Universidad del Rosario | Facultad de Economía

SERIE
DOCUMENTOS
DE TRABAJO

No. 247

Mayo de 2020

The perils of misusing remote sensing data

The case of forest cover

Leopoldo Fergusson

Santiago Saavedra

Juan F. Vargas

Universidad del Rosario, Facultad de Economía
Cl. 12c #4-59. Bogotá, D. C., Colombia
PBX: (031) 2970200 ext. 4163

facultadeconomia@urosario.edu.co
https://www.urosario.edu.co/Facultad-de-Economia/Inicio/
https://repository.urosario.edu.co/handle/10336/20291

Title: The perils of misusing remote sensing data

The case of forest cover

One Sentence Summary: Forest increased with conflict according to automated data, but it declined using validated data; difference is due to misclassifying oil palm.

Authors: Leopoldo Fergusson[a] Santiago Saavedra[b] Juan F. Vargas[b]

Affiliations:[a]Facultad de Economía, Universidad de los Andes [b]Facultad de Economía, Universidad del Rosario.

Abstract

Research on deforestation has grown exponentially due to the availability of satellite-based measures of forest cover. One of the most popular is *Global Forest Change* (GFC). Using GFC, we estimate that the Colombian civil conflict increases 'forest cover'. Using an alternative source that validates the same remote sensing images in the ground, we find the opposite effect. This occurs because, in spite of its name, GFC measures *tree* cover, including vegetation other than native forest. Most users of GFC seem unaware of this. In our case, most of the conflicting results are explained by GFC's misclassification of oil palm crops as 'forest'. Our findings call for caution when using automated classification of imagery for specific research questions.

Main Text: Academic research on deforestation has grown exponentially in recent years, largely due to the availability of satellite-based measurements. For instance, through the the *Global Forest Change* (GFC) project, [1] provide freely accessible data available for the entire world, during a relatively long period, and at a high spatial resolution. By March 2020, the paper had been cited over 5,000 times, a telling indicator of the dataset's widespread influence and use.[1]

Remote-sensing global datasets are attractive because they provide consistent measures through time and space. This is particularly important in regions with armed conflict, where field-based measurement of outcomes of interest is often unfeasible. However, measures that rely on automated image classification might not capture the specific phenomena that researchers set to study [2]. This paper argues this is the case of GFC for *forest* cover, despite the dataset's name and the supporting paper's language.[2] In the paper's *Supplementary Materials* document, instead, the authors acknowledge that, in their study, "the term 'forest' refers to tree cover and not land use unless explicitly stated, e.g. 'forest land use'." (p. 2). The distinction is however meaningful. In specific settings *tree cover* likely confounds native forest with human-transformed vegetation that features a similar canopy density.

While GFC's failure to distinguish forest from plantations was already highlighted by [3], most researchers have continued to misuse GFC. In a review of the recent economics literature, we found 32 papers published in top economics journals or top-field development or environmental economics journals between 2015 and 2019 that cite [1]. Of those, 15 use GFC data in their main empirical specification but only 3 cite [3]'s critique. Further, from reading the 12 papers that seem to be unaware of what GFC actually measures, we conclude that the findings of 9 could be affected by correctly coding forest

[1] Google Scholar count accessed on March 19, 2020.

[2] [1]'s paper is written using terms such as *forest cover*, *forest change*, *forest loss/gain* and *forest ecosystems*.

change (see *Supplementary Materials* Table A-4).

To show the potentially misleading conclusions that an incorrect usage of GFC entails, we study the effect of conflict on forest cover in Colombia. Using GFC we find that conflict *increases* 'forest' cover. When we revisit our estimates using a different data source based on the same satellite input as GFC but that includes a field validation protocol carried out by experts, we find that conflict *decreases* forest cover. The alternative dataset is provided by Colombia's Institute of Hydrology, Meteorology and Environmental Studies (IDEAM from the Spanish acronym).[3]

Understanding the source of this discrepancy is of foremost policy importance. Does conflict prevent deforestation by discouraging economic activity or do some of the financing activities of illegal armed groups constitute a force of deforestation? We show that the conflicting results reflect that GFC captures plantations. In particular, most of the discrepancy is explained by GFC's inclusion of areas planted with oil palm (some of which are also particularly violent areas in our context).

It is of course difficult to distinguish between plantations and natural forests, "even using the most advanced remote-sensing technology" [4].[4] As a first step to address this issue, GFC recently released a map of plantations for seven countries, including Colombia.[5] The comprehensiveness of this plantations map, especially in some parts of the world where the quality of secondary data is questionable, is yet to be assured. Notably, it does not solve the discrepancy that we highlight in this paper.

