A Landsat-based analysis of tropical forest dynamics in the Central Ecuadorian Amazon

Patterns and causes of deforestation and reforestation

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Fabián Santos

aus

Quito, Ecuador

Bonn 2018

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Patrick Hostert
2. Gutachter: Prof. Dr. Klaus Greve

Tag der Promotion: 26.11.2018
Erscheinungsjahr: 2018

ACKNOWLEDGEMENTS

A Ph.d. thesis is the result of a long and constant learning, which cannot succeed without the help of wonderful people. I want to first thank Prof. Dr. Gunther Menz who gave me an opportunity to develop my research project. I will be forever grateful for his support during my first Ph.d. years and am sad that he will not see me graduate. I want to also thank Prof. Dr. Patrick Hostert and Prof. Dr. Klaus Greve for also supporting me. Their guidance, knowledge, patience and compromise was a worthy motivation for me to complete this thesis.

I would like also to thank the teams of Center of Remote Sensing of Land Surfaces (ZFL) and Research Group Remote Sensing (RSRG) for their friendship and enjoyable atmosphere. I appreciate all your staff-out days, talking, conferences, lunch and coffee breaks, parties and other events we shared. They are memorable moments and I learnt a lot from you dear friends. Special thanks to Ellen Götz and Bärbel Konermann-Krüger for always keeping me informed about everything (even in perfect Spanish) and always supporting me during my studies. Thanks to my colleagues: Adrian Strauch, Annette Ortwein, Andreas Tewes, Andreas Rienow, Ayman Abdel-Hamid, Carlos García, Carsten Oldenburg, Esther Amler, Frank Thonfeld, Fridah Kirimi, Gohar Ghazaryan, Javier Gonzales, Javier Muro, Jessica Ferner, Jens Holberg, Jonas Schreier, Bartholomew Thiong'o Kuria, Olena Dubovyk, Paul Heid, Rui Liu, Ruben Piroska, Sabine Derdau, Stefanie Steinbach, Konrad Hentze, Reginald T. Guuroh and Valerie Graw. Also, I would like to thank the Geographical Institute of the Humboldt University for its valuable feedback and support, and the Bonner Graduiertenzentrum, European Space Agency (ESA) and the R community for inviting me to participate in different seminars, conferences and events.

Furthermore, I would like to thank Dr. Manske and Miake Retat-Amin who welcomed and invited me to join the Center for Development Research (ZEF) program. I am really happy for this experience, it opened my mind to new scientific advances and global challenges. Special thanks to Guido Lüchters for his statistical advices and Margaret Jend for her comments in this thesis. During this time, I met many inspiring colleagues and friends, to whom I wish to express my gratitude and admiration:

Alejandro Mora, Alicia Bustillos, Aline Barbosa, Anna Yuwen, Bibiana Betancur, Cesar Canon, Darius Mwingyine, Dufera Gerba, Emmanuel Rukundo, Ephraim Sekyi-Annan, Essa Chanie Mussa, Susanne Mundorf, Florent Noulekoun, Galia Figueroa, Getachew Feye, Giuduce Renzo, Marcos Glauser, Johannes Schielein, Juliet Kamau, Leny Martini, Marcos Jimenez, Melissa Quispe, Miranda A. Javier, Nina Pkhikidze, Dorothee Schmitt, Sonam Tashi, and Julia Zaadstra. Special thanks to my fellow-countrymen Jose Jara, Silvana Moreno, Paty Paz y Miguel Samaniego for their support and friendship during my Ph.d. years.

In my hometown, I want to thanks all my family: Ernesto, Sara, Juan Carlos, Natalia, Ignacio and Mauricio for their support and love during these years of travelling and studying abroad. Warm thanks to the Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) scholarship for supporting my research. During my first fieldwork campaign, where Prof. Menz visited me in Ecuador, I would like to thank the IKIAM, YACHAY and Estatal Amazónica Universities for receiving us and discussing common research interests. Special thanks to Fabian Englert and Raul Galeas for their sincere friendship and support in different ways. Finally, I would like to thank Liliana Herdoiza for all your help and support during our ascension to Sumaco volcano but also doing my fieldwork.

For all of you: my most sincere gratitude!

ABSTRACT

Tropical deforestation constitutes a major threat to the Amazon rainforest. Monitoring forest dynamics is therefore necessary for sustainable management of forest resources in this region. However, cloudiness results in scarce good quality satellite observations, and is therefore a major challenge for monitoring deforestation and for detecting subtle processes such as reforestation. Furthermore, varying human pressure highlights the importance of understanding the underlying forces behind these processes at multiple scales but also from an inter- and transdisciplinary perspective. Against this background, this study analyzes and recommends different methodologies for accomplishing these goals, exemplifying their use with Landsat time-series and socio-economic data. The study cases were located in the Central Ecuadorian Amazon (CEA), an area characterized by different deforestation and reforestation processes and socio-economic and landscape settings. Three objectives guided this research. First, processing and time-series analysis algorithms for forest dynamics monitoring in areas with limited Landsat data were evaluated, using an innovative approach based in genetic algorithms. Second, a methodology based in image compositing, multi-sensor data fusion and post-classification change detection is proposed to address the limitations observed in forest dynamics monitoring with time-series analysis algorithms. Third, the evaluation of the underlying driving forces of deforestation and reforestation in the CEA are conducted using a novel modelling technique called geographically weight ridge regression for improving processing and analysis of socio-economic data. The methodology for forest dynamics monitoring demonstrates that despite abundant data gaps in the Landsat archive for the CEA, historical patterns of deforestation and reforestation can still be reported biennially with overall accuracies above 70%. Furthermore, the improved methodology for analyzing underlying driving forces of forest dynamics identified local drivers and specific socio-economic settings that improved the explanations for the high deforestation and reforestation rates in the CEA. The results indicate that the proposed methodologies are an alternative for monitoring and analyzing forest dynamics, particularly in areas where data scarcity and landscape complexity require approaches that are more specialized.

KURZFASSUNG

Landsat-basierte Analyse der Dynamik tropischer Wälder im Zentral-Ecuadorianischen Amazonasgebiet: Muster und Ursachen von Abholzung und Wiederaufforstung

Die tropische Entwaldung stellt eine große Bedrohung für den Amazonas-Regenwald dar. Daher ist die Überwachung von Walddynamiken eine notwendige Maßnahme, um eine nachhaltige Bewirtschaftung der Waldressourcen in dieser Region zu gewährleisten. Jedoch verschlechtert Bewölkung die Qualität der Satellitenaufnahmen und stellt die hauptsächliche Herausforderung für die Überwachung der Entwaldung sowie die Detektierung einhergehender Prozesse, wie der Wiederaufforstung, dar. Darüber hinaus zeigt der unterschiedliche menschliche Nutzungsdruck, wie wichtig es ist, die zugrundeliegenden Kräfte hinter diesen Prozessen auf mehreren Ebenen, aber auch inter- und transdisziplinär, zu verstehen. Variierender anthropogener Einfluss unterstreicht die Notwendigkeit, unterschwellige Prozesse (oder „Driver“) auf multiplen Skalen aus inter- und transdisziplinärer Sicht zu verstehen. Darauf basierend analysiert und empfiehlt die vorliegende Studie unterschiedliche Methoden, welche unter Verwendung von Landsat-Zeitreihen und sozioökonomischen Daten zur Erreichung dieser Ziele beitragen. Die Untersuchungsgebiete befinden sich im Zentral-Ecuadorianischen Amazonasgebiet (CEA). Einem Gebiet, das einerseits durch differenzierte Entwaldungs- und Aufforstungsprozesse, andererseits durch seine sozioökonomischen und landschaftlichen Gegebenheiten geprägt ist. Das Forschungsprojekt hat drei Zielvorgaben. Erstens werden auf genetischen Algorithmen basierten Verfahren zur Verarbeitung der Zeitreihenanalyse für die Überwachung der Walddynamik in Gebieten, für die nur begrenzte Landsat-Daten vorhanden waren, bewertet. Zweitens soll eine Methode in Anlehnung an Satellitenbildkompositen, Datenfusion von mehreren Satellitenbildern und Veränderungsdetektion gefunden werden, die Einschränkungen der Walddynamik durch Entwaldung mithilfe von Zeitreihen-Algorithmen thematisiert. Drittens werden die Ursachen der Entwaldung/Abholzung im CEA anhand der geographischen gewichteten Ridge-Regression, die zur einen verbesserten Analyse der sozioökonomischen Information beiträgt, bewertet. Die Methodik für das Walddynamik-Monitoring zeigt, dass trotz

umfangreicher Datenlücken im Landsat-Archiv für das CEA alle zwei Jahre die historischen Entwaldungs- und Wiederaufforstungsmuster mit einer Genauigkeit von über 70% gemeldet werden können. Eine verbesserte Analysemethode trägt außerdem dazu bei, die für die Walddynamik verantwortlichen treibenden Kräfte zu identifizieren, sowie lokale Treiber und spezifische sozioökonomische Rahmenbedingungen auszumachen, die eine bessere Erklärung für die hohen Entwaldungs- und Wiederaufforstungsraten im CEA aufzeigen. Die erzielten Ergebnisse machen deutlich, dass die vorgeschlagenen Methoden eine Alternative zum Monitoring und zur Analyse der Walddynamik darstellen; Insbesondere in Gebieten, in denen Datenknappheit und Landschaftskomplexität spezialisierte Ansätze erforderlich machen.

RESUMEN

Análisis basado en la constelación Landsat para el análisis de la dinámica forestal en la Amazonía Central del Ecuador: patrones y causas de la deforestación y reforestación

La deforestación en los trópicos constituye una de las mayores amenazas para los bosques húmedos amazónicos. Es por ello que el monitoreo de la dinámica forestal es necesaria para el manejo sustentable de los recursos forestales en la región. Sin embargo, la alta nubosidad limita la adquisición de información satelital y esto constituye un gran obstáculo y reto para el monitoreo de la deforestación y cambios sutiles como la reforestación. Por otro lado, distintas presiones humanas resaltan la importancia de entender los factores subyacentes de estos procesos en múltiples escalas, pero también desde una perspectiva inter- y transdisciplinaria. Por ello, este estudio analiza y recomienda diferentes metodologías para cumplir estos objetivos, ejemplificando su uso con series temporales de la constelación Landsat y datos socio-económicos. Los estudios de caso se localizaron en la Amazonía Central Ecuatoriana (CEA), un área caracterizada por diferentes procesos de deforestación y reforestación; así como por diferentes entornos económicos y paisajísticos. Tres objetivos guiaron esta investigación. El primero, evalúa algoritmos de procesamiento y análisis de series temporales para el monitoreo de la dinámica forestal en áreas con poca información de la constelación Landsat, usando un innovador enfoque basado en algoritmos genéticos. El segundo, propone una metodología basada en composición de imágenes, fusión de datos de múltiples sensores, y la detección de cambios post-clasificación para superar las limitaciones observadas para el monitoreo de la dinámica forestal con algoritmos de análisis de series temporales. El tercero, evalúa las fuerzas conductoras subyacentes de deforestación y reforestación en la CEA, enfocándose en una nueva técnica llamada regresión de cresta ponderada geográficamente, para mejorar el procesamiento y el análisis de información socio-económica. La metodología de monitoreo de la dinámica forestal demostró que a pesar de los abundantes vacíos de información en el archivo Landsat en la CEA, los patrones históricos de deforestación y reforestación pueden aún ser reportados bianualmente con precisiones globales mayores al 70%. Por otro lado, la metodología mejorada para el análisis de las fuerzas conductoras subyacentes identificaron determinantes locales y entornos socio-económicos específicos que mejoraron las explicaciones de las más altas tasas de deforestación y reforestación en la CEA. Estos resultados indican que las metodologías propuestas son una alternativa para el monitoreo y el análisis de la dinámica forestal, particularmente en áreas donde la escases de datos satelitales y la complejidad del paisaje requieren de métodos especializados.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS I

ABSTRACT III

KURZFASSUNG IV

RESUMEN VI

TABLE OF CONTENTS VIII

LIST OF FIGURES XII

LIST OF TABLES XV

ACRONYMS AND ABBREVIATIONS XVI

1 INTRODUCTION 1

1.1 Background 1

1.1.1 Dynamics of tropical forests 1

1.1.2 Central Ecuadorian Amazon 2

1.1.3 Challenges of monitoring tropical forest dynamics 5

1.1.4 Challenges of analyzing the causes of tropical forest dynamics 7

1.2 Thesis structure 9

1.2.1 Research problem 9

1.2.2 Research objectives 10

1.2.3 Chapters published as articles 11

2 MONITORING FOREST DYNAMICS IN THE ANDEAN AMAZON: THE APPLICABILITY OF BREAKPOINT DETECTION USING LANDSAT TIME-SERIES AND GENETIC ALGORITHMS 15

2.1 Introduction 15

2.2 Materials and Methods 17

2.2.1 Study Area 17

2.2.2 Landsat Data Acquisition and Geometry Correction Assessment 19

2.2.3 Ancillary Data for Sampling, Validation, and Pre-Processing 21

2.3 Additional Landsat Pre-Processing Calculations 22

2.3.1 Topographic Correction and Radiometric Normalization 22

2.3.2 Vegetation Indices Derivatives 25

2.3.3 Composites Calculation 25

2.4 Description of the PCG Function and Its Algorithms Integrated 26

2.4.1 Step 1: Time-Series Compilation 26

2.4.2 Step 2: Noise Reduction and Gap-Filling 30

2.4.3 Step 3: Breakpoint Detection 31

2.5 Genetic Algorithm Implementation 32

2.5.1 Genetic Algorithms 32

2.5.2 Training Sampling Design 35

2.5.3 Samples Interpretation and Fitness Evaluation 36

2.5.4 GA Calibration Settings 38

2.6 Results 39

2.6.1 GA Search Approaches and Calibrations Found 39

2.6.2 Execution of Calibration Solutions and Visual Assessment 42

2.6.3 Manual Optimization of ECP Calibration Solution and Execution of the PCG for a Larger Area 43

2.6.4 Accuracy Assessment 44

2.6.5 Speed Performance 46

2.7 Discussion 47

2.7.1 Objections Regarding Breakpoint Detection and Its Applicability in the Andean Amazon 47

2.7.2 Considerations When Data Quantity Allows the Implementation of Breakpoint Detection 48

2.7.3 Some Advice for Improved Optimization of Complex Functions with GA 49

2.8 Conclusions 50

Acknowledgments 51

3 MONITORING LONG-TERM FOREST DYNAMICS WITH SCARCE DATA: A MULTI-DATE CLASSIFICATION IMPLEMENTATION IN THE ECUADORIAN AMAZON 52

3.1 Introduction 52

3.2 Study Area 54

3.3 Materials and Methods 56

3.3.1 Data and Pre-Processing 57

3.3.2 Standardized Biennial Compositing 60
3.3.3 Model Training 62
3.3.4 Post-classification change detection 64
3.3.5 Accuracy Assessment 65
3.4 Results 66
3.4.1 Variable Importance and svmPoly Optimization Report 66
3.4.2 Accuracy Metrics Results 67
3.4.3 Deforestation and reforestation maps 68
3.4.4 Comparison with other sources 70
3.5 Discussion 71
3.5.1 Biennial Image Compositing and Pre-Processing Effects on Results 71
3.5.2 Post-Classification Change Detection Performance 72
3.5.3 Multi-Sensor Fusion Benefits and Landsat Archive Limitations 73
3.5.4 Long-Term Forest Dynamics in the UNW and Feasibility of Multi-Date Classification in the Andean Amazon 73
3.6 Conclusions and outlook 75
Acknowledgments 76

4 ANALYZING UNDERLYING CAUSES OF DEFORESTATION AND REFORESTATION IN THE CENTRAL ECUADORIAN AMAZON: A GEOGRAPHICALLY WEIGHTED RIDGE REGRESSION APPROACH 77

4.1 Introduction 77
4.2 Study area 80
4.3 Materials and Methods 82
4.3.1 Deforestation and reforestation maps 82
4.3.2 Dasymetric mapping and population allocation algorithm 84
4.3.3 Socio-economic, accessibility, proximity and biophysical variables 86
4.3.4 Geographically weighted ridge regression 89
4.4 Results 92
4.4.1 GWRR collinearity assessment, filtering and regression models 92
4.4.2 Forest change rates and regression coefficients 94
4.4.3 Regression coefficient surfaces – globally and locally influencing variables 95
4.4.4 Spatial regimes and underlying causes 97

4.5 Discussion 100
4.5.1 Achievements of dasymetric mapping for census data processing........ 100
4.5.2 GWRR advantages and limitations 101
4.5.3 Forest dynamics drivers, population growth and agricultural expansion in the CEA 102
4.6 Conclusions 104
Acknowledgments 105

5 RESEARCH SYNTHESIS AND CONCLUSIONS 106

5.1 Summary 106
5.1.1 Objective I: Evaluation of time-series algorithms for monitoring forest dynamics with genetic algorithms 106
5.1.2 Objective II: Implementation of multi-date classification for long-term forest dynamics monitoring with scarce data 108
5.1.3 Objective III: Assessment of underlying causes of deforestation and reforestation through geographically weighted ridge regression 110
5.2 Conclusions and outlook 113

REFERENCES 115

SUPLEMENTARY MATERIAL 142

CURRICULUM VITAE 172

LIST OF FIGURES

FIGURE 1. ANNUAL NET FOREST GAIN/LOSS BY COUNTRY, 1990 - 2015. ADAPTED FROM FAO, 2016. 2

FIGURE 2. ECUADOR MAIN EXPORTS FOR 2016, ACCOUNTING FOR U$24.4 BILLON. ADAPTED FROM SIMOES AND HIDALGO (2011). DATA FROM UN COMTRADE DATABASE (2012). 3

FIGURE 3. (A) ECUADOR FOREST REFERENCE EMISSION LEVEL MAP OF LAND-USE CHANGE FOR THE PERIOD 2001-2008 WITH AREAS BY LAND-USE CHANGE CATEGORY. ADAPTED FROM MAE, 2017. (B) SPATIAL SPREAD OF FRONTIER SETTLEMENTS IN THE ECUADORIAN AMAZON. ADAPTED FROM BROWN ET AL., 1994. CENTRAL ECUADORIAN AMAZON DEFINED BY DASHED LINES. 4

FIGURE 4. (A) USGS LANDSAT ARCHIVE HOLDINGS AS OF JANUARY 1, 2015. ADAPTED FROM WULDER ET AL., 2012. (B) APPROXIMATE NUMBER OF CLOUD-FREE OBSERVATIONS ACQUIRED PER YEAR INTEGRATING LANDSAT, RESOURCESAT-2 (AWIFS) AND CBERS CONSTELLATIONS. ADAPTED FROM WULDER ET AL., 2016. CENTRAL ECUADORIAN AMAZON POINT BY THE ARROW. 6

FIGURE 5. (A) THREE TYPES OF BOUNDARY CHANGES IN A CENSUS FROM 2000 (BLACK) TO 2010 (RED). ADAPTED FROM LOGAN, XU AND STULTS, 2014. (B) SCALE EFFECTS INDUCED BY DATA AGGREGATION. ADAPTED FROM SALMIVAARA ET AL. 2015. (C) SPATIAL HETEROGENITY PATTERNS. ADAPTED FROM ESRI, 2010. 8

FIGURE 6. (A) LOCATION OF STUDY AREAS. (B) PHOTOGRAPHS OF A, B, AND C STUDY AREAS. A IS LOCATED IN THE VICINITY OF A PROTECTED AREA, B IS LOCATED CLOSE TO A SETTLEMENT, AND C IS LOCATED IN A PRIVATE FOREST RESERVE SURROUNDED BY AGRICULTURE. MAP SOURCE: MAE 2008. 18

FIGURE 7. (A) NO-DATA PERCENTAGE OBSERVED FOR PERIOD 1996–2015 FOR THE SUBSET OF TWO LANDSAT FOOTPRINTS USED (149 IMAGES SELECTED). (B) SPATIAL DISTRIBUTION OF NO-DATA FREQUENCY. 19

FIGURE 8. EDGE MASK FROM REFERENCE IMAGE (NOVEMBER 2000) AND CONTROL, DISPLACEMENTS, AND MATCHING POINTS WITH OVERLAP GREATER THAN 80% IN: (A) STUDY AREA A, (B) STUDY AREA B, AND (C) STUDY AREA C. 21

FIGURE 9. (A) EXAMPLES OF DIFFERENT PRE-PROCESSING APPROACHES USING YEARLY COMPOSITES AND THE NBR VEGETATION INDEX. BOXPLOTS WERE BASED ON SAMPLES LOCATED IN SUN-EXPOSED AND SHADOW AREAS. RED LINES HIGHLIGHT MEAN VALUES IN THE TIME-SERIES. (B) MAP OF THE SAMPLES USED IN BOXPLOTS. BLUE DOTS IDENTIFY SHADE-EXPOSED SAMPLES WHILE RED DOTS IDENTIFY SUN-EXPOSED SAMPLES. 24

FIGURE 10. REMAINING NO DATA PERCENTAGES IN COMPOSITES TEMPORALITIES. 26

FIGURE 11. PCG WORKFLOW. 27

FIGURE 12. GA APPLIED TO PCG WORKFLOW 34

FIGURE 13. PROCESSING TIME MEASURED FOR DIFFERENT SAMPLE SIZE SETS. THE PCG WAS EQUALLY CONFIGURED IN ALL CASES, EXECUTING GA FOR 10 TIMES TO AVERAGE ITS PROCESSING TIME. FOR TESTING PURPOSES, GA POPULATION SIZE WAS SET AT 30 INDIVIDUALS AND 10 GENERATIONS. 35

FIGURE 14. INTERPHASE FOR SAMPLES INTERPRETATION. FIRST ROW (FROM LEFT TO RIGHT) REFERS TO HIGH RESOLUTION IMAGERY COMPOSED BY AERIAL PHOTOS (1982) AND IMAGES FROM ASTER (2001 AND 2006), RAPID EYE (2011) AND GOOGLE EARTH (2015). THE REST OF THE IMAGE CHIPS CORRESPOND TO THE LTS AS YEARLY RBG COMPOSITES. FOLLOWING PROTOCOLS, WE INTERPRET THESE CASES AS FOLLOWS: (A) SAMPLE REJECTED BY UNMASKED RIVER, (B) FOREST GAIN AFTER DISTURBANCE OBSERVED IN THE 3RD ROW, (C) FOREST LOST WITH A SHORT REGROWN PERIOD, AND (D) STABLE FOREST SAMPLE. 37

FIGURE 15. SEARCH SPACE SIZE ALLOWED FOR THE EXPLORATIVE AND EXPLOITATION MODES. 38

FIGURE 16. BDA FITNESS VALUE EVOLUTION. 41

Figure 17. Spatial representations of the best BDA solutions. Study areas are shown in rows (A, B, C) and columns refer to (a) Change 1997–2015 mask, (b) BreakoutDetection, (c) Changepoint, and (d) ECP results. In all cases, no-forest areas before 1997 are shown in black, while areas not analyzed are shown in white. 43

Figure 18. Application of the ECP-optimized processing chain. Study areas are shown in rows (A, B, C) and columns refer to (a) LCCM, (b) The forest dynamics map obtained, and (c) its application to a larger area. 44

Figure 19. Execution time for: (a) Each BDAs optimized processing chain, running 10 times with a sample set of 100 pixels. (b) Total GA execution time obtained for each BDA package and approaches applied. Acronyms used for BDAs are: BRE for BreakoutDetection; CHP for Changepoint; and ECP for the package of the same name. 46

Figure 20. (a) The UNW study area, localization in the context of the Amazon Basin and Tropical Andes. Data from The National Information System of Ecuador (2017) and digital elevation model of the UNW. (b) Boxplot of the canopy closure index and estimated forest age; and hemispherical photographs with derived canopy closure index for different forest development stages 56

Figure 21. Methods and workflow for implementing the multi-date classification. 57

Figure 22. (a) Spatial distribution of no-data pixel frequencies and (b) no-data percentage of selected images. Gray areas in the plots refer to the 1992–1996 data gap and dashed lines to effective 23-year time-series. 59

Figure 23. (a) Number of images used for compositing and no-data percentage in resulting composites and (b) Tasseled Cap Wetness (TCW) mean value of composites with (Matched) and without histogram matching (Raw). Gray areas in the plots refer to the 1992–1996 data gap and dashed lines to effective 23-year time-series. 61

Figure 24. (a) The feature space of the training samples using the first and second principal components (PC1 and PC2) obtained from Landsat bands, derivatives, and terrain parameters. (b) Boxplots of classifiers out-of-bag error using training samples. 63

Figure 25. Example of a forest-loss sample (3 pixels). (a) Location of the sample in the map. (b) High-resolution image chips from ASTER (2001–2008 and 2012), and aerial color photography (2010). (c) Landsat color composites image chips (1992–2014, with a data gap in 1996), and (d) overview showing sample area and analysis period. 66

Figure 26. (a) Variable importance by class during classification with svmPoly. (b) Optimization of the svmPoly classifier during the model-training phase. 67

Figure 27. (a) Deforestation year map for the UNW. (b) Magnified areas show: (1) Linear deforestation along the E45 highway; (2) landslide scars; and (3) false-positive errors in the mixed forest and pasture areas. 69

Figure 28. (a) Reforestation year map for the UNW. (b) Magnified areas show: (1) Jatun-Sacha Biological Reserve, which is known for reforestation since the 1990s; (2) forest succession after landslides; and (3) false-positive errors in a stable-forest area. 69

Figure 29. (a) GFC2015 and Multi-date deforestation areas by biennials for the period 2000 – 2014. (b) MLUX and Multi-date deforestation and reforestation areas for the periods 2000 – 2008 and 2008 – 2014, together with their (c) stable classes areas. 70

Figure 30. Deforestation / reforestation area for the period 1992-2014 (a) along the highway E45 and (b) along the highway E20. 74

FIGURE 31. (A) STUDY AREA IN THE AMAZON BASIN. (B) CANTONS. (C) LAND COVER FOR 2008. DATA FROM MAE, 2013.81

FIGURE 32. (A) CELLS OF DEFORESTATION AND (B) REFORESTATION RATES FOR 2000 - 2010 IN THE CEA. DEFORESTATION RATES SIGN WERE INVERTED TO ESTABLISH A COMMON SCALE IN THE MAPS. (C) BOXPLOTS OF ANNUAL RATES FOR EACH CANTON IN THE CEA WITH MEAN DEFORESTATION (VALUES WITH NEGATIVE SIGN AT THE LEFT SIDE) AND REFORESTATION (VALUE WITH POSITIVE SIGN AT THE RIGHT SIDE) RATES. 83

FIGURE 33. (A) TOTAL POPULATION FOR THE CENSUS 2010 (AGE 15 - 72) AND (B) AFTER ALLOCATION. (C) BOXPLOT OF ALLOCATED POPULATION AND SUM FOR EACH CANTON. 86

FIGURE 34. (A) GAUSSIAN KERNEL FUNCTION SCHEMA AND (B) CORRELATION PLOT FOR DEFORESTATION VARIABLES 90

FIGURE 35. (A) LCN BOXPLOT OBTAINED FOR EACH GWRR MODEL AND VARIABLE GROUP. THE RED DOTTED LINE INDICATES THE 30 LCN THRESHOLD. OBSERVATION REMOVAL FREQUENCY FROM GWRR MODELS IN (B) DEFORESTATION AND (C) REFORESTATION CELLS 92

FIGURE 36. BOXPLOTS OF REGRESSION COEFFICIENTS FOR DEFORESTATION (UPPER PLOT) AND REFORESTATION (LOWER PLOT) CELLS, TOGETHER WITH THEIR CORRESPONDENT ANNUAL FOREST CHANGE RATE (COLORED POINTS) 94

FIGURE 37. SURFACE REGRESSION COEFFICIENTS AND VARIANCE FOR VARIABLES BEHAVING MORE LOCALLY OR GLOBALLY IN DEFORESTATION CELLS 96

FIGURE 38. SURFACE REGRESSION COEFFICIENTS AND VARIANCE FOR VARIABLES BEHAVING MORE LOCALLY OR GLOBALLY IN REFORESTATION CELLS 97

FIGURE 39. REGRESSION COEFFICIENTS FOR DEFORESTATION CLUSTERS AND THEIR LOCALIZATION IN THE CEA 98

FIGURE 40. REGRESSION COEFFICIENTS FOR REFORESTATION CLUSTERS AND THEIR LOCALIZATION IN THE CEA 99

LIST OF TABLES

TABLE 1. ACQUISITION PARAMETERS OF THE USED IMAGES. 20

TABLE 2. CONTROL POINTS, NEW-COREGISTER DISPLACEMENT AND POINT MATCHES WITH OVERLAP 20

TABLE 3. SUMMARY OF PCG STEPS, METHODS, ALGORITHMS, PACKAGES, PARAMETERS, VALUES, STEPS, AND THEIR SEARCH SPACE SIZE. 28

TABLE 4. PARAMETER SELECTION BY GA AS BEST CALIBRATION SOLUTIONS. 42

TABLE 5. CONFUSION MATRIX FOR THE FOREST DYNAMICS MAP. 45

TABLE 6. ACCURACY MEASURES FOR THE THREE STUDY AREAS. 45

TABLE 7. LANDSAT BANDS AND DERIVATIVES. 59

TABLE 8. CLASSES CONSIDERED AND TRAINING SAMPLE SIZE. 62

TABLE 9. TRANSITION RULES. 64

TABLE 10. CLASSES, MINIMUM DISTANCES, AND SAMPLE SIZES. 65

TABLE 11. OVERALL ACCURACY OF DEFORESTATION AND REFORESTATION MAPS BY PROCESSING APPROACHES. 67

TABLE 12. COMMISSION AND OMISSION ERRORS FOR THE SURFACE REFLECTANCE – FILTERED APPROACH. 68

TABLE 13. DESCRIPTIVE STATISTICS FOR DEFORESTATION AND REFORESTATION CELLS 84

TABLE 14. ROAD ACCESSIBILITY TIME AND WEIGHTS ASSIGNED 85

TABLE 15. DESCRIPTIVE STATISTICS OF VARIABLE GROUPS OBTAINED FOR DEFORESTATION CELLS 87

TABLE 16. RESULTS OF GWRR MODELS FOR VARIABLE GROUPS 93

ACRONYMS AND ABBREVIATIONS

CEA	Central Ecuadorian Amazon
EFR	Ecuador Forest Reference
GWRR	Geographically Weighted Ridge Regression
REDD+	Reducing Emissions from Deforestation and Forest Degradation
AFRI16	Aerosol-Free Vegetation Index 1.6 μm band
AFRI21	Aerosol-Free Vegetation Index 2.1 μm band
AIC	Akaike information criterion
BDA	Breakpoint detection algorithms
DEM	Digital elevation model
ECP	Non-Parametric Multiple Change-Point Analysis of Multivariate Data Package
ETM+	Enhanced Thematic Mapper Plus
FDD	Forest dynamics drivers
FLY	Forest Loss Year
GA	Genetic algorithms
GEMI	Global Environmental Monitoring Index
HPC	High-performance computing
LCCM	Land-cover change map
LCLUC	Land-Cover and Land-Use Change
LCN	Local condition numbers
LOF	Local outlier factor algorithm
LTS	Landsat time-series
MODIS	Moderate Resolution Imager Spectroradiometer
MSAVI	Modified Soil Adjusted Vegetation Index
NBR	Natural Burn Ratio
NBR	Normalized burn ratio
NDVI	Normalized Difference Vegetation Index
NREF	Radiometric normalization of surface reflectance
NTOPO	Radiometric normalization of topographic correction
OLI	Operational Land Imager
PCG	Processing chain generator

R43, R54, R57	Band ratios: TM4/TM3,TM5/TM4, TM5/TM7
REF	Surface reflectance
RMSE	Root mean square error
SD	Standard deviation
SRTM	Shuttle Radar Topography Mission
svmPoly	Least squares support vector machine with polynomial kernel
SVMs	Support vector machines
TCTB	Tasseled Cap brightness
TCTG	Tasseled Cap greenness
TCTW	Tasseled Cap wetness
TM	Thematic Mapper
TOPO	Topographic correction
UNW	Upper Napo Watershed

1 INTRODUCTION

1.1 Background

1.1.1 Dynamics of tropical forests

Tropical forests are fundamental for providing essential services to all forms of life, and play an unquestionable role in climate regulation, biochemical cycles and biological diversity (Joseph et al. 2011). With respect to their significant importance in achieving the Sustainable Development Goals (United Nations 2015) it is recognized that social and economic development depends on their sustainable management. Historical trends of land-use change have revealed the relationship between population growth, agricultural expansion and deforestation. This characterized the economic development in the temperate climatic regions until the 19th century; however, this relationship now has the greatest influence in the tropical climatic regions (Figure 1) (FAO 2016b). According to the FAO (2016a), tropical forests showed the highest deforestation rate between 1990 and 2015, where Africa (-0.49%), South America (-0.40%) and Asia (-0.24%) came first. These regions are mainly represented by developing countries of low to middle incomes, growing rural populations and net gains in agricultural area (Sachs 2001). The latter is associated with about 80% of the deforestation worldwide, where commercial agriculture, cattle ranching and subsistence agriculture are referred to as the most important causes (Kissinger et al. 2012).

In contrast, reforestation is characterized by reduced pressure on forests that lead to net losses in agricultural areas (FAO 2016b). In this regard, forest spontaneously regenerates in abandoned lands leading to net gains in forest areas. This has been observed in high-income countries where pressure on forests has ceased such as Europe (0.01%), but also where afforestation policies (or forest planting) have been implemented (e.g. Asia and North America) (FAO 2016a). Nevertheless, there is evidence that some tropical countries are experimenting with similar trends (e.g. Ghana, Costa Rica, Vietnam) but not specifically by afforestation, as this implies policies for deliberately expanding forest cover (e.g. Bangladesh, China, India) (Rudel et al. 2005). Thus, economic growth, declining rural populations and agricultural intensification have been pointed out as being the main causes of spontaneous reforestation (FAO 2016a).

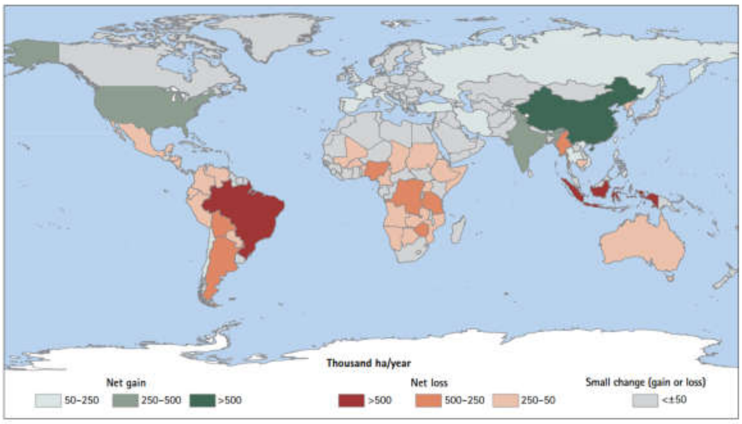


Figure 1. Annual net forest gain/loss by country, 1990 - 2015. Adapted from FAO, 2016.

Both processes, i.e. deforestation and reforestation, define forest dynamics in this thesis. Identifying their patterns and discussing their causes is the base for sustainable land use and balanced decisions on forest conversion. Patterns thus refer to human footprints observable from satellite data that help us to monitor forest gains and losses. Causes also can be direct or indirect human activities and their complex socio-economic interactions, which act as driving forces behind forest dynamics. Both aspects are challenging in regions such as the tropics, where data availability is poor and methodological transfer is not always applicable. The latter is related to the uneven distribution of the studies and methodologies developed, where the vast majority were carried out within the Brazilian region (Da Ponte et al. 2015). However, forest conversion is much more extreme in other countries, e.g. Paraguay, Nigeria, and Myanmar, but the small number of studies demonstrates the lack of concern and local expertise. Against this background, this study focuses on the Central Ecuadorian Amazon (CEA), a region located at the foothills of the Andean range as a study case. This region has been little investigated, but important land-cover changes have been reported since oil resources were discovered and the agrarian reform and colonization deeply reshaped its landscapes (Pierre et al. 1988).

1.1.2 Central Ecuadorian Amazon

Ecuador is considered as a country with one of the highest species and ecosystem diversities in the world (Sierra et al. 2002; Bass et al. 2010). Its geographic position and the Andean Range result in many different forest types including tropical rainforest, dry

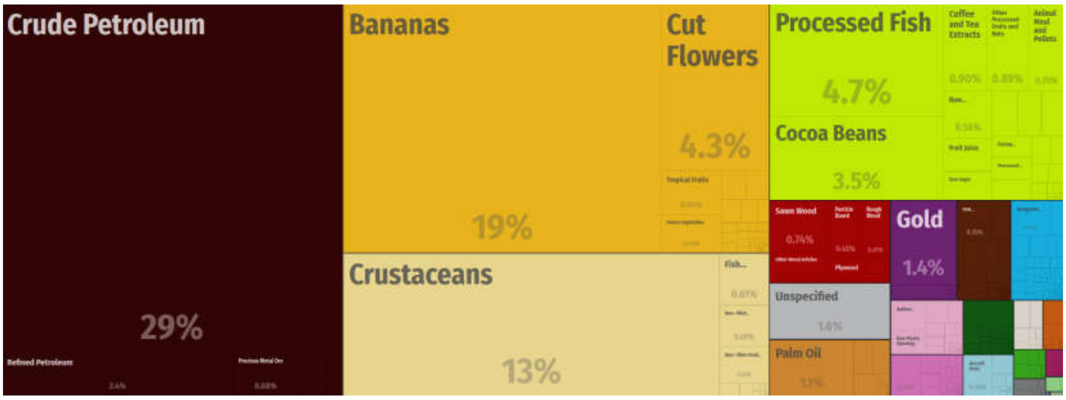
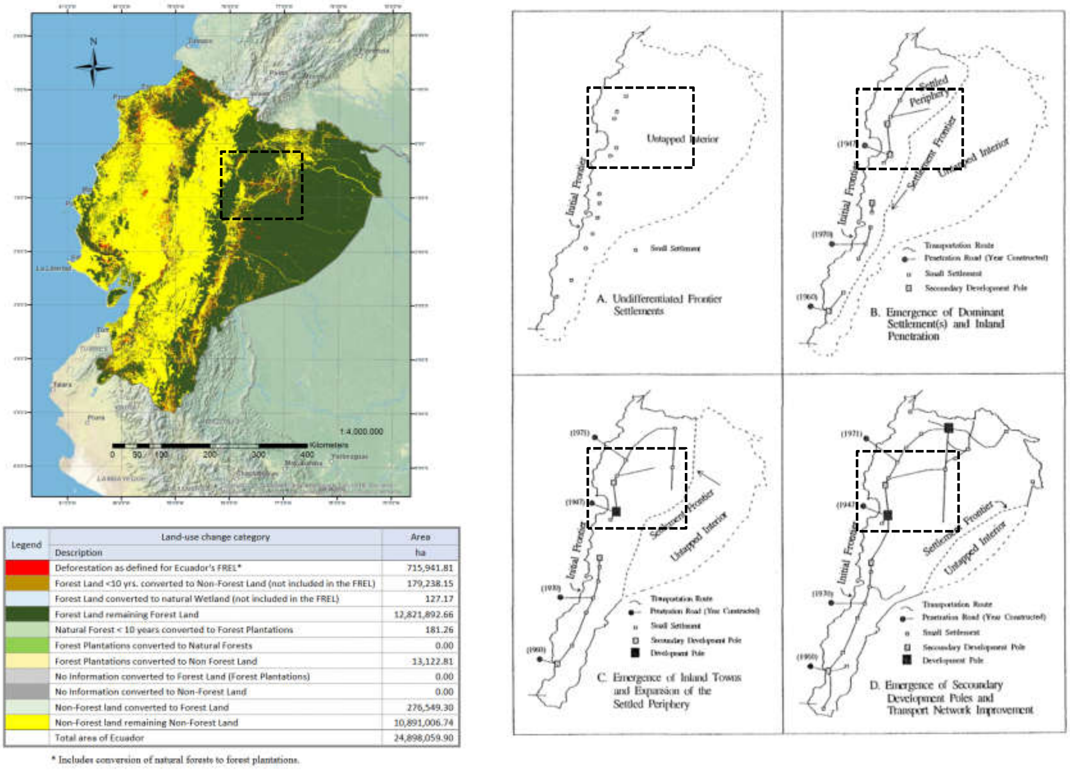


Figure 2. Ecuador main exports for 2016, accounting for U$24.4 billon. Adapted from Simoes and Hidalgo (2011). Data from UN Comtrade Database (2012).

forest and montane forest, among others (MAE 2012b). Moreover, most of the Ecuador territory forms part of two of the most important hotspots in the world known as Tumbes-Chocó-Magdalena and Tropical Andes corridors (Myers et al. 2000). However, these are seriously threatened by deforestation and Ecuador lost 0.6% of its forest cover annually for the period 1990 – 2015 (FAO 2016).

Other sources focusing exclusively on the Amazon biome, where most of the intact forests are located in Ecuador (i.e. 97,530 km^2), indicate that the country has suffered the fourth highest deforestation rate in South America. This represented a total 10.7% forest loss for the period 2000 – 2013, and it is estimated that between 1970 and 2013, 10.470 km^2 of original forest were lost in the Ecuadorian Amazon (RAISG 2015). While reforestation areas are less known, the FAO (2016a) reported that forests expanded on average 33,000 ha year^{-1} for the period 1990 – 2010.

As the Ecuadorian economy is based on raw material production and export, much depends on the use of natural resources. Extractive sectors include most prominently the oil industry, agriculture, fisheries, aquaculture and forestry representing 40% of gross domestic product and 80% of its exports (Figure 2) (Carrion & Chíu 2011; Simoes & Hidalgo 2011; United Nations 2012). Therefore, deforestation in the Ecuadorian Amazon has been associated with crude oil export and the agrarian reform and colonization, which promoted timber exploitation and changes in land-use practices (RAISG 2015). This was particularly important between 1964 and


N
1:4.000.000
0
50
100
200
300
400
Kilometers
Untapped Interior
Initial Frontier
Small Settlement
A. Undifferentiated Frontier Settlements
Settled Periphery
Settlement Frontier
(1947)
(1970)
(1960)
Transportation Route
Penetration Road (Year Constructed)
Secondary Development Pole
B. Emergence of Dominant Settlement(s) and Inland Penetration
(1975)
(1930)
Development Pole
C. Emergence of Inland Towns and Expansion of the Settled Periphery
(1971)
(1943)
D. Emergence of Secondary Development Poles and Transport Network Improvement

Legend	Land-use change category	Area
	Description	ha
🟥	Deforestation as defined for Ecuador's FREL*	715,941.81
🟫	Forest Land <10 yrs. converted to Non-Forest Land (not included in the FREL)	179,238.15
▨	Forest Land converted to natural Wetland (not included in the FREL)	127.17
🟩	Forest Land remaining Forest Land	12,821,892.66
▨	Natural Forest < 10 years converted to Forest Plantations	181.26
🟩	Forest Plantations converted to Natural Forests	0.00
▨	Forest Plantations converted to Non Forest Land	13,122.81
▨	No Information converted to Forest Land (Forest Plantations)	0.00
▨	No Information converted to Non-Forest Land	0.00
▨	Non-Forest land converted to Forest Land	276,549.30
🟨	Non-Forest land remaining Non-Forest Land	10,891,006.74
	Total area of Ecuador	24,898,059.90

* Includes conversion of natural forests to forest plantations.

(a) (b)

Figure 3. (a) Ecuador Forest Reference Emission Level Map of land-use change for the period 2001-2008 with areas by land-use change category. Adapted from MAE, 2017. **(b)** Spatial spread of frontier settlements in the Ecuadorian Amazon. Adapted from Brown et al., 1994. Central Ecuadorian Amazon defined by dashed lines.

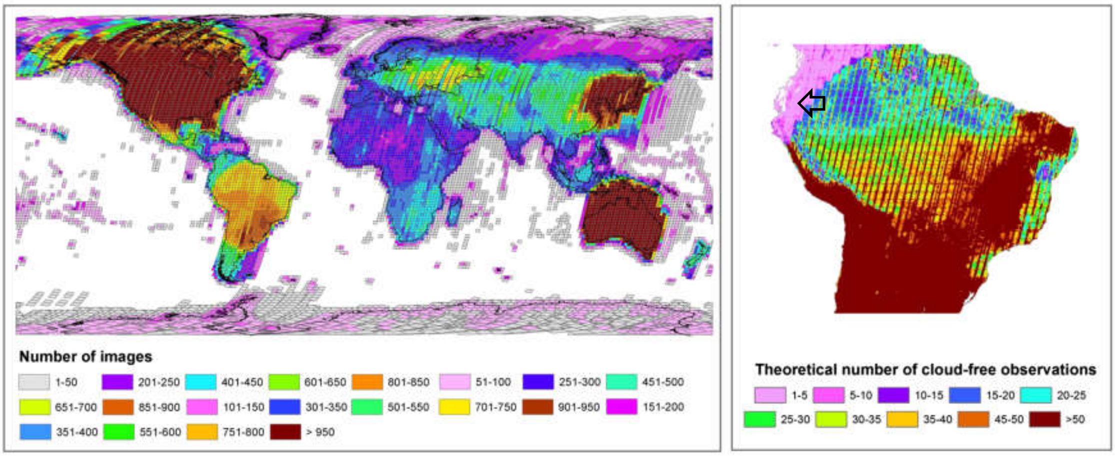
1994, as oil discoveries motivated road construction, new settlements and migration of farmers pursuing agriculture and cattle ranching (Bass et al. 2010). Consequently, spontaneous and largely undirected colonization transformed the Ecuadorian Amazon, exacerbating land conflicts. In this regard, the northern part was one of the most deforested regions in the world (Myers 1993). This provoked unrestrained grabbing of lands considered as “vacant” from the traditional territories used by the Cofan, Siona-Secoya and Huaorani people (Wasserstrom & Southgate 2013). Farther south, in the central part, deforestation was less intense since there was little or no oil there, but roads were built to encourage settlement (Figure 3). As these lands were the traditional territory of the Kichwa people, land conflicts with the new colonists ended in land claims. To avoid territory loss, the Kichwa people abandoned their traditional subsistence economy and subdivided their communal land into individual parcels. Transformed into pastures for cattle ranching, these lands could be titled and protected

(Macdonald 1981; Wasserstrom & Southgate 2013). Nevertheless, some studies suggest that after the agrarian reform and colonization had ended, reforestation characterized some regions of the Ecuadorian Amazon. For instance, the Shuar people, living in the south, abandoned land-extensive cattle ranching for land-intensive cash cropping, which was also combined with forestry in some cases (Rudel et al. 2002). To the best knowledge of the author, few studies have been conducted to confirm whether this was also the case in other regions of the Ecuadorian Amazon.

Due to these massive social and environmental transformations in recent decades (Perreault 2001), an investigation focusing on the patterns of forest dynamics and the causes of changes in the Ecuadorian Amazon may reveal important aspects regarding protection of the remaining forests. In this regard, the CEA represents an interesting study case, as pressures on the forest either through agricultural expansion, oil extraction or population growth vary locally. This makes it difficult to provide a single explanation for the forest dynamics, while monitoring is not a straightforward task as explained in following sections.

1.1.3 Challenges of monitoring tropical forest dynamics

Monitoring tropical forest dynamics refers to evaluation of periodical information on the status of the forests. This requires continuous observations, normally costly and even impossible to collect in the field (Fragal et al. 2016). In this regard, satellite data offers a unique source as it covers large areas in different spectral, spatial, and temporal resolutions for characterizing forest cover. Thanks to the free and accessible data policy of the Landsat (USGS 2014) and Copernicus (ESA 2018) programs, it is today possible to access to global-scale medium spatial resolution satellite data (Hansen & Loveland 2011), making near-real-time forest monitoring possible. In this context, remote sensing has played an important role as a scientific discipline providing methods for land-cover change monitoring. Methods range from simpler approaches based on comparing images of two different times (bi-temporal) to the time-series analysis that provides the entire spectral history of the pixel (Coppin et al. 2004). The number of algorithms can be overwhelming.



(a) (b)

Figure 4. (a) USGS Landsat archive holdings as of January 1, 2015. Adapted from Wulder et al., 2012. **(b)** Approximate number of cloud-free observations acquired per year integrating Landsat, ResourceSat-2 (AWIFS) and CBERS constellations. Adapted from Wulder et al., 2016. Central Ecuadorian Amazon point by the arrow.

Nevertheless, deriving accurate estimates of forest change in the tropics is fraught with difficulties and uncertainties. From the conceptual discrepancies of what is a forest to the technical difficulties in the identification of forest boundaries, variation in estimates is a constant (Foody 2003). Moreover, persistent cloudiness over the year (Wulder et al. 2016; Li et al. 2018) coupled with topographically complex landscapes (Asner et al. 2014) drastically reduce data quantity and quality of the satellite archives (Figure 4). While radar sensors could be more suitable in cloudy regions, their unavailability and scarce studies have placed more emphasis on optical sensors, where the Landsat family is the preferred option for tropical forest monitoring (Da Ponte et al. 2015). In this regard, the use of bi-temporal approaches has been mostly with Landsat data, where visual interpretation is a common procedure to derive land-cover maps and subsequently evaluate land-cover change. While the method is relatively simple and easy to implement, further development has included semi-automatic image classification (Matricardi et al. 2007), image thresholding (Greenberg et al. 2005) or post-classification change detection (Caldas et al. 2015) to reduce workload and facilitate interpretation.

