UNIVERSITÀ DEGLI STUDI DI MACERATA

CORSO DI DOTTORATO DI RICERCA IN
METODI QUANTITATIVI PER LA POLITICA ECONOMICA
CICLO XXXIV

TITOLO DELLA TESI

"AN ECONOMETRIC ANALYSIS OF COLOMBIA'S NARCODEFORESTATION AND GLYPHOSATE ASPERSION POLICY"

SUPERVISORI DI TESI
Chiar.ma Prof.ssa Luisa Scaccia
Chiar.mo Prof. Luca Salvati

DOTTORANDA
Dott.ssa Perla Irasema Rivadeneyra García

COORDINATORE
Chiar.mo Prof. Luca DeBenedictis

ANNO 2022

Dedication

I dedicate this thesis to my husband Daniele and my son Leonardo, for their unconditional love, support, encouragement and company. They are my motive and the core of my world. A special feeling of gratitude to my loving parents, Gerardo and Blanca, whose trust and constant support have always been my safe place. Finally, a special thanks to my parents-in-law for their support and affection.

Acknowledgments

I would like to express my sincere appreciation to all those who have helped me put the ideas for this thesis, well beyond the level of simplicity and into something concrete.
I would like to express my special thanks of gratitude to my supervisor Luisa Scaccia, for her effort, guidance, teachings and support, and who also became a friend in this journey.
A special thanks to professor Luca Salvati for his valuable ideas, corrections and guidance that were essential to this thesis.
And last, but not least, a sincere thank you to professor De Benedictis for his support and understanding.

Table of Contents

List of Figures

List of Tables

Introduction

Colombia is the second most biodiverse country in the world and the most biodiverse per square kilometer. It is home to 10% of the species worldwide, many of which are endemic (World Wildlife Fund, 2017). Half of the Colombian territory is covered by forest, which amounts to around 59 million hectares (FAO, 2016). The country is divided into five natural regions, the region with the largest portion of forest being the Amazon (it owns two-thirds of the total forests in the country), followed by the Andean region, the Pacific region, Orinoquia and lastly, the Caribbean region (Ministerio de Ambiente y Desarrollo Sostenible, 2020). Fifty-nine natural protected areas belong to the National Parks System, which extend over an area of 142.682 km^2. Twenty-six of these areas fall within indigenous communities and afro descendent territories (SINAP, 2020).

(a) (b)

Figure 1. Pictures of (a) Colombian Amazon rainforest (Credits: kimkim.com) and (b) *Ramphastos brevis*, known as Choco Toucan, an endemic species (Credits: Ondrej Prosicky / shutterstock).

Nevertheless, deforestation is threatening many forested ecosystems worldwide, accelerating climate change, and Colombia makes no exception. In turn, the negative consequences of extreme climate events exacerbate deforestation, hence, the cycles of deforestation and climate change are feeding each other in circle. Colombia is one of the three most vulnerable countries to extreme weather events in South America based on the Climate Risk Index (Eckstein et al., 2019). Besides, the country has registered an enormous biodiversity loss. There are 10,349 species under threat of extinction in

Colombia, and 60% of them live in forests (International Union for Conservation of Nature, 2020). The Colombian forest harbors numerous plants and animals, but also vulnerable communities like indigenous tribes and afro-descendent settlements.

The determinants of deforestation are numerous and some of them are well established and studied. Little is known, however, about the consequences of illegal economies on the environment. Although, it has been estimated that 40% of the deforestation worldwide is illegal, a figure that reaches 80% in Indonesia and Brazil (Tellman et al., 2020). Disclosing this relationship is therefore essential to ensure the compliance of the *17 Sustainable Development Goals* of the decade to reduce the overwhelming amount of human footprint and avoid imminent and irreversible damages to our planet (Griggs et al., 2013; Robert et al., 2005). As illegal markets hide from the public record, data is scarce and mostly unreliable. However, new technologies and statistical methodologies are being developed to provide indirect measurements to shed light on the shadow economy. According to the North American think tank *Global Financial Integrity*, the sum of the value gaps identified in trade[1] between 171 countries in 2017 could amount to some US$817.6 billions (Kar and Spanjers, 2017). This quantity is slightly greater than the Gross Domestic Product of Saudi Arabia in 2019 (World Bank, 2020).

Within illegal markets, drugs market represents the second most lucrative business after counterfeiting, with a revenue of US$426 to US$658 billions in 2017[2] (Kar and Spanjers, 2017). The drug market that generates the second highest revenues is cocaine, with an annual income of some US$94 to US$143 billions. Its price depends on the elasticity of the demand and the law enforcement effectiveness (repression), faced during trafficking until its point of sell. Countries that inflict more repression against drug-related crimes, usually have higher prices of drugs (Becker et al., 2006). This mechanism inflates prices and provides high profits for drug cartels.

Colombia is the world's biggest producer of cocaine, ahead of Peru and Bolivia. In addition, drugs market relates to many serious crimes and other illegal markets. Many criminal acts are sponsored or derived from drug trafficking to increase profit or to stabilize the market. In Colombia, the drugs market has been linked to armed rebellion (*las FARC* - Fuerzas Armadas Revolucionarias de Colombia) either by tax or collaboration (Holmes et al., 2010), to kidnapping, illegal logging, land grabbing (Van Dexter and Visseren-Hamakers, 2019) and environmental crimes (Nellemann et al.,

[1] The identification of value gaps in trade is a methodology used by this institution to measure the size of the illegal economy.

[2] This value does not consider new psychoactive substances (NPS) and is therefore a conservative guess of the size of the drugs market according to Global Financial Integrity.

2014). In particular, the relationship between forest, coca cultivation and *las FARC* is multifaceted and complex. Forests provide a myriad of natural assets that have an enormous economic trade capacity, especially in extensively forested countries of the developing world, where forests are depleted to obtain raw materials. Deforestation and forest tradable goods are the source of subsistence of many *guerrillas*, not only in Colombia but in developing countries in general. Cases in Africa and Latin America are abundant. In these regions of the world, it is possible to observe a pattern of forest abundance and armed conflict (Price et al., 2003; Sánchez-Cuervo and Aide, 2013; Castro-Nunez et al., 2017; McNeely, 2003). Forests provide shelter for mobilized conflicts, hidden in the innermost locations (Fearon and Laitin, 2003). For this reason, guerrillas flourish, survive, and even thrive in forests, and especially in tropical forests (Price et al., 2003). The Colombian guerrilla lasted for 54 years until a peace treaty was signed in 2016 after years of arduous negotiations between the Colombian government and *las FARC*. The word "forest" appears three times in the Peace Agreement of *La Habana*, mostly related to conservation and protection of the woodlands (Acuerdo de Paz, 2016). The former inhabitants of the unreachable lands of *las FARC*, were holding back large forest loss by conceding limited permits for deforestation to some illegal farm-holders like *coqueros*[3] and even legal or illegal miners, farmers or peasants (Reardon, 2018; Brodzinsky, 2017). *Las FARC* was deterring bigger private investors from colonizing and destroying wide extensions of forests, as investors feared the potential extortion that violent groups could exert on their businesses (Arias et al., 2014). Nevertheless, the role of *las FARC* in deforestation is a conflicting and opposite force: on one hand, *las FARC* prevented private investors to access their conquered territories, and on the other, boosted illegal crops and expanded the illegal crops frontier into the forest. A complex social scenario has been unfolding in the forests of Colombia, where the links between armed conflict, crime, drugs, and gold mining led to one of the most enduring resource-fueled conflicts in the world. In the process, the actors have transformed and influenced social, political, and economic institutions to pursuit wartime illicit markets (Rettberg and Ortiz-Riomalo, 2016).

While conflict invariably caused negative impacts on biodiversity, peace was even worse, as it enabled forest exploitation to operate with impunity (McNeely, 2003). *Las FARC*'s role in deforestation containment became apparent after the Peace Agreement of 2016. The years following the reach of the Agreement a steep rise in deforestation took place. In 2018, narco-organizations or Drug Trafficking Organizations (DTO's)

[3]Coca growers.

took over *las FARC* former territories. A part of the land was used for growing coca, which explains the simultaneous increase in deforestation and coca crops. As said, *las FARC* acted both, as a conservationist but also as a driving force for forest clearance. It was a competing force that toggled in opposite directions. When *las FARC* no longer played a role, small conservationist groups and a weak State could not fill in the gap and were unable to impart order. Thus, deforestation jumped to historic heights and drug cartels took advantage of the order void by literally grabbing the land that the *guerrilla* left behind. DTO's are not only disturbing forest patches by directly planting coca or poppy, but also by financing other illegal activities that disrupt forests, like cattle ranching (Devine et al., 2020; Van Dexter and Visseren-Hamakers, 2019), and agribusiness in general (Van Dexter and Visseren-Hamakers, 2019; Tellman et al., 2020). This "legal" land-related businesses are used as a facade to off-load currency (Devine et al., 2020; Salama, 2000). Rural areas provide ideal scenarios for drug-traffickers to establish businesses in land, labor, livestock, processing facilities, but also to hide, traffic, grow plant-based drugs and ultimately gain from the low cost of human labor. DTO's accomplish very high profits from a low-cost product, enhanced by illegality.

Despite the Peace Agreement, armed violence persists across the country. As a result, mass displacement almost tripled this year compared to 2020. Experts do not know if this consensual peace will last or how long it will take to stabilize and solve (at least partially) the social resentments that paved the way toward civil war, but the remnants of *las FARC* are sparse in the forests and are planning their comeback, even though they will not reach their former organization and power (The Guardian, 2019). The government has labeled these groups as *Grupos Armados Organizados Residuales*[4] (GAOR). While most of the former guerrilleros reintegrated into society and *las FARC* became a legal and recognized political force *Partido FARC*[5] (FARC party), the dissidents of *las FARC-EP* are growing, they doubled in number in only one year, from around 2300 combatants in 2019 to 4600 in 2020 according to El Tiempo (2020), the newspaper with the largest circulation in Colombia.

Obviously, the drivers of deforestation in Colombia are not only linked to illegal markets but also relate to legal or natural causes, such as agriculture, cattle-grazing, legal mining, population growth and clearance for government planned settlement schemes (Armenteras et al., 2013, 2011; Brodzinsky, 2017). The most relevant drivers of deforestation are resumed in Figure 2, where we establish relationships and the flow of external causes, forces, and actors that impinge forest clearance in Colombia, with an

[4]Organized residual armed groups.
[5]Their official website is: https://partidofarc.com.co/farc/

emphasis on illegal markets.

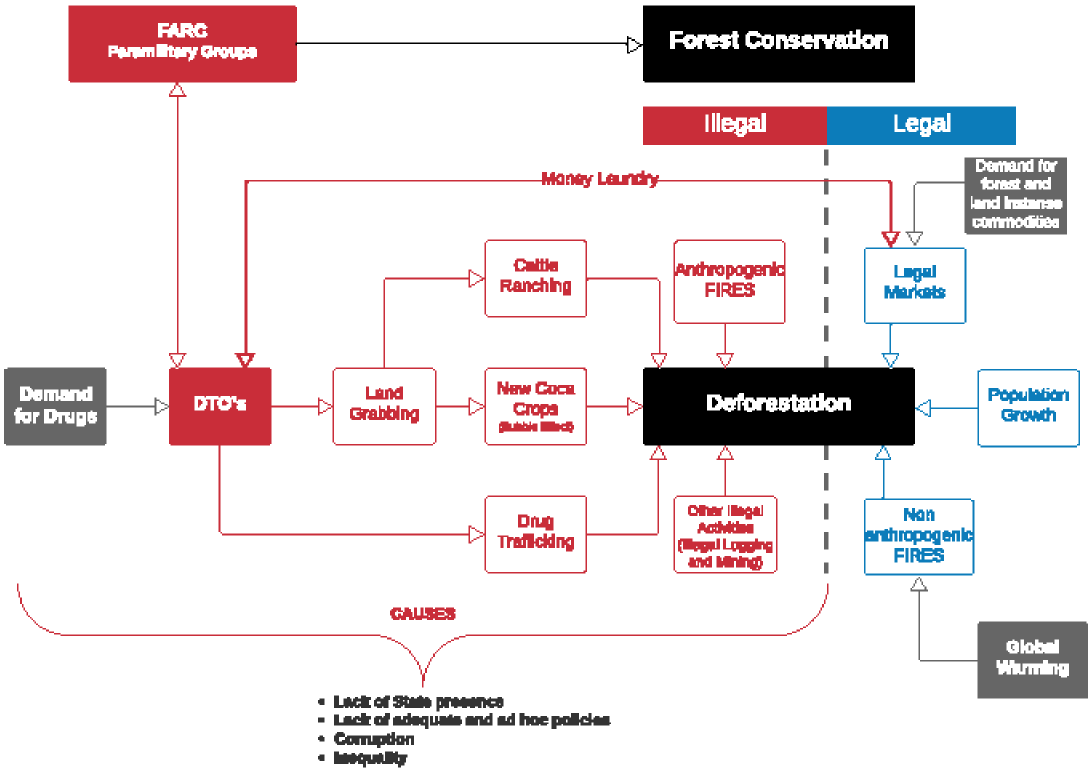


Figure 2. Diagram of deforestation drivers in Colombia. Legal markets that cause deforestation include mainly agro-pastoral activities, legal logging, road construction and legal mining.

On the left side of the diagram in Figure 2, illegal drivers are shown. As discussed, the peace process has worsened the conservation of the forests, and simultaneously DTO's have become more powerful by expanding their production, trafficking, and income (The Economist, 2019) with severe implications to forest dynamics and conservation. DTO's are directly and indirectly contributing to a large-scale land-change dynamics where "legal" entities make clandestine transactions (Tellman et al., 2020) for the unconscious and disruptive use of soil that ultimately damages landscapes permanently. Illegal logging is also a major cause for deforestation in Colombia, as nearly 50% of the wood in the market is illegal (Environmental Investigation Agency, 2019). A further deforestation driver represented in Figure 2 that contributes greatly to forest loss consists of anthropogenic and non-anthropogenic related fires. Non-anthropogenic fires, on the right side of the diagram, are fundamentally fires that are initiated by "natural causes". Warming and drying have significantly increased fire-season fuel aridity, fostering a more favorable fire environment across forested systems. Studies in the west of the United States have shown that human-caused global warming has doubled fire activity since 1970 (Abatzoglou and Williams, 2016). The multiple causes of deforestation, actors and long-lasting dynamics make deforestation a complex problem. As a consequence, the solution must be well-engineered, comprehensive, and innovative. The solution is not a single policy nor a single action, it will take a web of solutions, interconnected and locally conceived, that aspire to contemplate everyone involved and address the problem from its root.

In this thesis, we propose an econometric analysis of Colombia's narcodeforestation and the effectiveness of glyphosate aspersion policy to reduce coca crops in the country. The thesis consists of three chapters, corresponding to related but free-standing essays. Chapter 1 investigates the role of coca cultivation as a determinant of deforestation in Colombia at both national and sub-national levels, using cross-sectional data at the municipal level. For the sub-national level, a cluster analysis is employed to preliminary identify four homogeneous regions of coca cultivation and deforestation levels. Then, spatial models are used to test whether coca cultivation has a significant impact on deforestation at the national level and in each of the sub-regions. To control for factors affecting deforestation other than coca cultivation, we consider a set of forty-four control variables, including biophysical, anthropogenic, and socioeconomic variables. To avoid multi-collinearity, a spectral decomposition based on principal component extraction is implemented to reduce the dimensionality of the predictors' matrix, while retaining as much information as possible from the data. In this first chapter, deforestation is computed from the Global Forest Change Dataset, as it provides more updated data than other sources. However, this dataset has sometimes been criticized for over quantifying forest cover in different parts of the world, a problem that can be alleviated through a careful calibration of the forest cover percentage used to produce forest maps. The availability of reliable forest cover estimates is obviously of crucial relevance for researchers monitoring forest stocks. This necessity motivates the second chapter of this thesis.

In Chapter 2, we present an accuracy assessment of the three main datasets to predict forest cover in Colombia, namely, the one released by the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), which is the most widely used dataset for studies about deforestation in Colombia, the Global Forest Change Dataset (GFCD) from the University of Maryland, and the European Space Agency (ESA) dataset. The aim is (i) to investigate the accuracy of forest cover maps obtained from GFCD using different tree cover thresholds across different ecological zones in Colombia; (ii) to compare the accuracy of GFCD, after optimization of the threshold, with that of ESA and IDEAM; (iii) to provide users with guidelines for choosing the most reliable data set for forest mapping, both, for particular ecological zones and Colombia as a whole. For this purpose, we randomly draw a large number of pixels from each ecological zone and collect reference data for each of them relying on Collect Earth, which is a program that is used for field-based inventories, in particular for cross-checking land classifications. Then, we derive a confusion matrix between the forest cover map obtained from each database (and from each threshold for GFCD) and the reference data, and we use this matrix to compute various measures of accuracy.

Finally, in Chapter 3, we study the causal effect on coca cultivation of a specific antidrug policy: glyphosate aspersion of coca crops. This is a controversial antidrug policy for many reasons: not only is it risky for human health and the environment, it also accidentally hits livestock and legal plantations, targeting poor households, feeding social insecurity and discord, and determining displacement of entire communities. On top of this, its efficacy in reducing coca cultivation is doubtful and studies on this issue often reached contrasting results. Aerial aspersion was suspended in 2015, but despite its flaws, this policy is being reconsidered as a forced eradication measure. The aim of this chapter is, therefore, to contribute to the current literature on evaluating the real impact of glyphosate aspersion on coca crops. The main challenge is to address endogeneity issues caused by the fact that spraying is not randomly assigned but is higher in areas with more coca crops, thus generating simultaneity bias in OLS estimates. We use panel regression analysis on municipal data spanning the period 2000-2015, and we exploit exogenous source of variation in aerial aspersion through an instrumental variable approach. We instrument aerial aspersion with the number of days in the year in which wind was below a certain threshold. In fact, spraying is normally carried out when the surface wind speed is less than 6-7m/s, so that the relevance condition for the instrument is met. To satisfy also the exclusion restriction, we further control for climatic variables, which can be correlated with the wind and have a direct effect on coca cultivation, such as precipitations.

Chapter 1

A Spatial Regression Analysis of Colombia's Narcodeforestation with Factor Decomposition of Multiple Predictors

Abstract

In the current accelerated process of global warming, forest conservation is becoming more difficult to address in developing countries, where woodlands are often fueling the illegal economy. In Colombia, the issue of narcodeforestation is of great concern, because of the ramification of narco-activities that are affecting forests, such as agribusinesses and cattle ranching for money laundering. In this study, we use spatially explicit regressions incorporating a factor decomposition of predictors through principal component analysis to understand the impact of coca plantations on local-scale deforestation in Colombia. At a national level there is a positive and statistically significant relationship between coca crops and deforestation. We found out that in 2 out of 4 regions, coca is causing deforestation in the country, especially in the Department of Northern Santander and on the Pacific coast.

1 Introduction

Containing deforestation is a critical environmental issue to be tackled in the short term, especially in the tropics. Forest loss is a complex phenomenon with multiple sources, which range from social, economic, and ecological nature. A detailed analysis of the spatial and temporal behavior of deforestation drivers is required to grasp the true causes and potential consequences of forest loss in a given region. This is notably important in Colombia, a country with huge geographical, ecological, and cultural diversity, preserving a considerable forest stock for Latin America. More than 52%

of Colombia land is covered by forests, and 14% is classified as primary forest, the most biodiverse and carbon-dense form of forest. Colombia is also classified as the second most biodiverse country in the world holding 51,330 registered species until today (Ministerio de Ambiente y Desarrollo Sostenible, 2021). However, deforestation is one of the main challenges to overcome in the short term if the country wishes to achieve its aim to reduce greenhouse gases by 51% and black carbon emissions by 40% in 2030 compared to their 2014 levels (Climate Action Tracker, 2021). Today, deforestation accounts for around a quarter of Colombia's total emissions (Climate Home News, 2021). Colombia is also the 5$^{\text{th}}$ country in Latin America by deforestation rate, well ahead of Brazil or Mexico (Armenteras et al., 2017). In Colombia half of the ecosystems are in a critical state of deterioration or a state of danger, moreover, in these endangered ecosystems more than a third of Colombia's plants and 50% of its animals are at risk of extinction (WWF, 2017). Similarly, indigenous and Afrocolombian descendants, whose cultures and lifestyles are closely tied to these forested ecosystems, are also at risk of disappearing. Furthermore, according to Masiokas et al. (2020), the cryosphere of the Andes in Colombia has narrowed down by 62% in the last 50 years.

Earlier studies identify and discuss the main drivers of tropical deforestation in the world. Among these factors, there are demographic drivers like population growth that push the colonization fronts, economic elements like production boosts or extraction of raw material, infrastructure, and road development (Armenteras et al., 2013). Moreover, environmental factors, including geology, topography, and soil type, as well as climatic factors, like drought and rainfall, are important natural determinants of deforestation (Bax and Francesconi, 2018). In this sense, the relationship between the dry season and wildfires is especially conspicuous in the Amazon forest where scientists have observed that with the current climate change the length of the dry season has increased (Armenteras et al., 2013; Shukla et al., 1990; Fu et al., 2013), which has, in turn, led to an increase in wildfires. In the specific case of Colombia, the main drivers of deforestation are related to agriculture, fires, mining, cattle-grazing and population growth linked to clearance for government planned settlement schemes (Armenteras et al., 2011; Brodzinsky, 2017).

Several authors also consider the impact of coca cultivation on deforestation. As preliminary context, it is worthwhile noting that Colombia is the top producer of coca leaves in the world, the main input in the production of cocaine hydrochloride, an addictive psychostimulant drug whose use is illegal in most countries (UNODC, 2021; American Addiction Centers, 2021). Colombia produces around 65% of the coca in

Latin America, and with Peru and Bolivia, the three countries produce 98% of the coca worldwide (UNODC, 2021; Moreno-Sanchez et al., 2003). In 2020 the production of coca leaves reached 143,000 hectares only in Colombia (UNODC, 2021). The impact of coca crops on deforestation is complex because of the different actors involved, but also because of the indirect effects of the cultivation of coca. It has been observed in Central America, especially in Guatemala, Honduras, Nicaragua, and Costa Rica, that narcotrafficking itself is a cause of deforestation (Devine et al., 2021; McSweeney et al., 2014; Sesnie et al., 2017). However, this relationship has not been studied in Colombia, because coca plantations make it difficult to single out the effect of trafficking from crops. Central America is a corridor to get to the consumers, while Colombia is the factory. Therefore, the amount of forest loss due to the trafficking of illegal drugs in Colombia is unknown. There are also known indirect causes of forest clearance tied to cocaine trafficking, like cattle ranching or agribusiness for money laundry (Van Dexter and Visseren-Hamakers, 2019), and aerial spraying with glyphosate. This last measure intended for eradication was suspended in 2015 but has found support in the political sphere of Iván Duque in recent years and might resume if the bill gets approved (El Tiempo, 2021,).

Additionally, the long-lasting insurgency of *las FARC* in Colombia has linked forest dynamics to the guerrilla history process. Peace and war between the State and the armed groups have determined the level of forest clearance in Colombia. Since the peace treaty between *las FARC* and the State was signed in 2016, after 52 years of armed conflict, a drastic increase in forest loss occurred due to legal and illegal land grabbing of former *las FARC* occupied land (Van Dexter and Visseren-Hamakers, 2019), converted into large scale cattle-ranching, coca cultivation or other illegal land markets. After the peace treaty, new fragmented organizations from *las FARC* began to take shape during Duque's presidency, and the COVID pandemic struck the cocaine market (Sanín, 2020), due to radical restrictions in national and international transportation. The price of cocaine has dropped due to hard lock-downs that have paralyzed transport flows worldwide (The Washington Post, 2020; Sanín, 2020).

The present work aims at contributing to the literature about the impact of coca cultivation on deforestation rates. In particular, most of the existing literature considers the impact at the national level and does not investigate differences among regions. Other studies, however, focus on specific areas of Colombia (Anaya et al., 2020; Hoffmann et al., 2018). The only exception is provided by Armenteras et al. (2013), who recognize the importance of accounting for Colombia's biophysical, socioeconomic, and demographic heterogeneity, and explore the sub-national variation in deforestation dynamics and its drivers, within the country's four main natural lowland regions. Armenteras et al. (2013), however, do not focus on the particular role played by coca cultivation, but simply consider illicit crops as one of the many factors affecting deforestation. In addition, the spatial structure of the data is not considered. In this work, we thus investigate the role of coca cultivation as a possible determinant of deforestation in Colombia at both national and sub-national levels, using cross-sectional data at the municipal level. For the sub-national level, a cluster analysis is employed to preliminary identify four homogeneous regions of coca cultivation and deforestation levels. Then, different types of spatial models are fitted to the data and used to test whether coca cultivation has a significant impact on deforestation at the national level and in each of the sub-regions. To control for factors affecting deforestation other than coca cultivation, we consider a set of 44 control variables, including biophysical, anthropogenic, and socioeconomic variables. To avoid multi-collinearity, a spectral decomposition based on principal component extraction is implemented to reduce the dimensionality of the predictors' matrix, while retaining as much information as possible from the data.

The paper is organized as follows: Section 2 describes the data set, it enlists all the variables adopted in this study, the descriptive statistics, and some maps that help to geographically locate coca crops and deforested regions. Section 3 presents the methodology used. Section 4 illustrates data analysis and main results. Finally, Section 5 discusses the results, while Section 6 concludes by providing recommendations for future research.

2 Data

2.1 Variables

We use cross-sectional data at municipal level. The municipalities are obtained from the GADM database (version 3.6). The original 1,065 municipalities are then aggregated according to the supplementary material of Mendoza (2020), ultimately reducing the number of municipalities to 1,060. This allows to simplify the indexing and enhance the process of data matching and data fusion (Christen, 2012), as the 1,060 municipalities sub-division is the most widely used by national institutions.

We consider a set of 48 different variables, including deforestation and coca crops, which are the pivotal variables for our purpose. We include biophysical attributes, like slope and elevation, anthropogenic variables such as population density, wildfires, and remoteness (mean municipal pixels distance to the nearest road), and socioeconomic variables such as wealth distribution, added value created at the municipal scale, or the ratio of value added created by primary and secondary economic activities. All of the above mentioned variables have been shown to be correlated with forest loss either in Colombia or elsewhere. The database contains cross-sectional information of each variable mean for the case of variables that change over time, like deforestation, coca crops or precipitation, but we also have static variables like elevation. Moreover, variables that do not vary greatly over time are also referred to a single year, an example is the age structure of the population. Most of the variables are obtained from government sources, NGOs, and agencies. In Table 16, we display the reference year and the source of data, for each of the variables included in the analysis.

Table 1. Variables description, source of data and reference period

Identifier	Description	Source	Year(s)
PDET	Programa de Desarrollo con Enfoque Territorial	ArcGIS data base	2019
FARC07_16	FARC presence during 2007-2016	Misión de Observación Electoral	2007-2016
Wconflict	Victims of armed conflict	Centro Nacional de Memoria Histórica	2010-2016
mining	Mining area in percentage	Tierra Minada	2018
national_parks	National park area in percentage	Environmental Research Systems Institute (ESRI) Colombia	2020
other_parks	Regional and other typologies of parks	Environmental Research Systems Institute (ESRI) Colombia	2021
afrocolombian _protected_area	Percentage of afro-colombian territories on total area	Instituto Geográfico Agustín Codazzi	2018
elevation	Elevation mean	Zenodo	2017
elevstdev	Standard deviation of the elevation	Zenodo	2017
night_lights	Sum of night light events	NASA's Earth Observatory	2016

Table 16 continued: Variables description, source of data and reference period

indigenous _protected_area	Percentage of indigenous reserves area on total area	Instituto Geográfico Agustín Codazzi	2018
wealth	Relative Wealth Index predicts the relative standard of living (Chi et al., 2022)	United Nations Office for the Coordination of Humanitarian Affairs (OCHA)	2021
added_ value_economy	Added value of the municipality	Departamento Administrativo Nacional de Estadística	2018
primary_vs_ secondary_activities	Primary and secondary activities as a ratio of tertiary activities	Departamento Administrativo Nacional de Estadística	2018
defense	Military and police stations per 1000 inhabitants	Datos Abiertos Gobierno de Colombia	2016
agriculture	Hectares for agricultural use as percentage of the total area	Environmental Research Systems Institute (ESRI) Colombia	2013- 2014
years_coca	Years of coca crops (permanence of coca crops)	Observatorio de Drogas de Colombia	2000-2020
financial_ institutions	Financial institutions per 1000 inhabitants	United Nations Office for the Coordination of Humanitarian Affairs (OCHA)	2020
popDensity	Population density	NASA's Earthdata	2020

Table 16 continued: Variables description, source of data and reference period

voting_turnout	Participation in the elections of 2018 as percentage of the total population	Departamento Administrativo Nacional de Estadística	2018
duque	Percentage of total voters of Iván Duque (right-wing candidate)	Departamento Administrativo Nacional de Estadística	2018
mean_youth _age_of_man	Man between the age of 15 to 24 years old	Departamento Administrativo Nacional de Estadística	2017
early_adults_age	Man between the age of 25 to 40 years old	Departamento Administrativo Nacional de Estadística	2017
women2017	Female population percentage	Departamento Administrativo Nacional de Estadística	2017
TFR	Total fertility rate	Departamento Administrativo Nacional de Estadística	2017
internal_ displacements	Internally displaced persons	United Nations Office for the Coordination of Humanitarian Affairs (OCHA)	2000-2014
private_property	Private properties for agricultural use	Environmental Research Systems Institute (ESRI) Colombia	2013- 2014

Table 16 continued: Variables description, source of data and reference period

rented_property	Rented properties for agricultural use	Environmental Research Systems Institute (ESRI) Colombia	2013- 2014
usufruct_property	Usufructed properties for agricultural use	Environmental Research Systems Institute (ESRI) Colombia	2013- 2014
collective_property	Collective property for agricultural use	Environmental Research Systems Institute (ESRI) Colombia	2013- 2014
gHM	Global Human Modification Index (cumulative measure) (Kennedy et al., 2019)	Google Earth Engine Repository	2016
remoteness	Mean municipal pixels distance to nearest road	United Nations Office for the Coordination of Humanitarian Affairs (OCHA)	2020
water_capacity	Water holding capacity of soil, mean value	Zenodo	1990-2020
WCstdev	Water holding capacity of soil, standard deviation	Zenodo	1990-2020
health_ infrastructure	Hospital infrastructures per 1000 inhabitants	Departamento Administrativo Nacional de Estadística	2021
poverty	Global Multidimensional Poverty Index (MPI)	United Nations Office for the Coordination of Humanitarian Affairs (OCHA)	2021

Table 16 continued: Variables description, source of data and reference period

rainfall	30-year precipitation mean	Zenodo	1990-2020
cattle	Hectares intended for breeding of adult livestock	Departamento Administrativo Nacional de Estadística	2018
wildfires	30- year mean wildfires as percentage of total area	NASA	1990-2020
apersion_ glyphosate	Glyphosate aspersion as percentage of total area	Observatorio de Drogas de Colombia	2000-2015
temperature	30-year temperature mean	Zenodo	1990-2020
temperature_std	30-year temperature standard deviation	Zenodo	1990-2020
human_ settlements	Settlements as percentage of total area	European Space Agency (ESA)	2001-2020
palm_oil	Palm oil as percentage of total area	Google Earth Engine Repository	2019
cropland	Cropland as percentage of total area	European Space Agency (ESA)	2001-2020
other_ natural_land	Natural -non-forest- land percentage	European Space Agency (ESA)	2001-2020
coca_crops	20-year coca crop mean as percentage of total area	Observatorio de Drogas de Colombia	2000-2020

Table 16 continued: Variables description, source of data and reference period

deforestation	GFCD's 20-year mean deforestation as percentage of total area	University of Maryland	2000-2020

2.1.1 Deforestation

Deforestation and coca crops are the focus of our study, therefore, we characterize them in more detail in Table 2. Deforested area is calculated from the rasters of the University of Maryland, namely the Global Forest Change Dataset (GFCD). GFCD has commonly been criticized for over quantifying forest cover in different parts of the world, but it has also been demonstrated that the calibration through the forest cover percentage depends on the ecological zone treated. In the calculation of the forest cover in the GFCD data set, the canopy cover percentage is manually assigned by the researcher. The canopy cover threshold of 30% is widely used, but as investigated by Lwin et al. (2019), McRoberts et al. (2016) and Sannier et al. (2016), the optimum threshold for the GFCD data is higher for denser forest canopy regions. Lwin et al. (2019) determined that the accuracy of the forest cover in GFCD was dependent on selection of the appropriate forest cover threshold. The authors analyzed the case of Malaysia, where the main forest cover type is tropical rain forest, and concluded that the threshold should be set according to the ratio of tropical rainforest and total area. However, for different ecological zones, different thresholds are needed to reach higher accuracy. In summary, where tropical rainforest is dominant, selecting a higher threshold is advisable. The optimal threshold at a national level was around 40% for Malaysia (Lwin et al., 2019), 70% for Gabon (Sannier et al., 2016) and 95% for Santa Catalina in Brazil (McRoberts et al., 2016). As a result of the previous analyses and in consideration of the abundance of tropical forest in Colombia, we set the threshold at 70% in the present study. The decision to use this threshold is also guided by the comparison with the data provided by different sources, namely the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) and European Space Agency (ESA). However, further studies should be conducted to estimate the most accurate forest cover threshold of the GFCD data set for Colombia.

In the literature of Colombia's narcodeforestation, many scholars use the IDEAM forest data set and a smaller portion uses the GFCD data set, whereas no one, up to our knowledge, has used ESA so far. GFCD data set goal is to quantify forest cover change and it takes the year 2000 as a baseline to confront the loss per year, while gain of forest cover was only monitored until 2012. GFCD inspects global Landsat data at a 30-meter spatial resolution to identify forest extent worldwide (Hansen et al., 2013). The IDEAM data set is collected and prepared by a Colombian institution that aims at quantifying forest cover, and has the advantage that the information obtained from the satellite and classification algorithms is debugged and analyzed by geographers who know the area. In this data set, the forest cover is pre-established, without the possibility to modify it, at 30% (IDEAM, 2019). Also, the canopy height should be over 5 meters, which is a standard criterion compared to the other data sets. It is also a very easy-to-use file, it has only 3-pixel values, one for forest, another for non-forest, and one for no information, a relatively scarce value in this data set. The ESA data set is quite different from the previous two, it does not allow to establish the forest cover like in GFCD, but has more options of canopy cover thresholds than IDEAM, it all depends on the type of land one wishes to include and the canopy associated to that land classification. In essence, the goal of the ESA data set is to classify the type of land into some 17 categories, which is quite different from IDEAM or GFCD, whose only goal is to identify forest and forest cover changes. These substantial differences in the goal and classification method between ESA and IDEAM/GFCD lead to differences in the results. Furthermore, GFCD uses high-resolution satellite imagery from Landsat and this applies also to IDEAM after 2010, while ESA uses Sentinel as its imagery source. According to Astola et al. (2019) Sentinel-2 is slightly better than Lansat 8 in forest variable prediction. Figure 3 provides a comparison between ESA, GFCD and IDEAM forest cover for the year 2017.

