Remote sensing of the environment: an earth resource perspective**. [S.l.]: Prentice Hall: Upper Saddle River, NJ, 2007. 101

JUNQUEIRA, A. B.; SHEPARD, G. H.; CLEMENT, C. R. Secondary forests on anthropogenic soils of the middle Madeira river: valuation, local knowledge, and landscape domestication in Brazilian Amazonia. **Economic Botany**, v. 65, n. 1, p. 85–99, 2011. 1

KAMMESHEIDT, L. Perspectives on secondary forest management in tropical humid lowland America. **AMBIO: A Journal of the Human Environment**, v. 31, n. 3, p. 243–250, 2002. 9

KENNEDY, R. E.; YANG, Z.; COHEN, W. B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr–temporal segmentation algorithms. **Remote Sensing of Environment**, v. 114, n. 12, p. 2897–2910, 2010. 18

KOSMIDOU, V.; PETROU, Z.; BUNCE, R. G.; MüCHER, C. A.; JONGMAN, R. H.; BOGERS, M. M.; LUCAS, R. M.; TOMASELLI, V.; BLONDA, P.; PADOA-SCHIOPPA, E.; MANAKOS, I.; PETROU, M. Harmonization of the Land Cover Classification System (LCCS) with the General Habitat Categories (GHC) classification system. **Ecological Indicators**, v. 36, p. 290 – 300, 2014. 14

LI, G.; LU, D.; MORAN, E.; DUTRA, L.; BATISTELLA, M. A comparative analysis of ALOS PALSAR L-band and RADARSAT-2 C-band data for land-cover classification in a tropical moist region. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 70, p. 26 – 38, 2012. 101

LIU, H.; ZHOU, Q. Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. **International Journal of Remote Sensing**, v. 25, n. 5, p. 1037–1050, 2004. 22

LOBO, F.; ESCADA, M.; ALMEIDA, C. Análise do desflorestamento no distrito florestal sustentável da br-163, PA. In: ALBERNAZ, A. L. (Ed.). **Distrito Florestal Sustentável da BR-163**: dinâmicas sociais, mudanças ambientais e produção florestal. Belém: Museu Paraense Emílio Goeldi, 2015. 464 p. 3

LU, D.; LI, G.; MORAN, E. Current situation and needs of change detection techniques. **International Journal of Image and Data Fusion**, v. 5, n. 1, p. 13–38, 2014. 8, 20

LU, D.; MAUSEL, P.; BRONDÍZIO, E.; MORAN, E. Classification of successional forest stages in the Brazilian Amazon basin. **Forest Ecology and Management**, v. 181, n. 3, p. 301 – 312, 2003. ISSN 0378-1127. Available from: <http://www.sciencedirect.com/science/article/pii/S0378112703000033>. 52, 74

LU, D.; MAUSEL, P.; BRONDIZIO, E.; MORAN, E. Change detection techniques. **International Journal of Remote Sensing**, v. 25, n. 12, p. 2365–2401, 2004. 17, 18

MACIEL, A. L.; ALVES, D.; SANT'ANNA, S. J. S. Análise dos processos de desmatamento e regeneração nas Unidades de Conservação pertencentes no Vale do Ribeira (SP/PR) por meio de classificações de imagens de sensoriamento remoto. **Revista Brasileira de Cartografia**, v. 73, n. 1, p. 261–277, 2021. ISSN 0560-4613 and 1808-0936. 181

MACIEL, A. M. **Spatiotemporal interval logic for reasoning about land use change dynamics**. 2017. 131 p. Thesis (PhD in Applied Computing) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017. Available from: <http://urlib.net/rep/8JMKD3MGP3W34P/3Q44DE8>. Access in: 30 Mar. 2019. 2

MACIEL, A. M.; CAMARA, G.; VINHAS, L.; PICOLI, M. C. A.; BEGOTTI, R. A.; ASSIS, L. F. F. G. A spatiotemporal calculus for reasoning about land-use trajectories. **International Journal of Geographical Information Science**, v. 33, n. 1, p. 176–192, 2019. 25, 26

MAPBIOMAS. **Deforestation and Secondary Vegetation**: appendix - collection 6 version 1. 2022. Available from: <https://mapbiomas-br-site.s3.amazonaws.com/Metodologia/Deforestation___Secondary_Vegetation_Appendix_-_ATBD_Collection_6_v2.docx.pdf>. Access in: 22 Feb. 2022. 29

_____. **MapBiomas general "handbook"**: Theoretical Basis Document (ATBD) - Collection 6. 2022. Available from: <https://mapbiomas-br-site.s3.amazonaws.com/Metodologia/ATBD_Collection_6_v1_January_2022.pdf>. Access in: 22 Feb. 2022. 29

_____. **Plataforma MapBiomas**: fogo. 2022. Available from: <https://plataforma.brasil.mapbiomas.org/>. Access in: 04 Aug. 2022. 157, 159

MAURANO, L. E. P.; ESCADA, M. I. S. Comparação dos dados produzidos pelo PRODES versus dados do Mapbiomas para o bioma Amazônia. In: GHERARDI, D. F. M.; SANCHES, I. D.; ARAGÃO, L. E. O. e. C. d. (Ed.). **Anais do XIX Simposio Brasileiro de Sensoriamento Remoto**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2019. p. 735–738. ISBN 978-85-17-00097-3. Available from: <http://urlib.net/rep/8JMKD3MGP6W34M/3U24J5L>. Access in: 01 July 2020. 30