On the other hand, time-series analysis approaches have started to be applied more frequently in the tropical forests context (e.g. Müller et al. 2016; Ye et al. 2018;

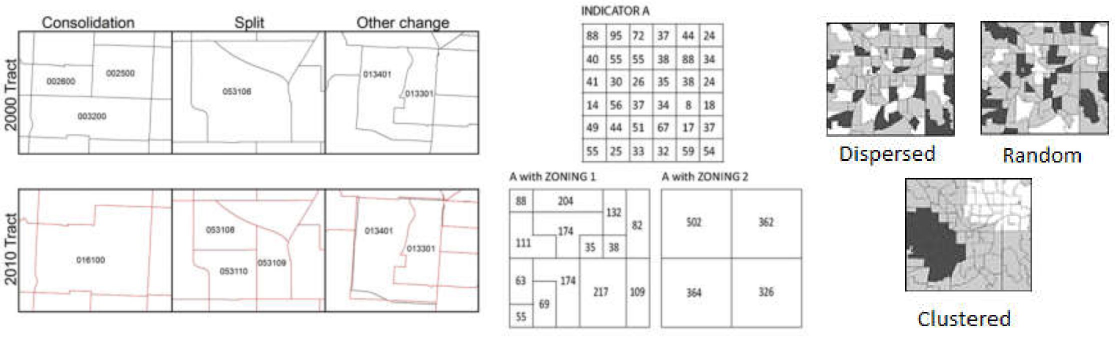
Potapov et al. 2012), demonstrating to facilitate land-cover studies. Since spectral-temporal trajectories are used for reconstructing forest disturbance and recovery, periods of several years can be observed. Moreover, analysis of large collections of images can be automated, and more informative and reliable results be obtained. However, more sophisticated algorithms and pre-processing procedures are required (Zhu 2017; Shimizu et al. 2018).

As such approaches for tropical forest monitoring exist, selecting an appropriate method for the CEA requires evaluation under the abovementioned conditions. While most studies conducted in the Ecuadorian Amazon have been based on bi-temporal and multi-temporal analysis (Sierra 2000; MAE 2012a), new algorithms using time-series analysis remain to be tested.

1.1.4 Challenges of analyzing the causes of tropical forest dynamics

Beyond the need of refined measurements of forest dynamics, identification of the causes depends critically on the ability to model the socio-economic determinants (Wood & Skole 1998). This is due to human activities and actions directly impacting forest cover (proximate causes), which in turn are linked to complex interactions between social, economic, political, cultural and technological processes (underlying forces) (Kissinger et al. 2012). Thus, different combinations of proximate causes and underlying forces in varying geographical and historical contexts result in explanations that are not universal (Geist & Lambin 2002). Therefore, explaining the causes of forest dynamics is not a straightforward task. However, this is greatly needed for policies targeting sustainable management of forests.

As proximate causes can mostly be identified through remote-sensing-based techniques (Da Ponte et al. 2015), underlying forces could be more complex as socio-economic data is required. Therefore, proximate causes are more frequently reported in such analyzes (Armenteras et al. 2017; Mon et al. 2012; Samndong et al. 2018). This highlights the importance of identifying which limitations impede or complicate studies on the underlying forces. First, population and agricultural censuses are in the most cases the unique source of socio-economic data, but differences in the reporting units complicate their integration and analysis (Figure 5a) (Logan et al. 2014). Moreover, as



(a) (b) (c)

Figure 5. (a) Three types of boundary changes in a census from 2000 (black) to 2010 (red). Adapted from Logan, Xu and Stults, 2014. (b) Scale effects induced by data aggregation. Adapted from Salmivaara et al. 2015. (c) Spatial heterogenity patterns. Adapted from ESRI, 2010.

aggregation of smaller units into larger observational units is a common practice for report census statistics, data loss and scale effects may lead to ecological fallacies (Figure 5b) (Holt et al. 1996; Salmivaara et al. 2015). Lastly, spatial heterogeneity has been regarded as an important issue in modelling relationships between variables (Figure 5c). Since global regression models are frequently used for analyzing the underlying forces of forest dynamics, it is assumed that the resulting parameter estimates are constant across space. This constitutes a disadvantage, as local variation is masked and results describe the study area as a single entity (Brunsdon et al. 2002).

The abovementioned difficulties require innovative methodologies; however, these have hardly been investigated in the context of forest dynamics. Nevertheless, some contributions have been proposed in other fields. In this regard, for overcoming areal-data integration conflicts and aggregation effects, algorithms such as pycnophylactic interpolation and kriging-based disaggregation have been developed (Krivoruchko et al. 2011; Tobler 1979). While these only require areal data as input, other more complex algorithms based on dasymetric mapping can recalculate areal data using spatial data (Stevens et al. 2015; Mennis & Hultgren 2006). Moreover, to address spatial heterogeneity, new methodologies such as geographically weighted regression have been proposed (Brunsdon et al. 1996). The capability of this regression to derive local surface representation of regression estimates has been discussed as an advance in geostatistics.

All these methodologies need to be evaluated in order to contribute to the analysis of the causes of forest dynamics. While proximate causes are well known in the Ecuadorian Amazon (i.e. agricultural expansion, population growth and oil extraction) (Armenteras et al. 2017; RAISG 2015), the underlying forces remain to be better understood.

1.2 Thesis structure

1.2.1 Research problem

The importance of monitoring tropical forest dynamics and understanding their causes is a necessary step forward to the Sustainable Development Goals, sustainable land use and balanced decisions on forest conversion (see section 1.1.1). However, there are important challenges in fulfilling this task. First, a remote sensing approach for tropical forest monitoring must be able to work with scarce data and topographically complex landscapes. While several methodologies exist for this task, their transferability still needs to be tested in regions like the CEA (see section 1.1.3). On the other hand, analysis of the causes of tropical forest dynamics requires integrating socio-economic data from various sources. This task faces challenges, as areal-based data inconsistencies and aggregation effects limit their analysis. Moreover, current approaches applying regression analysis ignore local differences, describing results as single entities (see section 1.1.4). This may not be suitable for regions such as the Central Ecuadorian Amazon, whose geographic particularities and contrasting socio-economic settings make it highly heterogeneous (see section 1.1.2). In this context, this research formulates the following three questions:

Q1. Which methodology and processing steps are needed for monitoring long-term forest dynamics in data-scarce and topographically complex regions of the Tropical Andes?

Q2. How can census data processing and regression analysis of forces driving deforestation/reforestation in heterogeneous regions of the Tropical Andes be improved?

Q3. Which are the patterns and causes of deforestation/reforestation in the CEA exemplified in a study case of the Tropical Andes?

1.2.2 Research objectives

The overarching research purpose of this study is to develop a methodology for monitoring tropical forest dynamics and evaluating their causes, and exemplifying its use in the CEA for the period 2000-2010. More specifically, using the Landsat archive as the main source of satellite information, a processing chain will be developed for monitoring tropical forest dynamics considering limitations induced by data scarcity and topographically complex landscapes. Such system will deliver historical deforestation and reforestation maps for the CEA, which are later integrated into a second processing chain. There, socio-economic and biophysical data will be processed and integrated with deforestation and reforestation maps. Then, an analysis of driving forces will be conducted to discuss these as causes of tropical forest dynamics in the CEA. Since challenges with respect to data integration, aggregation effects and spatial heterogeneity remain issues to be solved, this processing chain must provide specific solutions. Thus, the methodology should be developed considering reproducibility and further development. This will enable its applicability in other regions.

Three research objectives are formulated to address the different stages of the development of this methodology:

1. Evaluation of time-series analysis and pre-processing algorithms for monitoring tropical forest dynamics: Constitutes an exploration of different algorithms for data pre-processing and breakpoint detection using Landsat time-series. Applying genetic algorithms, experimental processing chains are created and optimized to determine best combination of algorithms and parameters. Moreover, time-series analysis suitability with Landsat data is evaluated for monitoring tropical forest dynamics in the CEA. Since a large of number of algorithms are analyzed, three test sites are studied to reduce processing and conclude results.
2. Implementation of multi-date classification for long-term tropical forest dynamics monitoring with scarce data: Derivation of a suitable methodology using multi-decade Landsat data. This considers techniques such as image compositing, multi-sensor data fusion and post-classification change detection

to overcome limitations induced by data scarcity and topographically complex landscapes. Here, deforestation and reforestation patterns are identified and discussed for the Upper Napo Watershed as a study area in the CEA.

3. Assessment of underlying causes of deforestation and reforestation through geographically weighted ridge regression: Completion of the methodology for evaluating the driving forces. In this regard, an innovative approach for integrating data and reducing aggregation effects with dasymetric mapping is presented. Moreover, using geographically weighted ridge regression, an analysis of local drivers is introduced to address spatial heterogeneity in the regression analysis. Finally, insights for explain causes of deforestation and reforestation in the CEA are discussed and concluded.

1.2.3 Chapters published as articles

This thesis includes one publication in a scientific journal (Chapter 2), one accepted for its publication in September 2018 (Chapter 3), and one submitted to consider its publication (Chapter 4). Minor changes have been made to the first publication to fit this thesis. A synthesis of main findings is presented in Chapter 5, targeting the research objectives and questions formulated, finalizing with conclusions and outlook.

Santos F., Dubovyk O. & Menz G., 2017. **Monitoring forest dynamics in the Andean Amazon: The applicability of breakpoint detection methods using Landsat time-series and genetic algorithms**. *Remote Sensing*, 9(1).

Abstract: The Andean Amazon is an endangered biodiversity hotspot but its forest dynamics are less studied than those of the Amazon lowland and forests in middle or high latitudes. This is because landscape variability, complex topography and cloudy conditions constitute a challenging environment for any remote-sensing assessment. Breakpoint detection with Landsat time-series data is an established robust approach for monitoring forest dynamics around the globe but has not been properly evaluated for implementation in the Andean Amazon. We analyzed breakpoint detection-generated forest dynamics in order to determine limitations when applied to three different study areas located along an altitude gradient in the Andean Amazon in Ecuador. Using the available Landsat imagery for the period 1997–2016, we evaluated different pre-processing approaches, noise reduction techniques, and breakpoint detection algorithms. These procedures were integrated into a complex function called the processing chain generator. Calibration was not straightforward since it required defining values for 24 parameters. To solve this problem, we implemented a novel approach using genetic algorithms. We calibrated the processing chain generator by applying a stratified training sampling and a reference dataset based on high resolution imagery. After the best calibration solution was found and the processing chain generator executed, we assessed accuracy and found that data gaps, inaccurate co-registration, radiometric variability in sensor calibration, unmasked cloud, and shadows can drastically affect the results, compromising the application of breakpoint detection in mountainous areas of the Andean Amazon. Moreover, since breakpoint detection analysis of landscape variability in the Andean Amazon requires a unique calibration of algorithms, the time required to optimize analysis could complicate its proper implementation and undermine its application for large-scale projects. In exceptional cases when data quality and quantity are adequate, we recommend the pre-processing approaches, noise reduction algorithms and breakpoint detection algorithms procedures that can enhance results. Finally, we include recommendations for achieving a faster and more accurate calibration of complex functions applied to remote sensing using genetic algorithms.

Santos F., Meneses P. & Hostert P., 2018**. Monitoring Long-Term Forest Dynamics with Scarce Data: A Multi-Date Classification Implementation in the Ecuadorian Amazon.** *European Journal of Remote Sensing (Manuscript accepted for publication).*

Abstract: Monitoring long-term forest dynamics is essential for assessing human-induced land-cover changes, and related studies are often based on the multi-decadal Landsat archive. However, in areas such as the Tropical Andes, scarce data and the resulting poor signal-to-noise ratio in time-series data renders the implementation of automated time-series analysis algorithms difficult. The aim of this research was to investigate a novel approach that combines image compositing, multi-sensor data fusion, and post-classification change detection that is applicable in data-scarce regions of the Tropical Andes, exemplified for a case study in Ecuador. We derived biennial deforestation and reforestation patterns for the period from 1992 to 2014, achieving accuracies of 82 ± 3% for deforestation and 71 ± 3% for reforestation mapping. Our research demonstrates that an adapted methodology allowed us to derive the forest dynamics from the Landsat time-series despite the abundant regional data gaps in the archive, namely across the Tropical Andes. This study therefore presents a novel methodology in support of monitoring long-term forest dynamics in areas with limited historical data availability.

Santos F. & Graw V., 2019. **Analyzing Underlying Causes of Deforestation and Reforestation in the Central Ecuadorian Amazon: A Geographically Weighted Ridge Regression Approach.** *PLOS ONE* (*Manuscript submitted for publication).*

Abstract: The Tropical Andes region encompasses endangered biodiversity hotspots with high conservation priority. Deforestation due to population growth and agriculture expansion is therefore one of the main threats to this region and thus highlights the importance of understanding the drivers of this process on multiple scales. On the other hand, the drivers of reforestation and their role in forest recovery are less known. Therefore, we propose an interdisciplinary approach to analyze both deforestation and reforestation drivers by applying geographically weighted ridge regression. This method evaluates spatial non-stationarity and provides surface representations of local parameter estimates to identify regions where drivers show higher significance for either deforestation or reforestation. Our analysis includes nine different variable groups and two predictors using socio-economic data from population censuses, accessibility models and biophysical features. Information on deforestation and reforestation were based on remote sensing input data. We used dasymetric mapping in conjunction with land-cover maps to downscale areal-based data and improve the

spatial resolution of our analysis. We conducted our research in the Tropical Andes of the Ecuadorian Amazon, a highly heterogeneous region, within the time period 2000 - 2010. Areas were highlighted where improved accessibility to palm oil, coffee, cacao and milk production facilities motivated deforestation, while reforestation seems to follow the opposite trend. Moreover, gender, ethnicity and household structure showed a high influence on untangled population dynamics and their relationship with forest change. This approach demonstrates the benefits of integrating remote sensing derived products and socio-economic data for understanding coupled socio-ecological systems from local to global scales.

2 MONITORING FOREST DYNAMICS IN THE ANDEAN AMAZON: THE APPLICABILITY OF BREAKPOINT DETECTION USING LANDSAT TIME-SERIES AND GENETIC ALGORITHMS

Fabián Santos, Olena Dubovyk and Gunter Menz

"Evolution continually innovates, but at each level it conserves the elements that are recombined to yield the innovations."

John Holland.

2.1 Introduction

Since an open data policy was adopted by the Landsat program, an increasing number of data-driven algorithms for monitoring forest dynamics using Landsat time-series (LTS) have been developed (Wulder et al. 2011). They highlight abrupt events (e.g., clear-cutting, crown fires) or slower processes (e.g., degradation, succession dynamics) that, within a longer time span, cause deviations or illustrate longer-duration changes from a presumably stable condition (Kennedy et al. 2010b). Typically, these algorithms are called breakpoint detection algorithms (BDA), and according to Banskota et al. (2014), they can be classified according to their methodological basis and scope. This research establishes breakpoint detection as a useful implementation of LTS analysis; however, conditions in the Andean Amazon—high topographic relief, cloud cover that prevents the production of more than a few clear observations throughout the year, and landscape variability—complicate this type of analysis.

Furthermore, pre-processing approaches and noise-reduction algorithms for Landsat have evolved in recent decades (Hansen & Loveland 2011), creating significant variation in preparing the data it generates to enhance breakpoint detection (Flood et al. 2013; Hermosilla et al. 2015a; Huang et al. 2010). For this reason, the interpreter's experience plays an important role in selecting, calibrating, and applying an algorithm. Inappropriate selection or calibration can introduce systematic errors that can be difficult to detect. This problem is particularly overwhelming, with a large number of complex algorithms interlinked. Breakpoint detection requires a dense time-series of Landsat images that are radiometrically homogeneous, free of cloud and shadow, and geometrically corrected. Achieving these levels of processing is challenging since parametrization of these algorithms is not an easy task, and computer processing LTS data is highly demanding.

For these reasons, it has become necessary to optimize the complex processing chains that combine pre-processing, noise reduction, and BDA procedures. There is little research in this field. Nonetheless, optimization has been shown to be successful in remote sensing. Examples vary according to the field and the search algorithm applied (Li et al. 2015; Wang et al. 2012; Dou et al. 2015). Genetic algorithms (GA) have a long history of refinement since they became popular though the work of Holland (1973); extensive research has reported it as a robust and efficient optimization algorithm with a wide range of application in areas such as engineering, numerical optimization, robotics, classification, pattern recognition, and product design, among others (Gibbs et al. 2008; Balakrishnan et al. 1996). Therefore, we chose a GA as our methodological approach to designing a processing chain based in BDA and to evaluate if this procedure might be a feasible approach for monitoring forest dynamics in the Andean Amazon. Since this is our first step before exploring other complex optimization procedures that, according to Eberhart et al. (1998), should be emphasized in the new hybrid implementations, we avoid extending our discussion in search of new algorithms. Such discussion is beyond the scope of this paper. We therefore recommend that interested readers review the references mentioned throughout this paper.

Our particular research objective focuses on conducting a GA optimization of different processing chains using different BDA in order to determine if it is possible to monitor the forest dynamics in the Andean Amazon. If the methodology is applicable, patterns of forest gain and forest loss should be evidenced during a period of time along a typical part of the Andean Amazon. Furthermore, since there are multiple methods of enhancing time-series quality, we also analyze different pre-processing approaches and noise reduction algorithms for improving breakpoint detection. In order to do this, we develop a function called a processing chain generator (PCG) to link these approaches as processing chains and evaluate if their results highlight patterns of forest dynamics. Since calibrating the PCG is not straightforward, due to the number of parameters involved in algorithms calibration (24 parameters with a total of 5.7491×10^{20} possible combinations), GA was used as the basis for exploring and designing an optimal calibration for the PCG. For this reason, our research also constitutes a novel approach for solving the calibration of complex models in remote sensing by reducing uncertainties through parametrization. This method is different from other optimization approaches in remote sensing, where the principal application is classification and pattern recognition (Li et al. 2015; Wang et al. 2012; Dou et al. 2015); however, it is closer to the approach of Iovine et al. (2005) for calibrating a model of cellular automata.

As this research considers landscape variability an important factor in monitoring the Andean Amazon, different study areas located in Ecuador were selected. They are characterized by frequent cloud cover, high topographic relief, and different forest management practices along a gradient of altitude. We found that these conditions mean the application of any remote sensing-based methodology is not straightforward.

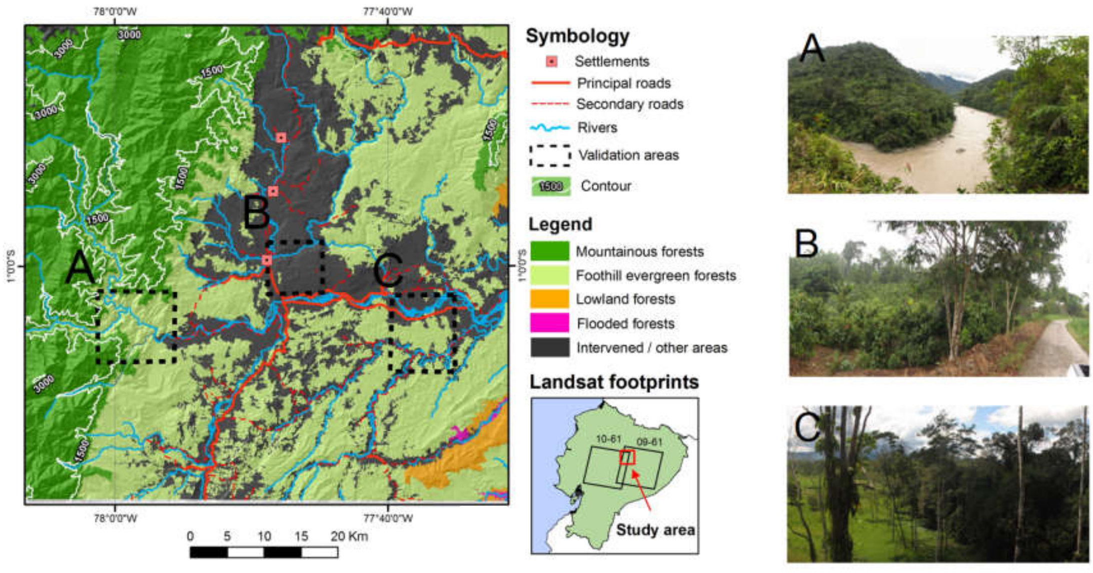
2.2 Materials and Methods

In this section, we describe general aspects of the study area, the Landsat data acquired for conduct our experiments, and the validation datasets used. Moreover, since a highly accurate co-registration is required for breakpoint detection, we include an assessment of the Landsat standard terrain correction (Level 1T) to ensure that images used were properly co-registered.

2.2.1 Study Area

The study area covers in total 241 km^2 distributed across three areas of 100, 52, and 89 km^2 termed A, B, and C, respectively. Their selection criteria were based on the different landscape configurations along a gradient of altitude and on forest management practices observed in the region. The areas studied are located in the central foothills of the Napo province in Ecuador (Figure 6), where mountainous terrain, foothills, and lowland evergreen forests constitute the main ecosystems (MAE 2013). The principal river is the Napo, which joins the Amazon River after 1800 km. The geomorphology is characterized by hilly (slopes 0°–26°) and mountainous (slopes greater than 26°) landscapes with high biodiversity (Beirne & Whitworth 2011). The altitude covers a range of 300–3875 m.a.s.l. Therefore, the region is characterized by a distinct climatic gradient with annual precipitation of 2000 mm –4000 mm and a mean temperature of 6 °C–24 °C. The main land-use systems are grasslands used for cattle grazing and croplands used for cacao, passion fruit, and corn production (Borja et al. 2015). The land-cover change area, which included forest loss and gain classes for the period 2000–2014, covered 3862 ha (MAE 2012a). The respective study areas measured 324 ha (Area A), 1575 ha (Area B), and 1963 ha (Area C) as a result of their different forest management practices.

Area A (mountainous, 2300–750 m.a.s.l.) is located in the vicinity of a protected area where forest loss is rare and is mainly caused by natural events (landslides or river floods); Area B (mainly hilly, 750–500 m.a.s.l.) is located in the vicinity of a settlement where forest loss is common and caused principally by expansion of the agricultural land; and Area C (mainly flat, 500–350 m.a.s.l.) is located in a private forest reserve, whose borders suffer forest lost as a result of road construction but the interior is experiencing forest gain as a result of ecological success after some areas were acquired for conservation.

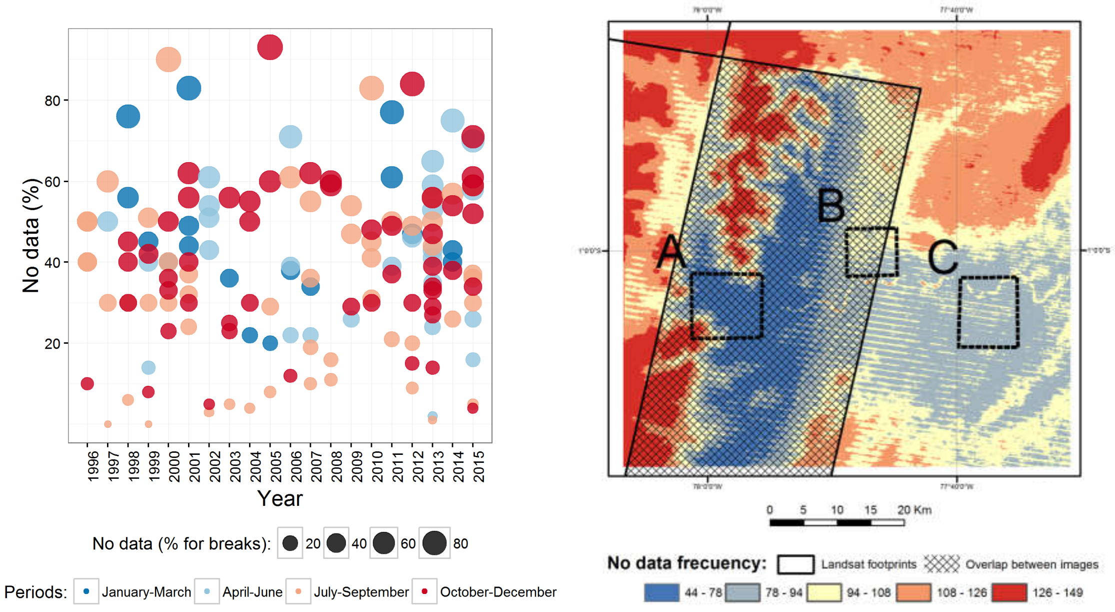


(a) (b)

Figure 6. (a) Location of study areas. **(b)** Photographs of A, B, and C study areas. A is located in the vicinity of a protected area, B is located close to a settlement, and C is located in a private forest reserve surrounded by agriculture. Map source: MAE 2008.

2.2.2 Landsat Data Acquisition and Geometry Correction Assessment

For this study, we obtained 826 surface reflectance images for a subset of two Landsat footprints (path and row: 09–61 and 10–61; total area: ~4400 km^2), which were processed through National Landsat Archive Processing System (LEDAPS) (Masek et al. 2012). They were downloaded from the US Geological Survey through their Internet website (USGS 2014). After applying the cloud mask “cfmask” product (Zhu & Woodcock 2012) 677 of these images were not used due to low image quality, including excessive no-data cover (>90%) (Figure 7). The remaining 149 images employed in this analysis were originally acquired for three Landsat sensors: 28 images by Thematic Mapper (TM), 94 images by Enhanced Thematic Mapper Plus (ETM+), and 27 images by Operational Land Imager (OLI). All imagery was processed to the standard terrain correction (Level 1T), and metadata reported a root mean square error (RMSE) of less than 7 ± 3 m. A summary of acquisition parameters for these images is shown in Table 1. The images selected for analysis covered an 18-year period (1996–2015) with a mean time interval between images of 47 days.



(a) (b)

Figure 7. (a) No-data percentage observed for period 1996–2015 for the subset of two Landsat footprints used (149 images selected). **(b)** Spatial distribution of no-data frequency.

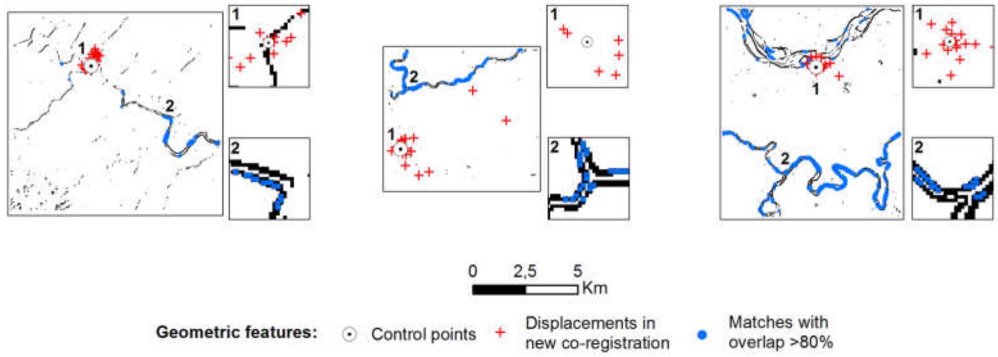
Table 1. Acquisition parameters of the used images.

Parameter	**Range**	**Average**	**Standard Deviation**
Sun Azimuth	45°–132°	92°	38°
Sun Elevation	53°–62°	58°	3°
Crossing time	14:40:08–15:28:42	15:12:50	9 min
Cloud cover (%)	19–89	68	16
Time interval between images (days)	1–304	47	53
Number of ground control points	9–240	75	49
RMSE of geometric residuals (meters)	3–11	7	3

As geometric accuracy of Landsat data is based on its footprint, we considered it necessary to evaluate it again for our study, as our research sites were less than a footprint. Therefore, we first applied a new co-registration to all images to evaluate if geometric accuracy would improve. For this, a Sobel filter was applied to the near-infrared channel (band 4) for each image in the LTS. We selected this band because it was less affected by cloud contamination compared to other bands. Then, edge masks were created thresholding these outputs and applying a linear registration to obtain match points and their transformation matrix. We used an affine model that uses 12 degrees of freedom to find match points and the nearest neighbor to resample images. This was done following the procedure and software R package NiftyReg (Clayden 2016b). To establish a geometric reference, a Landsat 5 image acquired in November 2000 was selected. For this image, the reported RMSE was 3.4 m with 240 points. To verify if co-registration improved results, displacements were calculated from control points to their new positions. Control points were identified in the reference image as stable areas during the 19 years of the LTS and they are described in Table 2.

Table 2. Control points, new-coregister displacement and point matches with overlap

Study Area	**Control Point Details**	**X and Y Mean Displacements in New Co-Registration [m]**	**Number of Point Matches with Overlap >80%**
A	River confluence	81.2	62
B	Road	576.4	182
C	Property boundary	97.3	667


1
2
0
2,5
5
Km
Geometric features:
Control points
Displacements in new co-registration
Matches with overlap >80%

(a) (b) (c)

Figure 8. Edge mask from reference image (November 2000) and control, displacements, and matching points with overlap greater than 80% in: **(a)** Study area A, **(b)** Study area B, and **(c)** Study area C.

The results of the described co-registration, however, did not improve the geometric accuracy of the images. On the contrary, the new co-registration increased displacements in the images; thus, this step was omitted from the analysis. In Figure 8, the results of the co-registration procedure displacements can be seen as red crosses, while control points as white dots. Results (Table 2) indicate that Area C was the easiest to co-register, but it was not enough to replace its existing geometric correction. Because of this, we applied another procedure for evaluating co-registration. It consisted of adding the edge masks created in the previous steps to observe if they overlapped along the LTS. Since the maximum overlap corresponded to areas where all images matched, it was normalized from 1 (no overlap) to 100 (all images in the LTS overlaps). Then, pixels whose overlap exceeded 80% were filtered. All three areas showed match points, except in areas where cloud cover was frequent or did not have relevant borders. As 20% of the images in the LTS remained uncertain, they were identified and visually inspected to manually improve their co-registration or eliminate them.

2.2.3 Ancillary Data for Sampling, Validation, and Pre-Processing

To establish a land-cover-change map (LCCM), two maps were created, i.e. forest and non-forest, based on years 1997 and 2015. Both maps were obtained using a trial-and-error threshold approach to classify the Natural Burn Ratio (NBR) index derived from the first and last images in the LTS. After visual inspection and correction, these maps were finalized and changes between 1997 and 2015 were established corresponding to five classes: stable non-forest, stable forest, forest lost, forest gain, and no data. This map was used later to extract the stratums needed for sampling (training and reference) and to perform visual assessments (see sections 2.5.3 and 2.6.4).

To validate training and reference samples, a dataset comprised of high-resolution imagery acquired from different sources was used. These images included aerial photos, RapidEye images, Google Earth, and ASTER imagery. These images were acquired for the years 1978, 1982, 2000, 2005, 2010, and 2015. Additionally, a field visit was done during the first semester of 2016 to corroborate the study area's land cover.

Finally, as topographic correction was indicated for Landsat pre-processing, we acquired a digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM) (CGIAR - CSI 2008) with a resolution of 90 m, as its interpolation quality was better than the 30-m version, especially in areas with data gaps of hilly landscapes.

2.3 Additional Landsat Pre-Processing Calculations

Because we need to evaluate different Landsat pre-processing approaches to determine the best calibration of the PCG, we describe in this section all the procedures implemented. These were applied to the subset of two Landsat footprints; therefore, its extension was larger than the study areas. As we calculated different vegetation indices and composites ensembles from each Landsat pre-processing approach, we also provide an overview of the procedures.

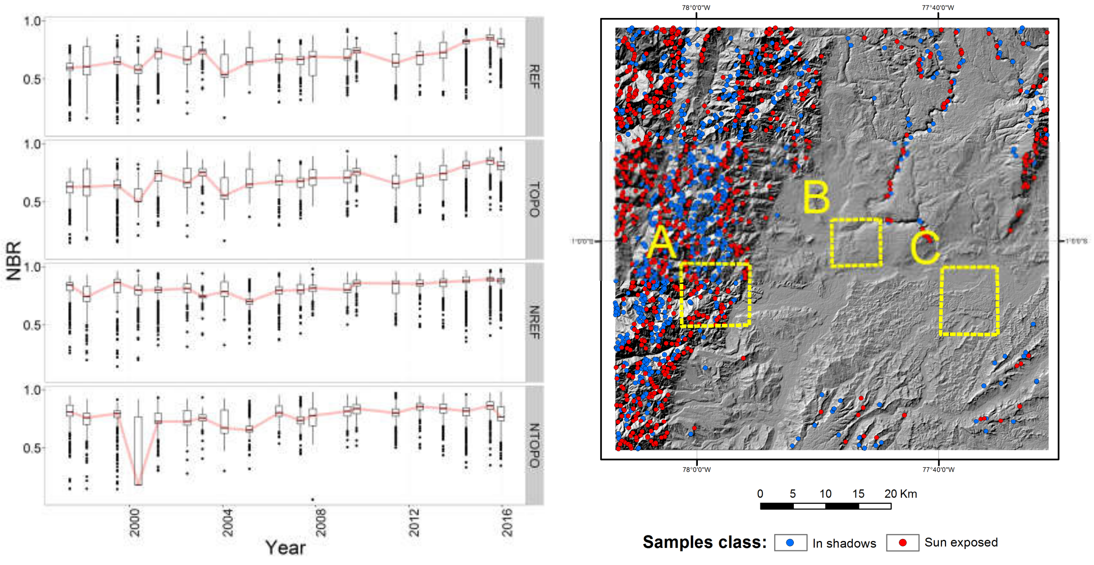
2.3.1 Topographic Correction and Radiometric Normalization

Topographic correction is considered a fundamental preliminary procedure before multi-temporal analyzes because solar zeniths and elevation angles, as well as direct topographic effects, differ from image to image (Flood et al. 2013). Both Richter et al. (2009) and Riaño et al. (2003) concluded that the C-correction method was preferred over other related topographic correction algorithms because it better preserves the spectral characteristics of the imagery. C-correction calculation incorporates the wavelength of each individual spectral band along with its diffuse irradiance. C-correction was originally proposed by Teillet et al. (1982), and we applied it to our LTS using the DEM in combination with the solar and azimuth elevation angles described in the imagery metadata.

Then, for radiometric normalization, we followed a procedure described by Hajj et al. (2008). This approach applies a relative radiometric normalization based on calculations of linear regressions between target and reference images and uses invariant targets to obtain regression coefficients. An important advantage of this technique is based on its absolute calculations. Vermote et al. (1997) described this methodology within the context of the 6S model. The 6S model relies on atmospheric data to reduce the effects of atmospheric and solar conditions relative to a reference image. Another advantage of this procedure is its ease of implementation and computation.

To implement radiometric normalization, a nearly cloud- and haze-free OLI surface reflectance image acquired by Landsat 8 on 2 September 2013 was used as reference. Its date corresponds to the low regional precipitation regime (with monthly rainfall below 250 mm) in the region (UNESCO 2010). As with all other images used, this reference image was cloud-masked and dimensionally homogeneous with the rest of the images. Afterwards, invariant target masks were generated from a Normalized Difference Vegetation Index (NDVI) ratio, where the NDVI of the target image was divided by the NDVI of the reference image (Rouse, R H Haas, et al. 1974). In most change-detection algorithms, the separation of no-change pixels is based on a histogram of outputs, where the mean or median value is used as the benchmark to define its range (Liu et al. 2004). This principle was observed in the NDVI ratio outputs and was implemented by computing different possible ranges in a loop calculation. This operation started from the mean value of the NDVI ratio, and the minimum and maximum values increased by a factor of 0.1% in each loop iteration. The process terminated when the number of pixels within the range exceeded 2% of the total number of pixels in the NDVI ratio. This percentage was considered sufficient to represent a wide diversity of features: primary forests, bare soils, water bodies, and urban infrastructures, among others. Hajj et al. (2008) and Mahiny et al. (2007) both applied percentages below 1% with the recommendation that this figure should be increased as much as possible. Finally, the range was used according to each image to reclassify and create the invariant target masks. After the calculation of invariant target masks for each target image (148 in total), the regression coefficients were calculated using a linear regression. For this step, we applied the invariant target mask to the reference and target images and then calculated the regression coefficients for each image's spectral band. In this procedure, each of the surface reflectance images normalizes its values according to the reference image.

To evaluate the performance of topographic correction, radiometric normalization and other combinations of these pre-processing approaches, different time-series boxplots were made for each case (Figure 9a). These plots used a sample set of 1678 pixels taken with a precision of 99%. They were randomly taken from the predominantly sun-exposed (839 samples, azimuth range areas: 45°–132°) and shadowed areas (839 samples, azimuth range areas: 225°–313°) (Figure 9b). Since variance should be similar in both sun-exposed and shadow-exposed areas if LTS achieves its temporal homogeneity, the standard deviation was calculated for each case. We obtained measurements of 0.11 for surface reflectance (REF), 0.11 for topographic correction (TOPO), and 0.07 for radiometric normalization of surface reflectance (NREF); and 0.15 for radiometric normalization of topographic correction (NTOPO). Based on these values, it was established that the NREF pre-processing approach reduced LTS variability, while other approaches, such as REF, TOPO and NTOPO, increased it.



(a) (b)

Figure 9. (a) Examples of different pre-processing approaches using yearly composites and the NBR vegetation index. Boxplots were based on samples located in sun-exposed and shadow areas. Red lines highlight mean values in the time-series. **(b)** Map of the samples used in boxplots. Blue dots identify shade-exposed samples while red dots identify sun-exposed samples.

2.3.2 Vegetation Indices Derivatives

With the pre-preprocessing outputs, a set of vegetation indices based on visible and infrared bands was calculated. Selection criteria were based on a literature review considering the atmospheric effects described by Myneni et al. and Matricardi et al. (1994; 2010). Moreover, because the time-series was based on the surface reflectance Landsat products, Tasseled Cap Transformations (TCTB, TCTG, and TCTW) were applied using the coefficients described by Baigab et al. (2014). The indices calculated can be classified in three groups:

- Using red and infrared bands: the NDVI (Rouse, R H Haas, et al. 1974), Global Environmental Monitoring Index (GEMI) (Pinty & Verstraete 1992), and Modified Soil Adjusted Vegetation Index (MSAVI) (Qi et al. 1994);
- Using only infrared bands: The NBR (Key & Benson 1999), the Aerosol Free Vegetation Index 1.6 µm band (AFRI16), and the Aerosol Free Vegetation Index 2.1 µm band (AFRI21) (Karnieli et al. 2001);
- Using the complete sensor bandwidth: Tasseled Cap Transformation Brightness (TCTB), Greenness (TCTG), and Wetness (TCTW) (Crist & Cicone 1984).

The topographic correction procedure can negatively impact the detection of low-magnitude changes in the landscape (Chance et al. 2016). Therefore, all vegetation indices were calculated for each pre-processing approach (REF, TOPO, NREF, and NTOPO) to expand the PCG calibration alternatives.

2.3.3 Composites Calculation

Following the calculation of the vegetation indices from the different preprocessing approaches, output images were composited. The image acquisition dates and yearly, semesterly, and trimesterly time periods were the principal considerations in structuring the image composite. The one-year timeframe is appropriate for use with Landsat archive data considering data acquisition limitations and the 16-day satellite repeat cycle (Kennedy et al. 2010b). However, consideration of the shorter temporalities for composites was necessary to validate this affirmation.

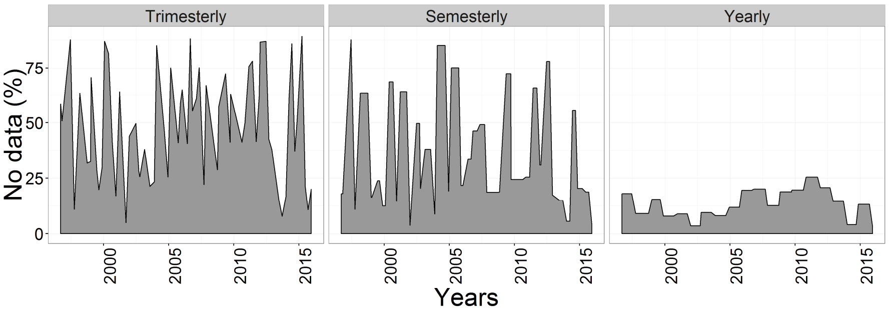


Figure 10. Remaining no data percentages in composites temporalities.

For this procedure, we followed recommendations described by Griffiths et al. (2013). Images were selected according to their acquisition date and the number of data gaps included in each. Images with fewer and less extensive data gaps are preferred within a particular time period. As each image is selected, any data gaps are filled with data from alternate images from the same time period. Twenty yearly composites, 37 semesterly composites, and 64 trimesterly composites were obtained to establish different periodicities and time-series arrangements. Figure 10 shows that no data percentages that remained after the composites were calculated.

2.4 Description of the PCG Function and Its Algorithms Integrated

In this section, we describe each algorithm integrated in the PCG before its optimization with GA. Algorithms were executed in three steps, of which the workflow in shown in Figure 11. Starting with the time-series compilation, all steps are further described below.

2.4.1 Step 1: Time-Series Compilation

To construct a time-series, the PCG needs to define which of the pre-processing approaches, vegetation indices and composites temporalities is used in this operation. All these procedures were previously described (see section 2.3); therefore, we do not extend the discussion. For the first step in the PCG, by combining 4 pre-processing approaches, 9 vegetation indices, and 3 composites temporalities, a total of 108 time-series compilation alternatives were available. With no clear reason to select a particular alternative, our decision was based on defining a combination of the three elements mentioned before. These combinations among others parameters required by the PCG are summarized in Table 3.

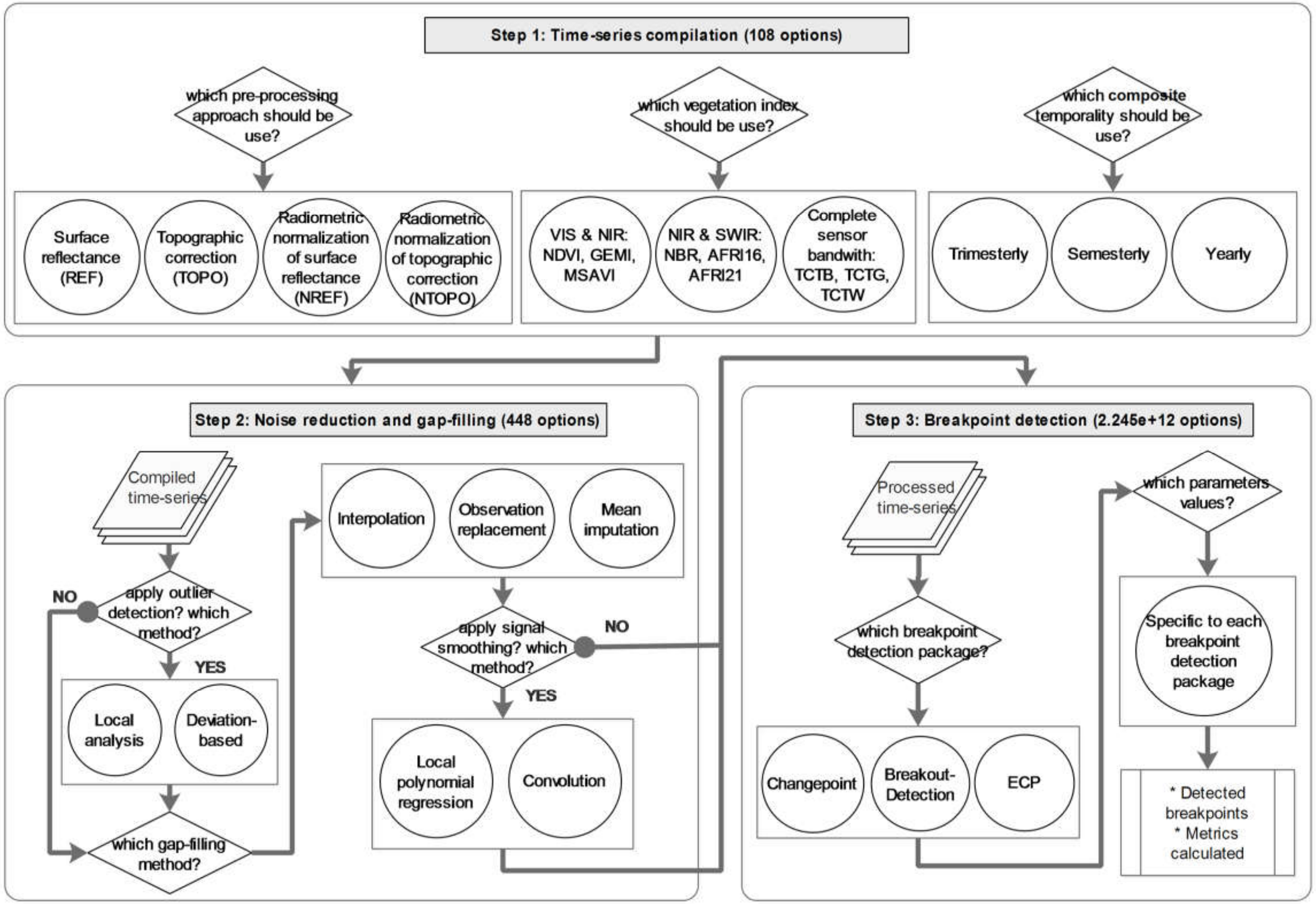

Step 1: Time-series compilation (108 options)
which pre-processing approach should be use?
which vegetation index should be use?
which composite temporality should be use?
Surface reflectance (REF)
Topographic correction (TOPO)
Radiometric normalization of surface reflectance (NREF)
Radiometric normalization of topographic correction (NTOPO)
VIS & NIR: NDVI, GEMI, MSAVI
NIR & SWIR: NBR, AFRI16, AFRI21
Complete sensor bandwith: TCTB, TCTG, TCTW
Trimesterly
Semesterly
Yearly
Step 2: Noise reduction and gap-filling (448 options)
Compiled time-series
NO
apply outlier detection? which method?
YES
Local analysis
Deviation-based
which gap-filling method?
Interpolation
Observation replacement
Mean imputation
apply signal smoothing? which method?
NO
YES
Local polynomial regression
Convolution
Step 3: Breakpoint detection (2.245e+12 options)
Processed time-series
which breakpoint detection package?
Changepoint
Breakout-Detection
ECP
which parameters values?
Specific to each breakpoint detection package
* Detected breakpoints
* Metrics calculated

Figure 11. PCG workflow.

Table 3. Summary of PCG steps, methods, algorithms, packages, parameters, values, steps, and their search space size.

Step	Method	Parameter [1]	Value	Step	Total [2]
1: Time-series compilation	Random selection	P1: Pre-processing and composites temporality	REF, TOPO, NREF, NTOPO	-	108
			Yearly, semesterly, trimesterly	-	
		P2: Vegetation indices	AFRI16, AFRI21, GEMI, MSAVI, NBR, NDVI,TCTB, TCTG, TCTW	-	
2: Outlier detection and removal [3]	Decision	P3: Apply it?	No, LOF, deviation	-	448
	Required for all	P4: Window size	2–5	1	
	Local outlier factor (LOF)	P5: LOF factor	2–15	2	
	Deviation-based test (deviation)	P6: Threshold	0–14	1	
3: Gap-filling	Decision	P7: Which method?	Interpolation, locf, mean imputation	-	54
	Interpolation	P8: Type	Lineal, spline, stine	-	
	Observation replacement (locf)	P9: Direction	Forward, backward	-	
	Mean imputation	P10: Type	Mean, median, mode	-	
4: Signal smoothing [3]	Decision	P11: Apply it?	No, polynomial, convolution		98
	Local polynomial regression fitting (polynomial)	P12: Degree smoothing	0.5–3	0.19	
	Savitzky–Golay filter (convolution)	P13: Filter order	1–10	1.3	

5: Breakpoint detection	Decision	P14: Which BDA package?	Changepoint, BreakoutDetection, ECP	-	2.245×10^{12}
	Required for all	P15: Segment size	3–10	1	
	Changepoint	P16: Statistical property	Mean, variance, mean and variance	-	
		P17: Algorithm	PELT, binary segmentation	-	
		P18: Max breakpoints	0–10	1	
		P19: Penalty mode	SIC, BIC, MBIC, AIC, Hannan-Quinn, Manual	-	
		P20: Penalty for manual mode [4]	0–1	0.01	
	BreakoutDetection	P21: Penalty percentage [4]	0–1	0.01	
		P22: Penalization polynomial	0–2	1	
	ECP	P23: Significance level [4]	0–1	0.01	
		P24: Moment index [4]	0.1–2	0.01	

[1] "P" and number indicate the code used in this manuscript for identifying parameters. [2] Refers to the total search space as number of alternatives. [3] Modules are optional in the processing chain. [4] Threshold parameter.

Following the stacked time-series, different noise reduction and gap-filling algorithms could be applied in the PCG. The following sub-sections summarize the available options, their characteristics, parameter space sizes, and developer references.

2.4.2 Step 2: Noise Reduction and Gap-Filling

As the PCG required us to define how and whether to enhance the quality of the compiled time-series, multiple algorithms were integrated in the PCG to enhance breakpoint detection:

1. Outlier detection, which constitutes the detection and removal of anomalous values in the time-series. Typically, they constitute cloud and shadow pixels that were not properly masked. This step was considered optional and its principal parameters were window size, which defines the moving window to analyze the time-series, and the threshold for sensibility control. In total, 448 alternatives were available in the parameter search space. Two algorithms were available for this step:
    - Local analysis, using the local outlier factor algorithm (LOF) (Breunig M. M. et al. 2000), which calculates for an object a metric to measure how isolated it is with respect to the surrounding neighborhood.
    - A deviation-based test using the Hampel filter, which calculates the median absolute deviation to find outliers (Borchers 2015).
2. Gap-filling, which constitutes the procedure for filling data gaps in the time-series caused by cloud and shadow masking or the Landsat 7 ETM+ sensor scan line error. Gap-filling was a compulsory step to enable breakpoint detection. The parameters in this step refer to the different methods and options for its execution. Thresholds were not required, so parameter search space was limited to 54 alternatives. Three algorithms were available:
    - By interpolation, which estimates new data points with methods such as linear, spline, and stine interpolation (Moritz 2016).
    - Observation replacement, through which missing values are replaced with a temporally adjacent value—the last observation being carried forward or the subsequent observation carried backward.

