GEOSPATIAL MODELING OF LAND COVER CHANGE IN THE CHOCÓ-DARIEN GLOBAL ECOREGION OF SOUTH AMERICA: ASSESSING PROXIMATE CAUSES AND UNDERLYING DRIVERS OF DEFORESTATION AND Reforestation

by

José Camilo Fagua

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY in

Ecology

Approved:

__________________________
R. Douglas Ramsey, Ph.D.
Major Professor

__________________________
James A. Lutz, Ph.D.
Committee Member

__________________________
Eugene W. Schupp, Ph.D.
Committee Member

__________________________
Laura C. Schneider, Ph.D.
Committee Member

__________________________
Jacopo Baggio, Ph.D.
Committee Member

__________________________
Laurens H. Smith, Ph.D.
Interim Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2018
ABSTRACT

Geospatial Modeling of Land Cover Change in the Chocó-Darien Global Ecoregion of South America: Assessing Proximate Causes and Underlying Drivers of Deforestation and Reforestation

by

José Camilo Fagua, Doctor of Philosophy

Utah State University, 2018

Major Professor: Dr. R. Douglas Ramsey
Department: Wildland Resources

The Tropical Rain Forests of northwest South America fall within the Chocó-Darien Global Ecoregion (CGE). The CGE is one of 25 global biodiversity hotspots prioritized for conservation due to its high biodiversity and endemism as well as threats due to deforestation. The analysis of land-use and land-cover (LULC) change within the CGE using remotely sensed imagery is challenging because this area is considered to be one of the rainiest places on the planet (hence high frequency of cloud cover). Furthermore, the availability of high-resolution remotely sensed data is low for developing countries prior to 2015, including Panamá, Colombia, and Ecuador. In this dissertation, I performed the first LULC change analysis for the entire CGE and address three main objectives: 1) Selection of the best available imagery to build annual LULC maps from 2001 to 2015 across the CGE. 2) Model LULC change across the CGE to assess forest change trends from 2002 to 2015 and identify the effect of proximate causes of deforestation and reforestation. 3) Estimate the effects of underlying drivers on
deforestation and reforestation across the CGE between 2002 and 2015. Using the Random Forest ensemble learning classification tree system, I developed annual LULC maps across the CGE from 2002 to 2015 using a time series of cloud-free MODIS vegetation index products. The MODIS data was selected and processed through a Gaussian weighted filter to further correct for cloud pollution and classification training areas matched to visual interpretations of LULC classes from available high spatial resolution imagery (WorldView-2, Quick Bird, Ikonos and GeoEye-1). Validation of LULC maps resulted in high accuracies (Kappa = 0.87; SD = 0.008). We detected a gradual replacement of forested areas with agriculture (mainly grassland planted to support livestock grazing) and secondary vegetation (agriculture reverting to early regeneration of natural vegetation) across the CGE. Forest loss, which included the change from forest to secondary vegetation, was higher between 2010-2015 (2.6% per year) when compared to 2002-2010 (1.78% per year). The primary proximate cause of deforestation was conversion to grassland in each country. Grassland was also the main proximate cause of reforestation in Colombia and Panamá while crop was the main proximate cause of reforestation in Ecuador. Population growth and road density were underlying drivers of deforestation. Armed conflicts, Gross Domestic Product, and average annual rain were underlying drivers related to reforestation.
PUBLIC ABSTRACT

Geospatial Modeling of Land Cover Change in the Chocó-Darien Global Ecoregion of South America: Assessing Proximate Causes and Underlying Drivers of Deforestation and Reforestation

José Camilo Fagua

The Chocó-Darien Global Ecoregion (CGE) in South America is one of 25 global biodiversity hotspots prioritized for conservation. I performed the first land-use and land-cover (LULC) change analysis for the entire CGE in this dissertation. There were three main objectives: 1) Select the best available imagery to build annual land-use and land-cover maps from 2001 to 2015 across the CGE. 2) Model LULC across the CGE to assess forest change trends from 2002 to 2015 and identify the effect of proximate causes of deforestation and reforestation. 3) Estimate the effects of underlying drivers on deforestation and reforestation across the CGE between 2002 and 2015. I developed annual LULC maps across the CGE from 2002 to 2015 using MODIS (Moderate Resolution Imaging Spectroradiometer) vegetation index products and random forest classification. The LULC maps resulted in high accuracies (Kappa = 0.87; SD = 0.008). We detected a gradual replacement of forested areas with agriculture and secondary vegetation (agriculture reverting to early regeneration of natural vegetation) across the CGE. Forest loss was higher between 2010-2015 when compared to 2002-2010. LULC change trends, proximate causes, and reforestation transitions varied according to administrative authority (countries: Panamanian CGE, Colombian CGE, and Ecuadorian CGE). Population growth and road density were underlying drivers of deforestation.
Armed conflicts, Gross Domestic Product, and average annual rain were proximate causes and underlying drivers related to reforestation.
ACKNOWLEDGMENTS

I would like to thank my advisor Dr. R. Douglas Ramsey for his support and guidance throughout my doctoral program. I thank my committee members – Dr. James A. Lutz, Dr. Jacopo Baggio, Dr. Eugene W. Schupp, and Dr. Laura C. Schneider for their support and assistance, as well as their guidance in understanding the nuances of academia. I also acknowledge Dr. Adele Cutler for her help for developing this dissertation.

I want to express my gratitude to Benjamin Crabb, Dr. Andy Kleinhesselink, and Jaime Florez for their collaboration and assistance in learning R programeing and Chris Garrard and Cristian Loaiza for assisting with Python programming. Kendall Becker (Writing Center of Utah State University) helped me improve English writing skills. I would like to thank the students and staff of the Remote Sensing and GIS Lab in the Department of Wildland Resources at USU for their constant collaboration. I acknowledge COLCIENCIAS-Colombia (529-2011) and Fulbright-US who supported this research, the Ecology Center and the Remote Sensing/GIS lab at Utah State University-US who provided additional financial and technical support, DigitalGlobe Foundation provided the high spatial resolution imagery, IDEAM-Colombia and IGAC-Colombia provided data and technical support, and IDEA WILD provided equipment.

I am thankful to all my friends in Logan during these past five years who have been a great mental support during my doctoral program. Most of my gratitude is for my mother and father who always are in my mind.

José Camilo Fagua
CONTENTS

Page

ABSTRACT ........................................................................................................................................ iii

PUBLIC ABSTRACT ......................................................................................................................... v

ACKNOWLEDGMENTS ..................................................................................................................... vii

LIST OF TABLES .............................................................................................................................. xi

LIST OF FIGURES ............................................................................................................................ xiii

CHAPTER

1. INTRODUCTION .......................................................................................................................... 1

   Literature Cited ............................................................................................................................. 6
   Figures ........................................................................................................................................... 11

2. COMPARING THE ACCURACY OF MODIS DATA PRODUCTS FOR VEGETATION DETECTION BETWEEN TWO ENVIRONMENTALLY DISSIMILAR ECOREGIONS: THE CHOCÓ-DARIEN OF SOUTH AMERICA AND THE GREAT BASIN OF NORTH AMERICA ........................................ 12

   Abstract .......................................................................................................................................... 12
   Introduction ..................................................................................................................................... 13
   Methods ......................................................................................................................................... 18

      Study area .................................................................................................................................. 18
      Identification of vegetation cover in the GB and the CGE ......................................................... 20
      MODIS products ......................................................................................................................... 22
      Post process optimization .......................................................................................................... 23
      Vegetation cover classification ................................................................................................. 25

   Results .......................................................................................................................................... 27
   Discussion ...................................................................................................................................... 29
   Conclusions ................................................................................................................................. 32
   Literature Cited ............................................................................................................................. 32
   Tables and Figures ....................................................................................................................... 45

3. GEOSPATIAL MODELING OF LAND COVER CHANGE IN THE CHOCÓ-DARIEN GLOBAL ECOREGION OF SOUTH AMERICA; ONE OF MOST BIODIVERSE AND RAINY AREAS IN THE WORLD. ............... 51
5. CONCLUSIONS AND SUMMARY .................................................................150

Remote sensing analysis .............................................................................150
LULC dynamic and proximate causes of forest changes .........................151
Hotspots and underlying causes of forest changes .................................152
Literature Cited .........................................................................................154

APPENDICES .................................................................................................156

Appendix A: Chapter 2 Supplemental Materials ........................................157
Appendix B: Chapter 3 Supplemental Materials ........................................159
Appendix C: Chapter 4 Supplemental Materials ........................................169

CURRICULUM VITAE ......................................................................................173
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>MODIS vegetation index products with the spatial resolution of 250 m</td>
</tr>
<tr>
<td>3.1</td>
<td>Correlations of land-use and land-cover change trends among administrative divisions; The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE). Pearson's correlation coefficient (R) are shown for two time periods 2002-2010 and 2010-25. Significant correlation are bolded</td>
</tr>
<tr>
<td>4.1</td>
<td>Variables evaluated as underlying causes of deforestation and reforestation</td>
</tr>
<tr>
<td>S2.1</td>
<td>Selection of the windows size for the temporal optimization using the Asymmetric Gaussian function. We assessed windows sizes from 3 to 7 data and selected the size that maximize the accuracy (Kappa-K) of the classification. To estimate K, 100 iterations were performed for every Random Forest classification, each with 5-fold cross-validation. Variation of window size did not affect Kappa significantly (P-values &gt; 0.05)</td>
</tr>
<tr>
<td>S2.2</td>
<td>Selection of filter order and Weighted Moving Average Filter (WMAF) for the temporal optimization using the Savitzky-Golay function. We assessed WMAF sizes from 3 to 9 and filter order from 2 to 4. We selected the WMAF and filter order that maximized Kappa (K). To estimate K, 100 iterations were performed for every Random Forest classification, each one with 5-fold cross-validation. Variation of filter order and WMAF by MODIS products did not significantly affect Kappa (P-values &gt; 0.05)</td>
</tr>
<tr>
<td>S3.1</td>
<td>High spatial resolution imagery used to the visually interpreting of Land-use and land-cover (LULC) classes</td>
</tr>
<tr>
<td>S3.2</td>
<td>Table of response and predictor variables used in the Random Forest classification. This table is in CSV format that can be download in the next link <a href="https://data.gis.usu.edu/?prefix=CDLCC/">https://data.gis.usu.edu/?prefix=CDLCC/</a></td>
</tr>
<tr>
<td>S3.3</td>
<td>Codes S3.3. R codes used for the Random Forest Classification (Run with Table S3.2)</td>
</tr>
<tr>
<td>S3.4</td>
<td>Distribution of samples across land-use/land-cover (LULC) classes before and after sampling reduction</td>
</tr>
</tbody>
</table>
S3.5 Confusion matrix of the cross validation from the original data. Kappa, commissions and omissions are in the matrix.................................................................164

S3.6 Confusion matrix of the cross validation from the data when forest class is reduced as the grassland class number. Kappa, commission and omission are in the matrix. .................................................................165

S3.7 Trends of land-use and land-cover (LULC) changes between 2002-2010. ....166

S3.8 Deforestation (proximate causes of deforestation) and reforestation transitions 2002-2010.................................................................................................167

S3.9 Deforestation (proximate causes of deforestation) and reforestation transitions 2010-2015.................................................................................................168
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The Chocó-Darien global ecoregion (CGE); global location (a) and the countries that share the ecoregion (b).</td>
<td>11</td>
</tr>
<tr>
<td>2.1</td>
<td>Study areas. Global location of the ecoregions analyzed in this study (a); the Great Basin global ecoregion (GB) in North America (b) and the Chocó-Darien global ecoregion (CGE) in South America.</td>
<td>46</td>
</tr>
<tr>
<td>2.2</td>
<td>Annual time series of NDVI and EVI for vegetation types in the Great Basin global ecoregion using MOD13Q1 data. Annual time series from raw data (a,b), annual time series after the asymmetric Gaussian function (c,d), and annual time series after the Savitzky-Golay function (e,f). Mean Absolute Error (MAE) values of NDVI and EVI for each vegetation type using MOD13Q1 data. These analyses were also performed for MYD13Q1, MODQ09Q1, and MYD13Q1.</td>
<td>47</td>
</tr>
<tr>
<td>2.3</td>
<td>Annual time series of NDVI and EVI for the vegetation types in the Chocó-Darien global ecoregion using MOD13Q1 data. Annual time series from raw data (a,b), annual time series after the asymmetric Gaussian Function (c,d), and annual time series after the Savitzky-Golay function (e,f). Mean Absolute Error (MAE) values of NDVI and EVI for each vegetation type using MOD13Q1 data. These analyses were also performed for MYD13Q1, MODQ09Q1, and MYD13Q1.</td>
<td>48</td>
</tr>
<tr>
<td>2.4</td>
<td>Accuracies (measured as kappa-K) of the different vegetation classifications for the Great Basin global ecoregion (North America). The classifications were generated using data from the MODIS products MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1. The original MODIS algorithm was used in the preprocessing of the 13Q1 products while the MAIAC algorithm was used in the preprocessing of the 09Q1 products.</td>
<td>49</td>
</tr>
<tr>
<td>2.5</td>
<td>Accuracies (measured as kappa-K) of the different vegetation classifications for the Chocó-Darrien global ecoregion (South America). The classifications were generated using data from the MODIS products MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1. The original MODIS algorithm was used in the preprocessing of the 13Q1 products while the MAIAC algorithm was used in the preprocessing of the 09Q1 product.</td>
<td>50</td>
</tr>
<tr>
<td>3.1</td>
<td>The Chocó-Darien Global Ecoregion (CGE): (a) estimated historical extent of Tropical Rain Forest (TRF) in South America: TRF-CGE (estimated TRF in CGE), TRF-Amz (estimated TRF in Amazon basin), and TRF-BrAt (estimated TRF in Brazilian Atlantic Forest). (b) The</td>
<td></td>
</tr>
</tbody>
</table>
Chocó-Darien Global Ecoregion (CGE): This global ecoregion is formed by three sub-ecoregions: Magdalena-Urabá Moist forests (MgU), Chocó-Darién Moist Forests (ChD), and Western Ecuador Moist Forests (WEc). ...97

3.2 An example of the time series filtering procedure from 2009 to 2013 using the Gaussian weighted filter (GWF). GWF improved the identification of land cover using the MODIS bands and indices. (a) A time series EVI pixel of woody vegetation before and after filtering; outliers are replaced with estimates calculated by the Gaussian weighted filter. (b) Temporal variation of 120 pixels corresponding to woody vegetation before filtering and (c) the same 120 pixels after filtering; the variance of these 120-time series is reduced. (d) Post filtering results—simplified to means, for 120 woody vegetation pixels, 116 grassland pixels, 542 palm plantation pixels, 99 settlement pixels, 101 water pixels, and 454 wetland pixels. GWF increased the differentiation among these land covers. The GWFs were applied using the Terrset software (Eastman 2017) with a temporal filter length of 5. .................................................................98

3.3 Land-use and land-cover change trends in The Chocó-Darien Global Ecoregion (CGE). Significant correlation coefficients (R) are shown for two time year periods 2002-2010 and 2010-2015 (a). Significant P range values; P<0.001(***), P<0.01(**), and P<0.05(*). Land-use and land-cover maps for 2002, 2010 and 2015 are showed (b, c, d). .....................99

3.4 Quantification of deforestation (proximate causes of deforestation) and reforestation transitions from 2002 to 2010. (a) Percentage of deforested area and net area deforested by every proximate cause of deforestation, and (b) percentage of reforested areas and net area reforested by every reforestation transition. The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE). .................................................................100

3.5 Quantification of deforestation (proximate causes of reforestation) and reforestation transitions from 2010 to 2015. (a) Percentage of deforested area and net area deforested by every proximate cause of reforestation, and (b) percentage of reforested areas and net area reforested by every reforestation transition. The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE). .................................................................101

4.1 The Chocó-Darien Global Ecoregion (CGE): (A) estimated historical extent of Tropical Rain Forest (TRF) in South America: TRF-CGE (estimated TRF in CGE), TRF-Amz (estimated TRF in Amazon basin), and TRF-BrAt (estimated TRF in Brazilian Atlantic Forest). Land-Use and Land-cover map for the CGE at 2002 (B) and 2015 (C), these maps
can be downloaded in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/.................................145

4.2 Deforestation (A) and reforestation (B) hotspots on the land Land-Use and Land-cover map at 2015. Spatial correlation between deforestation and reforestation hotspot areas were mapping (C). These maps can be downloaded in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/......146

4.3 Municipalities with significant clustering of deforestation and reforestation hotspots. (A) Moran scatter plot for deforestation; red diamonds represent the municipalities with significant influence on the Moran’s (I) global test for deforestation. (B) Moran scatter plot for reforestation; violet diamonds represent the municipalities with significant influence on the Moran’s (I) global test for reforestation. (C) Maps of municipalities with significant influence on the Moran’s (I) global test for deforestation. (D) Maps of municipalities with significant influence on the Moran’s (I) global test for reforestation..............................147

4.4 Bayesian Structural Equation model for deforestation. Unstandardized posterior path coefficients in black and standardized posterior path coefficients in red. .................................................................148

4.5 Bayesian Structural Equation model for reforestation. Unstandardized posterior path coefficients in black and standardized posterior path coefficients in red. ....................................................................................149

S4.1 K Relation between global Moran’s test and matrix size for A) deforestation (p=0.001; R= -0.94) and B) reforestation (p=0.001; R= -0.97). Both graphics showed a significant negative exponentially shape. ...........................................................................................................169
CHAPTER 1

INTRODUCTION

Life has proliferated on Earth forming levels of biodiversity that is not spread evenly across its surface. Biodiversity follows spatially cohesive patterns determined by climate, geology, and evolutionary histories; these patterns are used to define "ecoregions" (Mittermeier 1999, Myers et al. 2000). Although every ecoregion defines a distinct biodiversity that provides various ecosystem services fundamental to human life, the most diverse ecoregions on Earth (in terms of species richness) are those where Tropical Rain Forests occur. Tropical rain forests are characterized by dense evergreen vegetation with trees reaching more than 30 m in height. Tropical rain forests are typically located on lowlands of moist tropical climatic zones where the mean annual temperatures exceeds 25°C, the mean annual rainfall higher than 1680 mm, and the dry season is absent or less than two months in length (Primack and Corlett 2009). Global studies focusing on forest cover change have documented extensive deforestation and minor reforestation (regrowth of secondary vegetation) in tropical rain forests during the last few decades (Aide et al. 2013, Hansen et al. 2013, Li et al. 2016). This net loss in tropical rain forests is a primary concern since changes in forest density and structure are a main determinant in the degradation of ecosystem services, such as climate and soil regulation, carbon storage, water supply, and pollination (Lambin et al. 2003, Pfaff et al. 2013, Leblois et al. 2017). Identifying causes that affect deforestation and reforestation in tropical rain forests is essential to developing strategies for mitigating forest loss.

Causes of deforestation and reforestation can be classified into two types: 1) proximate causes (direct causes) are immediate actions that directly result in forest loss or
gain; in other words, proximate causes refer to land management that specifically replaces forests with another type of land cover or allows existing anthropogenic land use to convert to secondary vegetation; 2) underlaying drivers (indirect causes) are interactions of social, economic, political, cultural, or other processes that initiate proximate causes into action resulting in deforestation and reforestation (Geist and Lambin 2001, 2002). This statement has been the basis for a wide array of literature focused on studying proximate causes (e.g., Roberts et al. 2002, Flamenco-Sandoval et al. 2007, Tng et al. 2012, Asner et al. 2012, Houghton 2012, Alejandra Chadid et al. 2015, Armenteras et al. 2017) or underlying drivers (Armenteras et al. 2006, 2013, 2017, Hosonuma et al. 2012, Kissinger et al. 2012, Mueller et al. 2013, Leblois et al. 2017) to assess forest cover change at different spatial scales throughout the tropical rain forests domain. Studies that include both proximate causes and underlying drivers as well as their interactions have been less prevalent (Davalos et al. 2014, Richards 2015).

The Chocó-Darien Global Ecoregion (CGE) of South America holds the most diverse tropical rain forest in relation to vascular plants of the planet (Gentry 1986). The CGE is a lowland area (194,737 km²) located along the pacific coast of eastern Panamá, western Colombia, northern Ecuador and the lowlands of the Magdalena River in Colombia (Fig. 1.1). This ecoregion has been declared as one of the top 25 global hotspots for conservation priorities due to its high biodiversity, species endemism, and threats due to deforestation (Myers et al. 2000, Olson et al. 2001, Primack and Corlett 2009, WWF 2017). While forest cover change analysis is critically important for the preservation of the Tropical Rain Forest of the CGE, studies of this type have not been performed for the entire ecoregion. The analysis of forest cover change across the CGE
can most effectively be done using remotely sensed imagery, but this technique is challenging because this area is considered one of the rainiest on the planet (Poveda and Mesa 2000), resulting in a high percentage of cloud cover obscuring most satellites. In this dissertation, I investigated the forest cover change dynamic across the CGE from 2002 to 2015. The results of my analyses include information useful in the conservation of tropical rain forests of the CGE across and relative to its administrative distribution between three countries.

The first chapter of this dissertation focuses on a remote sensing analysis to develop land-use and land-cover (LULC) maps from 2001 to 2015 across the CGE. This analysis used the Moderate Resolution Imaging Spectroradiometer (MODIS) remote sensing products developed by the Earth Observation System administered by the US National Aeronautics and Space Administration (NASA). I chose the MODIS products since other satellite-based sensors such as Landsat, which have a better spatial resolution compared to MODIS, do not offer the temporal resolution required to reduce the effect of clouds necessary for creating annual LULC maps in the CGE. With its daily repeat cycle, MODIS offers superior temporal resolution. As a result, NASA has created various temporal MODIS products aimed at improving vegetation detection (Lyapustin et al. 2014, Zhang et al. 2017). Users of these products can, and often do, apply post-processing algorithms to optimize these data in order to further improve its utility for LULC mapping. Selecting the appropriate MODIS product to conduct a LULC classification as described here can be confusing since many products offer similar data, but with different processing parameters. As a result, MODIS vegetation products have three possible causes of variation that could affect the accuracy of LULC classifications:
1) the MODIS platform (Aqua or Terra) that collects the data (calibration variability between sensors, orbital characteristics, degradation etc.), 2) the pre-processing algorithm used to identify cloud, cloud shadow and assess atmospheric aerosol contamination, and 3) the post-processing of temporal sequences to reduce noise and optimize phenological curves (usually a time-series mathematical function) that users can perform. I examined in this first chapter the effect of these three causes of variation on the accuracy of vegetation classifications when data from seemingly equivalent MODIS products are used as predictor variables to map vegetation. The results of this analysis identified the most accurate combination (satellite, pre-processing, and post-processing) to detect vegetation and build annual LULC maps in the CGE.

The second chapter is a modeling of LULC change across the CGE to 1) estimate general forest change trends from 2002 to 2015 and determine its heterogeneity in time and space and to 2) identify the effect of proximate causes of deforestation and reforestation. Due to the high percentage of clouds that plague remote sensing analysis using higher resolution imagery, little was known about forest change trends, proximate causes of forest change, and their spatial and temporal variation across the three countries that share the CGE. Farming is estimated to be the primary proximate cause for approximately 80% of deforestation worldwide (Hosonuma et al. 2012, Kissinger et al. 2012); however, farming includes a range of different land uses that vary across regions or countries. On the other hand, abandonment of agriculture lands is considered to be the primary proximate cause of reforestation globally (Rudel et al. 2002, Lambin et al. 2003, Aide et al. 2013). In this chapter, I created annual LULC maps across the CGE from 2002 to 2015 using a time series of the vegetation index product provided by the MOD13Q1
MODIS product that resulted in the most accurate classification of vegetation cover in the first chapter. By analyzing these maps, I detected a gradual replacement of forested areas with agriculture (mainly grassland planted to support livestock grazing) and secondary vegetation (agriculture reverting to early regeneration of natural vegetation) across the CGE. However, this general forest change trend and the proximate causes of deforestation/reforestation varied temporally and spatially according to the primary administrative areas (countries: Panamanian CGE, Colombian CGE, and Ecuadorian CGE).

The third chapter is a novel approach to estimate the effects of underlying drivers of deforestation and reforestation across the CGE between 2002 and 2015 according to second administrative levels (municipalities). Using the LULC maps built in the second chapter coupled with Bayesian Structural Equation modeling (Bayesian SEM), I estimated the interaction between hotspots of deforestation (areas that exhibit significant spatial correlation of deforestation transitions) to their proximate causes and underlying drivers. I also performed an analogous Bayesian SEM to estimate the interaction between hotspots of reforestation and their proximate causes and underlying drivers. Both Bayesian SEMs were focused on the effects of underlying drivers on the hotspots and were built under the assumption that interactions between direct and underlying drivers would cluster forest cover changes forming hotspots. Eighteen municipalities located on the border between Colombia and Ecuador showed significant aggregations of deforestation hotspots while thirty-four municipalities in three areas of Colombia and the area between the Colombian and Ecuadorian border showed significant clustering of reforestation hotspots. Eleven of these municipalities presented significant
clustering of both reforestation and deforestation hotspots. The Bayesian SEM for deforestation showed that population growth and road density were underlying drivers of deforestation hotspots. The Bayesian SEM for reforestation found that armed conflicts, Gross Domestic Product (GDP), and average annual rain were underlying drivers related to reforestation hotspots.