The main reason why we do not use IDEAM in our study is that it does not provide annual information within the time interval for which data on coca crops are available (from 2000 to 2020). IDEAM released intermittent rasters before 2012 and yearly information since 2013. Moreover, before 2007 the spatial resolution of the imagery obtained through MODIS was lower than that obtained from the Landsat satellite, which was used by IDEAM after 2010 (IDEAM, 2019).

Figure 3. Comparison between ESA, GFCD and IDEAM forest cover maps for the year 2017.

Table 2 provides some descriptive statistics on the average annual deforested area (as percentage of the municipal area) in Colombia, during the period 2000-2020, on the basis of GFCD data. A graphical representation is, instead, given in the left panel of Figure 4. The north-west of Colombia, which includes a considerable portion of the departments of Northern Santander, Santander, and Antioquia, shows a high mean deforestation area in the last 20 years. On the other hand, Caquetá, Guaviare, Meta, and Putumayo display large patch disturbances. These forested areas, highly valuable from a biodiversity perspective, are the main target of deforestation during the studied period.

Table 2. Descriptive statistics on average annual deforestation area and coca crop extension, as percentage of municipality area, between 2000 and 2020

Statistic	Deforestation area (in %)	Coca crops (in %)
Null values	4	755
NA values	0	0
Min	0	0
Max	1.23	5.19
Median	0.06	0
Mean	0.12	0.06
Variance	0.03	0.07
Std.dev	0.16	0.27
Variation coef.	1.34	4.69
Skewness	2.89	10.03
Kurtosis	10.87	144.2

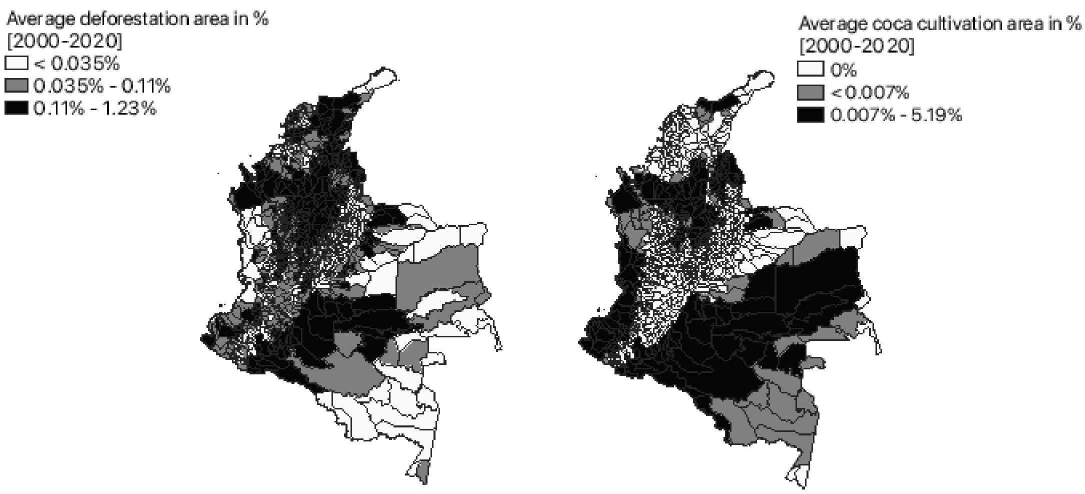


Figure 4. Average annual deforestation (left panel) and coca cultivation (right panel) area, as percentage of municipal area, between 2000 and 2020.

2.1.2 Coca crops

In this study, we use the data on coca crop extension published by the ODC (Observatorio de Drogas de Colombia), which in turn is supported by the SIMCI project (Integrated System for the Monitoring of Illicit Crops) of the United Nations. These data are produced through a multi-step process. Because of the illicit nature of the crops, various sources of information and validations are required. First, medium resolution satellite imagery is obtained from three different satellites: Landsat, Sentinel, and Worldview II, depending on which satellite image encounters fewer clouds and offers, in general, a better view of the area. The images cover the whole extension of the Colombian mainland, except for the isles of San Andrés and Providencia (excluded from our analysis). Next, the images are pre-processed to enhance object detection. Later, the images are visually interpreted and, in this step, historical data and information from government sources that account for potential or verified coca crops are cross-checked with the images. Afterward, overflights on the coca crops detected through imagery take place for validation. Finally, the information is edited and published (Biesimci, 2016).

We extract information on coca crops from 2000 until 2020, calculate the annual average, divide it by the area of the municipality and multiply it by 100. In this way, we obtain the annual average percentage of land covered by coca crops in each municipality. The result is shown in Figure 4, right panel, while some descriptive statistics are provided in Table 2.

From Figure 4, we can identify 3 macro-regions of coca crops during the studied period. One is located in the north of the country, encompassing Chocó, Antioquia, Santander and Northern Santander. The second one is on the west coast, which comprises the unique biodiversity hotspot of Tumbes-Chocó-Magdalena, where concerns about the impact of deforestation are particularly relevant (Sanchez-Cuervo and Aide, 2013). The third region, which extends over a very large area, wraps the strip that touches both the Andean foothills and the Amazon rain forest and includes the departments of Putumayo, Amazonia, Guaviare, Meta, and Vichada.

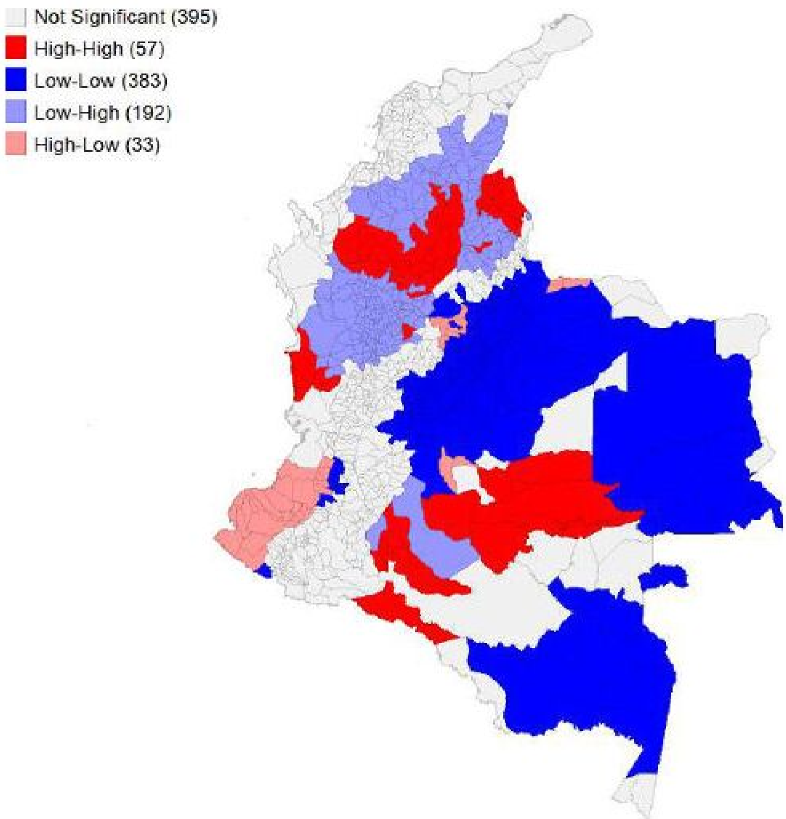


Figure 5. Bivariate Local Indicator of Spatial Association Map (Bilisa).

To conclude the descriptive analysis of the main variables of interest, Figure 5 shows a local indicator of spatial association (LISA) between deforestation and coca crops. In practice, it is a graphical representation of the local Moran I, with "High-High" (or "Low-Low") values representing high (or low) deforested municipalities surrounded by high (or low) coca cultivated neighbors and "High-Low" (or "Low-High") values representing high (or low) deforested municipalities surrounded by low (or high) coca cultivated neighbors. Out of the 1,060 municipalities, 57 fall into the High-High group and 383 in the Low-Low group, which means that these data points have a positive spatial autocorrelation. A negative autocorrelation is instead found for 225 municipalities falling in the Low-High or High-Low groups. The LISA map suggests interesting locations of analysis and eventually clusters of both Low-Low and High-High alike values. A more wholesome analysis is offered in a multivariate context, and this is why spatial regression analysis is essential to gain more conclusive results on the studied relationship.

3 Methodology

Spatial regression analysis is used to explore the relationship between coca cultivation and deforestation while allowing for the possible presence of spatial autocorrelation in the data. The fundamental interpretation of spatial autocorrelation is the existence of a spatial pattern. Positive spatial autocorrelation occurs when similar values are detected between the areal unit and its neighbors. Positive spatial autocorrelation does not mean that the observation and the neighborhood share positive values, instead, the observation and the neighbors show like-values, either positive or negative (Le Gallo and Ertur, 2003). Conversely, negative spatial autocorrelation occurs when neighboring values are mutually dissimilar (Anselin, 2013; Bivand and Wong, 2018). Spatial regression analysis is appropriate to investigate spatial patterns. Moreover, if spatial autocorrelation is present in the data and not specifically accounted for, the assumption of independent errors in OLS regression is violated, calling into question the validity of hypothesis testing procedures (Smith and Lee, 2012).

Spatial weights matrices are the fundamental elements in regression models where a representation of the spatial structure is needed. The spatial structure of the network contained in the weights matrix is determined and assumed *a priori*, which has often been criticized by economists, but was magnificently fenced by Corrado and Fingleton (2012), arguing, among other things, that weights matrices are “convenient, useful, and succinct representation[s] of spatial interaction[s], either in the form of endogenous or exogenous lagged variables, and/or as part of an explicit error process". We test different types of weights matrices and decide to use an inverse distance matrix, based on the Euclidean distance between centroids of the municipalities, with a cut-off of 200 kilometers. The coordinates of the centroids of the polygons in the system were originally in WGS84 (EPSG: 4326) coordinate system, later reprojected into WGS84 / Pseudo-Mercator (EPSG 3857). This reprojection allowed conversion from degrees to meters of the coordinates.

The weights matrix accompanies either the dependent variable, the independent variable, the error term, or a combination of the previous terms, and this determines the differences between the various spatial models. A regression with spatial weights captures the amount to which the changes in an observational unit impact its neighbors, the so-called spillover effect (Yustisia, 2017). Depending on the model, this effect can be measured at a global or local scale. We consider seven spatial regression models, hence, the Spatial Lag of X (SLX) model, the Simultaneous Autoregressive (SAR) model, the Spatial Error Model (SEM), the Spatial Autoregressive Combined (SAC) model, the

Spatial Durbin Model (SDM), the Spatial Durbin Error Model (SDEM), and the General Nested (GNS) model, which encompasses all previous models.

Using the notation proposed in Elhorst and Vega (2013), the GNS model has the following equation:

$$\begin{aligned} \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \alpha\mathbf{1}_N + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \mathbf{u}, \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \epsilon, \end{aligned} \tag{1.1}$$

where $\mathbf{W}$ is an $N \times N$ weights matrix that describes the spatial configuration of the units in the sample (Elhorst, 2014), with N being the number of observations. Thus, $\mathbf{Wy}$ captures the endogenous interaction effects among the dependent variable, $\mathbf{WX}$ the exogenous interaction effects between the independent variables, and $\mathbf{Wu}$ the interaction effects among the disturbance term of the different observational units. This model is the most general since it includes all types of interaction effects. The coefficient ρ is called the spatial autoregressive coefficient and reflects the spatial dependence inherent in our sample data, it measures the average influence on observations by their neighbors, while λ is a coefficient which captures the spatial correlation between errors, and θ, just as β, represents a $k \times 1$ vector of response parameters, k being the number of independent variables. The OLS regression, which is obtained from the GNS model when all the spatial coefficients are set to zero, is used as a diagnostic tool for model specification and as a fundamental benchmark for spatial models. We calculate the Moran I test on OLS residuals to test the null hypothesis of no spatial correlation, in which case the no rejection would imply the OLS model or non-spatial model to be adequate to explain the relationship between the variables (Golgher and Voss, 2016). On the other hand, upon rejection, we use Lagrange Multiplier (LM) tests for spatial dependence on OLS residuals to drive in the choice of an appropriate spatial model. The robust t-test or F-test on parameters ρ, γ, and θ in the various spatial models are also considered for this purpose. A graphical representation of the restrictions applied to the GNS model to obtain all the other spatial linear models is represented in Figure 6.

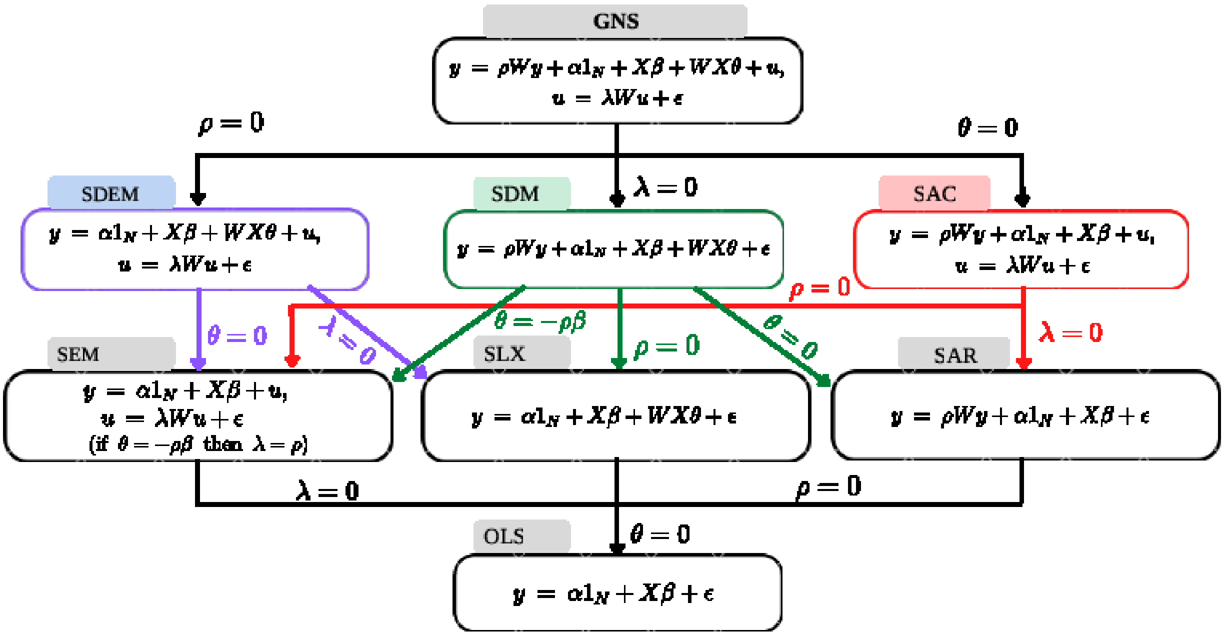


Figure 6. Comparison of different spatial econometric model specifications. *Modified from*: Elhorst and Vega (2013). OLS is the ordinary least squares model, SEM is the spatial error model, SLX is the spatial lag of X model, SAR is the spatial autoregressive model, SDEM is the Durbin error model, SDM is the spatial Durbin model, SAC is the spatial autoregressive combined model and GNS is the general nesting model.

We use spatial regression both on a national and subnational scale. Colombia has, in fact, a rich diversity in biophysical, cultural, social, and economic terms, thereby, dividing the country into macro-regions is essential to achieve a better understanding of the effect of coca cultivation on deforestation across the country. To define appropriate macro-regions, we adopt a two-step process. The first step is a K-means clusterization of the 1,060 municipalities, using coca crops and deforestation as input variables. Then, in the second step, we slightly readapt the results to obtain visually well-defined regions.

To control for factors affecting deforestation, other than coca cultivation, we consider many different variables of demographic, socioeconomic, anthropic, and biophysical nature, as described in Section 2. We adopt Principal Component Analysis (PCA) to reduce the dimensionality of the predictors' matrix and, thus, avoid multi-collinearity issues in the regression, while retaining as much information as possible from the data. In general terms, a PCA projects the original data into new variables, called Principal Components (PCs), that are orthogonal to each other and are ordered so that the first few retain most of the variation from the original variables. Thus, only the first few PCs are incorporated in the regression as control variables, avoiding multicollinearity issues. In this way, PCA helps delving into the data and identifying clusters and latent similarities between observations. We run one country-level PCA and use the results in the regional regressions, because we privilege the space in the regression models to be explicit. The PCA gives us a comprehensive decomposition of the underlying factors that act at a national level, simultaneously reducing the clear redundancy of the data set as well as allowing us to aggregate the variables in fewer meaningful factors. In other words, PCA analysis represents the first filter at a national scale, while the regression represents the regional level, which becomes spatially explicit when working with spatial models at a municipality level (i.e. elementary analysis' unit). All components from the PCA having an eigenvalue of 1 or greater are incorporated as control variables in the OLS and spatial regressions of deforestation on coca crops, at both national and sub-national levels. The only exception is region D (see Section 4.1), where we retain a smaller number of components because of the reduced number of observations in this region. We allow 1 variable for each 10-15 observations (Harrell et al., 2015), though retaining only components with statistically significant coefficients. We use the **R** software for most of the analysis. In particular, to fit spatial models, we exploit the *spdep* package that contains all spatial linear model functions (Bivand and Piras, 2021). We specify two options in the functions that make the results more robust. First, we change the default method to "LU"[1] which is a standard decomposition method for sparse matrices without permutations . The second specification is an additional control that uses the eigenvalue decomposition from the weights matrix.

4 Results

4.1 Municipalities clustering

We divide Colombia into four regions, as shown in Figure 7, that correspond broadly to the following geographical regions: (A) Caribbean/Sierra Nevada, (B) Pacific/Western Andean Foothills, (C) Andean Mountain Range, and (D) Amazonia/Eastern Andean Foothills/Orinoco. The four regions include, respectively, 133, 675, 173, and 117 municipalities.

[1] A lower–upper (LU) decomposition or factorization is a similar process to Gaussian elimination and is equivalent in terms of elementary row operations. Matrix A is decomposed so that $A = LU$ where L is a lower triangular matrix with a leading diagonal of ones and U is an upper triangular matrix. The product sometimes includes a permutation matrix as well, which is an available method in the *spdep* package, too. Computers usually solve square systems of linear equations using LU decomposition (Lindfield and Penny, 2019).

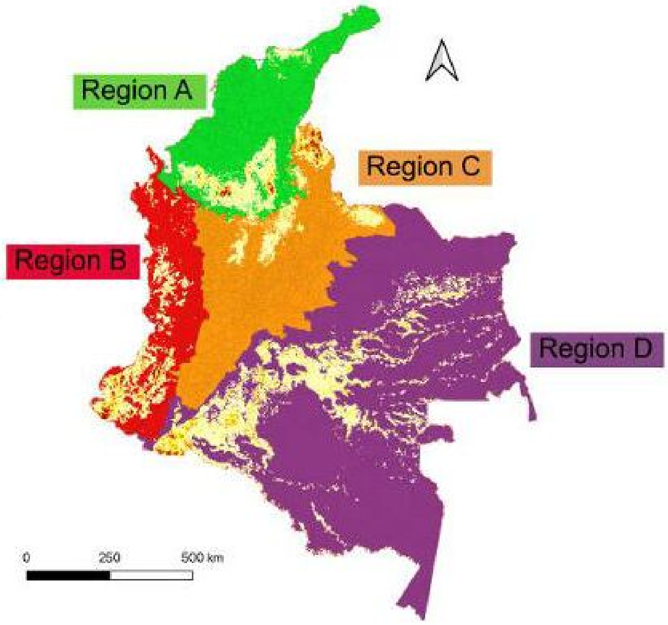


Figure 7. K-means clustering of Colombian municipalities

The partition into macro-regions adopted here is different from the one used in Armenteras et al. (2011, 2013), who consider the five Colombian natural macro-regions instead. We decide to rearrange the 5 natural regions of Colombia because there are regions, such as the Pacific, that have less than 30 municipalities, which is insufficient data to run a spatial regression. Also, Amazonia has a reduced number of observations, consequently, the results would be unreliable for both of these regions. On the other hand, there are regions where municipalities are very small and the territory extension is large. For instance, the region of the Andes has over 700 municipalities. Therefore, we balance the number of observations with clusterization criteria that would allow a spatial regression for each macro-region. A similar strategy is used in Dávalos et al. (2011), where the authors adopted for a partition based on geographic coordinates (North, Center, and South).

4.2 Principal Component Analysis

We select 44 variables for the direct construction of the principal components and 3 quantitative supplementary variables, namely the average annual deforestation area, the average annual coca cultivation area and the number of years during which coca crops were observed in the municipalities. These supplementary variables are highlighted in blue in Figure 8. Also, we consider a qualitative supplementary variable, defining five natural regions, which does not modify the PCA outcome but allows to find patterns in the data through graphical visualization. All variables were standardized, which is a customary procedure before performing PCA, considering that this technique is sensitive to the variance of the original variables.

The first component extracts most of the variability in the data (13.3%). This relatively low percentage of explained variance is normal in a large data set with a high number of poorly correlated variables. In this national PCA, deforestation, years of coca and coca crops place close each other in Figure 8 revealing a certain positive correlation between these variables, which is then further investigated. This is also the case of elevation and wealth. For example, Bogotá has a mean elevation of 2,640 meters above sea level (m.a.s.l.). Medellín has a mean altitude of 1,495 m.a.s.l., and Cali 1,018 m.a.s.l., and these are the municipalities that contribute the most to the national GDP, in the same order. On the other side, Fuente de Oro in Meta, with an altitude of 359 m.a.s.l. and Fundación Magdalena (10 m.a.s.l.) are, instead, the municipalities that contribute the least to the national GDP (DANE, 2020). For this reason, elevation and wealth are inversely correlated.

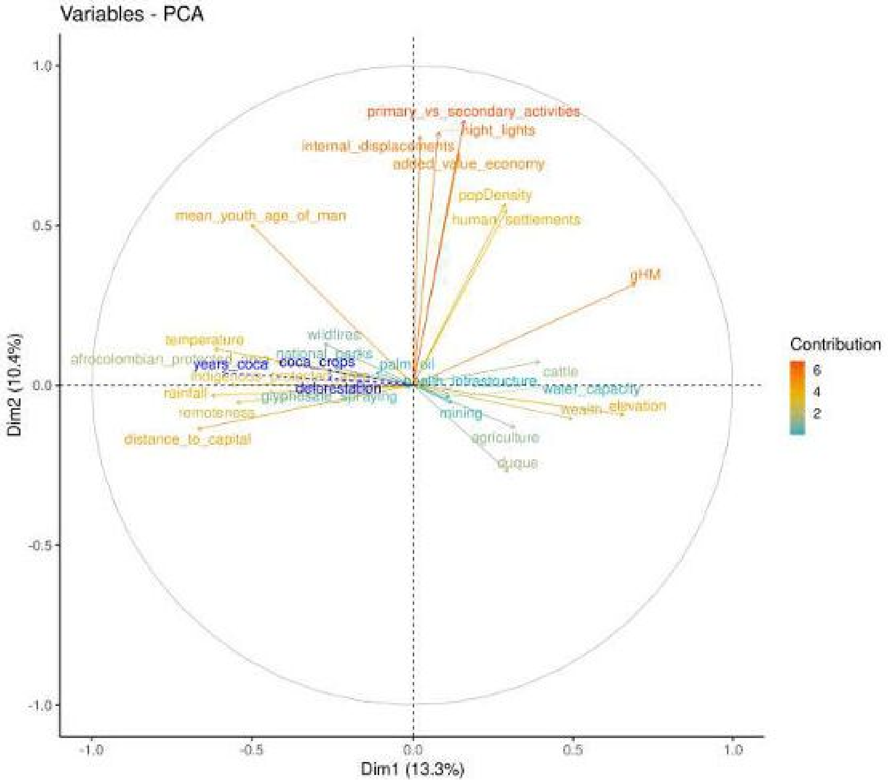


Figure 8. Variable contribution to the first two principal components.

Components 1 and 2, account jointly for 23.7% of the total variance. The plot of these two dimensions (Figure 9) suggests that the Pacific and Amazonia regions share more characteristics in common than the other three regions. On the other hand, the Caribbean and the Flatlands (Llanos) are more closely related to each other than to the Andes. The Andes is the furthest point on the graph and is quite opposite to Amazonia, outlining different socioeconomic and biophysical conditions. The ellipse around the symbols surrounds the 95% confidence interval for the centroid of the observations.

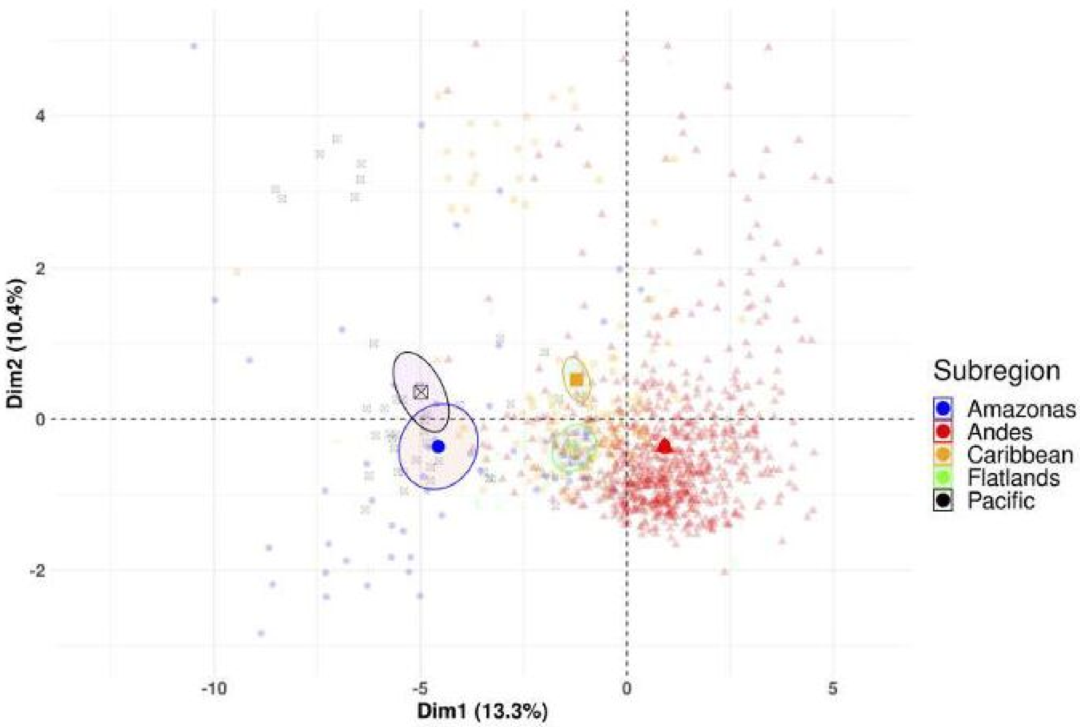


Figure 9. Municipalities and subregion centroids on the first two principal components.

Table 3 is a list of the first 14 components used in the regressions; they are selected as having an eigenvalue greater than 1. With these 14 components, we retain a cumulative variance of about 70%.

Table 3. First 14 components from the PCA

Component	Eigenvalue	Cumulative in %
Comp.1	5.87	13.34
Comp.2	4.58	23.75
Comp.3	3.36	31.39
Comp.4	3.24	38.75
Comp.5	2.13	43.61
Comp.6	1.84	47.79
Comp.7	1.60	51.43
Comp.8	1.36	54.53
Comp.9	1.27	57.42
Comp.10	1.22	60.19
Comp.11	1.15	62.81
Comp.12	1.14	65.42
Comp.13	1.04	67.78
Comp.14	1.00	70.05

4.3 Spatial Modelling

We test three different spatial weights matrices and find out that the one providing more significant and conceptually better results was W_d, determined by the inverse distance from the centroid of the polygon to a cut-off distance of 200 km, which means that all polygons whose centroids are less than 200 km away from each other are neighbors and their weight in the matrix is the inverse distance measured in km. We can illustrate the neighborhood of the inverse distance criterion with a practical example. In Figure 10, *Miraflores* is a middle-size municipality in southern Colombia that has 11 municipalities considered as neighbors. Under this criterion, some municipalities in the Andean region have over 400 neighbors, because the municipalities are quite small, roads are more abundant and the population density is higher. Once an appropriate weighting matrix is selected, spatial models are specified and estimated. In this next step, we use the Moran I test as a diagnostic tool to determine whether a spatial structure is present in the data. For this purpose, the null hypothesis of randomness in the OLS residuals can be tested

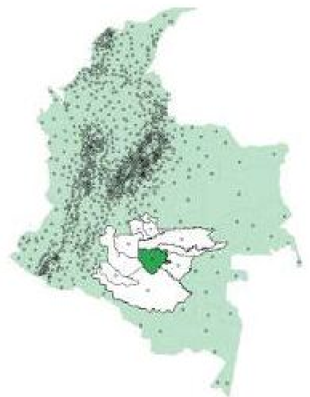


Figure 10. Inverse distance with 200 km cut-off neighbors of *Miraflores*, a municipality in the department of Guaviare.

against the alternative hypothesis of spatial autocorrelation. Table 4 shows the results of this test at both the national and regional levels.

From Table 4, all models exhibit spatial autocorrelation. The LM test can then be used to assist in selecting the appropriate specification of the spatial structure. However, using the LM test exclusively is not enough, and it should be considered jointly with the inferential results on the spatial regression coefficients (Anselin and Florax, 1995; Anselin and Bera, 1998).

Table 4. Specification tests: Moran I (two sided tail) and Lagrange Multiplier test

Statistic	National	A	B	C	D
Moran Std. Deviate	17.64^{***}	9.24^{***}	8.25^{***}	15.42^{***}	5.18^{***}
Observed Moran I	$8.11e-02$	$6.62e-02$	0.076	$5.92e-02$	0.25
Lmerr[1]	215.26^{***}	21.50^{***}	19.87^{***}	121.58^{***}	17.20^{***}
Lmlag[2]	128.12^{***}	13.63^{***}	20.95^{***}	47.60^{***}	22.80^{***}
RLMerr[3]	93.55^{***}	8.51^{**}	1.04	76.73^{***}	0.02
RLMlag[4]	6.40^{**}	0.63	2.13	2.75	5.62^{**}
SARMA[5]	221.66^{***}	22.13^{***}	21.99	124.33^{***}	22.83^{***}

$^{***}p \leq 0.001$; $^{**}p \leq 0.01$; $^{*}p \leq 0.05$

[1] Simple LM test for error dependence.

[2] Simple LM test for a missing spatially lagged dependent variable.

[3] Robust LM test for error dependence.

[4] Robust LM test for a missing spatially lagged dependent variable.

[5] LMerr + RLMlag.

Considering the results of LM test in Table 4, the inference on the coefficients and looking for the most parsimonious model, we select a final model for each region. The results for the final models are shown in Table 5 at the national and regional level, respectively. Complete results on the eight models estimated for each region are provided in the Appendix. The SLX model is estimated using Ordinary Least Squares, while the rest of the spatial models are estimated using Maximum Likelihood Estimation.