- Mean imputation, which replaces missing values with either median, mean, or mode data values.

3. Signal smoothing, which enables the increase of signal-to-noise ratio in a time-series and enhances the detection of breakpoints. This step was optional since it is not required in breakpoint detection. For all the algorithms included, only one parameter was required to define a threshold for sensibility control. In total, 98 alternatives constituted the parameter search space. Two algorithms were available:
   - Local Polynomial Regression Fitting (Chambers 1992), which involves the local fitting of a polynomial surface determined by one or more data points.
   - Convolution, using the Savitzky–Golay filter (Savitzky & Golay 1964), which fits a linear least-squares low-degree polynomial to successive subsets of adjacent data points.

2.4.3 Step 3: Breakpoint Detection

The ultimate step in the PCG is the detection of breakpoints. After it is executed, the PCG gives as outputs the position of the breakpoints found within the time-series analyzed. For this purpose, three different BDA packages were considered since they were programmed in the open source R language (R Development Core Team 2017), and this made it possible to observe their codes and integrate them easily to the GA used (Scrucca 2013). Other alternatives, such as BFAST (Verbesselt et al. 2010), were discarded as our LTS was not extensive enough to fit a harmonic model. Based on different approaches, two types of BDA were considered:

1. BDA based in parametric statistics:
   - Methods for Changepoint Detection R package (Changepoint) (Killick & Eckley 2014), which implements the detection of changes in mean and/or variance values of a univariate time-series based on a normality test or cumulative sums. Additionally, Changepoint requires a penalty approach provided by any of five included automatic methods: Hannan-Quinn or Schwarz (SIC), Bayesian (BIC), Modified Bayesian (MBIC), and the Akaike information criterion (AIC). Finally, it includes three algorithms, but we focus on only two. The first one, called binary segmentation, applies a statistical test to the entire dataset and splits it if a single

breakpoint is identified. Afterwards, it repeats this operation until no breakpoints are found. The second, called PELT (pruned exact linear time), is based on the algorithm of Jackson et al. (2005) but it reduces computational cost and improves accuracy by using a dynamic programming approach.

2. BDA based on non-parametric statistics:
    - Breakout detection via Robust E-statistics R package (BreakoutDetection) (James et al. 2014) is an algorithm based on energy statistics. It implements a novel statistical technique called energy divisive with medians (E-divise) for estimating the statistical significance of breakpoints through a permutation test. Its design aims to be robust against anomalies.
    - Non-Parametric Multiple Change-Point Analysis of Multivariate Data package (ECP) (James & Matteson 2015). This package constitutes an upgrade of the E-divise algorithm developed by the authors of BreakoutDetection. Its novelty is the use of E-divise together with the E-agglomerative technique to reduce processing time by segmenting data before analyzing it.

To avoid any bias caused by unequal search space sizes between BDA packages, threshold parameter ranges were set to similar sizes. Considering all parameters combinations, a total of 2.2×10^{12} alternatives were available in this step.

2.5 Genetic Algorithm Implementation

Since GA is the search algorithm for analyzing the PCG parameter search space, in this chapter we describe the concept behind it, the sampling needed for its implementation, the associated reference data, the fitness evaluation, and the calibration applied to GA.

2.5.1 Genetic Algorithms

GAs are a class of evolutionary algorithms that include scientific models of evolutionary process and search algorithms (Iovine et al. 2005). They became popular thanks to the work of John Holland and his colleges during the 1970s (Holland 1973), and are inspired by principles of natural selection and survival of the fittest. GAs have four principal advantages: the capability to solve complex problems, an emphasis on global searches, the provision of multiple solutions, and parallelism.

The workflow of GAs aims at evolving a population of feasible solutions for a target problem, or objective function. A fitness evaluation is then made in which each solution is executed and evaluated for its performance with respect to a validation dataset. This process involves an evolution-directed search in the parameter search space of the objective function to create new solutions. In each new generation, the best solutions are selected, reproduced, and mutated with respect to a set of operators used by the GA:

- Selection: Controls the survival of the best chromosomes according to their fitness value and selects them as parents of a new generation.
- Elitism: An optional selection rule that selects the best chromosomes directly for the next generations to guarantee their existence.
- Crossover: An operator that creates offspring of the pairs of parents from the selection step. Its probability determines if the offspring will represent a blend of the chromosomes of the parents.
- Mutation: Adds random changes to the chromosomes.

To facilitate computation and gene representation, integer encoding is used by GA. This allows the transformation of discrete and continuous parameter values to integer codes and permits GA operators to experiment with them as if they were genes. Since discrete parameters are represented by an exact number of alternatives, their representation as integer codes is relatively easy. However, for continuous parameters, it is necessary to subset their ranges by steps in order to facilitate their operation. Therefore, this process should be carefully defined according to the sensibility behavior of the parameter and the instructions provided by the developer of the algorithm analyzed.

Since GA terms require a homologation, the reader should keep in mind that the following terms refer to: (1) PCG as objective function; (2) all possible methods and parameter values in the PCG as search space; (3) accuracy assessment of detected breakpoints by the PCG as fitness evaluation; (4) any PCG calibration as chromosome and (5) a parameter value for any algorithm in the PCG as a gene. For a better understanding of these concepts, Figure 12 shows how the PCG is integrated with GA.

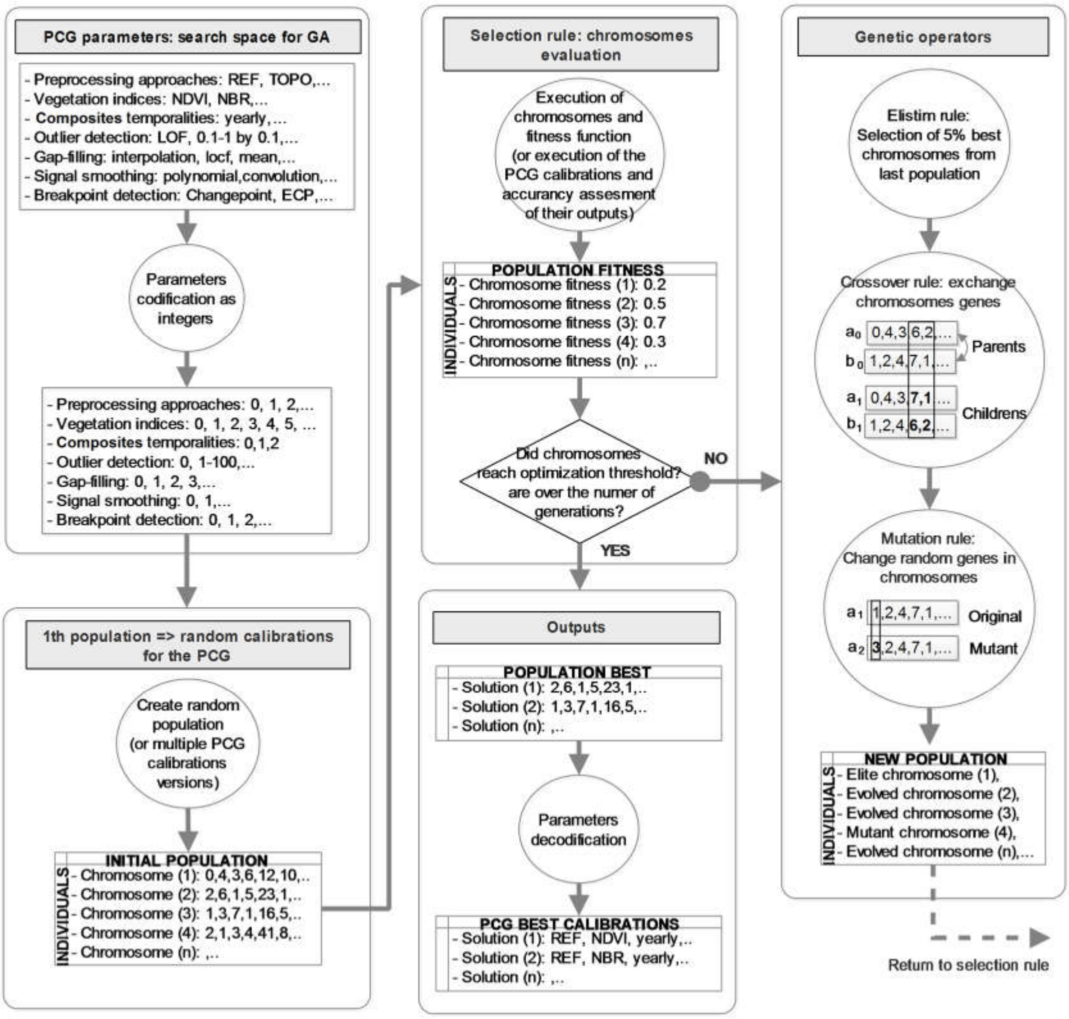

PCG parameters: search space for GA
- Preprocessing approaches: REF, TOPO,...
- Vegetation indices: NDVI, NBR,...
- Composites temporalities: yearly,...
- Outlier detection: LOF, 0.1-1 by 0.1,...
- Gap-filling: interpolation, locf, mean,...
- Signal smoothing: polynomial,convolution,...
- Breakpoint detection: Changepoint, ECP,...
Parameters codification as integers
- Preprocessing approaches: 0, 1, 2,...
- Vegetation indices: 0, 1, 2, 3, 4, 5, ...
- Composites temporalities: 0,1,2
- Outlier detection: 0, 1-100,...
- Gap-filling: 0, 1, 2, 3,...
- Signal smoothing: 0, 1,...
- Breakpoint detection: 0, 1, 2,...
1th population => random calibrations for the PCG
Create random population (or multiple PCG calibrations versions)
INITIAL POPULATION
INDIVIDUALS
- Chromosome (1): 0,4,3,6,12,10,..
- Chromosome (2): 2,6,1,5,23,1,..
- Chromosome (3): 1,3,7,1,16,5,..
- Chromosome (4): 2,1,3,4,41,8,..
- Chromosome (n): ,..
Selection rule: chromosomes evaluation
Execution of chromosomes and fitness function (or execution of the PCG calibrations and accurancy assesment of their outputs)
POPULATION FITNESS
INDIVIDUALS
- Chromosome fitness (1): 0.2
- Chromosome fitness (2): 0.5
- Chromosome fitness (3): 0.7
- Chromosome fitness (4): 0.3
- Chromosome fitness (n): ,..
Did chromosomes reach optimization threshold? are over the numer of generations?
NO
YES
Outputs
POPULATION BEST
- Solution (1): 2,6,1,5,23,1,..
- Solution (2): 1,3,7,1,16,5,..
- Solution (n): ,..
Parameters decodification
PCG BEST CALIBRATIONS
- Solution (1): REF, NDVI, yearly,..
- Solution (2): REF, NBR, yearly,..
- Solution (n): ,..
Genetic operators
Elistim rule: Selection of 5% best chromosomes from last population
Crossover rule: exchange chromosomes genes
a0 0,4,3 6,2,...
b0 1,2,4 7,1,...
Parents
a1 0,4,3 7,1 ...
b1 1,2,4 6,2 ...
Childrens
Mutation rule: Change random genes in chromosomes
a1 1,2,4,7,1,...
Original
a2 3,2,4,7,1,...
Mutant
NEW POPULATION
INDIVIDUALS
- Elite chromosome (1),
- Evolved chromosome (2),
- Evolved chromosome (3),
- Mutant chromosome (4),
- Evolved chromosome (n),...
Return to selection rule

Figure 12. GA applied to PCG workflow.

Depending on the complexity of the objective function, GA can be computing-intensive. In this instance, the GA R package (Scrucca 2013), which allows for flexible manipulation and analysis of the GA, was used. This package was developed for a parallel computing environment and has been proven to solve different cases of optimization problems, such as the Rastringin function, Andrews Sine function, curve fitting, and subset selection, among others.

There are other GA drawbacks. Convergence toward local rather than global optimization can be an infrequent problem. Larger populations and more generations are needed to obtain better results, but this demands large simulations. Moreover, due to GAs reliance on a stochastic search method, each run constitutes a different approximation to the solution as its initial population is randomly created (Maaranen et al. 2007). This can affect solution finding as GA can become stuck in a subset of parameter search space, depending on its execution of a random mutation for exploring new parameter combinations.

2.5.2 Training Sampling Design

To get training samples for the GA task, an optimal allocation technique to perform a stratified sampling using a random scheme was applied. This method minimizes the variance of the estimated overall thematic accuracy of the dataset and is recommended for use when the number of map classes is relatively limited (Olofsson et al. 2014).

To calculate sample sizes, the LCCM (see section 2.3) classes were used to extract two stratums, which we named change (merging forest gain and forest lost classes) and no-change (using stable forest class). The size of each strata was allocated according to its expected variances. For this step, we followed the sampling design procedure outlined by Olofsson et al. (2014), under which informed conjecture of the user´s accuracy can indicate these values. Applying a sampling precision of 70% and standard deviations of 1 and 4, the samples were calculated for all validation areas (100 in total, 54 for no-change and 46 for change areas). Since sample size constitutes an important factor in GA, because it controls the amount of data that chromosomes should ingest to produce breakpoints in a manageable processing time (approximately 5 to 15 min.) (Figure 13); we considered these samples sufficient, given our limited computing resources (4 cores, Intel i5 processor) for following the GA calibration recommendations (see section 2.5.4).

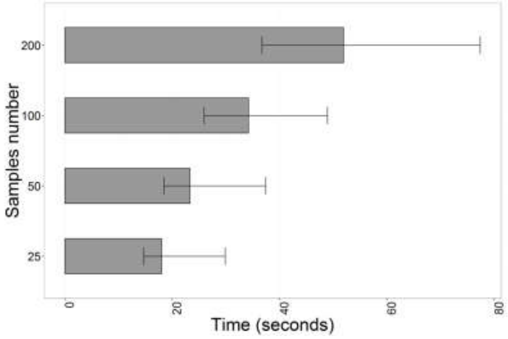


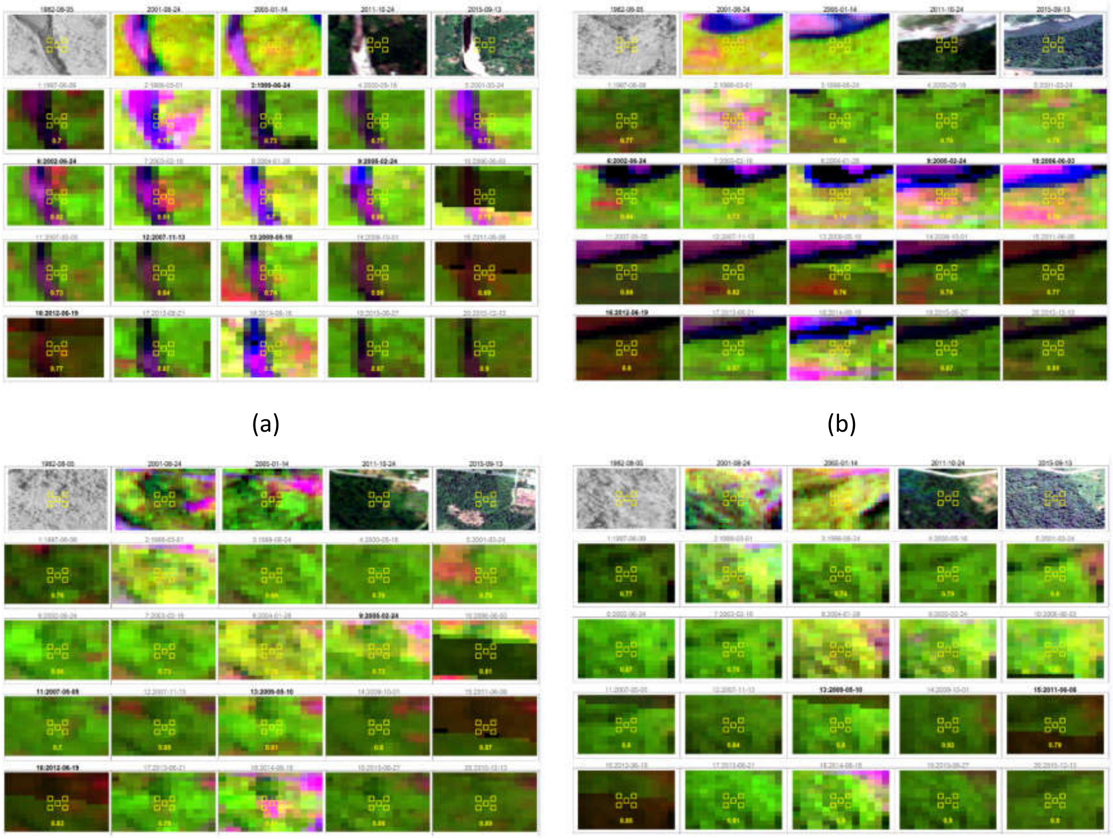
Figure 13. Processing time measured for different sample size sets. The PCG was equally configured in all cases, executing GA for 10 times to average its processing time. For testing purposes, GA population size was set at 30 individuals and 10 generations.

2.5.3 Samples Interpretation and Fitness Evaluation

The fitness evaluation includes the design and execution of a specific function for evaluating proposed solutions during the GA execution. The fitness evaluation in our case was formed by an accuracy assessment of the PCG outputs. Our implementation for this task again followed the recommendations of Olofsson et al. (2014) for two components: response design and accuracy analysis.

Starting with the response design, we chose an array of 5 neighbor pixels arranged in an “x” shape as the sampling unit and applied the visual interpretative approach. For this purpose, we plotted yearly composites of 5, 4, and 3 Landsat bands, composed as color compositions, along with high-resolution images. This task was facilitated by a developed tool similar to the TimeSync system (Cohen W. et al. 2010) but with specific functionalities (Figure 14). The protocol for interpreting each sample unit can be summarized as follows.

1. Any of four conditions was a basis for rejecting and replacing a sample, or in some cases, a neighbor pixel in the sample unit, if:
    - There was a lack of high-resolution images, obscuring most of the land-cover history of the sample unit;
    - Visual interpretation of the sample unit was compromised because of unmasked clouds, shadows, haze, or water bodies (Figure 14a);
    - A neighbor pixel exhibited a land-cover history different from the majority of pixels included in the sample unit;
    - The number of eliminated neighbor pixels in a sample unit exceeded 4 cases.
2. To consider a sample unit as a forest gain, clear evidence of continuous regeneration was needed. According to Guariguata and Ostertag (2001), early forest development within the tropical forest environment is achieved approximately five years after the disturbance event. We followed this condition to accept samples as forest gain (Figure 14b) or reject samples and cast replacements.
3. In the case of forest loss samples, an absence of the regeneration period was a necessary condition for validation (Figure 14c).



(a) (b)

(c) (d)

Figure 14. Interphase for samples interpretation. First row (from left to right) refers to high resolution imagery composed by aerial photos (1982) and images from ASTER (2001 and 2006), Rapid Eye (2011) and Google Earth (2015). The rest of the image chips correspond to the LTS as yearly RBG composites. Following protocols, we interpret these cases as follows: **(a)** Sample rejected by unmasked river, **(b)** Forest gain after disturbance observed in the 3rd row, **(c)** Forest lost with a short regrown period, and **(d)** Stable forest sample.

4. For stable forest samples, the unique condition required was an absence of disturbances (such as clear cuts or fires) during the period analyzed (Figure 14d).

As visual interpretation of breakpoints dates is not an easy task, we preferred to ensure that a sample unit belonged to the classes described in the protocol instead of introducing more uncertainties by subjective interpretation of disturbance dates. Subsequently, the samples were grouped into change and no-change classes to apply this information to training samples.

For the accuracy analysis, using a straightforward matching procedure, the fitness value of our samples was based on the calculation of their overall accuracy. This measure helped us to overcome local optimal convergence experienced when R^2 and adjusted R^2 were tested. Other measures such as kappa, omission and commission errors, etc., were avoided since only two classes were examined.

2.5.4 GA Calibration Settings

Following the interpretation of training samples, the GA was executed. The procedure for analyzing the search space of the PCG and calibrating GA according to its characteristics followed the recommendations of Kumar et al. (2012), and Eshelman et al. (1991). These recommendations guided the design of two types of searches. The first mode, called “explorative” was created to reduce the size of the search space of the processing chain and provide insights for its calibration as a whole. For this, BDA threshold parameters with a search space size of 100 alternatives were contracted, increasing step values and shortening its range to 12 alternatives (left side of Figure 15).

Other specific GA calibration parameters were applied following the recommendations of Scrucca (2013). This application indicated an elitism rate of 0.05 and a uniform crossover probability of 0.8. A random mutation rate was tested with 0.5 and 0.1 for experimental purposes. The tournament selection was the mechanism set to favor the selection of better individuals, as suggested by Miller et al. (1995).

The second mode, called “exploitation,” was created to analyze in detail BDA threshold parameters. Therefore, its step values were maintained as shown in Table 3, conserving its 100 alternatives in all cases (right side of Figure 15).

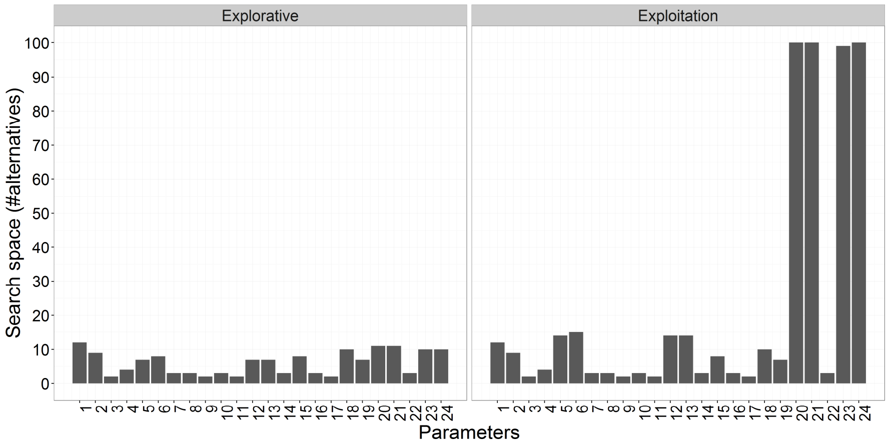


Figure 15. Search space size allowed for the explorative and exploitation modes.

To codify each parameter value as genes, integer codes were applied to discrete and continuous parameters. Details of the steps and ranges applied to continuous parameters are shown in Table 3. With this codification of parameters, the “real-value” GA type was chosen. This allowed us to obtain floating-point representations of numbers, which can be converted into integer values and define a parameter value from the search space.

2.6 Results

After the calibration of GA, in this section we describe the procedures for exploring PCG parameter search space, the optimized PCG calibrations found, their execution for visual and accuracy assessment, and their processing time.

2.6.1 GA Search Approaches and Calibrations Found

To find the best solutions, different GA executions and experiments were implemented. For all BDA packages, these procedures were similar; however, two approaches were applied to consider different management of the PCG parameters search space. For this step, our approach was divided, according to the parameters, by analysis and the type of GA search applied. Two different approaches were considered, where the main difference was the application or non-application of noise-reduction algorithms and a predefined pre-processing input.

1. Search with only BDA packages (reduced processing):
    - In this sequence, the BDA parameter search space (parameters P14-P24 as is noted in Table 3) is analyzed in detail, but the rest of parameters are defined by a default mode.
    - The default mode defined the NBR composites yearly, since short-wave infrared (SWIR) region metrics are more sensitive to forest changes (Kennedy et al. 2010b) and this composites temporality helped to reduce data gaps. They were calculated from the Landsat surface reflectance products for reducing uncertainties with respect to radiometric normalization and topographic correction performance. Noise reduction algorithms, including outlier detection and signal smoothing, were deactivated, which focused the GA on BDA parameter search space exclusively.

    - For this approach, the GA was set to the “exploitation” mode, executing it twice with each BDA package. The first run had a mutation rate of 0.1 and the second run, a rate of 0.5 (enhanced diversity). Each GA execution had 120 generations and 30 individuals.
2. Search with pre-processing and noise reduction + BDA packages search (extended processing):
    - This search consisted of split search space, executing shorter GA runs, and splitting pre-processing and noise reduction parameter search space from each BDA.
    - To find a calibration solution for the pre-processing and noise reduction algorithms (parameters P1-P13 as noted in Table 3), GA was set to the “explorative” mode during a first run of 60 generations and 30 individuals.
    - After determining a pre-processing and noise reduction calibration, it was applied before the BDA parameters search space was analyzed in detail. For this step, a second GA run was applied, changing the search mode to “exploitation” but only for 60 generations of 30 individuals.
    - Hence, GA was executed four times: two for pre-processing and noise reduction, and two for the BDA parameters search space, considering again the two mutation rates applied in the first approach (0.1 and 0.5).

Results from these sequences and the best solutions achieved are shown in Figure 16, where the fitness value and its peaks are highlighted as colored dots. These peaks suggest that the ECP and BreakoutDetection packages performed best when a reduced processing approach (left side of Figure 16) was applied. Furthermore, all packages improved their results drastically after applying the extended processing approach, which assigns a pre-processing and noise reduction calibration before the BDA is optimized by GA. The right side of Figure 16 shows the enhancement that Changepoint, BreakoutDetection, and ECP provided, increasing their overall accuracy: 0.03, 0.11 and 0.13, respectively. This demonstrates that the extended processing approach was better for optimization, providing us with better calibration solutions, so we chose these enhancements for further analysis.

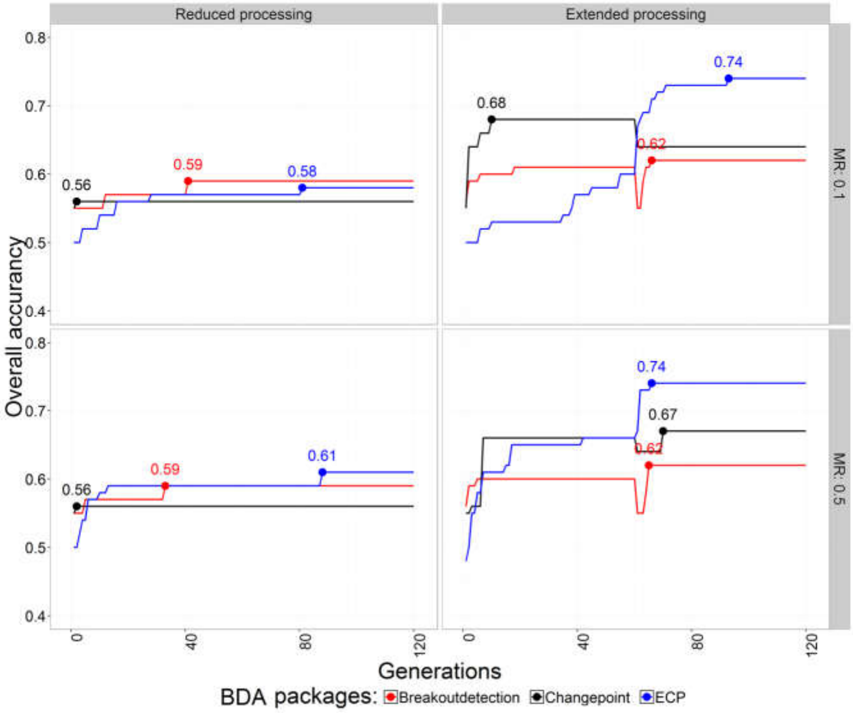


Figure 16. BDA fitness value evolution.

Details for the extended processing calibration solutions are shown in Table 4. The time-series compilation step shows that yearly composites and vegetation indices based on infrared bands (NBR and AFRI16) were mostly preferred; however, is not clear which radiometric correction was most beneficial since topographic correction (TOPO), surface reflectance (REF) and radiometric normalization of topographic correction (NTOPO) were all present in the different calibrations. Noise reduction algorithms, such as outlier detection and signal smoothing, showed that calibrations using LOF and polynomial regression methods gave improved results. This was not the case for the gap-filling step, where different methods were selected. Regarding BDA calibrations, the segments size parameter showed different behaviors between BDA packages. Because of this, the number of breaks obtained by each BDA package differed, but we maintained them because of our experimental design. The rest of the BDA parameters were unique and incomparable, as each of them was specific to each BDA package parameters.

Table 4. Parameter selection by GA as best calibration solutions.

PCG Step	BDA Package		
	Breakout Detection	Changepoint	ECP
1: Time-series compilation	TOPO, yearly, NBR	NTOPO, yearly, MSAVI	REF, Yearly, AFRI16
2.1: Outlier detection	LOF, window size 2, LOF factor 8	LOF, window size 4, LOF factor 2	LOF, window size 2, LOF factor 4
2.2: Gap-filling	Observation replacement, forward direction	By interpolation, spline method	By interpolation, linear method
2.3: Signal smoothing	Polynomial, degree smoothing 0.5	Polynomial, degree smoothing 1.7	Polynomial, degree smoothing 1.7
3: Breakpoint detection	Segment size 5, penalty percentage 0.09, penalization polynomial 1	Size segments 3, variance as statistical property, binary segmentation algorithm, 7 breakpoints, manual penalty mode, penalty in manual mode 0.5	Size segments 9, significance level 0.2, moment index 0.94
Overall accuracy	0.62	0.68	0.74

2.6.2 Execution of Calibration Solutions and Visual Assessment

After the best calibration profiles were determined, we proceeded to apply them to each study area (see Figure 17), considering that each LTS pixel must include a minimum of 10 observations (a full time-series pixel includes 20 observations). Therefore, it was possible to observe BDA performances under different conditions. Artifacts such as horizontal stripes indicated Landsat 7 scan-off error, while irregular patches showed areas of unmasked clouds and shadows. Despite these hindrances, some patterns of land-cover change were present in all outputs. In this regard, ECP (Figure 17, column d) looked less affected in comparison to the reference change 1997–2015 mask (Figure 17, column a). The rest of the BDA packages, such as Breakoutdetection and Changepoint, did not perform well in all areas as compared to the change 2000–2015 mask. Following

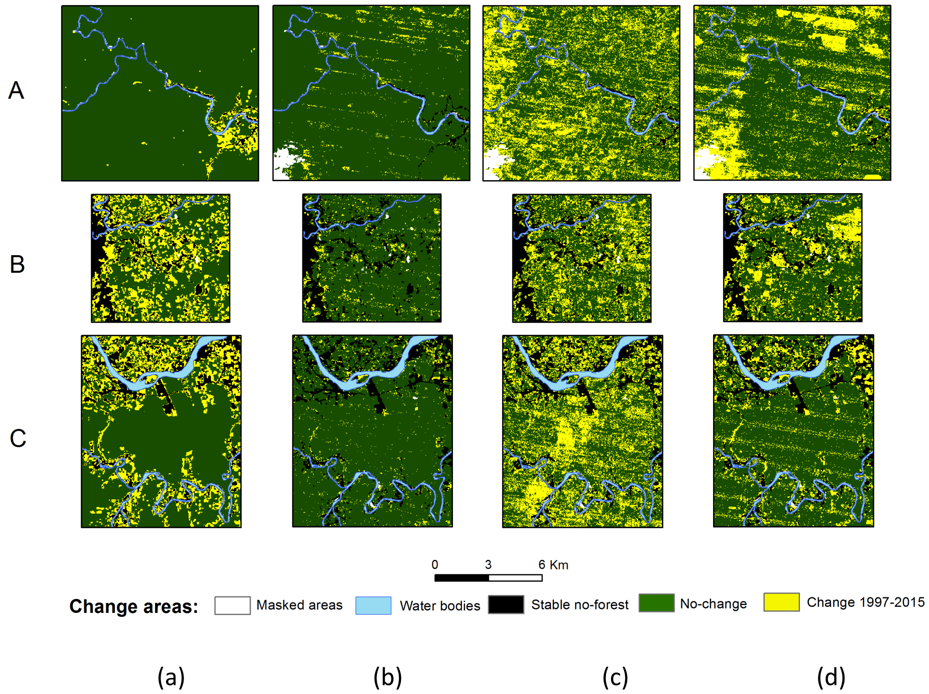
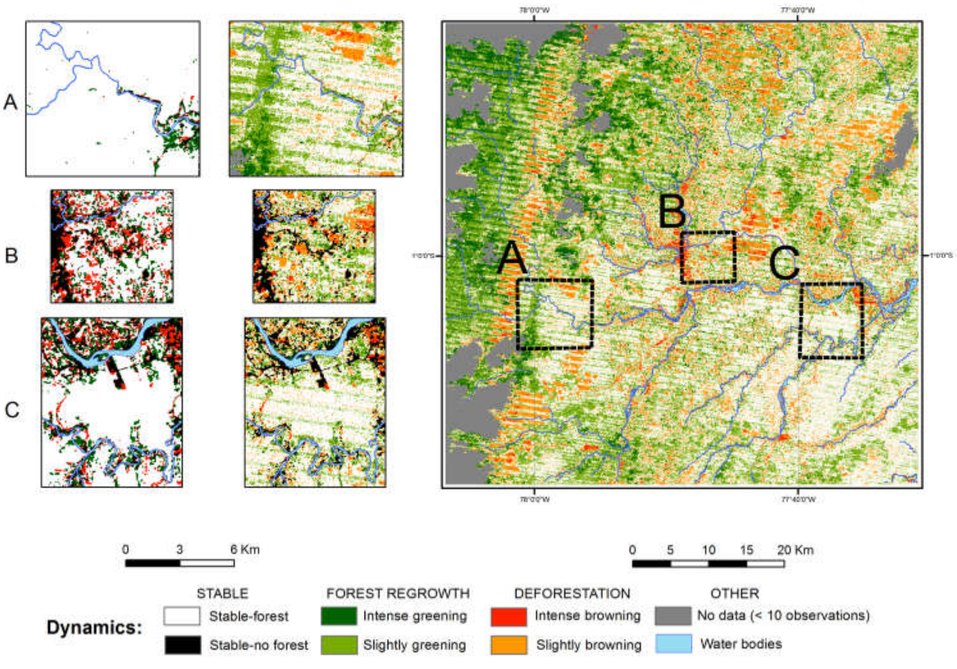


Figure 17. Spatial representations of the best BDA solutions. Study areas are shown in rows (A, B, C) and columns refer to **(a)** Change 1997–2015 mask, **(b)** BreakoutDetection, **(c)** Changepoint, and **(d)** ECP results. In all cases, no-forest areas before 1997 are shown in black, while areas not analyzed are shown in white.

these initial insights, the ECP package was selected for further analysis and manual calibration described in the next section.

2.6.3 Manual Optimization of ECP Calibration Solution and Execution of the PCG for a Larger Area

As GA provided us with a preliminary calibration of the ECP processing chain, we started to play with different parameters in order to see if we could improve results. After experimenting with different configurations, we observed that the only parameter that enhanced results was segment size, since it affected the number of breakpoints detected in the time-series. Therefore, we redefined this parameter into 5 (with the condition that the first breakpoint detected should be observed from the 2001 data) but we kept the other parameters as shown in Table 4. We then executed the PCG, but now applied to a larger area that contained the three study areas (Figure 18, column c). To differentiate forest dynamics classes (forest gain and forest loss), the breakpoints identified by the PCG were used to isolate two segments: (1) the “before-change” segment, which takes the segment in the time-series that happens before the first



(a) (b) (c)

Figure 18. Application of the ECP-optimized processing chain. Study areas are shown in rows (A, B, C) and columns refer to **(a)** LCCM, **(b)** The forest dynamics map obtained, and **(c)** its application to a larger area.

breakpoint was detected; and (2) the "after-change" segment, which takes the segment in the time-series that happens after the last breakpoint was detected. With each of them, the mean value was calculated and subtracted to obtain a residual. This was classified into 5 natural breaks classes for enhancing different levels of greening and browning patterns, which are associated with forest gain and forest loss dynamics [6] (Figure 18, column b). As this layer highlights these dynamics, it was used for accuracy assessment using the reference samples extracted from the LCCM (Figure 18, column a). This procedure is explained in the next section.

2.6.4 Accuracy Assessment

The accuracy assessment followed a similar procedure for training (see sections 2.5.2 and 2.5.3) but emphasized the results obtained for the forest dynamics map. To determine a sample size, we used a sampling precision of 75% and standard deviations of 1, 4, and 4 for the classes stable forest, forest gain and forest loss, respectively. The stratums were again extracted from the LCCM and this gave us a total of 551 samples, divided into 226 stable-forest, 159 forest-gain and 166 forest-loss samples. Reference

Table 5. Confusion matrix for the forest dynamics map.

Assigned Classes	Referenced Classes			User Accuracy
	Stable-Forest	**Forest Gain**	**Forest Loss**	
Stable-Forest	691	165	121	0.70
Forest Gain	230	278	23	0.52
Forest Loss	81	35	346	0.74
Producer Accuracy	0.68	0.58	0.70	Overall Accuracy: 0.66

information for interpreting the samples follows the same criteria explained in training samples interpretation (see section 2.5.3). Since the forest dynamics map contains five classes, it was reclassified for match the categories of the LCCM. This led to merging intense and slightly greening as forest gain, intense and slightly browning as forest loss, and stable class as stable forest. Results from the confusion matrix using the evaluated pixels (1970 in total) for each sample are shown in Table 5.

Overall accuracy for the three study areas indicates that this map achieved an overall accuracy of 0.66, where stable forest and forest loss were less inaccurate than forest gain (Table 5). This was less than we predicted (≥0.74 of overall accuracy) based on GA accuracy results. Continuing with the accuracy assessment, other measures were calculated for each study area to evaluate its specific precision. This is shown in Table 6.

Table 6. Accuracy measures for the three study areas.

Study Area	Class	Measure			Overall Accuracy
		Sensitivity	**Specificity**	**Balanced Accuracy**	
A	Stable forest	0.58	0.58	0.58	0.56
	Forest gain	0.34	0.71	0.52	
	Forest loss	0.50	0.88	0.69	
B	Stable forest	0.59	0.70	0.64	0.62
	Forest gain	0.56	0.83	0.69	
	Forest loss	0.70	0.89	0.80	
C	Stable forest	0.85	0.72	0.78	0.76
	Forest gain	0.63	0.89	0.76	
	Forest loss	0.72	0.97	0.84	

These results show that overall accuracies were particularly poor for study areas A and B, but better for Area C, so we further describe that section’s accuracy results by class. By class, stable-forest and forest-loss classes are shown to be easier to detect, as sensitivity and specificity were greater than 0.7; but forest gain is harder, since sensitivity was below 0.7. As balanced and overall accuracy was greater than 0.75, breakpoint detection seems to have worked better for this unique case.

2.6.5 Speed Performance

As we considered processing time a key issue in breakpoint detection, average times were measured for each BDA calibration solution, running each 10 times with a sample set of 100 pixels on an Intel i7-processor-based personal computer with 8 cores. The results are shown in Figure 19a: the ECP optimized processing chain took the longest (6.5 s), followed by the Changepoint and BreakoutDetection packages (5.11 and 5.03 s, respectively). Applied to our larger area (~4400 km^2, ~5 million pixels), processing time is estimated at ~4.9 h for BreakoutDetection, ~5.06 h for Changepoint and ~6.4 h for ECP. However, this depends on the algorithms associated to their processing chains and not on the breakpoint detection packages themselves.

Total time for optimization using the two GA approaches are shown in Figure 19b. It was observed that extended processing (0.5–1 h) more than doubles the time required by the reduced processing approach (0.13–0.27 h). This was partly because in the extended processing approach, GA must run the BDA packages twice, as well as running the rest of the algorithms included for noise reduction.

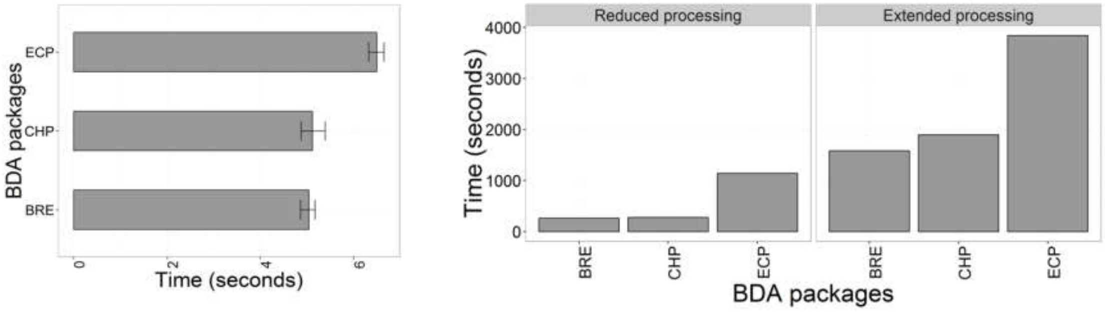


Figure 19. Execution time for: (a) Each BDAs optimized processing chain, running 10 times with a sample set of 100 pixels. (b) Total GA execution time obtained for each BDA package and approaches applied. Acronyms used for BDAs are: BRE for BreakoutDetection; CHP for Changepoint; and ECP for the package of the same name.

2.7 Discussion

In this paper, we present the application of a GA for optimizing a processing chain and evaluate if breakpoint detection is applicable for monitoring forest dynamics in the Andean Amazon in Ecuador. We tested this procedure on three different study areas characterized by limited data availability, high topographic relief, and landscape variability. To organize this discussion, we divide this chapter into three subsections.

2.7.1 Objections Regarding Breakpoint Detection and Its Applicability in the Andean Amazon

Our results show that breakpoint detection using LTS failed for the study areas A and B since breakpoints were not properly detected, while in Area C, results were better despite the fact that the forest-gain class was imprecise, as sensitivity was below 0.7. Furthermore, breakpoint dates detected by the ECP-optimized processing chain do not clearly reveal change events. Forest dynamics patterns we observed through the calculated metric constitute an approximation based on assumed dates of detected changes that could not be real. This contrasts the position of DeVries et al. (2015) using BFAST, where forest dynamics monitoring was applied to a tropical montane forest in Ethiopia. In that research, breakpoint dates and magnitudes allowed researchers to track down and classify forest changes, indicating; however, that excessive unmasked clouds and shadows, and limited data for fit a history model, would restrict the use of the algorithm. Both sources of errors impacted our case, since data availability in some areas was less than 41 images (reddish areas, see Figure 7b) with some of those affected by unmasked clouds and shadows, in some cases resulting in one or no usable observations for a year or more. This prevented our using that algorithm, supporting our objection regarding breakpoint detection and its applicability in some areas of the Andean Amazon.

On the other hand, our exercise of calibrating the PCG and implementing breakpoint detection with scarce and noisy data conditions introduced additional problems since parametrization was complex and had to be solved by a GA. On this point, if we consider the implementation of breakpoint detection for a larger area in the Andean Amazon, calibration could be a problem since parametrization will be unique to each landscape configuration and its specific data conditions. For this reason, the processing time required to optimize such processing chains also constitutes a drawback. As we have seen, GA is a computing-intense simulation the implementation of which was only possible when a small training sample size was applied to optimize the PCG. This certainly affected the effectiveness of the optimization, since overall accuracy between training and reference samples differed significantly (0.74 against 0.66, respectively, for the ECP-optimized processing chain). Finally, even if the calibration problem is solved, a complete Landsat footprint could require 49 h of processing time using a personal computer with 8 cores (Intel i7 processor) and the ECP-optimized processing chain. Considering these results, processing time is certainly excessive for a single Landsat footprint, whose noisy outputs, applied on a relatively larger scale in our research, did not satisfy our expectations regarding its use in the Andean Amazon.

2.7.2 Considerations When Data Quantity Allows the Implementation of Breakpoint Detection

Despite these objections, some results deserve to be mentioned as they seem to diminish the effect of some problems and improve breakpoint detection in some specific cases observed in our study areas. First of all, as Area C obtained an improved result, data quantity and geometric accuracy are the first factors to consider for a successful implementation. As seen in Figure 7b, the spatial distribution of no-data frequency in Area C was almost homogeneous. This suggests that the threshold required for using the ECP-optimized processing chain was around 55 to 71 observations. However, these observations should be distributed along the time-series in order to have yearly composites almost free of gaps and properly co-registered in order to allow breakpoint detection (Roy et al. 2010). As data quantity requirements are not always made explicit by developers of BDAs, and geometric accuracy of image subsets could be different than the value reported on its metadata, it is worth knowing these aspects before a BDA is considered for its use.

To diminish the effects of atmospheric contamination, the aerosol-free vegetation index (AFRI), and other indices derived from shortwave infrared bands using only the Landsat surface reflectance products seem to ameliorate this effect and enhance analysis of forest dynamics as Matricardi et al. (2010) also found. However, it is seen in Figure 18c that the application of the ECP-optimized processing chain for a larger area was affected by radiometric differences between Landsat footprints. This was concluded after observing the asymmetries between the left and right sides of the metric calculated. Therefore, users of Landsat surface reflectance products should note that sensor calibration is not homogeneous and radiometric normalization should be applied. In our case, the relative radiometric normalization procedure applied does not seem to reduce this effect; therefore, the authors do not recommend its application, and a better method independent from the footprint area should be considered. Furthermore, as cloud and shadow masks fail from time to time, improve masking is our first recommendation. However, as outlier detection with a local outlier factor algorithm helped to reduce the incidence of these pixels as false-positive breakpoints, its application should be considered but carefully calibrated, as forest changes could be erroneously interpreted as outliers in some cases.

With regard to enhancing the calibration sensibility of BDAs, signal smoothing with a local polynomial regression-fitting algorithm helped us to improve it, especially when threshold parameters did not react and a subtle sensibility to breakpoint detection is required. Other researchers found this procedure useful when LTS is used and the adjustment of models requires an improvement of their signal quality (Goodwin & Collett 2014; Powell et al. 2010).

Finally, it was seen that a non-parametric breakpoint algorithm behaved better in our case than parametric algorithms. Apparently, the absence of assumptions of data normality in the ECP package seemed to improve our results. This is also becoming an important research field, since non-parametric algorithms in remote sensing are more flexible for solving tasks where data quality and quantity are not exceptional (Mountrakis et al. 2011; Atkinson & Tatnall 1997). However, as other assumptions could exist, knowledge is a must for proper implementation of these algorithms.

2.7.3 Some Advice for Improved Optimization of Complex Functions with GA

As GA constitutes our innovative contribution to this body of research, some advice regarding its implementation should be considered. First, splitting the search space into modules is a useful strategy to avoid becoming stuck in local optimization and to produce relatively quickly. This is corroborated by Garibay et al. (2003), who indicate the benefits of splitting a search space into small modules instead of analyzing large and complex search spaces as a single module.

Regarding calibration, we found that GA was sped up when a higher mutation rate was applied. This condition predefines a more diverse population, consequently producing more diverse combinations of parameters. This conclusion is also supported by Montana et al. (1989) and Haines et al. (2014), wherein exhaustive GA experiments were conducted and similar results, i.e., where mutation rate and crossover probability significantly influenced GA success, were found. However, as calibration of GAs implies many other parameters, the authors recommend implementing other approaches, such as those analyzed by El-Mihoub et al. (2006), which can simplify this task.

Finally, because GA is computationally demanding, it is obvious but nonetheless important to mention that the workload introduced by the objective function and its training data should be carefully examined. If any of these elements introduce bottlenecks or code bugs during execution, GA will result in an overflow. This condition can limit calibration to small population sizes and lead to premature convergence (Eshelman et al. 1991) and less than satisfactory results. Therefore, processing time should be evaluated with different training data sizes and the objective function should be optimized as much as possible.

2.8 Conclusions

This research demonstrated that breakpoint detection should be carefully evaluated before its application for monitoring forest dynamics in the Andean Amazon using LTS, since insufficient data availability, inaccurate co-registration, radiometric variability in sensor calibration, and unmasked cloud and shadow pixels compromise its implementation in most areas of the Andean Amazon. Moreover, since landscape diversity in the Andean Amazon includes heterogeneous conditions, which require specific calibrations for breakpoint detection, its optimization could be compromised as more processing chains will be required for each specific landscape configuration. This is certainly costly in terms of software development and processing but also inefficient for large-scale monitoring projects in the region. Therefore, users should consider these limitations and suggestions before implementing breakpoint detection in similar landscape conditions.

Acknowledgments

This work was supported by a SENESCYT grant. The contents of this paper are the sole responsibility of the authors. Many thanks to three anonymous reviewers who helped us to re-organize and improve this paper and anonymous English language editors for improving it. Special thanks to the R community for sharing their knowledge and the Landsat and USGS teams for making the archive available. In memory of Gunter Menz, an inspiring and valued friend and mentor.

3 MONITORING LONG-TERM FOREST DYNAMICS WITH SCARCE DATA: A MULTI-DATE CLASSIFICATION IMPLEMENTATION IN THE ECUADORIAN AMAZON

Fabián Santos, Pablo Meneses and Patrick Hostert

"Adapt what is useful, reject what is useless, and add what is specifically your own."

Bruce Lee.

3.1 Introduction

The Amazon rainforest constitutes one of the biologically most diverse, structurally complex, and carbon-rich bioregions of the world (Asner et al. 2014). It performs essential global-scale functions and provides a multitude of ecosystem services (Paula et al. 2014). Tropical deforestation is a major threat to the region and a driver of climate change with potentially critical impacts on the biosphere (Fearnside 2005). Large-area deforestation assessments indicate that the Amazon Basin lost 13.3% of its forest from 2000–2013, where the headwater basins suffered most of the pressure (RAISG 2015). This is also particularly alarming for the region itself, as the highland Amazon (or Tropical Andes) is highly susceptible to global warming (Karmalkar et al. 2008), while being under-researched in deforestation studies (Armenteras et al. 2011). Moreover, only little is known about forest succession (Barbosa et al. 2014) or land-cover intensities (Kuemmerle et al. 2013) in this sub-region, which are also important components for understanding the impacts on ecological services (e.g., on biodiversity, carbon sequestration, or nutrient sinks) (Brown & Lugo 1990; Edwards et al. 2017; Poorter et al. 2016). Monitoring forest dynamics in the Tropical Andes therefore plays a key role for informing policymakers and resource managers in their decision making processes over the next few years (Angelsen & Wertz-Kanounnikoff 2008; De Koning et al. 2011). While lowland tropical forests have been well researched, closing the remaining knowledge gaps on forest dynamics in Andean tropical forests is of prime importance (Spracklen & Righelato 2014; Oliveira et al. 2014; Armenteras et al. 2017; Da Ponte et al. 2015).