MAUS, V.; CÂMARA, G.; CARTAXO, R.; SANCHEZ, A.; RAMOS, F. M.; QUEIROZ, G. R. D. 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**, v. 9, n. 8, p. 3729–3739, 2016. 19

MAUS, V. W.; CAMARA, G.; APPEL, M.; PEBESMA, E. dtwsat: time-weighted dynamic time warping for satellite image time series analysis in r. **Journal of Statistical Software**, v. 88, n. 5, p. 1–31, 2019. 19

MCCONNELL, W.; MORAN, E. **Meeting in the middle: the challenge of meso-level integration.**: LUCC report series no. 5. Rome, Italy, 2001. Available from: <http://www.comap.ca/kmland/getdocument.php?F=310836083_4c7195ce74a518.86955292>. Access in: 30 July 2020. 13, 14

MELLO, A. Y. I.; ALVES, D. S. Secondary vegetation dynamics in the Brazilian Amazon based on Thematic Mapper imagery. **Remote Sensing Letters**, v. 2, n. 3, p. 189–194, 2011. 12

MELLO, N. A. **Políticas territorias na Amazônia**. São Paulo: Annablume, 2006. 412 p. 36, 37, 39, 40, 41

MELLO, N. G. a. R.; ARTAXO, P. Evolução do Plano de Ação para Prevenção e Controle do Desmatamento na Amazônia Legal. **Revista do Instituto de Estudos Brasileiros**, p. 108 – 129, 04 2017. ISSN 0020-3874. Available from: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0020-38742017000100108&nrm=iso>. 11, 41

MENA, C. F. Trajectories of land-use and land-cover in the Northern Ecuadorian Amazon. **Photogrammetric Engineering & Remote Sensing**, v. 74, n. 6, p. 737–751, 2008. 3, 23

MERTENS, B.; LAMBIN, E. F. Land-cover-change trajectories in southern Cameroon. **Annals of the association of American Geographers**, v. 90, n. 3, p. 467–494, 2000. 3

MEYFROIDT, P.; LAMBIN, E. F. Global forest transition: prospects for an end to deforestation. **Annual Review of Environment and Resources**, v. 36, n. 1, p. 343–371, 2011. 2

MIRANDA, M. Colonização oficial da Amazônia: o caso de altamira. In: BECKER, B. B.; MACHADO, L. O.; MIRANDA, M. (Ed.). **Fronteira amazônica**: questões sobre a gestão do território. Brasília/Rio de Janeiro: UnB/UFRJ, 1990. 219 p. 39

MORAN, E. F.; BRONDIZIO, E. S.; TUCKER, J. M.; SILVA-FORSBERG, M. C. da; MCCRACKEN, S.; FALESI, I. Effects of soil fertility and land-use on forest succession in Amazonia. **Forest Ecology and Management**, v. 139, n. 1, p. 93–108, 2000. 2

MÜLLER, H.; RUFIN, P.; GRIFFITHS, P.; HISSA, L. d. B. V.; HOSTERT, P. Beyond deforestation: differences in long-term regrowth dynamics across land use regimes in southern Amazonia. **Remote Sensing of Environment**, Elsevier, v. 186, p. 652–662, 2016. 23, 108, 133, 175

MÜLLER-HANSEN, F.; CARDOSO, M. F.; DALLA-NORA, E. L.; DONGES, J. F.; HEITZIG, J.; KURTHS, J.; THONICKE, K. A matrix clustering method to explore patterns of land-cover transitions in satellite-derived maps of the Brazilian Amazon. **Nonlinear Processes in Geophysics**, v. 24, n. 1, p. 113–123, 2017. 23

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION (NASA). **MODIS**: Moderate Resolution Imaging Spectroradiometer. 2020. Available from: <https://modis.gsfc.nasa.gov/>. Access in: 18 Dec. 2020. 15

_____. **Terra**: the EOS Flagship. 2020. Available from: <https://terra.nasa.gov/>. Access in: 18 Dec. 2020. 15

NEGRI, R. G. **Avaliação de dados polarimétricos do sensor ALOS PALSAR para classificação da cobertura da terra da Amazônia**. 2009. 170 p. Thesis (Master in Applied Computing) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2009. 48, 75

NEVES, A. K.; KÖRTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S. Assessment of TerraClass and MapBiomas data on legend and map agreement for the Brazilian Amazon biome. **Acta Amazonica**, v. 50, n. 2, p. 170–182, 2020. 29

NG, A. Y.; JORDAN, M. I. On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. In: DIETTERICH, T.; BECKER, S.; GHAHRAMANI, Z. (Ed.). **Advances in Neural Information Processing Systems**. Cambridge: MIT Press, 2001. p. 841–848. 78, 79

NUNES, S.; OLIVEIRA JUNIOR, L.; SIQUEIRA, J.; MORTON, D. C.; SOUZA JUNIOR, C. M. Unmasking secondary vegetation dynamics in the Brazilian Amazon. **Environmental Research Letters**, v. 15, n. 3, p. 034057, 2020. 12, 23, 29

OLOFSSON, P.; FOODY, G. M.; HEROLD, M.; STEHMAN, S. V.; WOODCOCK, C. E.; WULDER, M. A. Good practices for estimating area and assessing accuracy of land change. **Remote Sensing of Environment**, Elsevier, v. 148, p. 42–57, 2014. 21, 117, 132