Figure 1 presents a visual representation of the two datasets. They coincide in their coding of the Amazon and the Pacific rainforests (respectively, the South and West region of

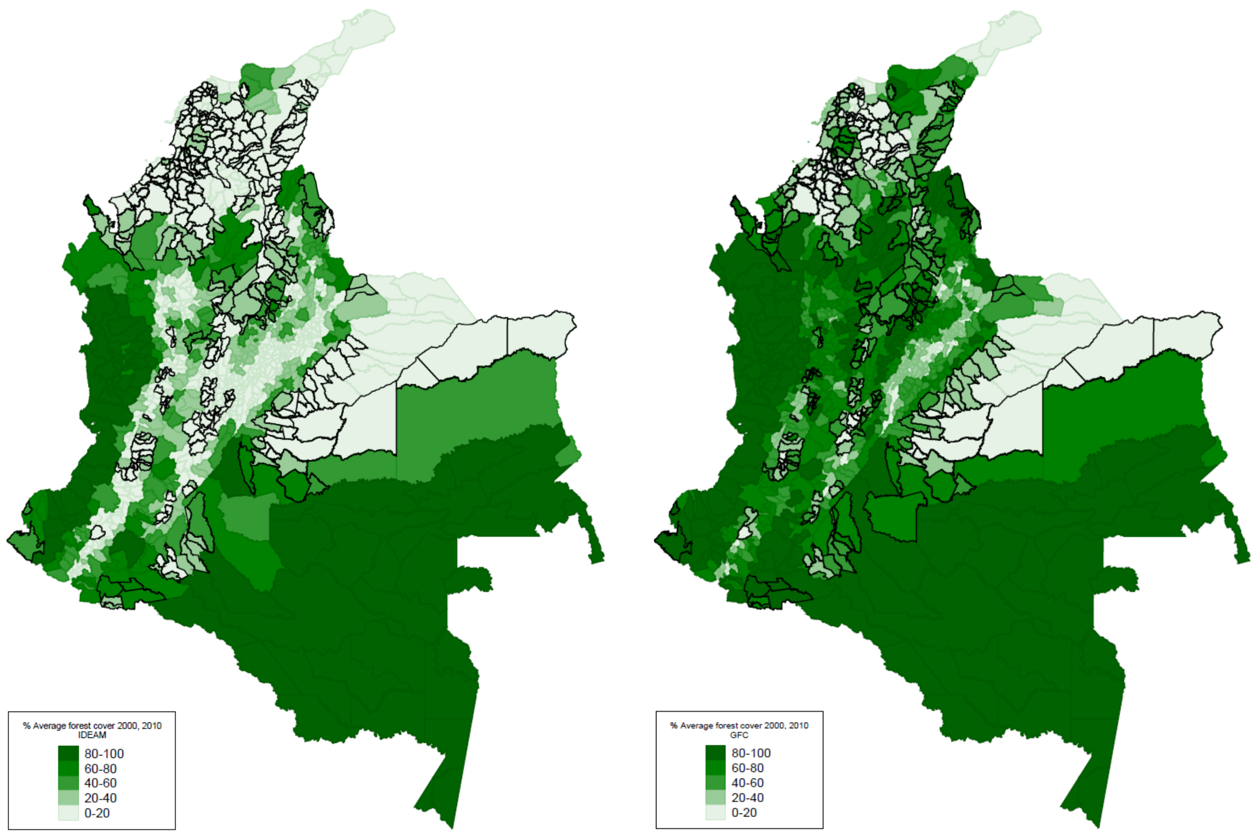
[3]See *Supplementary Materials* for details about data sources and their coding criteria.

[4]However, [5] shows promising results distinguishing plantations and natural forest for three Indian states using the new Sentinel satellites, that were launched on 2015.

[5]See `http://data.globalforestwatch.org/datasets/tree-plantations`.

the country). But marked differences can be observed in the Central and North regions, that feature high oil palm suitability (areas with a black border). Figure 2 further illus-

Figure 1: Forest cover and palm suitability



(a) IDEAM (b) GFC

Notes: The maps report municipal-level averages of forest cover in 2000 and 2010. Darker colors indicate a greater share of forest cover. In black contour are municipalities above the 75% of area suitable for palm cultivation.

trates the difference between GFC and IDEAM looking at two specific municipalities. The left-hand column is Espinal in the agriculture-intensive Andean region. The right-hand column is La Victoria, in the Amazon basin, where there is almost no agricultural activity. In Espinal, GFC identifies a large share of tree cover, mainly around river valleys. IDEAM however, identifies no forest cover. This is confirmed by the actual Google Earth satellite picture, which shows that the urbanized center is entirely surrounded by agricultural plots. In La Victoria, both GFC and IDEAM code the entire municipality as forest covered, and this is confirmed by the satellite image.

The literature that looks at the effect of conflict on deforestation consistently finds that conflict preserves forest cover [6, 7]. Our finding that conflict increases deforestation is thus puzzling. We posit that this effect of right-wing paramilitary activity on forest captures the deforestation patterns produced by the economic groups that have traditionally financed the paramilitary in Colombia.

Paramilitary groups exist since the 1970s, when the Colombian military armed and trained self-defense organizations with the purpose of fighting the extortion and ransom of communist guerrillas that were active since the mid 1960s. These groups were ruled illegal on 1989, but continued growing with the acquiescence of the military [8]. In the 1990s several splinter paramilitary armies joined forced under the umbrella organization of the *United Self-Defense Groups of Colombia* (AUC). Cattle-ranchers, landowners and drug lords provided most of the initial funding necessary to sustain the the AUC's expansion, which resulted in the forced displacement of millions in order to expropriate land and develop a model of resource extraction and extensive agriculture [9]. This likely led to a large forest loss.

Figure 2: Forest cover data and satellite images

GFC

IDEAM

Google Earth

(a) Espinal

(b) La Victoria

Notes: The maps show forest cover in 2000 according to GFC and IDEAM, and report the Google Earth map image of two municipalities in Colombia: Espinal (left) and La Victoria (right).