Comprehensive forest dynamics monitoring has traditionally implied decades of field observations (Fragal et al. 2016). Such observations are costly and it might even be impossible to collect the necessary data in the field. Remote sensing offers a unique alternative for supplying this information for large areas and in different spectral, spatial, and temporal resolutions. Terra and Aqua on board of the Moderate Resolution Imager Spectroradiometer (MODIS) have regularly been used for broad-scale mapping of forest dynamics (Hansen et al. 2008). However, the spatial resolution of MODIS data is limited when fine-scale disturbance regimes prevail such as from selective logging, skid trails, or disturbances related to landslides or local wind throw. The Landsat sensor family with its 30-m spatial resolution and its 45-year observation record is better suited for capturing long-term and fine-scale processes. It is the most widely used observation system for Land-Cover and Land-Use Change (LCLUC) assessments and forest dynamics monitoring programs (Camara 2013; Hansen & Loveland 2012). Since the launch of Landsat-1 in 1972, the Landsat Program has continuously collected data across the globe and—since the launch of the Landsat Thematic Mapper in 1982—in six spectral bands covering the optical, near-infrared, and shortwave infrared wavelength regions. For these reasons, it is the most long-term medium-resolution Earth observation satellite archive available. Due to the open data policy since 2008, Landsat data is available free-of-charge as a standard high-level product for long-term LCLUC analysis (Wulder et al. 2016). This development allowed major improvements in automated time-series analysis, leading to a range of novel algorithms such as TIMESAT, LandTrendr, BFAST, and CCDC (Eklundh & Jönsson 2015; Kennedy et al. 2010a; Verbesselt et al. 2010; Zhu & Woodcock 2014) that allow for extracting forest dynamics information. However, in some regions around the globe, the archive data density is considerably lower, mostly due to persistent cloudiness (Chance et al. 2016; Arvidson et al. 2001) reduce data quantity and quality. This result in time-series with poor signal-to-noise ratio, due useful information about forest status is weak and not significant to differentiate from random noise. This is a limitation for transferring these novel algorithms to data scarce regions, such as the Tropical Andes (Santos et al. 2017), as time-series analyzes require rather dense data stacks over time.

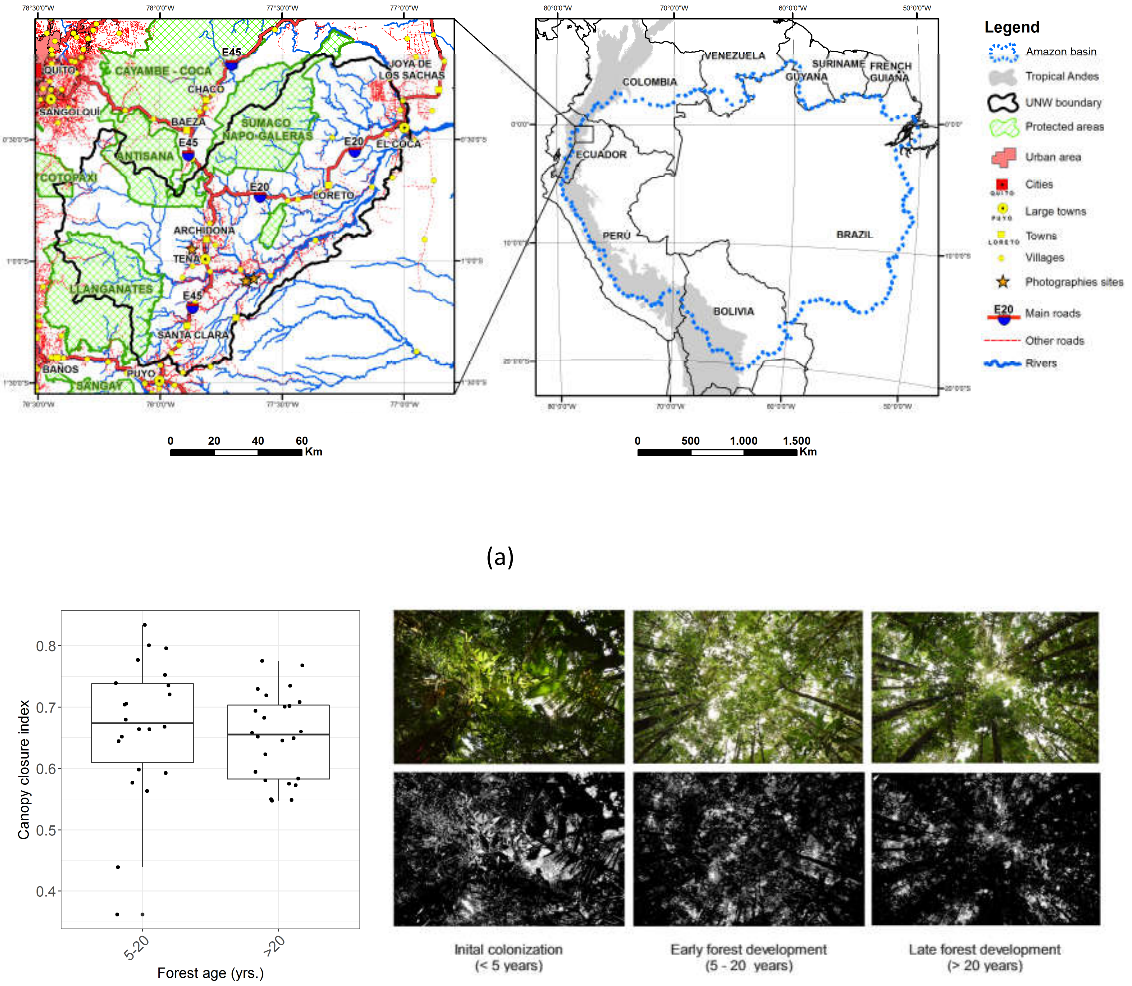
Conceptual approaches including those based on image compositing, multi-sensor fusion, and post-classification change detection have demonstrated their potential to overcome these limitations (Potapov et al. 2012; Hansen et al. 2013; Griffiths et al. 2014). Multi-date classification (Zhu 2017) has the advantage to cope with noise-prone observations, especially when based on cloud-free composites that not only allow identifying deforestation or reforestation processes, but also characterize land-use intensities from post-deforestation dynamics (Rufin et al. 2015). For these reasons, multi-date classification is the methodology of choice for monitoring long-term forest dynamics in areas such as the Tropical Andes. However, implementing a multi-date classification scheme under specific regional conditions can still be challenging, e.g. due to poor data availability in the past depending on historical data receiving strategies or increased cloudiness due to topography. Consequently, we applied this methodology to a study case located in the Amazon region of Ecuador, the Upper Napo Watershed (UNW), where heavy rainfall regimes (Espinoza et al. 2015) and complex landscapes (Asner et al. 2014) are major impediments. Our overarching objective was to monitor long-term forest dynamics and identify deforestation/reforestation for the period between 1992 and 2014. We pose the following research questions related to these objectives:

- Which processing steps and techniques are needed to implement multi-date classification in a data sparse region as exemplified in the UNW?
- How well does a multi-date classification approach perform when monitoring long-term forest dynamics in the environments of the Tropical Andes?

3.2 Study Area

The UNW is located between 78°25´W and 76°25´W longitude, and 0°10´N and 1°30´S latitude (Figure 20a). It covers an area of about 12,500 km^2 in the Ecuadorian Amazon, spreading across the three provinces Napo (63% of the watershed), Orellana (26%), and Pastaza (9%). The altitudinal gradient of the Andes covers ~260 to 5,600 m.a.s.l. Mean temperatures vary from -0 to 26 °C and annual precipitation from 1,100–5,300 mm (MAE 2013). The core rainy season extends from December to May, but fog and clouds are abundant throughout the year, especially at higher elevations (Ramírez et al. 2017). The complex geology creates diverse edaphic conditions that in combination with topographic gradients and climatological impacts result in a multitude of extremely species-rich ecosystems (Hoorn et al. 2010).

The percentage of natural vegetation in the UNW was estimated to be 79% in 2014, with 76% forest, 2% shrub-dominated landscapes, and 1 % grasslands (MAE 2017). Forests are generally evergreen, but forest ecosystems vary substantially in tree composition, flood regimes, and topographic and bioclimatic boundary conditions (MAE 2013). According to Guariguata and Ostertag (2001), reforestation occurs as quickly as in five years after a disturbance in the evergreen forest ecosystems of the Ecuadorian Amazon, with variations depending on past land-use practices and abiotic site conditions. This was verified during fieldwork in May 2017 at three reforestation sites in the UNW, where 46 hemispherical photographs were acquired. We followed Pueschel et al. (2012) to binarize these photographs and derive the canopy closure index. We found that canopy closure after five years can be greater than of a 20-year-old forest (Figure 20b). We accordingly used a time threshold of five years as a reference for mapping reforestation (see section 3.3). Four National Protected Areas, mainly created during the 1990s, are present in the UNW and cover 25% of the total area. 78,300 ha of native forest were reported to be converted to pastures and croplands between 2000 and 2014, resulting in an average annual deforestation rate of ~6,300 ha or 0.5% of the forested land (MAE 2017). While these numbers vary, deforestation rates do not exceed 1.2–1.6% yr^{-1} (Sierra 2000).



(a)

(b)

Figure 20. (a) The UNW study area, localization in the context of the Amazon Basin and Tropical Andes. Data from The National Information System of Ecuador (2017) and digital elevation model of the UNW. **(b)** Boxplot of the canopy closure index and estimated forest age; and hemispherical photographs with derived canopy closure index for different forest development stages.

3.3 Materials and Methods

We organized the processing in five main steps (Figure 21) when implementing our multi-date classification. All procedures were developed in the R language (R Development Core Team 2017) and different strategies were applied to improve the processing (e.g., parallelization, vectorization, c-code libraries) (Revolution Analytics & Weston 2015; Hijmans 2016; Bengtsson 2016; Clayden 2016a; GDAL Development Team 2017) in complex computations.

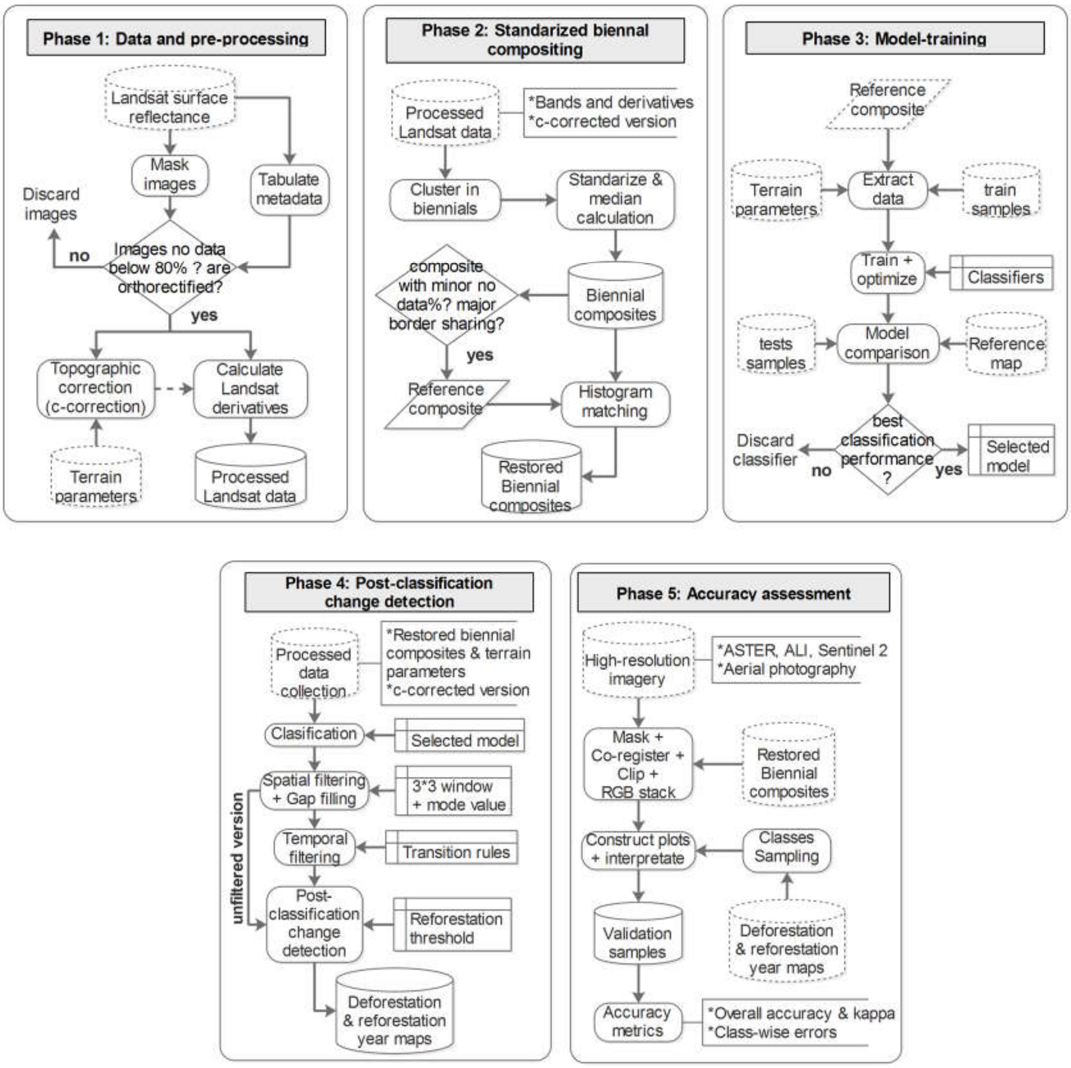
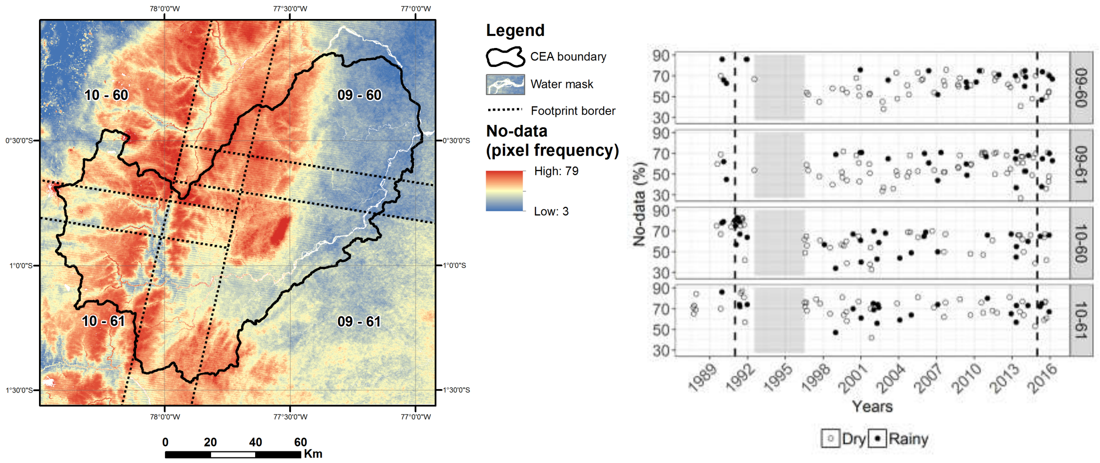

Phase 1: Data and pre-processing
Landsat surface reflectance
Mask images
Tabulate metadata
Discard images
no
Images no data below 80% ? are orthorectified?
yes
Topographic correction (c-correction)
Calculate Landsat derivatives
Terrain parameters
Processed Landsat data
Phase 2: Standarized biennal compositing
Processed Landsat data
*Bands and derivatives
*c-corrected version
Cluster in biennials
Standarize & median calculation
composite with minor no data%? major border sharing?
Biennial composites
yes
Reference composite
Histogram matching
Restored Biennial composites
Phase 3: Model-training
Reference composite
Terrain parameters
Extract data
train samples
Train + optimize
Classifiers
tests samples
Model comparison
Reference map
Discard classifier
no
best classification performance ?
yes
Selected model
Phase 4: Post-classification change detection
Processed data collection
*Restored biennial composites & terrain parameters
*c-corrected version
Clasification
Selected model
Spatial filtering + Gap filling
3*3 window + mode value
unfiltered version
Temporal filtering
Transition rules
Post-classification change detection
Reforestation threshold
Deforestation & reforestation year maps
Phase 5: Accuracy assessment
High-resolution imagery
*ASTER, ALI, Sentinel 2
*Aerial photography
Mask + Co-register + Clip + RGB stack
Restored Biennial composites
Construct plots + interpretate
Classes Sampling
Validation samples
Deforestation & reforestation year maps
Accuracy metrics
*Overall accuracy & kappa
*Class-wise errors

Figure 21. Methods and workflow for implementing the multi-date classification.

3.3.1 Data and Pre-Processing

For this study, we downloaded 1,350 images for four Landsat footprints (09-60, 09-61, 10-60, and 10-61) and for the period 1989–2016. They were processed to surface reflectance and acquired from the United States Geological Survey (USGS) Global Archive, sourced through the Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) (USGS 2014). This dataset included Landsat Thematic Mapper (TM), Enhanced Thematic Mapper plus (ETM+), and the Operational Imager (OLI) data. This ready-to-use dataset is radiometrically calibrated by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2012), and orthorectified using a digital elevation model (DEM) and ground control points (NASA

2011). The percentage of masked areas (or no-data pixels) for each image was calculated using C code based on the Function of Mask (Fmask) algorithm (CFmask) (Zhu & Woodcock 2012) and LEDAPS by merging cloud, cloud shadow, glacier, and water areas into a unique class (Figure 22a). Images with percentages above 90% of no-data pixels and images without orthorectification according to their metadata were omitted, reducing the time-series of 27 years to effectively 23 years since 1991 (section 3.3.4). In total, 288 images were used (i.e., 68, 79, 74, and 67 images per Landsat footprint, respectively) to complete the multitemporal composites. Data was mostly available during the dry season (67% of images, Figure 22b) while the fewer images acquired during the rainy season avoid additional data gaps in single years for our forest dynamics analyzes (Lunetta et al. 2004; Kimes et al. 1998). This is in contrast to other studies, which selected images from specific periods within a year (Müller et al. 2016). However, we preferred to maintain all images as a strategy to reduce data loss in this data-sparse environment. The average time interval between consecutive images for each footprint was 141, 122, 130 and 154 days, respectively. Nevertheless, in all footprints, a data gap from August 1992 to July 1996 in the Landsat archive introduced an interruption of 3.9 years in our time-series. Finally, a set of vegetation indices, band ratios, and Tasseled Cap transformation derivatives were calculated from the Landsat images (Table 7) following recommendations from similar forest dynamics studies (Kennedy et al. 2010a; Müller et al. 2016; Potapov et al. 2012). To overcome the topographic effects, a c-correction algorithm (Riaño et al. 2003) was applied to Landsat bands and its derivatives to evaluate if it contributed to improving classification results (Section 3.4.2).



(a) (b)

Figure 22. (a) Spatial distribution of no-data pixel frequencies and **(b)** no-data percentage of selected images. Gray areas in the plots refer to the 1992–1996 data gap and dashed lines to effective 23-year time-series.

Table 7. Landsat bands and derivatives.

Name	Abbreviation	Wavelength region	Reference
Normalized Difference Vegetation Index	NDVI	VIS, NIR	(Rouse, Haas, Scheel, & Deering, 1974)
Aerosol Free Vegetation Index 1.6 µm band	AFRI16	NIR	(Karnieli, Kaufman, Remer, & Wald, 2001)
Normalized burn ratio	NBR	NIR	(Key & Benson 2006)
Landsat bands 1-7	Bands 1-7	VIS, NIR, SWIR	-
Band ratios: TM4/TM3, TM5/TM4, TM5/TM7	R43, R54, R57	VIS, NIR, SWIR	(Krishna Bahadur 2009)
Tasseled cap: brightness, greenness, and wetness	TCB, TCG, TCW	VIS,NIR, SWIR	(Crist & Cicone 1984)

We collected 872 image chips from different sources acquired between April 2000 and August 2016 from high- and very high-resolution data for validating our medium-resolution remote sensing outputs (Olofsson et al. 2014): aerial photography (1 m spatial resolution), pan-sharpened images from the Advanced Land Imager (ALI, 10 m), Sentinel-2a (10 m), and Advanced Spaceborne Thermal Emission and Reflection

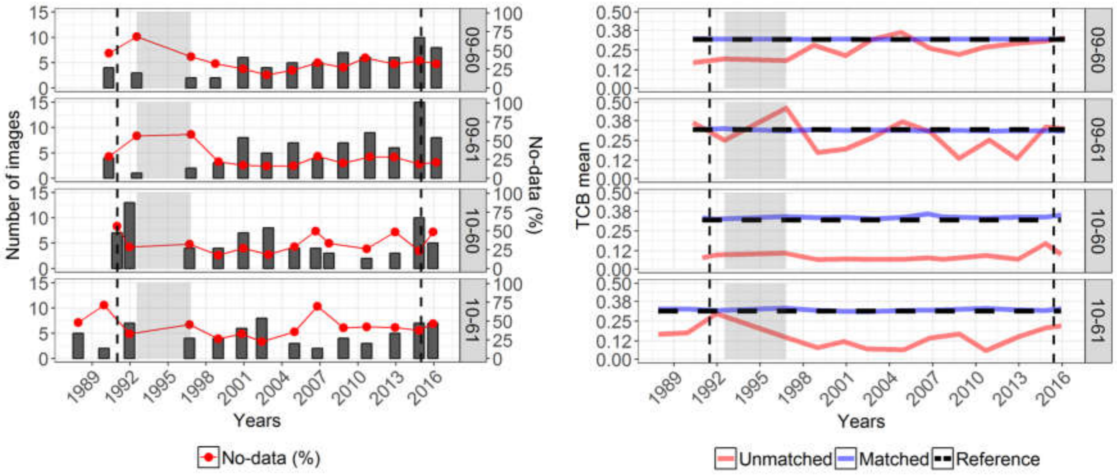
Radiometer (ASTER) imagery (15 m). The latter is available for free (National Institute of Advanced Industrial Science and Technology & Geological Survey Japan 2017) and we downloaded the whole archive to densify our high-resolution validation dataset. All images were pre-processed including manual co-registration (using Landsat imagery as a reference) and cloud masking including shadows (applying simple thresholds and manual screening in some cases). Finally, multispectral imagery was stacked, i.e., ASTER, ALI and Sentinel, for displaying them as false-color composites during the construction of validation sample plots (section 3.3.5).

To guide our implementation (section 3.3.4), the information used for the establishment of Ecuador´s Forest Reference Emission Level (MAE 2017) were collected. This dataset constitutes a series of land-cover and vegetation maps based on Landsat, ASTER, and Rapid Eye imagery that were visually interpreted with accuracies around 70% for different periods between 1990 and 2014.

For the elevation source, we used the three arc-second (90 m) digital elevation model from the Shuttle Radar Topography Mission (SRTM) (CGIAR - CSI 2008). This dataset corrected for data gaps and its quality is, especially for mountainous regions of Ecuador, higher than the one arc-second (30 m) resolution product. We derived elevation, slope, aspect, roughness, the topographic position index (TPI), and the terrain ruggedness index (TRI) (Wilson et al. 2007) to evaluate their contribution to classification performance.

3.3.2 Standardized Biennial Compositing

For compositing, we discarded Landsat bands 1 and 2 as they are known to be more sensitive to atmospheric effects (Zhang et al. 1999), while Landsat bands 3, 4, 5 and 7 and the calculated derivatives described in Table 7 were grouped in biennials according to the acquisition date of the image used in their calculation. The biannual time step was chosen as it resulted in 5 ± 2 images being available for most composite cases. This arrangement resulted in a no-data percentage average of 34 ± 13% (Figure 23).



(a)

(b)

Figure 23. (a) Number of images used for compositing and no-data percentage in resulting composites and **(b)** Tasseled Cap Wetness (TCW) mean value of composites with (Matched) and without histogram matching (Raw). Gray areas in the plots refer to the 1992–1996 data gap and dashed lines to effective 23-year time-series.

All biannual input pixels were z-transformed for the compositing:

$$\mathrm{Z}_{ij} = \rho_{ij} - \mu_j/\sigma_j \tag{1}$$

with ρ_{ij} being the pixel vector at position *i* and for date *j*, and μ and σ being the pixel's mean and standard deviation at date j. We then calculated the median $\bar{\mathrm{Z}}_i$ value for each vector Z_i as this metric is known to be less affected by atmospheric contamination or phenological variation in image compositing (Potapov et al. 2012). Other metrics (e.g. quantiles, maximum, minimum, variance) that are regularly applied in similar studies (De Fries et al. 1998) were also tested. However, the scarce data situation required a conservative approach using the median. Finally, we normalized $\bar{\mathrm{Z}}_i$ and stored it as the output value β for a given biennial composite according to:

$$\beta = \frac{\bar{\mathrm{Z}}_i - min(\bar{\mathrm{Z}}_{i:n})}{max\ (\bar{\mathrm{Z}}_{i:n}) - min(\bar{\mathrm{Z}}_{i:n})} \tag{2}$$

As residual radiometric offsets occurred in the overlap areas between footprints, we required a pixel-level radiometric alignment (Pflugmacher et al. 2012). We selected the 2002 composite from path-row 09-60 and 09-61 as reference composites, as they had few no-data values and a low atmospheric aerosol load. Values in the overlapping footprints (10-60 and 10-61) were aligned based on histogram matching across the overlap areas to neighboring footprints. The same procedure was then applied across the time-series within each footprint. In total, 52 biennial composites were aligned with references footprints, reducing for example the variance of the Tasseled Cap Brightness (TCB) from 0.1 to 0.01 after histogram matching (Figure 23b).

3.3.3 Model Training

We defined four classes to map permanent forest cover and deforested/reforested areas. Permanent forests included on the one hand, evergreen forests, (encompassing montane, foothill, lowland, and flooded forests) and on the other hand, *Guadua spp.* forests with their spectrally distinct patterns due their different species composition, canopy height, and overall lower biomass (Silman et al. 2003). Other non-forest vegetation above 3300 m.a.s.l. such as grasslands or shrubs, were not considered in this research. Conversely, change classes included human land use for agricultural production, i.e. pastures and croplands (including early revegetation, commercial, and subsistence plantations); and non-vegetated areas, i.e. bare soils and urban areas. We gained initial knowledge of the approximate class distribution by running a first and non-validated image classification with a few training samples, which built the basis for distributing the training samples for training in a guided fashion (Table 8):

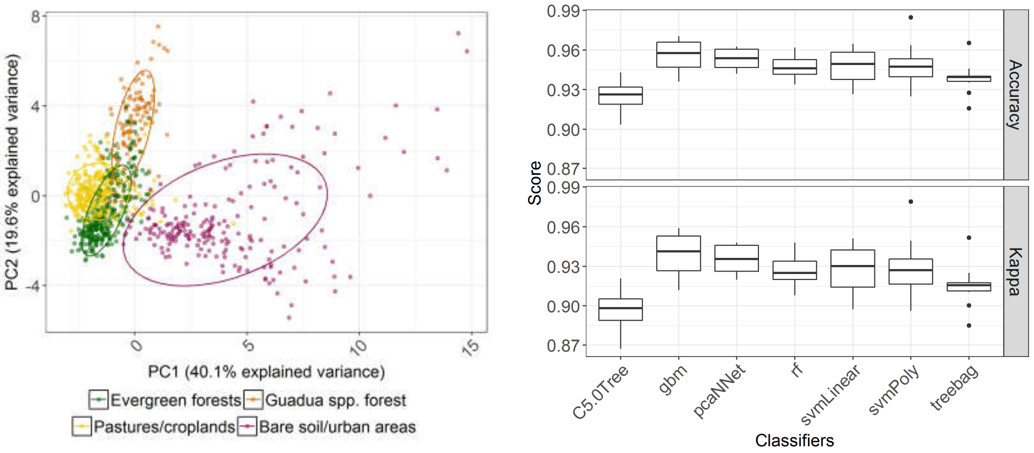
Table 8. Classes considered and training sample size.

Type	Classes	Approximate Area [km^2]	Samples [no.] [1]	Example photograph
Permanent classes	Evergreen forest	8053	307	
	Guadua spp. forest	181	122	
Change classes	Pastures / croplands	2047	279	

	Bare soil / urban areas	653	221	

[1] A class-wise standard deviation of 0.9 and a confidence interval of 0.9 were considered to obtain a reasonable class-wise sample size.

We then interpreted the almost 1000 training samples on-screen based on a mosaicked, biennial color composite from 2002 and on a recent vegetation map (MAE 2017). Since the four classes represented a complex spectral feature space and their visual separation was challenging (Figure 24a), we tested different classifiers using the Caret package (Kuhn 2016). This software applies a parameter tuning of classifiers, and bootstraps training samples to determine their effect on performance and decide which model perform best. As most classifiers provided similar out-of-bag errors (Figure 24b), we decided to use a least squares support vector machine with polynomial kernel (svmPoly) (Karatzoglou et al. 2004), as it achieved the highest overall correlation (0.697) with the land-cover reference maps (MAE 2017). Other classifiers including Random Forest (rf, 0.684), Stochastic Gradient Boosting (gbm, 0.682) or Neural Networks (pcaNNet, 0.680) showed good correlations, but they were not higher than svmPoly.



(a) (b)

Figure 24. (a) The feature space of the training samples using the first and second principal components (PC1 and PC2) obtained from Landsat bands, derivatives, and terrain parameters. **(b)** Boxplots of classifiers out-of-bag error using training samples.

3.3.4 Post-classification change detection

By classifying biennial composites, 13 land-cover maps were obtained for each footprint, covering the period 1989–2016. A 3 x 3 median filter was applied to eliminate spurious pixels within a land-cover map, but for data gaps, values were input calculating the per-pixel time-series mode from all land-cover maps. While random noise and discontinuities were eliminated, artefacts from clouds and cloud shadow remnants, sensor noise, or simply misclassified pixels were still present in the data. We therefore further applied a temporal filter with transition rules (Clark et al. 2010) to identify illogical land-cover and land-cover change patterns, and reclassified errors according to a set of rules based on contextual knowledge. For example, it is impossible for bare soil to become an evergreen forest in one year and return to the bare soil class the next year again. Instead, this may represent either cropped land (bare soil or cropped land in one year) in the case of agricultural land, or it simply may be a misclassification. In any case, it will not represent land change associated with forest cover. We accordingly implemented a temporal filter based on a moving window of three consecutive observations and a set of allowed transition rules (Table 9).

Since the first and the last land-cover maps could not be temporally filtered according to this scheme, we omitted those years (i.e. 1989–1991 and 2015–2016 periods), thereby reducing our time-series to the period 1992–2014, i.e. 11 observations for each footprint. We then derived deforestation and reforestation dates from the series of land-cover maps. We flagged the first year of a forest pixel being mapped as one of the non-forest classes as the deforestation year. Reforestation, though, is a continuous process that can only be mapped from satellite data once a certain threshold

Table 9. Transition rules.

Observation year / Class		**year $\eta + 1$**			
		Evergreen forest	Guadua spp. forest	Pastures / croplands	Bare soil / urban areas
years η and $\eta + 2$	Evergreen forest	Yes	No	No	No
	Guadua spp. forest	No	Yes	No	No
	Pastures/croplands	No	No	Yes	Yes
	Bare soil/urban areas	No	No	Yes	Yes

of “forestedness” has transgressed, i.e. a previously non-forested pixel spectrally resembles a forest class for a minimum period of time (defined in this paper in five years as described in section 3.2).

Finally, the outputs of the post-classification change detection were mosaicked and the UNW area extracted, considering the treeline (3300 m.a.s.l.) to exclude non-forested areas. As our aim was to evaluate different algorithms for multi-date forest change classification, we iterated all these steps for each pre-processing approach (i.e. surface reflectance or topographic correction) and omitted/applied temporal filtering to evaluate their individual contribution to the overall accuracies of these maps.

3.3.5 Accuracy Assessment

We calculated confusion error matrices for deforestation and reforestation maps (Olofsson et al. 2014; Thomas et al. 2011; Cohen et al. 2017). We employed a stratified random sampling based on the deforestation and reforestation classes, arranging samples of 5-3 pixels in a cross shape (Figure 25a). We sampled a minimum of 50 samples per class, and 100 for the larger classes of stable forest and stable non-forest. Since reforestation stable classes could be affected during post-classification change detection due to its assigned regrowth-time threshold, we sampled these classes independently to ensure their performance. Finally, as spatial autocorrelation can bias the accuracy assessment (Congalton 1991), a minimum threshold distance between samples of the same class was applied (Table 10).

Table 10. Classes, minimum distances, and sample sizes.

Map	Class name	Distance threshold [m]	Sample size [pixel]
Deforestation map	Stable-forest	2500	100
	Stable-non forest	2500	100
	Deforestation year	2500	50/year
Reforestation map	Stable-forest	2500	100
	Stable-non forest	2000	100
	Reforestation year	1500–2000	50/year

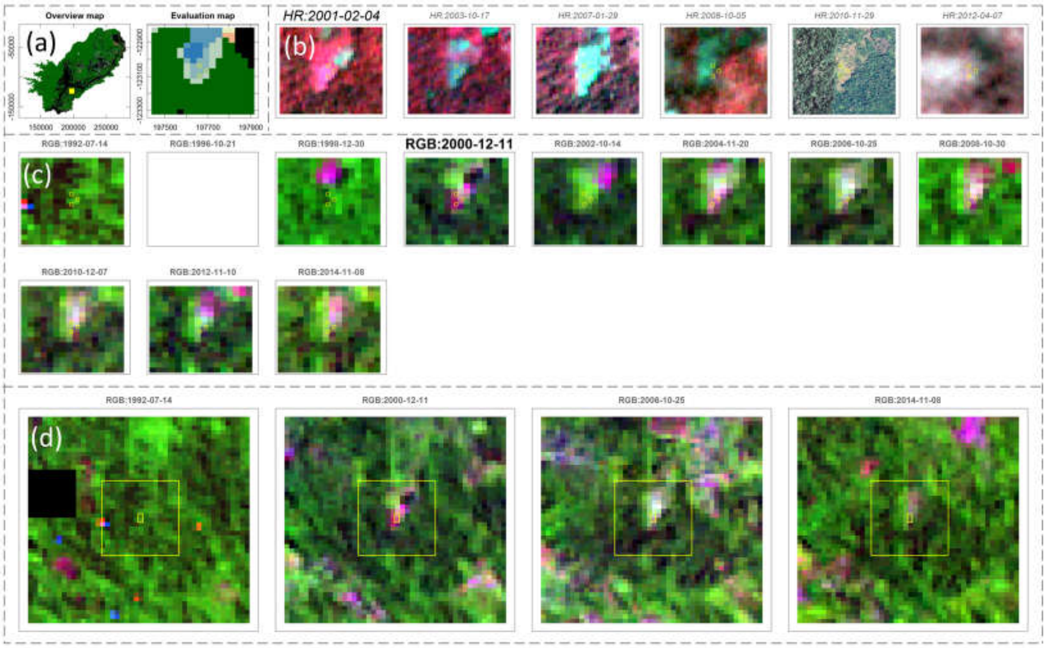


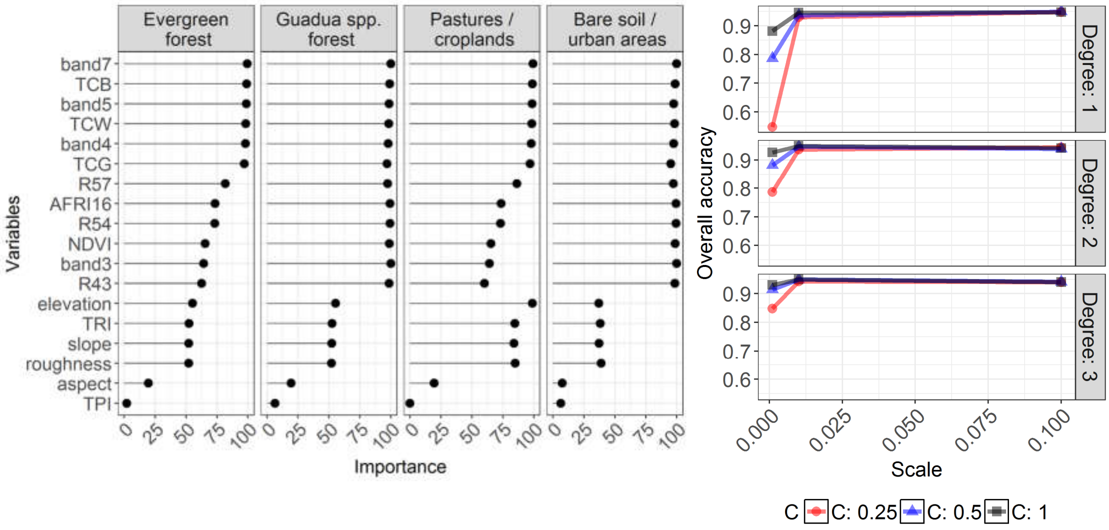
Figure 25. Example of a forest-loss sample (3 pixels). **(a)** Location of the sample in the map. **(b)** High-resolution image chips from ASTER (2001–2008 and 2012), and aerial color photography (2010). **(c)** Landsat color composites image chips (1992–2014, with a data gap in 1996), and **(d)** overview showing sample area and analysis period.

Each sample was interpreted on-screen based on high-resolution imagery (Figure 25b) and Landsat color composites (Figure 25c and Figure 25d). We calculated the overall accuracy, the kappa index, and class-wise commission and omission errors.

3.4 Results

3.4.1 Variable Importance and svmPoly Optimization Report

We decomposed the svmPoly model to observe which variables contributed to the respective results. Figure 26a reveals that Landsat bands 4–7 and Tasseled Cap brightness, wetness, and greenness yielded an importance above 90% in all classes. Conversely, vegetation indices and band ratios as well as terrain parameters were overall less significant, but contributed to separating pastures/croplands from the *Guadua spp.* forests. These overlapped spectrally, but separability improved when integrating terrain derivatives in the classification process. Regarding optimization, the Caret software explored three parameters of the svmPoly classifier (degree, scale, and cost) to maximize its classification accuracy (Figure 26b). Its final calibration yielded 274 support vectors with a cost of constraints (C) of one and the hyperparameter values were set to a degree of three, scale of 0.001, and a default offset of one.



(a) (b)

Figure 26. (a) Variable importance by class during classification with svmPoly. **(b)** Optimization of the svmPoly classifier during the model-training phase.

3.4.2 Accuracy Metrics Results

Deforestation and reforestation maps based on different filtering techniques varied substantially (Table 11). Overall accuracies were significantly better when applying surface reflectance and temporal filtering to land cover classifications, achieving 82 ± 3% and 71 ± 3% (calculated with a 95% confidence interval) for deforestation and reforestation maps, respectively. Topographic correction led to poorer overall accuracies by 13 ± 2% in both maps when compared to the best result. Temporal filtering improved accuracies substantially by 21 ± 1% for both deforestation and reforestation maps. Commission and omission errors for stable and change classes are shown for the

Table 11. Overall accuracy of deforestation and reforestation maps by processing approaches.

Map	Accuracy metrics	Processing approaches		
		Surface reflectance–not filtered [%]	Topographic correction–filtered [%]	Surface reflectance–filtered [%]
Deforestation	Overall	62	70	82
	Kappa	58	65	80
Reforestation	Overall	48	56	71
	Kappa	39	47	67

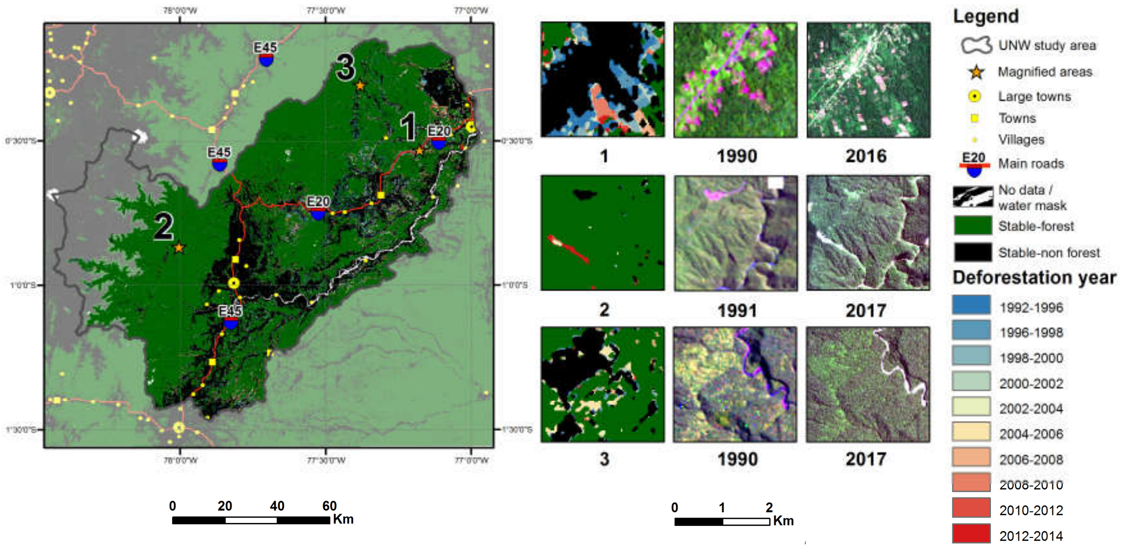
Table 12. Commission and omission errors for the surface reflectance – filtered approach.

Class name		Deforestation		Reforestation	
		Commission [%]	**Omission [%]**	**Commission [%]**	**Omission [%]**
Stable-forest		1	8	5	7
Stable-non forest		11	1	7	5
Change year	1992-1996	14	4	44	7
	1996-1998	20	13	44	3
	1998-2000	20	17	34	20
	2000-2002	16	2	42	19
	2002-2004	30	5	28	22
	2004-2006	26	3	28	10
	2006-2008	28	5	44	12
	2008-2010	20	0	36	9
	2010-2012	24	5	46	10
	2012-2014	24	0	-	-
Overall mean		19.5	5.25	32.54	11.27

filtered product (Table 12). The overall commission and omission errors were lower for the deforestation map (mean of 19% and 5%, respectively) than for the reforestation map (mean 32% and 11%, respectively). Moreover, stable classes were less prone to commission and omission errors (mean 1 – 11%) compared to change classes (mean 0 – 46%).

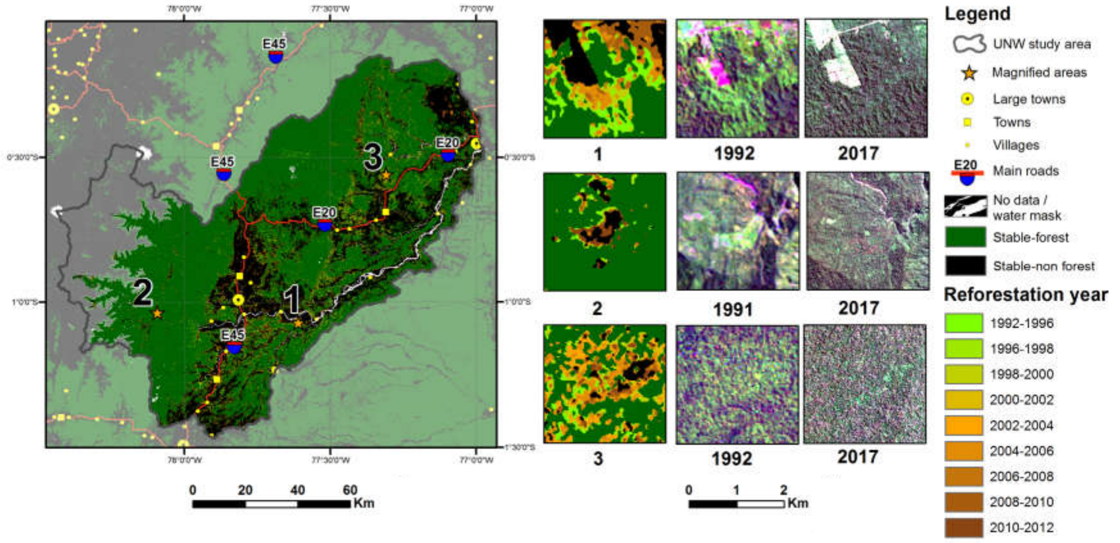
3.4.3 Deforestation and reforestation maps

Maps of deforestation and reforestation years are shown in Figure 27a and Figure 28a. In general, the patterns follow the description of Wasserstrom and Southgate (2013) for the Ecuadorian Amazon during its oil-related colonization (1964-1994). For instance, deforested areas along the E45 highway (built in 1975) and the banks of the Napo River relate to settlements that already existed before the period we analyzed. The age of the deforestation patches along the E20 highway (built in 1983) decreased with increasing distance from the highway (Figure 27b-1). In the mountainous areas, the detection of landslide scars (Figure 27b-2) was accurate, and topographic shadows did not apparently inhibit the change detection. However, false-positive errors were observed in areas with



(a) (b)

Figure 27. (a) Deforestation year map for the UNW. **(b)** Magnified areas show: **(1)** Linear deforestation along the E45 highway; **(2)** landslide scars; and **(3)** false-positive errors in the mixed forest and pasture areas.



(a) (b)

Figure 28. (a) Reforestation year map for the UNW. **(b)** Magnified areas show: **(1)** Jatun-Sacha Biological Reserve, which is known for reforestation since the 1990s; **(2)** forest succession after landslides; and **(3)** false-positive errors in a stable-forest area.

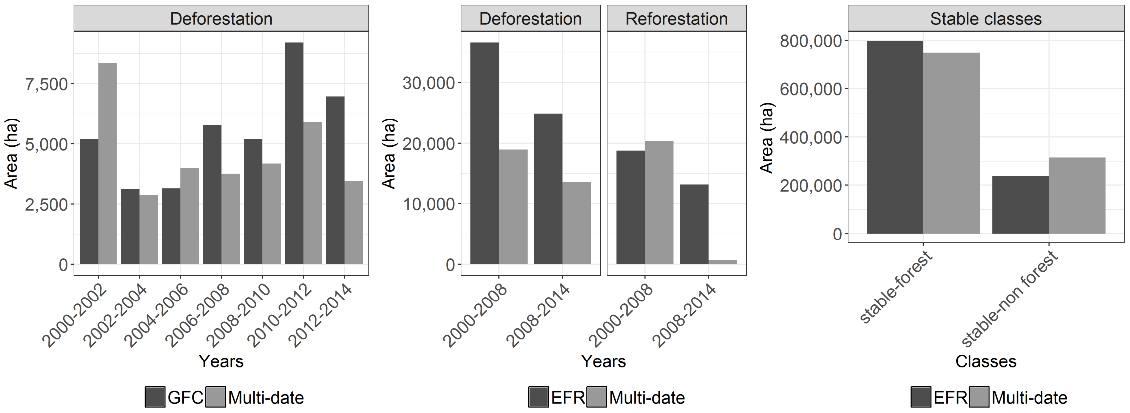
many mixed pixels where mostly evergreen forests and pasture/cropland occurred (Figure 27b-3), but also in areas with no-data values due to the occurrence of sparse observations, or an inaccurate water mask.

Areas of reforestation seemed to be more prominent along the E45 highway, where deforestation was less intense. Known areas of reforestation since the 1990s were well represented (Figure 28b-1), as was forest succession after landslides in mountainous areas (Figure 28b-2). Overall, the reforestation year map was more affected by mixed pixel problems and mask errors than the map of deforestation year (Figure 28b-3).

Following Rudel et al. (2002), we calculated overall deforestation and reforestation by applying a buffer distance of 3 km along the two main highways E45 and E20 in the UNW to corroborate our observations. Accumulated deforestation along highway E45 summed up to an area of 3320 ha, and 11,403 ha along highway E20. In contrast, reforestation along highway E45 accumulated to 7458 ha, and 5415 ha along highway E20.

3.4.4 Comparison with other sources

We compared our implementation with two different sources: Forest Loss Year (FLY) according to Hansen et al. (2013) and 2) Ecuador´s Forest Reference (EFR) Emission Level information (MAE 2017). Both sources were cropped with the UNW and re-labeled to match our classes (Figure 29). Results are similar across the three classifications. However, differences specifically exist with FLY for specific time periods such as 2000 – 2002 and 2010 – 2012 (Figure 29a), or EFR reforestation between 2008 – 2014 (Figure 29b). On average, deforestation was 2,757 ha year^{-1} for the period from 2000 to 2014,



(a) (b) (c)

Figure 29. (a) GFC2015 and Multi-date deforestation areas by biennials for the period 2000 – 2014. **(b)** MLUX and Multi-date deforestation and reforestation areas for the periods 2000 – 2008 and 2008 – 2014, together with their **(c)** stable classes areas.

in FLY data and 4,394 ha year^{-1} in EFR. Our estimates are comparably conservative with 2,319 ha year^{-1}. According to FAO (Puyravaud 2003), these values represented annual deforestation rates of -0.35%, -0.57% and -0.31% respectively.

Furthermore, reforestation summed up to 574 ha year^{-1} for the 2000-2014 period in FLY, indicated 2277 ha year^{-1} in EFR and 1504 ha year^{-1} in our analysis, representing annual reforestation rates of 0.07%, 0.28% and 0.19%, respectively in EFR, 796,982 ha stayed unchanged, while our analysis yielded 748,688 ha of stable forests (Figure 29c).