OLTHOF, I.; BUTSON, C.; FRASER, R. Signature extension through space for northern landcover classification: a comparison of radiometric correction methods. **Remote Sensing of Environment**, v. 95, n. 3, p. 290 – 302, 2005. 105

OVIEDO, A. F. P.; DOBLAS, J. **As florestas precisam das pessoas**. [S.l.], 2022. Available from: <https://acervo.socioambiental.org/acervo/documentos/florestas-precisam-das-pessoas>. Access in: 15 Oct 2022. 136, 174, 176

PANTALEÃO, E.; DUTRA, L. V.; SANDRI, S. Scenario analysis for image classification using multi-objective optimization. **INFOCOMP Journal of Computer Science**, v. 11, n. 3-4, p. 15–22, 2012. 13, 75

PARÁ. Secretaria de meio ambiente e sustentabilidade. instrução Normativa no 08, de 28 de outubro de 2015. Define procedimentos administrativos para a realização de limpeza e autorização de supressão, a serem realizadas nas áreas de vegetação secundária em estágio inicial de regeneração, localizadas fora da Reserva Legal e da área de Preservação Permanente - APP dos imóveis rurais, no âmbito do Estado do Pará e dá outras providências. **DOE 33.003**, 2015. Available from: <https://www.semas.pa.gov.br/2015/11/03/instrucao-normativa-no-08-de-28-de-outubro-de-2015/>. Access in: 21 Mar. 2021. 2, 12, 183

PARADELLA, W. R.; SANTOS, A. R. dos; VENEZIANI, P.; CUNHA, E. S. P. Radares imageadores nas Geociências: estado da arte e perspectivas. **Revista Brasileira de Cartografia**, v. 57, n. 1, 2005. 16

PAULA, D. S.; ESCADA, M. I. S.; ORTIZ, J. O. Análise multitemporal do uso e cobertura da terra na Amazônia: a expansão da agricultura de larga escala na bacia do Rio Curuá-Una. **Revista Brasileira de Cartografia**, v. 74, n. 2, p. 399–417, 2022. Available from: <http://dx.doi.org/10.14393/rbcv74n2-63206>. Access in: 04 ago. 2022. 168, 174

PEREIRA, L. O.; FREITAS, C. C.; SANT'ANNA, S. J. S.; REIS, M. S. Alos/palsar data evaluation for land use and land cover mapping in the Amazon region. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 12, p. 5413–5423, 2016. 13, 101

_____. Evaluation of optical and radar images integration methods for LULC classification in Amazon region. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 11, n. 9, p. 3062–3074, 2018. 13

PERZ, S. G.; SKOLE, D. L. Secondary forest expansion in the Brazilian Amazon and the refinement of forest transition theory. **Society & Natural Resources**, v. 16, n. 4, p. 277–294, 2003. 11

_____. Social determinants of secondary forests in the Brazilian Amazon. **Social Science Research**, v. 32, n. 1, p. 25–60, 2003. 174

PINHEIRO, T. F.; ESCADA, M. I. S.; VALERIANO, D. M.; HOSTERT, P.; GOLLNOW, F.; MÜLLER, H. Forest degradation associated with logging frontier expansion in the Amazon: the BR-163 region in Southwestern Pará, Brazil. **Earth Interactions**, v. 20, n. 17, p. 1–26, 2016. 24, 59

POWELL, R. L.; MATZKE, N.; JUNIOR SOUZA, C.; CLARK, M.; NUMATA, I.; HESS, L. L.; ROBERTS, D. A. Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. **Remote Sensing of Environment**, v. 90, n. 2, p. 221–234, 2004. 21, 132

PRATIHAST, A. K.; DEVRIES, B.; AVITABILE, V.; BRUIN, S.; KOOISTRA, L.; TEKLE, M.; HEROLD, M. Combining satellite data and community-based observations for forest monitoring. **Forests**, v. 5, n. 10, p. 2464–2489, 2014. 74

PUTZ, F. E.; REDFORD, K. H. The importance of defining 'forest': tropical forest degradation, deforestation, long-term phase shifts, and further transitions. **Biotropica**, v. 42, n. 1, p. 10–20, 2010. 9, 12, 51

RADAMBRASIL, P. **Folha SA.21 Santarém**: geologia, geomorfologia, pedologia, vegetação e uso potencial da terra. Rio de Janeiro: IBGE, 1976. 507p. 34, 108

RAMANKUTTY, N.; GIBBS, H. K.; ACHARD, F.; DEFRIES, R.; FOLEY, J. A.; HOUGHTON, R. A. Challenges to estimating carbon emissions from tropical deforestation. **Global Change Biology**, v. 13, n. 1, p. 51–66, 2007. 1

RASCLASS: supervised raster image classification (version 0.2.2). 2016. Available from: <https://cran.r-project.org/web/packages/rasclass/index.html>. 49

REIS, M. S.; DUTRA, L. V.; ESCADA, M. I. S.; SANT'ANNA, S. J. S. Avoiding invalid transitions in land cover trajectory classification with a Compound Maximum a Posteriori approach. **IEEE Access**, v. 8, p. 98787–98799, 2020a. 18, 22, 77

REIS, M. S.; DUTRA, L. V.; SANT'ANNA, S. J. S.; ESCADA, M. I. S. Análise das incertezas envolvidas em classificação multi-legendas da cobertura da terra com suporte de simulação Monte Carlo. **Revista Brasileira de Cartografia**, v. 69, n. 9, 2017a. 22, 133