LITERATURE CITED


Fig. 1.1. The Chocó-Darien global ecoregion (CGE); global location (a) and the countries that share the ecoregion (b).
CHAPTER 2

COMPARING THE ACCURACY OF MODIS DATA PRODUCTS FOR
VEGETATION DETECTION BETWEEN TWO ENVIRONMENTALLY
DISSIMILAR ECOREGIONS: THE CHOCÓ-DARIEN OF SOUTH
AMERICA AND THE GREAT BASIN OF NORTH AMERICA

ABSTRACT

The daily images produced by the MODIS (Moderate Resolution Imaging
Spectroradiometer) sensor aboard the Terra and Aqua satellites have been widely used to
monitor global vegetation. Using these data, the Earth Observing System operated by
U.S. National Aeronautics and Space Administration (NASA) has developed a variety of
MODIS products focused on the monitoring and evaluation of vegetation condition.
These products have three possible sources of variation that can affect the sensitivity of
vegetation detection: 1) orbital and mechanical differences between MODIS sensors
aboard Aqua or Terra, 2) the preprocessing algorithms used to generate multitemporal
cloud-free mosaics (MAIAC or original MODIS algorithm), and/or 3) post processing
algorithms applied by users to optimize vegetation index values derived from temporal
sequences of imagery. We evaluated these sources of variation by comparing the results
of a vegetation classification for two different ecoregions. The accuracies of vegetation
classifications utilizing either the Aqua or Terra MODIS sensors, the MAIAC or original
MODIS preprocessing algorithms, and two common post processing techniques
(Asymmetric Gaussian or Savitzky and Golay function) were compared to determine
which set of techniques or sensors yielded the best results. The ecoregions we chose to
use were the Great Basin of North America and Chocó-Darien of South America. We
compared four different MODIS data products (MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1) as predictor variables using Random Forest as the classification algorithm to generate a land cover map. We found that the accuracy of the vegetation classifications (using Kappa as measure of accuracy) changed significantly depending on the MODIS platform (Terra or Aqua), the preprocessing algorithm (MAIAC or MODIS), and the two postprocessing algorithms for both ecoregions. Our result suggest that comparative analyses are needed to optimize the results when equivalent MODIS products are used in vegetation detection and classification.

INTRODUCTION

Vegetation plays a major role in ecological cycles that govern the life on Earth (e.g. carbon, water, climate, and energy cycles) (Baldocchi et al. 2001, Law et al. 2002, Anderson-Teixeira et al. 2013). Natural and anthropogenic impacts to vegetation condition directly and indirectly influence these cycles (Matthews et al. 2004, Erb et al. 2018); thus, continuous global monitoring of vegetation change is critical to improving our ability to assess current conditions, forecast future scenarios, and, therefore, assist in the mitigation of societal impacts due to changes in ecological cycles (Zhang et al. 2017).

Global monitoring of vegetation using satellite imagery has been carried out primarily by two platforms due to their longevity, regular periodicity of data collection, and open accessibility. 1) The AVHRR (Advance Vey High Resolution Radiometer) has produced the longest continuous record of global image collection from 1978 to the present with a 1 km spatial resolution, five spectral bands, two of which are used for vegetation mapping (0.58 - 0.68 µm, and 0.725 - 1.10 µm), and a daily repeat cycle (Tucker et al. 2005). 2) The MODIS (MODerate resolution Imaging Spectroradiometer),
aboard the Terra and Aqua satellites, have recorded imagery of the entire Earth every 1 to 2 day starting in 2000 and continuing to the present. Both Aqua and Terra MODIS sensors acquire data in 36 spectral bands, including red (0.620 μm - 0.670 μm) and reflected near-infrared (0.841 μm – 0.876 μm) which are specifically used to generate vegetation indices (simple but robust measures of vegetation activity (Running et al. 2004)). MODIS has a variable spatial resolution (250 m, 500 m, and 1000 m) depending on spectral band; however, its red and near-infrared bands have a spatial resolution of 250 m. Due to its improved spatial resolution and radiometric fidelity, MODIS has functionally replaced the AVHRR as the primary global vegetation monitoring sensor (Marshall et al. 2016).

Using MODIS data, The Earth Observation System operated by U.S. National Aeronautics and Space Administration (NASA) generates products for use by scientists and managers interested in monitoring global environmental functions. These products include four vegetation index products with a spatial resolution of 250 m: 1) MOD13Q1 - Terra 16-Day L3 Global 250m vegetation index product (Didan 2015), 2) MYD13Q1 - Aqua 16-Day L3 Global 250m vegetation index product (Didan 2015), 3) MOD09Q1 - Terra Surface Reflectance 8-Day L3 Global 250 m product (Vermote 2015), and 4) MYD09Q1 - Aqua Surface Reflectance 8-Day L3 Global 250 m product (Vermote 2015).

MOD13Q1 and MYD13Q1 have been widely used to monitor vegetation at regional and global scales (Pringle et al. 2012, Sanchez-Cuervo et al. 2012, Aide et al. 2013, Estel et al. 2015, Jiang et al. 2015, Sangermano et al. 2015, Yin et al. 2018, Peng et al. 2017, Qin et al. 2017, Xue et al. 2017, Yin et al. 2017). Both products are 16-day composites of standardized reflectance that have been processed using an established
algorithm, called the MODIS algorithm (Ackerman et al. 1998, 2006), that masks cloud pixels, applies atmospheric correction of aerosol gases, adjusts for the Bidirectional Reflectance Distribution Function (BRDF), and calibrates images to surface reflectance (Didan 2017, Didan and Huete 2006). The MODIS algorithm filters the pixels affected by clouds using a series of reflectance and brightness temperature thresholds based on 14 of the 36 MODIS bands - since clouds are characterized by higher reflectance and lower temperature than the underlying Earth surface (Ackerman et al. 1998, 2006). For non-cloud pixels a quality ranking is employed to identify different levels of aerosol load. Pixels adjacent to clouds will have a higher potential aerosol load and are therefore ranked as low quality. The MODIS algorithm selects the best available pixel value from a 16-day “stack” of images using a criteria of cloud cover, view angle closest to NADIR, and the highest Normalized Difference Vegetation Index (NDVI) and/or Enhanced Vegetation Index (EVI) values (Didan and Huete 2006).

The MOD09Q1 and MYD09Q1 products are estimated surface reflectance for MODIS bands 1 and 2 (red and NIR) from which NDVI can be calculated. These products are 8-Day global cloud-free mosaics with a 250-m resolution (Vermote 2015 and 2015). The MOD09 products use the MAIAC (Multi-Angle Implementation of Atmospheric Correction) algorithm (Lyapustin et al. 2008), which is considered to be an improved process to estimate surface reflectance compared to the standard MODIS algorithms utilized for the MOD & MYD13Q1 products (Lyapustin et al. 2012). The MAIAC algorithm takes advantage of a MODIS 8-day time series using spatial/temporal analyses to detect clouds, retrieve aerosol content, and apply atmospheric corrections (Lyapustin et al. 2008). The MAIAC cloud detection algorithm relies on the fact that
clear-sky images of the same surface have a common textural pattern (even in the presence of snow) defined by surface topography, boundaries of rivers and lakes, distribution of soils and vegetation, etc. This pattern is functionally invariant given the daily rate of global MODIS observations whereas clouds introduce high-frequency random disturbances. Under clear skies, consecutive images of the same surface area have a high amount of covariance, whereas in presence of clouds the covariance is usually low (Lyapustin et al. 2012). Consequently, MAIAC builds cloud masks using pixel level and multi-day covariance analyses, and is considered more accurate than the standard MODIS algorithm over surfaces covered by snow and ice (Lyapustin et al. 2008, 2012). MAIAC was originally developed to improve the cloud classification of MODIS pixels over brighter surfaces such as snow, ice, playa bottoms, and deserts with little to no vegetation cover. These areas are spectrally similar to clouds and can be subject to temperature inversions frequent in the low troposphere during wintertime (Lyapustin et al. 2008). MAIAC has been assessed by others as sensitive to slight reflectance changes in forested areas such as variation in phenology (Ulsig et al. 2017), chlorophyll content (Hilker et al. 2017), and changes due to drought (Bi et al. 2016). Although a higher confidence is assumed for MODIS products that use the MAIAC algorithm, few assessments have compared the effectiveness of MAIAC for vegetation detection versus products that use the standard MODIS cloud and aerosol algorithm.

The effectiveness of MODIS cloud and aerosol algorithms should correlate with the product effectiveness in detecting and monitoring vegetation as well as properly characterizing natural (e.g. phenology) and anthropic (e.g. crop harvests) cycles. For instance, evergreen forest and tropical forest greenness is constant over years (Atzberger
and Eilers 2011); however, more temperate shrublands (Xian et al. 2015), grasslands (Anaya et al. 2009, Nitze et al. 2015, Xian et al. 2015), savannas (Jacquin et al. 2010), wetlands (Silio-Calzada et al. 2017, Fagua and Ramsey 2018) and crops (Nitze et al. 2015) present specific temporal greenness variations. The inclusion of these temporal patterns in an image-based vegetation cover classification process tends to improve accuracy (Jacquin et al. 2010, Verbesselt et al. 2010, Nitze et al. 2015), since different types of vegetation can be spectrally similar at any given time, but have dissimilar temporal spectral profiles. For example, undisturbed tropical forest, secondary vegetation (disturbed, but re-establishing tropical forests), and grassland have similar spectral profiles during the rainy seasons (Helmer et al. 2012), plantations and tropical forest are also spectrally similar (Hansen et al. 2013), and different types of shrublands can be indistinguishable spectrally (Xian et al. 2015). The inclusion of multi-temporal imagery such as these MODIS products can help separate these spectrally similar types due to variable phenologies or management (Jonsson and Eklundh 2002, Hird and McDermid 2009, Eklundh and Jönsson 2016).

MODIS products have been used to detect and monitor vegetation in many areas of research (e.g. land cover change, climate change, biodiversity, wildlife ecology, etc.), and often by non-remote sensing scientists who don’t completely understand nuanced technical issues related to pre-processing algorithms or platform orbital differences that affect data collection times (e.g. Terra vs Aqua). Comparisons between MODIS vegetation products relative to the satellite platform that produced the data have begun to be evaluated (Zhang et al. 2017). Comparisons of the effectiveness of MODIS products that are pre-processed with the MAIAC vs the standard MODIS cloud/aerosol algorithm
for vegetation detection are relatively sparse (Lyapustin et al. 2008). In contrast, the comparison of different optimization processes (data smoothing) to better utilize multi-temporal MODIS data to assess phenological flux in vegetation has received more attention (Hird and McDermid 2009). Here we assess if the ability to detect vegetation cover using different MODIS products varies depending on MODIS platform (Aqua or Terra), the pre-processing algorithm (MAIAC or MODIS algorithm), and post-processing, data optimization technique (Asymmetric Gaussian function or The Savitzky and Golay function). To answer these questions, we utilized these data and techniques to classify general vegetation cover types in two contrasting environments: 1) the Great Basin shrub steppe Global Ecoregion (GB) of North America and 2) the Chocó-Darien Global Ecoregion (CGE) of South America. The Great Basin is the most northerly of the four North American deserts. Detecting vegetation cover type in the GB is challenging due to the lack of vegetation, spectral similarities between different shrub communities and the climate-induced variability of vegetation phenology including extreme shifts in reflectance due to seasonal snow fall. The CGE, by contrast, is a tropical rain forest considered the rainiest place on the planet; another challenge for remote sensing due to the high proportion of clouds and subsequent cloud and aerosol “pollution” in the imagery.

METHODS

Study area

The GB is a contiguous endorheic basin composed of a series of uplifted mountain ranges and associated intervening valleys and dominated by arid and semiarid
bush and grassland vegetation (Bradley and Mustard 2005, WWF 2018) (Fig. 2.1a;b). The GB is bound on the east by the central Rocky Mountains, to the north by the Columbia Plateau, and to the west by the Cascade-Sierra Range. The southern boundary is generally placed at the confluence of the Colorado River drainage and the Mojave Desert of southern California and southernmost Nevada (Morris et al. 2013). Average annual rainfall is 250 mm in the GB but rainfall patterns are heterogeneous; 180 mm of rainfall per year with an inter-annual variance about that mean of 260 mm. Also, rainfall during wet years is frequently three to four times higher than during dry years. (Bradley and Mustard 2005). Thus, native perennial vegetation displays a unique inter-annual variability adapted to rainfall patterns. Vegetation productivity is coupled with variation in precipitation (mostly winter snow) but the range of variance is limited, with a 5% variance in live cover for a 200 mm variance in rainfall (Elmore et al. 2003). On the other hand, the CGE is a lowland area dominated by Tropical Rain Forest located along the Pacific coast spanning southern Panamá, western Colombia, and northeastern Ecuador. the lowlands of the Magdalena river valley in Colombia are also part of the CGE (Olson et al. 2001, Primack and Corlett 2009, WWF 2016) (Fig. 2.1a;c). The CGE is one of 25 global biodiversity hotspots prioritized for conservation due to its high biodiversity and endemism, as well as threats due to deforestation (Myers et al. 2000, Olson et al. 2001, Fagua and Ramsey 2018). The CGE has the highest records of annual average rainfall on the planet with averages above 12,700 mm (1952–1960) in Lloró (Choco-Colombia) and ranging from 8000 mm to 13000 mm across the CGE (Poveda and Mesa 2000).
Identification of vegetation cover in the GB and the CGE

Seven general vegetation cover types found in the GB were analyzed. 1) Basin big sagebrush: most extensive natural vegetation cover in the GB. This is formed by shrublands dominated by basin big sagebrush (*Artemesia tridentata*) and is located mainly in valleys where mean annual precipitation is >200 mm. Subspecies of *Artemisia tridentata* are also found along foothills and within dry mountain meadows and we include them in this cover class. This type can occur as a shrubland or as a shrub steppe when mixed with perennial bunch grasses and forbs (Bradley and Mustard 2008). 2) Herbaceous: these areas are dominated by natural or introduced graminoid or herbaceous forb vegetation and are not subject to intensive management such as tilling but is often utilized for grazing. 3) Forest: Conifer forests are established in the foothills and on the highest elevations of the GB. These forests are dominated by pinyon pine (*Pinus monophylla*) and juniper (*Juniperus occidentalis, Juniperus osteosperma*) at lower elevations and various spruce (*Picea engelmannii* and *Picea pungens*) and fir species (*Abies concolor* and *Abies bifolia*) at higher elevations. The forest cover is characterized by trees generally greater than 5 meters tall, and more than 75% of the tree species maintain their green foliage all year. 4) Deciduous Forest: this forest cover corresponds to areas dominated by trees generally greater than 5 meters tall (Aspen), where more than 75% of the tree species shed foliage in response to seasonal change. 5) Woody Wetlands: areas where forest or shrubland vegetation are periodically saturated with or covered with water. 6) Pasture: areas of cultivated grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or crops, typically on a perennial cycle. 7)
Bare: areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, and other natural accumulations of earthen material.

To train a classification algorithm, we identified 400 250-m sites (corresponding to MODIS pixel size) for each of the seven cover types. We distributed the training sites such that at least 80% of the sites for a specific cover type between 2006 and 2011 were located in relatively protected areas (e.g. National Parks, National Monuments, National Forest, State Parks, Bureau of Land Management lands, etc.) of the GB. The exception was woody wetlands, a rare type in the GB, where only 18 training sites were identified. Training sites were selected using the USGS National Land Cover Database (NLCD) for 2006 and 2011. Sites that were sufficiently large (> 250 m²) and whose NLCD classification did not change between 2006 and 2011 were selected. The USGS NLCD Shrubland Product was used to help identify shrubland training sites. We also utilized the MODIS Active Fire Maps to identify burn areas and, therefore, exclude areas from selection. Selecting training sites in relatively protected areas, as well as sites that did not change between 2006 and 2011, increased our confidence that temporal signatures derived from MODIS over that time period would properly represent those land cover types.

For the CGE, six general vegetation cover types were identified: (1) Tropical Rain Forest: evergreen vegetation dominated by trees that reach over 30 m in height. These forests are the primary natural vegetation cover type in the CGE (Etter et al. 2008, Primack and Corlett 2009, Rangel 2011). (2) Secondary vegetation: natural early regeneration of deforested tropical rain forest. (3) Wetland: swamps and shallow lakes where water saturates the soil generating a particular type of evergreen vegetation that
vary from shrubs to trees. These areas are located along the rivers, and their water levels vary during the year according to rain fall (Silio-Calzada et al. 2017). (4) Grassland: introduced grass species that are used primarily for cattle grazing (Leyva 1993). (5) Crops: agriculture consisting of annual or semiannual crops (corn, sugar canna, plantain, mainly). (6) Palm plantations: plantations of African palm (*Elaeis Guineensis* Jacq). These plantations are relatively stable vegetation because palm requires three years to mature and produce oil and its useful life is about 25 years at which point individuals are replanted with younger palms (Mingorance et al. 2004, FEDEPALMA 2011, Castiblanco et al. 2015). We initially selected 2885 250-m MODIS training sites (1137 for Tropical Rain Forest, 84 for secondary vegetation, 438 for wetland, 647 for grassland, 40 for crops, and 539 for palm plantation) by visually interpreting four types of high spatial resolution images: WorldView-2, Quick Bird, Ikonos and GeoEye-1. However, training site numbers were reduced due to pixels with no data across the MODIS time period for the different MODIS products (see the data reduction: https://s3-us-west-2.amazonaws.com/rsgis-public/MAVD/DATA+AND+CODES.7z).

**MODIS products**

We acquired from the USGS Earth Explorer web site the entire range of data for the four products (MOD13Q1, MYD13Q1, MOD09Q1 and MYD09Q) from 2006 through 2011 in the GB and the entire range from 2009 through 2013 in the CGE. The temporal range of MODIS data were different between the two ecoregions due to the availability of resources (NLCD for the GB and high resolution imagery for the GCE) to identify vegetation cover types within training sites.
The acronyms MOD or MYD refer to the Terra or the Aqua satellite, respectively. MOD13Q1 and MYD13Q1 use the traditional MODIS algorithm to identify the highest quality pixels from 16 daily images for four spectral bands: blue (459 nm – 479 nm), red (620 nm – 670 nm), near infrared (NIR: 841 nm – 876 nm), and mid-infrared (MIR: 2105 nm – 2155 nm). These products also include two indices: EVI and NDVI, as well as layers that estimate vegetation index quality, sensor view zenith, solar zenith angles, individual pixel Julian day, and pixel reliability rankings. The yearly collection of MOD13Q1 and MYD13Q1 consists of 23 temporally sequential periods (periods of 16 days) for every year from 2001 to the present. The MYD 16-day mosaic products are offset from the MOD products by 8 days and together, they functionally generate 8-day cloud free mosaics of the globe (Table 2.1).

The MOD09Q1 and MYD09Q1 use the MAIAC algorithm to identify the highest quality pixels from a sequence of 8 daily images for two spectral bands red (620 nm – 670 nm) and near infrared (NIR: 841 nm – 876 nm); from these two bands NDVI can be calculated but not EVI. Both products provide two layers with band quality control and their yearly collection consists of 46 temporally sequential periods (periods of 8 days) from 2001 to the present. The 8-day temporal period for the MYD and MOD09Q1 products are not offset like the 13Q1 products.

Post process optimization

The red, NIR, and NDVI pixels values from MOD13Q1, MYD13Q1, MOD09Q1 and MYD09Q1 were extracted for each training site in the GB and the CGE across the multi-year time span. The EVI values of MOD13Q1 and MYD13Q1 were also similarly extracted. These data were arranged temporally to form time series for each band, NDVI
and EVI for every training site from 2006 to 2011 in the GB and from 2009 to 2013 in the CGE. Since raw time series data from these MODIS products tend to contain noise due to residual atmospheric and/or BRDF issues (pre-processing algorithms only select the ‘best’ pixel), particularly in cloudy areas like the CGE (Nitze et al. 2015), we applied two filters to every MODIS time series to reduce the variation between observations. These filters consisted of the Asymmetric Gaussian function and the Savitzky-Golay function. Although different post process optimizations can be applied to MODIS time series (e.g., double logistic function, autoregressive running median function, Fourier function, mean-value iteration filter), we selected the asymmetric Gaussian function and the Savitzky-Golay function due to their successful application in other studies and simplicity to estimate (see: Hird and McDermid 2009).

Asymmetric Gaussian is a nonlinear function that generates a smoothed time series using mobile filter lengths or “windows” applied on the original temporally sequential values (Jonsson and Eklundh 2002). The Asymmetric Gaussian function reduced the temporal variation of the individual MODIS bands as well as the vegetation indices by replacing outlier values with estimates based on the window size. We used windows of 5 for every time series after a selection criteria based on the data (Table S2.1), cycles of 23 for the time series from MOD13Q1 and MYD13Q1, and cycles of 46 for the time series from MOD09Q1 and MYD09Q1. The Savitzky-Golay function is a simplified least squares-fit convolution for smoothing and computing derivatives of a set of consecutive values; the convolution is a weighted moving average filter with weighting given as a polynomial (Savitzky and Golay 1964). After a selection based on the data (see Chen et al. 2004) (Table S2.2), we used weighted moving average filters of
five, filter order of two, time scaling factor of 23 for the time series from MOD13Q1 and MYD13Q1, and time scaling factor of 46 for the time series from MOD09Q1 and MYD09Q1.

A measure used to compare the performance of these temporal filters in remote sensing analysis is the comparison of the preservation of the original data after applying a filter function; the best filter should result in the lowest differences between the original data and the filtered data (Nitze et al. 2015). We estimated the Mean Absolute Error (MAE) between the filtered data and the original data for both filters.

\[
MAE = \frac{1}{N} + \sum_{n=1}^{N} |\text{Filtered data} - \text{Raw data}|
\]

where N is the number of observations (Willmott 1982). MAE values were estimated for training sites within the GB and the CGE; the MAE values of EVI were only estimated for MOD13Q1 and MYD13Q1 because EVI for MOD09Q1 and MYD09Q1 cannot be calculated. We used t-tests to evaluate differences between MAE values of NDVI and EVI for each vegetation cover and for every MODIS products in both ecoregions.

Vegetation cover classification

Vegetation cover classifications were developed for the four MODIS products for both ecoregions using the raw data, the data after the asymmetric Gaussian filter, and the data after the Savitzky-Golay filter. All land cover outputs were generated using the randomForest machine leaning classifier found in the R statistical package (Breiman and Cutler 2015). For the GB, we used as predictor variables the 2006 winter mean and 2007 spring-summer mean values of every band and vegetation index since these systems have
defined phenological patterns consistent with temperate vegetation (Nitze et al. 2015). We used the data of the 2006 winter mean and 2007 spring-summer for the classifications since the resources used to identify the training sites (e.g., USGS-NLCD and USGS-NLCD Shrubland Product) corresponded to imagery from February-2005 to October-2007. For the CGE, we used as predictor variables the mean annual values of 2010 for every band and vegetation index (see: Atzberger and Eilers 2011) since the seasonal phenology of tropical vegetation is markedly less variable. We used data of 2010 for the classifications because most imagery used to identify the training sites corresponded to 2009 and 2010.

Validation (accuracy) of each land cover product was estimated using Kappa (K), where K is categorized into the following ranges of agreement: poor $K < 0.4$, good $0.4 < K < 0.75$, excellent $K > 0.75$ (Fielding and Bell 1997). The Random Forest algorithm performed 100 iterations, each with 5-fold cross-validation to estimate $K$. The supplementary material S3 shows the R code to run these classifications. This code runs with the data also found in the supplementary material S3. Finally, we compared the confidence of the vegetation classifications ($K$ estimations after the Random Forest classifications) using a factorial ANOVA in each ecoregion. The objective of the factorial ANOVA was evaluate the effect of the MODIS satellites (Aqua or Terra), the pre-preprocessing algorithm (MAIAC or original MODIS algorithm), and the optimization post-processing (Asymmetric Gaussian function and Savitzky and Golay function) on the $K$ of the vegetation classifications for both ecoregions.
RESULTS

The average NDVI, EVI, and band reflectance values for every vegetation type exhibited a distinct yearly summer-winter seasonal behavior in the GB with annual maxima during spring-summer and minima during winter (Fig. 2.2A-F), while in the CGE, these same curves showed no discernable seasonal pattern for each vegetation type through the years we included (Fig. 2.3A-F). The t-tests of the MAE values for NDVI and EVI for each vegetation type and for each MODIS product did not show a significant difference after applying the asymmetric Gaussian function and Savitzky and Golay function in either ecoregions (P values of the t-tests were > 0.1) (Fig. 2.2G, H; Fig. 2.3G, H); showing that the asymmetric Gaussian or the Savitzky-Golay filters preserved the data equally well.