3.5 Discussion

3.5.1 Biennial Image Compositing and Pre-Processing Effects on Results

Our image compositing technique was based on the standardization and median calculation, which is an effective strategy to maximize information extraction when the number of observations is limited. We chose a biennial classification scheme (Griffiths et al., in review). We thereby improved the signal-to-noise-ratio, as one-year composites may be inferior in data-scarce conditions (Potapov et al. 2011). Composites from longer time periods, though, may not be adequate to monitor subtle processes such as reforestation (Bustamante et al. 2016).

The histogram matching algorithm that we used for radiometrically aligning the composites enabled a regional-scale classification, and at the same time created consistency across the time-series. This was supported by the high consistency between the land-cover classifications from different years and the results after applying our post-classification change detection algorithm. However, cloud-free composites as references and sufficient spatial overlap between the target and reference footprints are mandatory for the proper functioning of the histogram matching (Benjamin & Leutner 2017).

We also accommodated for correcting the radiometric distortions due to topography. In our case, the c-correction algorithm principally improved the homogeneity of the imagery across sunlit and shaded slopes. However, commission errors increased after applying the topography correction. This is in line with the findings of Chance et al. (2016), who reported negative effects of a topographic correction on change detection analysis. Others found that the application of topography correction generally had a smaller influence on the overall accuracy of a classification when compared to the selection of a classifier (Vanonckelen et al. 2015). Future work should further improve the results from topographic correction by employing the best digital elevation models available (Pimple et al. 2017; Chance et al. 2016).

3.5.2 Post-Classification Change Detection Performance

Our post-classification change detection strategy was based on land-cover maps (MAE 2017) as reference to validate model training and classifier outputs. This allowed the selection of the most precise classifier based on the correlation between the classification and the land cover maps. While the limited size of our sampling set may not be representative for some classifiers (Zhu et al. 2016), support vector machines (SVMs) apparently performed well, as SVMs support small training samples (Wieland et al. 2016).

The original Landsat bands and derived Tasseled Cap components had a considerable predictive power (Figure 26). This was specifically true for bands 3–7, which are known to be important predictor variables in forest/non-forest classifications in the tropics (Potapov et al. 2012), but also in dry regions of the world (Mellor et al. 2013). Spectral mixtures and spectral similarity of land cover types (see Figure 24a) limit the separability at 30-m Landsat spatial resolution. In this regard, elevation and terrain derivatives from the digital elevation model (slope, aspect) contributed to class separation, despite their predictive power not being as high as that of the Landsat bands or Tasseled Cap components.

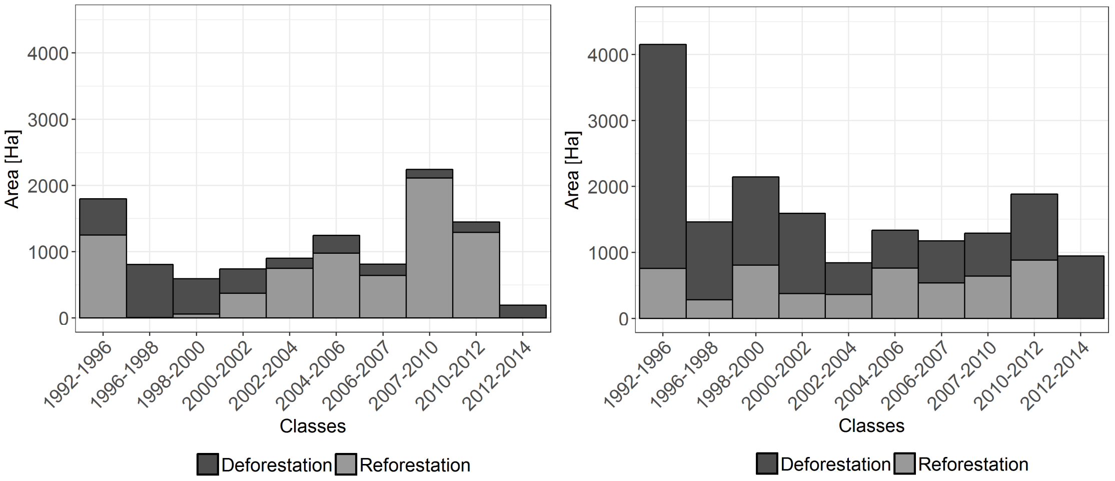
While some gaps related to cloud and cloud shadow remnants remained after the classification, we were able to demonstrate that temporal filtering is a powerful technique for removing these artefacts and considerably improving the results (comparison in Table 11). The set of transition rules allowed us to filter most of the illogical class transitions; however, some highly dynamic events were still missed due to the data-scarce setting and accordingly introduced omission errors.

3.5.3 Multi-Sensor Fusion Benefits and Landsat Archive Limitations

As long-term forest dynamics analyzes require historical satellite archives to observe land-surface changes retrospectively, multi-sensor fusion is an inevitable approach. In this respect, the Landsat archive and its surface reflectance products was demonstrated to be a valuable (and actually the only medium to high spatial resolution) source covering our observation period of interest. The Landsat sensor family effectively integrates multiple sensors and thereby provides the best possible data coverage over time, even in regions of high cloud cover. Two limitations, however, affected our results. First, the global data gap between 1991 and 1996 limited the quality of our findings for the mid-1990s, despite the time period not exceeding the forest-regrowth time threshold of five years that we defined for this study. It may specifically have affected the detection of relatively fast deforestation events (e.g. infrastructure construction, blown-downs, landslides). The second drawback was the lower geometric quality of single images in the Landsat archive, which were not always possible to identify from the metadata on image quality alone. These errors are related to the limitation of ground control-based geometric adjustments in cloud-prone areas. It will be easier to avoid ingesting critical imagery in the processing chain in the future as the new Landsat collection handles data quality more rigorously (Micijevic et al. 2017; Roy et al. 2014).

3.5.4 Long-Term Forest Dynamics in the UNW and Feasibility of Multi-Date Classification in the Andean Amazon

Despite the remaining limitations of our multi-date classification implementation, the spatially explicit forest dynamics patterns in the UNW allow novel insights beyond what was already known from previous satellite data analyzes relying on only two or just a few points in time (Sierra 2000), or only on spectral information (Stephen J. Walsh et al. 2008) ideally in the Tropical Andes. Dynamics along the E45 highway after 1992 mostly related to reforestation on peripheral lands, while deforestation rates were comparably low in that region (Figure 30a). This may be explained by the population census, where the population in the urban centers of Tena and Archidona increased by 233% between 1990–2010 (INEC 2010), suggesting a rearrangement in the population distribution between the rural and the urban areas. This assumption is supported by similar findings by Rudel et al. (2002) in the southern Ecuadorian Amazon. In contrast, deforestation was



(a) (b)

Figure 30. Deforestation / reforestation area for the period 1992-2014 **(a)** along the highway E45 and **(b)** along the highway E20.

principally identified along the E20 highway. Since this highway was constructed more recently, new settlements and commercial activities linked to oil extraction have triggered deforestation (Wasserstrom & Southgate 2013) (Figure 30b).

A comparison with other sources revealed further details. For instance, deforestation was seen to be more similar to FLY than to EFR, most likely because the FLY dataset is also based on a multi-date classification, while the EFR is based on an object-based classification that generalized deforestation patches. In the case of reforestation, all results differed markedly. Different conceptualizations of reforestation (Hansen et al. 2013; MAE 2017) and confusion with secondary forests are likely to be the main reasons. According to Cohen et al. (2017), it is not surprising that forest disturbance maps differ due to semantic and methods differences. Different accuracies and results accordingly relate to multiple factors such as differences in the change detection algorithm, in the quality of specific satellite imagery used, metrics and training information, the time-series density, or the thresholds applied to identify change.

Overall, our multi-date classification implementation was demonstrated to be far less sensitive to data scarcity and atmospheric contamination than other approaches using automated time-series analysis algorithms (Santos et al. 2017).

3.6 Conclusions and outlook

Forest dynamics in the complex and vulnerable regions of the Tropical Andes are still under-researched considering remote sensing data analyzes (Oliveras et al. 2014; Da Ponte et al. 2015). To the best of our knowledge, this was the first study of its kind that specifically focused on the challenges related to scarce data and the poor signal-to-noise-ratio in a long time-series for automated forest change analyzes in the Tropical Andes. We demonstrated that an adapted implementation of multi-date classification based on image compositing, multi-sensor fusion, and post-classification change detection could mitigate most of these limitations. Our findings add to the expanding body of literature on such approaches with a focus on data-scarce situations and highlight the importance of the Landsat archive for monitoring decadal land-cover change even in cloudy regions of the world.

Future research should focus on diversifying data sources and predictors, as our findings provide further evidence that classification results, specifically when using machine learners, will improve in data-rich environments. Moreover, the increasing web-based availability of high- and very high-resolution data will in the future allow further improving sample quantity and quality, while semi-automatic approaches (Huang et al. 2015) and temporal-spectral profiles sampling (Senf et al. 2015) are also promising alternatives. Furthermore, since our methodology requires a reforestation time threshold, it would be beneficial considering specific thresholds for different forest communities. This should ideally be based on forest growth models such as FORMIND (Paulick et al. 2017) that support specifying distributions for reforestation time thresholds. Additionally, improvements may well be possible with further refined transitioning rules in post-classification filtering or automated solutions with more complex transition rules with an increasing number of land cover classes (Abercrombie & Friedl 2016; Ahlqvist 2008).

As other areas may experience similar or even more severe data scarcity than the UNW, image compositing might be limited to lower observation frequencies. In this regard, regions such as north-central Africa and northern Russia, which have the sparsest Landsat coverage compared to Ecuador (Wulder et al. 2016) may constrain multi-date classification usability to frequencies greater than biennials. Such limitations may for example constrain its usability in the frame of Reducing Emissions from Deforestation and Forest Degradation (REDD+), which requires biennial updates reports for forest reference level information specially in developing countries (UN-REDD Programme 2015). Finally, implementing such a multi-date classification for larger study areas requires cloud-based or high-performance computing (HPC) environments, as the processing is demanding and it is more effective to "bring the algorithm to the data" than to download massive datasets. Currently, some alternatives are available (e.g. EODC, 2018; Gorelick et al., 2017; Open Foris, 2015), which allow implementing similar methodologies for large areas. Cloud-based or HPC environments also provide novel opportunities to develop monitoring systems based on sensor constellations, such as Landsat and Sentinel-2 (Wulder et al. 2015).

As optical remote sensing of the core tropics regularly suffers from high cloud cover, integrating newly available imagery will increase change map reliabilities. Linking the vast Landsat archive with the quickly expanding Sentinel-2 archive is therefore one of the cornerstones for future improvements (Drusch et al. 2012; Wulder et al. 2016). Such a strategy will also allow extending the applicability of our approach to larger regions such as the entire Tropical Andes, and to ecosystems with more diverse land cover.

Acknowledgments

This work was supported by a research grant from SENESCYT for F.S. and Spatial Statistical Analysis of Deforestation Drivers in the Upper Napo Watershed Project from Universidad Regional Amazónica Ikiam for P.M. The contents of this paper are the sole responsibility of the authors. Special thanks to the R community for sharing their knowledge and USGS-EROS for provide no-cost Landsat data; specific thanks to colleagues from ZFL and ZEF in Bonn, Germany, for their support and helpful comments. Further thanks for the support of the Gobierno Autónomo Descentralizado de Napo, Ecuador. Special thanks to Liliana Herdoiza and Ernesto Santos for their invaluable help during fieldwork. This paper contributes to research in support of the Landsat Science Team (https://landsat.usgs.gov/landsat-science-teams).

4 ANALYZING UNDERLYING CAUSES OF DEFORESTATION AND REFORESTATION IN THE CENTRAL ECUADORIAN AMAZON: A GEOGRAPHICALLY WEIGHTED RIDGE REGRESSION APPROACH

Fabián Santos and Valerie Graw

“Sadly, it´s much easier to create a desert than a forest”

James Lovelock

4.1 Introduction

The Tropical Andes is a mountainous region at the base of the Andes ridge. Due to its altitudinal gradient it is characterized by 23 ecoregions and 8 bioregions (Olson et al. 2001) providing important economic and ecological services to almost 40 million people (Armenteras et al. 2011). Recognized as an endangered biodiversity hotspot with a high conservation priority (Brooks et al. 2006; Myers et al. 2000), population growth and agriculture expansion (Cincotta et al. 2000; Armenteras et al. 2017) are the major driving forces of deforestation contributing to potential impacts of climate change (Buytaert et al. 2011). On the other hand, large-scale reforestation has been detected in some areas of Latin America (Grau & Aide 2008), especially along old colonization fronts (Rudel et al. 2002). However, these areas are less studied or understood, and their role in forest recovery and restoration of important environmental services is ignored (Nagendra 2007; Rudel et al. 2005). Therefore, the analysis of forest dynamics drivers (FDD) in the Tropical Andes is of prime importance for conservation, climate change adaptation and sustainability. This knowledge is decisive for countries like Ecuador, where most of the remaining native forests are located in the Tropical Andes and the deforestation rate has been the highest in South America for some years (Mosandl et al. 2008; FAO 2007).

Forest dynamics are shaped by complex societal and ecological interactions known as “causes”, “drivers” or “determinants”. Geist and Lambin (2002) proposed a conceptual framework to facilitate their understanding, classifying them into:

1) proximate causes (local level, direct agents);
2) underlying causes (different levels, socio-economic processes); and

3) other causes (determined by environmental factors and social trigger events).

Commonly adopted by countries participating in the Reducing Emissions from Deforestation and Forest Degradation (REDD+), recent research recognized that underlying causes are less frequently analyzed in Latin America (Armenteras et al. 2017; Salvini et al. 2014). Proximate causes are more easily identified through remote-sensing-based techniques (Da Ponte et al. 2015), while underlying causes could be more complex and rely on socio-economic data. This data is frequently not available or reliable at the scale needed (Grainger 2008). Moreover, impacts of globalization (Meyfroidt et al. 2013) and economic development (Mertens et al. 2000; Rudel et al. 2005) generate more complex scenarios that challenge their understanding.

In Ecuador, previous studies combined remote-sensing products and socio-economic data to identify the influencing FDD. For instance, Southgate et al. (1991) analyzed thematic cartography and census data in a regression analysis identifying agricultural rents, spontaneous settlements, and land tenure insecurity as deforestation drivers in eastern Ecuador. Following a similar approach but adding survey data, Rudel et al. (2002) discussed reforestation drivers observed among ethnic groups and their relationships between land-use practices, cultural background and distance to roads in southern Ecuador. More recently, Mena et al. (2006) used thematic cartography, census and survey data in a spatial regression model to conclude road accessibility and population density as the most important deforestation drivers in northern Ecuador. Similarly, Walsh et al. (2008) identified here that reforestation drivers were motivated by land security and distance to roads. From these studies, it can be observed that deforestation is commonly not studied together with reforestation. However, results suggest contrasting driving forces such as population increase vs. decrease, closeness to vs. remoteness from roads, better vs. worse land security, etc. Therefore, their integration and evaluation with regard to possible linkages defines a pending research gap which will be addressed in this study.

Integration, analysis and visualization of multi-source data is an important challenge in FDD analysis. Due to data aggregation, it is a required procedure to analyze areal-based data (such as census), where statistical bias can be introduced with significant impacts on the results. This is known as the modifiable areal unit problem and frequently has been a reason to criticize statistical hypothesis testing when spatial data is used (Holt et al. 1996). Moreover, boundary changes in areal units across time introduce additional inconsistencies that complicate analysis even more (Mennis 2003). Nevertheless, different approaches have been proposed to overcome these issues (Krivoruchko et al. 2011; Tobler 1979; Stevens et al. 2015; Semenov-Tian-Shansky 1923). However, dasymetric mapping is probably the most popular (Semenov-Tian-Shansky 1923; Petrov 2012). Furthermore, previous studies on FDD used linear regression models to generalize relationships between forest change and driving forces. These studies failed to capture the variability of these relationships over space (Deilami et al. 2016). Therefore, an approach for extending the traditional regression framework and considering spatial information is most desirable for FDD analysis. In this regard, approaches such as geographically weighted ridge regression (GWRR) (Fotheringham et al. 1998; Wheeler 2007) have demonstrated satisfying this objective (Pineda Jaimes et al. 2010; Clement et al. 2009; Mas et al. 2013).

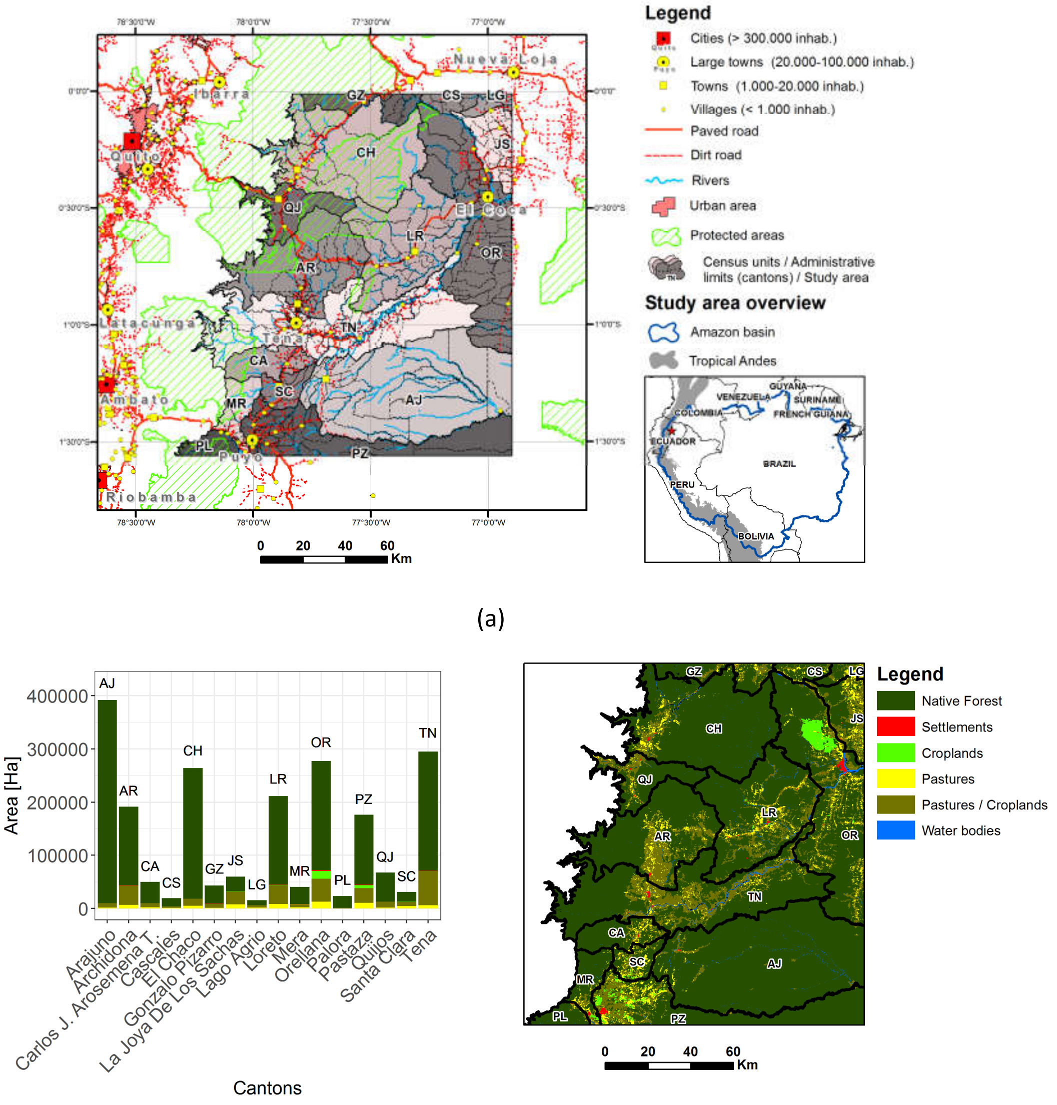
As the Tropical Andes constitutes a complex mosaic of landscapes, a workflow to analyze FDD is presented in this paper. Applying the abovementioned procedures, we conducted an analysis in the Central Ecuadorian Amazon (CEA), which is characterized by different colonization fronts with different socio-economic and biophysical settings. Our main objective is the exploration of a set of variable groups related to population growth and agricultural expansion to observe how they react with deforestation and reforestation rates during 2000-2010. Two research questions guided our analysis:

- Can dasymetric mapping and GWRR improve our understanding of the underlying causes of deforestation and reforestation?
- How do population growth/loss and agriculture expansion/reduction affect deforestation and reforestation in the CEA?

To answer these questions, we (i) explain how we calculated forest change rates, (ii) conducted dasymetric mapping for inter-census data processing, and (iii) further briefly describe the variable groups before we (iv) explain our implementation of GWRR. The discussion of the results will consider the benefits and limitations of the proposed approach and its contribution to the current knowledge of FDD in the CEA.

4.2 Study area

The CEA covers 21857 km^2 over an altitudinal gradient from 200 to 2800 m.a.s.l on the western slopes of the Andean Range (Figure 31a). It includes 16 cantons, i.e. second-level administrative units in Ecuador, which are used in this research for identifying specific zones in the CEA (Figure 31b). According to Olson et al. (2001), two ecoregions exist in the CEA, i.e. the Napo moist forests and Eastern Cordillera real montane forest. The latter has one of the highest conservation priorities in Ecuador as it covers less than 33% of its original area (Sierra et al. 2002). Moreover, the CEA is characterized by an extraordinary biodiversity, intense annual precipitation (1500-4500 mm), and a multitude of ecosystems (MAE 2013). Most of the soils are ferralitic with low fertility and high aluminium toxicity, although volcanic and alluvial soils can be an exception (Huttel et al. 1999; Eberhart 1998). Under these conditions, the agriculture limitations are well known; however, this does not prevent the native people from co-evolving with their natural environment (Coq-Huelva et al. 2017). Dramatic changes began in the 1970´s with the exploration and extraction of oil generating accelerated economic growth and industrialization (Pierre et al. 1988). Extensive road construction and the Agrarian and Colonization Reform of 1964 stimulated in-migration and rapid settlement over the whole Ecuadorian Amazon. According to Brown et al. (1994), the population grew by 432% by the end of 1990 resulting in an urban system that followed petroleum discovery and the related economic opportunities. This led to a disorganized and arbitrary colonization where land conflicts between the *colonos* (mestizos colonists) and native people were common. According to Perrault (2001), colonization of lands considered uninhabited by the reform but used ancestrally by the native people replaced traditional land-use practices with extensive agriculture and cattle ranching to secure land tenure.



(a)

(b)

(c)

Figure 31. (a) Study area in the Amazon basin. **(b)** Cantons. **(c)** Land cover for 2008. Data from MAE, 2013.

Forest clearing in the Ecuadorian Amazon thus peaked during 1970 - 1990 when the deforestation rate was one of the highest in South America (Mosandl et al. 2008). In the CEA, the forest areas reduced to 80.4%, i.e. 4130 km^2, by the end of 2014, principally due to pasture expansion for cattle ranching (MAE 2017) (Figure 31c). However, this was less intense than in the northeast of the CEA where oil fields were located (Sierra 2000). The declaration of protected areas, which accounted for 29% of the area and few oil discoveries (Wasserstrom & Southgate 2013) probably contributed to a reduced interest in colonization and to deforestation. Nevertheless, improved road connections between Quito and Nueva Loja and recent oil discoveries motivated further colonization of remote areas (Barrera 2014). Despite this, reports indicate a drop in deforestation rates from 92,800 to 74,000 ha year^{-1} (FAO 2015) in Ecuador since 1990.

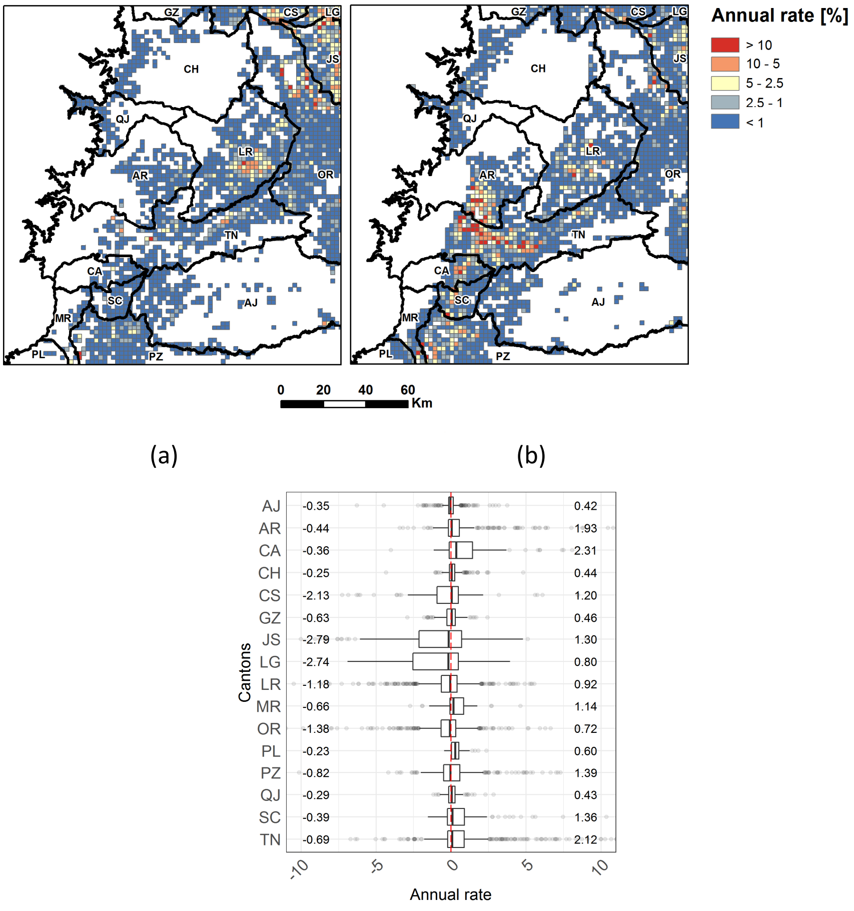
Later, financial instability led to a crisis that ended with the dollarization of the Ecuadorian economy in September of 2000. A reduction in the inflation rate from 96 to 7% was seen as an important sign of economic stabilization for the period 2000-2014 (Anderson 2016). However, Ecuador experienced an unprecedented wave of emigration, especially between 2000 and 2007 (around 483.000 migrants) (Bertoli et al. 2011). Nevertheless, the effects of migration and remittances through land-use change have been associated with a positive effect on agriculture rather than land abandonment and forest transition in the Andean region of Ecuador (Gray & Bilsborrow 2014).

4.3 Materials and Methods

This research was implemented in the R language (R Development Core Team 2017) using specific libraries for spatial data analysis (Pebesma et al. 2017; Hijmans et al. 2017), GWRR (Gollini et al. 2013; Bivand et al. 2017), database management (Dowle et al. 2017), parallel processing (Revolution Analytics & Weston 2015), and data visualization (Wickham & Chang 2016). For GIS analysis, we used ArcGis 10.3 software (ESRI 2010).

4.3.1 Deforestation and reforestation maps

We used maps from previous research that analyzed land-cover change in the CEA for the period 1990 - 2014. They were generated from a novel approach for monitoring long-term forest dynamics with scarce data (Santos et al. 2018), reporting overall accuracies around 78 ± 7%. Specifically, this method applies a post-classification change detection algorithm to a set of 13 biennial land-cover maps to derive deforestation and reforestation areas. In the case of deforestation, the algorithm looks for the date of conversion from forest to a human land use (i.e. pastures, croplands, infrastructures) while excluding areas identified as non-forest at the beginning of the time-series. Conversely, to identify reforestation, the algorithm looks for the date of conversion from human land use to forest considering that the area should be classified as forest for at least two consecutive observations (i.e. four years) before the time-series ends. From these maps, we derived the annual rate of deforestation and reforestation according to the FAO (Puyravaud 2003). We used an analysis grid with a cell size of 400 ha and extracted the forested area in 2000 and its change until 2010 to determine the rate. This time frame was selected in order to match the census databases described later. The cell size for the grid was selected to achieve an acceptable processing time during subsequent calculations. Moreover, it was able to represent our data and implement our analysis. These deforestation and reforestation rates constituted the set of dependent variables analyzed in this research (Figure 32a and Figure 32b). A boxplot of cell rates (Figure 32c) illustrates the average deforestation and reforestation rates for each canton in the CEA.



(a) (b)

(c)

Figure 32. (a) Cells of deforestation and **(b)** reforestation rates for 2000 - 2010 in the CEA. Deforestation rates sign were inverted to establish a common scale in the maps. **(c)** Boxplots of annual rates for each canton in the CEA with mean deforestation (values with negative sign at the left side) and reforestation (value with positive sign at the right side) rates.

Following Kennedy et al. (2015), we removed cells from the analysis grid where the deforestation or reforestation areas were less than 1 ha, and also cells associated to events not relevant to this study, e.g. landslides, local wind throw, floods (Table 13).

Table 13. Descriptive statistics for deforestation and reforestation cells

Forest change dynamics	**Total area [ha]**	**Prefix**	**Variable**	**Cells statistics**				
				Min.	**Mean**	**Max.**	**SD**[2]	**Total count**
Deforestation 2000-2010[1]	29,341.7	D1	Annual deforestation rate[3]	-100	-1.02	-0.02	2.7	2264
Reforestation 2000-2010	28,253.1	R1	Annual reforestation rate[3]	0,02	1.18	100	2.9	2375

[1] To avoid negative values and facilitate the reading of GWRR maps, absolute deforestation rates were considered. [2] SD = Standard deviation. [3] Together, these variables are referred as annual forest change rate.

4.3.2 Dasymetric mapping and population allocation algorithm

We used the 2001 and 2010 population censuses produced by the National Institute of Statistics and Census of Ecuador (INEC 2010) and processed them with a population allocation algorithm. These censuses are organized according to the administrative divisions of Ecuador, and focus on the demographic aspects of the surveyed households. For our analysis, we extracted the most detailed level of information, i.e. census blocks for reducing bias effects caused by the modifiable areal unit problem (Holt et al. 1996). Moreover, since rural population better explains conversion from forest to agriculture land (Carr 2009), we only used census blocks from rural areas and data from the population of the working age (15-72 years). Following Mennis (2003), we derived surface representations of censuses by dasymetric mapping and areal weighting. This technique allows redistributing population counts from a set of areal units into a grid using land cover maps. Therefore, we first identified habitable areas by reclassifying the deforestation map into two non-forest masks, each one for each census year. As non-habitable areas such as water bodies, rock surfaces and cliffs may introduce errors, we masked these areas before using non-forest masks. Consequently, we added a road

Table 14. Road accessibility time and weights assigned

Travel time class [hours]	**Road accessibility [weight]**
0-0.08	1
0.08-0.25	1/2
0.25-0.5	1/3
0.6-1	1/4
1-3	1/5
>3	1/6

accessibility map (SIGTIERRAS 2015) to non-forest masks in order to better represent habitable areas according to travel costs and geographic barriers (Pan et al. 2007). To simplify calculations, we reclassified the road accessibility map into six travel time classes $j = \{1, \ldots, 6\}$ and assigned to each one a unique weight $W = \{1, \ldots, \frac{1}{6}\}$ (Table 14).

These road accessibility weights follow a logarithmic growth, as deforestation and population density has been demonstrated in previous studies (Laurance et al. 2009; Barber et al. 2014). We rasterized census blocks according to the analysis grid considering all its cells (5635 cells in total) for calculating the accessible habitable area N_j for each $i = \{1, \ldots, n\}$ cell within the census block through:

$$N_j = \sum_{i=1}^{n} N_{ij} \tag{3}$$

Then we applied the road accessibility weights W_j to recalculate areas in N_j by:

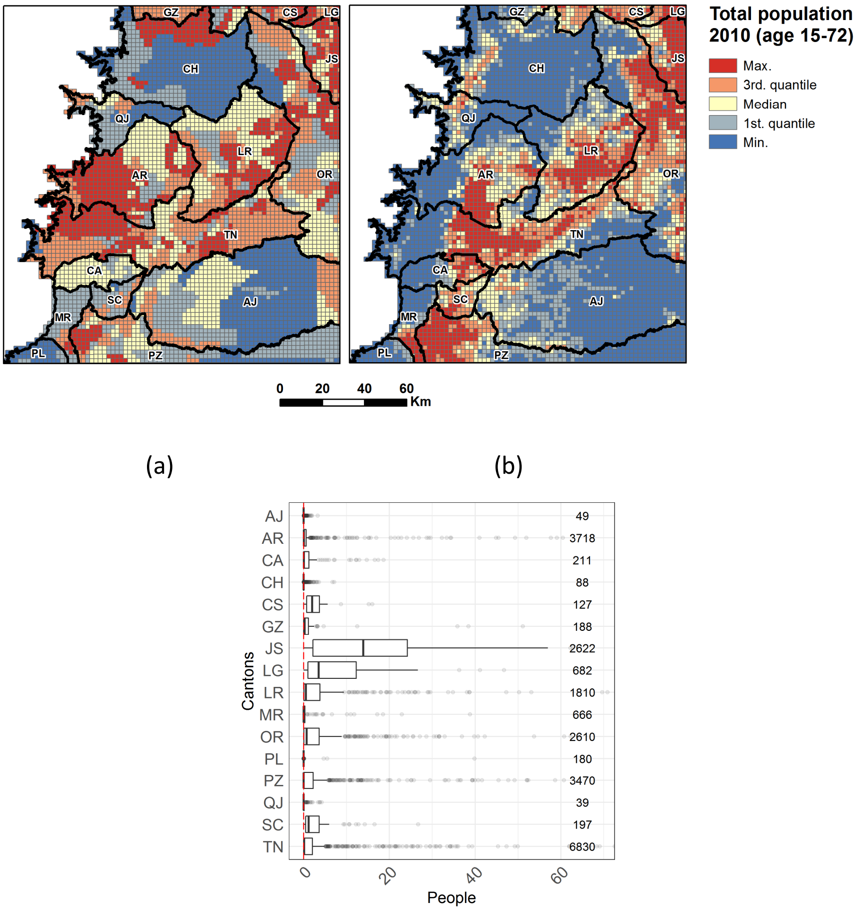
$$N'_j = N_j \times W_j \tag{4}$$

Since the sum of N'_j does not match the total area A in the census block, we transformed it to a percentage applying:

$$W'_j = (N'_j \times 100 \times A^{-1})\,(N_j)^{-1} \tag{5}$$

Hence, W'_j is the weight as a percentage for each cell and travel time class in the census block. As its sum represents A, it was summed to derive a unique weight UW_i for each cell. Finally, UW_i was applied to the population P count in the census block to redistribute it through:

$$P_i = UW_i \times P \times 100^{-1} \tag{6}$$



(a) (b)

(c)

Figure 33. (a) Total population for the census 2010 (age 15 - 72) and **(b)** after allocation. **(c)** Boxplot of allocated population and sum for each canton.

In consequence, P_i represents the allocated population in a census block cell, whose sum is the total population count. This algorithm was applied to each census year variable; an example is given in Figure 33. As during rasterization some census blocks with areas below 400 ha cell area in the analysis grid were omitted, we added the population to the overlapping census block cell before executing the algorithm. Moreover, incomplete census blocks (located at study area boundaries) were considered as complete units in the calculations to avoid erroneous allocations. The set of census variables processed with this algorithm are described in section 4.3.3.

4.3.3 Socio-economic, accessibility, proximity and biophysical variables

We derived 17 socio-economic variables from the processed censuses. Following the approach of Gray et al. (2008), they were selected according to cross-cultural aspects of the population in the Ecuadorian Amazon. Consequently, we classified them into 6 variable groups to describe household and families structure (Household), population age groups (Age), gender structures of households and total population sex (Gender), most frequently spoken languages in the CEA (Language), educational level by groups (Education), and work activities classified by economy sectors (Work). Since variables were from two different census years, we differentiated them to highlight areas of variable change in all cases. Furthermore, we collected an additional set of 11 variables following previous studies on the subject (Armenteras et al. 2017; Mena et al. 2006; Pineda Jaimes et al. 2010). These were classified into 3 additional variable groups describing travel time to collection centers and processing facilities to agricultural products (Accessibility), and Euclidean distance to oil infrastructures, mining sites and paved/dirt roads (Proximity). The latter variable group included biophysical features to describe specific landscape elements (Biophysical). Sources included different Ecuadorian government agencies and a digital elevation model. As these variables needed to be included in the analysis grid, we applied the average for each cell in the analysis grid. Moreover, as variables had different units and distributions, we standardized them before the GWRR was applied. This allowed us to compare outputs and discuss them as effect size according to standard deviations (SD) of the variables (see section 4.4.2 and 4.4.3). However, some variables were transformed from categorical to continuous, and recoded accordingly to the feature observed. A total 28 variables were available (Table 15; Annex 1).

Table 15. Descriptive statistics of variable groups obtained for deforestation cells

Variable group	Prefix	Variable	Cells statistics				Source
			Min.	Mean	Max.	SD	
Accessibility [hours]	A1	Accessibility to oil palm extraction facilities[1]	0-0.08	1-3	>3	-	(SIGTIERRAS 2015)
	A2	Accessibility to coffee and cacao collection centers[1]	0-0.08	0.5-1	>3	-	
	A3	Accessibility to fruit collection centers[1]	0.08-0.25	1-3	>3	-	

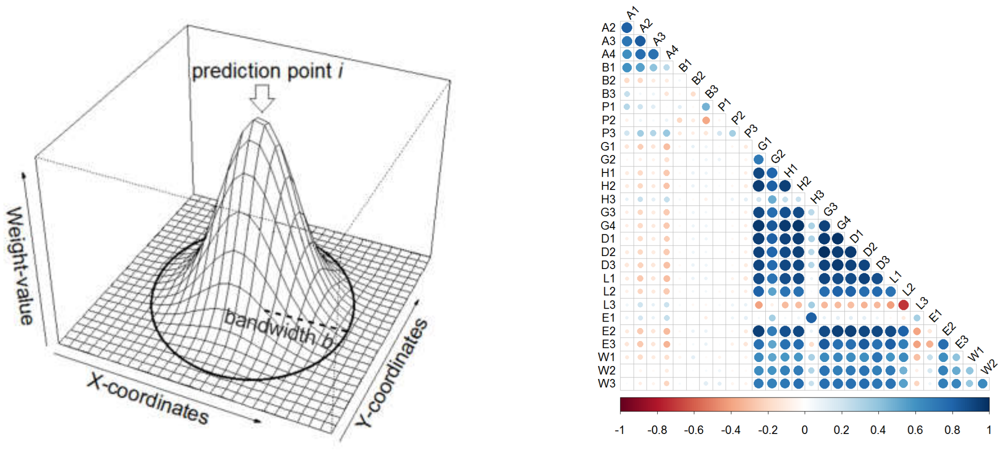
	A4	Accessibility to milk products collection centers [1]	0-0.08	1-3	>3	-	
Age [people]	D1	Younger population (age 15 - 25)	-27	0.8	253	8	(INEC 2001; INEC 2010)
	D2	Adult population (age 26 - 45)	-20	1.1	246	8	
	D3	Older adult population (age 45 - 72)	-19	0.7	152	4	
Biophysical [m.a.s.l.] [2] [unitless] [mm]	B1	Altitude	251	623	3007	471	(SNI 2017)
	B2	Soil fertility (>2% organic matter at max. value)	0	3	4	-	
	B3	Annual rainfall	1372	3576	5892	621	
Education [people]	E1	Basic education (1 - 6 years)	-65	-0.4	53	3	(INEC 2001; INEC 2010)
	E2	Secondary education (7 - 12 years)	-29	2.1	439	14	
	E3	Higher education (>13 years)	0	1.3	267	8	
Gender [people]	G1	Chief male household	-9	0.8	156	5	(INEC 2001; INEC 2010)
	G2	Chief female household	-18	0.2	56	2	
	G3	Male population	-33	1.4	319	11	
	G4	Female population	-34	1.3	333	11	
Household [people]	H1	Mothers with 1 - 2 children (small families)	-4	0.2	46	1	(INEC 2001; INEC 2010)
	H2	Mothers with 3 - 5 children (medium families)	-8	0.4	98	3	

	H3	Mothers with more than 5 children (large families)	-61	-0.3	26	2	
Language [people]	L1	Speak Spanish [3]	-30	3.1	668	22	(INEC 2001; INEC 2010)
	L2	Speak Kichwa [4]	-27	1.6	715	16	
	L3	Speak other languages [5]	-250	-0.5	43	6	
Proximity [meters]	P1	Distance to oil infrastructures	198	8670	46804	6879	(SNI 2017)
	P2	Distance to mining sites	167	8084	59025	8507	
	P3	Distance to paved and dirt roads	45	2072	28749	3031	
Work [people]	W1	Agriculture related workers	-36	0.5	47	3	(INEC 2001; INEC 2010)
	W2	Industry related workers	-17	0.1	72	2	
	W3	Services related workers	-12	0.1	85	2	

[1] Categorical variable ordered and recoded as continuous. [2] Meters above sea level. [3] Most spoken language by *colonos* in the CEA. [4] Second most spoken language and ethnicity in the CEA. [5] Includes 12 different native languages except Kichwa.

4.3.4 Geographically weighted ridge regression

GWRR is a statistical method to model spatial relationships under the assumption of spatial non-stationarity and location interdependency. It can be conceived as an extension of ordinary regression analysis while incorporating local estimates and surface representations of relationships among dependent and independent variables. According to Brudson et al. (1996), GWRR incorporates additional parameters in the linear regression referring to the spatial location (u_i, v_i) of a prediction point i in order



(a) (b)

Figure 34. (a) Gaussian kernel function schema and **(b)** correlation plot for deforestation variables

to model its spatial relationships through:

$$\gamma_i = \beta_0(u_i, v_i) + \sum_{k=1}^{m} \beta_k\,(u_i, v_i) \cdot \chi_{ki} + \varepsilon_i \tag{7}$$

where γ_i is the dependent variable, χ is a vector of $k = \{1, \dots, m\}$ independent variables, β_0 is the estimated intercept, β_k is a vector of regression coefficients, and ε is the error term of the estimation. As GWRR is spatially weighted, it applies a function to observations near to i in such way that closer ones influence more than those further away. Therefore, the weight function is defined by: 1) the type of distance between i and its neighbors, 2) a kernel function specifying the weighting scheme, and 3) the bandwidth distance to control the number of observations within the kernel. Following previous studies (Gao & Li 2011; Su et al. 2012), we used Euclidean distance to measure distances between i and its neighbors and a Gaussian kernel function (Figure 34a) as the weighting scheme, which is defined by:

$$W_{ij} = \exp\left(-\frac{1}{2}\left(\frac{d_{ij}}{b}\right)^2\right) \tag{8}$$

where W_{ij} is the weight assigned to observation j for the estimation of i, d_{ij} is the distance between j and i, and b is the bandwidth. The latter is a sensible parameter in GWRR calibration as large values result in global regression estimates, while small ones introduce randomness (Cahill & Mulligan 2009). To define b, we tried different bandwidth sizes as automatic procedures (Bivand 2017) failed with our data. Therefore, local condition numbers (LCN) were used for validation considering that regression coefficients do not exceed a recommended threshold of 30 LCN (Brunsdon et al. 2012), otherwise collinearity problems could be expected. In this regard, bandwidth size was defined as 15 km using around 172 observations within the kernel. This value represented around 7% of the total observations in the deforestation and reforestation datasets. Following Gollini et al. (2013), we applied ridge regression (Hoerl et al. 1975). This procedure can be perceived as an extended linear regression, which adds a penalty to regression estimates in order to reduce collinearity effects (Figure 34b). For this, it minimizes the equation:

$$\hat{\beta} = \min_{\beta} \sum_{i=1}^{n} (\gamma_i - \hat{\gamma}_i)^2 + \lambda \sum_{j=1}^{m} \beta_j^2 \qquad (9)$$

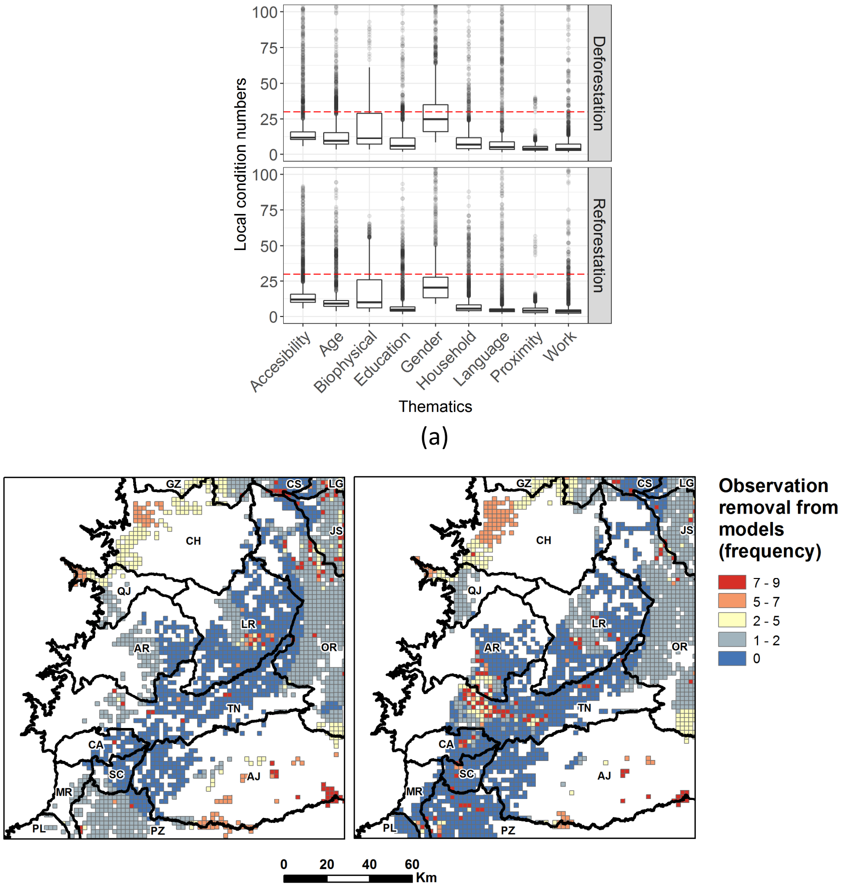
where $\hat{\beta}$ is the ridged regression estimate, β is the raw estimate, γ_i is the dependent variable, and $\hat{\gamma}_i$ its predicted value. In other terms, the summation on the left corresponds to the residual sum of squares of the linear regression, while the summation on the right corresponds to the sum of square of coefficients multiplied by the penalty parameter λ. The latter controls the size of regression estimates, decreasing the influence of correlated variables in the model (Wheeler 2007). For defining it, a cross-validation approach was applied. This is an iterative process that finds λ according to the best prediction result for $\hat{\gamma}_i$. Thus, we found that λ worked in the range [1.5×10^{-5}, 0.07] with our data. Since $\hat{\beta}$ becomes biased after its penalization, standard errors, t-values and associated p-values are no longer available. Therefore, to determine the relative importance of variables, we ran a different model for each variable group and extracted the predicted values to calculate accuracy metrics (i.e. r-square, root mean square error and Akaine information criterion). This allowed us to identify models that best predicted forest change rates and focused on their variables interpretation. To reduce uncertainty, observations that exceeded 30 LCN and predicted values with a cross-validation score outside the range [1, -1] SD were removed from the regression coefficients reports. Finally, to summarize them (28 in total), we followed the approach of de Freitas et al. (2013), and identified spatial regimes for deforestation and reforestation. Consequently, we clustered the regression coefficients, applying the k-means algorithm (Steinhaus 1957) and 3 clusters for both rates after evaluating the optimal number through the approach of Charrad et al. (2014).

4.4 Results

4.4.1 GWRR collinearity assessment, filtering and regression models

We produced nine models, one for each variable group, to filter suspicious observations among them. In this regard, most models reached LCN < 30. However, the models for gender and biophysical models were more affected by collinearity in both deforestation and reforestation (Figure 35a). We filtered them together with other observations with extreme predictions (i.e. outside the cross-validation score range [1, -1] SD). Island and channel patterns showed high removal frequency (Figure 35b and Figure 35c).

As the cantons CH, GZ and AJ were more prone to collinearity, regression coefficients reports were not used from these areas for most models (see also sections 4.4.2 and 4.4.3). In this regard, gender and biophysical models were the most filtered models, as they maintained 69 ± 5% of their observations. However, collinearity and



(a)

(b)

(c)

Figure 35. (a) LCN boxplot obtained for each GWRR model and variable group. The red dotted line indicates the 30 LCN threshold. Observation removal frequency from GWRR models in **(b)** deforestation and **(c)** reforestation cells extreme predictions were less intense for the remaining models, which kept 89 ± 6% of their observations (Table 16). The R-square values showed that accessibility was the most accurate model for deforestation and reforestation ($R^2 = 0.495$). The education and household models also indicated importance for both rates but explained reforestation rates better ($R^2 = 0.45$) than deforestation rates ($R^2 = 0.375$). Results differed in all other models, and the proximity model in deforestation ($R^2 = 0.37$) and the biophysical model in reforestation ($R^2 = 0.38$) were the most notable. Accordingly, the Akaike information criterion (AIC) indicated that the accessibility model achieved the lowest values (AIC < 6300) in both rates meaning that it was the model with the best performance. Moreover, the language and gender models in deforestation and language in reforestation were also highlighted as relevant models by the AIC. Finally, the root-mean-square errors (RMSE) indicate values less than 50% of one SD of their corresponding dependent variable for the models (RMSE = 0.295). This represents an error in rates predicted by the models of ± 0.9% for both deforestation (2.7%) and reforestation (2.9%) SD.