_____. Examining multi-legend change detection in amazon with pixel and region based methods. **Remote Sensing**, v. 9, n. 1, 2017b. ISSN 2072-4292. 49, 62, 75, 219

_____. Multi-source change detection with PALSAR data in the Southern of Pará state in the Brazilian Amazon. **International Journal of Applied Earth Observation and Geoinformation**, v. 84, p. 101945, 2020c. 90, 101

REIS, M. S.; ESCADA, M. I. S.; DUTRA, L. V.; SANT'ANNA, S. J.; VOGT, N. D. Towards a reproducible LULC hierarchical class legend for use in the Southwest of Pará State, Brazil: a comparison with remote sensing data-driven hierarchies. **Land**, v. 7, n. 2, p. 65, 2018. 43

REIS, M. S.; ESCADA, M. I. S.; SANT'ANNA, S. J. S.; DUTRA, L. V. Harmonização de legendas formalizadas em Land Cover Meta Language-LCML. In: GHERARDI, D. F. M.; ARAGÃO, L. E. O. C. (Ed.). **Anais do Simpósio Brasileiro de Sensoriamento Remoto.** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017c. p. 863–870. ISBN 978-85-17-00088-1. Available from: <http://urlib.net/sid.inpe.br/marte2/2017/10.23.19.25.18>. Access in: 08 Feb. 2018. 75

_____. Métodos de classificação e análise de trajetórias de uso e cobertura da terra na Amazônia: Implicações para estudos de regeneração florestal. **Revista Brasileira de Cartografia**, v. 72, n. ed. esp. 50 anos, p. 1087–1113, 2020b. ISSN 0560-4613 and 1808-0936. 1, 7, 136

REIS, M. S.; PANTALEÃO, E.; DUTRA, L. V.; SANT'ANNA, S. J. S.; ESCADA, M. I. S. Effects of different methods of radiometric calibration on the use of training data for supervised classification of Landsat5/TM images from other dates. In: IGARSS 2019 - 2019 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2019, Yokohama, Japan. **Proceedings...** [S.l.]: IEEE, 2019. p. 1566–1569. 103, 131

RICHARDS, J. A.; JIA, X. **Remote sensing digital image analysis**: an introduction. 4. ed. Germany: Springer, 2006. 223 p. 16

ROESSEL, J. W. V.; GODOY, R. C. SLAR mosaics for project RADAM. **Photogrammetric Engineering**, p. 583–595, 1974. 27

ROMERO-SANCHEZ, M. E.; PONCE-HERNANDEZ, R. Assessing and monitoring forest degradation in a deciduous tropical forest in Mexico via remote sensing indicators. **Forests**, v. 8, n. 9, 2017. 74

RORATO, A. C.; CAMARA, G.; ESCADA, M. I. S.; PICOLI, M. C. A.; MOREIRA, T.; VERSTEGEN, J. A. Brazilian amazon indigenous peoples threatened by mining bill. **Environmental Research Letters**, v. 15, n. 10, p. 1040a3, 2020. Available from: <https://doi.org/10.1088/1748-9326/abb428>. 1

RUFIN, P.; MÜLLER, H.; PFLUGMACHER, D.; HOSTERT, P. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. **International Journal of Applied Earth Observation and Geoinformation**, v. 41, p. 1–10, 2015. 17, 108, 131

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. Benchmark map of forest carbon stocks in tropical regions across three continents. **Proceedings of the National Academy of Sciences**, v. 108, n. 24, p. 9899–9904, 2011. 1

SALOMÃO, R. P.; VIEIRA, I. C. G.; BRIENZA JÚNIOR, S.; AMARAL, D. D.; SANTANA, A. C. Sistema capoeira classe: uma proposta de sistema de classificação de estágios sucessionais de florestas secundárias para o estado do Pará. **Boletim do Museu Paraense Emílio Goeldi. Ciências Naturais**, v. 7, n. 3, p. 297–317, 2012. 9, 52, 74

SANCHEZ, A. H.; PICOLI, M. C. A.; CAMARA, G.; ANDRADE, P. R.; CHAVES, M. E. D.; LECHLER, S.; SOARES, A. R.; MARUJO, R. F.; SIMÕES, R. E. O.; FERREIRA, K. R.; QUEIROZ, G. R. Comparison of cloud cover detection algorithms on Sentinel–2 images of the Amazon tropical forest. **Remote Sensing**, v. 12, n. 8, p. 1284, 2020. 17

SANTOS, A. C. **Uso do território e dinâmicas territoriais na região da rodovia Santarém-Curuá-Una (PA 370): o exemplo da realidade da comunidade Boa Esperança**. 2020. 253 p. Thesis (PhD in Environmental Sciences) — Universidade Federal do Oeste do Pará, Santarém, 2020. Available from: <https://repositorio.ufopa.edu.br/jspui/handle/123456789/357>. Access in: 19 June 2022. 36, 38, 41, 174

SANTOS, R. A. O. **Hitória econômica da Amazônia (1800-1920)**. São Paulo: T.A. Queiroz, 1980. 358 p. 35, 38

SCHMINK, M.; WOOD, C. **Contested frontiers in Amazonia**. New York: Columbia University Press, 1992. 387 p. 39

SCHOWENGERDT, R. **Remote sensing**: models and methods for image processing. 3. ed. USA: Academic Press, 2006. 560 p. 48, 90

SCIENCE PANEL FOR THE AMAZON (SPA). **Executive summary of the Amazon assessment report 2021**. New York, USA: United Nations Sustainable Development Solutions Network, 2021. 48 p. 1