In the GB, the Random Forest classification of vegetation type resulted in “good” accuracy (K >0.66 and < 0.74) using NDVI, red and NIR values from either MODIS products and both temporal optimizations (Fig. 2.4). However, the factorial ANOVA showed that the two MODIS platforms (F=928.4; P<0.001), the two pre-preprocessing algorithms (F=705; P<0.001), and the optimization postprocessing (F=86.6; P<0.001) had significant effects on the K of the classifications. The vegetation classifications generated from data collected by Terra (MOD09Q1 and MOD13Q1) had a higher K than the vegetation classifications produced by data from Aqua (P values of the Tukey test were <0.001) (Fig. 2.4). Vegetation classifications using raw data from products that use the MAIAC algorithm (MOD09Q1 and MYD09Q1) had higher K values than the classification using raw data from products that use the original MODIS algorithm (MOD13Q1 and MYD13Q1) (P values of the Tukey test were <0.001) (Fig. 2.4).
Vegetation classifications after either Gaussian and Golay temporal optimizations resulted in higher K values for MOD13Q1, MYD13Q1, and MOD09Q1 (P values of the Tukey test were <0.01), with the Gaussian optimization producing slightly better K values for both 13Q1 products (Fig. 2.4). When both EVI and NDVI values are included as predictor variables in the classification using MOD13Q1 and MYD13Q1 data, the accuracy of their vegetation classifications increased across all data treatments (P values of the Tukey test were <0.01) (Fig. 2.4). Vegetation classification using NDVI in the absence of EVI had higher Kappa than the classification using EVI in the absence of NDVI. The most accurate classification for the GB used EVI, NDVI, red and NIR values processed using the Gaussian function optimization on the MOD13Q1 product. The least accurate classification resulted from the MYD13Q1 data when EVI was not added as a predictor.

In the CGE, Random Forest produced vegetation classifications with “excellent” accuracy (K > 0.76 and < 0.88) using NDVI, red and NIR from either MODIS platforms and both temporal optimization (Fig. 2.5). However, the factorial ANOVA found that the MODIS platforms (F= 2437.6; P<0.001), the two pre-preprocessing algorithms (F= 79.8; P<0.001), and the temporal optimization post-processing (F= 38.2; P<0.001) had significant effects on K. Vegetation classifications produced by data from Terra (MOD09Q1 and MOD13Q1) had higher K than the classifications using data from Aqua (P values of the Tukey test were <0.001) (Fig. 2.5). Vegetation classifications using raw data from products that use the MAIAC algorithm (MOD09Q1 and MYD09Q1) also had higher K values than the classification using raw data from products that use the original MODIS algorithm (MOD13Q1 and MYD13Q1) (P values of the Tukey test were
Vegetation classifications after applying the Gaussian function optimization produced higher K values across the four MODIS products (P values of the Tukey test were <0.03) (Fig. 2.5). The Savitzky-Golay function optimization only increased the K of the classifications while using MOD13Q1 and MYD13Q1 data (P values of the Tukey test were <0.01) (Fig. 2.5). When the EVI and NDVI values were included as predictor variables in the MOD13Q1 and MYD13Q1 data, the K values of the vegetation classifications increased using raw data as well as data after both optimizations (Fig. 2.5). Vegetation classification using EVI in the absence of NDVI produced higher Kappa than classifications using NDVI in the absence of EVI in the CGE for MOD13Q1 and MYD13Q1. The highest K for the classifications of vegetation were produced by EVI, NDVI, red and NIR data from MOD13Q1 after the Gaussian function optimization. The least accurate classification were produced by MYD13Q1 data when EVI values were not added as predictors.

DISCUSSION

Our results show that the accuracy of a vegetation classification using these MODIS products changes significantly depending on the MODIS satellite (Aqua or Terra), the pre-processing algorithm (MAIAC or original MODIS), and the temporal optimization post-processing (Asymmetric Gaussian function and The Savitzky and Golay function). These results were similar across two ecoregions with contrasting environmental conditions and highly different vegetation conditions. The differences detected by our assessment show that comparative analyses are needed to optimize the results when different MODIS products can be used in vegetation detection and classification (e.g. MOD13Q1 and MYD13Q1 or MOD09Q1 and MYD09Q1).
The overriding source of variation in accuracy for our vegetation classification was the MODIS platform; the kappa of the classification that used data from the Aqua satellite (MYD) were consistently lower than the Kappa of the classifications from the Terra platform (MOD) for both ecoregions. We can only speculate at this point that these differences in accuracy between platforms could be a product of the different orbital characteristics with data collection by Terra occurring at mid-morning, and Aqua in the mid-afternoon, resulting in significant differences in solar and sensor view angles, which effects the bidirectional reflectance distribution function (BRDF) as well as atmospheric scattering (see Chang et al. 2018). Other factors could include differential degradation for each platform since both are well past their operational life expectancy of six years (NASA 2018, Lyapustin et al. 2014). Our results agree with Zhang et al. (2017) who found differences for NDVI and EVI trends of forest depending on which MODIS platform produced the data.

The classifications produced by the products that use the MAIAC algorithm (MOD09Q1 and MYD09Q1) had higher accuracy than the classification employing the products that use the original MODIS algorithm (MOD13Q1 and MYD13Q1). We expected this result, since the higher performance of the MAIAC algorithm has been already assessed against the original MODIS algorithm using data from the MODIS Terra satellite in tropical rain forest areas of Africa (Lyapustin et al. 2012). We expected that the classification of vegetation in the GB would have a proportionally higher increase in Kappa compared to the CGE since the MAIAC algorithm was developed to improve land cover detection on lands with high brightness produced by snow and/or bare areas. However, we found that the relative differences in Kappa between the classifications
from equivalent products for each algorithm were similar in both ecoregions, but with Kappa in the CGE being consistently higher than the GB.

Although the products that use the MAIAC algorithm produced higher classification accuracies in both ecoregions, slightly higher Kappa was generated by MOD13Q1 data (using the original MODIS algorithm) when EVI values were added as predictor variables. MOD13Q1 data allowed the use of EVI and NDVI as predictors at the same time, improving the accuracy of the vegetation classifications. EVI reduces atmospheric influences and improves the detection of vegetation in dense canopies, such as tropical forest, where NDVI tends to saturate (Huete et al. 1999). Vegetation classification using EVI in the absence of NDVI produced higher Kappa than classifications using NDVI in the absence of EVI in the CGE, confirming the higher performance of EVI in tropical forest regions. Conversely, NDVI has been documented as superior to EVI when detecting vegetation cover with lower biomass and less dense canopy, such as grassland, shrub, crop, and deciduous forests, etc. (Wardlow et al. 2007, Li et al. 2010, Wardlow and Egbert 2010, Breunig et al. 2015). Vegetation classification using NDVI in the absence of EVI had higher Kappa than the classification using EVI in the absence of NDVI in the GB, confirming the higher performance of NDVI in regions dominated by vegetation with low canopy and biomass. We conclude, then that MODIS products that utilize the MAIAC algorithm, but cannot generate EVI will not perform as well as products that can generate EVI in tropical areas.

Both temporal optimizations increased the accuracy of the classifications and were functionally equivalent in preserving the original data (MAE did not have significant differences between optimizations); however, the use of the asymmetric
Gaussian function consistently resulted in a slightly, though sometimes not significantly, higher Kappa value for every MODIS product in both ecoregions. The better performance of the asymmetric Gaussian function over other techniques (including the Savitzky-Golay function) in reducing the noise of NDVI time series data have already been found in other studies (Beck et al. 2006, Hird and McDermid 2009). The reduction in noise after applying the asymmetric Gaussian filter increased the differentiation between vegetation types resulting in the higher kappa values for our assessment.

CONCLUSIONS

By analyzing classifications of vegetation using four equivalent MODIS products, we found that the MODIS platform, the pre-processing algorithm, and the temporal optimization post-processing produced significant differences in the accuracy of detection vegetation classification. These sources of variation need to be considered in remote sensing analyses for vegetation. Thus, a comparative analyses between seemingly equivalent MODIS data products should be carried out to optimize classification accuracies in across any given landscape.

LITERATURE CITED


MODIS. Algorithm Theoretical Basis Document (MOD35). University of Wisconsin - NOAA, Madison, US.


Didan, K. 2015a. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set].

Didan, K. 2015b. MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 [Data set].


FEDEPALMA. 2011. Guía ambiental para el subsector de la agroindustria de la palma de aceite. FEDEPALMA, Bogota DC, Colombia.


Lyapustin, A., Y. Wang, and R. Frey. 2008. An automatic cloud mask algorithm based on
time series of MODIS measurements. Journal of Geophysical Research:
Atmospheres 113:n/a–n/a.
Lyapustin, A., Y. Wang, X. Xiong, G. Meister, S. Platnick, R. Levy, B. Franz, S. Korkin,
impact of MODIS C5 calibration degradation and C6+improvements.
Vegetation Index and Phenology Lab (VIP) and Global Inventory Modeling and
and anthropogenic climate change: incorporating historical land cover change,
Chocó Legalidad Ambiental, Territorial y Derechos Humanos. Human Right-
Programa Suizo para la promoción de la paz en Colombia, Bogotá DC, Colombia.
Morris, L. R., T. A. Monaco, R. Blank, and R. L. Sheley. 2013. Long-term
redevelopment of resource islands in shrublands of the Great Basin, USA.
Ecosphere 4:1–14.


http://wwf.panda.org/about_our_earth/ecoregions/chocodarien_moist_forests.cfm.


Table 2.1. MODIS vegetation index products with the spatial resolution of 250 m.

<table>
<thead>
<tr>
<th>Product</th>
<th>Satellite</th>
<th>Preprocessing Algorithm</th>
<th>Annual temporal sequences</th>
<th>Mosaic product</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOD13Q1</td>
<td>Terra</td>
<td>MODIS</td>
<td>23</td>
<td>16 days</td>
</tr>
<tr>
<td>MYD13Q1</td>
<td>Aqua</td>
<td>MODIS</td>
<td>23</td>
<td>16 days</td>
</tr>
<tr>
<td>MOD09Q1</td>
<td>Terra</td>
<td>MAIAC</td>
<td>46</td>
<td>8 days</td>
</tr>
<tr>
<td>MYD09Q1</td>
<td>Aqua</td>
<td>MAIAC</td>
<td>46</td>
<td>8 days</td>
</tr>
</tbody>
</table>
Fig. 2.1. Study areas. Global location of the ecoregions analyzed in this study (a); the Great Basin global ecoregion (GB) in North America (b) and the Chocó-Darien global ecoregion (CGE) in South America.
Fig. 2.2. Annual time series of NDVI and EVI for vegetation types in the Great Basin global ecoregion using MOD13Q1 data. Annual time series from raw data (a,b), annual time series after the asymmetric Gaussian function (c,d), and annual time series after the Savitzky-Golay function (e,f). Mean Absolute Error (MAE) values of NDVI and EVI for each vegetation type using MOD13Q1 data. These analyses were also performed for MYD13Q1, MODQ09Q1, and MYD13Q1).
Fig. 2.3. Annual time series of NDVI and EVI for the vegetation types in the Chocó-Darien global ecoregion using MOD13Q1 data. Annual time series from raw data (a,b), annual time series after the asymmetric Gaussian function (c,d), and annual time series after the Savitzky-Golay function (e,f). Mean Absolute Error (MAE) values of NDVI and EVI for each vegetation type using MOD13Q1 data. These analyses were also performed for MYD13Q1, MODQ09Q1, and MYD13Q1.)
Fig. 2.4. Accuracies (measured as kappa-K) of the different vegetation classifications for the Great Basin global ecoregion (North America). The classifications were generated using data from the MODIS products MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1. The original MODIS algorithm was used in the preprocessing of the 13Q1 products while the MAIAC algorithm was used in the preprocessing of the 09Q1 products.
Fig. 2.5. Accuracies (measured as kappa-K) of the different vegetation classifications for the Chocó-Darrien global ecoregion (South America). The classifications were generated using data from the MODIS products MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1. The original MODIS algorithm was used in the preprocessing of the 13Q1 products while the MAIAC algorithm was used in the preprocessing of the 09Q1 product.
CHAPTER 3:

GEOSPATIAL MODELING OF LAND COVER CHANGE IN THE CHOCÓ-DARIEN GLOBAL ECOREGION OF SOUTH AMERICA; ONE OF MOST BIODIVERSE AND RAINY AREAS IN THE WORLD.

ABSTRACT

The Tropical Rain Forests of northwest South America fall within the Chocó-Darien Global Ecoregion (CGE). The CGE is one of 25 global biodiversity hotspots prioritized for conservation due to its high biodiversity and endemism as well as threats due to deforestation. The analysis of land-use and land-cover (LULC) change within the CGE using remotely sensed imagery is challenging because this area is considered to be one of the rainiest places on the planet (hence high frequency of cloud cover). Furthermore, the availability of high-resolution remotely sensed data is low for Panamá, Colombia, and Ecuador before 2015. Using the Random Forest ensemble learning classification tree system, we developed annual LULC maps in the CGE from 2002 to 2015 using a time series of cloud-free MODIS vegetation index products. The MODIS imagery was processed through a Gaussian weighted filter to further correct for cloud pollution and matched to visual interpretations of LULC classes from available high spatial resolution imagery (WorldView-2, Quick Bird, Ikonos and GeoEye-1). Validation of LULC maps resulted in a Kappa of 0.87 (SD= 0.008). We detected a gradual replacement of forested areas with agriculture (mainly grassland planted to support livestock grazing), and secondary vegetation (agriculture reverting to early regeneration of natural vegetation) across the CGE. Forest loss was higher between 2010-2015 when compared to 2002-2010. The primary proximate cause of deforestation was grassland in
each country. Grassland also was the main proximate cause of reforestation in Colombia and Panamá while crop was the proximate cause of reforestation in Ecuador

INTRODUCTION

Land-use and land-cover (LULC) change brought about by human development are constantly reshaping natural regions at local, national to global scales (Olson et al. 2001, Lambin et al. 2003, Etter et al. 2006, Hansen et al. 2013). Evaluating these landscape level changes annually within regions where the natural condition is composed of tropical rain forests is difficult due to the high amounts of cloud cover obscuring remote sensing instruments. Tropical rain forests are commonly composed of dense evergreen vegetation with trees reaching 30 m in height, and are located in wet tropical climatic zones where mean annual temperatures exceed 25°C, mean annual rainfall is no less than 1680 mm, and a dry season is absent or less than 2 months (Primack and Corlett 2009). Globally, these regions are suffering significant LULC change (Aide et al. 2013, Hansen et al. 2013, Li et al. 2016) causing much concern due to its potential effect on climatic change, biodiversity loss, hydrologic alteration, soil degradation, and loss of ecosystem services (Lambin et al. 2003, Pfaff et al. 2013). Some national and global estimates have found that deforestation due to LULC change was significantly higher than reforestation in Central and South America (González et al. 2011, Sierra 2013, Hansen et al. 2013, Li et al. 2016). Conversely, other LULC change studies in the same region show a reforestation trend during similar time periods (Sanchez-Cuervo et al. 2012, Aide et al. 2013). Although the methodologies were different, these contradictory results suggest that LULC change could be highly heterogeneous in time and space in the
Tropical Rain Forest domain. It also shows that consistent and accurate information about the LULC dynamic is critical for the management and protection of tropical rain forests.

Tropical rain forests are currently the most biodiverse landscapes on our planet (Myers et al. 2000, Primack and Corlett 2009, Gardner et al. 2012). In South America, tropical rain forests form three well define natural regions; the Amazon Basin, the Brazilian Atlantic Forest, and the Chocó-Darien Global Ecoregion (CGE; also known as the Chocó Biogeographic Region) (Fig. 3.1a). The CGE is a lowland area located along the pacific coast of eastern Panamá, western Colombia, and northwestern Ecuador and has been declared as one of the top 25 global hotspots for conservation priorities (Fig. 3.1b) (Myers et al. 2000, Olson et al. 2001, Primack and Corlett 2009, WWF 2016).

Historically, most of the effort to estimate forest cover and the LULC dynamic have been focused on the Amazon Basin, the largest Tropical Rain Forests in the world (Soares et al. 2006, Etter et al. 2006, Armenteras et al. 2009, Rodriguez et al. 2012, Davidson et al. 2012, Dutra et al. 2012, INPE 2015, Swann et al. 2015). Likewise, LULC dynamic in the Brazilian Atlantic Forest has been well studied (Ribeiro et al. 2009, Calmon et al. 2011, Lira et al. 2012, Dutra et al. 2012). Despite the fact that the CGE is recognized as one of the world's most biologically diverse regions (Gentry 1986, Olson et al. 2001), it has not received the same level of study relative to its LULC dynamic. The countries that share the CGE and their research organizations have conducted studies of the CGE within their own boundaries (González et al. 2011, Ministerio del Ambiente del Ecuador 2012, Sanchez-Cuervo et al. 2012), but these studies have used different methodologies and/or sensors, and do not allow for valid comparisons to evaluate the region as a whole.

Furthermore, regional and global studies of LULC change have been done for large areas
that include the CGE (Lambin et al. 2003, DeFries et al. 2010, Friedl et al. 2010, Aide et al. 2013); however, these analyses have not focused specifically on the CGE and therefore only give a general idea about its LULC dynamics. The study of LULC change focused specifically on the CGE is a fundamental need to guide proper management and conservation.

Past LULC change studies in Neotropical rain forest regions have focused on the gain and/or loss of forest cover (González et al. 2011, Ministerio del Ambiente del Ecuador 2012, Sanchez-Cuervo et al. 2012, Aide et al. 2013, Hansen et al. 2013). Local studies within the Amazon Basin and Brazilian Atlantic Forest ecoregions have accurately identified proximate causes of deforestation and their temporal and spatial variation (Aldrich et al. 2006, Godar et al. 2012, Rodriguez et al. 2012, Davalos et al. 2014). However, within the CGE, much less is known about the proximate causes of deforestation. Farming is estimated to be a proximate cause for approximately 80% of deforestation worldwide (Hosonuma et al. 2012, Kissinger et al. 2012); however, farming, in general, includes a variety of different land uses that change across regions or countries depending on a host of environmental and social/economic factors such as soil quality, precipitation, temperature, topography, infrastructure, technology, and/or traditional knowledge. The United Nations Framework Convention on Climate Change negotiations has encouraged developing countries to spatially map proximate causes of deforestation (Viña and Leon 2014, UNFCCC 2015). While some studies have shown reforestation trends due to apparent abandonment of agriculture lands, this reforestation process has only been slightly studied in the Colombian CGE (Sanchez-Cuervo and Aide 2013).
An analysis of LULC change across the CGE is a challenge when remote sensing imagery is used because this area is considered one of the rainiest on the planet with an annual average rainfall between 8000 to 13000 mm (Poveda and Mesa 2000). Furthermore, the availability of high spatial resolution remote sensing data before 2015 is low for Panamá, Colombia and Ecuador. Consequently, the satellite images that are available (Landsat, Satellite Pour l'Observation de la Terre (SPOT), and RapidEye, for example) usually have a high percentage of cloud cover, making it difficult for the development of regional land cover maps. Nevertheless, over the past few years, new methodologies using MODIS (MODe rate resolution Imaging Spectroradiometer) data have generated standard periodic cloud-free products aimed at monitoring vegetation across the globe. Merging these MODIS products with available high spatial resolution (e.g. WorldView, Ikonos, QuickBird, GeoEye) imagery used as a reference data source with learning algorithms (e.g. Random Forest) (Clark and Aide 2011, Yin et al. 2017) offers potential for studying region-wide LULC in areas like the CGE.

We have applied a combination of these methodologies to multi-temporal MODIS imagery to generate yearly LULC maps across the CGE from 2002 to 2015. Our aim was to analyze LULC temporal dynamics across this ecoregion and address the following objectives: 1) Evaluate LULC change trends in the CGE and determine its heterogeneity in time and space. 2) Spatially identify proximate deforestation causes, reforestation transitions (cover types that represent secondary forest-like vegetation), and quantify their change in time and space. We discuss the types of information that are useful for conservation of biodiversity in the CGE relative to its administrative organization (countries).
METHODS

Study area

The Chocó-Darien Global Ecoregion (CGE) is a lowland area of Tropical Rain Forest located along the Pacific coast from Panamá, through Colombia, and into northwestern Ecuador. The CGE also includes the lowland of the Magdalena river valley (Fig. 3.1b) (Olson et al. 2001, Primack and Corlett 2009, WWF 2016). The CGE covers 194,737 km² and is recognized as one of the world's most biologically diverse tropical rain forest (Gentry 1986, Myers et al. 2000, Olson et al. 2001). The CGE became separated from the Amazon Tropical Rain Forest by the uplift of the Andes beginning around 25 million years ago. As a consequence, groups of endemic species emerged producing a significant impulse of diversification. Another substantial array of new species arose because of the relatively recent formation of the Isthmus of Panamá (3 million years ago), an extraordinary geological event that separated Atlantic and Pacific oceans forming a land bridge for plant and animals from North and South America (Gregory-Wodzicki 2000, Primack and Corlett 2009). Presently, it is estimated that the CGE has about 11,000 species of vascular plants (2,250 endemics), 900 species of birds, 350 species of amphibians (210 endemics), and 210 species of reptiles (63 endemic) (Myers et al. 2000, WWF 2016). It was estimated in the year 2000 that the remaining tropical rain forest within the CGE covers approximately 24% of its original distribution (Myers et al. 2000). Due to this level of deforestation and the high number of endemic species, the CGE was declared as one of the 25 global hotspots for conservation priorities (Olson et al. 2001). According to the World Wildlife Fund (2016), the CGE is formed by four smaller terrestrial ecoregions: Chocó-Darién Moist Forests (73028.6 km²), Eastern
Panamanian Montane Forests (2632.4 km²), Magdalena-Urabá Moist Forests (76396 km²), and Western Ecuador Moist Forests (33861.1 km²). Also, sections of three mangrove ecoregions are found along the CGE coast: South American Pacific Mangrove (6252.4 km²), Amazon-Orinoco-Southern Caribbean Mangrove (702.9 km²), and a small area of Mesoamerican Gulf-Caribbean Mangrove (50.2 km²) (Fig. 3.1b). We did not include mangrove ecoregions in our LULC change analysis because they are small areas of marine wetlands (3% of the CGE) and our study was focused on terrestrial tropical rain forest.

**LULC maps**

We generated a temporal set of LULC maps based on a Random Forest classification (Breiman 2001) in which we modeled a categorical response variable that identified eight LULC classes. Random Forest is an ensemble learning algorithm that constructs multiple classification trees (e.g. 500 individual trees) by bootstrapping samples from an input data set, and combines the predictions from all the trees to identify a modal response. Random forest is one of the most robust statistically-based classification techniques and presents two main advantages for our analysis; it has low sensitivity to the overfit produced by collinearity among predictors and allows for use of different types of response and predictor variables (e.g. numerical, binary, categorical) in the classification process (Cutler et al. 2007, Breiman and Cutler 2015, Matsuki et al. 2016).

The mapping of these LULC classes was accomplished by training MODIS cloud-free temporal image mosaics using 22 sampling sub-regions covering 20,708.6 km² of total land area within the CGE. These 22 sampling sub-regions corresponded with
locations of available high-resolution imagery. The cloud-free MODIS vegetation index products MOD13Q1.V006 (tiles h10v07, h10v08, h10v09, and h09v09) were downloaded from the NASA Distributed Active Archive Center and processed to transform the standard sinusoidal projection to WGS84 geographic coordinate system. This transformation resulted in a calculated pixel size of 231.3 m².

*Response variable (LULC classes)*

Training samples for each LULC class were collected by visually interpreting four types of high spatial resolution images: WorldView-2, Quick Bird, Ikonos and GeoEye-1. To improve visual interpretation of LULC classes, the multispectral bands from these sensors were fused to their corresponding panchromatic band (Table S3-1). We reviewed previous regional LULC studies within the CGE to help define our LULC classes (Friedl et al. 2010, González et al. 2011, Ministerio del Ambiente del Ecuador 2012, Sanchez-Cuervo and Aide 2013). From these studies, we established eight general LULC classes.