Table 16. Results of GWRR models for variable groups

Model	**Variable [number]**	**Observation kept [%]**		**R-square**		**AIC** [1]		**RMSE** [2]	
		Def. [3]	Ref. [4]	Def.	Ref.	Def.	Ref.	Def.	Ref.
Accessibility	4	84.0	83.8	0.50	0.49	6158	6118	0.30	0.31
Age	3	84.6	87.1	0.34	0.31	6313	6438	0.30	0.31
Biophysical	3	74.4	72.4	0.24	0.38	6309	6289	0.27	0.31
Education	3	88.5	86.0	0.37	0.45	6321	6376	0.29	0.30
Gender	4	64.6	75.8	0.33	0.36	6298	6416	0.31	0.31
Household	3	91.1	89.2	0.38	0.45	6300	6359	0.29	0.28
Language	3	88.6	88.5	0.34	0.32	6244	6352	0.30	0.30
Proximity	3	95.3	94.3	0.37	0.32	6317	6368	0.29	0.32
Work	3	91.6	90.4	0.34	0.27	6307	6442	0.29	0.32

[1] Akaine Information Criterion (AIC).[2] Root-mean-square-error (RMSE). [3] Deforestation cells.

[4] Reforestation cells.

4.4.2 Forest change rates and regression coefficients

After ensuring that the regression coefficients were free of outliers, we derived boxplots to facilitate reporting. Additionally, to observe the interaction of the coefficients with forest change rates, we added them as colored points behind the boxplots (Figure 36), where each model is plotted at each facet and grouped according to its information type. Since data was spread in gender and reached a range of [-8, 6] SD, the Y axis was zoomed to a fixed range of [-2, 2] SD (see Annex 2 for regression coefficients).

It is shown that improved accessibility to oil palm, coffee and cacao facilities (A1 and A2) influenced deforestation (upper section of Figure 36), while there was only a marginal impact of accessibility on fruit and almost none on milk facilities (A3 and A4). This is underlined by reduced proximity to roads (P3) but greater distance to oil and mine infrastructures (P1 and P2). As oil palm cultivation is suitable below 1500 m.a.sl., typically in flat but not waterlogged areas (Pirker et al. 2016), the biophysical model describes the tree’s demands, i.e. reduced altitude and annual rainfall (B1 and B3) but fertility increase (B2), which may suggest that the model indicates optimal requirements. Looking into the census-related models with regard to deforestation rates, increase in secondary education (E2) can be highlighted. Nevertheless, older

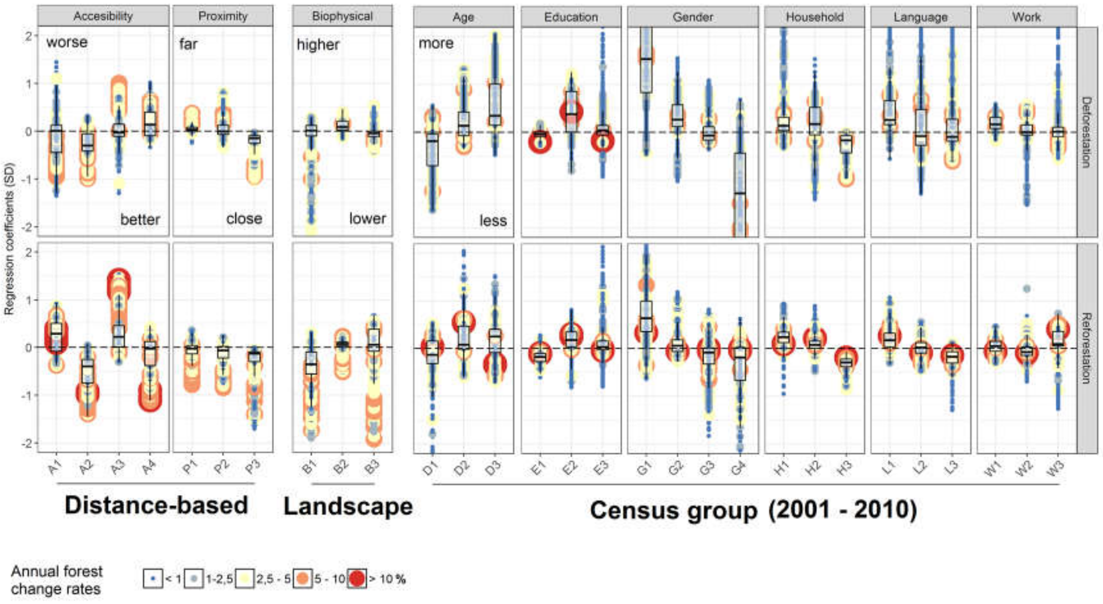


Figure 36. Boxplots of regression coefficients for deforestation (upper plot) and reforestation (lower plot) cells, together with their correspondent annual forest change rate (colored points).

adults and chief male households (D3 and G1) show an even higher increase with regard to the regression coefficients. The sharp decrease in the female population (G4) compared to the increase in the male population (G3) could further refer to population growth and its associated deforestation in the CEA linked to gender asymmetry, where the variable 'educated males above 25 years' dominates (D2 and D3). Additionally, models for medium families, Spanish language and agricultural workers (H2, L1 and W1) stand out. In this regard, deforestation in the CEA seems to be mainly associated with new medium-sized *colonos* families, whose principal activity was agriculture.

For reforestation (lower section of Figure 36), accessibility to oil palm and fruit facilities (A1 and A3) shows an opposite behavior when compared to deforestation. This suggests that reforestation was more prone to take place in areas with limited agricultural capabilities or where soils were depleted. This is also confirmed in the data, as decrease in soil fertility (B2) but increase in annual rainfall (B3) may describe landscape areas not suitable for oil palm but more suitable for coffee, cacao and cattle ranching. This may also explain why improved accessibility to the latter (A2 and A4) characterizes reforestation in the CEA. Regarding the census-related models, we observed marginal to negative population growth related to reforestation in most variables. Therefore, we will focus on this aspect, as rates seem to be higher here where a younger population (male and female), basic education and large families decreased (D1, E1, G3, G4 and H3). This may suggest migration or land-abandonment processes, which could have triggered reforestation. Moreover, as other languages decrease (L3), these people may belong to ethnicities other than the *colonos* and Kichwa. However, since industry workers also decrease (W2), another reading is possible as this can be linked to young workers with short-term contracts with oil companies or at mining sites. This could explain why considerable reforestation took place nearer to oil and mining infrastructures (P1 and P2) but also to roads (P3).

4.4.3 Regression coefficient surfaces – globally and locally influencing variables

To spatially describe underlying driving forces, regression coefficients surfaces are presented in Figure 37 and Figure 38. These were classified according to their influence on deforestation [>1, <-1] SD, and a distinction between those acting globally or almost

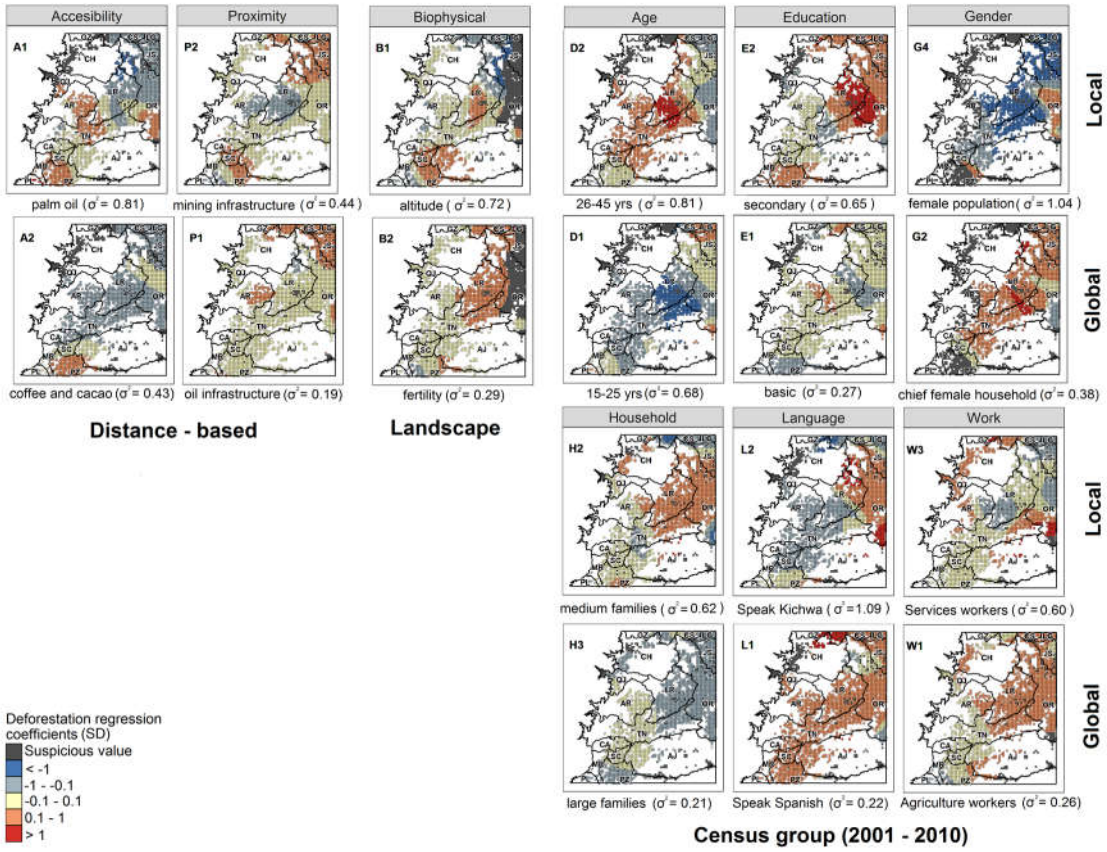

Accesibility
Proximity
Biophysical
Age
Education
Gender
A1
P2
B1
D2
E2
G4
Local
palm oil (σ^2 = 0.81)
mining infrastructure (σ^2 = 0.44)
altitude (σ^2 = 0.72)
26-45 yrs (σ^2 = 0.81)
secondary (σ^2 = 0.65)
female population (σ^2 = 1.04)
A2
P1
B2
D1
E1
G2
Global
coffee and cacao (σ^2 = 0.43)
oil infrastructure (σ^2 = 0.19)
fertility (σ^2 = 0.29)
15-25 yrs (σ^2 = 0.68)
basic (σ^2 = 0.27)
chief female household (σ^2 = 0.38)
Distance - based
Landscape
Household
Language
Work
H2
L2
W3
Local
medium families (σ^2 = 0.62)
Speak Kichwa (σ^2 = 1.09)
Services workers (σ^2 = 0.60)
H3
L1
W1
Global
large families (σ^2 = 0.21)
Speak Spanish (σ^2 = 0.22)
Agriculture workers (σ^2 = 0.26)
Census group (2001 - 2010)
Deforestation regression coefficients (SD)
Suspicious value
< -1
-1 - -0.1
-0.1 - 0.1
0.1 - 1
> 1
CH
OJ
AR
TN
CA
SC
PZ
JS
OR
LR
AJ

Figure 37. Surface regression coefficients and variance for variables behaving more locally or globally in deforestation cells

equally was made calculating their variance. Therefore, regression coefficient surfaces with high variances are plotted in the upper part (local behavior) and those with lower variances in the lower part of the figures (global behavior).

We observed that the variables distance to oil infrastructures (P1) ($\sigma^2 = 0.19$), large families (H3) ($\sigma^2 = 0.21$) and Spanish language (L1) ($\sigma^2 = 0.22$) achieved the lowest variance and behaved as the most globally influencing variables. On the other hand, the variables speak Kichwa (L2) ($\sigma^2 = 1.09$), female population (G4) ($\sigma^2 = 1.04$) and age 26-46 (D2) ($\sigma^2 = 0.81$) achieved the highest variance and behaved as the most locally influencing variables.

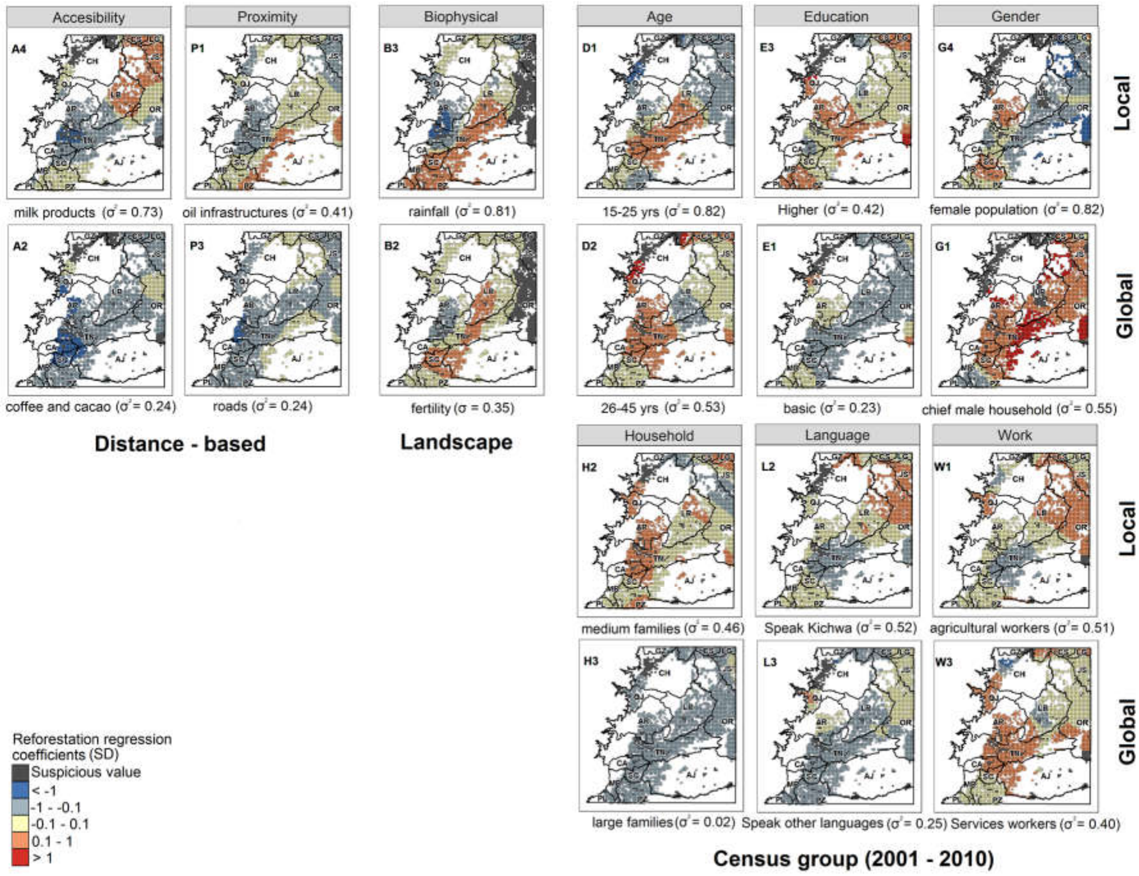

Accesibility
Proximity
Biophysical
Age
Education
Gender
A4
P1
B3
D1
E3
G4
Local
milk products ($\sigma^2 = 0.73$)
oil infrastructures ($\sigma^2 = 0.41$)
rainfall ($\sigma^2 = 0.81$)
15-25 yrs ($\sigma^2 = 0.82$)
Higher ($\sigma^2 = 0.42$)
female population ($\sigma^2 = 0.82$)
A2
P3
B2
D2
E1
G1
Global
coffee and cacao ($\sigma^2 = 0.24$)
roads ($\sigma^2 = 0.24$)
fertility ($\sigma = 0.35$)
26-45 yrs ($\sigma^2 = 0.53$)
basic ($\sigma^2 = 0.23$)
chief male household ($\sigma^2 = 0.55$)
Distance - based
Landscape
Household
Language
Work
H2
L2
W1
Local
medium families ($\sigma^2 = 0.46$)
Speak Kichwa ($\sigma^2 = 0.52$)
agricultural workers ($\sigma^2 = 0.51$)
H3
L3
W3
Global
large families ($\sigma^2 = 0.02$)
Speak other languages ($\sigma^2 = 0.25$)
Services workers ($\sigma^2 = 0.40$)
Census group (2001 - 2010)
CH
LB
OR
JS
TN
CA
AJ
PZ
Reforestation regression coefficients (SD)
Suspicious value
< -1
-1 - -0.1
-0.1 - 0.1
0.1 - 1
> 1

Figure 38. Surface regression coefficients and variance for variables behaving more locally or globally in reforestation cells

With respect to reforestation, we observed that the variables large families (H3) ($\sigma^2 = 0.02$), basic education (E1) ($\sigma^2 = 0.23$), and access to coffee and cacao (A2) ($\sigma^2 = 0.24$) achieved the lowest variance and behaved as the most globally influencing variables. Furthermore, the variables female population (G4) ($\sigma^2 = 0.82$), age 15-25 (D1) ($\sigma^2 = 0.82$) and annual rainfall (B3) ($\sigma^2 = 0.81$) achieved the highest variance and behaved as the most locally influencing variables.

4.4.4 Spatial regimes and underlying causes

We clustered the regression coefficients to determine the spatial regimes of deforestation and reforestation. For this, we also filtered suspicious regression coefficients from observations before reporting averaged regression coefficients for each cluster. However, we used all observations in the mapping to facilitate their localization and descriptions (deforestation Figure 39; reforestation Figure 40).

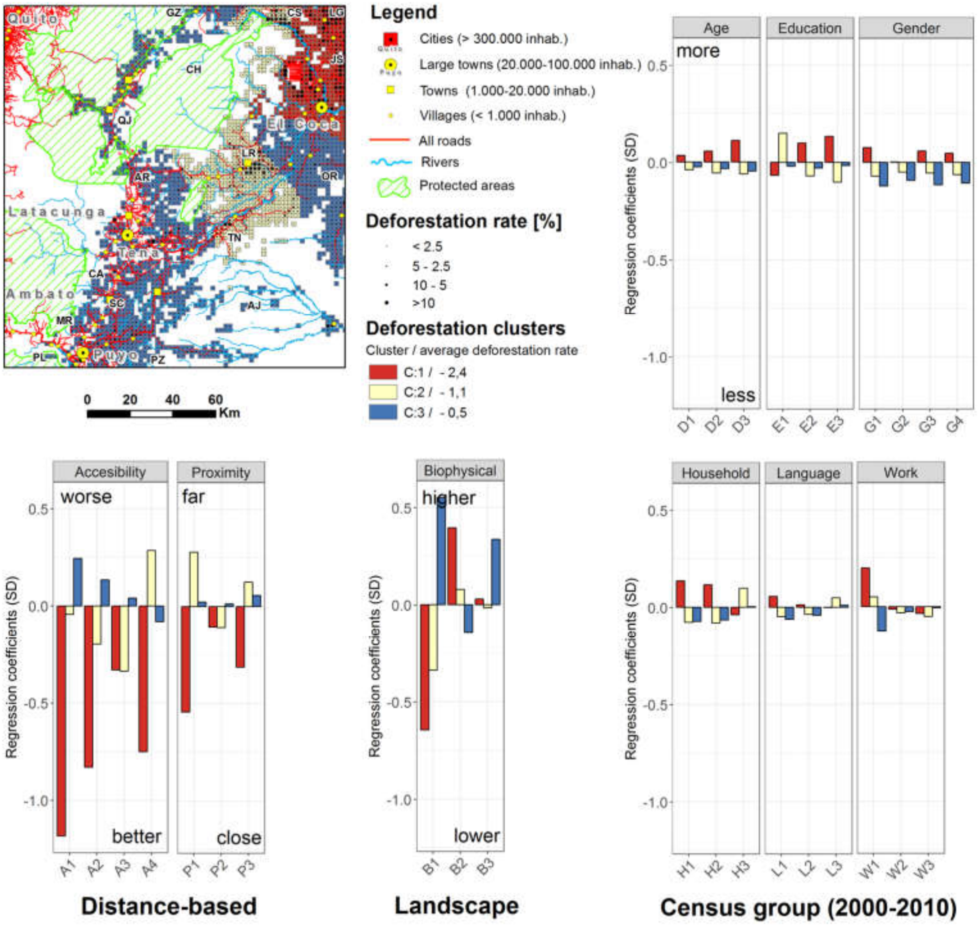


Figure 39. Regression coefficients for deforestation clusters and their localization in the CEA

It is shown that regression coefficients contrast gradually in clusters; however, for this discussion we focus only on cluster 1, as it achieved the highest deforestation average (-2.4% annual, C:1, red in Figure 39). In this regard, cluster 1 highlights a section of the colonization front related to oil fields in the CEA (Sierra 2000), whose proximity to oil infrastructures and roads (P1 and P3) is clearly evident. Moreover, an increased accessibility to all agriculture facilities (A1, A2 and A3) including cattle ranching (A4) indicates an intense land use besides oil extraction. Since this accessibility is located at low altitudes (B1) and on high-fertility soils (B2), its improved suitability for commercial agriculture seems to have attracted deforestation more than other areas of the CEA. Furthermore, variable groups related to population indicate similar patterns to those described for deforestation (section 4.4.2). In this regard, an increase in older adults, secondary education, chief male household, medium families, Spanish language and agricultural farmers (D3, E2, G1, H2, L1 and W1) seems to reinforce these findings.

Nevertheless, increases in higher education and small families (E3 and H1) add new features to population related to deforestation. Here, highly qualified workers in agro-industry or oil extraction companies may explain this association.

With respect to reforestation, a similar contrasting behavior between regression coefficients and clusters is observed. For this, we also focus on cluster 1 (1.9% annual, C:1, red in Figure 40), since it achieves the highest reforestation rate. Therefore, its localization is related to an old colonization front, which converges with territories of historical occupation (Muratorio 1998) of mainly Kichwa communities (L2). Since this area was not found suitable for oil extraction, increasing distances to infrastructures (P1) are observed; however, these are greater when related to mining sites (P2). Furthermore, poorer accessibility is representative in oil palm (A1) and slightly in fruit (A3) facilities, while it is otherwise minimal or even zero (A2 and A4). This may indicate that reforestation were more related to areas not suitable for commercial agriculture,

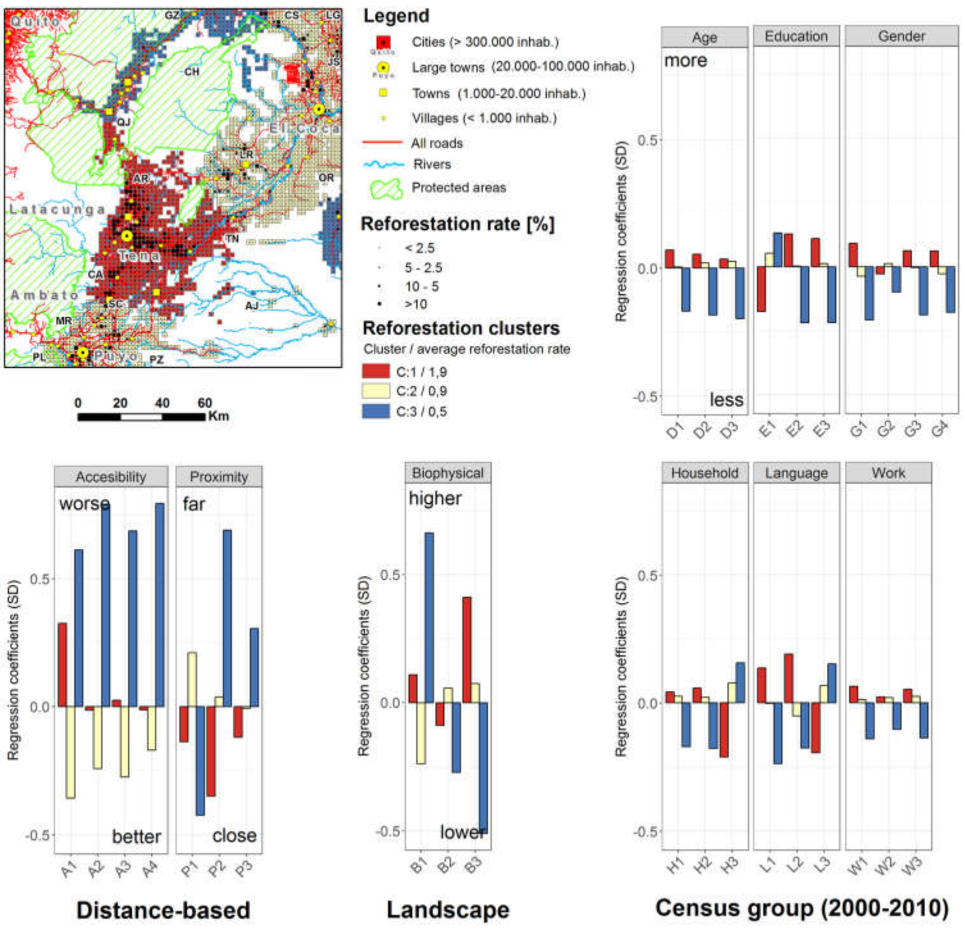


Figure 40. Regression coefficients for reforestation clusters and their localization in the CEA

as higher altitudes, lower soil fertility and intense annual rain (B1, B2 and B3) are the limiting factors. With respect to population, a slightly similar picture to that of deforestation can be observed (compare Figure 39 and Figure 40); however, there are some differences. On the one hand, the increase in the young to adult population related to industry and services works (D1, D2, W2 and W3) may indicate how this setting could be specifically related to reforestation when the population increases. On the other hand, the decrease in basic education, large families and other languages (E1, H3 and L3) may indicate a different setting similar to that described in section 4.4.2, which also favors reforestation but when population decreases. Since both processes may be related to land abandonment rather than to specific actions towards reforestation, migration from rural to urban areas may be a driving force.

4.5 Discussion

4.5.1 Achievements of dasymetric mapping for census data processing

Previous studies applying dasymatric mapping have shown its effectiveness and improved performance regarding interpolating population numbers (Lo & Yang 2002; Reibel & Agrawal 2007). This advantage was also experienced in this study, as multiple inconsistencies between boundaries of census blocks were solved using this approach. Moreover, inter-census analysis allowed us to identify areas of population change using the most detailed census information (INEC 2001; INEC 2010). This improved spatial resolution from an averaged census block area of 5547 ± 10525 ha to grid cells of 200 ha in resampled censuses. This increased the number of observations from 394 ± 81 census blocks to 5695 observations in the analysis grid. Therefore, GWRR improved performance as datasets with fewer observations are prone to errors (Páez et al. 2011). The use of land-cover and land-use change products is thus crucial to improve areal-based data processing, which could not be limited to population census but also to similar data sources (e.g. agricultural and migration censuses). However, researchers should be aware of those areas to be targeted to allocate respective census counts (e.g. croplands or pastures in agricultural censuses). Furthermore, as we used a road accessibility model to enhance our population allocation model, some observations are worth mentioning. First, this input data incorporates restrictions with regard to areas less probable to be inhabited. Therefore, its use is recommended when the assumption of the logarithmic relationship between population, roads and deforestation is valid. While this link was demonstrated in our study area (Etter et al. 2006; Pan et al. 2004), we cannot guarantee that errors are not present in the land-cover maps or in the road accessibility model. This certainly introduced noise in our results, which were expressed as anomalous populated areas in remote sites that required elimination. Moreover, since deforestation is caused not only by human activities but also by natural events (not focused on in this study), an automatic procedure to differentiate among them may reduce errors induced by their confusion. Several studies exist in this regard (Kennedy et al. 2015; Hermosilla et al. 2015b; Oeser et al. 2017) but to the best knowledge of the authors, its implementation is still pending evaluation in the Tropical Andes. Additionally, characterization of forest disturbances may also extend FDD analysis to specific land-cover change dynamics (de Freitas et al. 2013), of which the underlying forces are better understood when based on socio-economic data.

4.5.2 GWRR advantages and limitations

Our results indicate that despite some limitations, GWRR reduced collinearity effects and provided useful outputs for FDD analysis. Other similar studies have suggested alternative approaches, including analysis of one variable at a time (Tu 2011) or reducing variable redundancy by principal component analysis (PCA) (Pineda Jaimes et al. 2010). Both approaches were tested before: In the first case, we experienced a time-consuming task calibrating and verifying each model, while in the second case, interpretation of regression coefficients and their effect over forest change rates were not clear after PCA. Therefore, GWRR offered the best alternative in terms of accuracy, error identification and implementability. In this way, the FDD analysis benefited from a set of regression coefficients and spatial representations that better explained influence and behavior (local or global) of variables over forest change rates. However, the method has disadvantages worth mentioning. First, according to Myers (1990), ridge regression is a biased estimation technique where $\boldsymbol{\lambda}$ (i.e. penalty parameter) defines how much stability should be arbitrarily assigned to regression coefficients. Additionally, as GWRR requires defining the bandwidth size of its moving window, once again, an arbitrary decision must be taken regarding its size. While both parameters can be found through cross validation, a first attempt applying it to bandwidth parameter created collinearity problems and degraded our results. According to Cho et al. (2010), this may be caused by spatial autocorrelation induced by its selection, which is not properly adressed in GWRR and is currently object of further research. Nevertheless, our procedure, i.e. exploring different bandwidths, worked better for the definition of this parameter, while cross validation helped to identify $\boldsymbol{\lambda}$ in our case. Future implementations should consider that not only these parameters but also others such as bandwidth function and its type are the object of critisism, and data-driven approaches may not be the best choice in some cases. Moreover, as ridge regression may not be able to report the relative importance of variables, its implications should be considered if the objective of the study is to determine them according to conventional hypothesis testing (i.e. deriving p-values or significances). While we fill this gap with the presented approach, researchers applying GWRR should consider Bayesian spatially varying coefficients, which offer a richer inferential framework (Finley 2011; Wheeler & Calder 2007).

4.5.3 Forest dynamics drivers, population growth and agricultural expansion in the CEA

Our most significant contribution in this research demonstrates that despite the nature of the drivers and the region studied local differences exist that diversify and multiply FDD explanations. While other studies report these drivers (represented here as variable groups) as forces acting in the whole region (RAISG 2015), we highlighted which of them impacted among different areas and how much. Here, variable impacts could be differentiated in each cell compared to higher or lower influence with regard to deforestation and reforestation rates. Identifying this variable complexity offers the potential to design more precise strategies on the adminsitrative level to target deforestation reduction or reforestation encouragement.

Furthermore, our analysis reveals certain factors that deserve mentioning. Related to deforestation, an important effect of gender asymmetries in population could be observed, i.e. a higher number of males over females showed a link to higher deforestation rates. This relationship was also confirmed in other studies (Moran et al. 2003; Barbieri & Carr 2005). However, this aspect has been less investigated and should be included, not only in research but also in policy making. Nevertheless, it is recommended that gender is subject of different interpretations regarding ethnicity. Therefore, our approach applying language as a proxy of gender does not reveal anything already known in literature, i.e. *colonos* population growth was more related to deforestation. Combining gender and ethnicity could be a more valuable approach also integrating different roles with regard to different beliefs and ethnicity (Hutchison & Vallejo 2016; Villamor et al. 2015). For those variable groups not related to population, we observed that accessibility to palm oil facilities, followed by coffee, cacao and milk products, were the main attractors of deforestation in the CEA. Therefore, they represent a strong agriculturally but also economically important variable. High fertility of soils together with closer distance to roads and oil extraction infrastructures further demonstrated a contribution to increased deforestation rates. Nevertheless, our analysis highlights these areas especially in the northeast of the CEA, where other studies have concluded similar drivers (Mena et al. 2006; Sierra 2000).

In contrast, the south-western part of the CEA showed higher reforestation rates. In this regard, small to almost zero population growth was mostly observed but also variables with important decreases. A decrease in number of people with basic education, large families and other etnicities (as assumed from other native languages spoken) showed a relationship with high reforestation rates. This suggests specific processes related to depopulation or land-use change of specific ethnicites, not well known but observed in other regions of the Ecuadorian Amazon (Stephen J. Walsh et al. 2008). Contrary to deforestation, poor accessibility to palm oil facilities indicated higher reforestation rates, while other accessibility values tended to be of little or zero importance. This is also linked to areas not suitable for commercial agriculture such as higher altitudes, lower soil fertility and intense annual rain characterizing reforestation. However, closer distance to roads indicates a similar behavior with respect to deforestation. While this could be seen as being contradictory, Rudel et al. (2002) found different patterns between distance and reforestation among *colonos* and Shuar ethnicities in Ecuador. In this study, reforestation in the first case was more prone to happen at greater distances, while in the second only at shorter distances. In our case, this may imply that the Kichwa ethnicy may behave similarly to the Shuar, as our findings highlight areas inhabited mostly by these communities, and land-use practices could be similar (Torres et al. 2015). Nevertheless, overexplotation and soil degradation may be another explanation of such patterns (Rey Benayas 2007). Future approaches analyzing FDD may consider studying each ethnicity separately to gain more conclusive results.

4.6 Conclusions

This research underlines the importance of downscaling global problems to the local scale and assessing individual drivers of land-use change in coupled socio-ecological systems. Applying dasymetric mapping together with GWRR supported the analysis of the spatial distribution of population and forest dynamics in the Ecuadorian Amazon. Integrating forest dynamics with socio-economic variables helped to identify complex interactions among them. While at the global scale, key drivers can be identified and variable groups show their impact on forest dynamics, at the local scale they can differ significantly. Here, individual drivers can play more important roles than those at global scales. This is demonstrated in this study, as variable groups played different roles in forest change varying in magnitude and effect within different regions in the CEA. In this regard, accessibility to collection centers and processing facilities to agricultural products showed the most influential role in both deforestation and reforestation. However, a biophysical variable group cannot be minimized since it is an ancillary source that supports and corroborates findings focusing on it, i.e. describing suitable conditions for cultivation. Furthermore, gender, ethnicity and household structure showed high influence regarding untangling population dynamics and their relationship with forest change. However, this made interpretation of the results challenging and final statements more fuzzy. Nevertheless, combining forest dynamics and socio-economic information in a geospatial environment underlined variable complexity and their extent. Combination of individual aspects of livelihood patterns can be more meaningful than a proxy selected to represent one aspect of a livelihood. The results of this study also highlight the role of women and ethnicity in forest dynamics, which is more studied in social sciences. Interdisciplinary expertise and transdisciplinary exchange is needed to foster the understanding of coupled socio-ecological systems from local to global scales. With regard to the “underlying causes” as stated by Geist and Lambin (2002), the findings of this study show that analysis on small scales still needs further assessment to guide local actions carried out at larger administrative scales such as the Tropical Andes. This can only be facilitated by inter- and transdisciplinary research.

Acknowledgments

This work was supported by a research grant from SENESCYT for F.S. The contents of this paper are the sole responsibility of the authors. Special thanks to the R community and Lu BinBin for sharing their code and support. F.S. further thanks to his colleagues from ZFL, ZEF and RSRG in Bonn, Germany, for their support and helpful comments.

5 RESEARCH SYNTHESIS AND CONCLUSIONS

5.1 Summary

In this thesis, a methodological framework for monitoring tropical forest dynamics and evaluating its causes in a case study in the CEA are presented. A processing chain was developed for accomplishing these tasks. The main challenges data scarcity, topographic complexities and landscape heterogeneity were not an impediment for accomplishing the three research objectives. The first objective focused on the evaluation of time-series analysis and pre-processing algorithms in test sites. The second objective on the implementation of the multi-date classification in the Upper Napo Watershed. Finally, the third objective assessed the causes of deforestation and reforestation in the CEA.

5.1.1 Objective I: Evaluation of time-series algorithms for monitoring forest dynamics with genetic algorithms

An exploration of different time-series algorithms for forest monitoring was conducted. Empirical evidence led us to conclude that these were not adequate for the conditions of the study case. Despite testing several processing chains with different ensembles of algorithms, noise induced by atmospheric contamination and scarce data limited performance. Nevertheless, important insights and advances in the design and coding of the processing chain were achieved, which were later processed in a software (TFDynamics) for forest dynamics monitoring in the R language (Annex 3).

An alternative approach for designing processing chains in remote sensing was proposed. Genetic algorithms were demonstrated to be a powerful framework for test algorithms ensembles and for solving parametrization problems. While time-series analysis was concluded as not applicable in the CEA, some optimized processing chains resulted in prototypes that effectively detected deforestation/reforestation events in areas of dense Landsat time-series stacks. Their examination revealed specific routines that were later considered in this research. Findings that were more obvious indicated that radiometric normalization was a required processing step; however, topographic correction required further revision as degraded breakpoint detection. Moreover, indices based on shortwave infrared bands reduced atmospheric contamination effects better than those based on visible bands. On the other hand, less obvious findings indicate that image compositing with temporalities around a year produced the best results, while application of noise reduction routines enhanced breakpoint detection. In this regard, detection of outliers in Landsat time-series seems to reduce the effects of unmasked cloud and shadows, while gap-filling methods varied in all breakpoint detection algorithms. This indicated that gap-filled results were similar despite the algorithm applied; therefore, it was the less sensible processing step in the time-series analysis approach. However, this was different for the signal-smoothing algorithm, which enhanced detection of deforestation/reforestation events in all breakpoint detection algorithms. Concerning the latter, non-parametric statistics-based algorithms produced improved results; however, the number of detected breakpoints varied among them. Segment size seems to determine this difference, indicating that each algorithm has its own capability to detect changes. It is therefore concluded that optimization is an important processing step that is not frequently considered in time-series analysis and remote sensing. Additional research findings indicate that radiometric alignment between adjacent composites footprints were required for large-area processing, while the cloud cover filtering criteria needed to be adjusted as it was not rigorous enough.

Optimization with genetic algorithms required demanding computation simulations and Darwin evolution theory knowledge. Strategies such as ‘divide and conquer’, modular design, and application of functional and parallel coding paradigms were relevant in the first case. This stimulating and important learning experience led to a better implementation of the processing chain and improved coding practices in the following research objectives. With respect to the second case, genetic algorithms opened the door to a different programming paradigm, where evolution became a discovery engine of software improvements. While its specific terminology required a complicated abstraction in the context of remote sensing, admitting its flexibility to adapt to any optimization problem was not less important. Nevertheless, calibration and experimentation required to run multiple simulations before an algorithm was considered as ‘optimized’. This is due the stochastic basis of genetic algorithms, which rely on random variables (e.g. choosing chromosomes during tournament selection or during setting up the initial population). Therefore, optimization results varied from run to run, implying some uncertainty in the results. This involved a time-consuming calculation, especially in low-performance computing environments.

Finally, this research focused on specific testing sites in the CEA. Hence, forest dynamics and landscape diversity were observed at specific locations, but no conclusive patterns were identified at this point (**Q3**: third research question). Nevertheless, this research contributed to refining a methodology, identifying the processing steps for monitoring long-term forest dynamics, and coding the processing chain (**Q1**: first research question). Moreover, field data collection (interviews, field visits, collection of thematic maps, aerial photographs and documents, etc.) provided ancillary sources to better contextualize and validate the results of this research.

5.1.2 Objective II: Implementation of multi-date classification for long-term forest dynamics monitoring with scarce data

After identifying the limitations of time-series algorithms for forest dynamics monitoring, a new approach was implemented considering a different strategy. Defined as multi-date classification, it derived successfully deforestation and reforestation maps for the UNW. Its development followed findings discussed in the first research objective. Therefore, a more careful filtering of atmosphere-contaminated images was applied fusing Fmask (Zhu & Woodcock 2012) with LEDAPS (Masek et al. 2012) cloud mask to provide a more conservative masking approach. Moreover, only indices related to shortwave infrared bands were calculated, while additional derivatives such as Tasseled-Cap transformation bands were also calculated. Furthermore, the compositing algorithm was modified considering standardization and median calculation from the composite images set. This avoided applying radiometric normalization, which was done in the first research objective (i.e. linear regression with invariant features) but discarded in this work, as some erroneous images cases were found as high cloud cover impeded obtaining sufficient invariant features to perform regression. Additionally, a larger time step was considered (i.e. biennials), as more images per composite were available and derivation of “best pixel” value improved reducing residual noise. Since three reforestation sites were visited during field work for taking hemispherical photographs, it was found that this compositing time step was adequate, as canopy closure reached its maximum mostly after 3 to 5 years. Regarding radiometric alignment, histogram match algorithm allowed matching the four footprints needed to complete the UNW mosaic, while it enabled working at regional scales. These procedures completed the pre-processing of Landsat time-series and prepared them for the change detection step.

Data scarcity, atmospheric contamination and topographically complex regions were the main challenges that the multi-date classification algorithm aimed to solve. It first classified composites to later derive change from the semantic values; therefore, it did not depend on the dense images stacks time-series algorithms require. Due to this, a classifier was had to be trained and a set of samples for four different land-cover types (i.e. two forest and two non-forest classes) were interpreted and labelled. While, out-of-bag errors indicated that trained classifiers were above 90% accuracy, a test based on a correlation with a reference forest/non-forest map (MAE 2017) indicated that supporting vector machines were the more appropriate option for this dataset. Consequently, all composites were classified into land-cover maps; however, these still showed remaining noise from unmasked clouds and other artefacts. To eliminate it, a temporal filtering function was developed to recode illogical land cover transitions identified from a set of rules defined *a-priori*. Since the function worked pixel-wise, moreover considering a moving window of three consecutive observations, it had the disadvantage of removing end dates from the time-series in the analysis (i.e. all observations before 1992 and after 2014). Nevertheless, its application using not topographically corrected images resulted in enhanced overall accuracies from 62 to 82% in deforestation and from 48 to 71% in reforestation maps. The latter was identified as areas in a continuous process of “forestedness”, which were classified at least twice as forest (i.e. four years to accomplish the reforestation time threshold) after the area was deforested. In contrast, deforestation was flagged as the first year of a forest pixel being mapped as one of the non-forest classes.

Applying multi-classification in the most challenging areas of the UNW (i.e. almost permanent cloud cover and topographically irregular) was therefore qualified as successful considering the scarce data and difficulties experienced (**Q1**: first research question). Nevertheless, some of these difficulties were also highlighted as limitations for this research objective. In this regard, multi-classification requires training a classifier for deriving land cover maps, as its sampling is not fully automatized. Therefore, a proper interpretation and labeling of a large number of samples is required to achieve good results. This is a time-consuming task and not easily done due its on-screen interpretation. Furthermore, topographic correction with C-correction algorithms showed a visual reduction of topographic shadows; however, radiometric bias and accuracy reduction compared with not-corrected inputs (i.e. 82 to 70% in deforestation and from 71 to 56% in reforestation maps) were observed. This may be due to the low resolution of the digital elevation model (90 m) but a conclusive reason was not determined.

With respect to deforestation and reforestation patterns, the results of this study show for the first time the long-term forest dynamics in the UNW (**Q3**: third research question). Areas with deforestation can be observed mainly along the E20 highway, and reforestation along the E45 highway. A comparison with other sources indicates similar trends; however, figures indicate that these results are more conservative for deforestation (2,319 ha year^{-1}) compared to other sources (2,757 ha year^{-1} and 4,394 ha year^{-1}) (Hansen et al. 2013; MAE 2017). This is similar for reforestation, where results indicate a lower figure (1,504 ha year^{-1}) than the Ecuador Forest Reference Emission Level (2,277 ha year^{-1}) (MAE 2017). Finally, these maps reveal two different regions where deforestation was more frequent, i.e. in the north-eastern part, and reforestation was higher, i.e. in the south-western part. These regions are of interest as forest change in the CEA suggests different driving forces (Chapter 4).

5.1.3 Objective III: Assessment of underlying causes of deforestation and reforestation through geographically weighted ridge regression

A methodology for assessing census data processing and regression analysis was developed for analyzing the driving forces of deforestation and reforestation in the CEA based on two regional-based analyzes.

The first analysis focused on overcoming the boundary change problem in inter-census analysis and the modifiable areal unit problem in data aggregation, both with dasymetric mapping. Following this approach, an algorithm was designed to re-distribute census counts according to a weighting function and ancillary maps. This algorithm assumed that population density can be modeled according to the logarithmic relationship between road distance and deforestation areas. Using non-forest masks calculated for the Objective II (section 5.1.2) and a road accessibility model (SIGTIERRAS 2015), a weighted layer was derived to guide the algorithm to re-distribution of census counts. Since this operation required a common data structure to rasterize data inputs, an analysis grid was created to integrate results. Consequently, the two censuses (i.e. 2001 and 2010) were processed, and as they were calculated on the basis of the re-distributed census counts, their boundary inconsistencies were no longer a problem. Moreover, because they were not aggregated into larger units as the most detailed census level was used, the modifiable areal unit problem was minimized. This allowed operating them and obtaining population change for each of the census variables considered for the 2001 – 2010 period. This improved the spatial resolution of this processing task, increasing the number of observation from 394 ± 81 census blocks to 5695 cells in the analysis grid (**Q2**: second research question). An integration of these values with accessibility, proximity, and biophysical datasets resulted in a multivariate spatial database for regression analysis. Regarding the limitations, anomalous populated areas indicated that this approach was sensitive to misclassification errors introduced in non-forest masks. Moreover, deforestation caused by natural events and not properly masked on these added additional uncertainty.

The second analysis targeted the limitations of conventional regression analysis in studies on the driving forces of deforestation/reforestation. Motivated by the increasing number of studies using geographically weighted regression, an implementation using ridge regression was applied. This methodology allowed the identification of spatial variability of regression coefficients, reducing the collinearity effects attributed to other geographically weighted regression approaches based on ordinary linear regression. The results indicate that the driving forces of deforestation/reforestation are diverse in the CEA, and it was possible to identify which of them behaved more globally or locally. Moreover, a further clustering of regression coefficients allowed identifying spatial regimes of highest deforestation/reforestation rates. These clusters facilitated the description and discussion of underlying drivers considering similar socio-economic and biophysical settings in the study region (**Q2**: second research question). Nevertheless, there were some limitations. First, parametrization of bandwidth size in geographically weighted regression was not straightforward, and required careful evaluation before it was finally set. While automatic approaches exist (e.g. cross validation), the results obtained were not satisfying and required a manual approach for definition. Furthermore, as ridge regression is not able to report variable importance according to conventional hypothesis testing, some uncertainty could be expected. Since most of the studies on the driving forces of deforestation/reforestation reported drivers as p-values or significances, in this approach their absence is a disadvantage.

Two regions with specific socio-economic settings associated with high deforestation and reforestation rates were identified. The first was in the north-western part of the CEA . Here, an important effect of gender asymmetries in the population was identified, i.e. a higher number of males than females was linked to higher deforestation rates, specially among the *colonos* ethnicity. Moreover, better accessibility to palm oil facilities, followed by coffee, cacao and milk products enhanced the effect. Other aspects such as high soil fertility and shorter distance to roads and oil extraction infrastructures also contributed. In the second region, i.e. the south-western part of the CEA reforestation rates were higher. In contrast to deforestation, a low to almost zero population growth was associated. Nevertheless, decrease in the number of people with basic education, large families and etnicities other than *colonos* or Kichwas were related to reforestation. Moreover, low accessibility to palm oil facilities showed higher rates, while others accessibilities, i.e. to coffee, cacao, fruits and milk products, tended to be neither better nor worse. This allowed further association to areas not suitable for commercial agriculture, as higher altitudes, lower soil fertility and intense annual rain characterized reforestation. On the other hand, shorter distance to roads indicated a similar behavior to that of deforestation; however, a difference in ethnicity was found, indicating that Kichwas were more prone to reforestation activities than the *colonos*. These findings improved the explanation of causes of deforestation/reforestation in the CEA (**Q3**: third research question); however, not as definitive reasons as some limitations were observed. In this regard, further inter- and transdisciplinary research is required to better interpret the results. While contributions from different authors during this research are important here, a discussion with local experts is pending to be integrated. Moreover, as this study does not split the analysis by ethnicity, non-conclusive reasons can be associated to the different groups present in the CEA and the land-use practices observed.

5.2 Conclusions and outlook

Evaluating the patterns and drivers of forest dynamics in tropical regions is fundamental for balanced decisions on forest conversions, land planning and policy making. Limited data quantity and quality (e.g. satellite images, socio-economic training/validation data) increase the uncertainty with respect to the observation of deforestation or reforestation processes. The methodology presented in this study is capable of producing a series of products for facilitating this objective. In the first case, forest dynamics monitoring was partly automatized in a processing chain. When including routines for pre-processing, this can transform Landsat surface reflectance products into median composites for specific periods with diminished or even zero cloud cover. Moreover, a post classification change-detection algorithm, together with a temporal filtering technique, can further a synthesis of forest dynamics frontiers derived from land-cover maps collections. This data reveals patterns and landscape trajectories with an improved temporal frequency. However, accuracy depends on multiple factors (e.g. cloud masking, co-register precision, composite quality, algorithm training), which were documented and evaluated in this thesis. Although the results fulfill the monitoring objective of the study case, more research is needed for further improvements. Replication at other sites and development of improved routines (e.g. improved pixel selection in compositing, automatization of sample collection, and implementation of topographic correction) are recommended. Development of this methodology into a cloud-based or high-performance computing environment, which facilitates data processing but also integration with other archives (e.g. Sentinel 1-2, SPOT, ASTER) should be also considered in future research.

Furthermore, limitations regarding analysis of drivers of forest dynamics were identified and recommendations made. By applying dasymetric mapping and areal weighting, socio-economic data were integrated into a common spatial structure with increased spatial resolution, which solved border inconsistencies and aggregation effects. Accessibility, distance, biophysical and inter-census variables were collected and classified to derive nine regression models. Evaluated in a geographically weighted ridge regression, surfaces representations of regression coefficients were obtained. These layers summarized the effect and magnitude of model variables, and their clustering helped to identify spatial regimes. The latter were characterized by common socio-economic and biophysical settings, and this input facilitated discussion of the causes of deforestation and reforestation. In this regard, colonist migration, gender asymmetries, infrastructure accessibility to commercial crops and agricultural suitability are important drivers for high deforestation rates. This was observed where the influence of the oil industry and associated infrastructures was higher, i.e. in the north-eastern part of the CEA. This is in contrast to the high reforestation rates where low or zero population growth, poor accessibility to commercial crops facilities (especially oil palm) and agricultural limitations were the driving forces. While less related to the oil industry, road and mining industry were found to be somehow related to reforestation in the south-western part of the CEA. However, not all these findings can be considered as definitive “causes” as more inter- and transdisciplinary research is required. Future steps will consider integrating additional datasets derived from remote sensing (e.g. land cover trends) and socio-economic data (e.g. migration and agricultural census) to identify other relationships and extend the presented methodology.