SEXTON, J. O.; URBAN, D. L.; DONOHUE, M. J.; SONG, C. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. **Remote Sensing of Environment**, v. 128, p. 246–258, 2013. 21, 132

SHIMABUKURO, Y. E.; SMITH, J. A. Fraction images derived from Landsat TM and MSS data for monitoring reforested areas. **Canadian Journal of Remote Sensing**, v. 21, n. 1, p. 67–74, 1995. 17

SHIMADA, M.; ITOH, T.; MOTOOKA, T.; WATANABE, M.; SHIRAISHI, T.; THAPA, R.; LUCAS, R. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). **Remote Sensing of Environment**, v. 155, p. 13–31, 2014. 16

SILVA JUNIOR, C. H. L.; HEINRICH, V. H.; FREIRE, A. T.; BROGGIO, I. S.; ROSAN, T. M.; DOBLAS, J.; ANDERSON, L. O.; ROUSSEAU, G. X.; SHIMABUKURO, Y. E.; SILVA, C. A.; HOUSE, J. I.; AO, E. O. C. A. Benchmark maps of 33 years of secondary forest age for Brazil. **Scientific Data**, v. 7, n. 1, p. 1–9, 2020. 9, 10, 12, 23, 29, 30, 219

SILVA, W. B.; PEREIRA, L. O.; SANT'ANNA, S. J. S.; FREITAS, C. C.; GUIMARÃES, R. J. P. S.; FRERY, A. C. Land cover discrimination at Brazilian Amazon using region based classifier and stochastic distance. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS), 2011 IEEE INTERNATIONAL, 2011, Vancouver, Canada. **Proceedings...** [S.l.]: IEEE, 2011. p. 2900–2903. 13

SOUZA, A. R. **Economia e natureza: padrões de uso e cobertura da terra associados a atividades agropecuárias e extrativistas de comunidades do sudoeste do Pará**. 2017. 224 p. Thesis (Master in Remote Sensing) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2017. Available from: <http://urlib.net/ibi/8JMKD3MGP3W34P/3M8SD78>. Access in: 31 July 2022. 138, 140, 147, 169, 170, 174

SOUZA, A. R. d.; ESCADA, M. I. S.; MARUJO, R. d. F. B.; MONTEIRO, A. M. V. Cartografia do invisível: revelando a agricultura de pequena escala com imagens Rapideye na região do baixo Tocantins, PA. **Revista do Departamento de Geografia**, v. 38, p. 137–153, 2019. 28

SOUZA, C. M.; SHIMBO, J. Z.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A. A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; FERREIRA, L. G.; SOUZA-FILHO, P. W. M.; OLIVEIRA, S. W. de; ROCHA, W. F.; FONSECA, A. V.; MARQUES, C. B.; DINIZ, C. G.; COSTA, D.; MONTEIRO, D.; ROSA, E. R.; VÉLEZ-MARTIN, E.; WEBER, E. J.; LENTI, F. E. B.; PATERNOST, F. F.; PAREYN, F. G. C.; SIQUEIRA, J. V.; VIERA, J. L.; NETO, L. C. F.; SARAIVA, M. M.; SALES, M. H.; SALGADO, M. P. G.; VASCONCELOS, R.; GALANO, S.; MESQUITA, V. V.; AZEVEDO, T. Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat Archive and Earth Engine. **Remote Sensing**, v. 12, n. 17, p. 2735, 2020. 17, 19, 28, 112, 131, 133

STEININGER, M. K. Secondary forest structure and biomass following short and extended land-use in central and southern Amazonia. **Journal of Tropical Ecology**, p. 689–708, 2000. 2, 12

STRAND, J.; SOARES-FILHO, B.; COSTA, M. H.; OLIVEIRA, U.; RIBEIRO, S. C.; PIRES, G. F.; OLIVEIRA, A.; RAJAO, R.; MAY, P.; HOFF, R. van der et al. Spatially explicit valuation of the Brazilian Amazon forest's ecosystem services. **Nature Sustainability**, v. 1, n. 11, p. 657–664, 2018. 1

TEWKESBURY, A. P.; COMBER, A. J.; TATE, N. J.; LAMB, A.; FISHER, P. F. A critical synthesis of remotely sensed optical image change detection techniques. **Remote Sensing of Environment**, v. 160, p. 1–14, 2015. 18

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern recognition**. 4. ed. San Diego: Academic Press, 2009. 961 p. 79

THERY, H. Situações da Amazônia no Brasil e no continente. **Estudos Avançados**, v. 19, p. 37 – 49, 04 2005. ISSN 0103-4014. 41

TORRES, L.; SANT'ANNA, S. J.; FREITAS, C. C.; FRERY, A. C. Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means. **Pattern Recognition**, v. 47, n. 1, p. 141–157, 2014. 90

TRITSCH, I.; SIST, P.; NARVAES, I. d. S.; MAZZEI, L.; BLANC, L.; BOURGOIN, C.; CORNU, G.; GOND, V. Multiple patterns of forest disturbance and logging shape forest landscapes in Paragominas, Brazil. **Forests**, v. 7, n. 12, 2016. 74

TURNER, B.; MEYER, W. Global land-use and land-cover change: an overview. In: MEYER, W. B.; TURNER, B. L. (Ed.). **Changes in land use and land cover**: a global perspective. Cambridge: Cambridge University Press, 1994. 537 p. 3