1) Woody vegetation: this type of vegetation included tropical rain forest with trees taller than 30 m, secondary vegetation (shrubs and smaller trees) as well as mosaics of both. This is the primary natural cover type that occurs within the CGE (Etter et al. 2008, Primack and Corlett 2009, Rangel 2011). Initially, forest and secondary vegetation were established as two different LULC classes; however, the Random Forest classification could not adequately separate them. Likewise, we created a mixed woody class (pixels with 20%–80% of woody and the rest the pixel cover by agricultural land), but the Random Forest classification could not separate this cover type either.
Consequently, after doing a Fuzzy accuracy analysis (Lowry Jr. et al. 2008) of a preliminary classification, forest and shrub were merged into a woody vegetation class.

(2) Wetland: the CGE has a complex of river basins with swamps and shallow lakes ("ciénagas") covering large areas along the rivers. Wetland areas were absent in previous LULC work performed within portions of the CGE (Etter et al. 2006, González et al. 2011, Sanchez-Cuervo et al. 2012, Aide et al. 2013) and as a result have been markedly underestimated in global maps (Friedl et al. 2010, Hansen et al. 2013).

(3) Grassland: introduced grass species which are used primarily for cattle grazing (Leyva 1993). Within the CGE, large areas of native grasses do not occur as natural vegetation (Etter et al. 2008, Rangel 2011).

(4) Crops: agriculture consisting of annual or semiannual crops (corn, sugar canna, plantain, mainly).

(5) Palm plantations: Extensive areas of the CGE have been cultivated with African palm (*Elaeis Guineensis* Jacq) (Castiblanco et al. 2015). These palms take about three years to mature and produce oil. The useful life of a palm plantation is about 25 years at which point plantations are replanted with younger palms (Mingorance et al. 2004, FEDEPALMA 2011). In terms of remote sensing, this relatively stable structure of palm plantations allowed its identification as a LULC class using our imagery resources.

(6) Settlements and infrastructure.

(7) Continental waters including rivers and lakes.

(8) Bare areas; this class was not taken into account in the final analysis due to its low representation.
The 192,924 km² of land corresponding to the CGE was divided into square sample areas of 231.3 m x 231.3 m to match the MODIS pixel size. Based on this grid, a stratified sampling was applied to the area intersecting the aforementioned high spatial resolution images as follows: we visually identified sample squares with 100% of any of the eight LULC categories. We then superimposed a second grid of 1 km² as spatial filter to select one square of 231.3 m² for every 1 km² square. This spatial filter ensured that sample sites were separated by 693 m or more. By doing this, we identified 18,559 sample sites classified as one of the eight LULC classes. To estimate the error rate for the visual interpretation, we compared our visual interpretation with the visual interpretations of the Corine Land cover project for Colombia (IDEAM 2010), which used many resources (high spatial resolution imagery, aerial photos, Landsat, and field visits) to reach the best possible visual interpretation of land cover. We coupled 375 of our MODIS sampling sites to the corresponding interpretations from the Corine Land Cover effort for the years 2002, 2003, and 2007. The agreement between both interpretations resulted in a kappa of 0.93 (Accuracy = 0.9519), showing a high level of consistency between both interpretations.

**Predictor variables**

Five MODIS-based predictor variables were generated from the MOD13Q1 product (16-Day L3 Global 250 m Vegetation Indices). The MOD13Q1 product provides the highest quality pixels from 16 daily images for four spectral bands: blue (459 nm - 479 nm), red (620 nm –670 nm), near infrared (NIR: 841 nm – 876 nm), and mid-infrared (MIR: 2105 nm –2155 nm); as well as two indices: Enhanced Vegetation Index (EVI) and the Normalized Difference Vegetation Index (NDVI). EVI reduces atmospheric
influences on vegetation detection and improves identification of vegetation with dense
canopies, such as tropical forest, where NDVI tended to saturate (Huete et al. 1999).
However, we used both EVI and NDVI as predictors because NDVI has been equivalent
or better than EVI detecting vegetation covers with low biomass and canopies, such as
grassland, shrub, crop, and subtropical deciduous forests (Wardlow et al. 2007, Li et al.
2010, Wardlow and Egbert 2010, Breunig et al. 2015). The MOD13Q1 product also
provides layers that estimate vegetation index quality, sensor view zenith, solar zenith
angles, individual pixel Julian day, and a pixel reliability ranking. For our analysis we did
not use the blue spectral band due to its lower spatial resolution, 462.7 m². The yearly
collection of MOD13Q1 data consists of 23 temporally sequential periods (365 days and
16 days per period) for every year from 2001 to the present. We utilized the entire range
of data from 2001 through the end of 2015, for a total of 345 individual measurements of
red, NIR, MIR, NDVI and EVI for every 231.3 meter pixel in the CGE. Although the
MOD13Q1 product attempts to evaluate pixel quality as a function of radiometric and
atmospheric conditions (cloud interference), these data can still contain anomalies that
are caused by factors not relevant to the amount of photosynthetically active surface
cover, namely atmospheric conditions. To account for these anomalies and therefore the
uncertainty within the vegetation index products, we applied a Gaussian weighted filter to
the 23 temporal periods for each year and for each of the five spectral variables. This
filter reduced the variation of the MODIS bands and indices and replaced outlier values
with estimates calculated by the Gaussian weighted series (Fig. 3.2). We used the output
of the Gaussian weighted filter to estimate an annual mean for each band and index, and
we used these means as predictor variables (Hird and McDermid 2009, Nitze et al. 2015).
This analysis was performed using TerrSet Geospatial Monitoring and Modeling Software from Clark Labs (Eastman 2017), with each year from 2001 to 2015 representing a time series cycle (a total of 15 time series cycles) with a temporal filter length or “window” of 5. In addition to the MODIS-based predictor variables, we included the SRTM90 (NASA Shuttle Radar Topographic Mission) elevation data and its corresponding slope values as ancillary data in support of the image classification process (Varga et al. 2014, Li et al. 2017). Elevation and slope have been found to affect the type of land cover that occurs in a specific area; forest tends to be preserved in places with higher altitude and slope (due to a more difficult access) while crops and palm plantations occur in places with low slope (FEDEPALMA 2011, Fagua et al. 2013). As well, the wetlands in the CGE are located in areas with an altitude near or under the sea level (Rangel 2010, 2011). Additionally, the elevation data of SRTM90 could be affected by densely vegetated areas (Kellndorfer et al. 2004, Brown et al. 2005, Tanase et al. 2015, O’Loughlin et al. 2016) and wetlands (Simard et al. 2006). SRTM has a spatial resolution of 90 m² and was, therefore, resampled to 231.3 m² to match the MOD13Q1 pixels using a bilinear interpolation. From this resampled digital elevation model, SRTM elevation and topographic slope were extracted for each training side.

**Random Forest classification**

A total of 18,559 training sites located within the 22 sampling areas were classified visually into our eight LULC classes using high resolution imagery. The training site database, therefore, contained the interpreted LULC class as well as the predictor variables of temporal spectral and vegetation index values (NDVI and EVI) for each year along with SRTM values and topographic slope where columns represented
the response and predictor variables and rows consisted of the 18,559 observations (Table S3-2). The RF algorithm operates by constructing a large number of decision trees from random subsets of predictor variables and the resulting classification consists of the modal response of all trees for a particular outcome (Cutler et al. 2007, Freema et al. 2016). We used the R statistical packages ‘randomForest’ (Breiman and Cutler 2015) and “ModelMap” (Freema et al. 2016) to generate our yearly LULC maps (The R code to build a map is available in supplementary material S3-3 Codes). The randomForest utility in R generated a default number of trees (500), using a 80% subset of the samples (training subset) for every bootstrap iteration and the square root of the number of predictors as the number of predictors used to identify a split at each node. This RF model can provide accuracy estimates using OOB (or Out-of-Bag, a first independent subset from the training data) (Cutler et al. 2007) or using a second independent group formed by the 20% of samples that were not used as training subset (testing subset). We reported the kappa from the second independent group to reduce a possible overestimation in the accuracy (Witten et al. 2013). Accuracy estimates included Kappa (K), which was categorized into the following ranges of agreement: poor $K < 0.4$, good $0.4 < K < 0.75$, excellent $K > 0.7575$ (Fielding and Bell 1997), as well as percent omission and commission errors for each LULC class.

A RF classification is accomplished using available training data and therefore is subject to training data distribution amongst the different response classes. The dominant land cover category in our study area consisted of the woody vegetation class. Of our 18,559 sample sites, 14,228 samples (76%) consisted of the woody vegetation class. In order to detect the potential impact of this large sample size relative to other land cover
categories on the accuracy of minority classes, we randomly reduced the samples representing the woody vegetation class from 14,228 to 1,144 to match the sample size of the second most prevalent class, Grassland (Mellor et al. 2015) (S3-4 Table). RF classifications were run on both sample sets and the cross validation results showed that K for the original data was 0.872 (S3-5 Table) and K for the reduced samples of woody vegetation class was 0.876, commission and omission errors were similar for both sample distributions (S3-4 Table). Consequently, we decided to use the original sample set of 18,559 training samples. Using this methodology, we developed a RF-based LULC classification for each of the 15 years using SRTM values, topographic slope, combined with MOD13Q1 MODIS data for each year as predictors for that year. We used the R package ‘ModelMap’ which uses the ‘RGDAL’ libraries to generate LULC maps using the RF model outputs. For all of the 15 individual years, our LULC maps reached a high accuracy of K=0.872, with a standard deviation across all years of 0.008

Woody vegetation split

As we describe before, woody vegetation could not be separated into forest and shrub (secondary vegetation) classes. These two LULC classes had similar spectral and NDVI - EVI signatures provided by the MODIS data and consequently, RF classification could not separate them. This spectral similarity is common in tropical rain forests when using other multi-spectral sensors such as Landsat (Helmer et al. 2012). For that reason, a final refinement after the RF classifications was applied to the annual LULC maps generated from the annual sequence of MODIS imagery. Pixels classified as woody vegetation were converted to forest when that pixel was classified as woody vegetation for every year of our sequence (2001-2015). On the other hand, if a pixel was classified
as woody vegetation on the last year of our sequence (2015), but in previous years that pixel was classified as another type such as grassland, crop, or palm plantation, it was recoded as secondary vegetation. This seemed like a logical method of splitting the woody vegetation class into forest and secondary vegetation since the year-to-year accuracy of each LULC map was high and the average time for a forest canopy to reach maturity (tall forests) in other neotropical rain forest is between 190 - 217 years (Lieberman et al. 1985, Saldarriaga et al. 1988). Likewise, shrubby vegetation typically takes over 20 years in areas of tropical rain forest to develop arborescent structures (Brown and Lugo 1990, Guariguata and Ostertag 2001, Schlawin and Zahawi 2008). In other words, it is improbable that forests converted to a farm-like land-use will reach a forest-like stage in 15 years or less. Consequently, these pixels were considered secondary vegetation (landscapes converting from a farm-based land-use to natural vegetation). Based on this logic, we developed 14 final LULC maps (2002 to 2015) that included eight LULC classes: forest, secondary vegetation, wetland, grassland, crops, palm plantations, settlement, and continental water. The time series maps started in 2002 due to our methodology for splitting woody vegetation needs an initial sequence of annual maps, 2001-2002. To test accuracy of the secondary vegetation class, 191 pixels mapped as secondary vegetation were randomly selected and independently classified using visual interpretation of the available high resolution images. The accuracy of our secondary vegetation classification averaged 84.2% with a standard deviation across the years of 10.4.
Analysis of LULC change

LULC trends

To determine LULC trends, we estimated non-parametric Pearson correlations between the area occupied for every LULC class and the corresponding year in the annual sequence. A significant positive Pearson’s correlation coefficient indicates a significant increase in the trend of that specific LULC class while a negative correlation coefficient indicates a significant reduction in area as years progress. Pearson correlations were estimated at two spatial levels: 1) the entire CGE, and 2) the areas of the CGE corresponding to the countries that share this global ecoregion (Panamá, Colombia and Ecuador). Pearson correlations also were calculated in two time periods for each of the two spatial levels: from 2002 to 2010 and from 2010 to 2015. We chose these two time periods because we found that woody vegetation, forest, and secondary vegetation trends (main objectives of our analysis) significantly changed around 2010. Additionally, other studies that include the CGE analyzed LULC change from 2001 to 2010 (González et al. 2011, Sanchez-Cuervo et al. 2012, Sierra 2013, Aide et al. 2013): therefore, an analysis of temporal change between 2002-2010 provided an opportunity to compare our results with other studies.

Proximate causes of deforestation and farm conversion to secondary vegetation

To quantify the proximate causes of deforestation, we identified the following transitions. (1) Deforestation due to cattle grazing operations as indicated by areas of forest or secondary vegetation replaced by grassland. (2) Deforestation by annual or semiannual crops as indicated by areas of forest and secondary vegetation replaced by crops. (3) Deforestation by extensive palm plantations as indicated by areas of forest and
secondary vegetation replaced by palm. (4) Deforestation by infrastructure and urban expansion as indicated by areas of forest and secondary vegetation replaced by human development. Conversely, we quantified the conversion from every farming land-use (grassland, crop and palm plantations) to secondary vegetation as reforestation transitions. These deforestation and reforestation transitions were estimated for two time sequences 2002-2010 and 2010-2015 (using the first year of every sequence as the base year).

RESULTS

In 2002, 63.9% of the CGE (120,246 km²) was classified as woody vegetation (forest and secondary vegetation combined). In 2010, woody vegetation increased to 68.5% (128,801.8 km²), and in 2015, 65.5% (123,320.6 km²). In other words, woody vegetation increased 4.6% between 2002-2010 and reduced 3% between 2010-2015 (Fig. 3.3a). For woody vegetation, 90.4% was identified as forest in 2002, 72.1% in 2010 and 67.6% in 2015. LULC trends for the entire CGE shows that secondary vegetation increased significantly from 2002 to 2010 (R=0.94, P<0.01) whereas forest (R=-0.96, p<0.001) and agriculture (R=-0.64, P<0.05 for grassland; R=-0.64, p<0.06 for crop; R=-0.89, p<0.02 for palm) decreased showing a progressive replacement of forest and agriculture with secondary vegetation (Fig. 3.3a; Table S3.7). Some of these trends changed between 2010 to 2015; woody vegetation declined but not significantly, forest maintained its decreasing trend (R=-0.98, p<0.001), and grassland increased (R=0.85, p<0.02) while the other agricultural land use trends did not show significant trends (Fig. 3.3). These results show that deforestation transitions (changes from forest or secondary vegetation to farm covers) was higher between 2010-2015 than 2002-2010 and indicate
that grassland was the main land cover that replaced woody vegetation (forest and secondary vegetation) between 2010-2015.

When LULC trends are compared between political divisions during 2002-2010, we found that woody vegetation increased in the Colombian and Ecuadorian CGE during 2002-2010 (R=0.78, p< 0.01; R=0.64, p< 0.05) but did not show a significant trend in the Panamanian CGE. Secondary vegetation increased significantly in every national territory (R=0.95, p< 0.01 for Panamá; R=0.91 p< 0.01 for Colombia, and R=0.85, p< 0.01 for Ecuador) while forest decreased (R=-0.89, p< 0.01 for Panamá; R=-0.95 p< 0.001 for Colombia, and R=-0.94, p< 0.01 for Ecuador). Grassland decreased in the Colombian CGE (R=-0.68 p< 0.04) but it did not show a significant trend in Panamá and Ecuador. Crops decreased in the Ecuadorian CGE (R=-0.63 p< 0.05) and palm plantation decreased in the Colombian CGE (R=-0.86 p< 0.01). Between 2010-2015, some of the previous trends changed. Woody vegetation decreased significantly in Panamanian and Colombian CGE (R=-0.94, p< 0.01; R=-0.89, p< 0.02), forest maintained its decreasing trend in all three countries (R=-0.85, p< 0.02 for Panamá; R=-0.91 p< 0.01 for Colombia, and R=-0.96 p< 0.02 for Ecuador), grassland increased in Panamá and Colombia (R=0.86, p< 0.03 for Panamá; R=0.89 p< 0.02 for Colombia, and R=0.63 p< 0.03 for Ecuador), and crops tended to decrease non significantly in the three countries (Table 3.1).

The analysis of land cover transition showed that grassland was the most frequent proximate cause of deforestation between 2002 to 2010 for the entire CGE (63%) and for each country (73% in Panamá, 65% in Colombia, and 58% in Ecuador) (Fig. 3.4a; Table S3.8). Grassland was also the most frequent land cover that change to secondary forest
(reforestation) across the entire CGE (50%), in Panamá (65%), and Colombia (58%), while crops were the most frequent cover to convert to secondary vegetation in Ecuador (55%) (Fig. 3.4b; Table S3.8). Subsequently, from 2010 to 2015, LULCL transitions also showed that grassland was most frequent proximate cause of deforestation across the CGE (73%) as well as in every country (94% in Panamá, 76% in Colombia, and 59% in Ecuador) (Fig. 3.5a; Table S3.9). Grassland was also the most frequent land cover that converted to secondary vegetation during 2010-2015 for the CGE (47%) and in two countries (68% to Panamá and 53% to Colombia). In Ecuador, crops to secondary vegetation was the highest reforestation transition (55%) (Fig. 3.5b, Table S3.9). The net deforestation was almost two times higher during 2010-2015 (15,145 km^2) than 2002-2010 (7,228 km^2) in the CGE; this pattern was similar in every country (Figs. 3.4a, 3.5a). Conversely, net reforestation was higher between 2002-2010 (17783 km^2) than 2010-2015 (9120 km^2) in the CGE. As well, reforestation tended to be higher in every country during 2002-2010 compared to 2010-2015 (Figs. 3.4b, 3.5b).

DISCUSSION

LULC change trends have not been temporally homogeneous across the CGE. We identified an overall increase in woody vegetation driven mainly by an increase in secondary vegetation between 2002-2010, this increase, however, ceased between 2010-2015. Conversely, grassland showed an overall decrease between 2002-2010 and an overall increase between 2010-2015. These trend shifts around 2010 were similar between the Colombian and Ecuadorian portions (92% of CGE land) and suggest that underlying drivers could affected LULC change across the CGE. During the first decade of this century (2000-2010), the Colombian and Ecuadorian agricultural sectors declined,
thus reducing cultivated (grassland, crops, palm) area and allowing for the growth of secondary vegetation. The Colombian agriculture sector decreased 1.1% during this period (Minsalud 2011, Buitrago 2013, Marrugo 2013) while the Ecuadorian agricultural sector decreased by 1.8%. This was remarkable in Ecuador because its national agricultural sector had grown by 6.1% between 1990-2000 (BCE 2010). Increases in secondary vegetation were also found in several developing countries within Latin America during the first ten years of the present century (Aide et al. 2013). Some scholars have claimed that the globalization of markets negatively impacted the agriculture sectors of these countries resulting in abandonment of farm land and eventual reforestation (Aide and Grau 2004, Meyerson et al. 2007). Subsequently, from 2010 to 2015, Colombia and Ecuador showed a remarkable acceleration in their economic growth due to the global increase in the price of mining products (specially, oil, coal, energy, and gold). This acceleration could have a positive impact on all sectors of their economies (improving transportation routes, infrastructure in general, market for farming products) intensifying the use of farming areas. In Colombia, gross domestic agricultural product grew from negative values in 2009 to 5.5% in 2014 (Finagro 2014) and two important routes that cross large areas of the Colombian CGE were built (the route Tumaco-Tuquerres in Nariño department and the route Virginia-Quibdó in Risaralda and Choco departments). These routes correspond to some of the deforestation that we identified in our maps. In Ecuador, gross domestic agricultural product grew 6% from 2009 to 2013 (ProEcuador 2014). This increase in agricultural production should have had a negative effect on the regeneration of secondary vegetation thus increasing deforestation as our results indicate. Some authors have claimed that reforestation in the Colombian CGE
territory during 2002-2010 occurred principally due to land abandonment caused by internal armed conflicts in Colombia (Sanchez-Cuervo and Aide 2013). However, we found the same pattern in Ecuador during the same period (2002-2010), a country with no armed conflict. The regrowth of secondary vegetation across farming areas was proportionally higher in the Ecuadorian CGE compared to the Colombian CGE. Additionally, we found that reforestation has decreased significantly between 2010-2015 in the Colombian CGE while the armed conflict was still occurring and this area had a strong presence of the two main guerrilla groups in Colombia. This evidence suggests that economic growth could have a greater influence on the balance of deforestation and reforestation compared to local phenomenon such as armed conflicts. The Panamanian economy is not based on agriculture (main sectors in Panamá are transportation, communication, market, services and banking) (Fisher 2015). This could explain the flat trend for woody vegetation in the Panamanian CGE through 2002-2010; however, reductions in woody vegetation, secondary vegetation and forest also occurred in the Panamanian CGE during 2010-2015 indicating increased human land use driven by economic growth during this time period. Panamá had the highest economic growth in Latin America between 2000-2013 (7.2% on average) (Fisher 2015, WBG 2017). Only forest had an overall consistent temporal trend cross the CGE, and tended to decline during both time periods across the three countries. Our split of woody vegetation into secondary vegetation and forest allowed us to identify this progressive replacement of well-preserved forest primarily by grassland and secondary vegetation. Forest reduction has been documented in the Colombian CGE (González et al. 2011) and in the
Ecuadorian CGE (Ministerio del Ambiente del Ecuador 2012) between 2001-2010 using Landsat data and discriminations between forest and secondary vegetation.

Agricultural expansion was the most frequent proximate cause of deforestation during both time periods across the CGE; 98% of deforestation due to agricultural conversion and 1% by the establishment of settlement and infrastructure. Our results agree with other reports showing agricultural expansion as the main proximate cause of deforestation in the tropics (Gibbs et al. 2010, Hosonuma et al. 2012). In addition, we analyzed sub-categories of agricultural proximate causes of deforestation (grassland, crops, and palm plantations) and found that grassland conversion was the main cause of deforestation across the CGE during both time periods. Extensive cattle grazing is a main agricultural activity for the areas corresponding to Magdalena-Urabá Ecoregion (46% of the CGE land and this entire sub-ecoregion is in Colombia) and to Western Ecuador Ecoregion (17% of the CGE land and the entire sub-ecoregion is in Ecuador) (Fig. 3.1b) (INER 2003, MAG 2003, Castillo 2014, PNUD 2014). Other causes that explain the gradual replacement of forest by grassland and secondary vegetation cross the CGE during both time periods (2002-2010 and 2010-15) are the colonial process and the land possession policies of Colombia, Ecuador and Panamá. Basically, colonists are required to prove they are using land in order to become landowners. The cheapest and fastest method to prove land use is to convert forest to grassland. However, many of these deforested areas are underutilized and they consequently revert to secondary vegetation. Evidence supporting this hypothesis has been documented by other scholars; Davalos et al. (2014) found that forest conversion to grassland in several areas of the Amazon within Colombia were not related to beef production. They concluded that colonists were
removing forest to prove active land use, gain ownership of the property, and wait for land values to increase (Davalos et al. 2014). IGAC (2015) found that deforestation after colonization in areas with fragile soils, such as Choco-Darien ecoregion of the CGE (Fig. 3.1b), resulted in 38% of soils becoming unproductive in Colombia (IGAC 2015).

Historically, land possession has been a main source of economic and political power in Colombia and Ecuador resulting in land conflicts (Flórez et al. 2012, Sierra 2013). Consequently, future pressure on forest areas across the CGE could increase since this area hosts the largest population of colonist in Panamá and Ecuador (Argüelles 2010, Sierra 2013).

Reforestation transitions were also not homogeneous across the CGE. Grassland to secondary vegetation was the highest reforestation transition across the CGE; however, it was different in the Ecuadorian CGE (16% of the CGE land) where crop conversion to secondary vegetation was the highest reforestation transition during both time periods (2002-2010 and 2010-2015). Agriculture consisting of annual or semiannual crops (corn, plantain, coffee, rice) was the principal proximate cause of deforestation in the Ecuadorian CGE during 1990 and 2000 (Sierra 2013). Manabí, Esmeraldas (the south side), and Santo Domingo (the largest Ecuadorian provinces in the CGE) are provinces considered to be specialized in crop production, but cattle has increased since 2000 in this region while crops have decreased; presently, about 50% of the land consists of cultivated grassland and 18% by crops (PEC 2011, 2011b) and are consistent with our results.