Finally, this research contributes to the ongoing development of a national REDD+ Program in Ecuador and its efforts in reducing deforestation. As a country highly vulnerable to climate change, the next decades will be challenging and actions are crucial to guarantee the human right to a sustainable environment for future generations.

REFERENCES

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

Ahlqvist, O., 2008. 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*, 112(3), pp.1226–1241.

Anderson, A., 2016. Dollarization: A Case Study of Ecuador. *Imperial Journal of Interdisciplinary Research*, 2(5), pp.2454–1362.

Angelsen, A. & Wertz-Kanounnikoff, S., 2008. *Realising REDD+: National strategy and policy options*, Bogor, Indonesia: Center for International Forestry Research (CIFOR).

Armenteras, D. et al., 2017. Deforestation dynamics and drivers in different forest types in Latin America : Three decades of studies (1980 – 2010). *Global Environmental Change*, 46(June), pp.139–147. Available at: http://dx.doi.org/10.1016/j.gloenvcha.2017.09.002.

Armenteras, D. et al., 2011. Understanding deforestation in montane and lowland forests of the Colombian Andes. *Regional Environmental Change*, 11(3), pp.693–705.

Arvidson, T., Gasch, J. & Goward, S.N., 2001. Landsat 7's long-term acquisition plan - An innovative approach to building a global imagery archive. *Remote Sensing of Environment*, 78(1–2), pp.13–26.

Asner, G.P. et al., 2014. Landscape-scale changes in forest structure and functional traits along an Andes-to-Amazon elevation gradient. *Biogeosciences*, 11(3), pp.843–856.

Atkinson, P.M. & Tatnall, A.R.L., 1997. Introduction Neural Networks in Remote Sensing. *Http://Dx.Doi.Org/10.1080/014311697218700*, 18(4), pp.699–709.

Baig, M.H.A. et al., 2014. Derivation of a tasselled cap transformation based on Landsat 8 at satellite reflectance. *Remote Sensing Letters*, 5(5), pp.423–431.

Balakrishnan, S. et al., 1996. Genetic Algorithms for Product Design. , 42(8), pp.1105–1117.

Banskota, A. et al., 2014. Forest monitoring using Landsat time-series data - A review. *Canadian Journal of Remote Sensing*, 40, pp.362–384.

Barber, C.P. et al., 2014. Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation*, 177, pp.203–209. Available at: http://dx.doi.org/10.1016/j.biocon.2014.07.004.

Barbieri, A.F. & Carr, D.L., 2005. Gender-specific out-migration, deforestation and urbanization in the Ecuadorian Amazon. *Global and Planetary Change*, 47(2–4 SPEC. ISS.), pp.99–110.

Barbosa, J.M., Broadbent, E.N. & Bitencourt, M.D., 2014. Remote Sensing of Aboveground Biomass in Tropical Secondary Forests: A Review. *International Journal of Forestry Research*, 2014(ID 715796), pp.1–14. Available at: http://www.hindawi.com/journals/ijfr/2014/715796/.

Barrera, D., 2014. *Gestión del territorio y manejo de bienes comunes en contextos extractivos: una aproximación al caso de las comunidades Kichwas del Cantón Arajuno en la Provincia de Pastaza, Ecuador*. Facultad Latinoamericana de Ciencias Sociales (FLACSO).

Bass, M.S. et al., 2010. Global conservation significance of Ecuador’s Yasuní National Park. *PLoS ONE*, 5(1).

Beirne, C. & Whitworth, A., 2011. *Frogs of the Yachana Reserve*, Exeter, England: Global Vision International.

Bengtsson, H., 2016. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors), Version 0.50.2. Available at: https://cran.r-project.org/package=matrixStats.

Benjamin, A. & Leutner, M.B., 2017. RStoolbox: Toolbox for remote sensing image processing and analysis such as calculating spectral indices, principal component transformation, unsupervised and supervised classification or fractional cover analyses, Version 0.1.9. Available at: https://github.com/bleutner/RStoolbox.

Bertoli, S., Moraga, J.F.H. & Ortega, F., 2011. Immigration policies and the ecuadorian exodus. *World Bank Economic Review*, 25(1), pp.57–76.

Bivand, R., 2017. *Geographically Weighted Regression*, Available at: https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf.

Bivand, R., Yu, D. & Miquel-Angel Garcia-Lopez, 2017. Package ‘ spgwr ’, version 0.6-32.

Borchers, H.W., 2015. pracma: Practical Numerical Math Functions.

Borja, I. et al., 2015. Dinámicas de Deforestación y Regeneración en Ecuador.

Breunig M. M. et al., 2000. LOF: Identifying Density-Based Local Outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference On Management of Data*. Dallas, Texas.

Brooks, T.M. et al., 2006. Global Biodiversity Conservation Priorities. , 313(July), pp.58–62.

Brown, L. & Smith, R., 1994. Frameworks of Urban System Evolution in Frontier Settings and the Ecuador Amazon. , pp.72–96.

Brown, S. & Lugo, A., 1990. Tropical Secondary Forests. *Journal of Tropical Ecology*, 6(1), pp.1–32.

Brunsdon, C., Charlton, M. & Harris, P., 2012. Living with Collinearity in Local Regression Models. *International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Sciences*, pp.67–72.

Brunsdon, C., Fotheringham, A. & Charlton, M., 2002. Geographically weighted summary statistics - a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, 26(6), pp.501–524.

Brunsdon, C., Fotheringham, A. & Charlton, M.E., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4), pp.281–298. Available at: http://dx.doi.org/10.1111/j.1538-4632.1996.tb00936.x.

Bustamante, M.M.C. et al., 2016. Toward an integrated monitoring framework to assess the effects of tropical forest degradation and recovery on carbon stocks and biodiversity. *Global Change Biology*, 22(1), pp.92–109. Available at: http://doi.wiley.com/10.1111/gcb.13087.

Buytaert, W., Cuesta-Camacho, F. & Tobón, C., 2011. Potential impacts of climate change on the environmental services of humid tropical alpine regions. *Global Ecology and Biogeography*, 20(1), pp.19–33.

Cahill, M. & Mulligan, G., 2009. Local Crime Patterns. *Social Science Computer Review*, pp.174–193.

Caldas, M.M. et al., 2015. Land-cover change in the Paraguayan Chaco: 2000–2011. *Journal of Land Use Science*, 10(1), pp.1–18. Available at: http://dx.doi.org/10.1080/1747423X.2013.807314.

Camara, G., 2013. *Programa Amazônia - Projeto PRODES*, Instituto Nacional de Pesquisas Espaciais.

Carr, D., 2009. Population and deforestation: Why rural migration matters. *Progress in Human Geography*, 33(3), pp.355–378.

Carrion, D. & Chíu, M., 2011. *National Programme Document - Ecuador. UN-REDD Programme sixth policy board meeting*,

CGIAR - CSI, 2008. SRTM 90m DEM, Version 4. Available at: http://srtm.csi.cgiar.org/index.asp.

Chambers, J.M., 1992. Linear models. Chapter 4 of Statistical Models in S J. M. C. and T. J. Hastie, ed. *Local regression models*.

Chance, C.M. et al., 2016. Effect of topographic correction on forest change detection using spectral trend analysis of Landsat pixel-based composites. *International Journal of Applied Earth Observation and Geoinformation*, 44, pp.186–194. Available at: http://linkinghub.elsevier.com/retrieve/pii/S030324341530026X.

Charrad, M. et al., 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6). Available at: http://www.jstatsoft.org/v61/i06/.

Cho, S.-H., Lambert, D.M. & Chen, Z., 2010. Geographically weighted regression bandwidth selection and spatial autocorrelation: an empirical example using Chinese agriculture data. *Applied Economics Letters*, 17(8), pp.767–772. Available at: http://www.tandfonline.com/doi/abs/10.1080/13504850802314452.

Cincotta, R.P., Wisnewski, J. & Engelman, R., 2000. Human population in the biodiversity hotspots. *Nature (London)*, 404(6781), pp.990–992.

Clark, M.L. et al., 2010. A scalable approach to mapping annual land cover at 250 m using MODIS time series data: A case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment*, 114(11), pp.2816–2832. Available at: http://dx.doi.org/10.1016/j.rse.2010.07.001.

Clayden, J., 2016a. mmand: Mathematical Morphology in Any Number of Dimensions, Version 1.4.0. Available at: https://cran.r-project.org/package=mmand.

Clayden, J., 2016b. NiftyReg: Image Registration Using the NiftyReg Library, Version 2.6.1. Available at: https://github.com/jonclayden/RNiftyReg%0D.

Clement, F. et al., 2009. Drivers of afforestation in Northern Vietnam: Assessing local variations using geographically weighted regression. *Applied Geography*, 29(4), pp.561–576. Available at: http://dx.doi.org/10.1016/j.apgeog.2009.01.003.

Cohen W. et al., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation. *Remote Sensing of Environment*, 114, pp.2911–2924.

Cohen, W.B. et al., 2017. How Similar Are Forest Disturbance Maps Derived from Different Landsat Time Series Algorithms? *Forests*, 8(4), pp.1–19.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), pp.35–46.

Coppin, P. et al., 2004. Review ArticleDigital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25(9), pp.1565–1596. Available at: https://doi.org/10.1080/0143116031000101675.

Coq-Huelva, D., Torres-Navarrete, B. & Bueno-Suárez, C., 2017. Indigenous worldviews and Western conventions: Sumak Kawsay and cocoa production in Ecuadorian Amazonia. *Agriculture and Human Values*, 0(0), pp.1–17.

Crist, E.P. & Cicone, R.C., 1984. A Physically-Based Transformation of Thematic Mapper Data-The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(3), pp.256–263.

Deilami, K., Kamruzzaman, M. & Hayes, J.F., 2016. Correlation or causality between land cover patterns and the urban heat island effect? Evidence from Brisbane, Australia. *Remote Sensing*, 8(9).

DeVries, B. et al., 2015. Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sensing of Environment*, 161, pp.107–121. Available at: http://dx.doi.org/10.1016/j.rse.2015.02.012.

Dou, J. et al., 2015. Automatic case-based reasoning approach for landslide detection: Integration of object-oriented image analysis and a genetic algorithm. *Remote Sensing*, 7(4), pp.4318–4342.

Dowle, M. et al., 2017. data.table: Extension of “data frame”, Version 1.10.4-3. Available at: http://r-datatable.com.

Drusch, M. et al., 2012. Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, pp.25–36.

Eberhart, N., 1998. *Transformaciones agrarias en el frente de colonización de la Amazonia ecuatoriana*, Quito - Ecuador: Abya-Yala.

Eberhart, R.C. & Shi, Y., 1998. Comparison between Genetic Algorithms and Particle Swarm Optimization. *Lecture Notes in Computer Science*, pp.612–615.

Edwards, D.P. et al., 2017. Tropical secondary forest regeneration conserves high levels of avian phylogenetic diversity. *Biological Conservation*, 209, pp.432–439. Available at: http://dx.doi.org/10.1016/j.biocon.2017.03.006.

Eklundh, L. & Jönsson, P., 2015. TIMESAT : A software package for time-series processing and assessment of vegetation dynamics. In Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 141–158. Available at: http://dx.doi.org/10.1007/978-3-319-15967-6_7.

EODC, 2018. Earth Observation Data Centre for Water Resources Monitoring: An open and international cooperation to foster the use of Earth Observation data. Available at: https://www.eodc.eu/ [Accessed May 18, 2018].

ESA, 2018. Copernicus Open Access Hub. Available at: https://scihub.copernicus.eu/dhus/#/home [Accessed April 24, 2018].

Eshelman, L. et al., 1991. Preventing Premature Convergence in Genetic Algorithms by Preventing Incest. In Richard K. Belew & ashon B. Booker, eds. *Proceedings of the Fourth International Conference on Genetic Algorithms*. Morgan Kaufmann

Pubishers, pp. 115–122.

Espinoza, J.C. et al., 2015. Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation. *Water Resources Research*, 51(5), pp.3459–3475.

ESRI, 2010. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Etter, A. et al., 2006. Regional patterns of agricultural land use and deforestation in Colombia. *Agriculture, Ecosystems and Environment*, 114(2–4), pp.369–386.

FAO, 2015. *Global Forest Resources Assessment 2015: Desk Reference*, Rome - Italy: Food and Agriculture Organization of the United Nations (FAO). Available at: http://www.fao.org/3/a-i4808e.pdf.

FAO, 2016a. *Global Forest Resources Assessment 2015*, Rome - Italy: Food and Agriculture Organization of the United Nations (FAO). Available at: http://www.fao.org/forest-resources-assessment/current-assessment/en/.

FAO, 2007. *State of the World's Forests*, Rome - Italy: Food and Agriculture Organization of the United Nations (FAO).

FAO, 2016b. *State of the World's Forests 2016. Forests and agriculture: land-use challenges and opportunities*, Rome - Italy: Food and Agriculture Organization of the United Nations (FAO). Available at: http://ccafs.cgiar.org/news/press-releases/agriculture-and-food-production-contribute-29-percent-global-greenhouse-gas.

Fearnside, P.M., 2005. Deforestation in Brazilian Amazonia: History, rates, and consequences. *Conservation Biology*, 19(3), pp.680–688.

Finley, A.O., 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2(2), pp.143–154.

Flood, N. et al., 2013. An Operational Scheme for Deriving Standardised Surface Reflectance from Landsat TM/ETM+ and SPOT HRG Imagery for Eastern Australia. *Remote Sensing*, 5(1), pp.83–109.

Foody, G.M., 2003. Remote sensing of tropical forest environments: Towards the monitoring of environmental resources for sustainable development. *International Journal of Remote Sensing*, 24(20), pp.4035–4046.

Fotheringham, A., Charlton, M.E. & Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A*, 30(11), pp.1905–1927.

Fragal, E.H., Silva, T.S.F. & Novo, E.M.L. de M., 2016. Reconstructing historical forest cover change in the Lower Amazon floodplains using the LandTrendr algorithm. *Acta Amazonica*, 46(1), pp.13–24. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0044-59672016000100013&lng=en&nrm=iso&tlng=en.

de Freitas, M.W.D., Santos, J.R. dos & Alves, D.S., 2013. Land-use and land-cover change processes in the Upper Uruguay Basin: Linking environmental and socioeconomic variables. *Landscape Ecology*, 28(2), pp.311–327.

De Fries, R.S. et al., 1998. Global land cover classifications at 8 km spatial resolution: The use of training data derived from Landsat imagery in decision tree classifiers. *International Journal of Remote Sensing*, 19(16), pp.3141–3168.

Gao, J. & Li, S., 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography*, 31(1), pp.292–302. Available at: http://dx.doi.org/10.1016/j.apgeog.2010.06.003.

Garibay, O.O., Garibay, I.L. & Wu, A.S., 2003. The modular genetic algorithm: exploiting reguarities in the problem space. In *Computer and Information Sciences - ISCIS 2003*. pp. 584–591.

GDAL Development Team, 2017. GDAL - Geospatial Data Abstraction Library, Version 2.0.1. Available at: http://www.gdal.org.

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

Gibbs, M.S., Dandy, G.C. & Maier, H.R., 2008. A Genetic Algorithm Calibration Method Based on Convergence due to Genetic Drift. *Information Sciences*, 178(14), pp.2857–2869.

Gollini, I. et al., 2013. GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. Available at: http://arxiv.org/abs/1306.0413.

Goodwin, N.R. & Collett, L.J., 2014. Development of an automated method for mapping fire history captured in Landsat TM and ETM+ time series across Queensland, Australia. *Remote Sensing of Environment*, 148, pp.206–221.

Gorelick, N. et al., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. Available at: https://doi.org/10.1016/j.rse.2017.06.031.

Grainger, A., 2008. Difficulties in tracking the long-term global trend in tropical forest area. *Proceedings of the National Academy of Sciences*, 105(2), pp.818–823. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0703015105.

Grau, H.R. & Aide, M., 2008. Globalization and land-use transitions in Latin America. *Ecology and Society*, 13(2).

Gray, C.L. et al., 2008. Indigenous Land Use in the Ecuadorian Amazon: A Cross-cultural and Multilevel Analysis. *Human Ecology*, 36(1), pp.97–109. Available at: http://www.jstor.org/stable/27654258.

Gray, C.L. & Bilsborrow, R.E., 2014. Consequences of out-migration for land use in rural Ecuador. *Land Use Policy*, 36, pp.182–191.

Greenberg, J.A. et al., 2005. Survival analysis of a neotropical rainforest using multitemporal satellite imagery. *Remote Sensing of Environment*, 96(2), pp.202–211.

Griffiths, P. et al., 2013. A pixel-based landsat compositing algorithm for large area land cover mapping (IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(5), pp.2088–2101.

Griffiths, P. et al., 2014. Forest disturbances, forest recovery, and changes in forest types across the carpathian ecoregion from 1985 to 2010 based on landsat image composites. *Remote Sensing of Environment*, 151, pp.72–88.

Guariguata, M.R. & Ostertag, R., 2001. Neotropical secondary forest succession:

Changes in structural and functional characteristics. *Forest Ecology and Management*, 148(1–3), pp.185–206.

Haines, A.L., Mills, K.L. & Filliben, J.J., 2014. Determining Relative Importance and Best Settings for Genetic Algorithm Control Parameters. *Evolutionary computation*, pp.1–34.

El Hajj, M. et al., 2008. Relative Radiometric Normalization and Atmospheric Correction of a SPOT 5 Time Series. *Sensors*, 8, pp.2775–2791.

Hansen, M.C. et al., 2008. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment*, 112(5), pp.2495–2513.

Hansen, M.C. et al., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, 342(6160), pp.850–853. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1244693.

Hansen, M.C. & Loveland, T., 2011. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Enviroment*, 122, pp.66–74.

Hansen, M.C. & Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122, pp.66–74. Available at: http://dx.doi.org/10.1016/j.rse.2011.08.024.

Hermosilla, T. et al., 2015a. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. *Remote Sensing of Environment*, 158, pp.220–234.

Hermosilla, T. et al., 2015b. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. *Remote Sensing of Environment*, 170, pp.121–132. Available at: http://dx.doi.org/10.1016/j.rse.2015.09.004.

Hijmans, R. et al., 2017. raster: Geographic Data Analysis and Modeling, Version 2.6-7. Available at: http://www.rspatial.org/.

Hijmans, R., 2016. raster: Geographic Data Analysis and Modeling. Available at: https://cran.r-project.org/package=raster.

Hoerl, A.E., Kannard, R.W. & Baldwin, K.F., 1975. Ridge regression:some simulations.

Communications in Statistics, 4(2), pp.105–123. Available at: http://www.tandfonline.com/doi/abs/10.1080/03610927508827232.

Holland, J.H., 1973. Genetic Algorithms and the Optimal Allocation of Trials. *SIAM J. Comput.*, 2(2), pp.88–105.

Holt, D.T. et al., 1996. Aggregation and Ecological Effects in Geographically Based Data. *Geographical Analysis*, 28(3), pp.244–261.

Hoorn, C. et al., 2010. Amazonia Through Time : Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science*, 330(November), pp.927–931.

Huang, C. et al., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, 114, pp.183–198.

Huang, X. et al., 2015. Automatic labelling and selection of training samples for high-resolution remote sensing image classification over urban areas. *Remote Sensing*, 7(12), pp.16024–16044.

Hutchison, H.C. & Vallejo, I., 2016. *La deforestación y la participación de mujeres en el manejo de recursos naturales: una comparación de casos de estudio de comunidades indígenas y colonas en la provincia de Napo, Ecuador*. Facultad Latinoamericana de Ciencias Sociales (FLACSO).

Huttel, C., Zebrowski, C. & Gondard, P., 1999. *Paisajes Agrarios del Ecuador*, IGM.

INEC, 2001. Censo de Población y Vivienda 2001. Available at: http://www.inec.gob.ec/estadisticas/ [Accessed March 21, 2018].

INEC, 2010. Censo de Población y Vivienda 2010. Available at: http://www.inec.gob.ec/estadisticas/ [Accessed March 21, 2018].

Iovine, G. et al., 2005. Applying genetic algorithms for calibrating a hexagonal cellular automata model for the simulation of debris flows characterised by strong inertial effects. *Geomorphology*, 66, pp.287–303.

Jackson, B. et al., 2005. An Algorithm for Optimal Partitioning of Data on an Interval. *IEEE Signal Processing Letters*, 12(2), pp.105–108.

James, N.A., Arun, K. & David, S.M., 2014. BreakoutDetection: Breakout Detection via Robust E-Statistics. Version 1.0.1. Available at:

https://github.com/twitter/BreakoutDetection.

James, N.A. & Matteson, D.S., 2015. ecp: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *Journal of Statistical Software*, 62(7).

Joseph, S., Murthy, M.S.R. & Thomas, A.P., 2011. The progress on remote sensing technology in identifying tropical forest degradation: a synthesis of the present knowledge and future perspectives. *Environmental Earth Sciences*, 64(3), pp.731–741. Available at: https://doi.org/10.1007/s12665-010-0893-8.

Karatzoglou, A. et al., 2004. kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), pp.1–20. Available at: http://statistik.wu-wien.ac.at/%5Cnhttp://www.jstatsoft.org/v11/i09/paper.

Karmalkar, A. V., Bradley, R.S. & Diaz, H.F., 2008. Climate change scenario for Costa Rican montane forests. *Geophysical Research Letters*, 35(11), pp.1–5.

Karnieli, A. et al., 2001. AFRI - Aerosol free vegetation index. *Remote Sensing of Environment*, 77(1), pp.10–21.

Kennedy, R.E. et al., 2015. Attribution of disturbance change agent from Landsat time-series in support of habitat monitoring in the Puget Sound region, USA. *Remote Sensing of Environment*, 166, pp.271–285. Available at: http://dx.doi.org/10.1016/j.rse.2015.05.005.

Kennedy, R.E., Yang, Z. & Cohen, W.B., 2010a. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12), pp.2897–2910. Available at: http://dx.doi.org/10.1016/j.rse.2010.07.008.

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

Key, C.H. & Benson, N.C., 2006. *Landscape Assessment: Ground measure of severity, the Composite Burn Index; and Remote sensing of severity, the Normalized Burn Ratio*, Ogden, UT: USDA Forest Service, Rocky Mountain Research Station.

Key, C.H. & Benson, N.C., 1999. The Normalized Burn Ratio, a Landsat TM radiometric index of burn severity incorporating multi-temporal differencing USGS, ed.

Available at: http://nrmsc.usgs.gov/research/cbi.htm.

Killick, R. & Eckley, I.A., 2014. changepoint: An R Package for Changepoint Analysis. *JSS Journal of Statistical Software*, 58(3), pp.1–19.

Kimes, D.S. et al., 1998. Accuracies in mapping secondary tropical forest age from sequential satellite imagery. *Remote Sensing of Environment*, 65(1), pp.112–120.

Kissinger, G., Herold, M. & De Sy, V., 2012. Drivers of Deforestation and Forest Degradation. *A synthesis report for REDD+ Policymakers*, p.48. Available at: http://www.era-mx.org/biblio/Drivers of deforestation and forest degradation.pdf.

De Koning, F. et al., 2011. Bridging the gap between forest conservation and poverty alleviation: The Ecuadorian Socio Bosque program. *Environmental Science and Policy*, 14(5), pp.531–542.

Krishna Bahadur, K.C., 2009. Improving landsat and irs image classification: Evaluation of unsupervised and supervised classification through band ratios and dem in a mountainous landscape in Nepal. *Remote Sensing*, 1(4), pp.1257–1272.

Krivoruchko, K., Gribov, A. & Krause, E., 2011. Multivariate areal interpolation for continuous and count daata. *Procedia Environmental Sciences*, 3, pp.14–19. Available at: http://dx.doi.org/10.1016/j.proenv.2011.02.004.

Kuemmerle, T. et al., 2013. Challenges and opportunities in mapping land use intensity globally. *Current Opinion in Environmental Sustainability*, 5(5), pp.484–493.

Kuhn, M., 2016. caret: Classification and Regression Training, Version 6.0-71. Available at: https://cran.r-project.org/package=caret.

Kumar, R. & Jyotishree, 2012. Blending Roulette Wheel Selection & Rank Selection in Genetic Algorithms. *International Journal of Machine Learning and Computing*, 2, pp.365–370.

Laurance, W.F., Goosem, M. & Laurance, S.G.W., 2009. Impacts of roads and linear clearings on tropical forests. *Trends in Ecology and Evolution*, 24(12), pp.659–669.

Li, L. et al., 2015. Sub-pixel flood inundation mapping from multispectral remotely sensed images based on discrete particle swarm optimization. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, pp.10–21.

Li, P., Feng, Z. & Xiao, C., 2018. Acquisition probability differences in cloud coverage of the available Landsat observations over mainland Southeast Asia from 1986 to 2015. *International Journal of Digital Earth*, 11(5), pp.437–450. Available at: https://doi.org/10.1080/17538947.2017.1327619.

Liu, Y., Nishiyama, S. & Yano, T., 2004. Analysis of four change detection algorithms in bi-temporal space with a case study. *International Journal of Remote Sensing*, 25(11), pp.2121–2139.

Lo, C.P. & Yang, X., 2002. Drivers of Land-Use / Land-Cover Changes and Dynamic Modeling for the Atlanta , Georgia Metropolitan Area. *October*, 68(10), pp.1073–1082.

Logan, J.R., Xu, Z. & Stults, B., 2014. 1970 to 2010 : A Longtitudinal Tract Database. *Professional Geography*, 66(3), pp.412–420.

Lunetta, R. et al., 2004. Impacts of imagery temporal frequency on land-cover change detection monitoring. *Remote Sensing of Environment Environment*, pp.444–454.

Maaranen, H., Miettinen, K. & Penttinen, A., 2007. On initial populations of a genetic algorithm for continuous optimization problems. *Journal of Global Optimization*, 37(3), pp.405–436.

Macdonald, T., 1981. *Indigenous Responses to an Expand- ing Frontier: Jungle Quichua Economic Conversion to Cattle Ranching*, Urbana: University of Illinois Press.

MAE, 2017. Documentation of the information used for the establishment of Ecuador’s Forest Reference Emission Level. Available at: http://suia.ambiente.gob.ec/web/suia/anexos-nivel-referencia [Accessed August 16, 2017].

MAE, 2012a. *Línea Base de Deforestación del Ecuador Continental*, Quito - Ecuador: Ministerio del Ambiente (MAE).

MAE, 2013. *Metodología para la representación Cartográfica de los Ecosistemas del Ecuador Continental*, Quito - Ecuador: Ministerio del Ambiente del Ecuador (MAE).

MAE, 2012b. *Sistema de clasificación de los ecosistemas del Ecuador Continental*, Quito - Ecuador: Ministerio del Ambiente (MAE).

Mahiny, A.S. & Turner, B.J., 2007. A comparison of four common atmospheric correction methods. *Photogrammetric Engineering and Remote Sensing*, 73(4), pp.361–368.

Mas, J.-F. et al., 2013. Assessing Local Variations of Deforestation Processes in Mexico Using Geographically Weighted Regression. *International Cartographic Conference*.

Masek, J.G. et al., 2012. LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code, Version 2. Model product.

Matricardi, E.A.T. et al., 2010. Assessment of tropical forest degradation by selective logging and fire using Landsat imagery. *Remote Sensing of Environment*, 114(5), pp.1117–1129.

Matricardi, E.A.T. et al., 2007. Multi-temporal assessment of selective logging in the Brazilian Amazon using Landsat data. *International Journal of Remote Sensing*, 28(1), pp.63–82.

Mellor, A. et al., 2013. The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sensing*, 5(6), pp.2838–2856.

Mena, C.F., Bilsborrow, R.E. & McClain, M.E., 2006. Socioeconomic drivers of deforestation in the Northern Ecuadorian Amazon. *Environmental Management*, 37(6), pp.802–815.

Mennis, J., 2003. Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer*, 55(1), pp.31–42.

Mennis, J. & Hultgren, T., 2006. Intelligent Dasymetric Mapping and Its Application to Areal Interpolation. *Cartography and Geographic Information Science*, 33(3), pp.179–194. Available at: http://www.tandfonline.com/doi/abs/10.1559/152304006779077309.

Mertens, B. et al., 2000. Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. *World Development*, 28(6), pp.983–999.

Meyfroidt, P. et al., 2013. Globalization of land use: Distant drivers of land change and geographic displacement of land use. *Current Opinion in Environmental*

Sustainability, 5(5), pp.438–444. Available at: http://dx.doi.org/10.1016/j.cosust.2013.04.003.

Micijevic, E., Obaidul, M.H. & Mishra, N., 2017. Radiometric characterization of Landsat Collection 1 products. *Proc.SPIE*, 10402, pp.10402-10402–8. Available at: http://dx.doi.org/10.1117/12.2276065.

Miller, B. & Goldberg, D., 1995. Genetic Algorithms, Tournament Selection, and the Effects of Noise. *Complex Systems*, 9, pp.193–212.

Mon, M.S. et al., 2012. Factors affecting deforestation and forest degradation in selectively logged production forest: A case study in Myanmar. *Forest Ecology and Management*, 267, pp.190–198. Available at: http://www.sciencedirect.com/science/article/pii/S0378112711007213.

Montana, D. & Davis, L., 1989. Training feedforward neural networks using genetic algorithms. In *11th International Joint Conference on Artificial Intelligence*. pp. 762–767.

Moran, E.F., Siqueira, A. & Brondizio, E., 2003. Household Demographic Structure and Its Relationship to Deforestation in the Amazon Basin. In J. Fox et al., eds. *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Boston, MA: Springer US, pp. 61–89. Available at: https://doi.org/10.1007/0-306-48130-8_3.

Moritz, M.S., 2016. Package ‘ imputeTS .’

Mosandl, R. et al., 2008. Ecuador Suffers the Highest Deforestation Rate in South America. In E. Beck et al., eds. *Gradients in a Tropical Mountain Ecosystem of Ecuador*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 37–40. Available at: https://doi.org/10.1007/978-3-540-73526-7_4.

Mountrakis, G., Im, J. & Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, pp.247–259.

Müller, H., Griffiths, P. & Hostert, P., 2016. Long-term deforestation dynamics in the Brazilian Amazon — Uncovering historic frontier development along the Cuiabá – Santarém highway. *International Journal of Applied Earth Observations and Geoinformation*, 44, pp.61–69. Available at:

http://dx.doi.org/10.1016/j.jag.2015.07.005.

Muratorio, B., 1998. *Rucuyaya Alonso y la historia social y económica del Alto Napo 1850-1950* Abya-Yal., Quito - Ecuador.

Myers, N. et al., 2000. Biodiversity hotspots for conservation priorities. *Nature*, 403, pp.853–858.

Myers, N., 1993. Tropical Forests: The Main Deforestation Fronts. *Environmental Conservation*, 20(1), pp.9–16. Available at: https://www.cambridge.org/core/article/tropical-forests-the-main-deforestation-fronts/0953F8D81149DA3940E346BD08BC0DAD.

Myers, R., 1990. *Classical and Modern Regression with Applications*, PWS-KENT. Available at: https://books.google.de/books?id=X76TGwAACAAJ.

Myneni, R.B. & Asrar, G., 1994. Atmospheric effects and spectral vegetation indices. *Remote Sensing of Environment*, 47(3), pp.390–402.

Nagendra, H., 2007. Drivers of reforestation in human-dominated forests. *Proceedings of the National Academy of Sciences*, 104(39), pp.15218–15223. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.0702319104.

NASA, 2011. *Landsat 7 Science Data Users Handbook*, National Aeronautics and Space Administration (NASA). Available at: https://landsat.gsfc.nasa.gov.

National Institute of Advanced Industrial Science and Technology & Geological Survey Japan, 2017. MADAS: METI AIST satellite Data Archive System. Available at: https://gbank.gsj.jp/ [Accessed August 16, 2017].

Oeser, J. et al., 2017. Using intra-annual Landsat time series for attributing forest disturbance agents in Central Europe. *Forests*, 8(7).

Oliveira, R.S. et al., 2014. The hydroclimatic and ecophysiological basis of cloud forest distributions under current and projected climates. *Annals of Botany*, 113(6), pp.909–920.

Oliveras, I., Anderson, L.O. & Malhi, Y.S., 2014. Application of remote sensing to understanding fire regime and biomass burning emissions of the tropical Andes. *Global Biogeochemical Cycles*, 28, pp.480–496.

Olofsson, P. et al., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, pp.42–57. Available at: http://dx.doi.org/10.1016/j.rse.2014.02.015.

Olson, D.M. et al., 2001. Terrestrial Ecoregions of the World : A New Map of Life on Earth. , 51(11), pp.933–938.

Open Foris, 2015. System for earth observations, data access, processing & analysis for land monitoring (SEPAL). Available at: https://sepal.io/.

Páez, A., Farber, S. & Wheeler, D., 2011. A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning A*, 43(12), pp.2992–3010.

Pan, W. et al., 2004. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. *Agriculture, Ecosystems and Environment*, 101(2–3), pp.117–134.

Pan, W. et al., 2007. Forest clearing in the Ecuadorian Amazon: A study of patterns over space and time. *Population Research and Policy Review*, 26(5–6), pp.635–659.

Paula, F.S. et al., 2014. Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Molecular Ecology*, 23(12), pp.2988–2999.

Paulick, S. et al., 2017. The carbon fluxes in different successional stages: modelling the dynamics of tropical montane forests in South Ecuador. *Forest Ecosystems*, 4(1), p.5. Available at: http://forestecosyst.springeropen.com/articles/10.1186/s40663-017-0092-0.

Pebesma, E. et al., 2017. sp: Classes and Methods for Spatial Data, Version 1.2-5. Available at: https://github.com/edzer/sp/ https://edzer.github.io/sp/.

Perreault, T., 2001. Developing Identities : Indigenous Mobilization , Rural Livelihoods , and Resource Access in Ecuadorian Amazonia. *Ecumene*, 8(4).

Petrov, A., 2012. One Hundred Years of Dasymetric Mapping: Back to the Origin. *The Cartographic Journal*, 49(3), pp.256–264. Available at: http://www.tandfonline.com/doi/full/10.1179/1743277412Y.0000000001.

Pflugmacher, D., Cohen, W.B. & Kennedy, R.E., 2012. Using Landsat-derived disturbance history (1972-2010) to predict current forest structure. *Remote Sensing of Environment*, 122, pp.146–165.

Pierre, G. et al., 1988. *Transformaciones Agrarias En El Ecuador* L. Juan & P. Pierre, eds., Quito - Ecuador: IGM.

Pimple, U. et al., 2017. Topographic correction of Landsat TM-5 and Landsat OLI-8 imagery to improve the performance of forest classification in the mountainous terrain of Northeast Thailand. *Sustainability (Switzerland)*, 9(2), pp.1–26.

Pineda Jaimes, N.B. et al., 2010. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Applied Geography*, 30(4), pp.576–591.

Pinty, B. & Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetation*, 101, pp.15–20.

Da Ponte, E. et al., 2015. Tropical forest cover dynamics for Latin America using Earth observation data: a review covering the continental, regional, and local scale. *International Journal of Remote Sensing*, 36(12), pp.3196–3242. Available at: http://dx.doi.org/10.1080/01431161.2015.1058539.

Poorter, L. et al., 2016. Biomass resilience of Neotropical secondary forests. *Nature*, 530(7589), pp.211–214. Available at: http://www.nature.com/doifinder/10.1038/nature16512.

Potapov, P., Turubanova, S. & Hansen, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sensing of Environment*, 115(2), pp.548–561. Available at: http://dx.doi.org/10.1016/j.rse.2010.10.001.

Potapov, P. V. et al., 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000-2010, with Landsat ETM+ data. *Remote Sensing of Environment*, 122, pp.106–116. Available at: http://dx.doi.org/10.1016/j.rse.2011.08.027.

Powell, S.L. et al., 2010. Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, 114, pp.1053–1068.

Pueschel, P., Buddenbaum, H. & Hill, J., 2012. An efficient approach to standardizing the processing of hemispherical images for the estimation of forest structural attributes. *Agricultural and Forest Meteorology*, 160, pp.1–13. Available at: http://dx.doi.org/10.1016/j.agrformet.2012.02.007.

Puyravaud, J.P., 2003. Standardizing the calculation of the annual rate of deforestation. *Forest Ecology and Management*, 177(1–3), pp.593–596.

Qi, J. et al., 1994. A Modified Soil Adjusted Vegetation Index. *Remote Sensing of Environment*, 48, pp.119–126.

R Development Core Team, 2017. The R Project for Statistical Computing, Version 3.4.3. Available at: http://www.r-project.org/.

RAISG, 2015. *Deforestation in the Amazonia (1970-2013)*, Available at: www.raisg.socioambiental.org.

Ramírez, B. et al., 2017. Tropical Montane Cloud Forests: Hydrometeorological variability in three neighbouring catchments with different forest cover. *Journal of Hydrology*, 552, pp.151–167.

Reibel, M. & Agrawal, A., 2007. Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26(5–6), pp.619–633.

Revolution Analytics & Weston, S., 2015. foreach: Provides Foreach Looping Construct for R, Version 1.4.3. Available at: https://cran.r-project.org/package=foreach.

Rey Benayas, J., 2007. Abandonment of agricultural land: an overview of drivers and consequences. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 2(057). Available at: http://www.cabi.org/cabreviews/review/20073206799.

Riaño, D. et al., 2003. Assessment of different topographic corrections in landsat-TM data for mapping vegetation types (2003). *IEEE Transactions on Geoscience and Remote Sensing*, 41(5 PART 1), pp.1056–1061.

Richter, R., Kellenberger, T. & Kaufmann, H., 2009. Comparison of topographic correction methods. *Remote Sensing*, 1(3), pp.184–196.

Rouse, J.., Haas, R.H., et al., 1974. Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS Symposium*, SP-351 I, pp.309–317.

Rouse, J.., Haas, R.H., et al., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. In *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium*. pp. 48–62.

Roy, D.P. et al., 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, pp.154–172. Available at: http://dx.doi.org/10.1016/j.rse.2014.02.001.

Roy, D.P. et al., 2010. Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment*, 114, pp.35–49.

Rudel, T.K. et al., 2005. Forest transitions: Towards a global understanding of land use change. *Global Environmental Change*, 15(1), pp.23–31.

Rudel, T.K., Bates, D. & Machinguiashi, R., 2002. A tropical forest transition? Agricultural change, out-migration, and secondary forests in the Ecuadorian Amazon. *Annals of the Association of American Geographers*, 92(1), pp.87–102.

Rufin, P. et al., 2015. Land use intensity trajectories on Amazonian pastures derived from Landsat time series. *International Journal of Applied Earth Observation and Geoinformation*, 41, pp.1–10. Available at: http://dx.doi.org/10.1016/j.jag.2015.04.010.

Sachs, J., 2001. Tropical Underdevelopment. *NBER Working Paper*, 8119, pp.1–31.

Salmivaara, A. et al., 2015. Exploring the modifiable areal unit problem in spatial water assessments: A case of water shortage in Monsoon Asia. *Water (Switzerland)*, 7(3), pp.898–917.

Salvini, G. et al., 2014. How countries link REDD+ interventions to drivers in their readiness plans: implications for monitoring systems. *Environmental Research Letters*, 9(7), p.074004. Available at: http://stacks.iop.org/1748-9326/9/i=7/a=074004?key=crossref.919d82da0ee8ec3572acc36bf95c95d1.

Samndong, R.A. et al., 2018. Institutional analysis of causes of deforestation in REDD+ pilot sites in the Equateur province: Implication for REDD+ in the Democratic Republic of Congo. *Land Use Policy*. Available at: http://www.sciencedirect.com/science/article/pii/S0264837717305963.

Santos, F., Dubovyk, O. & Menz, G., 2017. Monitoring forest dynamics in the andean amazon: The applicability of breakpoint detection methods using landsat time-series and genetic algorithms. *Remote Sensing*, 9(1).

Santos, F., Meneses, P. & Hostert, P., 2018. Monitoring long-term forest dynamics with scarce data: a multi-date classification implementation in the Ecuadorian Amazon. *Manuscript accepted for publication*.

Savitzky, A. & Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8), pp.1627–1638.

Scrucca, L., 2013. GA: A Package for Genetic Algorithms in R. *Journal of Statistical Software*, 53(4), pp.1–37.

Semenov-Tian-Shansky, 1923. Metody dazimetrii (Methods of Dasymetric Mapping). In *Dazimetrichskaya Karta Evropeiskoi*. Petrograd: Scientific Chemistry and Technology Publishing, pp. 18–26.

Senf, C. et al., 2015. Characterizing spectral-temporal patterns of defoliator and bark beetle disturbances using Landsat time series. *Remote Sensing of Environment*, 170, pp.166–177.

Shimizu, K. et al., 2018. Assessments of preprocessing methods for Landsat time series images of mountainous forests in the tropics. *Journal of Forest Research*, 0(0), pp.1–10. Available at: https://doi.org/10.1080/13416979.2018.1434034.

Sierra, R., 2000. Dynamics and patterns of deforestation in the western Amazon: The Napo deforestation front, 1986-1996. *Applied Geography*, 20(1), pp.1–16.

Sierra, R., Campos, F. & Chamberlin, J., 2002. Assessing biodiversity conservation priorities: Ecosystem risk and representativeness in continental Ecuador. *Landscape and Urban Planning*, 59(2), pp.95–110.

SIGTIERRAS, U.-M.-P., 2015. Metodología Accesibilidad. Proyecto: “Levantamiento de cartografía temática escala 1:25000, lote 2”. Temáticas Nacionales. Available at: http://metadatos.sigtierras.gob.ec:8080/geonetwork/srv/spa/catalog.search#/search?resultType=details&any=accesibilidad&from=1&to=20&sortBy=relevance.

Silman, M.R., Ancaya, E.J. & Brinson, J., 2003. Los Bosques de Bambú en la Amazonía Occidental. In R. Leite, N. Pitman, & P. Álvarez, eds. *Alto Purús: Biodiversidad,* *Conservación y Manejo*. pp. 62–73.

Simoes, A. & Hidalgo, C., 2011. The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*. Available at: https://atlas.media.mit.edu/en/.

SNI, 2017. Secreataría Nacional de Planificación y Desarrollo. Sistema Nacional de Información (SNI): Archivos de Información Geográfica. Available at: http://sni.gob.ec/coberturas [Accessed December 20, 2017].

Southgate, D., Sierra, R. & Brown, L., 1991. The causes of tropical deforestation in Ecuador: A statistical analysis. *World Development*, 19(9), pp.1145–1151.

Spracklen, D. V. & Righelato, R., 2014. Tropical montane forests are a larger than expected global carbon store. *Biogeosciences*, 11(10), pp.2741–2754.

Steinhaus, H., 1957. Sur la division des corps matériels en parties. *Bull. Acad. Pol. Sci., Cl. III*, 4, pp.801–804.

Stephen J. Walsh et al., 2008. Integration of Hyperion Satellite Data and Household Social Survey to Characterize the Causes and Consequences of Reforestation Patterns in the Northern Ecuadorian Amazon. *Photogrammetric Engineering & Remote Sensing*, 74(6), p.725–735.

Stevens, F.R. et al., 2015. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*, 10(2), pp.1–22.

Su, S., Xiao, R. & Zhang, Y., 2012. Multi-scale analysis of spatially varying relationships between agricultural landscape patterns and urbanization using geographically weighted regression. *Applied Geography*, 32(2), pp.360–375. Available at: http://dx.doi.org/10.1016/j.apgeog.2011.06.005.

Tarek A. El-Mihoub et al., 2006. Hybrid Genetic Algorithms: A Review. *Engineering Letters*, 13(2).

Teillet, P.M. et al., 1982. On the Slope-Aspect Correction of Multispectral Scanner Data. *Canadian Journal of Remote Sensing*, 8, pp.84–106.

Thomas, N.E. et al., 2011. Validation of North American Forest Disturbance dynamics derived from Landsat time series stacks. *Remote Sensing of Environment*, 115(1), pp.19–32. Available at: http://dx.doi.org/10.1016/j.rse.2010.07.009.

Tobler, W.R., 1979. Smooth pycnopylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367), pp.519–530.

Torres, B. et al., 2015. The Contribution of Traditional Agroforestry to Climate Change Adaptation in the Ecuadorian Amazon: The Chakra System. In W. Leal Filho, ed. *Handbook of Climate Change Adaptation*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1973–1994. Available at: https://doi.org/10.1007/978-3-642-38670-1_102.

Tu, J., 2011. Spatially varying relationships between land use and water quality across an urbanization gradient explored by geographically weighted regression. *Applied Geography*, 31(1), pp.376–392. Available at: http://dx.doi.org/10.1016/j.apgeog.2010.08.001.

UN-REDD Programme, 2015. Technical considerations for Forest Reference Emission Level and / or Forest Reference Level construction for REDD+ under the UNFCCC. , p.31.

UNESCO, 2010. Atlas pluviométrico del Ecuador J. Cedeño & M. C. Donoso, eds. *PHI-LAC*, 21.

United Nations, 2015. Transforming our world: the 2030 Agenda for Sustainable Development. *General Assembley 70 session*, 16301(October), pp.1–35.

United Nations, 2012. UN Comtrade Database. Available at: https://comtrade.un.org/ [Accessed May 3, 2018].

USGS, 2014. Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) On Demand Interface. Available at: https://espa.cr.usgs.gov [Accessed January 14, 2017].

Vanonckelen, S., Lhermitte, S. & Rompaey, A. Van, 2015. The effect of atmospheric and topographic correction on pixel-basedimage composites: Improved forest cover detection in mountainenvironments. *International Journal of Applied Earth Observation and Geoinformation*, 35(PB), pp.320–328. Available at: http://dx.doi.org/10.1016/j.jag.2014.10.006.

Verbesselt, J. et al., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1), pp.106–115. Available at: http://dx.doi.org/10.1016/j.rse.2009.08.014.

Vermote, E.F. et al., 1997. Second Simulation of the Satellite Signal in the Solar Spectrum (6S). 6S User Guide Version 2. Appendix III: Description of the subroutines. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), pp.675–686.

Villamor, G.B. et al., 2015. Land use change and shifts in gender roles in central Sumatra , Indonesia. *International for*, 17(1), pp.1–15.

Wang, Q., Wang, L. & Liu, D., 2012. Particle swarm optimization-based sub-pixel mapping for remote-sensing imagery. *International Journal of Remote Sensing*, 33(20), pp.6480–6496.

Wasserstrom, R. & Southgate, D., 2013. Deforestation, Agrarian Reform and Oil Development in Ecuador, 1964-1994. *Natural Resources*, 04(01), pp.31–44. Available at: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/nr.2013.41004.

Wheeler, D., 2007. Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A*, 39(10), pp.2464–2481.

Wheeler, D.C. & Calder, C.A., 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems*, 9(2), pp.145–166.

Wickham, H. & Chang, W., 2016. Package ‘ggplot2’, version 2.2.1. Available at: https://ggplot2.tidyverse.org/.

Wieland, M. et al., 2016. Object-based urban structure type pattern recognition from Landsat TM with a Support Vector Machine. *International Journal of Remote Sensing*, 37(17), pp.4059–4083. Available at: http://dx.doi.org/10.1080/01431161.2016.1207261.

Wilson, M.F.J. et al., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30(1–2), pp.3–35.

Wood, C.H. & Skole, D., 1998. Linking Satellite, Census, and Survey Data to Study

Deforestation in the Brazilian Amazon. In D. Liverman et al., eds. *People and Pixels. Linking Remote Sensing and Social Science*. Washington D.C. - EEUU, pp. 70–93.

Wulder, M.A. et al., 2011. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122, pp.2–10.

Wulder, M.A. et al., 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, 185, pp.271–283. Available at: http://dx.doi.org/10.1016/j.rse.2015.11.032.

Wulder, M.A. et al., 2015. Virtual constellations for global terrestrial monitoring. *Remote Sensing of Environment*, 170, pp.62–76.

Ye, S., Rogan, J. & Sangermano, F., 2018. Monitoring rubber plantation expansion using Landsat data time series and a Shapelet-based approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 136, pp.134–143. Available at: http://www.sciencedirect.com/science/article/pii/S0924271618300029.

Zhang, M. et al., 1999. Noise reduction and atmospheric correction for coastal applications of Landsat Thematic Mapper imagery. *Remote Sensing of Environment*, 70(2), pp.167–180.

Zhu, Z., 2017. Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, pp.370–384. Available at: http://linkinghub.elsevier.com/retrieve/pii/S092427161730103X.

Zhu, Z. et al., 2016. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, pp.206–221. Available at: http://dx.doi.org/10.1016/j.isprsjprs.2016.11.004.

Zhu, Z. & Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, 144, pp.152–171. Available at: http://dx.doi.org/10.1016/j.rse.2014.01.011.

Zhu, Z. & Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in

Landsat imagery. *Remote Sensing of Environment*, 118, pp.83–94. Available at: http://dx.doi.org/10.1016/j.rse.2011.10.028.