UDDIN, K.; SHRESTHA, H. L.; MURTHY, M.; BAJRACHARYA, B.; SHRESTHA, B.; GILANI, H.; PRADHAN, S.; DANGOL, B. Development of 2010 national land cover database for the Nepal. **Journal of Environmental Management**, v. 148, p. 82 – 90, 2015. 14

UNITED STATES GEOLOGICAL SURVEY (USGS). **USGS EROS Archive - Products Overview**. 2020. Available from: <https://www.usgs.gov/centers/eros/science/usgs-eros-archive-products-overview?qt-science_center_objects=0#qt-science_center_object>. Access in: 18 Dec. 2020. 15

_____. **Landsat missions**. 2022. Available from: <https://www.usgs.gov/landsat-missions>. Access in: 18 June 2020. 107, 108, 132

VERBESSELT, J.; HYNDMAN, R.; NEWNHAM, G.; CULVENOR, D. Detecting trend and seasonal changes in satellite image time series. **Remote Sensing of Environment**, v. 114, n. 1, p. 106–115, 2010. 24

VIEIRA, I. C. G.; ALMEIDA, A. S. de; DAVIDSON, E. A.; STONE, T. A.; CARVALHO, C. J. R. de; GUERRERO, J. B. Classifying successional forests using landsat spectral properties and ecological characteristics in eastern Amazônia. **Remote Sensing of Environment**, v. 87, n. 4, p. 470 – 481, 2003. Available from: <http://www.sciencedirect.com/science/article/pii/S0034425703002062>. 52, 74

VIEIRA, I. C. G.; GARDNER, T.; FERREIRA, J.; LEES, A. C.; BARLOW, J. Challenges of governing second-growth forests: a case study from the Brazilian Amazonian State of Pará. **Forests**, v. 5, n. 7, p. 1737–1752, 2014. 2, 9, 11

WANDELLI, E. V.; FEARNSIDE, P. M. Secondary vegetation in central Amazonia: land-use history effects on aboveground biomass. **Forest Ecology and Management**, v. 347, p. 140–148, 2015. 176

WANG, Y.; ZIV, G.; ADAMI, M.; ALMEIDA, C. A. de; ANTUNES, J. F. G.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; GOMES, A. R.; GALBRAITH, D. Upturn in secondary forest clearing buffers primary forest loss in the Brazilian Amazon. **Nature Sustainability**, v. 3, n. 4, p. 290–295, 2020. 9, 12, 24, 28, 106

WEINSTEIN, B. **A borracha na Amazônia**: expansão e decadência (1850-1920). São Paulo: HUCITEC, Editora da Universidade de São Paulo, 1993. 371 p. 35, 38

WHITE, J.; WULDER, M.; HOBART, G.; LUTHER, J.; HERMOSILLA, T.; GRIFFITHS, P.; COOPS, N.; HALL, R.; HOSTERT, P.; DYK, A.; GUINDON, L. Pixel-based image compositing for large-area dense time series applications and science. **Canadian Journal of Remote Sensing**, v. 40, n. 3, p. 192–212, 2014. 17

WITTEN, I. H.; FRANK, E. **Data mining: practical Machine Learning tools and techniques**. 2. ed. San Francisco: Morgan Kaufmann, 2005. 49

WOODCOCK, C. E.; MACOMBER, S. A.; PAX-LENNEY, M.; COHEN, W. B. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. **Remote Sensing of Environment**, v. 78, n. 1, p. 194 – 203, 2001. Landsat 7. 103, 105

WULDER, M. A.; COOPS, N. C.; ROY, D. P.; WHITE, J. C.; HERMOSILLA, T. Land cover 2.0. **International Journal of Remote Sensing**, v. 39, n. 12, p. 4254–4284, 2018. 3, 7, 16, 22, 113, 131

XIAN, G.; HOMER, C. Updating the 2001 National Land Cover Database impervious surface products to 2006 using Landsat imagery change detection methods. **Remote Sensing of Environment**, Elsevier, v. 114, n. 8, p. 1676–1686, 2010. 18

YUAN, F.; SAWAYA, K. E.; LOEFFELHOLZ, B. C.; BAUER, M. E. Land cover classification and change analysis of the Twin Cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. **Remote Sensing of Environment**, v. 98, n. 2-3, p. 317–328, 2005. 23

ZADEH, L. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338 – 353, 1965. Available from: <http://www.sciencedirect.com/science/article/pii/S001999586590241X>. 74

ZHU, Z. Change detection using Landsat time series: a review of frequencies, preprocessing, algorithms, and applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 370 – 384, 2017. 7, 17, 20

ZHU, Z.; FU, Y.; WOODCOCK, C. E.; OLOFSSON, P.; VOGELMANN, J. E.; HOLDEN, C.; WANG, M.; DAI, S.; YU, Y. Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: a case study from Guangzhou, China (2000-2014). **Remote Sensing of Environment**, v. 185, p. 243 – 257, 2016. Landsat 8 Science Results. 22

ZHU, Z.; WOODCOCK, C. E. Continuous change detection and classification of land cover using all available landsat data. **Remote Sensing of Environment**, v. 144, p. 152 – 171, 2014. 18

APPENDIX A – FIELD DATA INFORMATION

Table A.1 - Field expeditions information.