Table 5. Spatial Regression Results

		Regions			
	National	A	B	C	D
	SDEM	SLX	SLX	SDEM	SLX
(Intercept)	0.116	0.016	0.122	0.220	−0.063

	(0.334)	(0.048)	(0.126)	(0.681)	(0.079)
Coca crops	0.164^{***}	-0.049	0.249^{**}	0.303^{***}	-0.063
	(0.028)	(0.063)	(0.081)	(0.039)	(0.11)
PC1	-0.317^{***}	0.252	-0.3	0.365^{***}	1.169^{***}
	(0.052)	(0.139)	(0.215)	(0.074)	(0.226)
PC2	0.050^{*}	-0.060	0.203^{*}	-0.180^{***}	—
	(0.026)	(0.074)	(0.09)	(0.045)	—
PC3	-0.336^{***}	-0.238^{*}	-0.192	-0.255^{***}	—
	(0.037)	(0.092)	(0.122)	(0.049)	—
PC4	-0.040	0.133	-0.236^{*}	-0.057	—
	(0.027)	(0.094)	(0.094)	(0.043)	—
PC5	-0.202^{***}	-0.131	-0.12	-0.196^{**}	-0.655^{**}
	(0.038)	(0.112)	(0.131)	(0.061)	(0.232)
PC6	-0.157^{***}	0.145	-0.43^{***}	0.016	—
	(0.039)	(0.119)	(0.125)	(0.060)	—
PC7	-0.037	-0.113	0.407^{***}	0.036	0.573^{***}
	(0.031)	(0.093)	(0.114)	(0.047)	(0.129)
PC8	-0.065^{*}	0.081	-0.233^{*}	-0.038	—
	(0.027)	(0.066)	(0.111)	(0.047)	—
PC9	0.033	-0.157^{*}	-0.006	-0.086^{*}	-0.441^{**}
	(0.025)	(0.07)	(0.125)	(0.034)	(0.152)
PC10	0.007	0.104	-0.009	0.016	—
	(0.029)	(0.084)	(0.096)	(0.044)	—
PC11	-0.073^{**}	-0.162^{*}	0.175	-0.109^{**}	—
	(0.027)	(0.081)	(0.109)	(0.038)	—
PC12	-0.069^{**}	-0.063	-0.095	-0.004	0.371^{*}
	(0.026)	(0.067)	(0.13)	(0.038)	(0.147)
PC13	-0.096^{***}	0.102	0.226^{**}	0.137^{***}	-0.433^{**}
	(0.024)	(0.06)	(0.086)	(0.037)	(0.147)
PC14	0.114^{***}	0.361^{***}	-0.203	0.031	—
	(0.025)	(0.084)	(0.129)	(0.042)	—
$\mathbf{W} \cdot \mathbf{u}$	0.932^{***}	—	—	0.959^{***}	—
	(0.028)	—	—	(0.027)	—

W·Coca crops	0.784**	0.466**	2.372*	−0.158	−0.358
	(0.253)	(0.146)	(0.931)	(0.368)	(0.376)
W·PC1	0.583**	0.171	−2.266*	1.063	−1.872**
	(0.198)	(0.212)	(0.076)	(0.909)	(0.676)
W·PC2	−0.070	−0.173	2.543*	−3.780***	—
	(0.337)	(0.155)	(1.077)	(0.909)	—
W·PC3	−0.042	0.121	0.877	−0.156	—
	(0.226)	(0.143)	(0.826)	(0.462)	—
W·PC4	−0.396	0.27	1.672	−4.312***	—
	(0.253)	(0.149)	(1.332)	(0.927)	—
W·PC5	−0.297	0.203	4.534***	0.268	0.859
	(0.189)	(0.192)	(1.343)	(0.813)	(0.533)
W·PC6	0.501**	0.036	−1.434	0.016	—
	(0.161)	(0.181)	(0.736)	(0.441)	—
W·PC7	−0.354	0.044	−0.967	−1.577**	−0.015
	(0.194)	(0.168)	(1.173)	(0.608)	(0.379)
W·PC8	−0.229	0.447***	−2.592*	−0.528	—
	(0.215)	(0.133)	(0.146)	(0.615)	—
W·PC9	0.005	−0.22	4.71*	−0.208	−0.294
	(0.261)	(0.189)	(1.892)	(0.526)	(0.427)
W·PC10	−0.251	0.181	−0.135	1.331**	—
	(0.180)	(0.139)	(1.065)	(0.438)	—
W·PC11	0.463	0.055	−3.071*	−0.893	—
	(0.242)	(0.152)	(1.293)	(0.515)	—
W·PC12	−0.625***	−0.087	−5.091**	−1.771***	−0.632
	(0.188)	(0.138)	(1.671)	(0.518)	(0.442)
W·PC13	0.107	0.12	4.141***	−1.063*	−0.489
	(0.232)	(0.13)	(1.043)	(0.494)	(0.489)
W·PC14	−0.173	0.407	2.154	0.219	—
	(0.137)	(0.232)	(1.871)	(0.545)	—
Adj. R^2	—	0.642	0.477	—	0.551
Num. obs.	1,060	196	173	613	78
Parameters	33	—	—	33	—

σ^2	0.517	—	—	0.454	—
Pseudo R^2	0.468	—	—	0.537	—
Wald test	1077.845^{***}	—	—	1247.482^{***}	—
Log.Lik(Linear)	−1207.901	—	—	−651.028	—
Log.Lik(Spatial)	−1168.683	—	—	−633.612	—
AIC (Linear)	2479.803	416.393	464.376	1366.056	187.16
AIC (Spatial)	2403.366	355.992	408.687	1333.223	174.29
LR test	78.437^{***}	—	—	34.833^{***}	—

Significance: $^{***}p \leq 0.001$; $^{**}p \leq 0.01$; $^{*}p \leq 0.05$

Table 6. Model comparison of the estimated direct and spillover effects of coca crops on deforestation

Model	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS	Mean	Std.Dev
					Direct effects					
National	0.23^{***}	0.16^{***}	0.23^{***}	0.19^{***}	0.19^{***}	0.19^{***}	0.16^{***}	0.19	0.19	0.03
	(0.03)	(0.03)	(0.03)	(0.028)	(0.04)	(0.03)	(0.03)	(0.19)		
Region A	0.12^{*}	0.11	0.11	0.06	0.07	0.01	0.01	0.01	0.06	0.05
	(0.06)	(0.06)	(0.11)	(0.06)	(0.22)	(0.11)	(0.05)	(0.1)		
Region B	0.24^{**}	0.25^{**}	0.28^{***}	0.24	0.25^{*}	0.3^{*}	0.25^{***}	0.3	0.26	0.03
	(0.08)	(0.08)	(0.09)	(0.082)	(0.09)	(0.1)	(0.07)	(0.43)		
Region C	0.32^{***}	0.32^{***}	0.27^{***}	0.32^{***}	0.3^{***}	0.31^{***}	0.3^{***}	0.3^{***}	0.31	0.02
	(0.03)	(0.04)	(0.03)	(0.036)	(0.04)	(0.05)	(0.04)	(0.04)		
Region D	0.1	-0.06	0.06	0.05	0.06	−0.06	−0.1	−0.09	0	0.07
	(0.11)	(0.11)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)		
					Spillover effects					
National	—	1.06^{***}	1.07^{**}	—	0.73	5.88	0.78^{**}	2.71	2.01	2.05
	—	(0.18)	(0.49)	—	(26.48)	(9.37)	(0.253)	(191.9)		
Region A	—	0.47^{**}	0.14	—	0.09	2.46	1.88^{*}	2.33	1.2	1.1
	—	(0.15)	(0.1)	—	(0.2)	(1.6)	(0.94)	(1.62)		
Region B	—	2.37^{*}	1.14	—	0.45	5.11	2.35^{***}	4.95	2.72	1.93
	—	(0.93)	(5.98)	—	(9.27)	(12.46)	(0.86)	(67.84)		
Region C	—	−0.084	0.94	—	2.05	−0.16	−0.16	−0.97	0.21	1.1
	—	(0.29)	(1.83)	—	(11.43)	(20.81)	(0.37)	(13.79)		
Region D	—	−0.36	0.07	—	0.04	−0.49	−0.38	−0.56	−0.27	0.27
	—	(0.38)	(0.18)	—	(1.38)	(0.58)	(0.31)	(2.13)		

Significance: $^{***}p \leq 0.001$; $^{**}p \leq 0.01$; $^{*}p \leq 0.05$

Note: Standard errors and p-values of the SAR, SAC, SDEM and GNS models are simulated values with an initial random value seed equal to 1.

We choose the SDEM for the National and Region C data. The SDEM, as well as the SLX and the SEM, is a local model, which is theoretically justified particularly in the case of the national level data. The purpose of choosing a local model is to avoid looped effects captured by the $\mathbf{Wy}$ vector, that could bias the spatial parameters and account for effects that are not in our defined neighborhood in $\mathbf{W}$. In regions A, B and D, we choose an SLX model, which is the simplest of all spatial models. For both the SLX and SDEM, the single municipality effects are captured by β and the surrounding effects by θ. In these models, β_j is a non spatial parameter that reflects the impact of the j^{th} variable at municipality i on the response at the same municipality i (direct effects), which is the usual interpretation in a linear non-spatial model, while the parameter θ_j reflects the surrounding effect on municipality i (indirect effect or spillover). Hence, a crucial element of SLX and SDEM is that there are no prior restrictions imposed on the ratio between the direct effects and spillover effects, which is a shortcoming of the SAR and SAC models. We estimate also the GNS model for completeness but do not consider it as a possible "best" model for any of the data sets, because it is often overparametrized and prone to overfitting. Similarly, the Durbin model has a global spillover specification which might be counter-intuitive in our study case. Besides, unless theoretically supported, global spillover specifications are difficult to justify and have been overused in applied studies (Elhorst and Vega, 2013).

The national model has an estimated direct effect of coca crops on deforestation of 0.164, which is significant at the 1‰ level: an increment of one standard deviation in coca cultivated area in a given municipality is associated with an increase of 0.164 standard deviations of deforested area in that municipality. As shown in Table 6 this direct impact is quite stable among different model specifications, ranging from 0.16 in the SLX and SDEM models to 0.23 in the SAR and non-spatial model, with a mean of 0.19 and a between models standard deviation of 0.03. The coefficient of coca crops in our SAC specification (equal to 0.183, see Table 7 in Appendix 1C) is also in line with those obtained by Mendoza (2020) under different spatial panel fixed effects SAC model specifications[2] (values ranging from 0.106 to 0.271 in low conflict intensity municipalities and from 0.095 to 0.250 in high conflict intensity municipalities). Note that Mendoza uses a per-pixel forest cover threshold of 30% for the GFCD data base, while we set it at 70%. So, at least at national level, results seem to be quite robust also to this threshold specification. Results are instead different from those in Armenteras et al. (2013), who use non-spatial generalized linear models to study the drivers of

[2]Mendoza (2020) does not provide direct effects but only coefficient values, so we can only compare these coefficients with those we obtain under the SAC specification.

deforestation in Colombia. Armenteras et al. provide only the sign of the coefficients, not their magnitude, and in their national model they find a positive coca crops coefficient, which is however not statistically significant. Finally, note that for the chosen model SDEM also the spillover effect of coca-crops is positive (equal to 0.78) and significant at the 1% level. The coefficient of the lagged error term is positive, significantly different from zero and takes values close to one in all the specifications that include a lagged error term, testifying a positive correlation between the error terms of neighboring municipalities.

In region A, the direct effect of coca crops is not statistically significant, in all the models considered. Therefore, for this region, the data do not seem to support the claim of an impact of coca crops on deforestation within municipalities. This result is in line with some previous findings. Dávalos et al. (2011) also conclude that coca cultivation is not associated with an increase in the probability of deforestation in the northern part of Colombia. On the contrary, Armenteras et al. (2013) find a significant and positive coefficient for illicit cultivation in their Caribbean region. Note, however, that their Caribbean region, with respect to our region A, includes a good portion of the Santander department, which is a department with a particularly strong association between coca cultivation and deforestation. See the discussion of the results for region C for further details. It is interesting to note that region A also includes a part of the department of Antioquia, an area where the *Global Human Modification Index* is high and the land used to grow coca is not always covered by forest, but already intended for other human uses, in most cases. For this reason, coca cultivation is not causing further deforestation in this area. The bivariate LISA map, being limited to the interaction of the two variables alone, provides results that can be somewhat in contrast with those obtained through multivariate regression. Nevertheless, there are positive and significant spillover effects under both the SDEM (at the 5% level) and the SLX (at the 1% level) model.

In Region B the estimate of the direct effect of coca crops on deforestation is quite stable among different models. For the SLX model chosen for this region, its value is 0.25, significantly different from zero at the 1% level. In addition, a positive and significant spatial local spillover can also be observed. Similarly, Dávalos et al. (2011) show that coca cultivation increases the probability of forest conversion in the Chocó department, included in our region B.

Also region C has a positive, statistically significant and stable across models direct effect of coca crops. In particular, this is the region showing the highest values of this effect, immediately followed by region B. For the SDEM model the estimated direct effect is equal to 0.303 and highly significant. For this region, the bivariate LISA map pinpoints a large hotspot in Northern Santander. This is an area, as already said, where coca cultivation seems to strongly contribute to deforestation. For example, moving this hotspot to region A would make the direct effect of coca crops in this region positive and highly significant, providing results in accordance with those of Armenteras et al. (2013) for their Caribbean region. Dávalos et al. (2011) also conclude that coca cultivation increases the probability of forest conversion in the northern Andes (a zone roughly included in our region C), even if they do not find a significant association for the central region as a whole. In the same vein, Armenteras et al. (2011) found a positive and significant relationship between illicit crops and the probability of deforestation in the Andean region. Differently from previous regions, spillover effects seem to be not significantly different from zero in region C. Lastly, the spatial autocorrelation coefficient λ has a value near to 1 and is statistically significant in those spatial specification in which it is present (Table 10 in Appendix 1C), implying that the error terms across neighboring spatial units are correlated.

In region D, neither the direct nor the indirect effects of coca crops appear to be significantly different from zero. Thus, data do not provide enough evidence to support the claim of an impact of coca crops on deforestation in Amazonia and Orinoco. This region has only 78 observations, for this reason, we removed non-significant principal components from the regression. In Amazonia and Orinoco, municipalities span over large territories, thus the lack of detail in the information may act as a source of errors, and the spatial models are probably capturing it. With respect to this area, Armenteras et al. (2013) found a non-significant relationship between illicit crops and deforestation in Amazonia, while they conclude in favor of a positive association between the two variables in Orinoco. In the Southern region, Dávalos et al. (2011) found that the probability of transition from forest to nonforest is found to increase significantly as coca cultivated area increases, and to decrease significantly as the distance from new coca crops increases. However, their Southern region encompasses only part of our Region D, while including portions of Region C and B, thus undermining results comparability.

5 Discussion

This study responds to the growing need to understand the main drivers of tropical deforestation, which are causing an alarming biodiversity loss worldwide, endangering half of Colombia's ecosystems, impacting the monetary land resources, and therefore negatively impinging on the economy of the country (WWF, 2017; Negrete et al., 2019).

In particular, the aim of this study is to assess if, and to what extent, coca crops contribute to deforestation. At this purpose, we used spatial models on cross-sectional municipality data, running regressions at both national and macro-regional levels, in order to account for possibly heterogeneous effects. Macro-regions were defined on the basis of coca crops and deforestation values, using the K-means method plus a slight readjustment to obtain four well-defined regions. A set of over 40 relevant control variables was considered and its dimensionality reduced by means of principal components analysis, while retaining more than 70% of the original variability in the data set.

The novelty of our approach, in comparison with previous studies, is manifold. Firstly, rather than relying on predefined and not necessarily suitable grouping of municipalities into a north-central-south classification or similar geographical classifications, we selected macro-regions through an unsupervised machine learning method of clusterization, thus defining more homogeneous groups of municipalities with respect to the two variables of interest. Secondly, to compute forest loss extension, we used a different threshold for the GFCD data and set this value to 70%, basing our choice on the consideration of the predominant forest type in Colombia and following the guidelines in Lwin et al. (2019). Thirdly, we gathered a relevant number of variables from very different sources, creating - up to our knowledge - the largest data base of control variables ever used in studies on deforestation in Colombia. Finally, we compared seven different spatial models to assess the stability of direct and spillover effects of coca crops on deforestation.

Our results indicate that the presence of coca cultivation increases notably the extent of deforestation in Colombia. This is particularly true in two regions that include two pivotal biodiversity hotspots in the Andes and the Pacific Coast. The other two regions, Amazon/Orinoco and Caribbean/Sierra Nevada did not show a significant direct effect of coca crops on forest loss, although significant spillover effects could be present in the Caribbean/Sierra Nevada region.

In the current debate on the negative environmental outreach of coca cultivation, there is a general acceptance that an effect on deforestation do exist. In this study, we showed that this impact is not negligible, and is amplified by spillover effects. These turned out to be particularly relevant, an aspect that surely deserves further investigation. A possible cause can be ascribed to trafficking, which is the second step in the commercialization of this plant-based drug. Even if not yet proved, probably coca trafficking is driving deforestation within the producing countries. What is certain is that this is happening in transiting countries, like Central America, where the effect has been singled out (McSweeney et al., 2014; Wrathall et al., 2020; Devine et al., 2021).

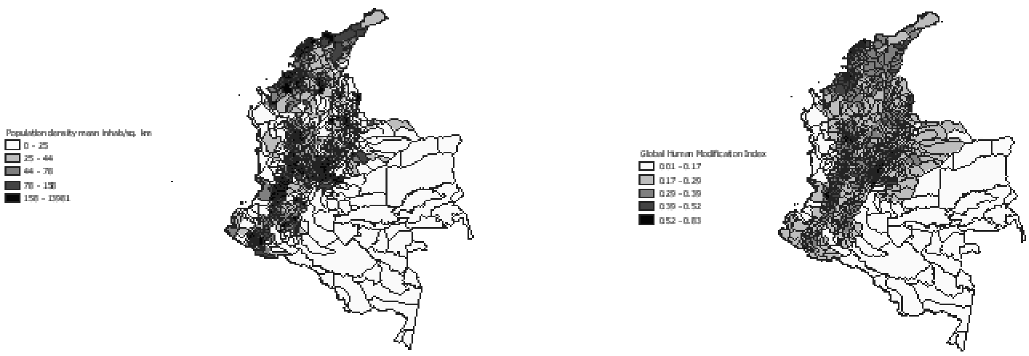
We believe that this work provides important findings for policymakers and civil institutions that are concerned with deforestation patterns and want to pinpoint the leading causes in Colombia. To mitigate deforestation it is crucial to address the underlying drivers and our work supports the claim that coca cultivation is one of them. Reducing coca crops can help in preserving forests and, as McSweeney et al. (2014) clearly state, "drug policy is conservation policy". Obviously, this is not a simple task and there is not a unique and permanent solution. The forces funneling coca-related activities and deforestation are changing constantly: the peace treaty, the COVID pandemic, the formation of new paramilitary groups, or the 2021 protests act as enhancers or diminishers of deforestation and coca cultivation. This was manifest with the peace treaty of 2016 that represented a watershed in the trend of deforestation and coca cultivation in the country. During conflict, many remote areas were occupied by the guerrilla, and this had the side effect of stunningly protecting forests from private interests and containing the agricultural frontier. Therefore, when this force disappeared, land grabbing and private interests took over the liberated lands, and the peace treaty unintentionally determined an impressive acceleration in deforestation. This was also fueled by the rapid increase in coca cultivation, which followed the peace treaty and which was partly due to the failures in the implementation of adequate substitution programs. So far, drug policies have failed to reduce coca cultivation and trafficking, sprawling on the contrary violence and corruption. This was also due to the fact that drug policies are wrongly focused only on the supply-side through interdiction and eradication, while a more unwavering solution would be to focus on the demand-side (McSweeney et al., 2014).

6 Conclusions

We add a few considerations and suggestions for further research. As we hinted before, the extremely varied and disparate sources of our seemingly large data set imposed a challenge in terms of data matching and data fusion, whereas our ultimate goal was to obtain a record with comparable, meaningful, clean and consistent information. However, the fact that our data set contains more variables than other comparable studies is an advantage in terms of robustness of the spatial regressions, but is also a disadvantage as it represents a source of multiple and varied types of intrinsic errors that were patent in some of the variable's sources. Colombia is in the process of constructing reliable statistics and data, but this is just happening in these latest years, hence, historical data on economic and social aspects, in the best case scenario, date back to the early 2000s. Moreover, data are not collected regularly or with a consistent methodology through the years, and in many cases the opposite is the rule, meaning that data are found in intermittent years and changing methodologies are quite common (International Monetary Fund, 2006). Considering that the quality and span of historical data are likely to enlarge and improve in the upcoming years, monitoring of the forest's state as well as future multivariate regressions to understand the etiology of forest loss should profit of this increased accuracy. In connection with these data availability and reliability worries, we posed in this work the first concerns regarding the most suitable data source, between GFCD, ESA or IDEAM, to estimate deforestation rates in Colombia, and we tried to provide a more reasoned choice of the threshold used for the GFCD. We believe, however, that a more thorough assessment of the accuracy of these three sources, through collection and comparison with reference data, is desirable and even essential to achieve more reliable analyses and results.

Appendices

1A National maps of different variables

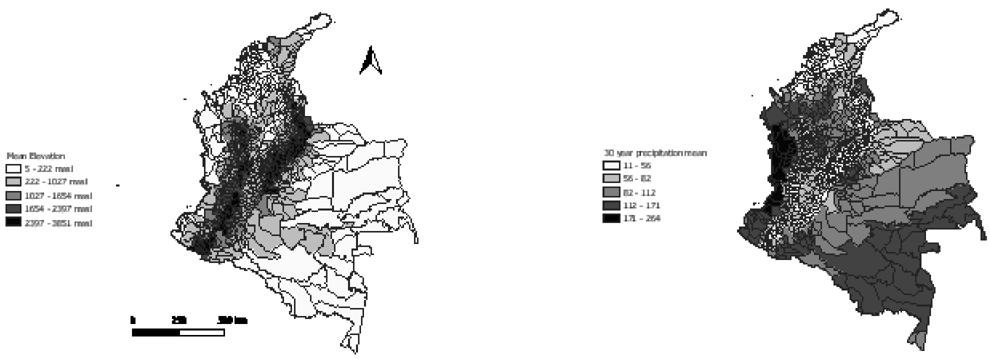


(a) (b)

Figure 11. Maps of (a) population density (b) Global Human Modification Index.

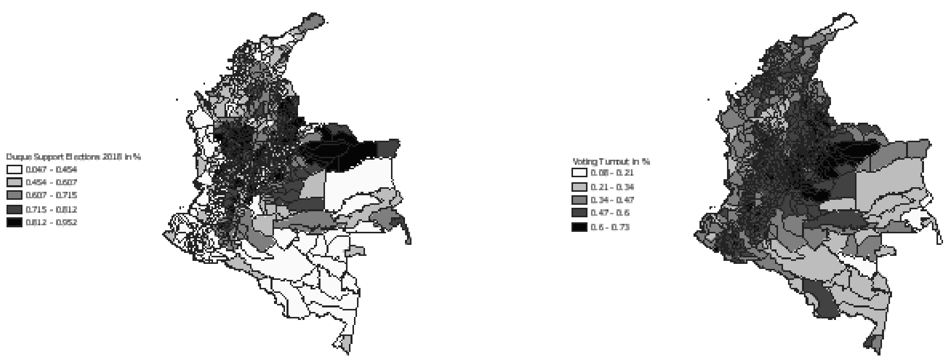
(a) (b)

Figure 12. Maps of (a) remoteness (b) soil water capacity.



(a) (b)

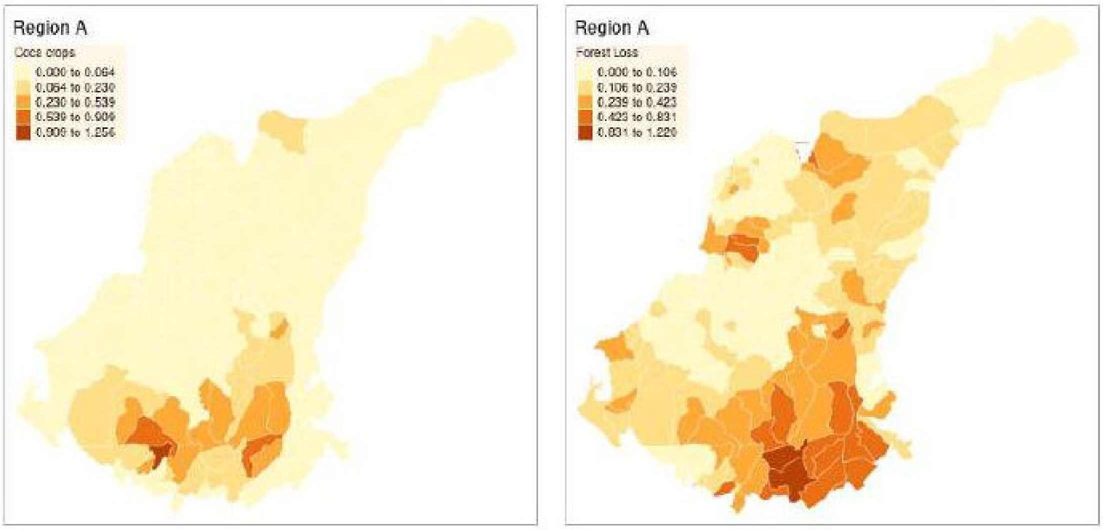
Figure 13. Maps of (a) elevation mean (b) 30 year precipitation mean.



(a) (b)

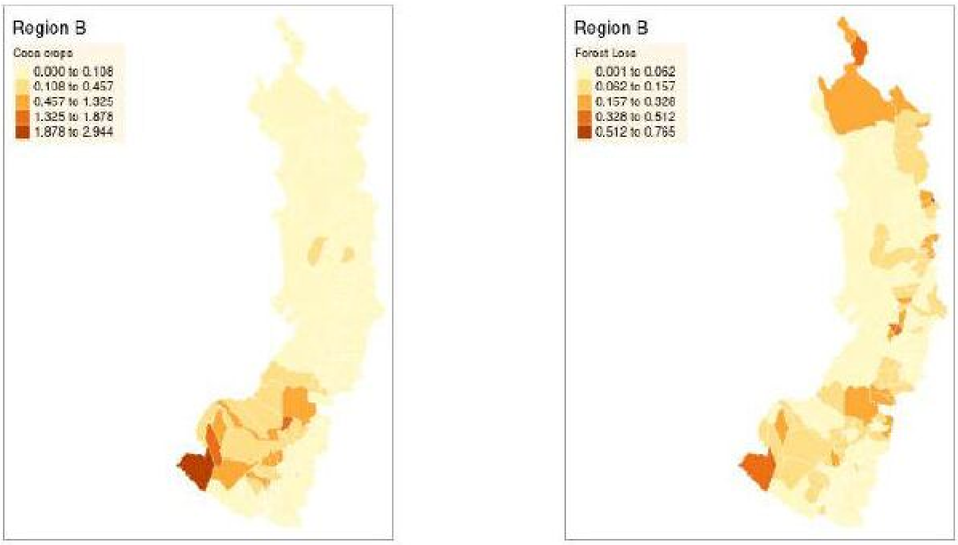
Figure 14. Maps of (a) Duque supporters in % (b) voting turnout in %.

1B Regional maps of coca cultivation and deforestation



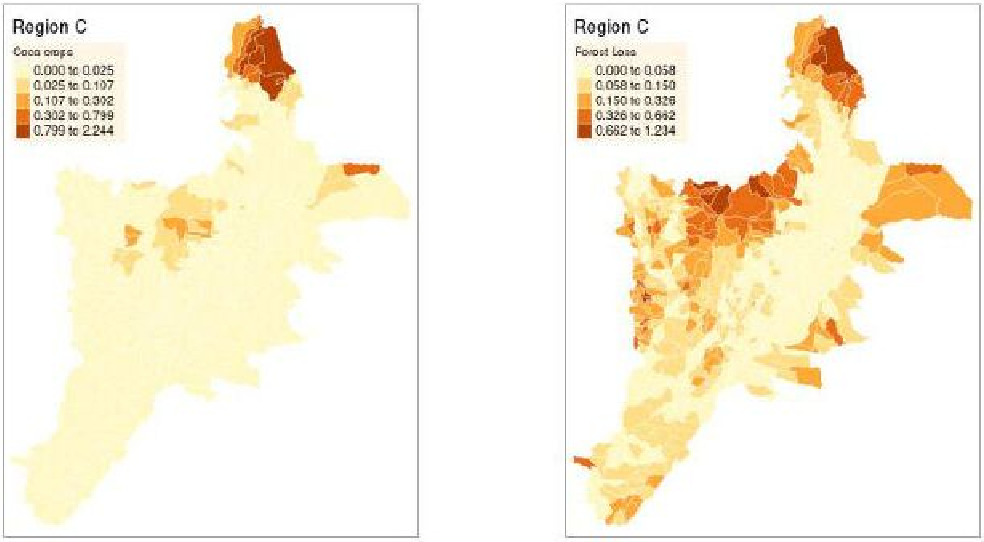
(a) (b)

Figure 15. Region A: (a) average annual coca cultivation and (b) deforestation area, as percentage of municipal area, between 2000 and 2020.



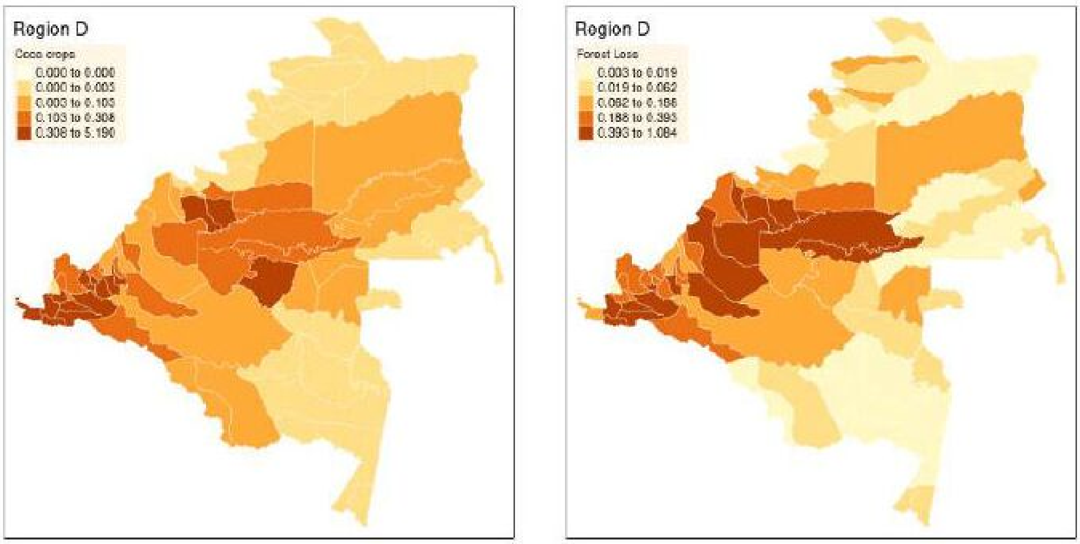
(a) (b)

Figure 16. Region B: (a) average annual coca cultivation and (b) deforestation area, as percentage of municipal area, between 2000 and 2020.



(a) (b)

Figure 17. Region A: (a) average annual coca cultivation and (b) deforestation area, as percentage of municipal area, between 2000 and 2020.



(a) (b)

Figure 18. Region D: (a) average annual coca cultivation and (b) deforestation area, as percentage of municipal area, between 2000 and 2020.

1C Spatial regression models: national and regional results

Table 7. National spatial regression results

Dependent variable: deforestation								
Independent variable	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS
(Intercept)	0.000	0.005	0.047^{*}	0.065	−0.014	0.014	0.116	0.059
	(0.025)	(0.03)	(0.023)	(0.476)	(0.279)	(0.028)	(0.334)	(0.239)
Coca crops	0.226^{***}	0.156^{***}	0.224^{***}	0.186^{***}	0.183^{***}	0.163^{***}	0.164^{***}	0.168^{***}
	(0.029)	(0.03)	(0.027)	(0.028)	(0.027)	(0.028)	(0.028)	(0.029)
PC1	-0.200^{***}	-0.301^{***}	-0.123^{***}	-0.294^{***}	-0.245^{***}	-0.299^{***}	-0.317^{***}	-0.326^{***}
	(0.026)	(0.051)	(0.026)	(0.040)	(0.040)	(0.049)	(0.052)	(0.052)
PC2	0.012	0.047	−0.006	0.049	0.047	0.046	0.050^{*}	0.048
	(0.025)	(0.027)	(0.024)	(0.025)	(0.024)	(0.025)	(0.026)	(0.025)
PC3	-0.252^{***}	-0.335^{***}	-0.177^{***}	-0.351^{***}	-0.314^{***}	-0.329^{***}	-0.336^{***}	-0.329^{***}
	(0.025)	(0.037)	(0.026)	(0.035)	(0.034)	(0.036)	(0.037)	(0.037)
PC4	-0.064^{*}	−0.026	-0.050^{*}	−0.038	−0.024	−0.029	−0.040	−0.041
	(0.025)	(0.028)	(0.023)	(0.026)	(0.025)	(0.027)	(0.027)	(0.027)
PC5	-0.288^{***}	-0.218^{***}	-0.192^{***}	-0.240^{***}	-0.217^{***}	-0.218^{***}	-0.202^{***}	-0.196^{***}
	(0.025)	(0.039)	(0.024)	(0.034)	(0.033)	(0.037)	(0.038)	(0.037)
PC6	0.072^{**}	-0.144^{***}	0.014	-0.090^{**}	-0.090^{**}	-0.124^{***}	-0.157^{***}	-0.145^{***}
	(0.025)	(0.039)	(0.023)	(0.035)	(0.033)	(0.037)	(0.039)	(0.038)
PC7	−0.020	−0.055	−0.040	−0.037	−0.028	−0.040	−0.037	−0.032
	(0.025)	(0.031)	(0.024)	(0.030)	(0.028)	(0.030)	(0.031)	(0.031)
PC8	-0.132^{***}	−0.069	-0.075^{**}	-0.104^{***}	-0.082^{**}	-0.068^{*}	-0.065^{*}	-0.064^{*}
	(0.025)	$(0.028)^{*}$	(0.023)	(0.026)	(0.025)	(0.027)	(0.027)	(0.027)
PC9	0.042	0.038	0.057^{*}	0.034	0.032	0.036	0.033	0.035
	(0.025)	(0.026)	(0.023)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)
PC10	−0.008	0.012	−0.014	0.001	−0.010	0.009	0.007	0.009
	(0.026)	(0.03)	(0.024)	(0.028)	(0.027)	(0.028)	(0.029)	(0.049)
PC11	-0.070^{**}	-0.083^{**}	−0.041	-0.056^{*}	-0.054^{*}	-0.073^{**}	-0.073^{**}	-0.065^{*}
	(0.025)	(0.028)	(0.024)	(0.026)	(0.025)	(0.027)	(0.027)	(0.026)
PC12	-0.150^{***}	-0.079^{**}	-0.099^{***}	-0.086^{***}	-0.069^{**}	-0.073^{**}	-0.069^{**}	-0.063^{*}
	(0.025)	(0.027)	(0.024)	(0.026)	(0.025)	(0.026)	(0.026)	(0.025)
PC13	-0.069^{**}	-0.097^{***}	-0.091^{***}	-0.081^{***}	-0.086^{***}	-0.094^{***}	-0.096^{***}	-0.094^{***}
	(0.025)	(0.026)	(0.023)	(0.024)	(0.023)	(0.024)	(0.024)	(0.025)
PC14	0.124^{***}	0.121^{***}	0.094^{***}	0.113^{***}	0.102^{***}	0.113^{***}	0.114^{***}	0.105^{***}
	(0.026)	(0.027)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
$\mathbf{W} \cdot y$	—	—	0.827^{***}	—	0.791^{***}	0.877^{***}	—	0.791^{***}
	—	—	(0.042)	—	(0.161)	(0.042)	—	(0.088)