To assess the effect of paramilitary violence on deforestation we combine the data from GFC and IDEAM with several other datasets described in the *Supplementary Materials*, including data on the dynamics of the Colombian armed conflict and geo-referenced measures of oil palm suitability. We estimate panel models with municipality and time fixed effects, and allow for differential municipal trends parametrized by several geographic characteristics. The methods are thoroughly described in the *Supplementary Materials*.

Table 1 presents the main results. Each column uses a different combination of forest data source and time period. Each panel present the effect of paramilitary attacks using a different lag. Columns 1 and 2 use IDEAM's data: cross sections for 1990, 2000 and 2010 in Column 1 and dropping 1992 in Column 2 for comparison with GFC (available only from 2000). Columns 3 to 6 use GFC. Column 3 looks at the entire yearly panel from 2000 to 2010. Column 4 is directly comparable to Column 2, as it uses only the 2000 and 2010 cross sections of GFC. Columns 5 and 6 are equivalent to 3 and 4 but exclude the plantations identified by GFC for Colombia. Finally, Column 7 intersects IDEAM's and GFC's forest definitions: it codes a pixel as forest only if both datasets agree on this classification.

Paramilitary violence decreases IDEAM forest, and the effect is increasing as we increase the cumulative paramilitary activity lag. In contrast, with the exception of the contemporaneous effects, the effect of paramilitary attacks on forest cover is positive in any of the specifications based on GFC (Columns 3 to 6). The point estimates are unchanged when GFC's outcome accounts for plantations.

The negative effect of violence on contemporaneous tree cover that we observe using GFC is consistent with the immediate forest clearing that occurs after paramilitary violence takes place. Indeed, as noted paramilitary violence often represents specific economic interests, including oil palm, which incidentally takes around 26 months to grow mature. The measure that combines both datasets is clearly noisier, but suggests that, in accordance with the results found using IDEAM, paramilitary violence reduces forest cover (Column 7). This implies that the positive effect that results from using GFC is mostly driven by places that IDEAM screens out as forest after its validation procedure. In terms of the magnitude, a one-standard-deviation increase in the two-year lag measure of paramilitary attacks reduces IDEAM-defined forest cover by 0.43 to 0.62 percentage points.

Table 1: Paramilitary attacks effect on forest cover: GFC vs Ideam

	Ordinary least squares regression						
Dependent variable: *Share of forest cover*							
Forest base	Ideam	Ideam	GFC	GFC	GFC NP	GFC NP	Ideam, GFC
Years	1990, 2000, 2010	2000, 2010	2000 to 2010	2000, 2010	2000 to 2010	2000, 2010	2000, 2010
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Contemporaneous							
Paramilitary Attacks	-0.097***	-0.093*	-0.012	-0.019	-0.011	-0.017	0.017
	(0.033)	(0.052)	(0.0072)	(0.034)	(0.0073)	(0.035)	(0.049)
N	2,692	1,802	9,911	1,802	9,911	1,802	1,802
R-squared	0.98	0.99	1.00	1.00	1.00	1.00	0.99
Panel B: Average Two Year Lag							
Paramilitary Attacks	-0.92***	-1.32***	0.099***	0.34**	0.095***	0.32**	-0.072
	(0.31)	(0.45)	(0.034)	(0.15)	(0.034)	(0.15)	(0.29)
N	2,686	1,802	9,911	1,802	9,911	1,802	1,802
R-squared	0.98	0.99	1.00	1.00	1.00	1.00	0.99
Panel C: Average Four Year Lag							
Paramilitary Attacks	-1.44***	-2.28***	0.15***	0.47**	0.14***	0.45**	-0.18
	(0.47)	(0.65)	(0.053)	(0.21)	(0.053)	(0.21)	(0.41)
N	2,686	1,802	9,911	1,802	9,911	1,802	1,802
R-squared	0.98	0.99	1.00	1.00	1.00	1.00	0.99

Notes: All estimations include municipality and year fixed effects, and controls for geographic characteristics and rents. Standard errors clustered at the municipality level in parentheses. ***, **, * is significant at the 1%, 5% and 10% level respectively.

We conclude that the positive effect of violence on GFC-identified tree coverage is driven by areas with non-native vegetation, and that the lag at which the effect becomes significant is consistent with the growing cycle of an adult oil palm tree. We now explore more explicitly the hypothesis that oil palm confounds the effect of paramilitary attacks on forest cover when GFC is used to estimate this relationship. Table 2 presents results on forest cover by oil palm suitability.[6] This sheds light on the extent to which the effect of paramilitary violence on forest/tree cover depends on the oil palm suitability level, exploiting within-municipality variation. For simplicity, we focus on combinations of data source/sample period equivalent to those in columns 2, 3 and 4 of Table 1.

Column 1 of Table 2 suggests that, when using IDEAM, paramilitary violence decreases forest cover regardless of the level of palm suitability. The magnitude of the effect, as well as the precision of the estimates, are however higher in areas of no oil palm suitability, which are likely the areas with a higher prevalence of natural forests. The effects are also stronger with longer lags of paramilitary attacks. In contrast, when using GFC (columns 2 and 3), the contemporaneous effect of paramilitary attacks on tree cover is negative and significant, and larger in areas with high oil palm suitability or with no suitability (Panel A). But when we explore potential lagged affects (panels B and C) we obtain an effect that is positive and significant in medium or high suitability areas only. We posit that it is precisely in these areas where GFC captures plantations instead of native forest.