Some scholars have claimed that Palm plantations were one of the main proximate causes of deforestation in the CGE (Mingorance et al. 2004, Goebertus 2008, Montaño 2008, Sabogal 2013). Our results showed that Palm plantation was the third
most significant proximate cause of deforestation across the CGE and its effect on forest and woody vegetation was different in every country; palm was the second proximate cause of deforestation in Colombia and the third in Ecuador. Panamá did not have palm plantations and thus it was not a factor in that country. Also, the reduction of forest as a result of palm plantations is substantial lower than the reduction produced by grassland cross the CGE. The zones that we identified as areas with palm plantation in Colombia coincide with the municipalities identified as areas with palm plantations by the Colombian Federation of Palm Farmers (FEDEPALMA 2017). Specifically, we found that palm plantations were concentrated in three areas: Near the Colombia-Ecuador border, around the Urabá gulf, and through Magdalena Valley. As well, we found that palm plantations are partially spread cross the Ecuadorian CRB, which agrees with Ecuadorian studies about palm distribution; the Ecuadorian CGE is the region with the most palm plantations in this country and these cultivated areas have doubled between 2000 and 2010 (Potter 2011).

Mining for mineral resources has been a primary historical economic activity along the Pacific coast of Colombia within the CGE. Due to the increasing price of gold (16% of annual increase in average), silver (21% of annual increase in average) and platinum (11% of annual increase in average) in international markets between 2001 and 2013, mining has increased with little governmental control in the Colombian Choco-Darien. Miners cut down forest, turn the soil, and separate minerals from soil material using mercury with water from nearby rivers. Additionally, areas are deforested to build roads to transport machinery (Mosquera 1978, Zapata 2013). Frequently, this mining activity occurs in smaller areas than our MODIS pixels size (231.3 m²); consequently, the
spatial scale of our analysis did not allow us to study this proximate cause of deforestation. Furthermore, up-to-date maps of mining activities do not exist and high resolution imagery for this portion of the study area are consistently cloud covered. Recently, the Colombian government has been using aerial cameras to document illegal mining in specific areas of the CGE, however, these methodologies are not applicable for an analysis of the entire region. Illegal farming activity, predominantly coca (*Erythroxylum coca*), is commonly found in the Colombian side (Nariño Department) near the border with Ecuador (UNODC 2015). These areas were coincident with one of the deforestation areas that we identify in our maps. Although, we cannot discriminate coca crops from other farming activities, the documented distribution of this crop is evidence of its significant influence as a proximate cause of deforestation within the CGE.

Developing annual maps of land cover across the CGE using satellite-based remote sensing instruments with higher spatial resolution than MODIS has not been successful. The United Nations Collaborative Program on reducing emissions from deforestation and forest degradation (REDD) in Colombia used available Landsat imagery to develop four forest/non-forest land cover maps for the years 2000, 2005, 2010 and 2012 (Phillips et al. 2011, IDEAM 2017). Each of these maps were developed using Landsat mosaics consisting of 3-4 contiguous years of imagery resulting in 13% of the area with no-information due to cloud cover. Our approach, using MODIS, allowed us to develop annual maps from 2002 to 2015 and identify land cover trends with a finer temporal grain. However, the MODIS pixel size cannot detect land cover change smaller than the 250 m² nominal pixel size which could affect our results. We therefore compared
the published trends of the four Landsat forest/no-forest maps from the Colombian REDD project with our MODIS maps for the same time periods. This analysis showed similar forest change trends between the Landsat and MODIS products; forest cover change trends were negatively correlated in similar proportions in the Landsat and MODIS maps (Landsat: $R=-0.99$, $p = 0.003$; MODIS: $R=-0.97$, $p = 0.02$).

We also compared the woody vegetation change (forest and secondary vegetation) of our 2002 and 2014 MODIS LULCL maps with the global forest change (GFC) maps of Hansen et al. (2013), which estimated loss and gain of tree cover between 2000 and 2014. To make an accurate comparison, we clipped the area classified as forest in our initial 2002 LULC map along with the LULC change between 2002 and 2014. We extracted the corresponding area of tree cover, tree loss and tree gain between 2000-2014 from the GFC database. The GFC product did not distinguish between forest (old forest) and secondary vegetation (young forests) as we did. Therefore, we combined these two classes into simply “woody vegetation” for the comparison. Our MODIS-based maps detected 6.35% woody vegetation loss between 2002 and 2014 compared to 3.9% for the GFC product. This level of non-agreement can be explained by the differences in spatial and temporal resolution as well as the definition of map classes between the GFC Landsat-based maps and our MODIS-based maps. The GFC database consists of two global maps of tree cover percentage for 2000 and 2014. The GFC database does not record the dynamics of tree cover between these two dates; consequently, the GFC does not discriminate between younger and older tree cover. Further, the increased spatial resolution of the GFC product compared to MODIS allows forest transitions to be identified at a finer scale. Small areas of non-forest within a matrix of forest tended to be
classified as secondary forest using MODIS whereas the GFC product seemed to identify these areas as non-forest. Consequently, our MODIS-based product seemed to overestimate deforestation as compared to the GFC database. However, this difference is mitigated by the inclusion of widespread palm plantations and wetlands as tree cover in the GFC product where we were able exclude them from our classification of forest. A direct comparison, therefore is difficult.

We used the GFW processed MODIS MOD13Q1 to build the LULC annual maps. The MOD13Q1 dataset is a 250 m resolution 16-day composite product calibrated to reflectance using an atmospheric correction for aerosol gases, and a BRDF (Bidirectional Reflectance Distribution Function) adjustment (Didan and Alfredo 2006, Didan 2017). MOD13Q1 adopts two cloud filters (Ackerman et al. 1998, 2006) and an aerosol quality filter. Recently, other MODIS products, such as MOD09 (MOD09Q1 and MOD09A1), have been developed with improved cloud filtering using the MAIAC algorithm (Multi-Angle Implementation of Atmospheric Correction) (Lyapustin et al. 2012). We chose to use the MOD13Q1 product over the MOD09 products after comparing annual time series of NDVIs of both products. We found that, overall, the pre-GFW NDVI temporal sequence of MOD13Q1 (original data) time series are less variable than the NDVI temporal sequence of MOD09Q1 within the CGE, and where pixels coincided temporally between the two products on the 16-day cycle, the calculated NDVI values were often identical between the two products. Consequently, the MOD13Q1 time series after GWF had significantly less variation (t = 5.54; p = 0.02), allowing for a better discrimination between land cover types. Additionally, MOD09Q1 consists of only the first two spectral MODIS bands (red and NIR) which would not provide an EVI
calculation and the MOD09A1 product, which allows for a calculation of EVI, has a spatial resolution of 500 m reducing our ability to discriminate between spatially adjacent land cover types. Therefore, for our purposes, we found the MOD13Q1 product superior to the MOD09 products.

Taking into account the high diversity and endemism of the CGE, the rapid reduction of forest is a primary concern for conservation activities. Currently the CGE still has significant reserves of original forest. FAO (2010) estimated that 64% of the global woody vegetation corresponded to forest regeneration following anthropogenic disturbances (FAO 2010). We estimate that 34% of woody vegetation in the CGE in 2015 corresponded to secondary vegetation, suggesting that the CGE has a higher proportion of well conserved forest (42% by our estimate) than other areas across the world. These areas support high levels of biodiversity making them important for conservation.

Tropical rain forest areas across the CGE occupied 83312 km$^2$ in 2015; therefore, the CGE contains the second largest mass of tropical rain forest in South America, after the Amazon Basin. However, the fast and gradual replacement of forest areas by secondary vegetation points to another main concern. The high levels diversity and endemism prior to deforestation in these forests cannot be recovered after reforestations. That is, secondary forests evolving from secondary vegetation will have decreased biodiversity and different species assemblages (Norden et al. 2015). Conserved forests in the CGE are located in Panamá and along the pacific coast of Colombia. Human colonization has been restricted in these areas by two main geographic barriers, the Andes Mountains in the east and the Pacific Ocean to the west. However, the deforestation line has moved forward in two primary locations: to the east of the Colombia-Panamá border (in the northeast of
these well-preserved forests) and on the Colombia-Ecuadorian border (to the south of these well-preserved forests). Mitigating deforestation in these two areas is critical to the conservation of the CGE.

CONCLUSIONS

By analyzing annual LULC change dynamics in the Chocó-Darien Global Ecoregion (CGE), we found that LULC change varied temporally and regionally. These regional/temporal variations need to be considered when developing CGE-wide management plans aimed at preserving biodiversity and ecosystem services.

Deforestation and reforestation occurred across the CGE; however, deforestation increased after 2010 showing an increased risks for CGE conservation. We detected a gradual replacement of forest areas by secondary vegetation and agriculture, mainly grassland, which would then transition to secondary vegetation. The increased loss of forest after 2010 should be an important concern for the preservation of CGE biodiversity because forests in this ecoregion have high levels of species richness and endemism which are difficult to recover through reforestation. In other words, secondary forests evolving from secondary vegetation would have decreased biodiversity and different species assemblages (Norden et al. 2015).

We also found spatial variations that are important to the CGE conservation effort. Across national boundaries, the Ecuadorian section had the smallest proportion of forest (11%; 3578.6 km²), for that reason, restoration programs are urgently needed in the Ecuadorian CGE. The Colombian CGE had the largest area of forest (66160 km²) but also the largest deforested area. The Panamanian CGE contains the largest proportion of forest (88%; 13569 km²) but this forested area is only 8% of the CGE. However, the
forest of the Panamanian CGE are fundamental to the connection of fauna and flora between Central and South America because these forests span the

LITERATURE CITED


Argüelles, L. 2010. Estrategia Institucional para la Prevención y Control de la Tala Ilegal en los Bosques del Darién en Panamá. WWF, Panama.


Didan, K. 2017. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m

Didan, K., and H. Alfredo. 2006. MODIS Vegetation Index Product Series Collection 5
Change Summary. NASA - The University of Arizona.

Dutra, A. P., J. P. Ometto, C. Nobre, D. M. Lapola, C. Almeida, I. C. Vieira, J. V. Soares,
spatial and temporal heterogeneity of deforestation-driven carbon emissions: the
INPE-EM framework applied to the Brazilian Amazon. GLOBAL CHANGE
BIOLOGY 18:3346–3366.


a tropical deforestation wave: a dynamic spatial analysis of a deforestation
hotspot in the Colombian Amazon. GLOBAL CHANGE BIOLOGY 12:1409–
1420.

Etter, A., C. McAlpine, and H. Possingham. 2008. Historical Patterns and Drivers of
Landscape Change in Colombia Since 1500: A Regionalized Spatial Approach.

Etter, A., C. McAlpine, D. Pullar, and H. P. Possingham. 2006b. Modelling the
conversion of Colombian lowland ecosystems since 1940: Drivers, patterns and

Fagua, J. C., E. Cabrera, and V. H. Gonzalez. 2013. The effect of highly variable
topography on the spatial distribution of Aniba perutilis (Lauraceae) in the

FEDEPALMA. 2011. Guía ambiental para el subsector de la agroindustria de la palma de aceite. FEDEPALMA, Bogota DC, Colombia.


Finagro, Bogota DC, Colombia.


Hosonuma, N., M. Herold, V. De Sy, R. S. De Fries, M. Brockhaus, L. Verchot, A. Angelsen, and E. Romijn. 2012. An assessment of deforestation and forest degradation drivers in developing countries. ENVIRONMENTAL RESEARCH LETTERS 7.

IDEAM. 2010. Leyenda nacional de coberturas de la tierra. IDEAM, Bogotá D.C., Colombia.


Minsalud, M. de la P. C. 2011. PRODUCTO INTERNO BRUTO E INFLACIÓN EN COLOMBIA. Bogota DC, Colombia.


ProEcuador, I. de promocion de exportaciones e inversiones de E. 2014. PERFIL SECTORIAL DE AGROINDUSTRIA PARA EL INVERSIONISTA. Quito, Ecuador.


http://wwf.panda.org/about_our_earth/ecoregions/chocodarien_moist_forests.cfm.


Scientific Reports 7:1375.

Table 3.1. Correlations of land-use and land-cover change trends among administrative divisions; The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE). Pearson’s correlation coefficient (R) are shown for two time periods 2002-2010 and 2010-2015. Significant correlation are bolded.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CGE</td>
<td>2002</td>
<td>0.81 (0.02)*</td>
<td>-0.96 (0.001)**</td>
<td>0.94 (0.01)**</td>
<td>-0.64 (0.05)*</td>
<td>-0.65 (0.06)</td>
<td>-0.89 (0.02)*</td>
<td>0.51 (0.16)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>-0.69 (0.13)</td>
<td>-0.98 (0.001)**</td>
<td>0.17 (0.75)</td>
<td>0.85 (0.03)*</td>
<td>-0.75 (0.07)</td>
<td>0.89 (0.07)</td>
<td>-0.56 (0.25)</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>0.78 (0.01)**</td>
<td>-0.95 (0.001)***</td>
<td>0.91 (0.01)**</td>
<td>-0.68 (0.04)*</td>
<td>-0.33 (0.39)</td>
<td>-0.86 (0.01)*</td>
<td>0.08 (0.86)</td>
</tr>
<tr>
<td>Col CGE</td>
<td>2002</td>
<td>-0.89 (0.02)*</td>
<td>-0.91 (0.01)**</td>
<td>-0.14 (0.8)</td>
<td>0.89 (0.02)*</td>
<td>-0.57 (0.24)</td>
<td>-0.01 (0.99)</td>
<td>-0.54 (0.27)</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>-0.64 (0.05)*</td>
<td>-0.94 (0.01)**</td>
<td>0.85 (0.01)**</td>
<td>0.14 (0.74)</td>
<td>-0.63 (0.05)*</td>
<td>-0.56 (0.15)</td>
<td>0.96 (0.01)**</td>
</tr>
<tr>
<td>Ecu CGE</td>
<td>2002</td>
<td>-0.11 (0.84)</td>
<td>-0.96 (0.02)*</td>
<td>0.22 (0.68)</td>
<td>0.03 (0.63)</td>
<td>-0.72 (0.11)</td>
<td>0.31 (0.54)</td>
<td>-0.14 (0.8)</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>-0.57 (0.11)</td>
<td>-0.89 (0.01)**</td>
<td>0.95 (0.01)**</td>
<td>0.39 (0.01)**</td>
<td>0.06 (0.88)</td>
<td>0.1 (0.8)</td>
<td>0.36 (0.34)</td>
</tr>
<tr>
<td>Pan CGE</td>
<td>2010</td>
<td>-0.94 (0.01)**</td>
<td>-0.85 (0.02)*</td>
<td>0.73 (0.1)</td>
<td>0.86 (0.03)*</td>
<td>-0.78 (0.07)</td>
<td>0.36 (0.22)</td>
<td>0.75 (0.09)</td>
</tr>
<tr>
<td></td>
<td>2015</td>
<td>-0.46 (0.11)</td>
<td>-0.89 (0.01)**</td>
<td>0.95 (0.01)**</td>
<td>0.39 (0.01)**</td>
<td>0.06 (0.88)</td>
<td>0.1 (0.8)</td>
<td>0.36 (0.34)</td>
</tr>
</tbody>
</table>
Fig. 3.1. The Chocó-Darien Global Ecoregion (CGE): (a) estimated historical extent of Tropical Rain Forest (TRF) in South America: TRF-CGE (estimated TRF in CGE), TRF-Amz (estimated TRF in Amazon basin), and TRF-BrAt (estimated TRF in Brazilian Atlantic Forest). (b) The Chocó-Darién Global Ecoregion (CGE): This global ecoregion is formed by three sub-ecoregions: Magdalena-Urabá Moist forests (MgU), Chocó-Darién Moist Forests (ChD), and Western Ecuador Moist Forests (WEc).
Fig. 3.2. An example of the time series filtering procedure from 2009 to 2013 using the Gaussian weighted filter (GWF). GWF improved the identification of land cover using the MODIS bands and indices. (a) A time series EVI pixel of woody vegetation before and after filtering; outliers are replaced with estimates calculated by the Gaussian weighted filter. (b) Temporal variation of 120 pixels corresponding to woody vegetation before filtering and (c) the same 120 pixels after filtering; the variance of these 120-time series is reduced. (d) Post filtering results –simplified to means, for 120 woody vegetation pixels, 116 grassland pixels, 542 palm plantation pixels, 99 settlement pixels, 101 water pixels, and 454 wetland pixels. GWF increased the differentiation among these land covers. The GWFs were applied using the Terrset software (Eastman 2017) with a temporal filter length of 5.
Fig. 3.3. Land-use and land-cover (LULC) change trends in The Chocó-Darien Global Ecoregion (CGE). Significant correlation coefficients (R) are shown for two time year periods 2002-2010 and 2010-2015 (a). Significant P range values; P<0.001(***), P<0.01(**), and P<0.05(*). Land-use and land-cover maps for 2002, 2010 and 2015 are showed (b, c, d).
Fig. 3.4. Quantification of deforestation (proximate causes of deforestation) and reforestation transitions from 2002 to 2010. (a) Percentage of deforested area and net area deforested by every proximate cause of deforestation, and (b) percentage of reforested areas and net area reforested by every reforestation transition. The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE)
Fig. 3.5. Quantification of deforestation (proximate causes of reforestation) and reforestation transitions from 2010 to 2015. (a) Percentage of deforested area and net area deforested by every proximate cause of reforestation, and (b) percentage of reforested areas and net area reforested by every reforestation transition. The Chocó-Darien Global Ecoregion (CGE), Colombian CGE (Col CGE), Ecuadorian CGE (Ecu CGE), and Panamanian CGE (Pan CGE).
CHAPTER 4

CAUSES OF FOREST COVER CHANGES IN THE CHOCÓ-DARIEN GLOBAL ECOREGION OF SOUTH AMERICA

ABSTRACT

Tropical rain forests are suffering the highest deforestation and reforestation ever recorded. Interactions between proximate and underlying (or underlying drivers) causes could cluster these forest cover changes forming hotspots (areas that exhibit significant spatial correlation of deforestation or reforestation transitions). Using land-use and land-cover (LULC) maps and global (I) and local (li) Moran’s tests, we identified these hotspots in the Chocó-Darien Global Ecoregion (CGE) of South America, a natural region that was declared one of the top 25 hotspots for conservation priorities in the world. Subsequently, we tested and studied the effects and interactions between deforestation and reforestation hotspots and their proximate and underlying causes using Bayesian Structural Equation modeling (Bayesian SEM). We found that deforestation and reforestation were spatially auto-correlated forming hotspots (I=0.49, P = 0.001 for deforestation transitions and I=0.48, P = 0.001 for reforestation transitions). Also, hotspots of deforestation and reforestation were auto-correlated within municipality borders (I=0.5, P = 0.001 for deforestation transitions; I=0.49, P = 0.001 for reforestation transitions). Eighteen municipalities located on the border between Colombia and Ecuador showed significant aggregations of deforestation hotspots while thirty-four municipalities in three areas of Colombia and the area between the Colombian and Ecuadorian border showed significant clustering of reforestation hotspots. Eleven of these municipalities presented significant clustering of both reforestation and
deforestation hotspots. The Bayesian SEM for deforestation showed that population growth and road density were underlying causes of deforestation hotspots (0.194 and 0.115 standard deviation units). The Bayesian SEM for reforestation found that armed conflicts, Gross Domestic Product (GDP), and average annual rain were underlying causes related reforestation hotspots (0.152, 0.051, and 0.034 standard deviation units respectively).

INTRODUCTION

Tropical rain forests consist of evergreen vegetation that reaches 30 m in height and are located in the lowlands of the wet tropical climatic zone of the planet (Primack and Corlett 2009). Different studies have documented extensive deforestation and minor reforestation (regrowth of secondary vegetation) in these forests during the first ten years of the 21st century. Consequently, the tropical rain forest domains have exhibited significant and constant forest loss (Sanchez-Cuervo et al. 2012, Aide et al. 2013b, Hansen et al. 2013, Keenan et al. 2015, Qin et al. 2017). Deforestation and reforestation dynamics play a key role in global environmental changes, degrading or protecting ecosystem services that are fundamental to human development, such as, climate and soil regulation, water supply, carbon storage, and biodiversity (Lambin et al. 2003b, Leblois et al. 2017). The identification of causes that affect the deforestation-reforestation dynamic in tropical rain forests is essential to developing strategies for mitigating forest loss.

Causes of deforestation are classified into two types: (1) proximate causes (or direct causes) which are immediate actions that directly result in forest loss; in other words, proximate causes refer to the land cover that replaces forest cover; (2) underlying
causes (underlying drivers) are interactions of social, economic, political, cultural, or other processes that indirectly affect deforestation (Geist and Lambin 2001 2002). The previous statement has been the basis for a wide array of literature focused on studying direct (e.g., Roberts et al. 2002, Flamenco-Sandoval et al. 2007, Houghton 2012, Mueller et al. 2012, Tng et al. 2012, Alejandra Chadid et al. 2015, Armenteras et al. 2017) or underlying causes (e.g., Armenteras et al. 2006, 2013, 2017, Hosonuma et al. 2012, Kissinger et al. 2012, Mueller et al. 2013, Leblois et al. 2017) to assess forest cover change at different spatial scales throughout the tropical rain forests domain. Studies that include both proximate and underlying causes as well as their interactions have been less prevalent (Davalos et al. 2014, Richards 2015).

In the Neotropics, the tropical terrestrial ecoregions of Central and South America, changes in forest cover are mainly affected by farming expansion, logging, mining and, more generally, by natural resource extractive activities (Lambin et al. 2003b, DeFries et al. 2010). However, farming is estimated to be the largest specific proximate cause of change in forest cover (Hosonuma et al. 2012, Kissinger et al. 2012). On the other hand, regional analyses of underlying causes of deforestation in the Neotropic have found that deforestation is related to socioeconomic (e.g., National Gross Domestic Product) and demographic indicators followed by accessibility (e.g., existence of roads, rivers) (Rudel and Roper 1997, Aide et al. 2013, Armenteras et al. 2017). More localized studies of deforestation also highlight the importance of socio-economic as well as demographic underlying causes (Sanchez-Cuervo and Aide 2013, Armenteras et al. 2013), as well, topography, soil fertility, and climate variables (Steininger et al. 2001, Laurance et al. 2001, Sanchez-Cuervo and Aide 2013).
The causes of reforestation or natural regeneration of secondary vegetation have been less studied in the Neotropics (Grau et al. 2003, 2004, Aide et al. 2013); land abandonment is considered the principal cause of reforestation (Rudel et al. 2002, Lambin et al. 2003, Aide et al. 2013). Reforestation has been related indirectly to regional and local factors (e.g., population density, socioeconomic indicators, climate, topography, soil fertility, accessibility, violent conflicts) (Rudel et al. 2002, Lambin et al. 2003, Grau et al. 2003, Sanchez-Cuervo et al. 2012, Aide et al. 2013b), suggesting that proximate and underlying causes also affect reforestation transitions; as it has been proposed for deforestation by Geist and Lambin (2001, 2002).

Proximate and underlying causes of deforestation and reforestation could vary spatially promoting or restricting forest transitions in natural landscapes (Grau et al. 2003, Sanchez-Cuervo and Aide 2013). In this case, deforestation and reforestation may exhibit statistically significant spatial clustering or “hotspots” (Ferreira et al. 2007, Reddy et al. 2016, Harris et al. 2017). The spatial identification of deforestation and reforestation hotspots is an important tool for the preservation and management of tropical rain forest in natural regions. For instance, deforestation hotspots identify areas where forest protection strategies should be implemented whereas reforestation hotspots detect areas in which forest restoration programs might be established (Sanchez-Cuervo and Aide 2013). In the Neotropics, hotspots are traditionally estimated based on forest change trends between administrative subdivisions (e.g. countries, states, municipalities, etc.) (Myers 1993, Sanchez-Cuervo and Aide 2013, Reddy et al. 2016). However, both types of transitions can exhibit statistically significant spatial clustering within the same administrative division (Hansen et al. 2013). Increasing the scale of analysis by
identifying hotspots at the spatial scale of land use-land cover (LULC) maps would allow more accurate estimations.

Using Bayesian Structural Equation modeling (Bayesian SEM), we estimated the relationships between deforestation and reforestation hotspots and their proximate and underlying causes in the Chocó-Darien Global Ecoregion (CGE) of South America. Here, we focus on the effects of underlying causes on deforestation and reforestation hotspots. CGE is a natural region that was declared as one of the top 25 hotspots for conservation priorities in the world (Myers et al. 2000, Olson et al. 2001, WWF 2016). Deforestation and reforestation transitions were studied in the CGE between 2002 and 2015, showing a significant forest loss trend (Fagua and Ramsey unpublished manuscript; Chapter 3). Our two main objectives were 1) to build a model to quantify and compare the effect of proximate and underlying causes on the appearance of deforestation hotspots using a method where all these causal variables could interact, and 2) to build an analogous model to compare proximate and underlying causes of reforestation and their relationship with the appearance of reforestation hotspots. To construct the models, we first assessed if deforestation transitions were spatially clustering to form deforestation hotspots between 2002 and 2015 at the original spatial scale (250 m pixel) of LULC maps developed by Fagua and Ramsey (2018) using MODIS imagery. Likewise, a similar assessment was performed to test if reforestation transitions were spatially clustered. Finally, we assessed whether deforestation and reforestation hotspots were spatially correlated with the delineation of different municipalities in the region. Following our analysis of spatial correlations, we addressed proximate and underlying causes of
deforestation and reforestation at the municipality level to explain the emergence of hotspots.