$\mathbf{W} \cdot u$	—	—	—	0.952***	0.893***	—	0.932***	0.893***
	—	—	—	(0.022)	(0.106)	—	(0.028)	(0.070)
W·coca_crops	—	1.058***	—	—	—	0.584***	0.784**	0.438
	—	(0.175)	—	—	—	(0.167)	(0.253)	(0.245)
W·PC1	—	0.376***	—	—	—	0.487***	0.583**	0.672**
	—	(0.119)	—	—	—	(0.112)	(0.198)	(0.209)
W·PC2	—	0.0329	—	—	—	−0.024	−0.070	−0.028
	—	(0.269)	—	—	—	(0.254)	(0.337)	(0.703)
W·PC3	—	0.04	—	—	—	0.226*	−0.042	0.136
	—	(0.098)	—	—	—	(0.093)	(0.226)	(0.217)
W·PC4	—	−0.105	—	—	—	−0.071	−0.396	−0.290
	—	(0.185)	—	—	—	(0.175)	(0.253)	(0.329)
W·PC5	—	−0.089	—	—	—	0.133	−0.297	−0.082
	—	(0.117)	—	—	—	(0.112)	(0.189)	(0.246)
W·PC6	—	0.371***	—	—	—	0.161	0.501**	0.300
	—	(0.094)	—	—	—	(0.090)	(0.161)	(0.173)
W·PC7	—	−0.187	—	—	—	−0.081	−0.354	−0.232
	—	(0.105)	—	—	—	(0.100)	(0.194)	(0.193)
W·PC8	—	−0.398**	—	—	—	−0.198	−0.229	−0.102
	—	(0.166)	—	—	—	(0.157)	(0.215)	(0.223)
W·PC9	—	−0.143	—	—	—	−0.135	0.005	−0.038
	—	(0.179)	—	—	—	(0.169)	(0.261)	(0.291)
W·PC10	—	−0.261*	—	—	—	−0.146	−0.251	−0.138
	—	(0.127)	—	—	—	(0.120)	(0.180)	(0.205)
W·PC11	—	0.598**	—	—	—	0.512**	0.463	0.477
	—	(0.183)	—	—	—	(0.173)	(0.242)	(0.260)
W·PC12	—	−0.478***	—	—	—	−0.217	−0.625***	−0.359
	—	(0.123)	—	—	—	(0.117)	(0.188)	(0.195)
W·PC13	—	0.164	—	—	—	0.053	0.107	0.062
	—	(0.201)	—	—	—	(0.190)	(0.232)	(0.247)
W·PC14	—	−0.154	—	—	—	−0.259*	−0.173	−0.224
	—	(0.114)	—	—	—	(0.108)	(0.137)	(0.141)
R^2	0.358	0.428	—	—	—	—	—	—
Adj. R^2	0.348	0.411	—	—	—	—	—	—
Num. obs.	1,060	1,060	1,060	1,060	1,060	1,060	1,060	1,060
Parameters	—	—	18	18	19	33	33	34
σ^2	—	—	0.573	0.542	0.508	0.524	0.517	0.493
Pseudo R^2	—	—	0.416	0.441	0.470	0.464	0.468	0.487
Wald test	—	—	393.533***	1832.173***	225.941***	426.847***	1077.845***	162.098***
Log.Lik(Linear)	—	—	−1268.975	−1268.975	−1268.975	−1207.901	−1207.901	−1268.975
Log.Lik(Spatial)	—	—	−1218.146	−1195.510	−1166.956	−1173.019	−1168.683	−1149.712
AIC (Linear)	2571.950	2571.950	2571.950	2571.950	2571.950	2479.803	2479.803	2571.950

AIC (Spatial)	—	2479.803}	2472.291	2427.020	2371.912	2412.038	2403.366	2367.424
LR test	—	—	101.659***	146.930***	204.038***	69.765***	78.437***	238.526***

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 8. Spatial regression results for region A

Dependent variable: deforestation								
Independent variable	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS
(Intercept)	−0.000	0.016	0.182***	0.069	0.234	0.139	0.170	0.132
	(0.048)	(0.048)	(0.043)	(0.091)	(0.132)	(0.386)	(0.392)	(0.412)
Coca crops	0.121*	−0.049	0.090	0.057	0.035	−0.009	−0.008	−0.009
	(0.059)	(0.063)	(0.054)	(0.055)	(0.053)	(0.049)	(0.050)	(0.053)
PC1	0.293**	0.252	0.246**	0.312***	0.266	0.147	0.144	0.149
	(0.094)	(0.139)	(0.085)	(0.090)	(0.087)	(0.106)	(0.106)	(0.107)
PC2	−0.035	−0.060	−0.068	−0.018	−0.035	−0.066	−0.067	−0.067
	(0.069)	(0.074)	(0.063)	(0.064)	(0.063)	(0.062)	(0.062)	(0.062)
PC3	−0.138*	−0.238*	−0.199***	−0.191**	−0.230	−0.114	−0.115	−0.116
	(0.062)	(0.092)	(0.056)	(0.064)	(0.064)	(0.067)	(0.068)	(0.068)
PC4	0.194*	0.133	0.099	0.170*	0.099	0.080	0.079	0.079
	(0.083)	(0.094)	(0.075)	(0.078)	(0.078)	(0.088)	(0.088)	(0.089)
PC5	−0.144	−0.131	−0.053	−0.211*	−0.156	−0.118	−0.120	−0.118
	(0.087)	(0.112)	(0.081)	(0.085)	(0.084)	(0.087)	(0.087)	(0.087)
PC6	0.014	0.145	0.154*	0.008	0.105	0.198*	0.200*	0.200*
	(0.075)	(0.119)	(0.069)	(0.077)	(0.076)	(0.092)	(0.092)	(0.092)
PC7	−0.098	−0.113	−0.054	−0.117	−0.087	−0.138	−0.139	−0.136
	(0.076)	(0.093)	(0.069)	(0.073)	(0.071)	(0.072)	(0.072)	(0.072)
PC8	0.210**	0.081	0.193**	0.128*	0.106	0.145*	0.147*	0.144*
	(0.065)	(0.066)	(0.060)	(0.061)	(0.06)	(0.059)	(0.059)	(0.058)
PC9	−0.174*	−0.157*	−0.108	−0.197**	−0.150	−0.200*	−0.197*	−0.197*
	(0.071)	(0.07)	(0.064)	(0.062)	(0.064)	(0.078)	(0.079)	(0.079)
PC10	0.250***	0.104	0.241***	0.216***	0.191	0.252***	0.251***	0.249***
	(0.060)	(0.084)	(0.055)	(0.062)	(0.058)	(0.066)	(0.066)	(0.066)
PC11	−0.262***	−0.162*	−0.144*	−0.264***	−0.179	−0.197**	−0.194**	−0.193**
	(0.064)	(0.081)	(0.059)	(0.064)	(0.066)	(0.068)	(0.068)	(0.069)
PC12	−0.121*	−0.063	−0.083	−0.132*	−0.111	−0.038	−0.039	−0.039

	(0.058)	(0.067)	(0.053)	(0.055)	(0.05)	(0.054)	(0.055)	(0.054)
PC13	0.130*	0.102	0.083	0.124*	0.098	0.101*	0.104*	0.101*
	(0.056)	(0.06)	(0.051)	(0.052)	(0.05)	(0.049)	(0.049)	(0.049)
PC14	0.478***	0.361***	0.467***	0.457***	0.444	0.457***	0.457***	0.456***
	(0.086)	(0.084)	(0.077)	(0.076)	(0.072)	(0.072)	(0.072)	
$\mathbf{W} \cdot y$	—	—	0.969***	—	0.964	0.311	—	0.292
	—	—	(0.069)	—	(0.59)	(0.207)	—	(0.291)
$\mathbf{W} \cdot u$	—	—	—	0.999***	0.999	—	0.213	0.146
	—	—	—	(0.061)	(0.61)	—	(0.249)	(0.394)
W·coca_-crops	—	0.466**	—	—	—	1.708	1.884*	1.666
	—	(0.146)	—	—	—	(0.944)	(0.947)	(0.995)
W·PC1	—	0.171	—	—	—	1.677	1.739	1.714
	—	(0.212)	—	—	—	(1.053)	(1.063)	(1.053)
W·PC2	—	−0.173	—	—	—	−1.918	−1.933	−1.919
	—	(0.155)	—	—	—	(1.031)	(1.044)	(1.045)
W·PC3	—	0.121	—	—	—	−0.620	−0.680	−0.588
	—	(0.143)	—	—	—	(0.504)	(0.509)	(0.527)
W·PC4	—	0.27	—	—	—	1.620	1.696	1.653
	—	(0.149)	—	—	—	(0.949)	(0.960)	(0.962)
W·PC5	—	0.203	—	—	—	1.657	1.614	1.653
	—	(0.192)	—	—	—	(0.866)	(0.879)	(0.869)
W·PC6	—	0.036	—	—	—	1.841*	1.880**	1.832*
	—	(0.181)	—	—	—	(0.717)	(0.730)	(0.730)
W·PC7	—	0.044	—	—	—	0.145	0.130	0.150
	—	(0.168)	—	—	—	(0.833)	(0.843)	(0.838)
W·PC8	—	0.447***	—	—	—	2.899**	3.007***	2.914**
	—	(0.133)	—	—	—	(0.895)	(0.904)	(0.902)
W·PC9	—	−0.22	—	—	—	−1.788*	−1.802*	−1.781*
	—	(0.189)	—	—	—	(0.872)	(0.876)	(0.875)
W·PC10	—	0.181	—	—	—	0.323	0.390	0.365
	—	(0.139)	—	—	—	(0.548)	(0.555)	(0.565)
W·PC11	—	0.055	—	—	—	0.631	0.607	0.626
	—	(0.152)	—	—	—	(0.712)	(0.719)	(0.717)
W·PC12	—	−0.087	—	—	—	−1.549*	−1.603*	−1.570*
	—	(0.138)	—	—	—	(0.687)	(0.700)	(0.698)
W·PC13	—	0.12	—	—	—	0.237	0.263	0.229
	—	(0.13)	—	—	—	(0.650)	(0.665)	(0.660)
W·PC14	—	0.407	—	—	—	3.906**	4.086**	3.941**

	—	(0.232)	—	—	—	(1.298)	(1.304)	(1.335)
R^2	0.586	0.697	—	—	—	—	—	—
Adj. R^2	0.551	0.642	—	—	—	—	—	—
Num. obs.	196	196	196	196	196	196	196	196
Parameters	—	—	18	18	19	34	34	35
σ^2	—	—	0.367	0.354	0.321	0.255	0.256	0.255
Pseudo R^2	—	—	0.624	0.637	0.663	0.743	0.742	0.743
Wald test	—	—	199.452***	264.933	12.82**	2.264	0.736	0.137
LogLik(Linear)	—	—	−191.197	−191.197	−191.197	−144.996	−144.996	−191.197
LogLik(Spatial)	—	—	−181.806	−178.398	−170.986	−144.513	−144.845	−144.451
AIC (Linear)	416.393	416.393	416.393	416.393	416.393	355.992	355.992	416.393
AIC (Spatial)	—	355.992	399.613	392.796	379.973	357.026	357.690	358.902
LR test	—	—	18.781***	25.598***	40.421***	0.967	0.302	93.492***

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 9. Spatial regression results for region B

Dependent variable: deforestation								
Independent variable	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS
(Intercept)	0.000	0.122	0.023	0.210	0.074	0.113	0.112	0.112
	(0.067)	(0.126)	(0.061)	(0.391)	(0.189)	(0.113)	(0.129)	(0.112)
Coca crops	0.235**	0.249**	0.263***	0.238**	0.240**	0.249***	0.249***	0.249***
	(0.084)	(0.081)	(0.075)	(0.082)	(0.079)	(0.072)	(0.072)	(0.071)
PC1	0.198	−0.3	0.071	0.042	0.049	−0.278	−0.288	−0.277
	(0.144)	(0.215)	(0.129)	(0.160)	(0.152)	(0.192)	(0.194)	(0.154)
PC2	0.150	0.203*	0.139	0.128	0.119	0.195*	0.201*	0.195*
	(0.086)	(0.09)	(0.077)	(0.084)	(0.081)	(0.081)	(0.082)	(0.077)
PC3	−0.395***	−0.192	−0.299**	−0.363**	−0.342**	−0.201	−0.194	−0.201
	(0.108)	(0.122)	(0.098)	(0.117)	(0.111)	(0.109)	(0.109)	(0.106)
PC4	−0.103	−0.236*	−0.143	−0.174*	−0.173*	−0.233**	−0.230**	−0.232**
	(0.093)	(0.094)	(0.083)	(0.087)	(0.085)	(0.084)	(0.085)	(0.079)
PC5	−0.355**	−0.12	−0.290**	−0.369**	−0.346**	−0.122	−0.117	−0.121
	(0.122)	(0.131)	(0.110)	(0.120)	(0.116)	(0.117)	(0.118)	(0.117)
PC6	−0.266*	−0.43***	−0.179	−0.247*	−0.207	−0.396***	−0.408***	−0.394***
	(0.110)	(0.125)	(0.100)	(0.117)	(0.112)	(0.112)	(0.114)	(0.110)
PC7	0.271*	0.407***	0.285**	0.307**	0.308**	0.390***	0.399***	0.389***

	(0.123)	(0.114)	(0.110)	(0.112)	(0.111)	(0.102)	(0.103)	(0.085)
PC8	−0.324**	−0.233*	−0.266*	−0.272*	−0.264*	−0.210*	−0.227*	−0.210*
	(0.121)	(0.111)	(0.109)	(0.111)	(0.109)	(0.099)	(0.100)	(0.085)
PC9	−0.155	−0.006	−0.168	−0.188	−0.195	−0.015	−0.002	−0.014
	(0.127)	(0.125)	(0.114)	(0.114)	(0.113)	(0.112)	(0.113)	(0.113)
PC10	−0.179	−0.009	−0.138	−0.117	−0.111	−0.003	−0.005	−0.002
	(0.097)	(0.096)	(0.087)	(0.091)	(0.089)	(0.086)	(0.087)	(0.086)
PC11	0.127	0.175	0.137	0.169	0.167	0.170	0.166	0.168
	(0.107)	(0.109)	(0.095)	(0.101)	(0.099)	(0.097)	(0.098)	(0.097)
PC12	−0.061	−0.095	0.016	0.030	0.049	−0.079	−0.095	−0.080
	(0.125)	(0.13)	(0.112)	(0.115)	(0.114)	(0.116)	(0.118)	(0.118)
PC13	0.255**	0.226**	0.190*	0.214*	0.194*	0.218**	0.226**	0.218***
	(0.086)	(0.086)	(0.077)	(0.085)	(0.083)	(0.077)	(0.077)	(0.058)
PC14	−0.132	−0.203	−0.156	−0.194	−0.186	−0.189	−0.191	−0.187*
	(0.135)	(0.129)	(0.121)	(0.126)	(0.124)	(0.115)	(0.116)	(0.088)
$\mathbf{W}\cdot y$	—	—	0.814***	—	0.653**	0.531**	—	0.517*
	—	—	(0.093)	—	(0.215)	(0.194)	—	(0.258)
$\mathbf{W}\cdot u$	—	—	—	0.845***	0.665**	—	0.363	0.066
	—	—	—	(0.083)	(0.222)	—	(0.250)	(0.94)
W·coca_crops	—	2.372*	—	—	—	2.287**	2.349**	2.288**
	—	(0.931)	—	—	—	(0.832)	(0.861)	(0.823)
W·PC1	—	−2.266*	—	—	—	−2.572**	−2.350**	−2.587***
	—	(0.076)	—	—	—	(0.786)	(0.805)	(0.655)
W·PC2	—	2.543*	—	—	—	2.546**	2.623**	2.566
	—	(1.077)	—	—	—	(0.962)	(0.983)	(0.97)
W·PC3	—	0.877	—	—	—	1.178	0.880	1.174
	—	(0.826)	—	—	—	(0.745)	(0.771)	(0.748)
W·PC4	—	1.672	—	—	—	1.695	1.691	1.699
	—	(1.332)	—	—	—	(1.190)	(1.213)	(1.138)
W·PC5	—	4.534***	—	—	—	4.798***	4.563***	4.810***
	—	(1.343)	—	—	—	(1.200)	(1.212)	(0.630)
W·PC6	—	−1.434	—	—	—	−1.374*	−1.451*	−1.379*
	—	(0.736)	—	—	—	(0.657)	(0.695)	(0.653)
W·PC7	—	−0.967	—	—	—	−1.503	−1.116	−1.523***
	—	(1.173)	—	—	—	(1.055)	(1.083)	(0.274)
W·PC8	—	−2.592*	—	—	—	−1.976	−2.417*	−1.957**
	—	(0.146)	—	—	—	(1.036)	(1.071)	(0.654)
W·PC9	—	4.71*	—	—	—	4.794**	4.812**	4.825***
	—	(1.892)	—	—	—	(1.689)	(1.712)	(0.161)

W·PC10	—	−0.135	—	—	—	0.252	−0.064	0.260
	—	(1.065)	—	—	—	(0.956)	(0.972)	(1.005)
W·PC11	—	-3.071^{*}	—	—	—	-3.137^{**}	-3.093^{**}	−3.150
	—	(1.293)	—	—	—	(1.154)	(1.170)	(1.15)
W·PC12	—	-5.091^{**}	—	—	—	-5.130^{***}	-5.109^{***}	-5.147^{***}
	—	(1.671)	—	—	—	(1.492)	(1.497)	(0.523)
W·PC13	—	4.141^{***}	—	—	—	3.906^{***}	4.010^{***}	3.898^{***}
	—	(1.043)	—	—	—	(0.933)	(0.954)	(0.868)
W·PC14	—	2.154	—	—	—	2.520	2.268	2.543^{**}
	—	(1.871)	—	—	—	(1.672)	(1.700)	(0.978)
R^2	0.291	0.568	—	—	—	—	—	—
Adj. R^2	0.224	0.477	—	—	—	—	—	—
Num. obs.	173	173	173	173	173	173	173	173
Parameters	—	—	18	18	19	33	33	34
σ^2	—	—	0.621	0.621	0.602	0.417	0.426	0.417
Pseudo R^2	—	—	0.356	0.354	0.375	0.577	0.570	0.577
Wald test	—	—	76.479^{***}	102.800^{***}	8.943^{*}	7.454^{**}	2.105	0.005
Log.Lik(Linear)	—	—	−215.188	−215.188	−215.188	−172.343	−172.343	−215.188
Log Lik(Spatial)	—	—	−206.916	−207.164	−204.338	−170.619	−171.932	−170.611
AIC (Linear)	464.376	464.376	464.376	464.376	464.376	408.687	408.687	464.376
AIC (Spatial)	—	408.687	449.832	450.329	446.677	407.239	409.863	409.221
LR test	—	—	16.543^{***}	16.047^{***}	21.699^{***}	3.448	0.823	89.155^{***}

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 10. Spatial regression results for region C

Dependent variable: deforestation								
Independent variable	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS
(Intercept)	0.000	−0.008	0.050	0.285	0.079	−0.003	0.220	0.126
	(0.031)	(0.048)	(0.029)	(0.682)	(0.660)	(0.045)	(0.681)	(0.319)
Coca crops	0.324^{***}	0.315^{***}	0.265^{***}	0.321^{***}	0.289^{***}	0.309^{***}	0.303^{***}	0.299^{***}
	(0.034)	(0.041)	(0.033)	(0.036)	(0.035)	(0.039)	(0.039)	(0.038)
PC1	0.249^{***}	0.328^{***}	0.224^{***}	0.265^{***}	0.231^{***}	0.326^{***}	0.365^{***}	0.355^{***}
	(0.061)	(0.077)	(0.058)	(0.061)	(0.060)	(0.073)	(0.074)	(0.070)
PC2	-0.210^{***}	-0.197^{***}	-0.165^{***}	-0.181^{***}	-0.167^{***}	-0.183^{***}	-0.180^{***}	-0.167^{***}

	(0.047)	(0.048)	(0.045)	(0.045)	(0.044)	(0.045)	(0.045)	(0.044)
PC3	−0.204***	−0.277***	−0.172***	−0.221***	−0.210***	−0.267***	−0.255***	−0.250
	(0.042)	(0.051)	(0.040)	(0.044)	(0.042)	(0.048)	(0.049)	0.048
PC4	−0.079	−0.073	−0.057	−0.045	−0.046	−0.061	−0.057	−0.049
	(0.045)	(0.046)	(0.043)	(0.043)	(0.042)	(0.043)	(0.043)	(0.040)
PC5	−0.152**	−0.193**	−0.117*	−0.148*	−0.155**	−0.201**	−0.196**	−0.202***
	(0.058)	(0.065)	(0.055)	(0.058)	(0.056)	(0.062)	(0.061)	(0.058)
PC6	−0.122**	0.019	−0.010	−0.024	0.010	0.007	0.016	0.008
	(0.042)	(0.063)	(0.041)	(0.049)	(0.045)	(0.059)	(0.060)	0.059
PC7	0.108*	0.038	0.065	0.082	0.067	0.046	0.036	0.042
	(0.045)	(0.049)	(0.043)	(0.045)	(0.044)	(0.047)	(0.047)	(0.026)
PC8	−0.066	−0.051	−0.035	−0.071	−0.060	−0.046	−0.038	−0.035
	(0.047)	(0.05)	(0.045)	(0.046)	(0.044)	(0.047)	(0.047)	(0.032)
PC9	−0.068*	−0.064	−0.097**	−0.082*	−0.082**	−0.071*	−0.086*	−0.089**
	(0.034)	(0.035)	(0.033)	(0.033)	(0.032)	(0.033)	(0.034)	(0.032)
PC10	0.059	0.005	0.051	0.040	0.042	0.012	0.016	0.017
	(0.038)	(0.046)	(0.036)	(0.038)	(0.037)	(0.044)	(0.044)	(0.031)
PC11	−0.158***	−0.108**	−0.125***	−0.125***	−0.116**	−0.110**	−0.109**	−0.109**
	(0.036)	(0.04)	(0.035)	(0.036)	(0.035)	(0.038)	(0.038)	(0.037)
PC12	−0.029	−0.006	−0.021	−0.013	0.003	−0.000	−0.004	0.000
	(0.036)	(0.04)	(0.04)	(0.035)	(0.036)	(0.042)	(0.038)	(0.038)
PC13	0.090*	0.132***	0.125***	0.118***	0.130***	0.133***	0.137***	0.136***
	(0.035)	(0.039)	(0.033)	(0.034)	(0.034)	(0.037)	(0.037)	(0.034)
PC14	0.030	0.029	0.043	0.038	0.042	0.032	0.031	0.033
	(0.043)	(0.045)	(0.041)	(0.041)	(0.040)	(0.042)	(0.042)	(0.034)
$\mathbf{W} \cdot y$	—	—	0.780***	—	0.877***	0.908***	—	0.870***
	—	—	(0.075)	—	(0.095)	(0.055)	—	(0.104)
$\mathbf{W} \cdot u$	—	—	—	0.958***	0.948***	—	0.959***	0.944***
	—	—	—	(0.028)	(0.048)	—	(0.027)	(0.058)
W·coca_-crops	—	−0.084	—	—	—	−0.314	−0.158	−0.386
	—	0.291	—	—	—	(0.275)	(0.368)	(0.357)
W·PC1	—	0.706	—	—	—	0.292	1.063	0.699
	—	(0.839)	—	—	—	(0.793)	(0.909)	(0.864)
W·PC2	—	−3.239***	—	—	—	−2.598***	−3.780***	−3.162***
	—	(0.826)	—	—	—	(0.781)	(0.909)	(0.870)
W·PC3	—	−0.304	—	—	—	−0.176	−0.156	−0.021
	—	(0.413)	—	—	—	(0.390)	(0.462)	0.449
W·PC4	—	−5.16***	—	—	—	−3.766***	−4.312***	−3.713***

	—	(0.872)	—	—	—	(0.824)	(0.927)	(0.865)
W·PC5	—	−0.521	—	—	—	−0.249	0.268	0.417
	—	(0.754)	—	—	—	(0.712)	(0.813)	(0.693)
W·PC6	—	0.266	—	—	—	0.555	0.016	0.317
	—	(0.375)	—	—	—	(0.355)	(0.441)	0.468
W·PC7	—	-2.258^{***}	—	—	—	-1.916^{***}	-1.577^{**}	-1.515^{***}
	—	(0.538)	—	—	—	(0.508)	(0.608)	(0.333)
W·PC8	—	−0.548	—	—	—	−0.631	−0.528	−0.597
	—	(0.561)	—	—	—	(0.530)	(0.615)	(0.520)
W·PC9	—	−0.497	—	—	—	−0.346	−0.208	−0.077
	—	(0.499)	—	—	—	(0.471)	(0.526)	(0.375)
W·PC10	—	1.781^{***}	—	—	—	1.359^{***}	1.331^{**}	1.062^{**}
	—	(0.379)	—	—	—	(0.358)	(0.438)	(0.410)
W·PC11	—	−0.923	—	—	—	−0.594	−0.893	−0.603
	—	(0.486)	—	—	—	(0.459)	(0.515)	0.514
W·PC12	—	-2.053^{***}	—	—	—	-1.867^{***}	-1.771^{***}	-1.646^{***}
	—	(0.491)	—	—	—	(0.463)	(0.518)	(0.231)
W·PC13	—	-0.998^{*}	—	—	—	-0.902^{*}	-1.063^{*}	-0.970^{*}
	—	(0.457)	—	—	—	(0.431)	(0.494)	(0.474)
W·PC14	—	0.418	—	—	—	0.410	0.219	0.263
	—	(0.497)	—	—	—	(0.469)	(0.545)	(0.517)
R^2	0.441	0.809	—	—	—	—	—	
Adj. R^2	0.427	0.484	—	—	—	—	—	
Num. obs.	613	613	613	613	613	613	613	613
Parameters	—	—	18	18	19	33	33	34
σ^2	—	—	0.519	0.498	0.468	0.460	0.454	0.434
Pseudo R^2	—	—	0.476	0.492	0.517	0.533	0.537	0.552
Wald test	—	—	108.959^{***}	1200.99^{***}	387.523^{***}	272.391^{***}	1247.482^{***}	264.525^{***}
LogLik (Linear)	—	—	−691.110	−691.110	−691.110	−651.028	−651.028	−691.110
LogLik (Spatial)	—	—	−671.372	−661.861	−646.466	−635.954	−633.612	−622.988
AIC (Linear)	1416.219	1416.219	1416.219	1416.219	1416.219	1366.056	1366.056	1416.219
AIC (Spatial)	—	1366.06	1378.744	1359.723	1330.933	1337.908	1333.223	1313.977
LR test	—	—	39.476^{***}	58.496^{***}	89.287^{***}	30.148^{***}	34.833^{***}	136.242^{***}

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 11. Spatial regression results for region D

Independent variable	OLS	SLX	SAR	SEM	SAC	SDM	SDEM	GNS
	Dependent variable: deforestation							
(Intercept)	0.000	−0.063	0.000	−0.056	−0.012	−0.056	−0.076	−0.069
	(0.086)	(0.079)	(0.070)	(0.241)	(0.114)	(0.069)	(0.056)	(0.046)
Coca crops	0.101	−0.063	0.050	0.046	0.054	−0.038	−0.083	−0.061
	(0.111)	(0.11)	(0.090)	(0.095)	(0.094)	(0.096)	(0.099)	(0.097)
PC1	0.829^{***}	1.169^{***}	0.857^{***}	1.035^{***}	0.942^{***}	1.154^{***}	1.277^{***}	1.314^{***}
	(0.210)	(0.226)	(0.172)	(0.190)	(0.195)	(0.198)	(0.208)	(0.224)
PC5	-0.843^{***}	-0.655^{**}	-0.679^{***}	-0.756^{***}	-0.740^{***}	-0.634^{**}	-0.651^{**}	-0.620^{**}
	(0.167)	(0.232)	(0.140)	(0.177)	(0.163)	(0.203)	(0.215)	(0.210)
PC7	0.497^{***}	0.573^{***}	0.423^{***}	0.468^{***}	0.449^{***}	0.554^{***}	0.596^{***}	0.584^{***}
	(0.126)	(0.129)	(0.104)	(0.102)	(0.105)	(0.113)	(0.115)	(0.111)
PC9	-0.478^{**}	-0.441^{**}	-0.397^{**}	-0.385^{**}	-0.399^{**}	-0.413^{**}	-0.438^{**}	-0.397^{**}
	(0.161)	(0.152)	(0.132)	(0.133)	(0.134)	(0.133)	(0.138)	(0.133)
PC12	0.369^{*}	0.371^{*}	0.360^{**}	0.394^{**}	0.386^{**}	0.374^{**}	0.387^{**}	0.396^{**}
	(0.153)	(0.147)	(0.125)	(0.123)	(0.125)	(0.129)	(0.132)	(0.126)
PC13	-0.512^{**}	-0.433^{**}	-0.374^{**}	-0.390^{**}	-0.388^{**}	-0.409^{**}	-0.441^{***}	-0.416^{***}
	(0.152)	(0.147)	(0.125)	(0.125)	(0.127)	(0.128)	(0.132)	(0.126)
$\mathbf{W} \cdot y$	—	—	0.598^{***}	—	0.426	0.361^{*}	—	0.439^{*}
	—	—	(0.106)	—	(0.240)	(0.157)	—	(0.182)
$\mathbf{W} \cdot u$	—	—	—	0.716^{***}	0.406	—	−0.306	−0.532
	—	—	—	(0.093)	(0.305)	—	(0.224)	(0.325)
W·coca_crops	—	−0.358	—	—	—	−0.310	−0.381	−0.305
	—	(0.376)	—	—	—	(0.330)	(0.305)	(0.269)
W·PC1	—	-1.872^{**}	—	—	—	-1.699^{**}	-2.322^{***}	-2.236^{***}
	—	(0.676)	—	—	—	(0.592)	(0.565)	(0.584)
W·PC5	—	0.859	—	—	—	0.876	1.038^{*}	1.089^{**}
	—	(0.533)	—	—	—	(0.468)	(0.447)	(0.406)
W·PC7	—	−0.015	—	—	—	−0.134	−0.185	−0.390
	—	(0.379)	—	—	—	(0.346)	(0.309)	(0.303)
W·PC9	—	−0.294	—	—	—	−0.062	−0.275	0.013
	—	(0.427)	—	—	—	(0.389)	(0.354)	(0.307)
W·PC12	—	−0.632	—	—	—	−0.515	-0.785^{*}	-0.657^{*}

	—	(0.442)	—	—	—	(0.387)	(0.356)	(0.320)
W·PC13	—	−0.489	—	—	—	−0.181	−0.526	−0.152
	—	(0.489)	—	—	—	(0.446)	(0.410)	(0.377)
R^2	0.481	0.632	—	—	—	—	—	
Adj. R^2	0.429	0.551	—	—	—	—	—	
Num. obs.	78	78	78	78	78	78	78	78
Parameters	—	—	10	10	11	17	17	18
σ^2	—	—	0.376	0.362	0.372	0.343	0.356	0.319
Pseudo R^2	—	—	0.593	0.593	0.601	0.645	0.635	0.655
Wald test	—	—	31.816^{***}	59.517^{***}	1.776	5.312^{*}	1.869	2.678
LogLik (Linear)	—	—	−84.579	−84.579	−84.579	−71.146	−71.146	−84.579
LogLik (Spatial)	—	—	−75.106	−75.147	−74.298	−69.779	−70.875	−68.710
AIC (Linear)	187.158	187.158	187.158	187.158	187.158	174.293	174.293	187.158
AIC (Spatial)	—	174.29	170.211	170.294	170.596	173.557	175.750	173.420
LR test	—	—	18.947^{***}	18.865^{***}	20.563^{***}	2.736	0.542	31.738^{***}

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Chapter 2

An accuracy assessment of three forest cover data bases of Colombia

Abstract

Accurate forest assessment is essential to detect and tackle deforestation especially in emerging economies. Unfortunately, creating forest cover maps from field surveys or raw remote sensing data is a prohibitive task for most researchers, practitioners and other individual users. An alternative is provided by existing global data sets on forest cover. However, these data sets are not characterized by the same level of accuracy in all countries and, in some cases, require specific calibration. Thus, the best way to approach forest cover estimation in a specific country is to assess the accuracy of different data sets against ground observations for the different land types. In this study, we compared three data sources available for forest monitoring in Colombia: the European Space Agency (ESA), the Institute for Hydrology, Meteorology and Environmental Studies (IDEAM), and the Global Forest Change Data (GFCD) from the University of Maryland. We considered thirty-three ecoregions, and for each of them we determined the optimal forest threshold for GFCD and then compared its accuracy with that of ESA and IDEAM, on the basis of an independent verification scheme exploiting Google Earth imagery and landscape photographs in a representative sample of pertinent sites. IDEAM and GFCD proved to be quite accurate in most cases, and each of them turned out to be the best forest map in about half of the ecoregions. GFCD's optimal threshold was found to be equal to 90% in almost all of the ecoregions in which it turned out to be the best performing data set. In two ecoregions, none of the data sets was able to accurately discriminate forests from non-forest area. ESA failed to estimate accurately the real forest cover in Colombia.

1 Introduction

An effective and precise forest monitoring system is essential to address forest loss and deforestation rates in the world. Early detection of forest loss is a key tool to reduce emissions from deforestation, preserve habitats, and reduce the alarming acceleration of climate change (Hansen et al., 2013; Tropek et al., 2014; Hansen et al., 2014). However, creating a reliable tool for forest monitoring is not an easy task, as complex remote sensing algorithms are still making progress to screen forests regularly (some examples of the different algorithms and techniques, are: Guo et al. (2022); Dutrieux et al. (2015); Hansen and DeFries (2004); Potapov et al. (2015); Ørka et al. (2022); Panta et al. (2008); Bajocco et al. (2012); Song et al. (2014, 2015). In addition, using raw satellite imagery data represents a serious obstacle for many researchers and practitioners, as creating forest maps from them needs specific expertise in remote sensing and specialized software (Fraser et al., 2005).

Databases such as the Global Forest Change Data (GFCD), developed by the University of Maryland (Hansen et al., 2013), or the global land cover maps from the European Space Agency (ESA), elaborated under the global climate initiative[1] (ESA, 2017), both represent a viable solution for researchers, but still require technical improvements and conceptual implementation to become more accurate. Moreover, the databases are released yearly, so real-time deforestation is not yet available for most of the stakeholders (Pettorelli et al., 2018). There are real-time detection systems of wildfires like the FIRMS initiative of NASA, but other causes of deforestation are not instantly detected.

GFCD has commonly been criticized for over-quantifying forest cover in different parts of the world. The problem partly depends on the fact that this database does not directly provide a forest cover map, and this needs to be retrieved from the percent tree cover layer of GFCD, by applying a suitable threshold. Inappropriate choices of this threshold can affect empirical results, undermining the accuracy of forest cover estimation. It has been demonstrated that the calibration of forest cover rate depends on the ecological zone and should be designed, based on the researcher's knowledge of forest cover in the area. However, a canopy cover of 30% is often arbitrarily chosen in most studies, determining forest cover overestimation. In fact, as investigated by Lwin et al. (2019), McRoberts et al. (2016) and Sannier et al. (2016) the optimal threshold for GFCD data should be higher for denser forest canopy regions, especially in tropical moist forests. Lwin et al. (2019), in particular, show that the accuracy of the forest

[1]The Climate Change Initiative's Land Cover data set is generally referred to as CCI-LC

cover in GFCD is highly dependent on the selection of the appropriate forest cover threshold. They consider the case of Myanmar, where the main forest cover type is tropical rainforest, and conclude that the threshold should be set according to the ratio of tropical rainforest and total area. However, for different ecological zones, different thresholds are needed to achieve higher accuracy. In essence, in places where tropical rainforest is dominant, a higher threshold is required. In the case of Myanmar, the optimal threshold at a national level is determined to be 40%, for Gabon 70% (Sannier et al., 2016) and in Santa Catalina, Brazil 95% (McRoberts et al., 2016).