Our results point to the importance of exercising caution when using GFC. Despite [1]'s explicitly writing in their data description manual that GFC captures *tree-cover*, which in specific settings may confound native forest with certain crops, most researchers appear to be unaware of this feature of the data, and their conclusions are likely to be affected by this. Moreover, GFC's recently released map of plantations is, at least in our context, insufficient to account for the discrepancies we document. While global datasets based

[6] See Supplementary Materials equation 2.

on satellite imagery are an extraordinarily useful tool, researchers should be well aware of their features and limitations, or else they risk reaching misleading conclusions, with potentially problematic policy implications.

Table 2: Paramilitary attacks effect on forest cover by levels of palm suitability

Ordinary least squares regression			
Dependent variable: *% Forest cover*			
Forest base	Ideam	GFC	GFC
Years	2000, 2010	2000 to 2010	2000, 2010
	(1)	(2)	(3)
Panel A: Contemporaneous			
Attacks X High Palm	-0.018	-0.028*	-0.037
	(0.077)	(0.016)	(0.081)
Attacks X Medium Palm	-0.18	-0.0078	0.012
	(0.11)	(0.011)	(0.051)
Attacks X Low Palm	-0.0057	0.0031	0.0082
	(0.093)	(0.0093)	(0.046)
Attacks X No Palm	-0.12**	-0.012**	-0.020
	(0.061)	(0.0058)	(0.019)
N	7,208	39,644	7,208
R-squared	0.93	1.00	0.99
Panel B: Average Two Year Lag			
Attacks X High Palm	-0.38	0.19**	0.99**
	(0.71)	(0.087)	(0.39)
Attacks X Medium Palm	-2.02**	0.13*	0.54*
	(0.82)	(0.077)	(0.30)
Attacks X Low Palm	-1.32	0.14	0.23
	(1.04)	(0.087)	(0.39)
Attacks X No Palm	-2.66***	0.042	0.034
	(0.69)	(0.026)	(0.12)
N	7,208	39,644	7,208
R-squared	0.93	1.00	0.99
Panel C: Average Four Year Lag			
Attacks X High Palm	-1.26	0.33**	1.39***
	(1.28)	(0.15)	(0.50)
Attacks X Medium Palm	-2.80**	0.24*	0.96**
	(1.40)	(0.13)	(0.45)
Attacks X Low Palm	-3.33*	0.27*	0.050
	(1.96)	(0.14)	(0.56)
Attacks X No Palm	-4.69***	0.059	0.045
	(1.21)	(0.043)	(0.17)
N	7,208	39,644	7,208
R-squared	0.93	1.00	0.99

Notes: All estimations include population, rents, year-suitability and municipalities-suitability fixed effects controls. Standard errors clustered at the municipality level in parentheses. ***, **, * is significant at the 1%, 5% and 10% level respectively.

References

[1] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.

[2] M. Jain. The benefits and pitfalls of using satellite data for causal inference. *Review of Environmental Economics and Policy*, 14(1):157–169, 2020.

[3] R. Tropek, O. Sedlácek, J. Beck, P. Keil, Z. Musilová, I. Símová, and D. Storch. Comment on "high-resolution global maps of 21st-century forest cover change". *Science*, 344(6187):981, 2014.

[4] R. Petersen, E. D. Goldman, M. Weisse, and D. Aksenov. Forest loss pushes far beyond plantation boundaries in south america and africa. Technical report, World Resources Institute, 2016.

[5] W. Zhou. *High-resolution Remote Sensing to Identify Tree Plantations from Natural Forests and Agriculture in Southern India*. PhD thesis, 2019.

[6] R. Burgess, E. Miguel, and C. Stanton. War and deforestation in sierra leone. *Environmental Research Letters*, 10(9):095014, 2015.

[7] S. J. Singh N. Grima. How the end of armed conflicts influence forest cover and subsequently ecosystem services provision? an analysis of four case studies in biodiversity hotspots. *Land Use Policy*, 81:267–275, 2019.

[8] G. Duncan. *Conflicto armado, seguridad y construcción de paz en Colombia*. 40 años ciencia política. Universidad de los Andes, Facultad de Ciencias Sociales, Departamento de Ciencia Política, Centro de Estudios Socioculturales e Internacionales (CESO), 2010.

[9] J. Goebertus. Palma de aceite y desplazamiento forzado en zona bananera: "trayectorias" entre recursos naturales y conflicto. Thesis, Universidad de los Andes, 2008.

Acknowledgments: We are grateful to the Colombian Institute of Hydrology, Meteorology and Environmental Studies for help understanding the data. Carolina Vélez provided excellent research assistance.

The perils of misusing remote sensing data

The case of forest cover

Leopoldo Fergusson[*] Santiago Saavedra[†] Juan F. Vargas[†]

March 29, 2020

Supplementary Materials

Appendix A Data and summary statistics

A.1 Global Forest Change

Hansen et al. (2013)'s GFC dataset measures yearly gains and losses in tree coverage from 2000 to 2017 around the world. The data is generated using remote sensing techniques to process LANDSAT's satellite images. In particular, the authors develop an algorithm to detect the removal or recovery of plant biomass taller than 5m, with a pixel classified as "deforested" or "recovered" based on a 50% threshold.[1] GFC data include the percentage of tree cover per pixel for the year 2000 as well as its loss (or gain) each year between 2001 and 2017 for the entire world.