SUPLEMENTARY MATERIAL

Annex 1. Descriptive statistics of variable groups obtained for reforestation cells

Variable groups	Prefix	Variable	Cells statistics				Source
			Min.	Mean	Max.	SD	
Accessibility [hours]	A1	Accessibility to oil palm extraction facilities[1]	0-0.08	1-3	>3	-	(SIGTIERRAS 2015)
	A2	Accessibility to coffee and cacao collection centers[1]	0-0.08	0.5-1	>3	-	
	A3	Accessibility to fruit collection centers[1]	0-0.08	1-3	>3	-	
	A4	Accessibility to milk products collection centers[1]	0-0.08	0.5-1	>3	-	
Age [people]	D1	Younger population (age 15 - 25)	-27	0.9	123	5	(INEC 2001; INEC 2010)
	D2	Adult population (age 26 - 45)	-20	1.2	134	6	
	D3	Older adult population (age 45 - 72)	-19	0.7	48	3	
Biophysical [m.a.s.l.][2] [unitless] [mm]	B1	Altitude	250	662	3177	478	(SNI 2017)
	B2	Soil fertility (>2% organic matter at max. value)[1]	1	3.2	4	0.4	
	B3	Annual rainfall	1372	3647	5892	647	
Education [people]	E1	Basic education (1 - 6 years)	-64	-0.8	53	3	(INEC 2001; INEC 2010)
	E2	Secondary education (7 - 12 years)	-29	2.3	161	9	
	E3	Higher education (>13 years)	0	1.6	100	6	
Gender [people]	G1	Chief male household	-9	0.8	76	3	(INEC 2001; INEC 2010)
	G2	Chief female household	-18	0.2	56	1	
	G3	Male population	-33	1.4	136	8	

	G4	Female population	-34	1.3	161	7	
Household [people]	H1	Mothers with 1 - 2 children (small families)	-4	0.2	28	1	(INEC 2001; INEC 2010)
	H2	Mothers with 3 - 5 children (medium families)	-13	0.5	63	2	
	H3	Mothers with more than 5 children (large families)	-33	-0.4	26	2	
Language [people]	L1	Speak Spanish [3]	-28	3.4	242	14	(INEC 2001; INEC 2010)
	L2	Speak Kichwa [4]	-27	2.2	245	11	
	L3	Speak other languages [5]	-102	-0.7	22	4	
Proximity [meters]	P1	Distance to oil infrastructures	198	8972	47460	8051	(SNI 2017)
	P2	Distance to mining sites	167	7016	55977	7819	
	P3	Distance to paved and dirt roads	45	1749	28749	2534	
Work [people]	W1	Agriculture related workers	-36	0.5	52	4	(INEC 2001; INEC 2010)
	W2	Industry related workers	-17	0.09	34	1	
	W3	Services related workers	-12	0.17	26	1	

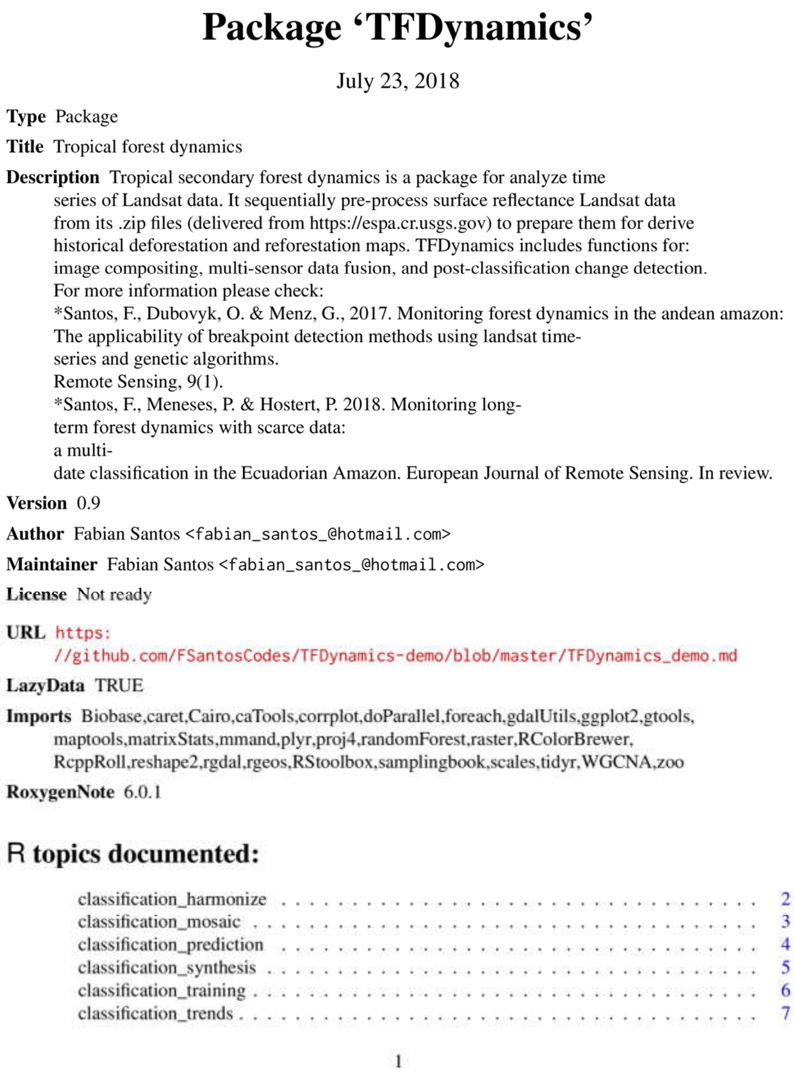
[1] Categorical variable ordered and recoded as continuous. [2] Stands for meters above sea level. [3] First spoken language by *colonos* in CEA. [4] Second spoken language and ethnicity at CEA. [5] Includes 12 different languages except Kichwa.

Annex 2. Descriptive statistics of regression coefficients obtained for deforestation and reforestation cells

Variable groups	**Prefix**	**Regression coefficients statistics**							
		Deforestation cells				**Reforestation cells**			
		Min.	**Mean**	**Max.**	**SD**	**Min.**	**Mean**	**Max.**	**SD**
Accessibility	A1	-1.34	-0.15	1.44	0.40	-0.38	0.21	0.92	0.32
	A2	-0.99	-0.26	0.32	0.28	-1.44	-0.51	0.06	0.37
	A3	-1.29	0.085	1.03	0.35	-0.30	0.28	1.55	0.37
	A4	-0.33	0.17	1.01	0.28	-1.16	-0.14	0.45	0.36
Age	D1	-1.63	-0.39	0.55	0.50	-4.02	-0.15	0.52	0.51
	D2	-0.34	0.24	1.32	0.40	-0.61	0.19	4.37	0.48
	D3	-0.47	0.56	2.04	0.59	-0.70	0.14	1.51	0.39

Biophysical	B1	-2.48	-0.09	0.29	0.40	-1.89	-0.36	0.31	0.35
	B2	-0.15	0.11	0.46	0.11	-0.51	0.042	0.23	0.12
	B3	-0.42	-0.051	0.53	0.13	-1.92	0.035	0.66	0.44
Education	E1	-0.32	-0.04	0.17	0.08	-0.60	-0.19	0.26	0.13
	E2	-0.84	0.43	1.25	0.45	-0.81	0.18	0.80	0.20
	E3	-0.92	0.09	3.43	0.37	-0.79	0.11	3.66	0.34
Gender	G1	-0.46	1.74	6.97	1.19	-0.63	0.69	2.98	0.56
	G2	-1.02	0.34	1.54	0.38	-0.28	0.085	1.18	0.24
	G3	-0.36	0.015	1.08	0.25	-2.25	-0.20	0.80	0.43
	G4	-8.64	-1.73	0.68	1.70	-3.51	-0.39	0.61	0.66
Household	H1	-0.33	0.23	2.78	0.44	-0.95	0.23	0.94	0.19
	H2	-1.40	0.20	1.52	0.41	-0.48	0.060	0.88	0.16
	H3	-0.97	-0.27	0.01	0.26	-0.88	-0.31	-0.07	0.14
Language	L1	-0.55	0.42	2.43	0.43	-0.31	0.17	1.06	0.21
	L2	-1.28	0.10	2.97	0.61	-0.49	0.020	0.57	0.19
	L3	-0.62	0.08	3.68	0.59	-1.29	-0.19	0.12	0.15
Proximity	P1	-0.24	0.05	0.38	0.08	-0.81	-0.062	0.35	0.17
	P2	-0.34	0.04	0.85	0.19	-0.84	-0.14	0.25	0.23
	P3	-1.03	-0.22	-0.003	0.22	-1.70	-0.24	-0.007	0.26
Work	W1	-0.18	0.18	0.58	0.13	-0.32	0.043	0.47	0.14
	W2	-1.51	0.02	0.61	0.27	-0.69	-0.050	1.25	0.18
	W3	-0.60	0.10	3.53	0.46	-1.26	0.14	0.72	0.25

Annex 3. Reference manual of TFDynamics Version 0.9



Package ‘TFDynamics’

July 23, 2018

Type Package

Title Tropical forest dynamics

Description Tropical secondary forest dynamics is a package for analyze time series of Landsat data. It sequentially pre-process surface reflectance Landsat data from its .zip files (delivered from https://espa.cr.usgs.gov) to prepare them for derive historical deforestation and reforestation maps. TFDynamics includes functions for: image compositing, multi-sensor data fusion, and post-classification change detection.
For more information please check:
*Santos, F., Dubovyk, O. & Menz, G., 2017. Monitoring forest dynamics in the andean amazon: The applicability of breakpoint detection methods using landsat time-series and genetic algorithms.
Remote Sensing, 9(1).
*Santos, F., Meneses, P. & Hostert, P. 2018. Monitoring long-term forest dynamics with scarce data:
a multi-date classification in the Ecuadorian Amazon. European Journal of Remote Sensing. In review.

Version 0.9

Author Fabian Santos <`fabian_santos_@hotmail.com`>

Maintainer Fabian Santos <`fabian_santos_@hotmail.com`>

License Not ready

URL `https://github.com/FSantosCodes/TFDynamics-demo/blob/master/TFDynamics_demo.md`

LazyData TRUE

Imports Biobase,caret,Cairo,caTools,corrplot,doParallel,foreach,gdalUtils,ggplot2,gtools, maptools,matrixStats,mmand,plyr,proj4,randomForest,raster,RColorBrewer, RcppRoll,reshape2,rgdal,rgeos,RStoolbox,samplingbook,scales,tidyr,WGCNA,zoo

RoxygenNote 6.0.1

R topics documented:

classification_harmonize 2
classification_mosaic 3
classification_prediction 4
classification_synthesis 5
classification_training 6
classification_trends 7

composites_match 8
composites_median 8
composites_modify 9
composites_percentiles 10
composites_RGB 11
composites_trends 11
images_clip 12
images_copyMove 13
images_delete 14
images_geometry 14
images_indices 15
images_mask 16
images_maskSum 16
images_metadata 17
images_organize 18
images_rename 18
images_RGB 19
images_trends 20
images_unzip 20
others_batchClip 21
others_metricsDEM 21
others_metricsDis 22
sampling_creation 22
sampling_extraction 23
sampling_fill 24
sampling_plots 25

Index 27

`classification_harmonize` *Harmonize RAW classifications*

Description

Improve classification outputs applying a temporal filtering. This function seeks time series patterns in classification files for recode them.

Usage

```
classification_harmonize(base_folder, crop_area = NA, pattern_recode,
  NA_fill = NA, median_filter = NA, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`crop_area`	String. If outputs need to be cropped, a shapefile filename should be introduced here

`pattern_recode`	Data frame. It should be contain two columns whose: 1) first column indicates the patterns to find in land cover maps (eg. "1,2,1"), and 2) second column indicates the values to replace for a specific pattern (eg. "1,1,1"). The number makes reference to the classes in land cover maps and its order the temporal pattern in found in time-series land cover maps. Patterns to find and replace should contain at least three values
`NA_fill`	String. For fill NA values to methods can be applied: 1) 'locf' where last observation carried forward is applied, and 2)'mode' which NAs are filled by mode value. Ignored if it is NA
`median_filter`	Numeric. An odd number for define a kernel size and apply a median filter. Ignored if it is NA
`chunksize`	Numeric. A number of cells to process at each iteration By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Post classification files stored at: '*base_folder*/ANALYSIS/CLASSIFICATION/FILTERED'

`classification_mosaic` *Mosaic '.envi' classification type maps or '.tif' rasters*

Description

Mosaic multiple '.envi' maps integrating their class names to corresponding pixel values. Colors are taken from metadata of the classification map with more classes (it is assumed that line 'class lookup' is located in the last metadata rows). If files are .tif,a simple mosaic is done. Files order indicates which layers goes up

Usage

```
classification_mosaic(output_folder, map_files, function_type = "merge",
  coordinate_system = NA, output_resolution = NA, median_filter = NA)
```

Arguments

`output_folder`	String. An output folder to store mosaics
`map_files`	String. A list of filename of maps in the '.envi' format. Their order indicates their priorization. Note that the legend applied in the mosaic is based in the map with more classes
`function_type`	String. Function to apply for mosaic. Can be: "merge", "mean", "max", "min"
`coordinate_system`	String. The coordinate system supported by gdal eg.("+proj=longlat +datum=WGS84"). If NA, no projection is made and coordinate system is taken from maps
`output_resolution`	Numeric. The output resolution for x and y in coordinate system units. If NA, spatial resolution is taken from maps
`median_filter`	Numeric. An odd number for define a kernel size and apply a median filter. Ignored if it is NA

Value

Mosaicked maps in the '.envi' classification type format

`classification_prediction` *Classify composites using algorithms and methods from 'caret' package*

Description

This function classifies specific or all composites dates according to the Random Forest model created. For that, if the model consider additional data, it should be provided as is defined in the model for run. For eliminate isolated pixels in outputs, a median filter can be specified. Additionally, an overlap classification mask is created for identify areas which have temporal gaps in the time series of classification outputs

Usage

```
classification_prediction(base_folder, model_name, crop_area = NA,
  specific_dates = NA, median_filter = NA, color_classes = NA,
  overlap_mask = T, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`model_name`	String or caret object. A filename specifing the *.Rdata caret model stored or the caret object to use
`crop_area`	String. String. If outputs need to be crop, a shapefile name should be introduced here
`specific_dates`	String. A list of composite dates for classify eg. c("1999-01-01","2005-12-22"). By default all dates are classified
`median_filter`	Numeric. An odd number for define a kernel size and apply a median filter. Ignored if it is NA
`color_classes`	String. For plots, a list of color names should be defined for each class sorted in an alphanumerical way. If colors are not defined, then a default pallete is applied
`overlap_mask`	Logical. If true, a overlap classification mask is created. This mask is the resultf of a sum of valid classified pixels
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Classification files stored at: '*base_folder*/ANALYSIS/CLASSIFICATION/RAW' and overlap classification mask at: '*base_folder*/ANALYSIS/MASK'"

`classification_synthesis` *Synthetize land cover maps into vegetation loss/regrowth and derive classes frequencies*

Description

This function synthesize land cover maps into: vegetation loss and regrowth maps. Optionally, it can calculate frequencies (in percentages) from land cover classes after change (in vegetation loss), before change (in vegetation regrowth) and for all the period (in stable classes).

Usage

```
classification_synthesis(base_folder, crop_area = NA,
  classification_folder = "FILTERED", specific_dates = NA,
  vegetation_classes, regrowth_observations = 3, calculate_frequencies = F,
  median_filter = NA, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA, restitute_files = F)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`crop_area`	String. If outputs need to be crop, a shapefile name should be introduced here
`classification_folder`	String. A folder name from the '/ANALYSIS/CLASSIFICATION/' outputs for change analysis. Two options can be possible: 'RAW' or 'FILTERED'. By default the latter is defined.
`specific_dates`	String. A list of specific dates to consider in processing eg. c("1999-01-01","2005-12-22","2010-11-01"). By default all analysis dates are considered (based in classifications dates)
`vegetation_classes`	Integer. A list of classes codes, which will be referred as not disturbed vegetation class. The rest of classes are considered as intervened areas
`regrowth_observations`	Integer. A number of consecutive observations to considered as vegetation regrowth after a disturbance. If this process is identified, the starting date is indicated.
`calculate_frequencies`	Logical. Calculate land cover frequencies? By default it is not calculated.
`median_filter`	Numeric. An odd number for define a kernel size and apply a median filter. Ignored if it is NA
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restitute_files`	Logical. Restitue files from an old run? .csv files should exist at '/ANALYSIS/CLASSIFICATION/SYNTHESIS/DATA' folder

Value

Vegetation loss, regrowth and classes frequencies are stored at: ’*base_folder*/ANALYSIS/CLASSIFICATION/SYNTH

`classification_training`

Train a model for classify composites

Description

Creates a model for classify composites using methods from caret package.

Usage

```
classification_training(base_folder, extracted_data = NA,
  algorithm_name = "rf", preprocess_method = "none",
  resampling_method = "boot", additional_sampling = "none",
  iteration_number = 10, number_cores = 2, return_outputs = F)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`extracted_data`	String or SpatialPointsDataFrame. A filename or a point-based shapefile from the ’sampling_extraction’ function. If NA, shapefile will be search at: ’ANALYSIS/SAMPLING/DATA_EXTRACTION/extracted_data.shp’
`algorithm_name`	String. A classification algorithm from the caret package. See: ’http://topepo.github.io/caret/available models.html’. Some tested are: "AdaBag", "evtree", "gbm", "kknn", "lssvmRadial", "multinom", "nb", "pcaNNet", "pda", "rf", "rFerns", "svmLinear", "svmPoly", "treebag"
`preprocess_method`	String. A pre-processing method to apply to predictors by caret::train function. Options: "none","BoxCox","YeoJohnson", "expoTrans", "center", "scale", "range", "knnImpute", "bagImpute", "medianImpute", "pca", "ica" and "spatialSign"
`resampling_method`	String. A resampling method defined by caret::trainControl function. Options are: "boot", "boot632", "cv", "repeatedcv", "LOOCV", "LGOCV"
`additional_sampling`	String. A Type of additional sampling by caret::trainControl function. Options are: "none", "down", "up", "smote", or "rose"
`iteration_number`	Numeric. The resampling iterations to apply
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`return_outputs`	Logical. Return model after calculations?

Value

Model file stored at: *base_folder*/ANALYSIS/CLASSIFICATION/MODEL

`classification_trends` *Get trends from a composited vegetation index using breakpoints from synthesis map*

Description

This function calculates pre/post trend averages and its differences using breakpoints from the synthesis map. As the latter is required, this should be calculated before.

Usage

```
classification_trends(base_folder, crop_area = NA, calculation_index,
  synthesis_map = "loss_date", specific_dates = NA, mean_type = "median",
  median_filter = NA, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA, restitute_files = F)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`crop_area`	String. If outputs need to be crop, a shapefile name should be introduced here
`calculation_index`	String. A index or transformation band prefix found at '/COMPOSITES/INDICES' to use in disturbace-regrowth calculations
`synthesis_map`	String. A synthesis map to use for trend-metrics. Note that stable-classes will be calculated using its half. Two options: "loss_date" or "regrowth_date". By default the first is used.
`specific_dates`	String. A list of classification dates to process eg. c("1999-01-01","2005-12-22"). By default all classification dates are processed
`mean_type`	String. A median type to apply in calculations. Two options: "mean" or "median"
`median_filter`	Numeric. An odd number for define a kernel size and apply a median filter. Ignored if it is NA
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restitute_files`	Logical. Restitue files from an old run? .csv files should exist at '/ANALYSIS/CLASSIFICATION/SYNTHESIS/DATA' folder

Value

Trends files are stored at: '*base_folder*/ANALYSIS/CLASSIFICATION/TRENDS'

`composites_match` *Histogram match for composites*

Description

As composites from different footprints could not respond to the same range of values, this function adjust histograms with a reference composite. Target and reference composites should have the same indices or bands but also a common area

Usage

```
composites_match(base_folder, reference_folder, match_statistics = F,
  number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`reference_folder`	String. A folder name with a list of composites to use them as a reference in histogram matching. By default, all tif files inside folders and subfolders inside are considered
`match_statistics`	Logical. Should be histogram match performance statistics calculated ?
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Histogram matched composite files are overwritten at: base_folder/ANALYSIS/COMPOSITES/INDICES & BANDS

`composites_median` *Median based composites for indices or bands*

Description

Composites bands or indices from a selected interval of time. Each pixel in the resulting composite constitute the median value of the images grouped.

Usage

```
composites_median(base_folder, RAD_COR_folder = "SR", time_step = 48,
  target_indices, target_bands, number_cores = 2, manage_cores = T,
  temp_folder = NA, restart_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RAD_COR_folder`	String. A subfolder prefix located at /RAD_COR and /INDICES for use in composites. By default the surface reflectance folder 'SR' is selected
`time_step`	Numerical. Defines the time interval for group indices of different dates in a composite. Units are months (eg. 48)
`target_indices`	String. A list of indices names to composite (eg. c("NDVI","NBR")). Calculations are avoid if it is NA
`target_bands`	Integer. A list of input bands to composite (eg. c(7,5,3)). Codification of input bands in all cases same as TM sensor: 1,2,3,4,5 and 7. Calculations are avoid if it is NA
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`manage_cores`	Logical. If true, 'number_cores' are reduced in demanding calculations for avoid memory overflow
`temp_folder`	String. A folder for temporary files
`restart_folder`	String. If function is required to start not from begging but from a specific folder (eg. "2014-07-18")

Value

Composite files are stored at: base_folder/ANALYSIS/COMPOSITES/

`composites_modify` *Mask composite values using a shapefile*

Description

This function mask composites values using a polygon shapefile. Each polygon should indicate the date of the composite to modify.

Usage

```
composites_modify(base_folder, nodata_shapefile, date_column,
  number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`nodata_shapefile`	String. A polygon shapefile to use for modify nodata in composites
`date_column`	String. The column name in the 'nodata_shapefile' which indicates the date of the composites to modify
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder name for temporary files

Value

Modified composites with the nodata values added

`composites_percentiles`

Percentiles based composites for indices or bands

Description

Composites bands or indices from a selected interval of time. Each pixel in the resulting composite constitute a series of values representing its percentiles.

Usage

```
composites_percentiles(base_folder, RAD_COR_folder = "SR", time_step = 48,
  perc_probs = c(0.2, 0.4, 0.6, 0.8), target_indices, target_bands,
  limit_images = NA, number_cores = 2, temp_folder = NA,
  restart_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RAD_COR_folder`	String. A subfolder prefix located at /RAD_COR and /INDICES for use in composites. By default the surface reflectance folder ’SR’ is selected
`time_step`	Numerical. Defines the time interval for group indices of different dates in a composite. Units are months (eg. 48)
`perc_probs`	Numberic. A list of percentiles probabilites to calculate from images sets in composites. Values should be from 0.01 to 0.99
`target_indices`	String. A list of indices names to composite (eg. c("NDVI","NBR")). Calculations are avoid if it is NA
`target_bands`	Integer. A list of input bands to composite (eg. c(7,5,3)). Codification of input bands in all cases same as TM sensor: 1,2,3,4,5 and 7. Calculations are avoid if it is NA
`limit_images`	Integer. As composites could be made by multiple images and consume all RAM memory during computation, this parameter control its number. By default all images available are used.
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restart_folder`	String. If function is required to start not from begging but from a specific folder (eg. "2014-07-18")

Value

Composite files are stored at: base_folder/ANALYSIS/COMPOSITES/

`composites_RGB` *Create RGB stacks from composites bands or indices*

Description

This function creates RBG stacks from composites calculated. In any case RGB can be created from indices or bands

Usage

```
composites_RGB(base_folder, RGB_bands = c("band7", "band4", "band3"),
  number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RGB_bands`	String. A list of composite bands or indices to stack as RGB. Their order should follow for assignation of red, green and blue channels. For bands apply as: eg. c("band7","band5","band3"); and indices: c("TCTW","TCTB","TCTG")
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Composite files are stored at: base_folder/ANALYSIS/COMPOSITES/RGB

`composites_trends` *Trends and other metrics derived from all composites*

Description

Calculates temporal metrics from all composites. Using a selected indice, this function calculates: max, min, mean, median, MAD, std, variance, nodata, quantiles and lineal regression

Usage

```
composites_trends(base_folder, indice_composite, specific_dates = NA,
  lineal_regression = NA, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`indice_composite`	String. A indice composite name to use in calculations eg. "NDVI". It should exist at /ANALYSIS/COMPOSITES/INDICES
`specific_dates`	String. A list of composites dates to use in calculations eg. c("1999-01-01","2005-12-22"). By default all classification dates are used
`lineal_regression`	String Should be calculated a robust lineal regression? OPtions: 1) "lm" for simple lineal regression; 2) "rlm" for robust lineal regression and 3) NA for ignore it
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Temporal metrics files stored at: '*base_folder*/ANALYSIS/METRICS/TEMPORAL'

`images_clip` *Clip images files*

Description

Clip images files (*.tif or *.envi) with a clip shapefile (*.shp). The latter should be projected in the same coordinate system as the files to clip

Usage

```
images_clip(base_folder, clip_filename, output_folder = NA, verbose = T,
  number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`clip_filename`	String. A file name of a shapefile to use for clipping
`output_folder`	String. A folder name to use for store clipped images (eg. C:/CLIP). If it is NA, then input images will be overwrote
`verbose`	Logical. List target files before clipping?
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder name for temporary files

Value

Clipped image files with the same name at 'base_folder/IMAGES' or at the 'output_folder'

`images_copyMove` *Move or copy files or folders found at 'base_folder/IMAGES' structure*

Description

This function move or copy images files or folders found by pattern at 'base_folder/IMAGES' (but not outside of it) to an specified output folder, conserving their folder structure. It requires a pattern for initiate its search but other conditions can be applied (e.g. time period, specific months, search by files, serach by folders)

Usage

```
images_copyMove(base_folder, operation_type = "copy", output_folder,
  search_pattern = "all", only_files = T, time_period = NA,
  specific_months = NA, number_cores = 2)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`operation_type`	String. If data require to be moved, write 'move'. Otherwise use 'copy' if data require to be copied
`output_folder`	String. An output folder name for store moved files
`search_pattern`	String. A list of patterns for search and move. If 'all' is set, then all folder or files will be selected. By defaul the latter is defined
`only_files`	Logical. Apply pattern search only for files? otherwise 'search_pattern' is applied for folders
`time_period`	String. Defines a time period for select data by specifing start and end dates (e.g. c('1985-01-01','2000-01-01'), then all images between this two dates will be selected). Ignored if NA
`specific_months`	Numeric. Select data of specific months in a year (e.g. c(7,9), then only months between july and september will be selected). Ignored if NA
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned

Value

Moved files or folders identified

`images_delete` *Delete files or folders found by pattern at 'base_folder/IMAGES' structure*

Description

Delete files or folders found by pattern at 'base_folder/IMAGES' structure (but not outside of it).

Usage

```
images_delete(base_folder, search_pattern, only_files = T)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene. For define a specific location inside the 'base_folder' and search on it, a subfolder name can be added eg: 'base_folder/ANALYSIS'
`search_pattern`	String. A list of file or folder patterns for seek & delete
`only_files`	Logical. Apply pattern search only for files? otherwise pattern search is applied only for folders

Value

Deleted files or folders identified

`images_geometry` *Evaluate geometry accurancy based in images water mask*

Description

This function evaluates the degree of geometry accurancy applying a correlation between images using the water mask derived from cfmask. A reference image should be selected in order to establish how much are correlated the rest of images in the collection with it. All images should belong to the same path-row and must contain its cfmask proceesed with 'images_masks' function

Usage

```
images_geometry(base_folder, reference_folder, RGB_name,
  correlation_threshold = 0.5, zoom_factor = 5000, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`reference_folder`	String. A folder name of an image to use it as a geometric reference in the evaluation (normally the best one, clear of clouds or noise) (eg. 'base_folder/IMAGES/*date*')
`RGB_name`	String. A prefix name to refer a file in 'base_folder/IMAGES/*date*/RGB' eg.(SR_743)

`correlation_threshold`	Numeric. A correlation threshold for select suspicious images whose correlation fails to be similar with the reference. All these images will be plotted for their evaluation by the user
`zoom_factor`	Numeric. A distance from samples to consider the extent of plots. By default is set in 500 mts
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder name for temporary files

Value

Output files are stored at: base_folder/ANALYSIS/GEOMETRY

`images_indices` *Calculate indices, ratios, and transformation bands*

Description

Using images bands, this script calculates indices, ratios, and transformation bands. Results can be optionally standardized

Usage

```
images_indices(base_folder, RAD_COR_folder = "SR", indices_names,
  standardized = F, apply_mask = F, number_cores = 2, temp_folder = NA,
  restart_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RAD_COR_folder`	String. A subfolder prefix located at ’/IMAGES/*date*/RAD_COR’ for use in indices calculations. By default the surface reflectance folder ’SR’ is selected
`indices_names`	String. A vector of indices to calculate. 1) indices: ("NDVI","GEMI","MSAVI","EVI","NBR","AFRI 2) tasseled cap: ("TCTB","TCTG","TCTW"); 3) ratios: ("R34","R43","R23","R32","R45","R54","R
`standardized`	Logical. Should results be standardized (i.e. mean 0 and SD 1)?
`apply_mask`	Logical. Apply the ’nodata’ mask to indices?
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restart_folder`	String. If function is required to start not from begging but from a specific folder (eg. "2014-07-18")

Value

Indices files are stored at: base_folder/IMAGES/*date*/INDICES

`images_mask` *Calculate 'nodata' and 'water' masks for each image*

Description

Merges all masks files into one called 'nodata' and isolate water bodies into one mask called 'water' for each image at 'base_folder/IMAGES

Usage

```
images_mask(base_folder, calculate_masks = T, include_water = T,
  apply_mask = T, RAD_COR_folder = "SR", number_cores = 2,
  temp_folder = NA, restart_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`calculate_masks`	Logical. If masks are already calculated, this option avoid their calculation
`include_water`	Logical. Should 'nodata' mask include water mask? *Relevant if calculate_masks is TRUE
`apply_mask`	Logical. Should be 'nodata' mask applied to selected RAD_COR_folder? WARNING: files will be overwritten
`RAD_COR_folder`	String. A subfolder prefix located at /RAD_COR outputs for apply masks. By default the surface reflectance folder 'SR' is selected
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restart_folder`	String. If function is required to start not from begging but from a specific folder (eg. "2014-07-18")

Value

Processed mask files are stored at: 'base_folder/IMAGES/*date*/MASK'.

`images_maskSum` *Sum 'nodata' and 'water' masks*

Description

This function takes all masks files from IMAGES/*date*/MASK, sum them and save into new files.

Usage

```
images_maskSum(base_folder, chunksize = 1e+06, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder name for temporary files

Value

Sum masks files are stored at: base_folder/ANALYSIS/MASKS_SUM

`images_metadata` *Metadata files tabulation*

Description

This script extract information from each image metadata file and tabulate it in an unique table. Additionally, a plot is created for visualize no data cover

Usage

```
images_metadata(base_folder, output_prefix, nodata_msk = T, scale_plot = 1,
  only_plot = F, return_outputs = F, number_cores = 2)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`output_prefix`	String. A prefix name for save outputs
`nodata_msk`	Logical. Use 'nodata' mask in plots? If false, 'CLOUD_COVER' value reported in metadata is used. To use 'nodata' mask, the script 'images_maskSum' must be ran before
`scale_plot`	Numerical. Defines size of plot. By defaul is set in 1
`only_plot`	Logical. Avoid tabulation and return only the plot. Metadata table should be located at: base_folder/ANALYSIS/METADATA
`return_outputs`	Logical. Return table and plot after calculations?
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned

Value

Metadata output files are stored at: base_folder/ANALYSIS/METADATA

`images_organize` *Organize, verify and get average extent from unzipped images*

Description

Organize unzipped images and verify if them are projected in the same coordinate system or if exist missing band files. Additionally, it creates a shapefile of the average extent of unzipped images

Usage

```
images_organize(base_folder, RAD_COR_folder = "SR")
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RAD_COR_folder`	String. A subfolder prefix located at '/IMAGES/*date*/RAD_COR' for organize and verify. By default the surface reflectance folder 'SR' is selected

Value

If fails, a list of images folders are listed. The average extent shapefile is stored at: 'base_folder/ANALYSIS/AVG_EXTE

`images_rename` *Rename files or folders found by pattern at 'base_folder/IMAGES' structure*

Description

Rename files or folders found by pattern at 'base_folder/IMAGES' structure (but not outside of it)

Usage

```
images_rename(base_folder, search_pattern, replace_pattern, only_files = T)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`search_pattern`	String. A file or folder pattern for seek & replace. It can be only one search pattern
`replace_pattern`	String. The replacement pattern of the file or folder. It can be only one replacement pattern
`only_files`	Logical. Apply pattern search only for files? otherwise pattern search is applied only for folders

Value

Renamed files or folders identified

`images_RGB` *Stack RGB for each image*

Description

Stack RGB for each image applying a band combination. This function also creates quick views from RGB files for a fast assesment of masking outputs or image quality

Usage

```
images_RGB(base_folder, RAD_COR_folder = "SR", input_bands = c(7, 4, 3),
  apply_mask = F, stretch_value = 0.05, quick_view = T,
  background_color = "yellow", zoom_point = NA, zoom_factor = 5000,
  number_cores = 2, temp_folder = NA, restart_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`RAD_COR_folder`	String. A subfolder prefix located at ’/IMAGES/*date*/RAD_COR’ for use in RGB stacks. By default the surface reflectance folder ’SR’ is selected
`input_bands`	Numeric or string. Bands names for combine them in a RGB. Their order specify channels: Red, Green and Blue. Codification of bands in all cases same as TM sensor: 1,2,3,4,5 and 7 eg. c(7,4,3)
`apply_mask`	Logical. Should ’nodata’ mask be applied to RGB?
`stretch_value`	Double. Probability value for stretch RBG histograms. By default, value is set to 0.05
`quick_view`	Logical. Create a quick view from RGBs files? By default is set as TRUE
`background_color`	String. If quickView is TRUE, then a color should be set for masked areas. By default it is set as ’yellow’
`zoom_point`	Numeric. If quickView is TRUE, XY coordinates (same coordinate system as images) of a point for zoom on its position. By defaul the center of the image is considered
`zoom_factor`	Numeric. If quickView is TRUE, this value controls zoom extent
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`restart_folder`	String. If function is required to start not from begging but from a specific folder (eg. "2014-07-18")

Value

RGB stacked files are stored at: ’base_folder/IMAGES/*date*/RGB’ and quick view files at: ’base_folder/ANALYSIS/Q

`images_trends` *Calculate trends from a index using data from 'base_folder/IMAGES' structure*

Description

This function takes all index files, puts on a time-series and fit a generalized Deming regression

Usage

```
images_trends(base_folder, target_index = "NDVI", crop_area = NA,
  chunksize = 1e+06, number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`target_index`	String. A index to use for calculate trends. Should exists at 'base_folder/*date*/INDICES/
`crop_area`	String. String. If inputs need to be crop, a shapefile name should be introduced here. Not recommended as it is computing demanding, instead use 'images_clip' function
`chunksize`	Numeric. A number of cells to process at each iteration. By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder name for temporary files

Value

Sum masks files are stored at: base_folder/ANALYSIS/TREND_CALC

`images_unzip` *Unzip Landsat zip files*

Description

Extract and organize data from zip files downloaded from ESPA - Ordering Interfase (see: http://espa.cr.usgs.gov/index/) for Landsat 5, 7 and 8. Surface reflectance, input product metadata and CFMask must be included in the order for next steps. If vegetation indexes are included in the order, they are automatically detected and organized.

Usage

```
images_unzip(zip_folder, number_cores = 2)
```

Arguments

`zip_folder`	String. A folder name containing the Landsat zip files
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned

Value

A structure of folders which organize unzipped Landsat data. Zip files are moved to a folder called: 'base_folder/ZIP/PASS' while corrupted zip files are moved to: 'base_folder/ZIP'

`others_batchClip` *Bach projection and clipping for shapefiles*

Description

This function applies a batch projection and clipping of a folder of multiple shapefiles (which can be in different projection systems). By default, the clipping shapefile defines the projection of outputs.

Usage

```
others_batchClip(shp_folder, clip_shp, output_folder, random_number = F)
```

Arguments

`shp_folder`	String. A folder name containing the shapefiles to clip
`clip_shp`	String. A shapefile name to use for project and clip the shapefiles inside 'shp_folder'
`output_folder`	String. A folder for store outputs
`random_number`	Logical. Add a random number to outputs names? Helpful if there are duplicated names

Value

Shapefiles projected (if they require) and clipped

`others_metricsDEM` *Calculate metrics from a DEM*

Description

Adjust a DEM (projection, resolution and extent) to a reference and calculates its metrics. Among them: aspect, flowdir, roughness, slope (in degrees), tpi (Topographic Position Index) and tri (Terrain Ruggedness Index)

Usage

```
others_metricsDEM(base_folder, DEM_filename, reference_filename = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`DEM_filename`	String. A filename of a DEM file, which should cover all the study area
`reference_filename`	String. A filename of a raster to take its coordinate system, resolution and extension. By default the first composite indice is taken for extract this information

Value

DEM metrics files stored at: '*base_folder*/ANALYSIS/METRICS/DEM'

`others_metricsDis` *Calculate euclidean distance metrics*

Description

Adjust a shapefile or shapefiles zo a raster file (projection, resolution and extent) using a reference for calculate distances from features. These metrics can be used to improve classification of Landsat composites

Usage

```
others_metricsDis(base_folder, shp_folder, reference_filename = NA,
  number_cores = 2, temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`shp_folder`	String. A folder name of points, lines or polygons shapefiles to use for calculate its euclidean distances. All shapefiles should cover all the study area
`reference_filename`	String. A filename of a raster to take its coordinate system, resolution and extension. By default a composite indice file is taken for take this information
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

distance metrics stored at: '*base_folder*/ANALYSIS/METRICS/DISTANCE'

`sampling_creation` *Stratified sampling based in a classmap*

Description

This script creates samples using a classmap and its classes for extract groups or strata. It first calculates the number of samples requiered using a stratified sampling formula, therefore its parameters should be defined. Optionally, manual samples sizes can be assigned as well. Distances between samples can be adjusted and as well from the border of the NA values in the classmap. Finally, it can add ancillary samples to the main sample for improve coverage of the sample

Usage

```
sampling_creation(output_folder, strata_map, classes_codes, precision = NA,
  SD_strata = NA, manual_mode = T, border_distance = NA,
  samples_size = NA, samples_distance = 1000, ancillary_distance = NA,
  verbose = T, number_cores = 2, temp_folder = NA)
```

Arguments

`output_folder`	String. A folder name for store output files
`strata_map`	String. A filename of the classmap for extract groups or strata. Should be a raster in a'.tif' format
`classes_codes`	Integer. A list of codes of the classes in the 'strata_map' to be considered in sampling
`precision`	Numeric. A number which specify sampling precision (0.01 to 0.99). Relevant if 'manual_mode' is FALSE
`SD_strata`	Numeric. A list of numbers which define standard deviation of strata samples sizes. Its order should be the same as 'classes_codes'. Relevant if 'manual_mode' is FALSE
`manual_mode`	Logical. If sample size should be assigned manually, this parameter should be TRUE.
`border_distance`	Numeric. A distance value from NA values in 'strata_map' for avoid sampling in border areas. Units should be in pixels
`samples_size`	Integer. A list of numbers of the sample sizes for each strata. Its order should be the same as 'classes_codes'. Relevant if 'manual_mode' is TRUE
`samples_distance`	Numeric. A minimun distance value between samples for avoid autocorrelation. Units should be in meters. If classes require different samples distance, it can be put as a vector of distances. Otherwise, an unique value is applied to all classes. If distance affect sample size it is reduced by 10 percent until best value is found
`ancillary_distance`	Numeric. If a minimun distance value is defined, ancillary samples will be created from the main sample . Units should be in pixels, Ignored if NA
`verbose`	Logical. Get samples sizes summary before sampling?
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Stratified samples shapefile stored at: *output_folder*

`sampling_extraction` *Extract data from Landsat composites and metrics for a sample point-based shapefile*

Description

This script extract data from a defined group of composites indices and optionally from a folder with additional data. It used a point-based shapefile, which should have a column with the class codes (numeric format). Additionally, it creates a boxplots for visualize each class data distribution

Usage

```
sampling_extraction(base_folder, date_composite, additional_metrics = NA,
  samples_shapefile, id_column, classes_column, scale_plot = 1,
  number_cores = 2, temp_folder = NA, return_outputs = F)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene
`date_composite`	String. A unique date for extract its values from its indices and bands composites if they exist eg. "1999-01-01"
`additional_metrics`	String. A list of metrics prefixes for extract its values and associate to training samples. Options are: 'TEMPORAL','DEM' or 'DISTANCE'. Ignored if it is NA
`samples_shapefile`	String or SpatialPointsDataFrame. A filename or a point-based shapefile to use in data extraction
`id_column`	String. Column name in the shapefile attribute table which define unique IDÂ´s for samples. IDÂ´s should be always codified as numbers. Column name is case sensitive
`classes_column`	String. Column name in the shapefile attribute table which define samples classes. Classes should be always codified as numbers. Column name is case sensitive
`scale_plot`	Numerical. For samples classes boxplots, this value controls its scale. By defaul is set in 1. Plot is ignored if it is NA
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files
`return_outputs`	Logical. Return values after calculations?

Value

Samples files with extracted data are stored at: *base_folder*/ANALYSIS/SAMPLING/DATA_EXTRACTION

`sampling_fill` *Fill attribute table from samples shapefile*

Description

This is an interactive program for fill attribute table from 'stratified_sampling' function outputs. It opens each sample and plots its units in order to delete or conserve units or simply recode them. This function is supouse to be used visualizing simultaneously the outputs of 'plot_samples' function

Usage

```
sampling_fill(output_folder, samples_shapefile = NA, id_sample_column = NA,
  class_code_column = NA, assign_classes = T, add_messages = F,
  restart_id = NA)
```

Arguments

`output_folder`	String. A folder name of a Landsat scene
`samples_shapefile`	String. A filename of a point-based shapefile to fill its attribute DBF table. By default it is assumed that the shapefile is an output from ’stratified_sampling’ function, therefore if it is NA, the shapefile at ’output_folder/ANALYSIS/SAMPLING/STRATIFIED/ is used
`id_sample_column`	String. The column name of samples identifier at ’samples_shapefile’. By default it is assumed that the shapefile is an output from ’stratified_sampling’ function, therefore if it is NA, ’id_sample’ string is used
`class_code_column`	String. The column name of samples class code at ’samples_shapefile’. By default it is assumed that the shapefile is an output from ’stratified_sampling’ function, therefore if it is NA, ’class_code’ string is used
`assign_classes`	Logical. During table filling, should be ask for assign classes?
`add_messages`	Logical. During table filling, should be ask for introduce messages?
`restart_id`	Numeric. If table fill must be restarted from a specific sample, its ’id’ should be defined. By default is assumed to start from the first id number identified

Value

Shapefile with filled DBF table at: *output_folder*/ANALYSIS/SAMPLING/TABLE_FILLED

`sampling_plots` *Plot samples using high resolution imagery and Landsat time series*

Description

This function creates a series of plots for each sample unit of a point-based shapefile. High resolution imagery can be any type of image but their coordinate system should be the same as Landsat RGB composites. This rule also applies to the shapefile used as samples

Usage

```
sampling_plots(base_folder, high_resolution_folder = NA, evaluation_map,
  samples_shapefile, id_sample_column = "id_sample",
  class_code_column = "class_code", time_frame = NA, sum_mask = F,
  color_sample_units = rep("yellow", 3), zoom_in_factor = 250,
  stretch_value = c(0.01, 0.01), chunksize = 1e+06, number_cores = 2,
  temp_folder = NA)
```

Arguments

`base_folder`	String. A folder name of a Landsat scene

`high_resolution_folder`	String. A folder name with high resolution files (unique raster or RGB composites). Files should apply date format ("YYYY-MM-DD", eg. "1999-01-01"), followed by an underscore "_" and its unique identificator (eg. "1999-01-01_AREA3753920180"). As demanding reading is needed, files are not recommended to be extreamly heavy for some calculations. All images must have the same coordinates system. Recommended pixel depth should be "INT1U" and image size between 250-700 MB. Ignored high resolution if NA
`evaluation_map`	String. A filename of the 'loss-regrowth' map to evaluate. It will be zoomed in and out for each sample unit plot
`samples_shapefile`	SpatialPointsDataFrame. A point-based SpatialPointsDataFrame to use it as sampling points to observe with high resolution images
`id_sample_column`	String. The column name of samples identifier at 'samples_shapefile'. By default it is assumed that it is 'id_sample'
`class_code_column`	String. The column name of samples class code at 'samples_shapefile'. By default it is assumed that it is 'class_code'
`time_frame`	String. The initial and end dates for specific periods in plots. Dates should apply date format ("YYYY-MM-DD", eg. "1999-01-01") and should be two (start and end dates).If period is NA, then they are extracted from first and last RGB composites.
`sum_mask`	Logical. Should be created a mask that sum all high resolution images extents? This output identifies where high resolution imagery is more frequent. By default it is not calculated
`color_sample_units`	String. A list of color names for samples units. Should be four, indicating: 1) for the evaluation map, 2) for the high resolution images, and 3) for the Landsat RGB. By default this is set as: c("yellow","yellow","yellow")
`zoom_in_factor`	Numeric. A factor defining zoom in for evaluation map, high resoultion images chips, and Landsat RGB. By default is set in 250
`stretch_value`	Numeric. A list of two numbers (between 0.01-0.99) for apply a stretch to: 1) high resolution imagery; and 2) Landsat RGB. By default a stretch of c(0.01,0.01) is applied
`chunksize`	Numeric. A number of cells to process at each iteration By default it is defined in 1000000
`number_cores`	Integer. A number indicating the number of cores to use for parallel processing. By default, 2 is assigned
`temp_folder`	String. A folder for temporary files

Value

Plots for each sample at: *base_folder*/ANALYSIS/SAMPLING/PLOTS

Index

classification_harmonize, 2
classification_mosaic, 3
classification_prediction, 4
classification_synthesis, 5
classification_training, 6
classification_trends, 7
composites_match, 8
composites_median, 8
composites_modify, 9
composites_percentiles, 10
composites_RGB, 11
composites_trends, 11

images_clip, 12
images_copyMove, 13
images_delete, 14
images_geometry, 14
images_indices, 15
images_mask, 16
images_maskSum, 16
images_metadata, 17
images_organize, 18
images_rename, 18
images_RGB, 19
images_trends, 20
images_unzip, 20

others_batchClip, 21
others_metricsDEM, 21
others_metricsDis, 22

sampling_creation, 22
sampling_extraction, 23
sampling_fill, 24
sampling_plots, 25

CURRICULUM VITAE

Fabián Santos

Curriculum vitae

Adress: ZFL. Walter Flex Str. 3, Bonn.
Telephone: (+49) 017637596279; (+593) 2506060
Email: fabian_santos_@hotmail.com
Git: https://github.com/FSantosCodes

Personal information

Date of birth: 22.08.1982
City of birth: Quito, Ecuador

Work experience

2012 – 2014 Technical in Remote Sensing and ecosystem mapping at the Environmental Ministry of Ecuador

- Programmer of routines in R, Python and ArcGIS for database management, ecosystem mapping and forest strata levels.

2010 – 2013 Consultant for Deforestation Mapping in the Amazon Basin. Socio-Environmental Geoinformation Network of the Amazon Watershed (RAISG) – Brasil

- Software tester of IDL routines for Landsat processing and change detection analysis. Development and report of deforestation maps 2000, 2005 and 2010 for the Ecuadorian Amazon.

2009 – 2010 Aid worker for the German Service of Technical Cooperation in the Government of Pastaza Province (GADPPz).

- Project design, database management and GIS analysis in the Land Planning Office of GADPPz. Tutoring and supervision of GIS related activities.

Education

2014 – 2018 Ph.d candidate at the University of Bonn. Center for Remote Sensing of Land Surfaces (ZFL)

- Thesis title: ‘A Landsat-based analysis of tropical forest dynamics in the Central Ecuadorian Amazon: patterns and causes of deforestation and reforestation.’

2011 Master in Geoinformation Sciences for Land Resources, San Simon University

- Dissertation title: ‘Deforestation monitoring with poor-quality Landsat 7 images in the Auca Sur sector of the Ecuadorian Amazon’

2002 – 2009 Engineer in Geography and Land Planning. Minor in Ecotourism. Pontifical Catholic University of Ecuador

- Thesis title: ‘Determination of the structure and functioning of the hydrogeological system of Santa Cruz island - Galapagos’

Paper contributions

In scientific journals:

- Santos F., Meneses P. & Hostert P., 2018. Monitoring Long-Term Forest Dynamics with Scarce Data: A Multi-Date Classification Implementation in the Ecuadorian Amazon. *European Journal of Remote Sensing* (Manuscript accepted for publication).
- Santos F., Dubovyk O. & Menz G., 2017. Monitoring Forest Dynamics in the Andean Amazon: The Applicability of

Breakpoint Detection Methods Using Landsat Time-Series and Genetic Algorithms. *Remote Sensing.* 2017, 9, 68.

Presented in conferences:

- Santos F. & Menz G., 2015. Insights for Manage Geospatial Big Data in Ecosystem Monitoring using Processing Chains and High Performance Computing. DCCLOSER Conference. 2015, Lisbon.

Courses and awards

2017 ESA. 7th advanced Training Course on Land Remote Sensing.

2015 University of Innsbruck. Close Range Sensing Techniques in Alpine Terrain.

2014. University of Bonn – ZEF. Interdisciplinary course 'Concepts and Theories of Development'.

2013 Silvacarbon and INPE. LiDAR processing

2010 JIICA, INPE and IBAMA. Forest monitoring and reporting training.

Programming skills

R - image processing, spatial analysis, data visualization, statistical analysis, web scraping, package development.

Python - image processing, spatial analysis.

GIS, remote sensing & other software skills

ArcGis, SNAP, ENVI, ERDAS, Ilwis, Grass, QGis, Saga Gis, Dinamica Ego, Fuzion (LiDAR), Microsoft Office, Gimp, Latex.

Field Skills

GPS navigation, terrestrial lidar, unmanned aerial vehicles, agroecology.

Communication Skills

Spanish (native) & English (fluently)

Research interests

Remote sensing, time series analysis, land cover dynamics, data mining, machine learning, high-performance computing.