Differently from GFCD, ESA is a world land cover data set that provides information on the different land types like shrublands, croplands, grassland, wetlands, settlements, and twelve different types of forests. In total, ESA uses twenty-three different categories of land types, and its objective is not to classify forest and non-forest pixels but to classify each pixel into one of these categories.

The present study contributes to the current literature on forest cover accuracy considering the relevant case of Colombia. Colombia is a crucial country in terms of forest conservation, because of the extension of its tropical forests (Keenan et al., 2015). According to the Food and Agriculture Organization (FAO), 54.5% of the country's territory is covered by forest, and of this extension, 14.1% is classified as primary forest, the most biodiverse and carbon-dense form of forest. Colombia is the second most biodiverse country in the world and the most biodiverse per square km. It is home to some 10% of the world's species and many of them are endemic (World Wildlife Fund, 2017). However, deforestation is a major problem in the country. According to Global Forest Watch (2021), Colombia lost 6.0% of its forests between 2000 and 2021, which is equivalent to releasing 2.76Gt of CO_2 emissions.

This study aims to determine the optimal threshold for the GFCD in the different ecological regions of Colombia and to compare its accuracy in producing forest cover maps with that of ESA and a local database, developed by the Institute for Hydrology, Meteorology and Environmental Studies (hereafter IDEAM). IDEAM has been widely used by researchers in the field of deforestation studies in Colombia. Its main advantage is the field validation protocol carried out by its experts, which allows, for instance, to exclude areas that grow commercial crops from the forest cover estimates. It has also the advantage that the vector files are easy to use as each pixel only presents three possible values: forest, non-forest, and no information. However, it is recognized to underestimate smaller forest patches or sparse forest. In a recent news release, IDEAM declared that they will be collaborating closely with ESA, which holds multiple satellite data sets and would help IDEAM to monitor deforestation more closely (Colombia Reports, 2018). Therefore, the specific objectives of our paper are: (i) to investigate the accuracy of forest cover maps obtained from GFCD using different tree cover thresholds across different ecological zones in Colombia; (ii) to compare the accuracy of GFCD, after optimization of the threshold, with that of ESA and IDEAM; (iii) to provide users with guidelines for choosing the most reliable data set for forest mapping, both, for particular ecological zones and Colombia as a whole.

The paper is organized as follows: in Section 2, we describe the data sources (GFCD, ESA, and IDEAM) and their distinctive characteristics; in Section 3, we illustrate the methodology used to assess accuracy; in Section 4, we present the empirical results of this study; in Section 5, we summarize the main findings, discussing advantages and disadvantages of each database considered here; and finally, conclusions are given in Section 6.

2 Data

The three databases considered in this study represent a valuable tool for researchers involved in forest change understanding and quantification and conservation planning. Nevertheless, they have distinct characteristics, purposes, and coverage. These peculiarities can lead to marked differences in the results when these databases are used to produce forest cover maps. Understanding their advantages or limitations, as well as uncertainties and inaccuracies in the estimates they produce within particular forest types and different canopy densities - which allows a basic quantification of ecosystem services - is vital to ensure their appropriate use for specific applications and local contexts (Blackman and Yuan, 2020). In this section, we briefly describe the three databases, as well as the area under study, for which the accuracy of forest cover maps obtained from the three databases is assessed.

2.1 Global Forest Change Database

The GFCD, developed by the University of Maryland (Hansen et al., 2013) and freely available to download from a dedicated data portal[2], is one of the most widely used data sets for forest change monitoring in the world, being the data source of the initiative Global Forest Watch. GFCD's goal is to quantify forest cover change. It is yearly updated and has been collecting forest data since 2000. In practice, GFCD inspects global Landsat

[2]The GFCD database can be downloaded at https://glad.earthengine.app/view/global-forest-change.

data at a 30-meter spatial resolution to identify forest extent worldwide. This resolution should be enough in most cases to detect deforested areas. However, when the areas of deforestation are fragmented into small patches, GFCD may underestimate total loss in such contexts. Another drawback of GFCD is that the forest definition used encompasses all vegetation taller than 5 meters in height. Under this definition, plantations such as oil palm, soybeans, tea, and monoculture crops or trees are classified as forests, though they cannot be considered natural forests and do not have their carbon or biodiversity value. In addition to this, information on forest and non-forest areas are not directly available. GFCD includes three layers: the first one estimating the percentage of tree cover in 2000 (0%–100%), the second including the annual forest cover loss (2000–2021), and the third reporting the forest cover gain. In practice, the year 2000 is taken as the baseline, and forest loss/gain is yearly computed with respect to this year[3]. Therefore, to produce forest cover maps, researchers must integrate forest cover knowledge from other sources. One possibility is to create a forest cover map from the percent tree cover layer of the GFCD, by choosing an appropriate threshold for tree cover percentage to discriminate between forest and non-forest areas. In this way, a forest cover map for the year 2000 is created, and from this, the forest cover map for a certain year is obtained by combining the annual loss layer and the annual gain layer between 2000 and that specific year. Since the definition of the threshold directly affects the areas of forest and non-forest, setting an appropriate threshold is crucial.

2.2 ESA Database

The ESA database is elaborated under the global climate initiative and is freely available for users[4]. It is also used worldwide and annually released, as GFCD. Information is collected since 1992, which makes ESA the richest database in terms of temporal coverage when compared to GFCD and IDEAM. Differently from GFCD, ESA uses Sentinel as its imagery source, instead of Landsat, and according to Astola et al. (2019), Sentinel-2 is slightly better than Lansat 8 in forest variable prediction. However, ESA has a lower resolution (300 m) than GFCD.

The main focus of ESA is on land cover classification, rather than forest cover monitoring. For this reason, it does not allow assigning the forest cover threshold like in GFCD but has different predefined options of canopy cover thresholds, depending on the type of land one wishes to study and the canopy associated with that land classification. In

[3]Gains in forest cover, however, were only monitored until 2012.

[4]The ESA database can be downloaded at http://maps.elie.ucl.ac.be/CCI/viewer/download.php

essence, the goal of ESA is to classify the type of land into different categories and ESA's bitmap file has 22 categories resulting from a Random Forest Classification process. Thus, its scope is much wider than that of GFCD, but this probably represents a disadvantage when using it as a forest cover monitoring database, and not as an approximation of the different land use extensions. To obtain forest cover maps, one has to reassign new values to the bitmap file and lower the number of categories. We consider seven categories (agriculture, forest, grassland, wetland, settlement, shrubland, and other land types), we mask all values different from forest and allow only the following values: 50, 60, 62, 70, 72, 80, 82, 90, 100 and 160. The table with ESA's original values and our converted values are shown in 26. The resulting forest map is shown in Figure 21, panel (d), for the area surrounding a specific location in the Caqueta moist forest in Amazonia. Clearly, ESA provides a broad approximation of forested areas, without precision on the forest limits.

2.3 IDEAM's Database

IDEAM is a Colombian institution that collects and prepares complete information on forest cover in the country. This database is thus specific to Colombia and no other regions or countries are covered. From 1990 until 2012, IDEAM's data were published every 5 years. The methods and imagery sources were redefined in 2012, and since then data were released annually. Unfortunately, at least two years elapse between the date to which the data refer and the date of release. For example, in 2022 the latest available data are those relative to the year 2019. For this reason, in this study, we compare the accuracy of the three databases in estimating forest cover maps for the year 2019.

Notwithstanding the limited availability of years for the IDEAM data set, it is the most recurrent data set used in deforestation studies in Colombia (Armenteras et al., 2013, 2017; González-González et al., 2021; Rodríguez-de-Francisco et al., 2021). Its great advantage is that the information obtained from satellite and classification algorithms is debugged and analyzed by geographers who know the area and prepare information only for this country. In doing so, the IDEAM data set aims to comply with the guidelines of the United Nations Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC) and create a system to detect and tackle deforestation, also to identify the causes of deforestation to reduce the emissions caused by it and avoid forest degradation in compliance with the REDD+ objectives.

Unlike GFCD, IDEAM's forest cover threshold is fixed at 30%, without the possibility of modifying it (IDEAM, 2019). A further difference is that IDEAM requires a minimum of 10,000 m^2 of continuous forest to label an area extension as forest, thus excluding small forest patches. In contrast, GFCD sets no minimum area size, and for this reason, it detects small forest patches that IDEAM ignores. On the other hand, when the forest is continuous, depending on the choice of the threshold, GFCD's forest areas may be a subset of IDEAM's, resulting in type I and type II mismatching errors (Fergusson et al., 2020). A relevant feature of IDEAM, however, is that it excludes palm oil cultivation areas from forest extension, which is not the case in GFCD. A further advantage is that the vector file is user-friendly, with only 3-pixel values, one for the forest, another for non-forest, and one with missing information (see Figure 13 in Appendix 2A).

2.4 Study area

Colombia is located in the northwest of South America between latitudes $4°13'30''$S and $12°27'46''$N (Banco de Colombia, 2022). With a total area of 1,141,750 km^2, it is the fourth largest country in South America after Peru. Colombia borders the Pacific Ocean and Panama to the west, with the Caribbean Sea to the North, Venezuela, and Brazil to the East, and Peru and Ecuador to the South (FAO, 2015). As of 2022, Colombia has a population of 52 million people (Worldometer, 2022) and a cultivable area of 9% of the country (FAO, 2015). Colombia is the second most biodiverse country in the world, with 54,871 registered species. Sadly, it is also one of the countries that contribute the most to biodiversity destruction in the world, precisely because of their immense biodiversity stock (Rodrigues et al., 2014). In particular, the forest cover in the country decreased from 58.9% to 53.3% between 1990 and 2020, with a 0.2% net deforestation rate per year (World Bank, 2022). In general, the Latin American continent has one of the highest deforestation rates in the world (Da Ponte et al., 2015). There are 34 ecoregions in Colombia (Dinerstein et al., 2017) if we include the isles of San Andrés and Providencia, which are located in the Caribbean Sea near the Nicaraguan coast. Excluding these isles, the mainland holds 33 different ecoregions. The Caqueta moist forest is the largest ecoregion, followed by the Orinoquia lowlands. The Caqueta moist forest is situated in the transition zone between the Guiana and Amazon regions, which are very diverse in terms of flora and fauna. Despite being relatively intact, a large area of this ecoregion is dedicated to cattle ranching, which has led to the clearing of vast stretches of forest (One Earth, 2022). Grasslands, savannas, and shrublands are abundant in the lowlands of Orinoquia. As a result of its large reserves of natural gas and petroleum, this region is crucial to the country's economy (Gobierno de Colombia, 2020). The 33 mainland ecoregions will be used in Section 3 and 4 to compare the accuracy of the three different databases, according to the type of ecological zone under study.

3 Methodology

To determine the optimal threshold for GFCD and compare the accuracy of the different databases in classifying pixels as forest or non-forest, we use observations at selected sites as reference data. In practice, we selected a fixed number of random points from each ecoregion, collecting reference data for each of them. Then, we derive a confusion matrix between the forest cover map obtained from each database (and from each threshold for GFCD) and the reference data. Both forest cover maps and reference data pertain to the year 2019, the last year for which IDEAM data are available. Finally, we use the confusion matrix to compute various measures of accuracy.

For the GFCD database, we consider different thresholds, corresponding to ten-step percentages, from 10% to 90%. However, if we observe that 90% is close to the optimal threshold, but a higher percentage is needed, we also try higher thresholds, up to 98%. These cases are, however, rare. In Colombia, we find 95% to be the maximum value before the classification of forest exhibits a random behavior. On the other hand, for those areas in which desert zones are prevalent, we also consider a 1% threshold. Then, to obtain independent data on deforestation, we follow the operational process described in the Global Forest Change documentation of Google Earth Engine, specifically the section Quantifying Forest Change.

Stratified random sampling is used to select the sample points (Marchi et al., 2017; Ferrara et al., 2020; Bajocco et al., 2011; Conedera et al., 2011). A predefined number of points is randomly sampled in each ecoregion. We use 33 ecoregions as defined by Dinerstein et al. (2017) in the Colombian mainland, leaving the isles of Providencia and San Andrés out of the study because IDEAM does not cover these areas. Earlier studies point out that a minimum of 50 sample points for each land-cover category is needed for land classification (see, for instance, Myint et al., 2011). We opt for an equal allocation and use the same sample size, i.e. 60 points, for each ecoregion. Even if ecoregions have different extensions (see Figure 19), we believe them to be quite homogeneous internally with respect to forest cover and to show approximately the same variability (Hanssen et al., 2021; Song et al., 2015; Smiraglia et al., 2016; Masek et al., 2015; Salvati et al., 2017; Ferrara et al., 2017). If this assumption was true, the equal allocation would not differ much from the Neyman optimal allocation. We use QGIS to randomly sample the points.

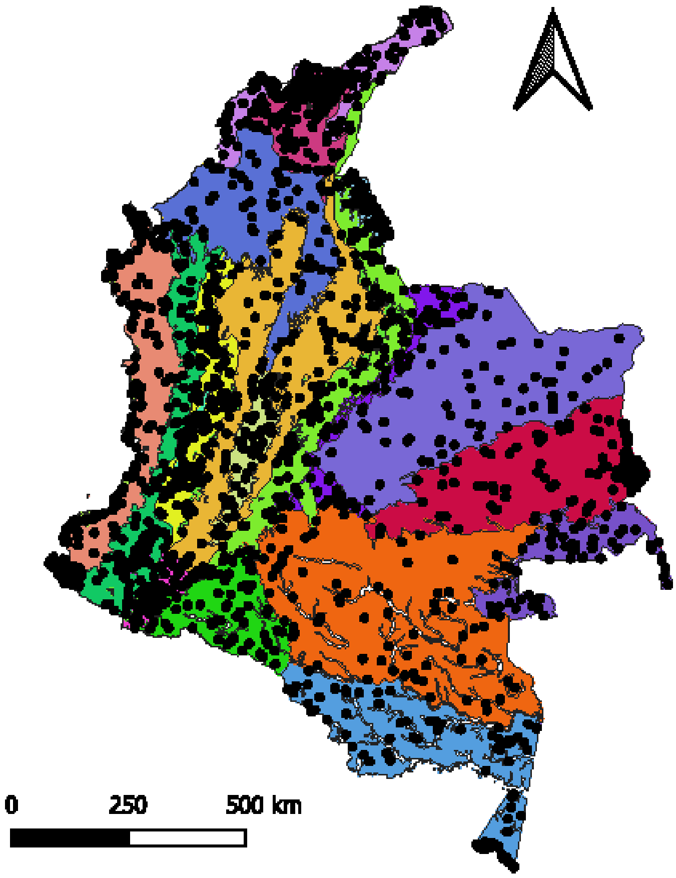


Figure 19. Stratified random sample points drawn in the 33 ecoregions as defined in Ecoregions 2017 (Dinerstein et al., 2017).

The next step consists in collecting reference data for the points at a ground level with the available imagery. For this purpose, we rely on Collect Earth, which is a program that is used for field-based inventories, in particular for cross-checking land classifications. We use the database points in the Google Earth Pro interface to compare against high-resolution imagery from Google Earth and Bing Maps as well as middle-quality images from Google Earth Engine, Landsat 8, and Landsat 7.

According to FAO (2000), forest is defined as <Land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10%, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use>. In agreement with this definition, we set grids of 0.5 ha or 70m × 70m, around the sample point coordinates, corresponding to an internal systematic grid of 49 points (7 × 7 points), as shown in Figure 20 b). Points are classified as forest if the crown cover of the trees occupies at least 5 out of the 49 points, which equals 10,2% of tree cover within the square. If a point does not meet this criterion, it is classified as non-forest.

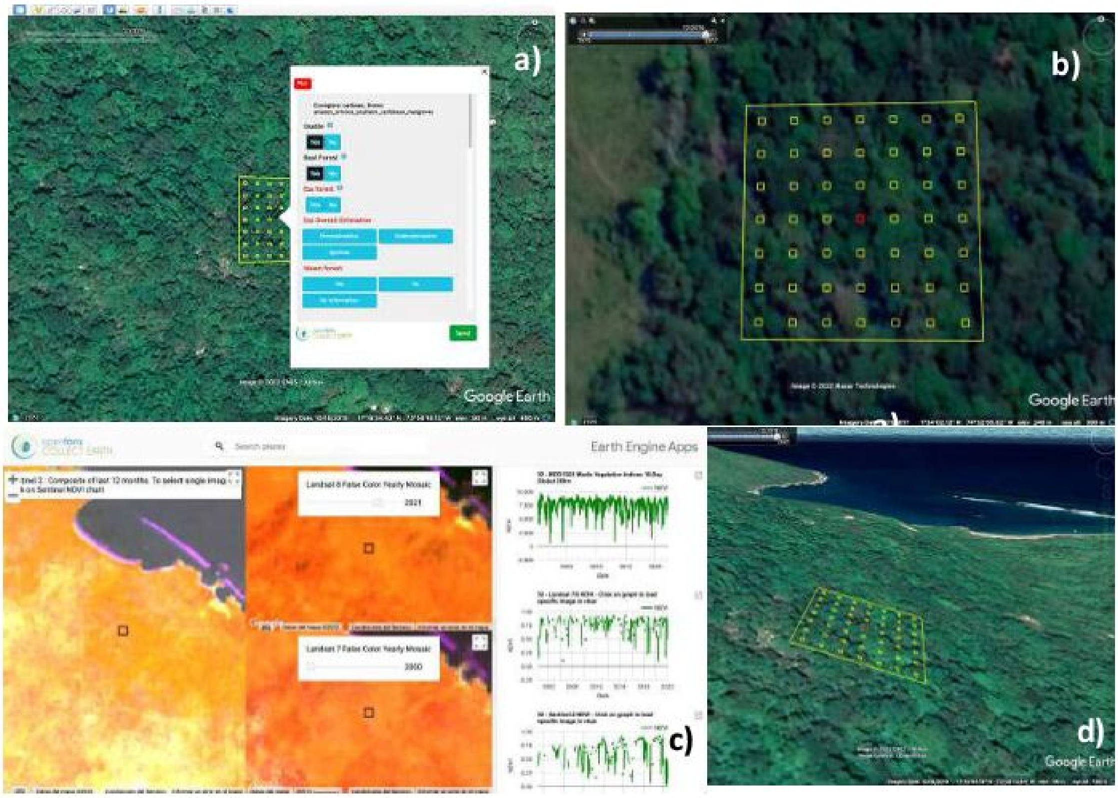


Figure 20. View of a specific sample point using Collect Earth software. **a)** Example of our customized Collect Earth survey opening at a sample point in Amazonia in Google Earth Pro Software. **b)** The 70m × 70m square grid in a forest point in Amazonia. **c)** Normalized difference vegetation index (NDVI) displayed with Sentinel 2, Landsat 7 and 8, as well as the NDVI calculation results with the 3 different satellites in the graphs at the right of the image. **d)** Sightseeing with a different angle available in Google Earth Engine, allowing to explore sample points and understand the terrain's slope and elevation.

After customizing the survey in Open Foris Collect, we transfer the *.cep* file to Collect Earth which works within Google Earth Engine. We compare the points with more than one satellite image, we use different high imagery sources synchronized with Google Earth Pro, like Bing maps, Google maps, and medium resolution imagery like Earth maps. Next, with the ground land analysis, defining each reference datum as forest or non-forest is a mechanical procedure consisting of counting points within the grid, according to the criterion explained above.

Using the reference data set, we generate confusion matrices between reference data and each forest cover map. In practice, for each ecoregion, we generate one confusion matrix for ESA, one for IDEAM and one for each of the thresholds considered for GFCD. These matrices are then used to compute different accuracy measures (Congalton and Green, 2019), namely:

- *Overall accuracy*, calculated by summing the number of correctly classified points and dividing by the total number of points.
- *User's accuracy*, calculated as the number of points that are correctly classified in the class forest (or non-forest) divided by the number of points classified in that class. This is a measure of true positives. It is also the complement of the *error of commission*[5] or type II error.
- *Producer's accuracy*, calculated as the number of points that are correctly classified in the class forest (or non-forest) divided by the number of points belonging to that class in the reference data. It is the complement of the *error of omission*[6] or type I error.
- *Cohen's Kappa Accuracy statistic* (Cohen, 2022), calculated as the difference between the actual agreement between the map and the reference data, and the agreement one would expect if the map were obtained through a random classification. The difference is then normalized to make the index vary between -1 and 1, with positive values meaning that the map performs better than the random classification.

These measures are used to determine the optimal data set (and the optimal threshold for GFCD) for each ecoregion, i.e. the one producing the most accurate forest map.

Besides the accuracy statistics (see results in Table 16), we collect further information of the surroundings of the sample points. In practice, we enlarge the map up to 10km × 8km to determine - based solely on observation - which data set better captures the forest extension. An example of this operation is available in Figure 21. Figure 21, panel (a) is an overview of an area where forest is highly fragmented in the Amazonian region. In panel (b), we overlay it with the IDEAM mask, then, in panel (c) with the GFCD mask, as obtained using the <best> threshold, which turned out to be equal to 90% for this specific ecoregion. Finally, in panel (d), we observe the same area with ESA's information. From the images, we conclude that the GFCD is overestimating the forest quantity, ESA is not accurately indicating the borders between forest and non-forest and is missing small forest patches, on the other hand, IDEAM is slightly under-quantifying forest extension.

[5]The error of commission is calculated as the number of observations incorrectly classified in a given class, divided by the total number of observations classified in that class.

[6]The error of omission is calculated as the number of observations belonging to a given class but classified in a different class, divided by the total number of observations belonging to that given class.

Thus, in this specific situation, we would choose IDEAM as the preferred data set. If all the layers fail to approach the real amount of forest for a given sample point, we attribute a <no optimal> classification label to that point. These qualitative results are available in Table 15 in Appendix 2C and are used to confirm the quantitative ones based on accuracy measures.

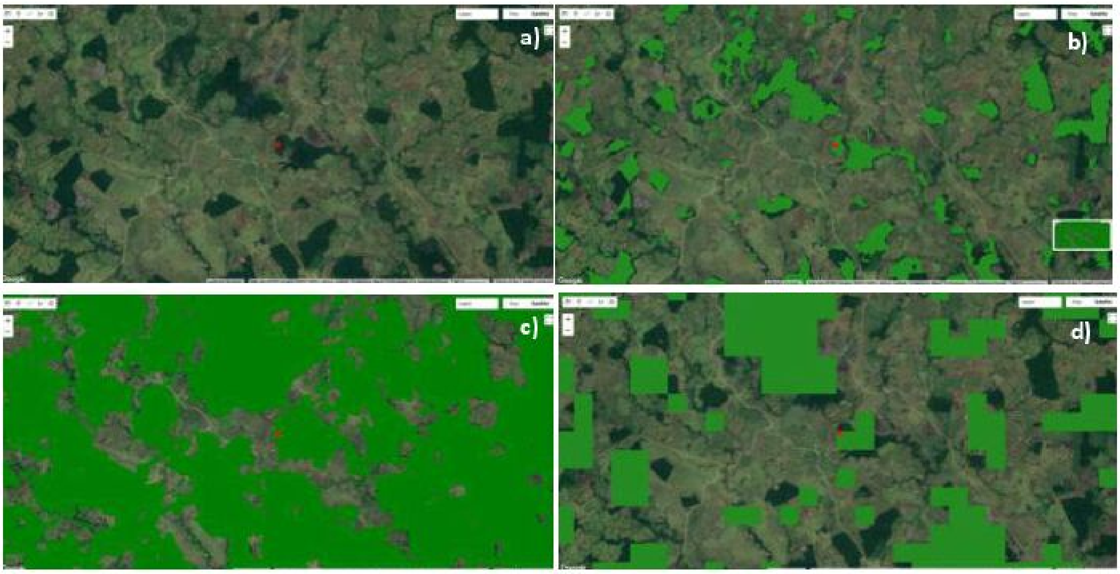


Figure 21. Visual information on the surroundings of a sample point. **a)** Overview of a point located in the Caqueta moist forest in Amazonia, coordinates (-74.54844705, 1.567342579) in EPSG 4326 projection. **b)** Same point as in *a)* with the IDEAM mask data set. **c)** Same point as in *a)* with the GFCD mask data set with a 90% threshold (95% does not yield a better result and non-forest points above 90% start to become random). **d)** The same point as in *a)* overlaid with the ESA mask data set.

4 Results

The main results of the analysis are summarized in Table 16, where, for each ecoregion, we report the accuracy measures for the three databases, as well as the mode and standard deviation of the optimal thresholds for GFCD over the sample points in the ecoregion. For the overall accuracy, we also provide significance levels (***p-value ≤ 0.001; **p-value ≤ 0.01; *p-value ≤ 0.05) of a one-sided binomial test to evaluate if the accuracy is better than the <no information rate>, which is taken to be equal to the largest class percentage in the reference data. The user's and producer's accuracy statistics are only computed for forest, as this is the main objective of this study.

The third column in Table 16 shows the effective sample size, n, for each ecoregion. In some of them, the effective sample size is smaller than 60, as cloud coverage made it impossible to classify a few points even with the different imagery sources at our disposal. In the Guaianan piedmont and lowland moist forest, which is a small region essentially surrounding the Orinoco river, we are only able to effectively observe 45 points, however, 93% of the points indicate that IDEAM is the best database to classify forest in this ecoregion, therefore, we do not hold indispensable to replace the unobserved points with others randomly drawn. The rest of the regions have more than 50 effectively observed points, and most of them have all 60 observed points.

The optimal threshold for GFCD is determined as the one producing the map with the best overall accuracy. In most cases, different thresholds produce the same overall accuracy, so a unique threshold is chosen based on qualitative information from visual observation of enlarged maps. In practice, the best threshold for each sample point is visually determined and the mode over the sample points in the ecoregion is taken as the optimal threshold. In doing so, we check that the threshold mode is equal to one of the thresholds that produce the maximal overall accuracy for that ecoregion. This turns out to be always the case for all the 33 ecoregions. The threshold mode is provided in Table 16, together with the standard deviation of the best thresholds over different sample points (their mean is given in Table 15 in Appendix 2C).

Based on the accuracy statistics and the observational results, we conclude that ESA's database is not completely suited to forest mapping in Colombia and produces, in most cases, maps that are less accurate than those obtained from the other two databases. We have satisfactory results for nearly all ecoregions. In two ecoregions, the Patia Valley Dry Forest and Sinu Valley Dry forest, we qualitatively observe a predominant amount of <no optimal> predicted areas surrounding the points (Appendix 2C), even though the overall accuracy in the former and latter region reaches 95% and 80%, respectively. However, the overall accuracy is not significantly better than the <no information rate> in both cases. This is due to the fact that both Patia and Sinu Valley are mostly characterized by non forest land, so that a classifier assigning all points to non-forest would easily classify most of them correctly, with an accuracy not significantly different from that obtained using one of the databases under consideration. When considering a wider observation of the map (10km × 8km), we notice that IDEAM underestimates forest cover (due to the presence of small woodlots) while GFCD overestimates it, even using a 90% threshold. Based on qualitative ground observations, the majority of the points fall into the <no optimal> category, thus none of the three databases can accurately predict forest or non-forest conditions in the two ecoregions. Researchers should be aware of the problems raising with use of these data sets in mapping forests, especially when estimating land-use accuracy.

Other ecoregions show an overall accuracy with unsatisfactory large p-values. However, in all cases, this is due to the imbalance between the number of forest and non-forest points in the reference data. For instance, in the Central American dry forests, all points observed are non-forest and they are correctly classified by all databases (with a 20% threshold for GFCD), determining an overall accuracy equal to 100% but undefined user's and producer's accuracy for forest, and Kappa statistic. On the other hand, in the Japura-Solimoes-Negro moist forests, only 2 reference points out of 60 are forest and IDEAM is the only database correctly identifying both forest and non-forest points, achieving 100% accuracy statistics. The other two databases are unable to categorise correctly these two non-forest points (not even selecting a 95% threshold for GFCD), as clearly shown by their Kappa value equal to 0. Qualitative inspection of these ecoregions does not reveal, however, particular problems in estimating forest extension.

In other cases, differences between the three databases are quite remarkable. For instance, in the Magdalena Valley dry forests, GFCD (80% threshold) has a significant overall accuracy of 88.3% and a Kappa value of 75.9%. The other two databases perform much worse for the same ecoregion. IDEAM in particular has an overall, user, and producer accuracy of respectively 65%, 100%, and 16%, and a Kappa statistic of 18.2%, indicating a substantial underestimation of forest cover and a limited classification capability in this specific case. A similar underestimation problem arises in Cauca Valley dry forests and, to a less extent, in the Apure-Villavicencio dry forests. All these areas of dry forests are characterized by a long dry season followed by a season of heavy rainfall. In this context, dry forests receiving more rain tend to be taller and have more evergreen species, while those located in dry areas generally have a lower, less dense canopy and more species that shed their leaves during the dry season. The structure of these forests and their scarce density make GFCD a more accurate tool for forest mapping, as a result of the minimum requirement of 10 acres of continuous forest to label an area as forest land. Even more so, in Colombia's ecoregions characterized by dry forests, where much of the original habitat has been destroyed by agriculture and overgrazing, leaving behind very fragmented forest patches.

A similar situation, in which GFCD shows better performances than IDEAM, was found in Cauca Valley montane forests, Cordillera Oriental montane forests, Magdalena Valley montane forests, Santa Marta montane forests and the Venezuelan Andes montane forests. All these ecoregions fall in the Northern Andean Montane Forests' global ecoregion and share similar characteristics. Montane forests have been fragmented by logging, farming, and ranching, particularly in the lower areas, but higher up there are still sizeable blocks of forest, some of which are protected. While in fragmented areas GFCD manages to better classify small patches, the difference between the two databases reduces in more intact zones of forest, so that in montane forests the discrepancy between the two is less accentuated than in dry forests. The performance of the two databases is inverted in Eastern Panamanian montane forests - a region largely intact with a substantial amount of endemism due to its inaccessibility - with IDEAM producing more reliable forest maps. A similar situation is observed in Northwestern Andean montane forests, an environment that has been greatly modified, although still presents sizable stands of continuous forest, with 50% of flora that is strictly endemic.

Other ecoregions where IDEAM outperforms the other two databases are, for instance, the Choco-Darien moist forests, Napo moist forests, Negro-Branco moist forests, Solimoes-Japura moist forests and Southwest Amazon moist forests. Even if not closely located to each other, these moist forests are all characterized by tall vegetation and dense evergreen tropical rainforest and are relatively intact (at least in part), which are the conditions under which IDEAM produces the most accurate results. Some of these ecoregions also present different large whitewater and blackwater rivers. These last ones, as well as lakes, may cause problems to GFCD, which easily miss-classifies them as forest, especially at thresholds lower than 90%. Among these moist forests, Choco-Darien represents an interesting case. It has extremely high rainfall, and the forests hold great biodiversity. Its northern and southern parts have been considerably modified for ranching and farming, and there are threats from logging for paper pulp, uncontrolled gold mining, coca-growing, and industrialization, but the central part of the ecoregion is relatively intact. The northern parts have mostly been replaced by banana plantations and cattle ranches, while those in the south by oil palm plantations and are being deforested for paper pulp. Oil palm and banana plantations represent another problem when using GFCD for forest mapping, as they are detected as forest, leading to substantial overestimation (see later discussion in this section). Some moist forest ecoregions represent, though, an exception, in that forest seems to be best predicted by GCFD. Magdalena-Uraba is one of them. It is surrounded by the most populated region of Colombia and is threatened by farming, ranching, logging, oil exploitation, and water pollution in the main rivers, so that areas where there has been little human impact, represent a residual portion. Habitat fragmentation and the presence of human settlements make GCFD a more accurate forest predictor.

Table 12. Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion in Colombia.