A.2 IDEAM

As a government agency, IDEAM follows the guidelines of the *United Framework Convention on Climate Change* (UNFCCC) and the *Intergovernmental Panel on Climate Change* (IPCC) to quantify natural forests and deforested areas in Colombia. To that end, IDEAM performs semi-automated digital processing of the same satellite imagery used by GFC (LANDSAT). The process is called "semi-automated" because, in contrast to GFC, the image classification algorithm is complemented by expert validation of randomly selected spots. This step helps excluding plantation areas where, for instance, palms or

[*]Facultad de Economía, Universidad de los Andes. E-mail: lfergusson@uniandes.edu.co.

[†]School of Economics, Universidad del Rosario. E-mails: santiago.saavedrap@urosario.edu.co & juan.vargas@urosario.edu.co

[1]The resolution of GFC's data is $30m \times 30m$ pixels.

fruit trees are grown (Galindo, Espejo, Rubiano, Vergara, & Cabrera, 2014). Currently, IDEAM data is available for four different period (1990, 2000, 2005, 2010).

There are also important differences between GFC and IDEAM regarding the image classification algorithm (see summary on Table A-1). For instance, IDEAM sets the following minimum criteria: a pixel canopy share of 30%, a canopy height of 5m, and at least 10,000 m^2 in area size. In contrast, GFC requires a canopy density threshold of 50% and sets no minimum area size. This creates type I and type II mismatches: GFC's forest areas are therefore a subset of IDEAM's when the forest is continuous, while GFC counts as forest small patches of forest that IDEAM ignores.

Figure A-1 presents a scatter plot that compares the two sources according to the share of forest cover by municipality, and includes the 45 degree line. With very few exceptions, we find that IDEAM reports *less* forest cover, and the discrepancy in some cases is very large. For instance, in some places where IDEAM reports no forest cover, GDC reports over 50%.

Generally speaking, authors that use GFC for their research seem unaware of the fact that this source does not necessarily capture *forest cover*. The year following the publication of Hansen et al. (2013)'s paper, Tropek et al. (2014) published a short memo highlighting that GFC failed to distinguish tropical forests from plantations, which implied "a substantial underestimate of forest loss and compromises its value for local policy decisions." Despite this call for caution, the misuse of GFC continues to generate potentially misleading conclusions. In a review of the recent economics literature, we found 32 papers published in top economics journals or top-field development or environmental economics journals between 2015 and 2019 that cite Hansen et al. (2013). Of those, 15 use GFC data in their main empirical specification (either as dependent or independent variable), but only 3 cite Tropek et al. (2014) to call for caution about their results. Moreover, from reading the 12 papers that seem to be unaware of what GFC actually measures, we conclude that the findings of 9 could be affected by correctly coding forest change (see Table A-4).

A.3 Other data sources

A.3.1 Oil palm

The Agricultural Rural Planning Unit (UPRA, from the Spanish acronym), classifies the area suitable for the commercial cultivation of oil palm in Colombia. UPRA uses a zoning methodology of aptitude for commercial crops at a scale of 1:100,000. This tool identifies the zones with aptitude for the establishment and development of the crop in a 3-scale classification: low, medium, and high. This classification is based on physical, ecosystem and socioeconomic variables. For each municipality, we compute the share of area within each palm suitability.

A.3.2 Paramilitary attacks

Our data on paramilitary activity comes from a detailed event-based dataset originally compiled by (Restrepo, Spagat, & Vargas, 2004), and updated through 2014 by Universidad del Rosario. This dataset codes violent events recorded in the *Noche y Niebla* reports from the NGO *Centro de Investigación y Educación Popular* (CINEP) of the Company of Jesus in Colombia, which provides a detailed description of the violent event, its date of occurrence, the municipality in which it took place, the identity of the perpetrator, and the count of the victims involved in the incident.[2] We extract from this source the count of paramilitary attacks per municipality and year.

A.3.3 Control variables

We include several municipality level controls. First, as a scale control, we include both the municipal population and the share of population settled in the urban part of the municipality (both from DANE, Colombia's Statistics Department). Second, we include various (time-invariant) geographical characteristics: the municipal surface area, its elevation, the average rainfall and the availability of water (rivers and lakes), the erosion and quality of the soil and the distance to the department's capital. All the geo-ecological controls come from IDEAM and the *Instituto Geográfico Agustín Codazzi* (IGAC). These are the official bureaus in charge of the climate and geographic monitoring, respectively.[3] Third, we include both the mining royalties received by the municipality and the municipal income tax revenue per 100,000 inhabitants (both variables come from the National Planning Department).

A.4 Summary statistics

Table A-2 presents summary statistics for our main variables. Panel A presents the statistics obtained from using IDEAM's data. Instead, panel B focuses on GFC. Both data sets suggest that forest cover has decreased in Colombia. While according to IDEAM only a quarter of the average municipal area is covered with forest, according to GFC this figure is around 50%. Panel C reports descriptive statistics for palm suitability and suggests that around 85% of the area of the average municipality is not suitable for oil palm cultivation. But there is substantial heterogeneity across municipalities, with some municipalities featuring over 70% of the area with medium or high suitability. Descriptive statistics for control variables are in reported on Table A-3.