Date	Team	Collected data
September 2009	Raimundo Oliveira dos Santos, Ramirez de Aguiar Souza, Rogério Galante Negri, Sidnei João Siqueira Sant'Anna,	Notes regarding LULC class identification and georeferenced photos (ground level)
September 2010	Elivan Oliveira dos Santos ("Galo"), Luciana de Oliveira Pereira, Luciano Vieira Dutra , Ramirez de Aguiar Souza, Rogério Galante Negri	Notes regarding LULC class identification and georeferenced photos (ground level)
August 2013	Daniela Souza dos Anjos, Felipe Rodolfo Santos Corrêa, Joventino Santos ("Jovem"), Mariane Souza Reis, Nathan David Vogt, Paulo Cesar Gurgel de Albuquerque, Sidnei João Siqueira Sant'Anna	Standardized notes on LULC class identification/estimates of height and coverage of LULC elements, trees height measurement (randomly selected), and georeferenced photos (ground level)
March 2015	Bruna Cristina Braga, Jony Martins Oliveira ("Peu") Maria Antônia Falcão de Oliveira, Mariane Souza Reis, Paulo Cesar Gurgel de Albuquerque, Sidnei João Siqueira Sant'Anna	Standardized notes on LULC class identification/estimates of height and coverage of LULC elements, non-georeferenced photos (ground level) and notes about photos coordinates
August-September 2016	Bruna Cristina Braga, Jony Martins Oliveira ("Peu"), Maria Antônia Falcão de Oliveira, Mariane Souza Reis, Noeli Aline Particcelli Moreira, Sidnei João Siqueira Sant'Anna, Paulo Cesar Gurgel de Albuquerque, Sidnei João Siqueira Sant'Anna	Standardized notes on LULC class identification/estimates of height and coverage of LULC elements and georeferenced photos (ground level and aerial/drone)
September 2017	Bruna Cristina Braga, Mariane Souza Reis, Paulo Cesar Gurgel de Albuquerque, Rebeca Suely Gabriella Soares Carneiro, Sidnei João Siqueira Sant'Anna, Tiago Martins Oliveira	Standardized notes on LULC class identification/estimates of height and coverage of LULC elements and georeferenced photos (ground level and aerial/drone)

SOURCE: Based on Dutra et al. (2021).

APPENDIX B – SUPPLEMENTAL INFORMATION FOR CHAPTER 6

B.1 Classification mask

Each channel of the yearly Landsat mosaics (images from June-November) was normalized to mean=127 and standard deviation=42. These normalized mosaics were segmented using the Multiresolution Segmentation algorithm within eCognition Developer version 8 (DEFINIENS, 2009), with the parameters Scale=30, Shape=0.3, and Compactness=0.5. These parameters were previously selected by Reis et al. (2017b) and were considered suitable for this data set after a visual inspection of the results. The mask includes pixels/segments labeled as:

a) *Aquatic or regularly flooded* (1984-2020): segments mostly covered by water in at least one year between 1984 and 2020. This analysis was made yearly, with each year inhering the mask from the previous year;

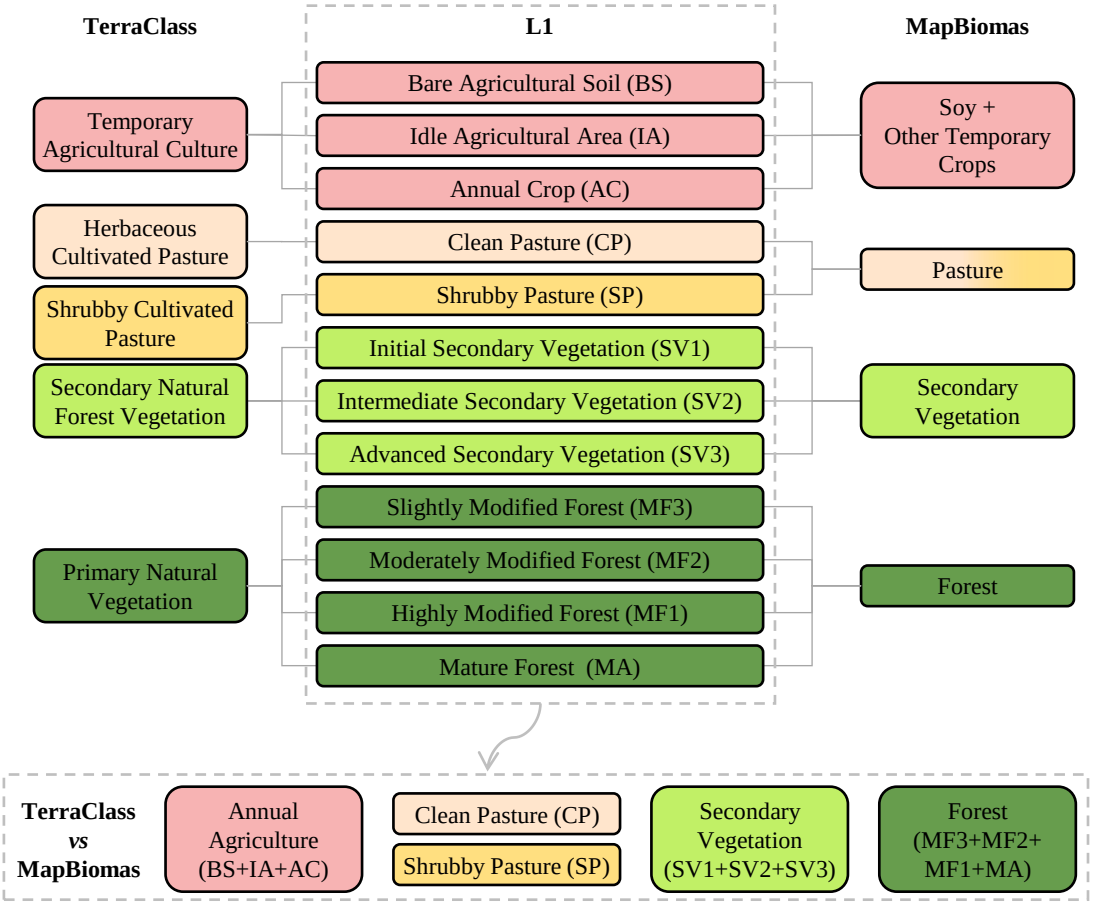
b) *Primarily Non-Vegetated/Non-Forested Area*: segments that cover areas mostly not originally covered by forests, so not prone to forest regeneration processes. These areas were selected considering the visual analysis of the Landsat mosaics from 1984 and 1985;