				GFCD's optimal		Best data set			
Ecoregion	**Region**	**n**	**Data set**	Mode	Std. Dev.	Overall Accuracy	User's Accuracy (Forest)	Producer's Accuracy (Forest)	Kappa
National		1940	GFCD	90	18.2	88.2***	88.8	91.2	75.5
			IDEAM			85.8***	96.5	78.5	71.9
			ESA			81.8***	84.9	83.7	62.7
Amazon-Orinoco-Southern Caribbean mangroves	Caribean	59	GFCD	1	39.9	77.6**	74.4	94.1	51.1
			IDEAM			72.4*	84.6	64.7	45.8
			ESA			51.7	66.7	35.3	9.4
Apure-Villavicencio dry forests	Andes	60	GFCD	70	20.8	79.7***	80.7	80.7	59.2
			IDEAM			74.6***	100	51.6	50.3
			ESA			61	83.3	32.3	24.3
Caqueta moist forests	Amazonia	60	GFCD	90	3.6	96.7	98.3	98.3	48.3
			IDEAM			90	98.2	91.4	21.1

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			93.3	100	93.1	47.4
Catatumbo moist forests	Andes	59	GFCD	90	12.5	79.7	84	91.3	33.5
			IDEAM			69.5	96.7	63	38.4
			ESA			72.9	81.3	84.8	16.5
Cauca Valley dry forests	Andes	60	GFCD	90	9.2	81.4*	75	63.2	55.5
			IDEAM			72.9	100	15.8	20.3
			ESA			69.5	52.9	47.4	28.2
Cauca Valley montane forests	Andes	60	GFCD	90	7.4	78.6**	81.1	85.7	53.4
			IDEAM			62.5	88.9	45.7	31.2
			ESA			64.3	68.3	80	19.2
Central American dry forests	Caribean	60	GFCD	20	0	100	–	–	–
			IDEAM			100	–	–	–
			ESA			100	–	–	–
Choco-Darien moist forests	Andes	60	GFCD	95	2.0	91.5	92.7	98.1	50.3
			IDEAM			98.3**	98.1	100	91.4

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			96.6*	96.3	100	81.5
Cordillera Oriental montane forests	Amazonia	60	GFCD	90	9.4	88.3***	94.4	87.2	75.2
			IDEAM			85***	100	76.9	70
			ESA			76.7*	79.1	87.2	46.4
Eastern Cordillera real montane forests	Amazonia	60	GFCD	90	1.8	94.9	96.3	98.1	69.9
			IDEAM			89.8	94.3	94.3	44.3
			ESA			89.8	89.8	100	0.00
Eastern Panamanian montane forests	Pacific	54	GFCD	90	1.7	92.2	92	100	31.1
			IDEAM			96.1	97.8	97.8	77.8
			ESA			92.2	92	100	31.1
Guajira-Barranquilla xeric scrub	Andes	60	GFCD	90	16.4	88.3*	73.3	78.6	68.2
			IDEAM			81.7	80	28.6	34
			ESA			80	58.3	50	41.2
Guianan piedmont and lowland moist forests	Orinoquia	45	GFCD	90	2.5	97.8**	97.4	100	91
			IDEAM			97.8**	97.4	100	91

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			95.6*	95	100	80.9
Iquitos varzea	Amazonia	53	GFCD	95	1.3	98.1***	97.4	100	95.2
			IDEAM			80.8	100	73	60.9
			ESA			78.8	90.6	78.4	53.1
Japura-Solimoes-Negro moist forests	Amazonia	60	GFCD	95	0.9	96.7	96.7	100	0
			IDEAM			100	100	100	100
			ESA			96.7	96.7	100	0
Llanos	Orinoquia	60	GFCD	70	4.9	94.9***	100	78.6	84.8
			IDEAM			98.3***	100	92.9	95.2
			ESA			93.2***	100	71.4	79.2
Magdalena Valley dry forests	Andes	60	GFCD	80	7.8	88.3***	87.5	84	75.9
			IDEAM			65	100	16	18.2
			ESA			70*	81.8	36	32.9
Magdalena Valley montane forests	Amazonia	60	GFCD	90	6.1	84.7***	82.9	90.6	69
			IDEAM			79.7***	91.7	68.8	59.9

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			74.6**	73	84.4	48
Magdalena-Uraba moist forests	Andes	60	GFCD	90	8.4	88.3***	78.3	90	74.7
			IDEAM			80*	100	40	47.1
			ESA			75	69.2	45	38.4
Napo moist forests	Amazonia	60	GFCD	90	2.7	88.3***	85.11	100	71.2
			IDEAM			93.3***	100	90	85.7
			ESA			86.7***	84.8	97.5	67.6
Negro-Branco moist forests	Amazonia	60	GFCD	95	2.5	90	90.9	98	52
			IDEAM			91.7	97.9	92.2	71.3
			ESA			85	86.2	98	13.5
Northern Andean paramo	Andes	60	GFCD	90	8.7	90**	100	60	69.2
			IDEAM			91.7***	81.3	86.7	78.3
			ESA			86.7*	70.6	80	66
Northwestern Andean montane forests	Andes	60	GFCD	90	4.0	85*	85.4	95.4	59.5
			IDEAM			93.3***	100	90.7	84.7

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			85*	87	93	61
Patia Valley Dry Forest	Pacific	60	GFCD	90	4.4	90	55	85.7	61.1
			IDEAM			91.7	66.7	57.1	56.9
			ESA			95	100	57.1	70.2
Rio Negro campinarana	Amazonia	60	GFCD	90	0.6	83.3***	79.1	97.1	64.1
			IDEAM			93.3***	91.9	97.1	86.1
			ESA			66.7	64.2	97.1	23.6
Santa Marta montane forests	Andes	60	GFCD	90	1.9	95	100	94.3	79.6
			IDEAM			68.3	100	64.2	29.5
			ESA			80	93.6	83	29.3
Santa Marta paramo	Andes	60	GFCD	90	0.0	93.3	62.5	83.3	67.7
			IDEAM			98.3*	100	83.3	90
			ESA			93.3	62.5	83.3	67.7
Sinu Valley dry forests	Andes	60	GFCD	90	6.9	80	60	60	46.67
			IDEAM			75	50	6.7	6.3

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

			ESA			75	50	40	28.6
Solimoes-Japura moist forests	Amazonia	60	GFCD	90	0	98.3	98.3	100	65.9
			IDEAM			100	100	100	100
			ESA			96.7	96.7	100	0.00
South American Pacific mangroves	Pacific	60	GFCD	90	0.9	86.7	86.2	100	29.4
			IDEAM			93.3*	94.2	98	73.9
			ESA			75	85.7	84	13.5
Southwest Amazon moist forests	Amazonia	57	GFCD	90	0.0	87.7	60	37.5	39.6
			IDEAM			89.5	100	25	36.4
			ESA			82.5	25	12.5	8.1
Venezuelan Andes montane forests	Andes	55	GFCD	90	0.0	87.3*	88.6	95.1	63.9
			IDEAM			78.2	93.9	75.6	51.6
			ESA			76.4	85	82.9	39.2
Western Ecuador moist forests	Pacific	58	GFCD	90	7.9	82.8	83.9	97.9	11.6
			IDEAM			81	100	77.1	53.7

Table 16 continued: Accuracy measures of the forest cover maps obtained from the GFCD, ESA and IDEAM databases, for each of the ecoregions.

	ESA	86.2	90	93.8	47.5

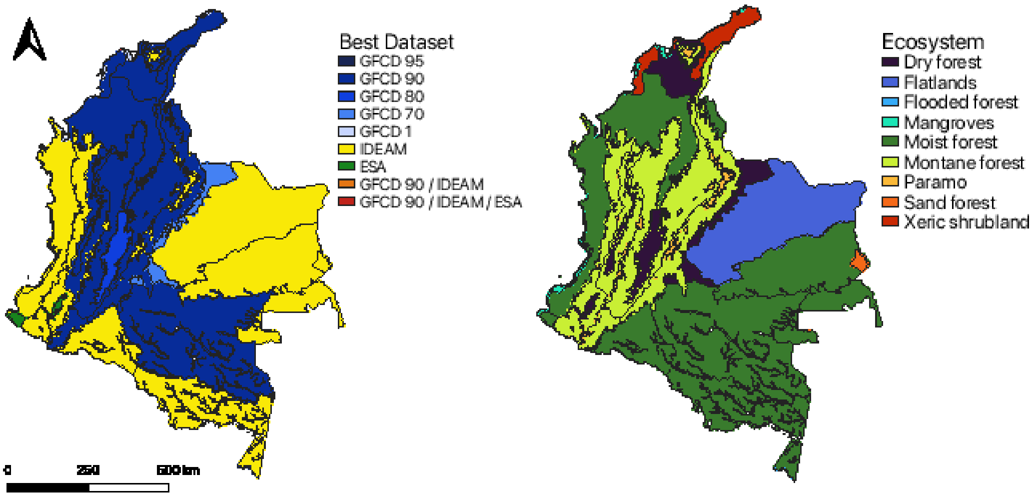


Figure 22. ***Left:*** Best data set per ecoregion. ***Right:*** Ecoregions by ecosystem.

To summarize the results, we group the 33 ecoregions into 9 ecosystems sharing similar landscape features (Figure 22, right panel) and compare the map of the ecosystems with that of the best performing database (on the base of the overall accuracy) on each ecoregion (Figure 22, left panel). From Figure 22, comparative performances of GFCD and IDEAM in different densely forested areas emerge clearly. Each of them proves superior in around half of the Colombian territory. GFCD is mostly representative of forests in the Andean region and most of the montane forest districts are better captured by GFCD, while IDEAM is better suited for denser forest settings. In particular, IDEAM is an excellent tool to predict tropical forests, especially when fragmentation is not a predominant feature of the landscape, and there are no steep slopes. In Amazonia, in the western and northern moist forests, and the northeastern flatlands, IDEAM is the recommended data set for these reasons. Note that the area in Amazonia that seems better predicted by GFCD is the large ecoregion of Caqueta moist forests. However, qualitative analysis of the surroundings of the sample points for this ecoregion pointed to the use of IDEAM (see Table 15 in Appendix 2C), despite a slightly smaller overall accuracy. On the other hand, IDEAM tends to underestimate forests in fragmented landscapes, like in the Andes, where crops, pastures, and settlements are present. As a result, the database in these areas sometimes fails to detect any forest at all, resulting in a strong underestimation of forest cover. GFCD tends, instead, to overestimate tree cover in tropical forest settings or, in general, in very dense forms of forests. In addition, we remark that the IDEAM algorithm can agglomerate points much better to make clear-cut distinctions of the forest and non-forest areas, matching the actual borders. GFCD tends to agglomerate pixels less, this is why the edges or borders are less defined in general and IDEAM is often preferable, according to the qualitative analysis (see Table 15 in Appendix 2C).

Complementary to Figure 22, Figure 23 shows the overall accuracy for each ecosystem and for each data set. In ecosystems where forest is more dense and where steep slops are absent, IDEAM shows a higher percentage of correctly predicted points. This happens in the wide lowland ecosystems, and in the areas where moist forests are more dense and better preserved. It also happens in smaller ecosystems, such as sand forests, mangroves, and paramos. However, IDEAM under-performs in the vast ecosystem of dry forests and in the smaller ecosystems of montane forests, xeric shrubland, and flooded forests, where GFCD is able to better identify the fragmented forests in these ecosystems. ESA, on average, does not outperform the other two databases in any of the ecosystems.

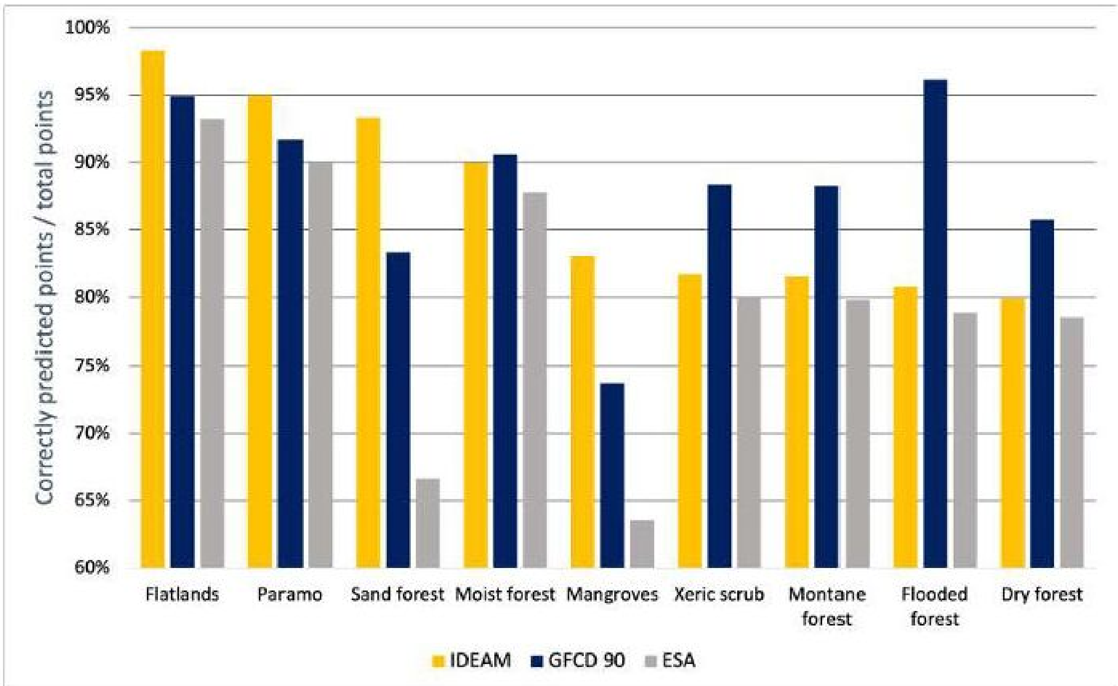


Figure 23. Percentage of correctly predicted points by ecosystem for IDEAM, ESA and GFCD (90% threshold).

Besides forest overestimation in densely forested areas, there is a relevant problem concerning GFCD, which we mentioned before and discuss here in more detail, namely, the fact that palm oil plantations and other crops are detected as forest (see also Tropek et al., 2014). In the present study, we do not consider palm oil plantations as forests, as many papers point to the disruption that this monoculture is causing in tropical forests, endangering species and biodiversity (Meijaard et al., 2020). Palm oil plantations are mostly located in the Andean region, and we report in Figure 24 two examples of zones where this cultivation is present in 2019. We observe that forest has been cleared out to make space for palm oil crops and the surrounding areas are also deforested. In both cases, GFCD classifies these areas as forest, while IDEAM does not.

Figure 24. Deforestation caused by palm oil plantation in two moist forest areas in the Andes: **a)** Magdalena Uraba and **b)** Catatumbo.

To conclude, Figure 25 gathers information on the producer's, user's and overall accuracy at the national level for all data sets and GFCD's selected thresholds. The producer's and user's accuracy statistics are only displayed for forest, as this is the main objective of our study. The GFCD shows an increasing overall and user's accuracy and a decreasing producer's accuracy, as the threshold increases. These trends can be explained as follows. As the threshold increases, the predicted forest decreases and so does the number of points erroneously classified as forest, decreasing the commission error and increasing the user's accuracy. As the predicted forest decreases, also the number of points correctly classified as forest decreases, increasing the omission error and decreasing the producer's accuracy. Altogether, however, as the predicted forest decreases, the number of points erroneously classified decreases more than those correctly classified, so that, the overall accuracy increases. This, in addition to the fact that the user's accuracy is below the producer's accuracy for all the thresholds considered, corroborates the tendency of GFCD to overestimate forest cover. The overall, producer's and user's accuracy statistics tend to converge to the same value for the 90% threshold. IDEAM has the highest user's accuracy of the three data sets and the lowest producer's accuracy, confirming the tendency of this database to underestimate, rather than overestimate forest. Its overall accuracy is, however, in line with the values observed for GFCD when using thresholds from 70% to 90%. Finally, ESA has a low user's accuracy when compared with the other two data sets, and an overall accuracy smaller than those of IDEAM and GFCD with a threshold of at least 20%. In summary, at the national level, IDEAM is preferable over GFCD as the overall accuracy is very close to that of GFCD with an optimal threshold, but the user's accuracy is visibly higher, reducing the risk of overestimation of forest cover, which is a more serious type of error for deforestation research.

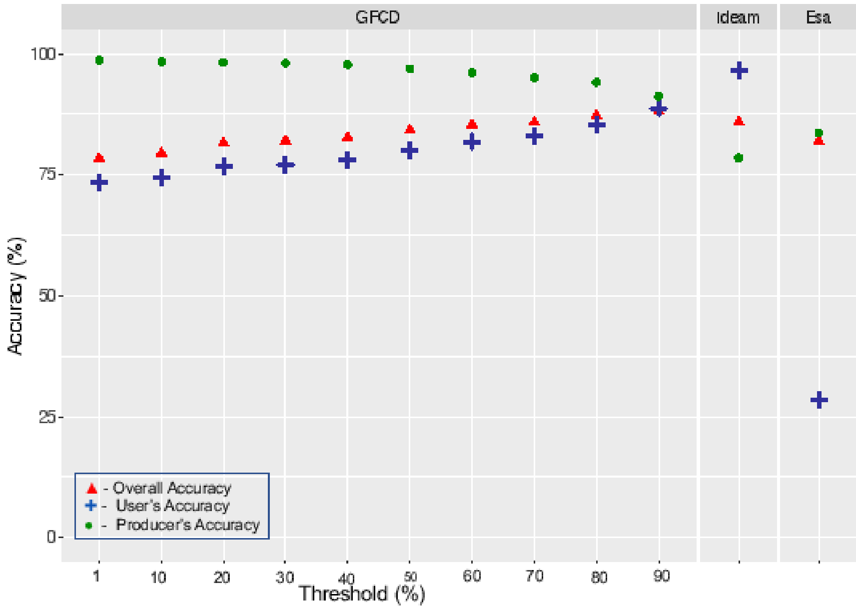


Figure 25. Overall accuracy, user's accuracy and producer's accuracy for GFCD with selected thresholds, IDEAM and ESA at national level. The producer's and user's accuracy statistics are displayed for the forest case.

5 Discussion

Deriving accurate and updated forest cover maps is essential in monitoring forest stocks and contrasting deforestation, an issue that is becoming more and more urgent to solve. Reducing deforestation is absolutely crucial in large forest systems, such as the Amazon rainforest, as they play a significant role not only in converting carbon dioxide into oxygen but also in regularizing global water cycles. Deforestation in these critical regions has a direct impact on the availability of freshwater supplies worldwide. Using raw remote sensing data to produce forest cover maps is a prohibitive task for most researchers. Freely available land cover map products represent, in this regard, an invaluable tool. Understanding their peculiarities and assessing their accuracy is indispensable for obtaining - for as far as possible - accurate calculations of forest extension.

In this study, we compared the performance of different freely available databases in producing forest cover maps in Colombia, namely, GFCD, IDEAM, and ESA. For GFCD we also considered the accuracy obtained using different tree cover thresholds to produce maps. A careful choice of this threshold is of practical importance as it affects forest cover estimates. To assess the accuracy of forest mapping for the different databases, we collected reference data on sample points from each ecoregion in Colombia, calculated confusion matrices between each forest map and reference data, and obtained accuracy metrics from them. We also enriched our conclusions with additional (observational) information that helped us identify the drawbacks and benefits of each data set.

Regarding GFCD, we showed that different tree cover thresholds are needed to maximize overall accuracy in the different ecoregions, although a 90% threshold is required in most of them and is recommended at the national level. This result depends on the fact that, in Colombia, the tropical moist forest is dominant and for this type of dense forest a higher threshold is necessary to avoid incurring in forest overestimation. McRoberts et al. (2016) demonstrate that a 95% threshold determines the highest overall accuracy in Santa Catalina in Brazil, Sannier et al. (2016) suggest a 70% threshold in Gabon, and Lwin et al. (2019) advocate an 80% threshold in the tropical rainforest of Myanmar. Despite these studies, much lower thresholds have been used in literature for Colombia. Just to cite a few, Negret et al. (2021) and Mendoza (2020) use a 30% threshold, and Harding et al. (2022) a 50% threshold. González-González et al. (2021) use a 60% threshold in studying mining contribution to Colombian deforestation, but they also consider different coverage levels (10%, 30%, 50%, and 70%), as well as IDEAM data, for checking results' robustness. In light of this, the present study helps practitioners towards a more aware choice of tree cover threshold when using GFCD for forest mapping, to account for the type of ecoregion under study, and avoid under (or over) estimation biases.

As far as the comparison between the three databases is concerned, we found no overall optimal database and, again, the characteristics of the region under consideration should instead drive the choice. GFCD was found to have the highest overall accuracy in 15 ecoregions, with a threshold of 90% in 11 of them, out of the 33 ecoregions into which Colombia is divided. Next is IDEAM, showing the highest overall accuracy in 14 ecoregions. ESA turned out to be more accurate than the other two databases in just 2 small ecoregions, while for the 2 remaining ecoregions more than one database proved adequate. In particular, IDEAM was perceived as more accurate in those areas where there are no steep slopes and forest is more dense and continuous. This is partly a consequence of IDEAM's minimum requirement of 10 acres of continuous forest to categorize the area as forest. Moist forests, lowlands, sand forests, paramos and mangroves are, therefore, the Colombian ecosystems where IDEAM performs best. On the other hand, GFCD proved superior in areas where settlements are present, as well as in mountainous and highly fragmented forest areas. Montane forests, dry forests, flooded forests and xeric shrubs are the ecosystems for which GFCD should be preferred for forest cover mapping. In general, we found that IDEAM tends to underestimate forest cover, showing a user's accuracy generally higher than the producer's accuracy. The opposite situation was true for GFCD, which tends to overestimate forest, above all for small values of the threshold. Since deforestation rate overestimation is considered as a more worrying bias, for this reason we suggest using IDEAM at national level, as having comparable overall accuracy to GFCD but larger user's accuracy. In addition, GFCD has the relevant drawback of miss-classifying lakes, rivers, palm oil and other plantations as forest. Despite the definition of forest given in GFCD, even vegetation lower than 5m (e.g., pineapple, soybeans, or tea plantations) is often classified as forest (Tropek et al., 2014). Erroneous consideration of plantations represents a serious problem as they do not have the carbon value of forests and often contribute to endanger species and biodiversity. Certainly GFCD can be improved by deleting non-seasonal lakes, rivers, and plantations from the final mask, as vector layers containing this information are available. However, this would complicate the use of this database making it less accessible. On the other hand, the principal shortcoming of IDEAM is that the latest available data dates to two years back or more. For example, the latest available data set in 2022 is the one referred to the year 2019. Meanwhile, ESA and GFCD offer very up-to-date information.

6 Conclusions

Remote sensing of forest cover is still far from being precise and faultless. The availability of free products, coming with a user-friendly online application, further enhances the potential risk of uncritical use. This might produce biased conclusions to research questions and mislead conservation policies, with potentially serious consequences for biodiversity and socioeconomic issues. An example is provided in Fergusson et al. (2020), where it is shown how the use of different sources of data (GCFD or IDEAM) can determine opposite answers to a specific research question. Our study finally calls for caution when using an automated classification of satellite imagery and aims to provide scholars involved in monitoring deforestation in Latin America with a proper accuracy assessment of various available sources of data, as well as guidelines for a more careful choice of the tree cover threshold, for the case of the GFCD database.

Remote sensing is expected to improve in the coming years, as more accurate satellite images and better algorithms will be available. However, at present, studies should consider and acknowledge the limitations of the different data sources; data including a field validation protocol carried out by experts, such as IDEAM, should be preferred;

checking robustness of results to different data sources would be highly advisable; accuracy assessment of forest cover maps based on different products should be carried out for all countries, with a high priority for those hosting precious and vulnerable tropical forest.

Appendices

2A IDEAM's vector file values and ESA categories

Table 13. IDEAM layer values

Forest	1
Non-Forest	2
No information	3

Min (ESA scale)	Max (ESA scale)	Esa's Category	Adapted Category (Value)	Adapted Category (Name)	Code
0	0	No Data	0	No Data	.-
10	12	Cropland, rainfed	1	Agriculture	AG
20	20	Cropland, irrigated or post-flooding	1	Agriculture	AG
30	30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	1	Agriculture	AG
40	40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	1	Agriculture	AG
50	50	Tree cover, broadleaved, evergreen, closed to open (>15%)	2	Forest	FO
60	62	Tree cover, broadleaved, deciduous, closed to open (>15%)	2	Forest	FO
70	72	Tree cover, needleleaved, evergreen, closed to open (>15%)	2	Forest	FO
80	82	Tree cover, needleleaved, deciduous, closed to open (>15%)	2	Forest	FO
90	90	Tree cover, mixed leaf type (broadleaved and needleleaved)	2	Forest	FO
100	100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	2	Forest	FO
160	160	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	2	Forest	FO
110	110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	3	Grassland	GR
130	130	Grassland	3	Grassland	GR
180	180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water	4	Wetland	WE
190	190	Urban areas	5	Settlement	SE
120	122	Shrubland	6	Shrubland	SH
140	140	Lichens and mosses	7	Other	OTH
150	153	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	7	Other	OTH
200	202	Bare areas	7	Other	OTH
210	210	Water bodies	7	Other	OTH
220	220	Permanent snow and ice	7	Other	OTH

Figure 26. Level 1 (or global) legends of the CCI LC maps and our converted categories

2B Cohen's Kappa Table

Table 14. Interpretation of Cohen's kappa

Value of Kappa	Level of agreement	% of data that are reliable
0–.20	None	0–4%
.21–.39	Minimal	4–15%
.40–.59	Weak	15–35%
.60–.79	Moderate	35–63%
.80–.90	Strong	64–81%
Above 90	Almost Perfect	82–100%

2C Qualitative results from visual observation of enlarged maps

Table 15. Qualitative results on accuracy of forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion.

Ecoregion	Region	n	Best data set	GFCD's optimal threshold (mean)	No optimals	General characteristics of observed land
Amazon-Orinoco-Southern Caribbean mangroves	Caribean	59	IDEAM	39	1	Swamps/ fragmented forest
Apure-Villavicencio dry forests	Orinoquia	60	GFCD	67	0	Crops/ fragmented forest
Caqueta moist forests	Amazonia/ Orinoquia	60	IDEAM	91	1	Dense forest
Catatumbo moist forests	Andes	59	IDEAM	90	17	Dense forest/ Fragmented forest
Cauca Valley dry forests	Andes/ Pacific	60	GFCD	83	0	Crops/ pastures/ settlements
Cauca Valley montane forests	Andes/ Pacific	60	GFCD	89	6	Mountainous forest
Central American dry forests	Caribean	60	GFCD	20	0	Swamp/ Asphalt/Small area
Choco-Darien moist forests	Pacific	60	IDEAM	94	10	Dense forest

Appendix table 15 continued: Qualitative results on accuracy of forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion.

Cordillera Oriental montane forests	Andes/ Orinoquia	60	GFCD	84	4	Mountainous fragmented forest/ Sparse forest
Eastern Cordillera real montane forests	Andes	60	IDEAM	90	4	Dense and fragmented forest. Problem with IDEAM in this area: many <no information points>
Eastern Panamanian montane forests	Pacific	54	IDEAM	91	2	Mostly dense forest with some fragmented forest
Guajira-Barranquilla xeric scrub	Andes/ Caribean	60	IDEAM	74	9	Shrubland/ sparse forest
Guianan piedmont and lowland moist forests	Orinoquia	45	IDEAM	92	0	Dense forest
Iquitos varzea	Amazonia	53	GFCD	95	4	Dense forest
Japura-Solimoes-Negro moist forests	Amazonia	60	IDEAM	95	1	Dense forest
Llanos	Orinoquia	60	IDEAM	70	0	Forests in river-like form

Appendix table 15 continued: Qualitative results on accuracy of forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion.

Magdalena Valley dry forests	Andes	60	GFCD	84	11	Highly fragmented forest/ Crops/ Grazing land
Magdalena Valley montane forests	Andes	60	GFCD	89	3	Fragmented forest/ Settlements/ Crops/ Pastures
Magdalena-Uraba moist forests	Andes/ Caribean	60	GFCD	89	9	Swamps/ Sparse forest
Napo moist forests	Amazonia/ Andes	60	GFCD	91	13	Dense and fragmented forest
Negro-Branco moist forests	Amazonia/ Orinoquia	60	IDEAM	93	1	Dense forest
Northern Andean paramo	Andes/ Pacific	60	IDEAM	86	2	High mountain hill
Northwestern Andean montane forests	Andes/ Pacific	60	IDEAM	90	1	Dense and fragmented forest
Patia Valley Dry Forest	Pacific	60	No opti- mal	88	30	Shrubland/ fragmented forest

Appendix table 15 continued: Qualitative results on accuracy of forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion.

Rio Negro campinarana	Amazonia	60	IDEAM and GFCD	90	1	Forests in river-like form
Santa Marta montane forests	Andes/ Caribean	60	GFCD	90	1	Montanous fragmented forest
Santa Marta paramo	Andes/ Caribean	60	IDEAM	90	0	Rocky mountain
Sinu Valley dry forests	Andes/ Caribean	60	No opti-mal	88	36	Savanna/ Shrubland
Solimoes-Japura moist forests	Amazonia	60	IDEAM	90	0	Dense forest
South American Pacific mangroves	Pacific	60	IDEAM	90	7	Dense forest
Southwest Amazon moist forests	Amazonia	57	GFCD	90	8	Settlement
Venezuelan Andes montane forests	Andes	55	GFCD	90	3	Fragmented forest

Appendix table 15 continued: Qualitative results on accuracy of forest cover maps obtained from the GFCD, ESA and IDEAM databases, by ecoregion.

Western Ecuador moist forests	Pacific	58	IDEAM	89	10	Dense forest with multiple small disturbances

Note: "No optimals" indicates the number of points for which the best data set did not show the best visual accuracy.

Chapter 3

Evaluating the causal effect of glyphosate aspersion on coca crops in Colombia

Abstract

In Colombia, aerial fumigation with glyphosate was a central policy to curb coca crops between 1994 and 2015, when this measure was banned following years of opposition and questionings about its effectiveness as well as concerns about its collateral effects on local communities and the environment. Many studies, environmental groups, and individual legal complaints argue that health issues and environmental degradation are a direct consequence of the use of glyphosate, however, there is no universal agreement about the carcinogenic potential of glyphosate in humans among international institutions, and its environmental impact is also debated. Discussions about bringing back the use of glyphosate have recently resumed, and some groups are determined to reintroduce it. In this context, this study aims to investigate the effectiveness of using fumigation with glyphosate to reduce the spread of coca plantations. We use a 15-year panel (2000-2015) of the municipalities where coca was detected in 2000, and a regression model including fixed effects for time and municipalities. To deal with endogeneity issues, we instrument aspersion with the number of days in which the strength of wind was below a certain threshold and aerial fumigation could take place. Estimates provide some evidence that aspersion enhances coca crops extension, in most of model specifications.

1 Introduction

In 2020, the area under coca bush cultivation in Colombia decreased by 7%, compared with the previous year, confirming a downward trend that began in 2017. Despite this reduction, the annual yield of coca leaf per hectare increased by around 10%, going from 5.8 mt/ha/year[1] in 2019 to 6.4 mt/ha/year in 2020, so that the production of cocaine hydrochloride globally increased to 1.228 tons, 8% higher than in 2019 (UNODC, 2021). As a result, Colombia continues to be the largest producer of cocaine in the world, well ahead of Peru and Bolivia. The effects of coca cultivation and trafficking bear enormous economic, safety and social costs to Colombians.

To offset the increasing levels of coca production, the country has mostly relied on eradication measures, like glyphosate aspersion and manual eradication, to tackle the production at its root, but the effectiveness of these policies is contested. Empirical studies about the efficacy of glyphosate aspersion are not abundant, and their results are sometimes contradictory. Moreno-Sanchez et al. (2003) use data at national level for the period 1987-2001 and conclude that eradication is an ineffective means of supply control. Rincón-Ruiz and Kallis (2013) consider instead municipal data between 2001 and 2008. Employing Moran's I index, they conclude for a spatial correlation between the area fumigated per municipality in a given year and the area under cultivation in the neighborhood of the same municipality in the following year. Therefore, they provide evidence for the so-called "balloon effect", i.e. the fact that fumigation, instead of eradicating coca production, simply displaces it to other regions, and particularly in areas of primary forest of great environmental importance. Also Dion and Russler (2008), using cross-section data for 32 sub-national departments between 2001 and 2005, conclude that aerial eradication is not effective in reducing coca cultivation and that the limited impact is temporary and mostly due to displacement associated with fumigation, thus generating additional human and economic costs. Based on their findings, they suggest to develop public infrastructure in conjunction with poverty reduction programs in the affected localities as a more effective way to curb coca crops.

The papers above, however, do not address the issue of endogeneity, which is a serious concern in this context and, in general, in all those studies evaluating the effect on illegal activities of law enforcement and policies (Allers and Hoeben, 2010; Cawley et al., 2018; Weber and Key, 2012; Rodriguez, 2020). Eradication efforts are, in fact, concentrated in particular regions of the country, depending on their levels of coca cultivation. Therefore, while eradication may or may not have an impact on coca crops, it is quite likely that coca cultivation has an impact in determining which areas, and to what extent, are targeted by fumigation, giving rise to simultaneity problems. In recent papers, this issue has been addressed. Reyes (2014), for example, uses municipal panel

[1] Metric tones per hectare per year.

data from 2001 to 2006 to estimate the effect of eradication on coca cultivation, by exploiting exogenous sources of variation in eradication. In particular, in an instrumental variable approach, he instruments aerial aspersion with the distance from the airport. Such a distance reflects differences in the expected costs of eradication: as air-crafts used for spraying get far from the area in which Anti-narcotics Police can protect them, the probability of being shot down by illegal groups increases. His conclusion is that a 1% increase in eradication determines a 1% increase in coca cultivation. A similar causal effect is found in Bogliacino and Naranjo (2012), who also use municipal panel data from 2000 to 2008 but instrument spraying with crime rate and arrival of refugees from forced displacement. On the other hand, Rozo (2014) finds evidence of a small negative effect of aerial spraying on cultivation, but also sizable unintended negative effects on the welfare conditions of the treated areas. She uses a data set that tracks coca cultivation and eradication on grid points of one square kilometer from 2000 to 2010, and exploits the exogenous variation created by restrictions to spraying in protected areas (i.e., indigenous territories and natural parks) and the time variation of U.S. international anti-drug expenditures. Analogously, Abadie et al. (2015), employing a linear dynamic panel model at the municipal level and the Arellano and Bond (1991) estimator, find that aerial spraying reduces the cultivated area, but significantly increases guerilla-led violence. Mejía et al. (2017) also conclude that aspersion has a negative and significant impact on coca cultivation, but the size of this effect clearly identifies aerial spraying as a cost-ineffective strategy. Using the same data set as Rozo (2014), they apply conditional difference-in-difference and regression discontinuity design estimators by exploiting exogenous cessation of aerial fumigation on the Colombia-Ecuador border resulting from diplomatic tensions. Finally, Cote (2019), studies the causal effect of various interdiction strategies[2] as well as eradication policies on coca cultivation. He applies an Arellano-Bond GMM approach to estimate a dynamic linear model on municipal panel data, using lagged levels of the dependent variable as instrument. In addition, he also includes external instruments that explain both interdiction and eradication, such as: the interaction between the distance of the municipality to a military base, and the US military aid to Colombia in billions of dollars; the interaction of the percentage of municipal area outside protected areas and US military aid to Colombia (similarly to Rozo, 2014); the interactions of the distance to military base and the percentage of area outside protected areas with year dummy variables. He does not find a robust effect of aerial spraying on coca cultivation. Although the coefficient of this variable is always

[2] The interdiction policies considered encompass dismantling of laboratories, seizures of coca base, seizures of coca leaves, seizures of cocaine.

negative, it is only significant at the 10% level in two of the specifications.

The aim of this study is to contribute to the current debate on whether glyphosate aspersion is an effective policy to reduce coca crops. We implement a panel regression analysis at the municipal level that comprises the years from 2000 to 2015, which is the last year when aspersion took place in Colombia, and we use an instrumental variable approach to account for the endogeneity of the aspersion policy. In particular, we use the strength of wind as an instrument, as spraying cannot take place under unfavorable weather conditions. A similar approach is used in Rodriguez (2020) to estimate the causal effect of aerial aspersion on child labor and education. Since the instrumental variable used varies in both time and space, we can also include municipal and time fixed effects in the model. In addition, we control for several covariates, which might be correlated with wind strength and coca cultivation.

This paper is divided into five sections: Section 2 describes the variables used for this research, with a special focus on aspersion and the instrumental variable; Section 3 details the methodology applied and justifies the choice of using wind speed as an instrument; Section 4 outlines the results; Section 5 discusses our main findings; Section 6 concludes and provides scope for future research.