[2] *Noche y Niebla* sources include "1. Press articles from more than 20 daily newspapers of both national and regional coverage. 2. Reports gathered directly by members of human rights NGOs and other organizations on the ground such as local public ombudsmen and, particularly, the clergy." ((Restrepo et al., 2004), p. 404). Notably, since the Catholic Church is present in even the most remote areas of Colombia, we have extensive coverage of violent events across the entire country.

[3] While the geographical characteristics are time-invariant, in the specifications that include the municipality fixed effects we include those interacted with the time dummy. This flexibly controls for differential time trends common to municipalities that have similar geo-ecological conditions.

Appendix B Empirical strategy and results

B.1 Empirical strategy

We examine the relationship between conflict and forest cover by estimating the following specification:

$$Y_{m,t} = \beta_1 Para_{m,t} + \beta_2 X_{m,t} + \delta_m + \delta_t + \sum_t \kappa'_m \bar{\omega}_t + \epsilon_{m,t} \quad (1)$$

Where $Y_{m,t}$ is the outcome of interest (namely forest or tree cover) measured as a share of the total area in municipality m at time t, for each of the approximately 1,000 Colombian municipalities. $Para_{m,t}$ are paramilitary attacks in municipality m during the years leading up to period t. We check the sensitivity of our results against alternative time-windows to calculate $Para_{m,t}$. To control for possible omitted variable bias, $X_{m,t}$ includes municipality-level time-varying controls like the municipal population and fiscal variables. We also include time (year) fixed-effects, δ_t, that absorb any events affecting the rate of forest cover change in all municipalities in Colombia, as well as municipality fixed effects δ_m, that control for any fixed, municipality-specific characteristics which may influence forest cover. For further robustness, we also include differential trends depending on fixed geographical characteristics of municipalities. Thus, $\bar{\omega}_t = 1$ in year t and zero otherwise, and κ'_m are time-invariant geographical characteristics of municipality m. We cluster the errors at the municipality level.

We also explore the evolution of forest/tree cover within municipality distinguishing between areas with high, medium, low or no suitability for oil palm cultivation. This specification is similar to equation 1 with the following adjustments. First, the dependent variable, $Y_{s,m,t}$, varies within municipalities across zones with palm suitability s. Second, $Para_{m,t}$ is now interacted with ϕ_s, which captures the share of forest cover to total municipal area in zones with oil palm suitability s. Third, $\delta_{m\times s}$ and $\delta_{t\times s}$ are municipality-suitability and time-suitability fixed effects, respectively. That is, our second specification is:

$$Y_{s,m,t} = \beta_0 + \beta Para_{m,t} * \phi_s + \delta_{m\times s} + \delta_{t\times s} + \sum_t X_{m,t}\phi_s + \sum_t \kappa'_m \bar{\omega}_t \phi_a + \epsilon_{m,t} \quad (2)$$

Table A-1: Differences between GFC and IDEAM forest classification

	Criteria to classify as forest	
Variable	GFC	IDEAM
Pixel resolution	30 m	30 m
Minimum canopy height	5 m	5 m
Tree density	50%	30%
Minimum area	0	10,000 m^2

Table A-2: Descriptive statistics: Main variables

	Obs	Mean	Stdv.	Min.	Max.
Panel A: *IDEAM Forest*					
% mun. area with forest (1990)	890	27.08	24.08	0.00	97.98
% mun. area with forest (2000)	901	25.03	23.53	0.00	96.46
% mun. area with forest (2010)	901	22.33	22.88	0.00	96.27
Panel B: *GFC Forest*					
% mun. area with forest (2000)	901	56.11	25.39	0.17	99.17
% mun. area with forest (2010)	901	53.86	25.10	0.09	98.82
% mun. area with plantations	901	0.35	2.09	0.00	33.68
Panel C: *Palm Suitability*					
% mun. area with high aptitude	901	2.64	8.10	0.00	72.64
% mun. area with medium aptitude	901	9.09	13.90	0.00	80.11
% mun. area with low aptitude	901	2.87	6.21	0.00	62.68
% mun. area with no aptitude	901	85.40	19.44	12.94	100.00
Panel D: *IDEAM Forest and Oil Palm Suitability (2000,2010)*					
% mun. area with forest and high apt.	1,802	0.111	0.63	0.00	9.30
% mun. area with forest and medium apt.	1,802	0.134	0.56	0.00	11.42
% mun. area with forest and low apt.	1,802	0.121	0.59	0.00	8.81
% mun. area with forest and no apt.	1,802	23.314	23.00	0.00	96.45
Panel E: *GFC Forest and Oil Palm Suitability (2000-2010)*					
% mun. area with forest and high apt.	9,911	0.64	2.49	0.00	31.51
% mun. area with forest and medium apt.	9,911	2.53	4.25	0.00	39.87
% mun. area with forest and low apt.	9,911	1.14	2.95	0.00	43.95
% mun. area with forest and no apt.	9,911	50.63	25.41	0.05	97.66
Panel F: *Paramilitary Attacks (PA)*					
Contemporaneous	9,911	0.05	0.28	0.00	10.10
Two Year Lag	9,911	0.13	0.47	0.00	7.50
Four Year Lag	9,911	0.13	0.39	0.00	6.75

Notes: In Panels A, B and C an observation is a municipality. In Panels D, E and F an observation is a municipality-year.