METHODS

Study area

The Chocó-Darien Global Ecoregion (CGE) is located along the Pacific Coast from southern Panamá to northeastern Ecuador. The CGE includes the lowlands of the Magdalena River Valley between the central and western branches of the Colombian Andes (Fig. 4.1A) (Olson et al. 2001, WWF 2017). Past geologic events have turned the CGE into one of the most diverse tropical rain forests on the planet (Gentry 1986, Gregory-Wodzicki 2000, WWF 2017). The CGE is also the rainiest area on Earth (Poveda and Mesa 2000); consequently, LULC analyses within the territories of the countries that share the CGE are limited due to cloud cover (González et al. 2011, Ministerio del Ambiente del Ecuador 2012, Sanchez-Cuervo et al. 2012). The CGE has also been included in undetailed regional and global analyses of forest change (Sanchez-Cuervo et al. 2012, Aide et al. 2013b, Hansen et al. 2013, Keenan et al. 2015, Qin et al. 2017). However, the only LULC analysis that studied the CGE as a unit found a gradual replacement of forested areas (23%) by: (1) secondary vegetation (14%) and (2) agriculture (5%), which then transitioned to secondary vegetation (Fagua and Ramsey unpublished manuscript; Chapter 3). Here we use the maps of this analysis to assess LULC changes (Figs. 4.1B,C). These maps allowed for a better understanding of direct and indirect effects on reforestation and deforestation from 2002 to 2015 because they have eight LULC classes: forest, secondary vegetation, grassland, crop, palm plantations,
settlement, wetland, and water. Thus, LULC transitions from these maps estimated the
deforestation caused by agriculture expansion and reforestation caused by the lack of
agricultural use in farming areas. These LULC maps can be download as GeoTIFF
format in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/

The three countries that share the CGE have had different land management
histories. Eight percent of the CGE is located in Panamá. This area has the lowest human
population density in that country (8 people/km²). Most of this territory consists of
indigenous reservations (56%) and national parks (55%) (INEC-Panama 2017). The main
sectors of the Panamanian economy are services, banking, and tourism (Fisher 2015).
The extraction of natural resources is a minor sector; however, legal and illegal logging
are main problems for the conservation of the forests in Panamá (Argüelles 2010).
Seventy percent of the CGE falls in Colombia. The area along the Pacific Coast has one
of the lowest human population densities of this country (10 people/km²); forms part of
its second largest forest reserve (58343 km²); and contains several afro-Colombian
reservations, (58%), indigenous reservations (22%) and national parks (5%) (IGAC
2017). On the other hand, the areas located in the Caribbean region and the Magdalena
Valley have higher human population density (23 people/km²) and different farming
activities that have been established historically (cattle and, in less proportion, palm
plantations and other crops) (PNUD 2014, IGAC 2017). Colombian economy is based on
the extraction of natural resources; legal and illegal crops, mining, and logging are main
problems for the conservation of the forest in the CGE of Colombia (Rangel 2011,
UNODC 2015). Sixteen percent of the CGE is located in Ecuador. This country is one of
the most populated in Latin-America, and its CGE section has the highest density of
human population (34 people/km²) compared with the Panamanian and Colombian parts (INEC-Ecuador 2017). The Ecuadorian economy is based on the extraction of natural resources (BCE 2010). The forests in the CGE of Ecuador have been affected mainly by different farming activities (annual-semiannual crops, cattle, and palm plantations) (Sierra 2013).

Spatial identification of deforestation and reforestation hotspots

We followed two methodological steps to identify areas where deforestation and reforestation transitions were significantly clustered. In the first step we ran two global Moran’s (I) tests, one for the deforestation transitions and another for the reforestation transitions, to test if these transitions were spatially autocorrelated between the 2002 and the 2015 LULC maps of the CGE (Fagua and Ramsey 2018). We considered deforestation as transitions from forest to secondary vegetation, grassland, crops, palm plantations, or settlement and reforestation as transitions between non-forest land cover types into secondary vegetation. Moran’s I is a standardized measure of correlation between observations in neighboring areas; it is commonly employed in the analysis of spatial data (Cliff and Ord 1981, Bivand et al. 2013). I is a linear correlation coefficient calculated as a ratio of the product of the variable of interest and its spatial lag, with the cross-product of the variable interest, and adjust for the spatial weights used (Bivand et al. 2013):

\[
I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \overline{y})(y_j - \overline{y})}{\sum_{i=1}^{n} (y_i - \overline{y})^2}
\]

Where \( n \) is the number of features, \( y_i \) is the \( i \)th observation, \( \overline{y} \) is the mean of the variable of interest, and \( w_{ij} \) is the spatial weight of the link between \( i \) and \( j \). I ranges from
−1 to +1; I values significantly below -1/(n-1) indicate negative spatial autocorrelation and I values significantly above -1/(n-1) indicate positive spatial autocorrelation (Cliff and Ord 1981). In our case, n corresponded to rectangular matrix of 3x3 pixels in the 2002-2015 transition LULC raster map where y represents the deforestation or reforestation value in the 3x3 matrix. To arrive at the 3x3 matrix size, we tested a variety of matrices from 3x3 pixels through 37x37 pixels. The relation I vs. matrix size for deforestation (p=0.001; R= -0.94) and reforestation (p=0.001 ; R= -0.97) showed a typical negative exponential shape of autocorrelated relationships (Appendix S4-1: Fig. 4.1); we therefore selected the 3x3 pixel matrix for our analysis due to its higher I (Dale and Fortin 2002). Deforestation values were calculated by assigning values of 1 to deforestation transitions and 0 to the other transitions in the 2002-2015 transition LULC map. Following the same schema, reforestation values were calculated giving values of 1 to reforestation transitions and 0 to the others (Shortridge 2007). Reforestation transitions were considered as the conversion from farming land-uses (grassland, crop and palm plantations) to secondary vegetation.

For our second methodological step, after testing the global spatial autocorrelation for deforestation and reforestation transitions, we ran local Moran’s ($I_i$) tests to identify the 3x3 matrices where deforestation or reforestation transitions were statistically significantly clustered. Since I is the slope of a linear relation, $I_i$ detects the 3x3 matrices with significant influence on the slope. Consequently, the formula for $I_i$ is similar to I, but $I_i$ is calculated separately for each feature, in our case each 3x3 matrix:

$$I_i = \frac{(y_i + \bar{y}) \sum_{j=1}^{n} w_{ij} (y_j - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$
The results of these analyses are shown in a map of deforestation hotspots and another of reforestation of hotspots (Figs 4.2A,B) and were implemented using the R “raster” package (Hijmans et al. 2016, R Core Team 2014). These maps of deforestation and reforestation hotspots can be downloaded as GeoTIFF format in the next link:
http://data.gis.usu.edu/CGE/?prefix=CGE/

Spatial autocorrelation of hotspots in relation to municipalities

We used a two-step process to identify municipalities where deforestation and reforestation hotspots were significantly clustered. The first step was to calculate the global Moran’s ($I$) to determine if there was spatial autocorrelation between deforestation and reforestation hotspots with municipality borders within the CGE. Because size and shape of the municipalities vary, we used a distance-based neighbor method to evaluate spatial clustering regardless of the distance at which the tests were applied. Distance-based neighbor selects the nearest neighbor(s) given the distance within the feature centroid. In our case, the centroid of the municipality (Bivand et al. 2013); we ran these $I$ tests from one to six neighbors with the null hypotheses stating that the deforestation or reforestation hotspots were randomly distributed among the municipalities. As before, $I$ values significantly below $-1/(n-1)$ indicated negative spatial autocorrelation and $I$ values significantly above $-1/(n-1)$ indicated positive spatial autocorrelation, where $n$ is the number of municipalities (Cliff and Ord 1981). To identify the municipalities where deforestation or reforestation hotspots were significantly clustered, we ran local Moran’s ($I_r$) tests. Since $I$ is the slope of a linear relationship, $I_r$ detected the municipalities with a significant influence on this slope. The results of these analyses are shown in Moran
scatter plots and maps of municipalities (Fig. 4.3). We used the R package “spdep” (Bivand and Piras 2015) for these analyses.

**Underlying causes of deforestation and reforestation**

Traditionally, the effect of proximate and underlying causes on forest transitions are estimated using different types of regressions (linear regression, logistic regression, generalized linear models, random forest regression, etc.) (e.g., Roberts et al. 2002, Armenteras et al. 2006, 2013, 2017, Flamenco-Sandoval et al. 2007, Hosonuma et al. 2012, Kissinger et al. 2012, Tng et al. 2012, Mueller et al. 2013, 2012, Houghton 2012, Sanchez-Cuervo and Aide 2013, Alejandra Chadid et al. 2015, Leblois et al. 2017); these statistical tests assume that all proximate and underlying causes as predictors have the same potential direct effect on deforestation. However, the “equality of effect” assumption is not consistent with the analysis of proximate and underlying causes of deforestation (see Geist and Lambin (2001, 2002)).

We propose that Bayesian SEM utilizing available LULC maps makes it possible to estimate 1) the effects of proximate and underlying causes, 2) the effects of proximate causes on hotspots, 3) the effects of underlying causes on hotspots, and 4) other possible interactions in the set of variables. Bayesian SEM, therefore, is more appropriate than the aforementioned techniques to evaluate proximate and underlying causes of deforestation and reforestation because Bayesian SEM presents fewer limitations given by sample size, underlying distribution, and assumptions of normality, allowing the modelling of complex networks of variables.

Municipalities (called Municipalities in Colombia, Districts in Panamá, and Cantones in Ecuador), were the primary units in our Bayesian SEM. Based on the groups
of underlying causes studied on different ecoregions of tropical forest (Geist and Lambin 2001, Meyfroidt et al. 2018) and specific characteristics of the CGE, we selected 10 different variables as underlying causes. All variables were defined at the municipality (Table 4.1) and can be classified into six categories.

(1) Accessibility: transportation networks and rivers (Geist and Lambin 2002). The accessibility variables in our analysis included paved and unpaved roads from national cartographic data from each country (IGAC 2017, IGNTG 2017, SNI 2017). Rivers were extracted from HydroSHEDS (Hydrologic data derived from Elevation Derivatives produced by the NASA Shuttle Radar Topographic Mission at various scales) (Lehner et al. 2008). Density was estimated as the length of routes or rivers (km) per km².

(2) Biophysical environment: Landscape physical characteristics can affect access to forested areas, thus influencing deforestation and reforestation transitions (Geist and Lambin 2002). We used the mean of topographic slope (degrees) and standard deviation of altitude (estimated as MASL) as biophysical variables; both variables were assembled from SRTM90 (NASA Shuttle Radar Topographic Mission) (Jarvis et al. 2008).

(3) Climate: CGE has important variations in precipitation and temperature that may affect deforestation and reforestation transitions. Although the CGE is one of the rainiest places in the world (Poveda and Mesa 2000), its northern and southern extremes have lower precipitation than the central section. Temperature also varies across the CGE, increasing from south to north. We selected mean annual precipitation (mm/year) and mean annual temperature (°C) from the 30 arc-second WorldClim layers (Fick and Hijmans 2017).
(4) Demography: population growth is considered a primary indirect cause of deforestation (Ehrhardt-Martinez 1998, Geist and Lambin 2001). We estimated this variable as:

\[ P_g = \frac{P_f - P_o}{P_f} \]

Where \( P_o \) and \( P_f \) are the population for every municipality in 2000 and 2010 respectively. Population was extracted from the official national census of each country (DANE 2017, INEC-Ecuador 2017, INEC-Panamá 2017). Panamá and Ecuador have census for 2000 and 2010 and Colombian has census for 1984 and 2005. We used the official projections of population for 2000 and 2010 for Colombia (DANE 2017).

(5) Armed conflicts: these type of conflicts have occurred in several areas of the CGE, especially within Colombia, produced by land possession, social and economic inequality, politic issues, illegal crops and drug production, and illegal mining (Yaffe 2011, Guzmán et al. 2016, Fajardo 2017). The effects of armed conflicts on environment, wildlands, and biodiversity remain complex (Hammill et al. 2016). Some scholars have found that armed conflicts put conservation at risk by reducing the effectiveness of protection of wild life or natural vegetation (Beyers et al. 2011, Hammill et al. 2016); others have showed beneficial effects on forest protection by creating exclusion zones (John 1998) or hindering extractive industries (McNeely 2003). Also, armed conflicts have been related to land abandonment and posterior forest regeneration in Colombia (Sanchez-Cuervo and Aide 2013). We used two variables to include armed conflicts in our analysis: number of armed conflicts and number of reported fatalities produced by the armed conflicts between 2000-2015 (Sundberg and Melander 2013, Croicu and Sundberg 2017).
(6) Economy: we included Gross Domestic Product (GDP) growth between 2000 and 2010 as our main economic variable. Each country has official GDP estimates for the country as well as the first administrative level (Departments in Colombia and Provinces in Panamá and Ecuador), but only Ecuador has an estimate of GDP for municipalities (BCE 2017, DANE 2017, INEC-Panamá 2017). To estimate the GDP in the municipalities of Colombia and Panamá, we allocated the GDPs of every Department in Colombia and Province in Panamá to their corresponding municipalities according to the proportion of population of that municipality within the Department or Province. Variables evaluated as underlying causes of deforestation and reforestation were listed in the Table 4.1.

Bayesian SEM

We applied a linear Bayesian structural equation model (Bayesian SEM) to examine the relationships among the measured variables at the municipality level. One model was built for deforestation where the primary variable was the number of hotspots of deforestation. Another model was constructed for reforestation in which the main primary variable was the number of hotspots of reforestation. To reduce endogeneity issues, all variables considered as independent were estimated earlier (underlying causes; from 2000 to 2010) than the dependent (proximate causes and deforestation hotspots; from 2002-2015) in both models (Aron 2000, Baggio and Papyrakis 2010). SEM uses the variances and covariances in the dataset to test the most probable path of linear relationships among the variables based on an initial hypothesis (Merkle and Rosseel 2016). Because our main objective was to estimate indirect effects of several variables on deforestation and reforestation hotspots, the linear Bayesian SEM offered us
methodological advantages including: (1) Bayesian SEM can explicitly examine both direct and indirect relationships between variables. (2) The sample size is less restrictive in Bayesian SEM than in the traditional likelihood SEM to produce reliable results in complex networks of variables. (3) The ability to include latent variables to represent theoretical variables that cannot be measured (Lee 2007). We used the R package “blavaan” (Merkle and Rosseel 2016) to run the Bayesian SEM model. We used the default settings of “blavaan” to define the a-priory distributions for each variable and parameters in the model. We selected “manual” convergence to add the rest of the settings according to Plummer (2015): the number of burnin iterations was set to 4000, the number of adaptive iterations to use at the start of the simulation was 1000, and the total number of samples to take after burnin was 10000 (Plummer 2015, Denwood 2016). We also recoded the set of variables to units of no more than two digits to reduce the variance as per a requirement to run SEM (Rosseel 2012). In the Bayesian SEM, the goodness-of-fit of a hypothesized model is evaluated by the posterior predictive P-value (PPP); PPPs higher than 0.05 indicate a good model fit. The direct effect of a variable on another is measured by unstandardized posterior path coefficients (Post.Mean values) and the standardized posterior path coefficients (Std.all values). Post.Mean represents the slope in the linear relation between a couple of endogenous and exogenous variables while the other exogenous variables are constant. In other words, Post.Means shows the percentage of change of the endogenous variable when the exogenous variable changes one raw unit. Post.Means are also estimated with the posterior standard deviation (Post.SD), the 95% highest posterior density interval (HPD.025 and HPD.975), and the potential scale reduction factor for assessing chain converge (PSRF). PSRF lower than
1.2 indicates that convergence was reached by the variable or parameter (Lee 2007, Merkle and Rosseel 2016). On the other hand, Std.all shows the change of standard deviation units in the endogenous variable when the exogenous variable changes one standard deviation unit while the other exogenous variables are constant. Due to Std.all estimation in constant units (standard deviations), Std.all can be used to estimate indirect effects and to compare the effects among variables. Indirect effects are estimated by multiplying the Std.all through a path (Grace and Bollen 2005). Bayesian SEM also estimates global fit measures that help select the best model among several hypothesized models. These global fit measures are Deviance Information Criterion (DIC), Widely Applicable Information Criterion (WAIC) and Leave-one-out cross-validation (LOOIC); smaller values of DIC, WAIC and LOOIC are the models with a better global fit (Merkle and Rosseel 2016, Vehtari et al. 2016, 2016b). In SEM analysis, Bayesian or Likelihood, the point of view of the researcher is the other criteria to select a model among hypothesized models with PPP higher than 0.05 (Grace et al. 2012). The R code to build our Bayesian SEMs and the data to run these codes is available in appendices S2 and S3. Because the sample units of our Bayesian SEMs, the municipalities, correspond to three different countries, we performed Bayesian SEMs adding Panamá, Colombia, and Ecuador as a categorical variable to estimate the effect of these countries in our models. We found that the effect of the countries in the final models did not affect notoriously the relation among the other variables in both models. The R code to build our Bayesian SEMs adding country as a categorical variable are available also in appendix S3.
RESULTS

Global Moran’s ($I$) tests showed that both deforestation and reforestation transitions were spatially auto-correlated within the CGE ($I$=0.49, $P = 0.001$ for deforestation transitions; $I$=0.48, $P = 0.001$ for reforestation transitions). Local Moran’s ($I_i$) tests detected hotspots of deforestation and reforestation, that is, areas where deforestation and reforestation transitions were significantly auto-correlated across the CGE (Fig. 4.2A,B). We also found that areas identified as hotspots of deforestation were spatially correlated to areas identified as hotspots of reforestation ($R= 0.8$, $p = 0.01$), showing that both transition types tended to occur in close proximity (Fig. 4.2C).

Additional Global Moran’s ($I$) tests determined that hotspots of deforestation and reforestation were auto-correlated with municipalities across the CGE ($I$=0.5, $P = 0.001$ for deforestation transitions; $I$=0.49, $P = 0.001$ for reforestation transitions). The corresponded Local Moran’s ($I_i$) tests (Fig. 4.3A,B) identified municipalities with significant hotspot clustering for both deforestation and reforestation. Eighteen municipalities located near the border of Colombia and Ecuador presented the most significant aggregation of deforestation hotspots; three in the Colombian side and 15 in the Ecuadorian side (Fig. 4.3C). On the other hand, 34 municipalities located in three areas of Colombia as well as the area around the Colombian and Ecuadorian border showed significant clustering of reforestation transitions (Fig. 4.3D). We found that 11 municipalities presented significant clustering of both deforestation and reforestation hotspots: in Colombia - Barbacoas, Mosquera, and Tumaco; and in Ecuador - Eloy Alfaro, Las Golondrinas, Pedernales, Puerto Quito, Quininde, Rio Verde, San Lorenzo and Tulcan.
The selected Bayesian SEM for deforestation reached a PPP of 0.952 (Fig. 4.4, Appendix S2: Data S1, Appendix S3: Metadata S1), indicating a very good model fit. The selected model showed the lowest values of DIC (2758), WAIC (2804) and LOOIC (2801) compared with other models. All relationships between variables and their parameters in the model had a PSRF lower than 1.2, indicating that they reached convergence. Standardized posterior path coefficients between proximate and underlying causes of deforestation showed positive relationships between forest replaced by farming and population growth, as well as forest replaced by farming and road density. Conversely, rain, GDP, topographic slope and temperature showed a negative relationship with forest replaced by farming. Population growth had the largest total correlation with deforestation hotspots (0.179), which included direct (0.101) and indirect effects (0.07). Road density had a high indirect effect on deforestation hotspots (0.115). All other underlying causes of deforestation had negative relationships with deforestation hotspots (Fig. 4.4).

The selected Bayesian SEM for reforestation also resulted in a very good model fit with a PPP of 0.266 (Fig. 4.5, Appendix 4-S2: Data 4-S1, Appendix 4-S3: Metadata 4-S1). This model also showed the lowest values of DIC (2165), WAIC (2177), LOOIC (2178) compared to other models. All relationships between variables and their parameters in the model reached convergence (PSRF < 1.2). Standardized posterior path coefficients between proximate and underlying causes of reforestation found positive relationships between farming that had transitioned to secondary vegetation and areas of armed conflicts, GDP, and rain (Fig. 4.5); armed conflicts showed the highest indirect
positive effect on reforestation hotspots (0.152). Road density and temperature were negatively related with farming that had transitioned to secondary vegetation.

DISCUSSION

*Spatial correlation*

Our results confirm that deforestation and reforestation in the CGE tended to be spatially clustered forming hotspots. This suggests that forest changes were not accidental processes; forest changes responded to causes that determined their geographical location and intensity. We also detected that many of the reforestation hotspots were spatially adjacent to deforestation hotspots, suggesting that after harvesting mature forest (deforestation) many of these lands were not used for other purposes in subsequent years, thus allowing the generation of secondary vegetation. Fagua and Ramsey (2018) showed that approximately 60% of harvested forests converted to secondary vegetation between 2001 and 2015 across the CGE. This relatively high proportion of forest harvest immediately followed by secondary vegetation growth is explained by three main factors within the CGE and the surrounding tropical forests: (1) Deforestation occurs in areas with fragile and unproductive soils, such as the Choco-Darien, an ecoregion within the CGE (IGAC 2015); (2) colonists remove older forests to demonstrate land use in order to gain ownership of the property (Davalos et al. 2014); and (3) the decline of Colombian and Ecuadorian agricultural sectors between 2000 and 2010 reduced cultivated areas (grassland, crops, palm), allowing for encroachment of secondary vegetation (BCE 2010, Buitrago 2013, Marrugo 2013).
We found that hotspots of deforestation and reforestation were autocorrelated with municipalities. The former was an expected result since municipalities are the basic administrative and governing units in the countries; thus, social, economic, political, and cultural processes (i.e., underlying causes or drivers of forest change) vary with municipalities, producing spatial aggregations of deforestation or reforestation transitions (Aide et al. 2013b, Sanchez-Cuervo and Aide 2013). We also identified the municipalities with significant clustering of deforestation or reforestation; surprisingly, we detected that 11 municipalities presented significant clustering of both deforestation and reforestation hotspots, corroborating the aforementioned statement that a high proportion of areas where mature forest has been deforested are not used in agriculture. These results also demonstrate that hotspot identification from land cover change maps allows more accurate estimations than those based on forest change trends, where the analyzed administrative unit can only be a deforestation or a reforestation hotspot, but not both (e.g., Armenteras et al. 2013, 2017, Sanchez-Cuervo and Aide 2013, T. Mitchell Aide et al. 2013).

**Deforestation**

The Bayesian SEM model of deforestation for the CGE showed that human population growth was most related to deforestation hotspots between 2002 and 2015. Deforestation hotspots were principally located in municipalities found in the northern portion of Ecuador and around the Ecuadorian and Colombian border. These municipalities tended to have higher population densities and population growth rates compared to other municipalities in the CGE. Our results agree with several scholars who have found demographic variables as cause of deforestation in tropical forests across
Latin America at national (Sanchez-Cuervo and Aide 2013, Armenteras et al. 2013) and regional levels (Laurance et al. 2002, Aide et al. 2013). We included population growth as an underlying cause (a factor that influences the expansion of farming) as well as a proximate cause of deforestation due to its influence on other land uses not documented in the MODIS derived LULC maps. These activities include urban/rural infrastructure development as well as mining and logging, which occurs historically in the CGE (Mosquera 1978, Zapata 2013).

Road density was positively and indirectly related to deforestation hotspots. From 2002 to 2015, the best-preserved forest of the CGE were located along the Pacific coast from Panamá and Gulf of Urabá (Colombia) to the border between Colombia and Ecuador. This belt-like section of well-preserved forest has only two main roads that connect two Colombian cities (Quibdó in the Choco Department and Buenaventura in the Valle del Cauca Department) with the center of the country; the rest of this area is essentially roadless. The southern and northern ends of this well-preserved forest, where more deforestation occurs, are characterized with higher road densities, providing evidence of how roads are related to deforestation. Other studies in the tropical forests of South America have also related road development to increases in forest loss (Kleinschroth and Healey 2017). Roads are considered the first to penetrate well-preserved forests, opening the potential for forest harvests and subsequent environmental changes (Laurance and Useche 2009).