2 Data

In this study, we use a panel data set at the municipality level and spanning the period 2000-2015. In the panel regression, the response variable is coca crops and the predictor of interest, for which we want to determine the causal effect, is aspersion. Wind speed is considered as an instrumental variable to deal with endogeneity issues. In addition, we include several variables that contribute to the changes in coca crops, as controls. We add environmental variables like rainfall, and other policies to reduce coca crops like seizures of clorohidrate of cocaine, dismanteled labs of cocaine and manual eradication. We also consider the degree of violence and conflict at the municipality level, proxied by the number of homicides and displaced people. In this section, we provide details on the variables considered, with a particular regard to those of direct interest.

2.1 Variables and municipalities under study

Table 16 displays all the variables used in this research, together with their definition and source. In particular, the response variable coca crops is defined as the land extension, measured in hectares, under this cultivation at the end of a given year. Data are obtained from the Observatory of Drugs of Colombia (from spanish OCD), which collected information from the year 2000 until 2015.

Table 16. Definition and data source for the variables used in this study

Variable	Description	Source
Aspersion	Hectares of coca eradicated with glyphosate aspersion	Centro de Estudios sobre Desarrollo Económico
Coca	Hectares of cultivated coca	Centro de Estudios sobre Desarrollo Económico
Manual eradication	Hectares of coca manually eradicated	Centro de Estudios sobre Desarrollo Económico
Displaced people	Number of forced displacements	Centro de Estudios sobre Desarrollo Económico
Seizures	Number of seizures	Observatorio de Drogas de Colombia
Wind	Mean number of days with average daily wind below 3 m/s in the municipality	Global Land Data Assimilation System 2.1
Homicides	Total numbers of homicides	Centro de Estudios sobre Desarrollo Económico
Rainfall	Mean number of days where the maximum daily rainfall exceeded 0.3m in the municipality	European Centre for Medium-Range Weather Forecasts (ERA5)
Credit	Subsidies to small producers	Centro de Estudios sobre Desarrollo Económico

Variables in Table 16 are used to build a panel dataset comprising the years from 2000 to 2015, that includes only the municipalities where coca is detected in the year 2000. Such a restriction is also used in Cote (2019) and Abadie et al. (2015), to have less zero-inflated data. As a result, the original number of 1060 municipalities is reduced to 172, for a total of 16072 observations. In Figure 27, we show the 172 municipalities that compose the dataset. The municipalities are obtained from the GADM database (version 3.6). Based on the supplementary material of Mendoza (2020), the original 1065 municipalities are aggregated, reducing the number of municipalities to 1060. As a result, the indexing and data matching processes are simplified (Christen, 2012).

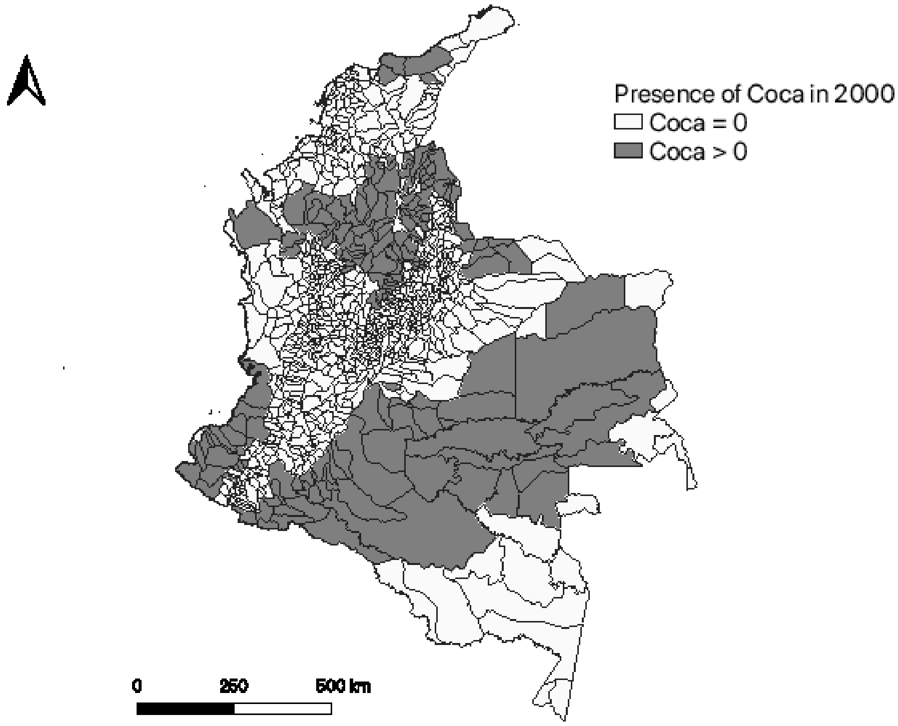


Figure 27. The 172 municipalities where coca was detected at the beginning of the studied period in 2000. Data source: Observatorio de drogas de Colombia.

Table 17. Descriptive Statistics

Variable	Unit	Mean	St. Dev.	Min	Max
Coca	Hectares	85.1	546.2	0	24,507
Aspersion	Hectares	104.9	797.1	0	33,814
Rainfall	Num. days	179.5	42.7	26	350
Manual Eradication	Hectares	30	329.1	0	19,906.6
Seizures	Kilograms	120.8	1,209.1	0	43,346
Dismanteled Labs	Num. labs	0.3	4	0	378
Credits small producers	Colombian Pesos	776.9	1,239	0	17,283
Homicides	Num. homicides	14	76.6	0	2,163
Displaced people	Num. people	248.6	1,535.3	0	56,087
Area	Km^2	1,100.7	3,606.6	15.9	66,272

In Table 17 we present the descriptive statistics for the variables used in this study for the 172 municipalities selected. We shortly explain some of the most extreme values present in this table. First, the maximum value of coca extension registered in a municipality during the 15 years of our study, equal to 24,507 hectares, refers to the year

2000 and the municipality of Valle de Guemez in Putumayo, in the South of the country, very near to the Ecuadorian border, an area which has been historically highly affected by the presence of coca crops. Similarly, the largest area fumigated is the municipality of Orito in Putumayo in 2002. The value is higher than the extension of coca crops identified during the same year in that municipality because a crop can be fumigated more than once a year, and it is also likely that other illegal crops might have been growing in Orito. From the statistical description of the variables, it is also evident that the manual eradication effort in terms of hectares is smaller than aerial spraying, because of the safety issues implied in that measure and the physical work needed to eliminate the bushes by hand. The maximum number of homicides, equal to 2163, is registered in the municipality of Cali in Valle del Cauca in 2004. Likewise, the highest number of displaced people is observed in the capital city of Bogotá in 2008, and is likely due to the global economic crisis that began in that year. Finally, the smallest area corresponds to the municipality of Busbanzá in Boyacá and the largest to Cumaribo in the department of Vichada.

For the main variables in our study, coca crops and aspersion, we illustrate in Figure 28 the municipalities in which the presence of coca has been more persistent and the number of years during which aspersion took place in each of them. A darker shade of gray indicates a higher number of years of coca cultivation and aerial spraying.

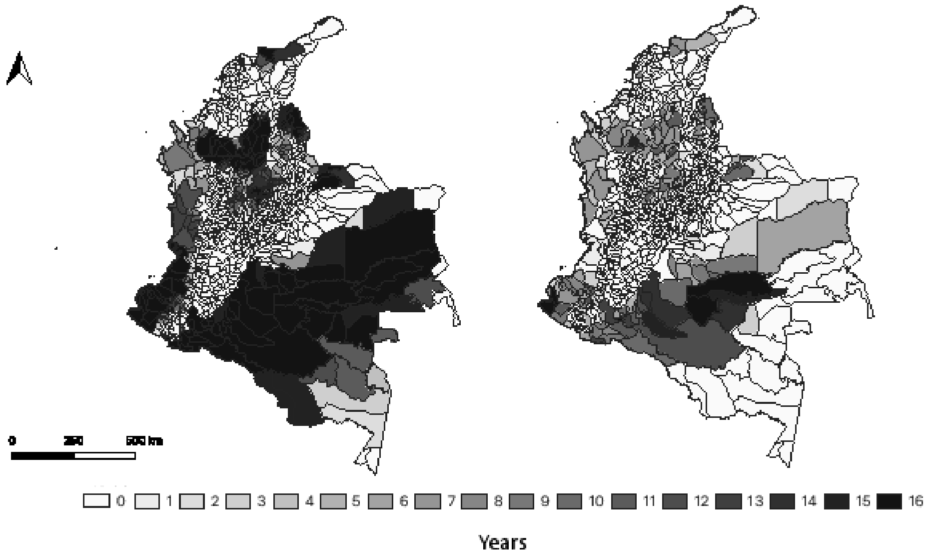


Figure 28. **Left:** Number of years of coca crops detected in each municipality during 2000-2015. **Right:** Number of years where fumigation took place in each municipality during 2000-2015. Data source: Observatorio de Drogas de Colombia.

2.2 Glyphosate

The herbicide glyphosate is a broad-spectrum systemic herbicide and crop dissecant with the highest production volume among all herbicides. It was developed to control the growth of weed species, and is used in agriculture, forestry, urban, and home applications (IARC, 2017), but the use of this chemical is highly controversial. On one side, different studies warn of the increased risk in humans of non-Hodgkin lymphoma (Zhang et al., 2019; Weisenburger, 2021; Taioli et al., 2020), follicular lymphoma (Meloni et al., 2021), multiple myeloma, acute myeloid leukemia (Andreotti et al., 2018) or proliferative effects in hormone-dependent breast cancer (Thongprakaisang et al., 2013) derived from the exposure to glyphosate or glyphosate-based herbicides for different groups of people exposed to the substance (like applicators, farmers or other exposed groups). In 2015 the International Agency for Research on Cancer (IARC, 2017), which is an intergovernmental agency forming part of the World Health Organization (WHO), carries out a study on 1000 people, and concludes that there is strong evidence that glyphosate is operating through two characteristics: 1) genotoxicity, referred to damage to the deoxyribonucleic acid (DNA), 2) oxiditive stress, signifying that cellular damage occurs through the presence of free radicals, both, for pure glyphosate and for glyphosate formulations. The agency labels glyphosate as probably carcinogenic for humans, and classifies it into group A2 (IARC, 2017) on the basis of "*limited* evidence of cancer in humans (from real-world exposures that actually occurred) and *sufficient* evidence of cancer in experimental animals (from studies of pure glyphosate)". These findings are contested in new studies that claim that glyphosate is not toxic to humans and, especially, that there is no supporting evidence that the herbicide actually causes cancer (Mink et al., 2011; Tarone, 2018; Berry, 2020; Williams et al., 2000). These conclusions are supported by the governments of most countries in the world, including almost all countries in the European Union, except for Luxembourg. It is evident that the differing conclusions about glyphosate's carcinogenic potential among international regulatory agencies, governments and independent studies have further incited controversy[3].

In more recent years, as studies and positions have polarized, some governments have decided to ban or regulate the use of glyphosate. The European Food Safety Authority (EFSA) has labeled glyphosate "as causing serious eye damage and being toxic to aquatic life [...]. [The European Chemical's] Agency Committee for Risk Assessment (RAC) also concludes that classifying glyphosate as a carcinogenic, mutagenic or reprotoxic

[3]The Wikipedia article on glyphosate reflects the diametric opinions and conflicting interests regarding glyphosate, considering that the discussion (talk section) consists of 19 archives until now.

substance is not justified" (EFSA, 2022). One of the largest citizen initiatives to ban glyphosate in the European Union "The stop glyphosate initiative" was presented in 2017, but the European Union has not acknowledged the potential health issues of the substance. European countries have differing legislation regarding glyphosate, for example, in Italy, the use of glyphosate is allowed in agriculture, but not during ripening and harvesting (G.U. n.193 19/08/2016). Other European countries have introduced a progressive prohibition on glyphosate, like Germany, France or Austria, but only Luxembourg has completely banned glyphosate from agricultural use. Mexico has set a progressive ban on glyphosate to be completed in 2024 (de Miguel, 2021).

The adverse effects of glyphosate in aquatic environments is more widely recognized. For this reason, the majority of glyphosate-based herbicides are not approved for use in aquatic environments. Nevertheless, measurable amounts of the active ingredient and surfactants have been detected in surface waters, making them potentially harmful for aquatic organisms (Annett et al., 2014). A glyphosate resistance transgene alone is highly unlikely to pose a threat to wild plant populations, but when linked to a transgene that confers a fitness benefit outside of agriculture (for instance, insect resistance), natural ecosystems could be affected (Cerdeira and Duke, 2006).

Banning glyphosate imposes serious challenges regarding crops' yields and rentability. The Glyphosate Renewal Group of France, a pro-glyphosate group, calculated that if glyphosate were banned, rentability of crops would decrease by 25%, and yields of corn, wheat, and barley would decline by 22% (Glyphosate Renewal Group, 2022). Similarly, there would also be a loss of $6.76 billion in global farm income each year, as well as a drop in international soybean, corn, and canola production levels (Brookes et al., 2017).

In conclusion, an unanimous consent about the effects of glyphosate on human health has not been reached, and the environmental impact has to be determined case by case.

2.3 Aerial aspersion with glyphosate

In this study the variable aspersion is defined as the hectares of land yearly sprayed with glyphosate in each municipality during the years 2000 to 2015. Colombia is the only country in the world to have performed aerial fumigation to combat illegal crops, and it was the single anti-coca and poppy eradication method for many years, until manual eradication was introduced in 2003. Aerial fumigation of illegal crops was introduced in 1978 to curb their propagation, which coincided with the hayday of the Cartel of Medellin, the largest drug cartel at that moment, and whose main merchandise was cocaine.

Fumigation in Colombia was carried out with a substance consisting of a mix of Glyfos (Roundup-SL®, Fuete-SL® and Gly-41®), and diluted with 2.3% v/v of the adjuvant Cosmo-Flux®411F, which increases penetration and enhances the dissecation of the coca plants (Solomon et al., 2009). Finally, water is added in the proportion of 50% of the original solution. A condition for fumigation is that it has to take place during the ripening of the plant to become effective. It has no effect on germinating seeds, because its action unfolds on the foliage of the plant (Henao Muñoz et al., 2015).

Spraying operations are conducted in Colombia under pre-established parameters, which are discussed in Section 2.4 and which are meant to reduce spray drift and maximize the efficacy of the treatment. The phenomenon of drift, inherent to aerial application of herbicides, refers to the fact that the substance intended to reach a certain area, actually affects also the surrounding areas. Spray drift depends essentially on wind speed and direction, as well as on a number of other atmospheric factors including temperature, relative humidity and atmospheric stability. Spraying should be performed at low altitudes, some 25 meters above the ground, but for security reasons, in some occasions, the aircrafts have to spray at higher altitudes, which increases the risk of drift and that means affecting the surrounding areas or legal crops (Ecuador v. Colombia, 2012). There is, in fact, a wealth of risks for the pilots during the fumigation operations: over 10 pilots were killed during aspersion missions and many aircrafts were hit from the ground in over 1000 occasions (El Tiempo, 2007).

The complaints about the effects of aerial fumigation from the affected communities were not only contained nationally. In 2008, Ecuador brought a suit before the International Court of Justice against Colombia on the grounds of glyphosate fumigations conducted on the Ecuadorian border by the Colombian government. Ecuador alledged that fumigation caused severe damages to population, crops, animals and the environment on ecuadorian territory (Government of Ecuador, 2008). In the end, Colombia accepted to establish a 10 km strip on the Colombian side along the border with Ecuador where aspersion was banned, in addition to an economic indemnization and a revisit of the fumigation policies of Colombia. This international suit increased the debate within Colombia about the collateral damages experienced by the affected communities and fumigation operations started to slow down in 2012 until a complete ban was passed on May 13 of 2015. Figure 29 shows the aspersion and coca crops trends during the period 2000-2020.

Figure 29. Hectares of coca crops and fumigated area (2000-2020). Data source: Observatorio de Drogas de Colombia.

Since the cultivations started to increase again after the suspension of the aerial fumigation, many assumed that this was a direct effect of the ban. Nevertheless, another shock acted upon the increase of coca cultivation. The paramilitary group *las FARC* and the Colombian State signed a peace treaty only one year after the ban was introduced, and this put an end to the longest insurgency in the world that lasted over 50 years. This event changed land dynamics deeply: it changed tenure and disputes increased over large territories while propitiating land grabbing for illegal purposes (Van Dexter and Visseren-Hamakers, 2019). As studied by Muñoz Mora et al. (2018) the lack of formalization of land property rights in the war against illicit crops increases the probability in the allocation of coca in these territories. Furthermore, the peace treaty introduced a series of programs and initiatives like the Programs for Development with a Territorial Approach (from spanish PDET) and an initiative to reintroduce into society the *ex-guerrilleros*. Therefore, as the ban and the peace treaty took place in a short distance of time from one another, the effects of both shocks were difficult to discern. For this reason, panel regression analysis, as well as other econometric methods, are better suited to investigate the actual contribution of certain policies, like aspersion on coca crops. Also the government of Duque simplistically concluded that the ban increased coca cultivation and initiated efforts to reintroduce aspersion with glyphosate. On January of 2022, the Colombian Court of Justice ruled that Duque's government was not authorized to reinstate the operations and emphasized the need to consult the communities involved

(Sedano, 2022). However, the political landscape shifted enormously when Gustavo Petro won the electoral contest on June 19 of 2022. The now elected president has a strong conviction that the aspersion with glyphosate is not the right policy to combat illegal drugs, he stated in 2019 on his Twitter account, that "Aerial aspersion with glyphosate is mortal, it is only an imposition from the United States Government captured in their agencies with commercial interests. To end with the cultivations of coca an agreement with the peasant communities and an agricultural reform is needed"[4].

2.4 Wind speed

In this study, wind speed is used as instrumental variable for aerial aspersion and in this section we motivate our choice. The availability of a well designed instrumental variable is, in fact, essential to properly deal with endogeneity issues and deliver more accurate regression results. Our choice originates from considering the guidelines for an effective and safe spraying (U.S. Department of State, 2002). Operations are, in fact, planned on a daily basis and only start after an assessment of whether the minimum required security and weather conditions are met. Missions are canceled if wind speed is greater than 10mph, if relative humidity is below 75 percent, or if temperature is over 32 degrees Celsius. Wind speed has to be ideally 3-4 m/s around the time when the spraying takes place, in order to reduce spray-drift as much as possible. In DIRAN (2010) it is clearly stated that "wind conditions are constantly monitored by the aircraft and if they are not within the parameters allowed, the mission is annulled or postponed". Therefore, we identify the strength of wind as a possible instrument. We expect this variable to meet both the relevance and the exclusion conditions. As the number of non windy days increases, the probability that aspersion takes place also increases and so does the area sprayed, so that the instrument is correlated with the endogenous variable. As reported in a document of the U.S. Department of State (2002), "in 1998 and 1999, spraying took place on an average of 125 days out of the year. During the other 240 days, the spray planes were grounded, with the majority of cancellations due to bad weather". As for the exclusion condition, it is unlikely that wind strength directly affects coca cultivation, as thunderstorm winds, tornado or other forms of damaging winds are not present in Colombia. Nevertheless, wind could be correlated with other climatic conditions, such as precipitations, which directly affect cultivation in general. As discussed in Section 3, in our modeling strategy we control for climatic variables, to be sure that the strength of

[4]Original text: "La aspersión aérea con glifosato es mortal, es solo una imposición del gobierno de EEUU capturado en sus agencias por intereses comerciales. Para acabar con cultivos de hoja de Coca lo que se necesita es acuerdos con las comunidades campesinas y reforma agraria"

the wind is not correlated with the error term.

As said in Section 1, wind strength is also employed in Rodriguez (2020) as an instrument for aerial aspersion to estimate the causal effect of spraying on child labor and education. We use the Global Land Data Assimilation System (GLDAS) from NASA which provides information at a 1-degree and 0.25-degree resolution from 1948 until present simulations of different land surface models. Our definition of the instrumental variable is, though, slightly different. In Rodriguez (2020), the instrumental variable is defined as the number of days in which wind speed is one standard deviation above the municipality's monthly average. However, these days might be days in which flying is perfectly reasonable if monthly average and standard deviation are small. Therefore, we rather prefer to identify a unique threshold and define the instrumental variable as the mean number of days (over the pixels of each municipality) in which the average daily wind speed is below a given value, set at 3 m/s. We observe that using either 3 or 4 m/s as thresholds gives rise to an instrument which correlates positively with the area sprayed, as these are probably commonly used thresholds to decide if operations should take place or not. Values below such thresholds are, in fact, considered as safer wind speed, because air-crafts can fly at higher altitudes to spray with a lower drift risk and reduce the threat of confrontation with the counter-groups on the ground[5]. As we observe a higher correlation between instrument and endogenous variable when the threshold is equal to 3 m/s rather than to 4 m/s, we decide to retain this first value in the analysis.

3 Methodology

The research question is addressed empirically by using a municipality level panel regression and instrumenting for the endogenous predictor aspersion. As deeply motivated in Section 2.4, we choose as instrument the mean number of days in each year and in each municipality, during which the wind speed is below 3 m/s. Exploiting time and space variability of the instrumental variable, we can safely make use of both municipality and year fixed effects. Municipality fixed effects enables us to control for all those variables that vary across municipalities but remain stable or change slowly over time. For example, size of the municipality, morphological characteristics, fertility, urbanization level, historical presence of State or armed groups, existence of trading routes, markets, experience in coca cultivation and processing. Time fixed effects, on the other hand, control for shocks affecting all municipalities in a given year, such as modifications in

[5]In 2013, not long before the program's suspension, *las FARC* guerrillas shot down two spray planes within the space of two weeks.

the world demand for cocaine or in the US funding for the war on drugs. In this regard, our instrumental variable presents an important advantage over the one used in Reyes (2014). Even though well engineered, his instrumental variable only varies in time and at regional level but not at municipal level, so that he can only include time and regional fixed effects in regression models but not municipality ones.

We implement a two-stage least-squares regression, where the first stage model is given by:

$$A_{nt} = \alpha + \beta I_{nt} + X'_{nt}\gamma + \delta_n + \lambda_t + \epsilon_{nt} \tag{3.1}$$

where A_{nt} is the number of hectares sprayed in municipality n and year t, I_{nt} is the instrumental variable, X_{nt} is a column vector of control variables, δ_n and λ_t are, respectively, municipality and year first stage fixed effects, α, β and γ are first stage parameters, and ϵ_{nt} is the first stage error term with zero mean. The second stage regression is, instead, given by:

$$Y_{nt} = \xi + \theta A_{nt} + X'_{nt}\phi + \omega_n + \rho_t + \mu_{nt} \tag{3.2}$$

where Y_{nt} is the number of coca hectares in municipality n at the end of year t, ω_n and ρ_t are, respectively, municipality and year second stage fixed effects, ξ, θ and ϕ are second stage parameters, and μ_{nt} is the second stage error term with zero mean. We further remark that, while the OLS estimate of θ might be biased due to correlation between the area sprayed and the error term, if I_{nt} is not correlated with μ_{nt}, the instrumental variable estimator[6] should produce unbiased estimates for the causal effect of aerial spraying on coca crops.

However, the fact that wind speed is uncorrelated with the error term in Equation (3.2) is not straightforward, as partially observed in Section 2.4. In fact, even if wind is not expected to directly impact the extension of coca cultivation, it might still be correlated with other variables that affect coca crops. In this regard, not only climatic variables are called into question. Also manual eradication might, for example, be correlated with coca cultivation, as well as with the instrumental variable. Coca eradication crews, in fact, require Anti-narcotics Police to protect them from illegal armed groups during eradication, and this is often ensured through helicopters controlling the area. Strong winds might then be correlated also with manual eradication, making crews protection more complicated. If variables correlated with both coca crops and wind speed are not included in the regression model, the instrumental variable estimator could still suffer from omitted variable bias. For this reason, and for robustness checking, we include these

[6]The instrumental variable estimator of θ is obtained by replacing A_{nt} in Equation (3.2), with its fitted value recovered from the OLS estimate of Equation (3.1).

variable as control variables in the vector X_{nt}, under some of the model specifications considered in Section 4.

4 Results

With the assumptions and elements documented in the previous section, we estimate different model specifications, whose results are displayed in Table 18. We estimate all models using robust standard errors, and to achieve this, we rely on the package *estimatr* in R. In particular, we use the function *lm_robust* and *iv_robust* for the OLS and two stage least squares regressions respectively.

Table 18. Panel Regression Results (2000-2015)

Variable	OLS 1	OLS 2	IV 1	IV 2	IV 3	IV 4
Aspersion	0.271***	0.073^{+}	0.262*	0.258*	0.251	0.283^{+}
	(0.037)	(0.037)	(0.119)	(0.117)	(0.155)	(0.164)
Rainfall				-0.085	-0.084	-0.126
				(0.334)	(0.313)	(0.328)
Manual Eradication					-0.085	-0.091
					(0.066)	(0.068)
Seizures					0.001	0.002
					(0.025)	(0.026)
Dismanteled Labs					15.030^{+}	13.562
					(7.796)	(8.365)
Credits					0.011	0.012
					(0.009)	(0.010)
Homicides						-0.032
						(0.088)
Displaced						0.014*
						(0.007)
Area	0.037***					
	(0.004)					
Municipality fixed effects	No	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Instrumental variable	No	No	Yes	Yes	Yes	Yes
Sample size	16072	16072	16072	16072	16072	16072
Adj. R-squared	0.271	0.542	0.489	0.492	0.509	0.489
F statistics	110.050***	3.841$^+$	3.052***	3.061***	3.146***	3.007***
Weak instruments			12.114***	14.136***	10.672**	10.129**
Wu-Hausman			3.740	3.730	2.826	3.491

(1) Significancy levels: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.1$.

(2) Robust Standard Errors are reported in parenthesis.

The first specification (OLS 1) in Table 18 does not consider the instrumental variable. It is mainly presented in order to compare results with those obtained by Reyes (2014) in a similar OLS specification. For this reason it does not contain municipality fixed effects[7] but controls for municipality area that obviously correlates with both coca crops and aspersion. The coefficient of aerial spraying is highly significant and positive, suggesting that an increase of 1 hectare in the area sprayed leads to an increase in the coca cultivated area of 0.271 hectares. This is quite similar to the value obtained by Reyes (2014), equal to 0.238. The second specification (OLS 2) adds municipality fixed effects and, thus, exclude municipality area to avoid perfect collinearity. The value of the coefficient of aerial aspersion changes drastically, testifying that its estimate in OLS 1 is biased due to correlation between aspersion and the error term. The coefficient in OLS 2 remains in any case positive but only significant at the 10% level. When using the instrumental variable (IV 1 to IV 4), the value of the coefficient sharply increases, showing that the OLS 2 estimator is downward biased. Similarly, also Reyes (2014) observes an increase when moving from the OLS to the instrumental variable estimator. Although his coefficient is about three times higher than ours and much more unstable under the different specifications. We hypothesize this might reveal that regional fixed effects and covariates used are not enough to guarantee the instrumental variable be uncorrelated with the error term. Municipality fixed effects are more suitable for this purpose. The different specifications IV 1 to IV 4 include various control variables in order to assure that the exclusion condition is met and to check the robustness of the estimates obtained for the coefficient of interest. Model IV 2 includes climatic variables. As already underlined, these variables may be correlated with both the instrument and the response variable and their exclusion might induce correlation between the instrument and the error term, thus mining the exclusion restriction. In particular, we include precipitations but exclude

[7] Reyes (2014) does not consider municipality fixed effects but includes fixed effects for the seven UNODC coca-growing regions.

temperature due to a correlation close to 1 between these two variables. The coefficient of aspersion does not modify sensibly, when compared with specification IV 1. Both estimates are close to 0.26 and significant at the 5% level. Model IV 3 further adds control variables related with other eradication, interdiction or subsidy policies. Manual eradication, as discussed in Section 2.4 may be correlated with both wind speed and coca crops. Finally, model IV 4 includes control variables that are generally associated with coca cultivation, such as the number of homicides, which captures incidents of violence and the presence of armed groups in the municipality, and the number of internally displaced population, which determines availability of cheap labor force. Both variables are used, for example, in Bogliacino and Naranjo (2012). In all specifications including instrumental variables, the coefficient of aspersion remains stable, even if in the last two specifications it is not significant or only significant at the 10% level. Its standard error, in fact, increases in these two models, probably due to collinearity issues.

Thus, our analysis seems to indicate that an increase in aspersion leads to an increase in coca crops. This same conclusion is provided by Reyes (2014); Moreno-Sanchez et al. (2003); Rincón-Ruiz and Kallis (2013); Bogliacino and Naranjo (2012) and Dion and Russler (2008). In particular, our estimates seem to indicate that as one more hectare of coca is sprayed, coca crops increases by 0,28 hectares. As for the coefficients of the other variables, some of them are significantly different from 0 in some of the specification. However, as we simply include them as control variables, they might be correlated with the error term and, thus, their coefficients might be biased. We finally remark that the first stage F-statistic (given in the Weak instrument row in Table 18) is highly significant and larger than the rule of thumb value of 10 in all specifications, thus confirming the relevance of the instrument. The Wu-Hausman test, which tests the null hypothesis that the aspersion is exogenous, does not provide evidence of correlation between this variable and the error term. This test is though known to have poor power in the presence of many covariates (or many fixed effects in our case).

5 Discussion

In this paper we explore the effect of aspersion on coca cultivation in Colombia during the period 2000-2015, using a panel regression with time and municipality fixed effects and an instrumental variable estimator to account for endogeneity of the policy variable. The main motivation for this paper is to contribute to the discussion about effective measures to reduce illegal crops in Colombia. Moreover, the Colombian experience can also be taken as a reference for other Latin American countries fighting illegal cultivation like Peru, Bolivia and Mexico. In fact, Colombian policies are already a reference for policymakers across Latin America, because of the diversity of measures that this country has used to fight illegal crops. According to our results, fumigation with glyphosate in Colombia further increases coca cultivation. The same conclusion can be found in a number of different papers on the same subject (Reyes, 2014; Moreno-Sanchez et al., 2003; Rincón-Ruiz and Kallis, 2013; Bogliacino and Naranjo, 2012; Dion and Russler, 2008). Other studies find, instead, opposite results but always conclude that the effect is very small and the costs are particularly high (Rozo, 2014; Abadie et al., 2015; Mejía et al., 2017; Cote, 2019). Beside this, aerial spraying has a wide range of negative impacts, making it an ineffective policy to reduce coca crops. There are testimonies and studies of the affected communities, that posit that aerial fumigation has reduced the economic opportunities by destroying legal crops, exposing people to potential health problems and damaging important natural environments. This policy is also driving people to displace out of their communities to find better opportunities. In consideration of these elements, there is sufficient evidence to claim that aerial fumigation is a cost-ineffective policy to curb the spread of coca crops in Colombia. Moreno-Sanchez et al. (2003) suggest that crop substitution programs are more effective to tackle this problem, furthermore, the collateral effects are expected to be positive and more importantly, do not affect negatively the local communities.

6 Conclusions and further developments

In conclusion, we have found that for each fumigated hectare of coca, the expected increase in the crop amounts to about 0.28 hectares. Therefore, we conclude that this measure has the opposite intended effect and is thus ineffective to combat illegal crops. This study can further be extended in many interesting ways. It has been claimed that one of the effects of fumigation has been to shift coca cultivation within protected areas, where spraying is forbidden, with a consequent negative environmental effect on zones of high biodiversity value. In our framework, it would be possible to separate, within each municipality, the extension of coca cultivation inside national parks from that outside national parks, and use two different regressions to evaluate whether fumigation has different effects on the two different areas. A larger effect in national parks would testify the so-called "balloon effect". Another enhancement of the present work would consist in accounting for heterogeneity and for spatial correlation. Spatial quantile regression for panel data could provide a valid analytical tool in this respect. Finally, since one of the reasons for resuming aerial spraying has been to reduce coca crops and, thus, protect forest, it would be crucial to investigate the extent of the effect of aerial spraying on deforestation, after controlling for coca crops. Up to our knowledge, econometric studies on this issue do not exist. Evidence of a negative effect on forests would provide one more strong argument against fumigation.

Bibliography

[1] Alberto Abadie, Maria C Acevedo, Maurice Kugler, and Juan Vargas. Inside the War on Drugs: Effectiveness and Unintended Consequences of a Large Illicit Crops Eradication Program in Colombia, 2015.

[2] John T Abatzoglou and A Park Williams. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42):11770–11775, 2016.

[3] Acuerdo de Paz. Acuerdo final para la terminación del conflicto y la construcción de una paz estable y duradera. 2016.

[4] A. Maarten Allers and Corine Hoeben. Effects of unit-based garbage pricing: a differences-in-differences approach. *Environmental and Resource Economics*, 45 (3):405–428, 2010.

[5] American Addiction Centers. Stimulant drug addiction: Types, Effects & Treatment Near Me, December 2021. URL https://americanaddictioncenters.org/stimulant-drugs.

[6] Jesús A Anaya, Víctor H Gutiérrez-Vélez, Ana M Pacheco-Pascagaza, Sebastián Palomino-Ángel, Natasha Han, and Heiko Balzter. Drivers of forest loss in a megadiverse hotspot on the pacific Coast of Colombia. *Remote Sensing*, 12(8): 1235, 2020.

[7] Gabriella Andreotti, Stella Koutros, Jonathan N Hofmann, Dale P Sandler, Jay H Lubin, Charles F Lynch, Catherine C Lerro, Anneclaire J De Roos, Christine G Parks, Michael C Alavanja, et al. Glyphosate use and cancer incidence in the agricultural health study. *JNCI: Journal of the National Cancer Institute*, 110(5): 509–516, 2018.

[8] Robert Annett, Hamid R Habibi, and Alice Hontela. Impact of glyphosate and

glyphosate-based herbicides on the freshwater environment. *Journal of Applied Toxicology*, 34(5):458–479, 2014.

[9] Luc Anselin. *Spatial econometrics: methods and models*. Springer Science & Business Media, 2013.

[10] Luc Anselin and Anil K Bera. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In Aman Ullah, editor, *Handbook of Applied Economic Statistics*, pages 237–290. CRC Press, 1998.

[11] Luc Anselin and Raymond JGM Florax. Small sample properties of tests for spatial dependence in regression models: Some further results. In Luc Anselin and Raymond J. G. M. Florax, editors, *New directions in spatial econometrics*, pages 21–74. Springer, 1995.

[12] Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58:277–297, 1991.

[13] María Alejandra Arias, Adriana Camacho, Ana María Ibáñez, Danuel Mejía, and Catherine Rodríguez. *Costos económicos y sociales del conflicto en Colombia:¿ Cómo construir un posconflicto sostenible?* Ediciones Uniandes-Universidad de los Andes, 2014.