Table A-3: Descriptive statistics: Controls

	Obs	Mean	Stdv.	Min.	Max.
		Panel A: *Population*			
Total population (2000-2010)	9,911	44,583.31	259,191.96	1,133.00	7,363,782.00
Share urban pop. (2000-2010)	9,911	0.43	0.24	0.02	1.00
		Panel B: *Geography*			
Municipality area (km^2)	901	941.57	2,938.10	15.39	65,618.92
Average elevation (masl)	901	1,115.30	858.10	2.00	3,087.00
Average rainfall (mm)	901	1,987.89	1,074.59	160.00	9,200.00
Water availability index	901	0.60	0.09	0.35	1.00
Erosion index	901	0.39	0.20	0.00	1.00
Quality of soil index	901	0.33	0.15	0.00	1.00
Linear Distance to the state's capital	901	122.46	98.33	0.00	600.00
		Panel C: *Rents*			
Log royalties per 100K people (2000-2010)	9,911	2.94	3.76	0.00	13.31
Log tax income per 100K people (2000-2010)	9,911	10.39	1.74	0.00	13.59

Notes: In Panels A and C an observation is a municipality-year. In Panel B an observation is a municipality. See sub-section A.3.3 for data sources.

Table A-4: Papers using GFC and not correcting for plantations

Paper	Correcting for plantations might affect conclusions?	Notes
Alix-Garcia, Sims, and Yañez-Pagans (2015)	No	The looks at the effect of the Payment for Ecosystem Services Program on environmental and poverty outcomes in Mexico. It uses Normal Difference Vegetation Index (NDVI) as dependent variable. According to Kou et al. (2015), NDVI is able to distinguish between plantations and natural forest, consequently the conclusions are likely unaffected. GFC's data is only used to corroborate the main results.
Berazneva and Byker (2017)	Yes	The paper looks at the effect of deforestation on malaria in Nigeria and finds that forest loss increases the incidence of Malaria. Deforestation is likely underestimated by not distinguishing native forest from plantations.
Blackman, Goff, and Planter (2018)	Yes	The paper studies whether Forest Stewardship Council (FSC) certification affects deforestation in Mexico. However the measure of deforestation does not distinguish natural forests from plantations. This distinction is important because the FSC principles mention that well-managed plantations can ease the pressure on natural forests.[4]
Chervier and Costedoat (2017)	Yes	This paper studies the impact of a collective Payment for Environmental Services (PES) scheme on deforestation in Cambodia. As in the case of Blackman et al. (2018), it is clearly important to distinguish natural forest from plantations when assessing the effectiveness of such schemes.
Cook, Wright, and Andersson (2017)	Yes	The paper assess how the way local governments deal with environmental issues affect public perceptions in Bolivia and Guatemala. Forest cover enters as a control with the motivation that "it is reasonable to expect that the extent of forest resources in a municipality will impact the choices made by local officials in regard to forest governance." While this variable turns out not to be significant, and thus becomes a secondary variable in the paper's discussion, it is clear in the theoretical discussion that the authors intend to measure forest, not cultivated areas. Thus, the results could be sensitive to a correction of the dataset.
Damania, Russ, Wheeler, and Barra (2018)	Yes	This paper examines the tradeoffs between economic growth, deforestation, and biodiversity loss of building roads in the Democratic Republic of Congo. To the extent that part of the natural forest near roads is replaced by plantations such, then the paper is likely to underestimate the effect of roads on deforestation.

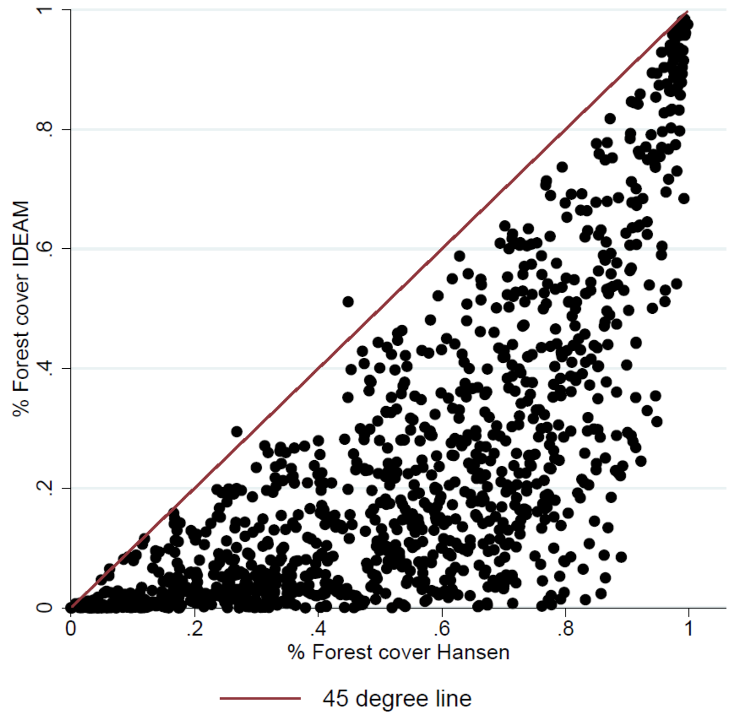
Continued on next page

[4] Principle 10 of FSC, available from `https://us.fsc.org/en-us/what-we-do/mission-and-vision` (last accessed Match 24, 2020).