c) *Artificial Surface and Associated Areas*: these areas were identified mainly using the Landsat mosaic from 2020 and derived segmentation. The visual interpretation was done based on the auxiliary use of the monthly mosaics of Planet images to help verify these areas and guided by the locality data from IBGE (IBGE, 2019), as well as by the maps from TerraClass - second legend (2014) and MapBiomas collection 5 (2019).

B.2 Legend levels and harmonization to auxiliary data

To incorporate the maps generated by PRODES, TerraClass (second legend), and MapBiomas (collection 5) in Chapter 6, an important step was establishing the correspondence between the adopted legends. This process is illustrated in Figure B.1. Firstly, we divided areas of *Forest Formation* within the MapBiomas data set into the classes *Secondary Vegetation* and *Forest*, by re-labeling pixels based on the classification of a previous time. Here, any pixel classified as *Forest Formation* in a classification of a given year is relabeled as *Forest* if all classifications of previous years correspond to *Forest Formation*, and as *Secondary Vegetation* otherwise. This process is similar to the one used by Silva Junior et al. (2020).

Figure B.1 - Harmonization among the legends used in TerraClass (second legend), MapBiomas Collection 5, and adopted legend L1.



Only classes occurring within the study area and with correspondence with the classes in L1 (excluding those used to determine the mask) were considered for harmonization.

SOURCE: The author.

We then harmonized the legends (Figure B.1) to enable the identification of the agreement between TerraClass and MapBiomas classifications of the same year. When comparing TerraClass and MapBiomas, we kept the classes *Clean Pasture* and *Shrubby Pasture* labels from the TerraClass data, if the correspondent pixels in the MapBiomas data of the same year were classified as *Pasture*. The amount of pixels with correspondent classifications between TerraClass and MapBiomas datasets is registered in Table B.1.

Table B.1 - Number of labeled samples derived from the agreement between TerraClass and MapBiomas.

Class/Year	**1991**	**2000**	**2004**	**2008**	**2010**	**2012**	**2014**
Annual Agriculture	0	84	18575	120745	144990	98701	184932
Clean Pasture	338839	1993213	2550117	3043316	2325671	3070295	3153172
Shrubby Pasture	76501	2332366	1513040	1503844	2431435	1228883	2082434
Secondary Vegetation	174447	1303629	1133699	1780416	2109314	2526979	2739002
Forest	3849113	12574171	11367820	9759003	8722770	8358627	8256326

SOURCE: The author.

B.3 Labeled samples collection

Firstly, we normalized and segmented each one of these mosaics, with the same segmenter/parameters described in Section B.1. Then, whole segments were selected as labeled samples for each class of legend L2, based on the following criteria:

a) visual interpretation of the yearly Landsat mosaics from August to November (1984-2020);

b) visual interpretation of the monthly Landsat mosaics from June to November (of the correspondent year);

c) the majority of the segment was classified as the correspondent class in TerraCLass and MapBiomas. Exclusively for the subclasses derived from *Forest*, we also considered the PRODES data set to guide the collection of samples. As such, samples from the years 1993 (TM mosaic) and 2016 (ETM+ and OLI mosaics) were included in the data set.

This process was guided by a regular grid of 10 $\times$ 10 cells, fitted to divide the study area roughly into 100 equal subsets. The goal was to try to select labeled samples for the majority of classes in each cell, in order to improve class representation and sample distribution. To avoid mislabeled samples caused by eventual sub-segmentation, pixels from the manually collected pixels that did not match a correspondent class in TerraClass/MapBiomas classifications were excluded. Note that although samples were collected considering the segment, the training of CMAP is done using pixels as samples. The number of samples collected for each class (i.e total of pixels per class/year) is presented in Table B.2.

Table B.2 - Number of manually collected labeled samples in L2 legend.

Class/Year	**1993**	**2004**	**2008**	**2010**	**2014**	**2016 (ETM+)**	**2016 (OLI)**
Bare Agricultural Soil	0	7577	60998	65185	91007	0	0
Idle Agricultural Area	0	0	0	0	0	0	0
Annual Crop	0	0	0	0	0	0	0
Clean Pasture	0	42324	94433	68581	92544	0	0
Shrubby Pasture	0	15312	30679	29012	40556	0	0
Secondary Vegetation	0	30447	49443	39177	88181	0	0
Highly Modified Forest	64690	0	0	0	0	148830	284292
Slightly-Moderately Modified Forest	0	18487	64141	38893	45609	0	0
Mature Forest	0	191354	264127	120589	212128	0	0

SOURCE: The author.