[14] Dolors Armenteras, Nelly Rodríguez, Javier Retana, and Mónica Morales. Understanding deforestation in montane and lowland forests of the Colombian Andes. *Regional Environmental Change*, 11(3):693–705, 2011.

[15] Dolors Armenteras, Edersson Cabrera, Nelly Rodríguez, and Javier Retana. National and regional determinants of tropical deforestation in Colombia. *Regional Environmental Change*, 13(6):1181–1193, 2013.

[16] Dolors Armenteras, Josep María Espelta, Nelly Rodríguez, and Javier Retana. Deforestation dynamics and drivers in different forest types in Latin America: Three decades of studies (1980–2010). *Global Environmental Change*, 46:139–147, 2017.

[17] Dolors Armenteras, Cerian Gibbes, Jesús A Anaya, and Liliana M Dávalos. Integrating remotely sensed fires for predicting deforestation for REDD+. *Ecological applications*, 27(4):1294–1304, 2017.

[18] Heikki Astola, Tuomas Häme, Laura Sirro, Matthieu Molinier, and Jorma Kilpi. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, 223:257–273, 2019.

[19] Sofia Bajocco, Luca Salvati, and Carlo Ricotta. Land degradation versus fire: A spiral process? *Progress in Physical Geography*, 35(1):3–18, 2011.

[20] Sofia Bajocco, Antonella De Angelis, and Luca Salvati. A satellite-based green index as a proxy for vegetation cover quality in a Mediterranean region. *Ecological Indicators*, 23:578–587, 2012.

[21] Banco de Colombia. Posición astronómica y geográfica de Colombia, 2022. URL https://enciclopedia.banrepcultural.org/index.php/Posici%C3%B3n_astron%C3%B3mica_y_geogr%C3%A1fica_de_Colombia#:~:text=Colombia%20se%20extiende%20desde%20los,mismo%20meridiano%2C%20por%20el%20occidente.

[22] Vincent Bax and Wendy Francesconi. Environmental predictors of forest change: An analysis of natural predisposition to deforestation in the tropical Andes region, Peru. *Applied geography*, 91:99–110, 2018.

[23] Gary S Becker, Kevin M Murphy, and Michael Grossman. The market for illegal goods: the case of drugs. *Revista de Econom´Institucional*, 8(15):17–42, 2006.

[24] Colin Berry. Glyphosate and cancer: the importance of the whole picture. *Pest management science*, 76(9):2874–2877, 2020.

[25] Biesimci. Metodología, 2016. URL https://www.biesimci.org/index.php?id=89.

[26] Roger Bivand and Gianfranco Piras. Spatialreg: Spatial Regression Analysis, 2021. URL https://CRAN.R-project.org/package=spatialreg.

[27] Roger S Bivand and David WS Wong. Comparing implementations of global and local indicators of spatial association. *Test*, 27(3):716–748, 2018.

[28] Raoul Blackman and Fei Yuan. Detecting long-term urban forest cover change and impacts of natural disasters using high-resolution aerial images and LiDAR data. *Remote Sensing*, 12(11):1820, 2020.

[29] Francesco Bogliacino and Alberto J. Naranjo. Coca Leaves Production and Eradication: A General Equilibrium Analysis. *Economics Bulletin*, 32:382–397, 2012.

[30] Sibylla Brodzinsky. Deforestation soars in Colombia after Farc rebels' demobilization, July 2017. URL https://www.theguardian.com/world/2017/jul/11/colombia-deforestation-farc.

[31] Graham Brookes, Farzad Taheripour, and Wallace E Tyner. The contribution of glyphosate to agriculture and potential impact of restrictions on use at the global level. *GM crops & food*, 8(4):216–228, 2017.

[32] Augusto Castro-Nunez, Ole Mertz, Alexander Buritica, Chrystian C Sosa, and Stephanie T Lee. Land related grievances shape tropical forest-cover in areas affected by armed-conflict. *Applied Geography*, 85:39–50, 2017.

[33] Anthony Cawley, Cathal O'Donoghue, Kevin Heanue, Rachel Hilliard, and Maura Sheehan. The impact of extension services on farm-level income: An instrumental variable approach to combat endogeneity concerns. *Applied Economic Perspectives and Policy*, 40(4):585–612, 2018.

[34] Antonio L Cerdeira and Stephen O Duke. The current status and environmental impacts of glyphosate-resistant crops: a review. *Journal of environmental quality*, 35(5):1633–1658, 2006.

[35] Guanghua Chi, Han Fang, Sourav Chatterjee, and Joshua E. Blumenstock. Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3):e2113658119, 2022.

[36] Peter Christen. *The data matching process.* Springer, 2012.

[37] Climate Action Tracker. Climate Action Tracker: Colombia, 2021. URL https://climateactiontracker.org/climate-target-update-tracker/colombia/.

[38] Climate Home News. Colombia banks on forest economy to Deliver Climate Ambition leap, Jan 2021. URL https://www.climatechangenews.com/2021/01/14/colombia-banks-forest-economy-deliver-climate-ambition-leap/.

[39] Kox Cohen. Measuring the level of agreement: Kappa, 2022. URL https://www.sheffield.ac.uk/polopoly_fs/1.885169!/file/88_Kappa.pdf.

[40] Colombia Reports. European Space Agency to help Colombia Monitor deforestation, Mar 2018. URL https://colombiareports.com/european-space-agency-to-help-colombia-monitor-deforestation/.

[41] Marco Conedera, Damiano Torriani, Christophe Neff, Carlo Ricotta, Sofia Bajocco, and Gianni Boris Pezzatti. Using Monte Carlo simulations to estimate relative fire ignition danger in a low-to-medium fire-prone region. *Forest Ecology and Management*, 261(12):2179–2187, 2011.

[42] Russell G Congalton and Kass Green. *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press, 2019.

[43] Luisa Corrado and Bernard Fingleton. Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2):210–239, 2012.

[44] Juan Pablo Cote. The effect of interdiction on coca cultivation in Colombia. *Documento CEDE*, (2019-18), 2019.

[45] Emmanuel Da Ponte, Martina Fleckenstein, Patrick Leinenkugel, Amanda Parker, Natascha Oppelt, and Claudia Künzer. Tropical forest cover dynamics for Latin America using Earth observation data: a review covering the continental, regional, and local scale. *International Journal of Remote Sensing*, 36(12):3196–3242, 2015.

[46] DANE. Cuentas Nacionales, 2020. URL https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales.

[47] Liliana M Dávalos, Adriana C Bejarano, Mark A Hall, H Leonardo Correa, Angelique Corthals, and Oscar J Espejo. Forests and drugs: Coca-driven deforestation in tropical biodiversity hotspots. *Environmental science & technology*, 45 (4):1219–1227, 2011.

[48] Teresa de Miguel. México prohíbe el glifosato para frenar sus efectos nocivos en la salud, Jan 2021. URL https://elpais.com/mexico/2021-01-12/mexico-prohibe-el-glifosato-para-frenar-sus-efectos-nocivos-en-la-salud.html.

[49] Jennifer A Devine, David Wrathall, Nate Currit, Beth Tellman, and Yunuen Reygadas Langarica. Narco-cattle ranching in political forests. *Antipode*, 52(4): 1018–1038, 2020.

[50] Jennifer A Devine, David Wrathall, Bernardo Aguilar-González, Karina Benessaiah, Beth Tellman, Zahra Ghaffari, and Daria Ponstingel. Narco-degradation: Cocaine trafficking's environmental impacts in Central America's protected areas. *World Development*, 144:105474, 2021.

[51] Eric Dinerstein, David Olson, Anup Joshi, Carly Vynne, Neil D Burgess, Eric Wikramanayake, Nathan Hahn, Suzanne Palminteri, Prashant Hedao, Reed Noss, et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6):534–545, 2017.

[52] Michelle L Dion and Catherine Russler. Eradication efforts, the state, displacement and poverty: Explaining coca cultivation in Colombia during Plan Colombia. *Journal of Latin American Studies*, 40(3):399–421, 2008.

[53] DIRAN. Report by the Anti-Narcotics Direction of the Colombian National Police, 2010.

[54] Loïc Paul Dutrieux, Jan Verbesselt, Lammert Kooistra, and Martin Herold. Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia. *ISPRS Journal of Photogrammetry and Remote Sensing*, 107: 112–125, 2015.

[55] David Eckstein, Vera Künzel, Laura Schäfer, and Maik Winges. Global climate risk index 2020. *Germanwatch Available at: https://germanwatch. org/sites/germanwatch. org/files/20-2-01e% 20Global*, 20, 2019.

[56] Ecuador v. Colombia. volume VI, Annexes 34-55. 2012.

[57] EFSA. Glyphosate, 2022. URL https://www.efsa.europa.eu/en/topics/topic/glyphosate.

[58] El Tiempo. Policía Antinarcóticos Lanza Campaña Para Evitar Muerte de pilotos de Aviones Fumigadores, Mar 2007. URL https://www.eltiempo.com/archivo/documento/CMS-3487812.

[59] El Tiempo. Disidencias de las Farc duplican su número de hombres en solo 12 meses, May 2020. URL https://www.eltiempo.com/unidad-

investigativa/disidencias-de-las-farc-duplican-su-numero-de-hombres-en-armas-solo-12-meses-501426.

[60] El Tiempo. Cae Proyecto de Ley Que Buscaba Prohibir el uso del glifosato en colombia, Oct 2021. URL https://www.eltiempo.com/politica/congreso/congreso-cae-proyecto-de-ley-para-prohibir-el-uso-del-glifosato-623551.

[61] El Tiempo. El tortuoso camino para el retorno de la aspersión, Sep 2021. URL https://www.eltiempo.com/justicia/investigacion/aspersion-con-glifosato-asi-va-el-tortuoso-camino-para-el-retorno-620760.

[62] J Paul Elhorst. *Spatial econometrics: from cross-sectional data to spatial panels*. Springer, 2014.

[63] Paul Elhorst and Solmaria Halleck Vega. On spatial econometric models, spillover effects, and W. Ersa conference papers, 53rd ersa congress, palermo, italy, 2013.

[64] Environmental Investigation Agency. *Condenando el bosque: Ilegalidad y falta de gobernanza en la Amazonía colombiana*. 2019. URL https://content.eia-global.org/posts/documents/000/000/894/original/Condenando_el_Bosque.pdf?1561565558.

[65] ESA. Land Cover CCI Product User Guide Version 2. Technical Report, 2017. URL http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

[66] FAO. FRA 2000 ON DEFINITIONS OF FOREST AND FOREST CHANGE, 2000. URL https://www.fao.org/3/ad665e/ad665e02.htm.

[67] FAO. Country Profile - Colombia, 2015. URL https://www.fao.org/aquastat/en/countries-and-basins/country-profiles/country/COL.

[68] FAO. Colombia, 2016. URL http://www.fao.org/countryprofiles/index/en/?iso3=COL.

[69] James D Fearon and David D Laitin. Ethnicity, insurgency, and civil war. *American political science review*, 97(1):75–90, 2003.

[70] Leopoldo Fergusson, Santiago Saavedra, and Juan F Vargas. The perils of misusing remote sensing data: The case of forest cover. *Documento CEDE*, (15), 2020.

[71] Agostino Ferrara, Constantinos Kosmas, Luca Salvati, Antonietta Padula, Giuseppe Mancino, and Angelo Nolè. Updating the MEDALUS-ESA framework for worldwide land degradation and desertification assessment. *Land Degradation & Development*, 31(12):1593–1607, 2020.

[72] Carlotta Ferrara, Margherita Carlucci, Efstathios Grigoriadis, Piermaria Corona, and Luca Salvati. A comprehensive insight into the geography of forest cover in Italy: Exploring the importance of socioeconomic local contexts. *Forest Policy and Economics*, 75:12–22, 2017.

[73] RH Fraser, A Abuelgasim, and R Latifovic. A method for detecting large-scale forest cover change using coarse spatial resolution imagery. *Remote sensing of environment*, 95(4):414–427, 2005.

[74] Rong Fu, Lei Yin, Wenhong Li, Paola A Arias, Robert E Dickinson, Lei Huang, Sudip Chakraborty, Katia Fernandes, Brant Liebmann, Rosie Fisher, et al. Increased dry-season length over southern Amazonia in recent decades and its implication for future climate projection. *Proceedings of the National Academy of Sciences*, 110(45):18110–18115, 2013.

[75] Global Forest Watch. Colombia deforestation rates amp; statistics: GFW, 2021. URL https://bit.ly/3w0Rtls.

[76] Glyphosate Renewal Group. Glyphosate Renewal Group (GRG), Jun 2022. URL https://www.glyphosate.eu/fr/?gclid=CjwKCAjwtcCVBhA0EiwAT1fY7wO9CC5XcLae04sjZdv4_joCncHE-KzJ0FAW4kY8QI-FPSeUeBsA-xoCbxcQAvD_BwE.

[77] Gobierno de Colombia. Región Orinoquia de Colombiana, May 2020. URL https://www.colombia.co/pais-colombia/geografia-y-medio-ambiente/region-de-la-orinoquia/.

[78] André Braz Golgher and Paul R Voss. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, 4(3):175–205, 2016.

[79] Andrés González-González, Nicola Clerici, and Benjamin Quesada. Growing mining contribution to Colombian deforestation. *Environmental Research Letters*, 16(6):064046, 2021.

[80] Government of Ecuador. Application of Instituting Proceedings: Ecuador v. Colombia, Mar 2008.

[81] David Griggs, Mark Stafford-Smith, Owen Gaffney, Johan Rockström, Marcus C Öhman, Priya Shyamsundar, Will Steffen, Gisbert Glaser, Norichika Kanie, and Ian Noble. Sustainable development goals for people and planet. *Nature*, 495 (7441):305–307, 2013.

[82] G.U. n.193 19/08/2016. Decreto 9 agosto 2016. *Ministero della Salute*, 2016. URL https://www.gazzettaufficiale.it/eli/id/2016/08/19/16A06170/sg.

[83] Jing Guo, Peng Gong, Iryna Dronova, and Zhiliang Zhu. Forest cover change in China from 2000 to 2016. *International Journal of Remote Sensing*, 43(2): 593–606, 2022.

[84] M. Hansen, P. Potapov, B. Margono, S. Stehman, S. Turubanova, and A. Tyukavina. Response to Comment on "High-resolution global maps of 21st-century forest cover change". *Science*, 344(6187):981–981, 2014.

[85] Matthew C Hansen and Ruth S DeFries. Detecting long-term global forest change using continuous fields of tree-cover maps from 8-km advanced very high resolution radiometer (AVHRR) data for the years 1982–99. *Ecosystems*, 7(7):695–716, 2004.

[86] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.

[87] Frank Hanssen, David N Barton, Zander S Venter, Megan S Nowell, and Zofie Cimburova. Utilizing LiDAR data to map tree canopy for urban ecosystem extent and condition accounts in Oslo. *Ecological Indicators*, 130:108007, 2021.

[88] Robin Harding, Mounu Prem, Nelson A. Ruiz, and David L. Vargas. Buying a Blind Eye: Campaign Donations, Regulatory Enforcement, and Deforestation in Colombia. IAST Working Papers 22-136, Institute for Advanced

Study in Toulouse (IAST), 2022. URL https://ideas.repec.org/p/tse/iastwp/126766.html.

[89] Frank E Harrell et al. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, 2015.

[90] Liliana Marcela Henao Muñoz, Claudia Marsela Montes Rojas, and Manuel Hernando Bernal Bautista. Acute toxicity and sublethal effects of the mixture glyphosate (Roundup®Active) and Cosmo-Flux®411F to anuran embryos and tadpoles of four Colombian species. *Revista de Biologia Tropical*, 63(1):223–233, 2015.

[91] Carolin Hoffmann, Jaime Ricardo García Márquez, and Tobias Krueger. A local perspective on drivers and measures to slow deforestation in the Andean-Amazonian foothills of Colombia. *Land Use Policy*, 77:379–391, 2018.

[92] Jennifer S Holmes, Sheila Amin Gutiérrez De Piñeres, and Kevin M Curtin. *Guns, drugs, and development in Colombia*. University of Texas Press, 2010.

[93] IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 112, Jan 2017. URL https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Some-Organophosphate-Insecticides-And-Herbicides-2017.

[94] IDEAM. Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM DOCUMENTO metodológico Operación Estadística Monitoreo de la Superficie de Bosque Natural en Colombia, Jun 2019.

[95] International Monetary Fund. Colombia: Detailed assessments using the Data Quality Assessment Framework (DQAF), Oct 2006. URL https://www.elibrary.imf.org/view/journals/002/2006/356/article-A003-en.xml.

[96] International Union for Conservation of Nature. The IUCN Red List of Threatened Species, 2020. URL https://www.iucnredlist.org/.

[97] Dev Kar and Joseph Spanjers. Transnational crime and the developing world. *Washington: Global Financial Integrity*, 30:2019, 2017.

[98] Rodney J Keenan, Gregory A Reams, Frédéric Achard, Joberto V de Freitas, Alan Grainger, and Erik Lindquist. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology and Management*, 352:9–20, 2015.

[99] Christina M Kennedy, James R Oakleaf, David M Theobald, Sharon Baruch-Mordo, and Joseph Kiesecker. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology*, 25(3):811–826, 2019.

[100] Julie Le Gallo and Cem Ertur. Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980–1995. *Papers in regional science*, 82 (2):175–201, 2003.

[101] George Lindfield and John Penny. Chapter 2 - Linear Equations and Eigensystems. In George Lindfield and John Penny, editors, *Numerical Methods (Fourth Edition)*, pages 73–156. Academic Press, fourth edition edition, 2019.

[102] Kay Khaing Lwin, Tetsuji Ota, Katsuto Shimizu, and Nobuya Mizoue. Assessing the importance of tree cover threshold for forest cover mapping derived from global forest cover in Myanmar. *Forests*, 10(12):1062, 2019.

[103] Maurizio Marchi, Carlotta Ferrara, Giada Bertini, Silvano Fares, and Luca Salvati. A sampling design strategy to reduce survey costs in forest monitoring. *Ecological Indicators*, 81:182–191, 2017.

[104] Jeffrey G Masek, Daniel J Hayes, M Joseph Hughes, Sean P Healey, and David P Turner. The role of remote sensing in process-scaling studies of managed forest ecosystems. *Forest Ecology and Management*, 355:109–123, 2015.

[105] Mariano Hugo Masiokas, Antoine Rabatel, Andres Rivera, Lucas Ruiz, Pierre Pitte, Jorge Luis Ceballos, Gonzalo Barcaza, Alvaro Soruco, Francisca Bown, Etienne Berthier, et al. A review of the current state and recent changes of the Andean cryosphere. *Frontiers in Earth Science*, 8:1–27, 2020.

[106] Jeffrey A McNeely. Biodiversity, war, and tropical forests. *Journal of Sustainable Forestry*, 16(3-4):1–20, 2003.

[107] Jeffrey A McNeely. Conserving forest biodiversity in times of violent conflict. *Oryx*, 37(2):142–152, 2003.

[108] Ronald E McRoberts, Alexander C Vibrans, Christophe Sannier, Erik Næsset, Matthew C Hansen, Brian F Walters, and Débora V Lingner. Methods for evaluating the utilities of local and global maps for increasing the precision of estimates of subtropical forest area. *Canadian Journal of Forest Research*, 46(7):924–932, 2016.

[109] Kendra McSweeney, Erik A Nielsen, Matthew J Taylor, David J Wrathall, Zoe Pearson, Ophelia Wang, and Spencer T Plumb. Drug policy as conservation policy: narco-deforestation. *Science*, 343(6170):489–490, 2014.

[110] Erik Meijaard, Thomas M Brooks, Kimberly M Carlson, Eleanor M Slade, John Garcia-Ulloa, David LA Gaveau, Janice Ser Huay Lee, Truly Santika, Diego Juffe-Bignoli, Matthew J Struebig, et al. The environmental impacts of palm oil in context. *Nature plants*, 6(12):1418–1426, 2020.

[111] Daniel Mejía, Pascual Restrepo, and Sandra Rozo. On the Effects of Enforcement on Illegal Markets: Evidence from a Quasi-experiment in Colombia. *The World Bank Economic Review*, 31:570–594, 2017.

[112] Federico Meloni, Giannina Satta, Marina Padoan, Andrea Montagna, Ilaria Pilia, Alessandra Argiolas, Sara Piro, Corrado Magnani, Angela Gambelunghe, Giacomo Muzi, et al. Occupational exposure to glyphosate and risk of lymphoma: results of an Italian multicenter case-control study. *Environmental Health*, 20(1): 1–8, 2021.

[113] Juan P. Mendoza. Colombia's transition to peace is enhancing coca-driven deforestation. *Environmental Research Letters*, 15:104071, 2020.

[114] Juan P Mendoza. Colombia's transition to peace is enhancing coca-driven deforestation. *Environmental Research Letters*, 15(10):104071, 2020.

[115] Ministerio de Ambiente y Desarrollo Sostenible. Importancia de los bosques, Colombia tercer país de la región en cobertura boscosa, 2020. URL https://www.minambiente.gov.co/index.php/component/content/article/noticias/1210-el-uso-sostenible-de-los-bosques-prioridad-de-minambiente-513.

[116] Ministerio de Ambiente y Desarrollo Sostenible. Colombia, el segundo país más biodiverso del mundo, celebra el Día Mundial de la Biodiversidad, 2021.

URL https://minciencias.gov.co/sala_de_prensa/colombia-el-segundo-pais-mas-biodiverso-del-mundo.

[117] Pamela J Mink, Jack S Mandel, Jessica I Lundin, and Bonnielin K Sceurman. Epidemiologic studies of glyphosate and non-cancer health outcomes: a review. *Regulatory Toxicology and Pharmacology*, 61(2):172–184, 2011.

[118] Rocio Moreno-Sanchez, David S Kraybill, and Stanley R Thompson. An econometric analysis of coca eradication policy in Colombia. *World Development*, 31 (2):375–383, 2003.

[119] Juan Carlos Muñoz Mora, Santiago Tobón, and Jesse Willem d'Anjou. The role of land property rights in the war on illicit crops: Evidence from Colombia. *World Development*, 103:268–283, 2018.

[120] Soe W Myint, Patricia Gober, Anthony Brazel, Susanne Grossman-Clarke, and Qihao Weng. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, 115 (5):1145–1161, 2011.

[121] Pablo Jose Negret, Martine Maron, Richard A. Fuller, Hugh P. Possingham, James E.M. Watson, and Jeremy S. Simmonds. Deforestation and bird habitat loss in Colombia. *Biological Conservation*, 257:109044, 2021.

[122] Pablo Jose Negrete, Laura Sonter, James EM Watson, Hugh P Possingham, Kendall R Jones, Cesar Suarez, Jose Manuel Ochoa-Quintero, and Martine Maron. Emerging evidence that armed conflict and coca cultivation influence deforestation patterns. *Biological Conservation*, 239:108176, 2019.

[123] Christian Nellemann, Rune Henriksen, Patricia Raxter, Neville Ash, Elizabeth Mrema, et al. *The environmental crime crisis: threats to sustainable development from illegal exploitation and trade in wildlife and forest resources.* United Nations Environment Programme (UNEP), 2014.

[124] One Earth. Caqueta moist forests, 2022. URL https://www.oneearth.org/ecoregions/caqueta-moist-forests/.

[125] Hans Ole Ørka, Marie-Claude Jutras-Perreault, Erik Næsset, and Terje Gobakken. A framework for a forest ecological base map–An example from Norway. *Ecological Indicators*, 136:108636, 2022.

[126] Menaka Panta, Kyehyun Kim, and Chudamani Joshi. Temporal mapping of deforestation and forest degradation in Nepal: Applications to forest conservation. *Forest Ecology and Management*, 256(9):1587–1595, 2008.

[127] N. Pettorelli, H. Schulte toBühne, A. C. Shapiro, and P. Glover-Kapfer. Satellite Remote Sensing for Conservation. Conservation technology series 1(4), WWF, 2018.

[128] Peter V Potapov, SA Turubanova, A Tyukavina, AM Krylov, JL McCarty, VC Radeloff, and MC Hansen. Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive. *Remote Sensing of Environment*, 159:28–43, 2015.

[129] Steven V Price et al. *War and tropical forests: conservation in areas of armed conflict.* Food Products Press, 2003.

[130] Sara Reardon. FARC and the forest: Peace is destroying Colombia's jungle–and opening it to science. *Nature*, 558(7709):169–169, 2018.

[131] Angelika Rettberg and Juan Felipe Ortiz-Riomalo. Golden opportunity, or a new twist on the resource–conflict relationship: Links between the drug trade and illegal gold mining in Colombia. *World Development*, 84:82–96, 2016.

[132] Luis Carlos Reyes. Estimating the causal effect of forced eradication on coca cultivation in Colombian municipalities. *World Development*, 61:70–84, 2014.

[133] Alexander Rincón-Ruiz and Giorgos Kallis. Caught in the middle, Colombia's war on drugs and its effects on forest and people. *Geoforum*, 46:60–78, 2013.

[134] Kates W Robert, Thomas M Parris, and Anthony A Leiserowitz. What is sustainable development? Goals, indicators, values, and practice. *Environment: science and policy for sustainable development*, 47(3):8–21, 2005.

[135] Ana SL Rodrigues, Thomas M Brooks, Stuart HM Butchart, Janice Chanson, Neil Cox, Michael Hoffmann, and Simon N Stuart. Spatially explicit trends in the global conservation status of vertebrates. *PloS one*, 9(11):e113934, 2014.

[136] Claudia Rodriguez. The Effects of Aerial Spraying of Coca Crops on Child Labor, School Attendance, and Educational Lag in Colombia, 2008-2012. 2020.

[137] Jean Carlo Rodríguez-de-Francisco, Carlos del Cairo, Daniel Ortiz-Gallego, Juan Sebastian Velez-Triana, Tomás Vergara-Gutiérrez, and Jonas Hein. Post-conflict transition and REDD+ in Colombia: Challenges to reducing deforestation in the Amazon. *Forest Policy and Economics*, 127:102450, 2021.

[138] Sandra Rozo. On the Unintended Consequences of Anti-Drug Eradication Programs in Producing Countries (APPAM Working Paper), 2014. URL http://www.appam.org/assets/1/7/On_the_Unintended_Effects_of_Spraying.pdf.

[139] Pierre Salama. The economy of narco-dollars: from production to recycling of earnings. *International Journal of Politics, Culture, and Society*, pages 183–203, 2000.

[140] Luca Salvati, Flavia Ranalli, Margherita Carlucci, Achille Ippolito, Agostino Ferrara, and Piermaria Corona. Forest and the city: A multivariate analysis of peri-urban forest land cover patterns in 283 European metropolitan areas. *Ecological Indicators*, 73:369–377, 2017.

[141] Ana María Sánchez-Cuervo and T Mitchell Aide. Consequences of the armed conflict, forced human displacement, and land abandonment on forest cover change in Colombia: a multi-scaled analysis. *Ecosystems*, 16(6):1052–1070, 2013.

[142] Ana Maria Sanchez-Cuervo and T Mitchell Aide. Identifying hotspots of deforestation and reforestation in Colombia (2001–2010): implications for protected areas. *Ecosphere*, 4(11):1–21, 2013.

[143] Francisco Gutiérrez Sanín. Eradication in the time of Covid: The case of Colombia. *The International Journal on Drug Policy*, 83:102902, 2020.

[144] Christophe Sannier, Ronald E McRoberts, and Louis-Vincent Fichet. Suitability of Global Forest Change data to report forest cover estimates at national level in Gabon. *Remote Sensing of Environment*, 173:326–338, 2016.

[145] Rodrigo Sedano. La Corte Constitucional de Colombia da su negativa a la aspersión de cultivos ilícitos con glifosato, Jan 2022. URL https://www.france24.com/es/am%C3%A9rica-latina/20220120-colombia-corte-constitucional-prohibicion-glifosato.

[146] Steven E Sesnie, Beth Tellman, David Wrathall, Kendra McSweeney, Erik Nielsen, Karina Benessaiah, Ophelia Wang, and Luis Rey. A spatio-temporal analysis of forest loss related to cocaine trafficking in Central America. *Environmental Research Letters*, 12(5):054015, 2017.

[147] Jagadish Shukla, Carlos Nobre, and Piers Sellers. Amazon deforestation and climate change. *Science*, 247(4948):1322–1325, 1990.

[148] SINAP. Sistema de Parques Nacionales Naturales, 2020. URL http://www.parquesnacionales.gov.co/portal/es/sistema-de-parques-nacionales-naturales/.

[149] Daniela Smiraglia, Tomaso Ceccarelli, Sofia Bajocco, Luca Salvati, and Luigi Perini. Linking trajectories of land change, land degradation processes and ecosystem services. *Environmental research*, 147:590–600, 2016.

[150] Tony E Smith and Ka Lok Lee. The effects of spatial autoregressive dependencies on inference in ordinary least squares: a geometric approach. *Journal of geographical systems*, 14(1):91–124, 2012.

[151] Keith R Solomon, EJP Marshall, and Gabriel Carrasquilla. Human health and environmental risks from the use of glyphosate formulations to control the production of coca in Colombia: overview and conclusions. *Journal of Toxicology and Environmental Health, Part A*, 72(15-16):914–920, 2009.

[152] Dan-Xia Song, Chengquan Huang, Joseph O Sexton, Saurabh Channan, Min Feng, and John R Townshend. Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: Case studies from the Eastern United States and Central Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103: 81–92, 2015.

[153] Xiao-Peng Song, Chengquan Huang, Joseph O Sexton, Saurabh Channan, and John R Townshend. Annual detection of forest cover loss using time series satellite measurements of percent tree cover. *Remote Sensing*, 6(9):8878–8903, 2014.

[154] Emanuela Taioli, Luoping Zhang, et al. Weeding out inaccurate information on glyphosate-based herbicides and risk of non-Hodgkin lymphoma. *Environmental research*, 191:110140, 2020.

[155] Robert E Tarone. On the International Agency for Research on Cancer classification of glyphosate as a probable human carcinogen. *European Journal of Cancer Prevention*, 27(1):82–87, 2018.

[156] Beth Tellman, Nicholas R Magliocca, BL Turner, and Peter H Verburg. Understanding the role of illicit transactions in land-change dynamics. *Nature Sustainability*, pages 1–7, 2020.

[157] The Economist. Cocaine production in Colombia is at historic highs, Jul 2019. URL https://www.economist.com/the-americas/2019/07/06/cocaine-production-in-colombia-is-at-historic-highs.

[158] The Guardian. Former Farc commanders say they are returning to war despite 2016 peace deal, Aug 2019. URL https://www.theguardian.com/world/2019/aug/29/ex-farc-rebels-announce-offensive-despite-peace-deal-colombia-video.

[159] The Washington Post. The coronavirus has gutted the price of coca. It could reshape the cocaine trade., Jun 2020. URL https://www.washingtonpost.com/world/the_americas/coronavirus-coca-crash-bolivia-colombia-peru-latin-america/2020/06/09/8c7da42c-a11f-11ea-be06-af5514ee0385_story.html.

[160] Siriporn Thongprakaisang, Apinya Thiantanawat, Nuchanart Rangkadilok, Tawit Suriyo, and Jutamaad Satayavivad. Glyphosate induces human breast cancer cells growth via estrogen receptors. *Food and chemical toxicology*, 59:129–136, 2013.

[161] Robert Tropek, Ondřej Sedláček, Jan Beck, Petr Keil, Zuzana Musilová, Irena Šímová, and David Storch. Comment on "High-resolution global maps of 21st-century forest cover change". *Science*, 344(6187):981–981, 2014.

[162] UNODC. SUMMARY FACT SHEET – COLOMBIA COCA CULTIVATION SURVEY, 2020, 2021. URL www.unodc.org/documents/colombia/2021/Julio/Resumen_Ejecutivo_y_Hoja_de_Datos_en_Ingles.pdf.

[163] UNODC. Monitoreo de Territorios Afectados por Cultivos Il´2020, 2021.

[164] UNODC. World Drug Report 2021, 2021.

[165] U.S. Department of State. Chemicals Used for the Aerial Eradication of Illicit Coca in Colombia and Conditions of Application, Spray Report to Congress, 2002. URL https://2009-2017.state.gov/j/inl/rls/rpt/aeicc/c7470.htm.

[166] Kristina Van Dexter and Ingrid Visseren-Hamakers. Forests in the time of peace. *Journal of Land Use Science*, pages 1–16, 2019.

[167] Jeremy G Weber and Nigel Key. How much do decoupled payments affect production? An instrumental variable approach with panel data. *American Journal of Agricultural Economics*, 94(1):52–66, 2012.

[168] Dennis D Weisenburger. A review and update with perspective of evidence that the herbicide glyphosate (Roundup) is a Cause of non-Hodgkin lymphoma. *Clinical Lymphoma Myeloma and Leukemia*, 21(9):621–630, 2021.

[169] Gary M Williams, Robert Kroes, and Ian C Munro. Safety evaluation and risk assessment of the herbicide Roundup and its active ingredient, glyphosate, for humans. *Regulatory toxicology and pharmacology*, 31(2):117–165, 2000.

[170] World Bank. Forest area (% of land area), 2022. URL https://data.worldbank.org/indicator/AG.LND.FRST.ZS.

[171] World Wildlife Fund. A look at the natural world of Colombia, 2017. URL https://www.worldwildlife.org/magazine/issues/winter-2017/articles/a-look-at-the-natural-world-of-colombia.

[172] Worldometer. Colombia population, 2022. URL https://www.worldometers.info/world-population/colombia-population/.

[173] David J Wrathall, Jennifer Devine, Bernardo Aguilar-González, Karina Benessaiah, Elizabeth Tellman, Steve Sesnie, Erik Nielsen, Nicholas Magliocca, Kendra McSweeney, Zoe Pearson, et al. The impacts of cocaine-trafficking on conservation governance in Central America. *Global Environmental Change*, 63:102098, 2020.

[174] WWF. Oct 2017. URL https://wwflac.awsassets.panda.org/downloads/colombia_viva__informe_2017_1.pdf.

[175] Grissila Yustisia. Geographically Weighted Regression (GWR) Modelling with Weighted Fixed Gaussian Kernel and Queen Contiguity for Dengue Fever Case Data. *CAUCHY*, 5(1):15–19, 2017.

[176] Luoping Zhang, Iemaan Rana, Rachel M Shaffer, Emanuela Taioli, and Lianne Sheppard. Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: a meta-analysis and supporting evidence. *Mutation Research/Reviews in Mutation Research*, 781:186–206, 2019.