Average annual rainfall was negatively and indirectly related to deforestation hotspots. The rainiest place in the world corresponds to the belt of well-preserved forest of the CGE where the average annual rainfall ranges between 8000 to 13000 mm (Poveda
and Mesa 2000). This amount of rainfall on thin soils with low nutrient levels result in non-productive soils not well suited for farming. The interaction between high rainfall, non-productive soils, and low accessibility (lack of roads) results in a reduction of deforestation events, explaining the negative relationship between rainfall and deforestation hotspots. High rainfall has been related to a reduction in local deforestation in the Bolivian Amazon, where the expansion of mechanized agriculture occurs mainly in response to access to export markets, fertile soil, and intermediate rainfall conditions (Mueller et al. 2012). Likewise, the negative and indirect relationship between annual temperature and deforestation hotspots in our study area could occur due to the limitations imposed on farming activities by high temperatures, high rainfall, and soil degradation of the northern and southern portions of the CGE. Also, topographic slope was negatively and indirectly correlated to hotspots of deforestation; some scholars have found that places with abrupt topography tend to not be deforested in tropical areas due to the physical restrictions on the establishment of agriculture and grazing operations, as well as difficult access to markets for agricultural products (Coblentz and Keating 2008, Fagua et al. 2013, Sandel and Svenning 2013).

Our results show that Gross Domestic Product (GDP) at the municipality level was negatively and indirectly related to deforestation hotspots. Similar negative relations between deforestation and national GDP have been documented on the islands of Cuba, Hispaniola, Puerto Rico, Barbados, St. Kitts, and Nevis and Grenada and has been attributed to a decline of agriculture, rural population migration, and robust forest protection (Grau et al. 2003, Helmer et al. 2008, Álvarez-Berríos et al. 2013, Newman et al. 2018). Further, forest loss has been found to decrease with increases in national GDPs
in the protected areas of 56 countries over 4 continents, indicating that the effectiveness of environmental protection improves with better national economies because more resources can be invested in forest protection (Spracklen et al. 2015).

The CGE is located in some of the poorest municipalities of three developing countries where local governance and environmental institutions are weak, with few resources invested in environment conservation (DANE 2015, INEC-Ecuador 2015, MEF and WBG 2017). Large extensions of these poor municipalities are located adjacent to or within protected areas (such as, national forests, indigenous or black community reservations, national parks, etc.). These areas include the entire Panamanian CGE (15,335 km\(^2\)), the area along the pacific coast of Colombia (58,343 km\(^2\)), and four national parks in Ecuador (669 km\(^2\)). Consequently, forest protection is weak explaining the negative relationship between municipal GDPs and deforestation hot spots. Additionally, we observed that economies of the municipalities with significant clustering of deforestation hotspots in Ecuador and Colombia are based totally on extraction of natural resources (farming, mining, logging, principally), increasing pressure on forest cover. Not one municipality in Panamá presented significant clustering of deforestation hotspots indicating lower pressure on forest cover. Panamá has the lowest corruption (Transparency international 2015), the highest per capita GDP (WBG 2017), and some Panamanian municipalities in the CGE obtain significant resources from ecotourism (Brown 2007, Klytchnikova and Dorosh 2009, Mapes 2009).

Reforestation

The Bayesian SEM model of reforestation across the CGE showed that armed conflict was the variable most indirectly related to reforestation hotspots. Colombia
occupies 75.6% of the CGE, and this country has suffered internal armed conflicts that have been especially strong in several areas of its portion of the CGE. The Colombian municipalities where reforestation hotspots were significantly clustered (departments of Nariño, Choco, Antioquia, Cordoba) were zones strongly disputed among different armed groups, such as guerrillas (FARC and ELN), paramilitaries, drug dealers, and the Colombian army (CPDH 2006, MOE et al. 2008, Barreto 2009, FIP et al. 2014). These violent confrontations have resulted in lands abandoned by farmers, and subsequently producing significant regrowth of secondary vegetation. The relationship between armed conflict and growth of secondary vegetation in Colombia has also been assessed at the national level (Sanchez-Cuervo and Aide 2013).

Our results also found that several municipalities in Ecuador had a significant clustering of reforestation hotspots. Ecuador is a country without armed conflicts, suggesting that other underlying causes promote LULC transitions from farming to secondary vegetation. Colombian and Ecuadorian agricultural sectors declined during the first ten years of the current century, thus reducing the cultivated (grassland, crops, palm) area and allowing for the growth of secondary vegetation (BCE 2010, Minsalud 2011, Buitrago 2013, Marrugo 2013). Further, the Colombian and Ecuadorian CGE have been affected by floods during La Niña years, which were especially strong in 2010-2011. Consequently, many farming areas were lost allowing for secondary vegetation growth (BE 2010, IGAC 2011, SGR 2014). We also found that higher municipal GDP was related indirectly to reforestation hotspots. In regions like the Ecuadorian CGE, higher GDPs could indicate technical improvements in agriculture and cattle production,
resulting in intensification of these land uses on more productive lands and a reduction of agricultural land use on less productive lands where secondary vegetation could grow.

There is no evidence, however, that reforestation will lead to new mature forests. In fact, reforestation, generated by land abandonment represent earlier forest seral stages that will be deforested before they can become mature forests (Fagua and Ramsey 2018). When famers are displaced during violent confrontations in Colombia, other landowners supported by armed groups appropriate their lands, and in subsequent years remove regenerated natural vegetation to reestablish agricultural activities (PNUD 2011, Chaparro 2017). Also, when agricultural sectors declined in Colombia and Ecuador, farmers did not abandon their lands (since land possession is a main source of economic and political power in these countries), and forest regrowth is removed during periods of better agricultural return (Flórez et al. 2012, Sierra 2013).

Our Bayesian SEM modelling also showed that temperature was indirectly and negatively related to reforestation hotspots. Temperature (Clark et al. 2003) and water supply (Álvarez-Dávila et al. 2017) strongly determine vegetation growth in the tropics; the southern and northern sides of the CGE present higher annual temperature and lower annual rain, reducing secondary vegetation growth.

CONCLUSIONS

Our spatial analysis of LULC maps showed that deforestation and reforestation transitions were spatially clustered across the CGE forming hotspots. These hotspots were clustered around municipalities which provided many of the proximate and underlying causes of LULC change. While our analysis focused on the municipality level, our results indicate that causes related to other political or ecological subdivisions
of the landscape (country boundaries, sub-ecoregions, etc.) at different scales can be similarly assessed. The challenge to use these type of models is the availability of homogeneous information at a consistent temporal and spatial scale. By using a consistent LULC maps generated yearly using MODIS imagery, we were able to relate deforestation and reforestation hotspots to proximate and underlying causes. We show that increases in population growth and density of roads were the primary underlying causes related to deforestation hotspots, whereas underlying causes that limited access to forested lands, such as topographic slope and climate, were negatively related to these hotspots. Where reforestation is concerned, we identified three underlying causes that were positively related to hotspots. These causes included armed conflicts, gross domestic product, and average annual rainfall.

DATA ACCESSIBILITY

The land cover maps of the Chocó-Darien Global Ecoregion (South America) for 2002 and 2015, the map of deforestation hotspots 2002-2015, and the map of reforestation hotspots 2002-2015 are available from the repository of the Remote Sensing and GIS Laboratory at the Utah State University, College of Natural Resources in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/

LITERATURE CITED


Argüelles, L. 2010. Estrategia Institucional para la Prevención y Control de la Tala Ilegal en los Bosques del Darién en Panamá. WWF, Panama.


Armenteras, D., J. Maria Espelta, N. Rodriguez, and J. Retana. 2017. Deforestation dynamics and drivers in different forest types in Latin America: Three decades of


Fajardo, D. 2017. Estudio sobre los orígenes del conflicto social armado, razones de su persistencia y sus efectos más profundos en la sociedad colombiana. Espacio crítico, Bogota DC, Colombia.

Ferreira, N. C., L. G. Ferreira, and F. Miziara. 2007. Deforestation hotspots in the Brazilian Amazon: Evidence and causes as assessed from remot sensing and census data. Earth Interactions 11.


case study evidence – (LUCC Report Series; 4). LUCC International Project
Office, Louvain, Belgium.

Geist, H. J., and E. F. Lambin. 2002. Proximate causes and underlying driving forces of

Gentry, A. H. 1986. Species richness and floristic composition of Choco Region plant
communities. Caldasia 15:5.

Garcia, and M. F. Ordoñez. 2011. Análisis de tendencias y patrones espaciales de
deforestación en Colombia. Instituto de Hidrología, Meteorología y Estudios
Ambientales-IDEAM, Bogotá D.C., Colombia.

Grace, J. B., and K. A. Bollen. 2005. Interpreting the Results from Multiple Regression
and Structural Equation Models. The Bulletin of the Ecological Society of
America 86:283–295.

Grace, J. B., D. R. Schoolmaster Jr., G. R. Guntenspergen, A. M. Little, B. R. Mitchell,
implementation of structural equation modeling. Ecosphere 3.

scenarios of the carbon budget in postagricultural Puerto Rico (1936-2060).

2003. The ecological consequences of socioeconomic and land-use changes in


http://srtm.csi.cgiar.org/.


Plummer, M. 2015. JAGS Version 4.0.0 user manual.


http://wwf.panda.org/about_our_earth/ecoregions/chocodarien_moist_forests.cfm.


http://wwf.panda.org/about_our_earth/ecoregions/chocodarien_moist_forests.cfm.


Table 4.1. Variables evaluated as underlying causes of deforestation and reforestation.

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Description</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>Density of roads</td>
<td>Length of road (km) of first, second and third level per Km²</td>
<td>(IGAC 2017, IGNTG 2017, SNI 2017)</td>
</tr>
<tr>
<td>Accessibility</td>
<td>Density of rivers</td>
<td>Length of rivers (km) per Km²</td>
<td>(Lehner et al. 2008)</td>
</tr>
<tr>
<td>Biophysical environment</td>
<td>Topographic slope</td>
<td>Degrees of topographic slope from STRM 90</td>
<td>(Jarvis et al. 2008)</td>
</tr>
<tr>
<td>Biophysical environment</td>
<td>Standard deviation of altitude</td>
<td>Standard deviation of pixel values of STRM 90</td>
<td>(Jarvis et al. 2008)</td>
</tr>
<tr>
<td>Climate</td>
<td>Temperature</td>
<td>Mean annual temperature (°C) from 1950 to 2000</td>
<td>(Fick and Hijmans 2017)</td>
</tr>
<tr>
<td>Climate</td>
<td>Precipitation</td>
<td>Mean annual precipitation (mm/year) from 1970 to 2000</td>
<td>(Fick and Hijmans 2017)</td>
</tr>
<tr>
<td>Demography</td>
<td>Population growth</td>
<td>Human population change from 2000 to 2010</td>
<td>(DANE 2017a, INEC-Ecuador 2017, INEC-Panamá 2017a)</td>
</tr>
<tr>
<td>Armed conflicts</td>
<td>Armed conflicts</td>
<td>Number of armed conflicts between 2000 and 2015</td>
<td>(Sundberg and Melander 2013, Croicu and Sundberg 2017)</td>
</tr>
<tr>
<td>Armed conflicts</td>
<td>Fatalities of armed conflicts</td>
<td>Reported fatalities produced by the armed conflicts between 2000 and 2015</td>
<td>(Sundberg and Melander 2013, Croicu and Sundberg 2017)</td>
</tr>
<tr>
<td>Economy</td>
<td>Gross Domestic Product (GDP)</td>
<td>Gross Domestic Product growth at the municipality level between 2000 and 2010</td>
<td>(BCE 2017, DANE 2017b, INEC-Panamá 2017b)</td>
</tr>
</tbody>
</table>
Fig. 4.1. The Chocó-Darien Global Ecoregion (CGE): (A) estimated historical extent of Tropical Rain Forest (TRF) in South America: TRF-CGE (estimated TRF in CGE), TRF-Amz (estimated TRF in Amazon basin), and TRF-BrAt (estimated TRF in Brazilian Atlantic Forest). Land-use and land-cover (LULC) maps for the CGE at 2002 (B) and 2015 (C), these maps can be downloaded in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/
Fig. 4.2. Deforestation (A) and reforestation (B) hotspots on the land land-use and land-cover (LULC) map at 2015. Spatial correlation between deforestation and reforestation hotspot areas were mapping (C). These maps can be downloaded in the next link: http://data.gis.usu.edu/CGE/?prefix=CGE/
Fig. 4.3. Municipalities with significant clustering of deforestation and reforestation hotspots. (A) Moran scatter plot for deforestation; red diamonds represent the municipalities with significant influence on the Moran’s (I) global test for deforestation. (B) Moran scatter plot for reforestation; violet diamonds represent the municipalities with significant influence on the Moran’s (I) global test for reforestation. (C) Maps of municipalities with significant influence on the Moran’s (I) global test for deforestation. (D) Maps of municipalities with significant influence on the Moran’s (I) global test for reforestation.
Fig. 4.4. Bayesian Structural Equation model for deforestation. Unstandardized posterior path coefficients in black and standardized posterior path coefficients in red.
Fig. 4.5. Bayesian Structural Equation model for reforestation. Unstandardized posterior path coefficients in black and standardized posterior path coefficients in red.
CHAPTER 5
CONCLUSIONS AND SUMMARY

REMOTE SENSING ANALYSIS

By analyzing classification accuracies of land cover maps produced by four MODIS products (MOD13Q1, MYD13Q1, MOD09Q1, and MYD09Q1), I found that the MODIS platform (Terra or Aqua), the preprocessing algorithm (original MODIS or MAIAC), and the temporal optimization post-processing (Asymmetric Gaussian function or Savitzky and Golay function) could produce statistically significant differences in the ability of these products to map land cover. Data from the Terra platform had more accurate classifications than the data from the Aqua platform. Classifications from data that used the MAIAC algorithm had more accurate results than data generated by the original MODIS algorithm (09Q1 vs 13Q1 products). The asymmetric Gaussian function performed slightly better than the Savitzky and Golay function. These analyses included the identification of vegetation in two contrasting ecoregions (the Chocó-Darien in South America and the Great Basin in North America). The results show that these three sources of variation need to be considered in remote sensing analyses of vegetation when MODIS products are used.

The MOD13Q1 and MOD09Q1 products (Terra data preprocessed with the original MODIS and MAIAC algorithms respectively) coupled with the asymmetric Gaussian function to optimize the temporal data produced the most accurate classification of vegetation. I selected the MOD13Q1 product coupled with the Gaussian function to develop the land-use and land-cover maps for the CGE. While the MOD09Q1 product is pre-processed with a superior cloud, cloud shadow, aerosol algorithm (MAIAC), the
MOD13Q1 product includes the Enhanced Vegetation Index (EVI), which functions better in areas of dense canopy compared to the standard NDVI (Huete et al. 1999). The MOD09Q1 product does not include the spectral bands necessary to generate the EVI. The MOD13Q1 product provides the NDVI, EVI as well as three spectral bands (red, NIR, and MIR) while the MOD09Q1 only provides the red and NIR spectral bands, from which only NDVI can be calculated.

**LULC DYNAMIC AND PROXIMATE CAUSES OF FOREST CHANGES**

The generation and validation of land-use and land-cover (LULC) change annual maps from 2002 to 2015 resulted in high accuracy (Kappa of 0.87; SD= 0.008). After analyzing these maps, I found that LULC varied temporally within the CGE. Secondary vegetation increased from 2002 to 2010 whereas forest and agriculture (grassland, crop, and palm) decreased, showing a progressive replacement of forest and agriculture with secondary vegetation. However, some of these trends changed between 2010 to 2015; forest maintained its decreasing trend, but grassland increased while the other agricultural land use trends did not show significant changes. These results showed that deforestation transitions (changes from forest or secondary vegetation to farm land use) was higher between 2010-2015 compared to 2002-2010 with grassland as the main agriculture land cover that replaced woody vegetation (forest and secondary vegetation) between 2010-2015 across the CGE. The increased loss of forest after 2010 should be an important concern for the preservation of CGE biodiversity due to the high levels of species richness and endemism which are difficult to recover through reforestation. In other words, secondary forests evolving from secondary vegetation would have decreased biodiversity and different species assemblages (Norden et al. 2015).
The temporal LULC dynamic shown in this study identified some variation between the countries that share the CGE. Grassland conversion was the most frequent cause of deforestation from 2002 to 2010 for the entire CGE (63%) and for each country (73% in Panamá, 65% in Colombia, and 58% in Ecuador). Grassland was also the most frequent land cover that reverted to secondary forest (reforestation) across the entire CGE (50%), as well as in Panamá (65%), and Colombia (58%). Ecuador, however, showed that crops were the most frequent land use type to convert to secondary vegetation (55%).

For the 2010 to 2015 period, grassland was also the most frequent proximate cause of deforestation across the CGE (73%) as well as in every country (94% in Panamá, 76% in Colombia, and 59% in Ecuador). Grassland was also the most frequent land cover that converted to secondary vegetation during 2010-2015 for the CGE (47%) and in two countries (68% to Panamá and 53% to Colombia). In Ecuador, agriculture to secondary vegetation was again the highest reforestation transition (55%). These temporal/regional variations of LULC change need to be considered when developing CGE-wide management plans aimed at preserving biodiversity and ecosystem services.

HOTSPOTS AND UNDERLYING CAUSES OF FOREST CHANGES

Based on the LULC maps from 2002 to 2015, I found that deforestation and reforestation transitions were spatially clustered across the CGE forming hotspots; areas that exhibit significant spatial correlation of deforestation or reforestation transitions, showing that proximate and underlying causes of forest change could vary spatially across the CGE. I also found that areas identified as hotspots of deforestation were spatially correlated to areas identified as hotspots of reforestation, showing that both
transition types tended to occur in close proximity. This may suggest that after cutting down mature forest, much of the land was allowed to transition to secondary vegetation.

Hotspots were also clustered around municipalities which suggested the causes of LULC change. Eighteen municipalities located near the border of Colombia and Ecuador presented the most significant aggregation of deforestation hotspots; three in the Colombian side and 15 in the Ecuadorian side. On the other hand, 34 municipalities located in three areas of Colombia as well as the area around the Colombian and Ecuadorian border showed significant clustering of reforestation. Interestingly, I found that 11 municipalities presented significant clustering of both deforestation and reforestation hotspots: in Colombia - Barbacoas, Mosquera, and Tumaco; and in Ecuador - Eloy Alfaro, Las Golondrinas, Pedernales, Puerto Quito, Quininde, Rio Verde, San Lorenzo and Tulcan. The aggregation of deforestation and reforestation hotspots in the same municipalities also support the statement that much of the deforested land would be not used in agriculture allowing the subsequent growth of secondary vegetation.

Increases in population growth and road density were the primary underlying causes related to deforestation hotspots in the CGE, whereas underlying causes that limited access to forested lands, such as topography and climate, were negatively related to these hotspots. Armed conflicts, gross domestic product, and average annual rainfall were the three main underlying causes positively related to hotspots of reforestation.

The previous relations were identified by my novel analysis using a Bayesian SEM to relate hotspots with causes of forest changes estimated in the municipality level. The variables that I used as underlying causes of forest change have been analyzed in many other studies, showing that their interactions and effects on forest changes are
complex and change through the land systems (Meyfroidt et al. 2018). Thus, more research is necessary to establish causal relations among the underlying causes that I researched in the CGE.

My results indicate that underlying causes related to other political or ecological subdivisions of the landscape (country boundaries, sub-ecoregions, etc.) at different scales can be similarly assessed using my novel approach. However, the challenge to use these type of models is the availability of homogeneous information at a consistent temporal and spatial scale. By using a consistent LULC maps, as the maps generated in this dissertation, it is possible to relate deforestation and reforestation hotspots to direct and indirect causes.

LITERATURE CITED


Neotropical forests are as uncertain as they are predictable. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 112:8013–8018.
Table S2.1. Selection of the windows size for the temporal optimization using the Asymmetric Gaussian function. We assessed windows sizes from 3 to 7 data and selected the size that maximize the accuracy (Kappa-K) of the classification. To estimate K, 100 iterations were performed for every Random Forest classification, each with 5-fold cross-validation. Variation of window size did not affect Kappa significantly (P-values > 0.05).

<table>
<thead>
<tr>
<th>Window Size</th>
<th>The Great Basin</th>
<th>The Chocó-Darien</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MO 13Q1</td>
<td>MOD 09Q1</td>
</tr>
<tr>
<td>3</td>
<td>71.8 +/- 0.6</td>
<td>68.8 +/- 0.7</td>
</tr>
<tr>
<td>5</td>
<td>71.9 +/- 0.6</td>
<td>68.9 +/- 0.6</td>
</tr>
<tr>
<td>7</td>
<td>71.9 +/- 0.5</td>
<td>68.7 +/- 0.5</td>
</tr>
<tr>
<td>9</td>
<td>71.8 +/- 0.5</td>
<td>68.8 +/- 0.6</td>
</tr>
</tbody>
</table>
Table S2.2. Selection of filter order and Weighted Moving Average Filter (WMAF) for the temporal optimization using the Savitzky-Golay function. We assessed WMAF sizes from 3 to 9 and filter order from 2 to 4. We selected the WMAF and filter order that maximized Kappa (K). To estimate K, 100 iterations were performed for every Random Forest classification, each one with 5-fold cross-validation. Variation of filter order and WMAF by MODIS products did not significantly affect Kappa (P-values > 0.05).