Table A-4 – Papers using GFC and not correcting for plantations (Continues from previous page)

de Souza Cunha, Börner, Wunder, Cosenza, and Lucena (2016)	Yes	This paper measures the costs of conservation policies in the Brazilian Amazon. It uses GFC as a measure of national-level deforestation. Clearly, the interest is in understanding conservation, where plantations are different to natural forests, so the results are potentially sensitive to the correction.
Gallemore and Jespersen (2016)	No	The paper studies the support of donors to environmental organizations and uses GFC as a control variable. Although the authors do not show results of the model without controls, it is unlikely plantations corrections will affect the results.
Gibson (2018)	No	This paper studies the effect of deforestation on inequality in the Salomon Islands. In the Solomon Islands most forest cover is native forest so, inadvertently, the distinction does not seem to matter.
Jung and Polasky (2018)	Yes	This paper evaluates the Responsible Soy Project, a partnership to curb deforestation in the Brazilian Amazon.The main deforestation data is GFC, so the conclusions are sensitive to plantations misclassified as forest. (In practice it may not make a difference to the extent that the overwhelming majority of plantations in this area is soybean and this is not miss-classified by GFC).
Richards (2017)	Yes	The paper seeks to understand the expansion of agriculture in Matto Grosso, Brazil, and specifically the roles of economies of scale and spatial clustering. GFC is used to measure forest cover in 2001. Clearly, forest cover in this context is intended to measure actual forest, not crop trees.
Song (2018)	Yes	The paper estimates the change in value of ecosystem services using GFC net change in forest cover. By not considering the fact that native forest could be replaced by plantations that cannot be screened out by GFC, the change in the value of the ecosystem is probably biased. This because a plantation does not provide the same value of ecosystem services (Taki, Yamaura, Okabe, & Maeto, 2011)

Figure A-1: Share of forest cover GFC vs. IDEAM (2000)



Notes: The graphs represents a scatter plot of share of forest cover according to GFC and IDEAM data for the year 2000.

References

Alix-Garcia, J. M., Sims, K. R., & Yañez-Pagans, P. (2015). Only one tree from each seed? environmental effectiveness and poverty alleviation in mexico's payments for ecosystem services program. *American Economic Journal: Economic Policy*, *7*(4), 1–40.

Berazneva, J., & Byker, T. S. (2017). Does forest loss increase human disease? evidence from nigeria. *American Economic Review*, *107*(5), 516–21.

Blackman, A., Goff, L., & Planter, M. R. (2018). Does eco-certification stem tropical deforestation? forest stewardship council certification in mexico. *Journal of Environmental Economics and Management*, *89*, 306–333.

Chervier, C., & Costedoat, S. (2017). Heterogeneous impact of a collective payment for environmental services scheme on reducing deforestation in cambodia. *World Development*, *98*, 148–159.

Cook, N. J., Wright, G. D., & Andersson, K. P. (2017). Local politics of forest governance: why ngo support can reduce local government responsiveness. *World Development*, *92*, 203–214.

Damania, R., Russ, J., Wheeler, D., & Barra, A. F. (2018). The road to growth: Measuring the tradeoffs between economic growth and ecological destruction. *World Development*, *101*, 351–376.

de Souza Cunha, F. A. F., Börner, J., Wunder, S., Cosenza, C. A. N., & Lucena, A. F. (2016). The implementation costs of forest conservation policies in brazil. *Ecological economics*, *130*, 209–220.

Galindo, G., Espejo, O., Rubiano, J., Vergara, L., & Cabrera, E. (2014). Protocolo de procesamiento digital de imágenes para la cuantificación de la deforestación en colombia. *V2. 0. IDEAM, Bogota*.

Gallemore, C., & Jespersen, K. (2016). Transnational markets for sustainable development governance: The case of redd+. *World Development*, *86*, 79–94.

Gibson, J. (2018). Forest loss and economic inequality in the solomon islands: using small-area estimation to link environmental change to welfare outcomes. *Ecological economics*, *148*, 66–76.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853. Retrieved from http://science.sciencemag.org/content/342/6160/850

Jung, S., & Polasky, S. (2018). Partnerships to prevent deforestation in the amazon. *Journal of Environmental Economics and Management*, *92*, 498–516.

Kou, W., Xiao, X., Dong, J., Gan, S., Zhai, D., Zhang, G., ... Li, L. (2015). Mapping deciduous rubber plantation areas and stand ages with palsar and landsat images. *Remote Sensing*, *7*(1), 1048–1073.

Restrepo, J., Spagat, M., & Vargas, J. (2004). The dynamics of the columbian civil conflict: A new dataset. *Homo Oeconomicus*, *21*, 396–429.

Richards, P. (2017). It's not just where you farm; it's whether your neighbor does too. how agglomeration economies are shaping new agricultural landscapes. *Journal of Economic Geography*, *18*(1), 87–110.

Song, X.-P. (2018). Global estimates of ecosystem service value and change: Taking into account uncertainties in satellite-based land cover data. *Ecological Economics*, *143*, 227–235.

Taki, H., Yamaura, Y., Okabe, K., & Maeto, K. (2011). Plantation vs. natural forest: Matrix quality determines pollinator abundance in crop fields. *Scientific Reports*, *1*, 132.

Tropek, R., Sedlácek, O., Beck, J., Keil, P., Musilová, Z., Símová, I., & Storch, D. (2014). Comment on "high-resolution global maps of 21st-century forest cover change". *Science*, *344*(6187), 981.