<table>
<thead>
<tr>
<th>Filter order</th>
<th>WMAF</th>
<th>The Great Basin</th>
<th>The Choco Global</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MOD 13Q1</td>
<td>MYD 13Q1</td>
<td>MOD 09Q1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>70.3</td>
<td>68.9</td>
<td>72.8</td>
</tr>
<tr>
<td></td>
<td>+/-0.6</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>2</td>
<td>71.9</td>
<td>67.6</td>
<td>72.6</td>
</tr>
<tr>
<td>5</td>
<td>+/-0.4</td>
<td>+/-0.6</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>2</td>
<td>70.2</td>
<td>66.9</td>
<td>72.6</td>
</tr>
<tr>
<td>7</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
<td>+/-0.4</td>
</tr>
<tr>
<td>2</td>
<td>69.4</td>
<td>67.4</td>
<td>72.5</td>
</tr>
<tr>
<td>9</td>
<td>+/-0.5</td>
<td>+/-0.6</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>3</td>
<td>70.2</td>
<td>67.5</td>
<td>72.5</td>
</tr>
<tr>
<td>3</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>3</td>
<td>70.3</td>
<td>66.8</td>
<td>72.6</td>
</tr>
<tr>
<td>7</td>
<td>+/-0.4</td>
<td>+/-0.6</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>3</td>
<td>69.4</td>
<td>67.4</td>
<td>72.5</td>
</tr>
<tr>
<td>9</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
<td>+/-0.4</td>
</tr>
<tr>
<td>4</td>
<td>71.9</td>
<td>68.7</td>
<td>72.6</td>
</tr>
<tr>
<td>3</td>
<td>+/-0.5</td>
<td>+/-0.6</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>4</td>
<td>71.8</td>
<td>68</td>
<td>72.9</td>
</tr>
<tr>
<td>7</td>
<td>+/-0.5</td>
<td>+/-0.6</td>
<td>+/-0.5</td>
</tr>
<tr>
<td>4</td>
<td>70.6</td>
<td>66.9</td>
<td>72.4</td>
</tr>
<tr>
<td>9</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
<td>+/-0.5</td>
</tr>
</tbody>
</table>
Table S3.1. High spatial resolution imagery used to the visually interpreting of Land-use and land-cover (LULC) classes.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Band Resolution</th>
<th>Bands</th>
<th>Panchromatic Resolution</th>
<th>Cover (Km²)</th>
<th>Year</th>
<th>Image name</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeoEye-1</td>
<td>2.2</td>
<td>4</td>
<td>0.5</td>
<td>1000</td>
<td>2012</td>
<td>12SEP03155046-P2AS-053711282080</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11MAR12153945-M2AS-053711282090</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12APR05154853-M2AS-054758991030</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12APR02153914-P2AS-053738769050</td>
</tr>
<tr>
<td>Ikonos</td>
<td>3.2</td>
<td>4</td>
<td>0.8</td>
<td>568</td>
<td>2006</td>
<td>po1473219</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>189</td>
<td>2005</td>
<td>po_1472872</td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>516</td>
<td>2007</td>
<td>07AUG01154941-M2AS-0546</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12JUN29145741-M2AS-0546</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12FEB17145933-M2AS-053738769040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>02JUN12154300-M2AS-053738769030</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>03FEB01154326-M2AS-053711282070</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10SEP101001000B9-0ED00-MS1</td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>400</td>
<td>2002</td>
<td>Panamá_net_point_463.csv</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>633</td>
<td>2003</td>
<td>Panamá2_net_point_463.csv</td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>172</td>
<td>2010</td>
<td>10SEP101001000B9-0ED00-MS1</td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>38</td>
<td>2011</td>
<td>Panamá_net_point_463.csv</td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>170</td>
<td>2011</td>
<td>Panamá2_net_point_463.csv</td>
</tr>
<tr>
<td>Product</td>
<td>Geo1</td>
<td>Geo2</td>
<td>Geo3</td>
<td>Date</td>
<td>Area</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>725 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QuickBird</td>
<td>2.4</td>
<td>4</td>
<td>0.6</td>
<td>54 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>2526 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>222 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>270 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>2018 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>2106 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WORLDVI EW-2</td>
<td>2.2</td>
<td>8</td>
<td>0.5</td>
<td>2320 2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area</td>
<td></td>
<td></td>
<td></td>
<td>20708</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S3.2. Table of response and predictor variables used in the Random Forest classification. This table is in CSV format that can be download in the next link https://data.gis.usu.edu/?prefix=CDLCC/

Table S3.3. Codes S3.3. R codes used for the Random Forest Classification (Run with Table S3.2).

```
########################################### Random Forest classification
library(randomForest)
library(ModelMap)

model.type <- "RF"
qdatafn <- "LA_TABLA S2.csv"
qdata.trainfn <- "VModelMapData_TRAIN_categorical.csv"
qdata.testfn <- "VModelMapData_TEST_categorical.csv"
folder <- getwd()
get.test( proportion.test=0.2, #Percentage for the cross validation
  qdatafn=qdatafn,
  seed=42,
  folder=folder,
  qdata.trainfn=qdata.trainfn,
  qdata.testfn=qdata.testfn)

MODELfn <- "VMModelMapEx3"

predList <- c( "evi",
  "mir",
  "ndvi",
  "nir",
  "red",
  "slope",
  "altitud"
)

response.name <- "cover"
response.type <- "categorical"
seed <- 44
unique.rowname <- "ID"

##########To build maps (Use only if you have the raster predictors)
#rastLUTfn <- "VMModelMapData_LUT.csv"
#rastLUTfn <- read.table( rastLUTfn,
#  #header=FALSE,
#  #sep="",
#  #stringsAsFactors=FALSE)```
#rastLUTfn[,1] <- paste(folder,rastLUTfn[,1],sep="/")

# Creation of the model
model.obj.ex3 <- model.build( model.type=model.type,
    qdata.trainfn=qdata.trainfn,
    folder=folder,
    unique.rowname=unique.rowname,
    MODELfn=MODELfn,
    predList=predList,
    predFactor=FALSE,
    response.name=response.name,
    response.type=response.type,
    seed=seed)

# Model diagnosis
model.pred.ex3 <- model.diagnostics( model.obj=model.obj.ex3,
    qdata.testfn=qdata.testfn,
    folder=folder,
    MODELfn=MODELfn,
    unique.rowname=unique.rowname,
    prediction.type="TEST", # By type TEST, the validation predictions
    # will be made on the test set provided by qdata.testfn.
    # or by type "OOB" you get the kappa of the out-of-bag (OOB)
    validations
    device.type="jpeg",
    cex=1.2)

# Commands for doing the map (Use only if you have the raster predictors)
#model.mapmake( model.obj=model.obj.ex3,
#    folder=folder,
#    MODELfn=MODELfn,
#    rastLUTfn=rastLUTfn,
#    na.action="na.omit")

# To allocate codes to my the land covers (Use only if you have the raster predictors)

#MAP.CODES <- read.table( paste(MODELfn,"_map_key.csv",sep=""),
#    header=TRUE,
#    sep="",
#    stringsAsFactors=FALSE)

#MAP.CODES
#write.csv(MAP.CODES, file = "MAP_CODES")

# To allocate codes to my the land covers (Use only if you have the raster predictors)
Table S3.4. Distribution of samples across land-use/land-cover (LULC) classes before and after sampling reduction.

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Class distribution</th>
<th>Class distribution after Woody vegetation reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woody vegetation</td>
<td>14228</td>
<td>1144</td>
</tr>
<tr>
<td>Grassland</td>
<td>1144</td>
<td>1144</td>
</tr>
<tr>
<td>Crop</td>
<td>404</td>
<td>404</td>
</tr>
<tr>
<td>Palm</td>
<td>743</td>
<td>743</td>
</tr>
<tr>
<td>Urban</td>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>Water</td>
<td>1123</td>
<td>1123</td>
</tr>
<tr>
<td>Wetland</td>
<td>796</td>
<td>796</td>
</tr>
</tbody>
</table>
Table S3.5. Confusion matrix of the cross validation from the original data. Kappa, commissions and omissions are in the matrix.

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Kappa.sd</th>
<th>Observed</th>
<th>Woody Vegetation</th>
<th>Grassland</th>
<th>Crop</th>
<th>Palm</th>
<th>Urban</th>
<th>Water</th>
<th>Wetland</th>
<th>total</th>
<th>Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.876</td>
<td>0.008</td>
<td></td>
<td>Woody vegetation</td>
<td>283</td>
<td>3</td>
<td>39</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>15</td>
<td>2918</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Grassland</td>
<td>14</td>
<td>151</td>
<td>11</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Crop</td>
<td>2</td>
<td>6</td>
<td>42</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Palm</td>
<td>11</td>
<td>6</td>
<td>1</td>
<td>108</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Urban</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Water</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>229</td>
<td>2</td>
<td>233</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Wetland</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>154</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>total</td>
<td>287</td>
<td>2</td>
<td>205</td>
<td>65</td>
<td>144</td>
<td>21</td>
<td>232</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Omission</td>
<td>0.01</td>
<td>0.26</td>
<td>0.35</td>
<td>0.25</td>
<td>0.05</td>
<td>0.01</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MAUC</td>
<td>0.983</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
</tr>
</tbody>
</table>
Table S3.6. Confusion matrix of the cross validation from the data when forest class is reduced as the grassland class number. Kappa, commission and omission are in the matrix.

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Kappa.sd</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.873</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Woody Vegetation</th>
<th>Grassland</th>
<th>Crop</th>
<th>Palm</th>
<th>Urban</th>
<th>Water</th>
<th>Wetland</th>
<th>Total</th>
<th>Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woody vegetation</td>
<td>185</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>201</td>
<td>0.08</td>
</tr>
<tr>
<td>Grassland</td>
<td>7</td>
<td>200</td>
<td>17</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>238</td>
<td>0.16</td>
</tr>
<tr>
<td>Crop</td>
<td>1</td>
<td>11</td>
<td>57</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>84</td>
<td>0.32</td>
</tr>
<tr>
<td>Palm</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>125</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>140</td>
<td>0.11</td>
</tr>
<tr>
<td>Urban</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>25</td>
<td>0.12</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>224</td>
<td>1</td>
<td>226</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Wetland</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>156</td>
<td>169</td>
<td>0.08</td>
</tr>
<tr>
<td>Total</td>
<td>202</td>
<td>229</td>
<td>83</td>
<td>148</td>
<td>23</td>
<td>230</td>
<td>168</td>
<td>1083</td>
<td>PCC</td>
</tr>
<tr>
<td>Omission</td>
<td>0.08</td>
<td>0.13</td>
<td>0.31</td>
<td>0.16</td>
<td>0.04</td>
<td>0.03</td>
<td>0.07</td>
<td>PCC</td>
<td>0.89</td>
</tr>
<tr>
<td>MAUC</td>
<td>0.985</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
<td>cmx</td>
</tr>
</tbody>
</table>
Table S3.7. Trends of land-use and land-cover (LULC) changes between 2002-2010.

<table>
<thead>
<tr>
<th>Year</th>
<th>Forest</th>
<th>Secondary vegetation</th>
<th>Grass</th>
<th>Crop</th>
<th>Palm plantation</th>
<th>Settlement</th>
<th>Water</th>
<th>Wetland</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>111270</td>
<td>12473</td>
<td>33367</td>
<td>11970</td>
<td>9603</td>
<td>1666</td>
<td>4950</td>
<td>9439</td>
<td>194738</td>
</tr>
<tr>
<td>2003</td>
<td>105564</td>
<td>17588</td>
<td>33281</td>
<td>13802</td>
<td>8561</td>
<td>1263</td>
<td>5010</td>
<td>9669</td>
<td>194738</td>
</tr>
<tr>
<td>2004</td>
<td>102818</td>
<td>23689</td>
<td>33798</td>
<td>9853</td>
<td>6163</td>
<td>1662</td>
<td>5523</td>
<td>11231</td>
<td>194738</td>
</tr>
<tr>
<td>2005</td>
<td>100213</td>
<td>26925</td>
<td>31315</td>
<td>10191</td>
<td>5673</td>
<td>2810</td>
<td>6795</td>
<td>10816</td>
<td>194738</td>
</tr>
<tr>
<td>2006</td>
<td>98508</td>
<td>32471</td>
<td>29323</td>
<td>7682</td>
<td>5889</td>
<td>3267</td>
<td>6994</td>
<td>10604</td>
<td>194738</td>
</tr>
<tr>
<td>2007</td>
<td>96820</td>
<td>30259</td>
<td>29233</td>
<td>12370</td>
<td>5534</td>
<td>1508</td>
<td>8083</td>
<td>10930</td>
<td>194738</td>
</tr>
<tr>
<td>2008</td>
<td>95657</td>
<td>35839</td>
<td>28343</td>
<td>8640</td>
<td>3752</td>
<td>1962</td>
<td>8789</td>
<td>11755</td>
<td>194738</td>
</tr>
<tr>
<td>2009</td>
<td>94382</td>
<td>34413</td>
<td>33378</td>
<td>6802</td>
<td>4552</td>
<td>3381</td>
<td>7718</td>
<td>10112</td>
<td>194738</td>
</tr>
<tr>
<td>2010</td>
<td>93430</td>
<td>39140</td>
<td>28362</td>
<td>9157</td>
<td>3615</td>
<td>2219</td>
<td>7678</td>
<td>11137</td>
<td>194738</td>
</tr>
<tr>
<td>2011</td>
<td>91661</td>
<td>39607</td>
<td>25995</td>
<td>8621</td>
<td>2537</td>
<td>3282</td>
<td>9972</td>
<td>13063</td>
<td>194738</td>
</tr>
<tr>
<td>2012</td>
<td>89674</td>
<td>34482</td>
<td>35336</td>
<td>6532</td>
<td>2353</td>
<td>3245</td>
<td>9746</td>
<td>13370</td>
<td>194738</td>
</tr>
<tr>
<td>2013</td>
<td>85928</td>
<td>34569</td>
<td>39661</td>
<td>8710</td>
<td>1877</td>
<td>4490</td>
<td>7763</td>
<td>11740</td>
<td>194738</td>
</tr>
<tr>
<td>2014</td>
<td>84443</td>
<td>36334</td>
<td>42882</td>
<td>6066</td>
<td>1909</td>
<td>4383</td>
<td>7194</td>
<td>11527</td>
<td>194738</td>
</tr>
<tr>
<td>2015</td>
<td>83312</td>
<td>43894</td>
<td>38477</td>
<td>6567</td>
<td>4151</td>
<td>2910</td>
<td>5115</td>
<td>10311</td>
<td>194738</td>
</tr>
</tbody>
</table>
Table S3.8. Deforestation (proximate causes of deforestation) and reforestation transitions 2002-2010.

<table>
<thead>
<tr>
<th></th>
<th>Deforested</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGE</td>
<td>Col</td>
<td>Ecu</td>
<td>Pan</td>
<td>CGE</td>
<td>Col</td>
<td>Ecu</td>
<td>Pan</td>
<td>CGE</td>
<td>Col</td>
<td>Ecu</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>4571</td>
<td>2970</td>
<td>1507</td>
<td>94</td>
<td>9008</td>
<td>7551</td>
<td>1405</td>
<td>51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(63.2)</td>
<td>(65.8)</td>
<td>(58.3)</td>
<td>(73)</td>
<td>(50.7)</td>
<td>(58)</td>
<td>(30)</td>
<td>(65.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>1583</td>
<td>629</td>
<td>919</td>
<td>35</td>
<td>4029</td>
<td>1398</td>
<td>2603</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(21.9)</td>
<td>(13.9)</td>
<td>(35.6)</td>
<td>(27)</td>
<td>(22.7)</td>
<td>(10.7)</td>
<td>(55.5)</td>
<td>(34.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>994</td>
<td>851</td>
<td>143</td>
<td>0</td>
<td>4744</td>
<td>4064</td>
<td>680</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(13.7)</td>
<td>(18.8)</td>
<td>(5.5)</td>
<td>(0)</td>
<td>(26.7)</td>
<td>(31.2)</td>
<td>(14.5)</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Settlement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>80</td>
<td>64</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(1.1)</td>
<td>(1.4)</td>
<td>(0.6)</td>
<td>(0.1)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>7228</td>
<td>4514</td>
<td>2585</td>
<td>129</td>
<td>17783</td>
<td>13014</td>
<td>4689</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S3.9. Deforestation (proximate causes of deforestation) and reforestation transitions 2010-2015.

<table>
<thead>
<tr>
<th></th>
<th>Deforested</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGE</td>
<td>Col</td>
<td>Ecu</td>
<td>Pan</td>
<td>CGE</td>
<td>Col</td>
<td>Ecu</td>
<td>Pan</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>11141</td>
<td>8888</td>
<td>1909</td>
<td>344</td>
<td>4325</td>
<td>2667</td>
<td>1590</td>
<td>68</td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(73.6)</td>
<td>(77)</td>
<td>(59.3)</td>
<td>(94.6)</td>
<td>(47.4)</td>
<td>(53.1)</td>
<td>(39.8)</td>
<td>(68.7)</td>
</tr>
<tr>
<td>Crop</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>1823</td>
<td>882</td>
<td>926</td>
<td>14</td>
<td>3347</td>
<td>1084</td>
<td>2231</td>
<td>31</td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(12)</td>
<td>(7.6)</td>
<td>(28.8)</td>
<td>(4.1)</td>
<td>(36.7)</td>
<td>(21.6)</td>
<td>(55.8)</td>
<td>(31.3)</td>
</tr>
<tr>
<td>Palm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>1953</td>
<td>1604</td>
<td>348</td>
<td>0</td>
<td>1447</td>
<td>1268</td>
<td>178</td>
<td>0</td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(13)</td>
<td>(14)</td>
<td>(10.8)</td>
<td>(0)</td>
<td>(15.9)</td>
<td>(25.3)</td>
<td>(4.5)</td>
<td>(0)</td>
</tr>
<tr>
<td>Settlement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>228</td>
<td>187</td>
<td>36</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(1.5)</td>
<td>(1.6)</td>
<td>(1.1)</td>
<td>(1.3)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area (km²)</td>
<td>15145</td>
<td>11561</td>
<td>3220</td>
<td>363</td>
<td>9120</td>
<td>5020</td>
<td>4000</td>
<td>99</td>
</tr>
<tr>
<td>Prop. of total (%)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
<td>(100)</td>
</tr>
</tbody>
</table>
Appendix S1: **Fig. S4.1.** Relation between global Moran’s test and matrix size for A) deforestation (p=0.001; $R= -0.94$) and B) reforestation (p=0.001; $R= -0.97$). Both graphics showed a significant negative exponentially shape.

Appendix S2: Data S1. Data to run the Bayesian SEMs for deforestation and reforestation (CSV file). This data can be download in the next link: https://data.gis.usu.edu/?prefix=CDLCC/
Appendix S3: Metadata S1. R code to build the Bayesian SEM for deforestation and reforestation; Run with data of the Appendix S2: Data S1 (code below or in the .R file of the appendix).

library(lavaan)
library(blavaan)
library("runjags")
library(qgraph)
library(standardize)

print(runjags.options())
runjags.options(silent.jags=TRUE, silent.runjags=TRUE)

#Set directory
setwd(""")

TAB = read.csv("Appendix B.csv")
str(TAB)
ncol(TAB)

#To select municipalities with more than 20km2 in the CGE and more than 15km2 of forest at 2002
TABL= subset(TAB,AREA >= 20 & Forest_in_2002 >= 15)
str(TABL)
head(TABL)

TABLA = TABL[,c(1,4:23)]
attach(TABLA)

# To reduce variance, the variables were divide (10,100,or 1000) to reach one or two digits # (Rosseel 2012)

sd_alt   = TABLA$sd_alt/100
mean_slope = TABLA$mean_slope
mean_tempe = TABLA$mean_tempe/10
mean_rain = TABLA$mean_rain/1000
mean_road = TABLA$mean_road /10
mean_river = TABLA$mean_river/10
violent_events = TABLA$vviolent_events/10
deaths = TABLA$deaths/100

HOTSPOTS_defo_15 = TABLA$HOTSPOTS_defo_15/1000
HOTSPOTS_REFO_15 = TABLA$HOTSPOTS_REFO_15/1000

#These variable show the farming that changed to secondary vegetation
Ch_SEC_VEG_2015 = TABLA$Ch_SEC_VEG_2015/1000

# These variables show that forest that changed to farming covers
Ch_GRASS_2015 = TABLA$Ch_GRASS_2015/1000
Ch_CROP_2015 = TABLA$Ch_CROP_2015/1000
Ch_PALM_2015 = TABLA$Ch_PALM_2015/1000
Ch_farm_2015 = Ch_GRASS_2015 + Ch_CROP_2015 + Ch_PALM_2015

AREA = TABLA$AREA
Forest_in_2002 = TABLA$Forest_in_2002

POPULATION_2000 = TABLA$POPULATION_2000
POPULATION_2010 = TABLA$POPULATION_2010
GDP_Growth_Mun_2000_10 = TABLA$GDP_Growth_Mun_2000_10

###
TABLA_2 =
data.frame(Country, sd_alt, mean_slope, mean_tempe, mean_rain, mean_road, mean_river, violent_events, deaths, HOTSPOTS_defo_15, HOTSPOTS_REFO_15, Ch_SEC_VEG_2015, Ch_GRASS_2015, Ch_CROP_2015, Ch_PALM_2015, Ch_farm_2015, AREA, POPULATION_2000, POPULATION_2010, Population.Growth.Rate, GDP_Growth_Mun_2000_10)

TABLA2 = TABLA_2[complete.cases(TABLA_2), ]
str(TABLA2)

# SEM model for deforestation.
# To run the Bayesian SEM, we follow the codes by (Merkle and Rosseel 2016).

modelo.3.7 <- 'HOTSPOTS_defo_15 ~ Ch_farm_2015
Ch_farm_2015 ~ Population.Growth.Rate + mean_slope + mean_road + GDP_Growth_Mun_2000_10 + mean_rain + mean_tempe
HOTSPOTS_defo_15 ~ Population.Growth.Rate'
modelo.3.7.fit <- bsem(modelo.3.7, data = TABLA2, convergence = "manual",
                      adapt = 1000, burnin = 4000, sample = 10000)
summary(modelo.3.7.fit, stand = T)
fitMeasures(modelo.3.7.fit)

# SEM model for Reforestation.
modelo.3.2 <- 'HOTSPOTS_REFO_15 ~ Ch_SEC_VEG_2015
Ch_SEC_VEG_2015 ~ Population.Growth.Rate+mean_road+GDP_Growth_Mun_2000_10+violent_events'
modelo.3.2.fit <- bsem(modelo.3.2, data=TABLA2,convergence="manual",
                        adapt=1000,burnin=4000,sample=10000)
summary(modelo.3.2.fit, stand=T)
fitMeasures(modelo.3.2.fit)

#SEM model for deforestation adding countries as categorical variable.

modelo.4.9 <- 'HOTSPOTS_defo_15 ~ Ch_farm_2015
Ch_farm_2015 ~ Population.Growth.Rate+mean_slope+mean_road+GDP_Growth_Mun_2000_10+mean_rain+mean_tempe+prior("dunif(0,3)")*Country
HOTSPOTS_defo_15 ~ Population.Growth.Rate'
modelo.4.9.fit <- bsem(modelo.4.9, data=TABLA2,convergence="manual",
                       adapt=1000,burnin=4000,sample=10000)
summary(modelo.4.9.fit, stand=T)
fitMeasures(modelo.4.9.fit)

#SEM model for Reforestation adding countries as categorical variable.

modelo.5.2 <- 'HOTSPOTS_REFO_15 ~ Ch_SEC_VEG_2015
Ch_SEC_VEG_2015 ~ Population.Growth.Rate+mean_road+GDP_Growth_Mun_2000_10+violent_events+prior("dunif(0,3)")*Country'
modelo.5.2.fit <- bsem(modelo.5.2, data=TABLA2,convergence="manual",
                       adapt=1000,burnin=4000,sample=10000)
summary(modelo.5.2.fit, stand=T)
fitMeasures(modelo.5.2.fit)
CURRICULUM VITAE

JOSE CAMILO FAGUA
GIS and Remote Sensing lab, Department of Wildland Resources, Utah State University
5230 Old Main Hill, NR 206, Logan, UT
Telephone: (435)799-7064
camilo.fagua@aggiemail.usu.edu

EDUCATION
  Dissertation: Geospatial Modeling of Land Cover Change in the Chocó-Darien Global Ecoregion of South America, one of most rainy areas in the world.
- M.S. 2010. University of Puerto Rico, Río Piedras Campus, San Juan, PR.
- Specialization in Geographic Information Systems. 2010. Universidad Distrital, Bogotá, Colombia.

RESEARCH INTERESTS
Spatial Analysis; Remote Sensing; Geographic Information Systems; Species Distribution Models, Ecology.

HONORS AND SCHOLARSHIPS
- Ecology Center Graduate Research Award ($5,000). 2017. Utah State University, Logan, UT, US.
- Digital Globe Award (20,000 km² of imagery). 2014. Imagery of high spatial resolution. Digital Globe, Washington DC, US.
- Ecology Center Graduate Research Award ($2,500). 2014. Utah State University, Logan, UT, US.
- Award to attend the course: Conservation genetics and their applications ($2,000). 2009. Red Latino Americana de Botánica, Bariloche, Argentina.
- Thesis research award ($12,000). 2008-2009. Graduate Studies Office, University of Puerto Rico. San Juan, PR, US.
- Award to attend the course: Conservation of important areas for plants in Latin America (2,000). 2008. Red Latino Americana de Botánica, Santo Domingo, República Dominicana.
• Award to attend to the course: Ecology of Amazon ecosystems ($3,600). 2003. Organization for Tropical Studies, Iquitos, Peru.

PUBLICATIONS
Fagua, J. C., Ferreira, R.B, & Guarino, C. A new approach to estimate robust absence records for species distribution modeling (target journal ECOLOGY AND EVOLUTION).
Fagua, J.C., Baggio, J.A, & Ramsey, R.D. Drivers of forest cover changes in the Chocó-Darien Global Ecoregion of South America. (in revision ECOSPHERE).
Fagua, J.C., Ramsey, R.D. Geospatial Modeling of Land Cover Change in the Chocó-Darien Global Ecoregion of South America; one of most rainy areas in the world. (in revision PLOS ONE).

Fagua, J. C., & Fagua, G. 2004. Protocolos para el manejo de las plantas amazónicas medicinales sangre de drago (Croton lechleri) y uña de gato (Uncaria tomentosa), en sistemas de aprovechamiento comercial. In M. T. Becerra (Ed.), Lineamientos para hacer planes de manejo de sistemas de aprovechamiento de recursos naturales en condiciones in-situ (pp. 115–132). Bogota DC, Colombia.

PROFESSIONAL EXPERIENCE


**TEACHING EXPERIENCE**

- Special guess to teach Imagery classification for Geospatial Analysis in R. 4/2018. Department of Environment and Society, Utah State University. Logan, Utah, US.

**PROFESSIONAL PRESENTATIONS**

- Annual Meeting Ecological Society of America. Portland, Oregon, US. 2017. Geospatial Modeling of Land Cover Change in the rainiest area on the earth, the Chocó-Darien Global Ecoregion (South America) (Contributed talk).
- Colombian Geomatics. Bogotá DC, Colombia. 2011. Two methodologies for generating digital terrain models (DTM) and digital elevation models (DEM) using LiDAR data and freeware (Contributed talk).

TRAINING & WORKSHOPS
• Species Distribution Modelling Using R. Utah State University. Logan, UT, US. 2017.
• Analysis of land change modeler and land use using TerrSet. Clark University, IGAC, Gordon and Betty Moore Foundation. Bogotá, Colombia. 2011.

SKILLS & TECHNIQUES
• Languages:
  English (Fluent)
  Spanish (Native)
• Programming Languages:
  R: Advanced proficiency in R (4 years of experience).
  Python: handling datasets, statistical analysis, and spatial analysis ArcPy.
• Software:
  ArcGIS, QGIS, ERDAS, FRAGSTATS, TerrSet, Microsoft Office suite.

PROFESSIONAL SERVICE ACTIVITIES
• Mentor, Ecological Society of America, SEEDS Program, 2017.
• Manuscript Reviewer, Revista de Biología Tropical, 2016.
• Manuscript Reviewer, Biodiversity and Conservation, 2015.