Department of Economics- FEA/USP

The Impacts of Palm Oil Expansion on Deforestation and Economic Activity in the Eastern Amazon

Pedro Henrique B. de Barros

Ariaster B. Chimeli

WORKING PAPER SERIES Nº 2025-03

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2025-03

The Impacts of Palm Oil Expansion on Deforestation and Economic Activity in the Eastern Amazon

Pedro Henrique Batista de Barros (p.batista-de-barros@exeter.ac.uk)

Ariaster Chimeli (chimeli@usp.br)

Research Group: [CEMAR; NEREUS]

Abstract:

In recent years, the Brazilian government has designed policies to promote the palm oil industry and forest protection, limiting oil palm plantations to already degraded areas. As a consequence, oil palm crops have increased rapidly in the eastern Amazon region and contributed to a low-carbon energy transition. However, little is known about the effectiveness of these policies in avoiding oil palm-induced deforestation. This paper estimates the impact of oil palm plantations on deforestation and nightlight intensity, a proxy for less land-intensive economic activities that could contribute further to forest protection. We do so in two steps. First, we combined optical spectral bands from Landsat-8 and radar backscatter values from Sentinel-1 to produce a more accurate map of oil palm plantations with a random forest machine learning algorithm. Next, we used the maximum agro-climatically attainable palm oil yield from the Global Agro-Ecological Zoning (GAEZ) as an instrument for oil palm expansion between 2014 and 2020, and estimated the impact of the crop on deforestation and nightlights. Oil palms expanded mainly on pastures, but also contributed to deforestation. We do not find any evidence that the crop stimulates less land-intensive economic activities.

Keywords: Oil Palm; Deforestation; Amazon; Remote Sensing.

JEL Codes: Q15, Q23, Q28, Q56.

The Impacts of Palm Oil Expansion on Deforestation and Economic Activity in the Eastern Amazon

Pedro Henrique Batista de Barros* Ariaster Baumgratz Chimeli†

Abstract

In recent years, the Brazilian government has designed policies to promote the palm oil industry and forest protection, limiting oil palm plantations to already degraded areas. As a consequence, oil palm crops have increased rapidly in the eastern Amazon region and contributed to a low-carbon energy transition. However, little is known about the effectiveness of these policies in avoiding oil palm-induced deforestation. This paper estimates the impact of oil palm plantations on deforestation and nightlight intensity, a proxy for less land-intensive economic activities that could contribute further to forest protection. We do so in two steps. First, we combined optical spectral bands from Landsat-8 and radar backscatter values from Sentinel-1 to produce a more accurate map of oil palm plantations with a random forest machine learning algorithm. Next, we used the maximum agro-climatically attainable palm oil yield from the Global Agro-Ecological Zoning (GAEZ) as an instrument for oil palm expansion between 2014 and 2020, and estimated the impact of the crop on deforestation and nightlights. Oil palms expanded mainly on pastures, but also contributed to deforestation. We do not find any evidence that the crop stimulates less land-intensive economic activities.

Keywords: Oil Palm; Deforestation; Amazon; Remote Sensing.
JEL codes: Q15, Q23, Q28, Q56.

*Postdoctoral Research Fellow in the Land, Environment, Economics and Policy Institute (LEEP) at the University of Exeter Business School Xfi Building, Rennes Drive, Exeter, Devon, EX4 4PU, United Kingdom. Email: p.batista-de-barros@exeter.ac.uk

†Full Professor in the Department of Economics at the University of São Paulo, Av. Prof. Luciano Gualberto, 908, Butantã, São Paulo, São Paulo State, Brazil. Email: chimeli@usp.br

1 Introduction

Palm oil is the most consumed and exported vegetable oil in the world and is used mainly as food and in biodiesel production (Villela et al. 2014; Chong et al. 2017). The growing world demand for this commodity reflects a high production potential, a low production cost, and incentives to replace fossil fuels with biofuels, resulting in a rapid expansion in palm plantations (Xu et al. 2021).

Palm oil production is mainly located in tropical Southeast Asian countries such as Indonesia and Malaysia. Brazil is the 10th largest producer in the world, with most production and expressive growth in the cultivated area concentrated in the state of Pará. More specifically, in the period 2004-2014, Pará increased its cultivation area by more than 200%, reaching 2190 km^2 in extension and a production level of 900,000 tons of palm oil per year. This growth consolidated the state as the main producer in the country with more than 95% of national production (Villela et al. 2014; Carvalho et al. 2015; Benami et al. 2018; Nahum, Santos, and Santos 2020; Almeida, Vieira, and Ferraz 2020). In any case, the cultivated area is still small compared to the country's productive potential. This fact is mainly due to high production costs, a learning curve for oil palm cultivation in its early stages, and problems in the sector's governance model (Englund et al. 2015; Benami et al. 2018; Brandão et al. 2021).

Recognizing the strategic importance of palm oil, Brazil has implemented several programs to encourage its production. These initiatives include the National Program for the Production and Use of Biodiesel (*Programa Nacional de Produção e Uso do Biodiesel* - PNPB), the Agroecological Zoning of Palm Oil Cultivation (*Zoneamento Agroecológico da Cultura de Palma de Óleo* - ZAE), the Sustainable Palm Oil Production Program (*Programa de Produção Sustentável de Óleo de Palma* - PPSP), and the Pronaf Eco Palm Oil program (*Pronaf Eco Dendê*) (Carvalho et al. 2015; Englund et al. 2015; Lameira, Vieira, and Toledo 2016; Nahum, Santos, and Santos 2020; Brandão et al. 2021). The ZAE, PPSP, and Pronaf Eco Dendê, in particular, aim to promote oil palm cultivation in degraded areas while preventing crop-driven deforestation. Despite the economic benefits associated with palm oil expansion, concerns have been raised about its possible environmental consequences, particularly the illegal clearing of tropical forests (Koh and Wilcove 2008; Englund et al. 2015; Xu et al. 2021; Brandão et al. 2021).

The recent expansion of oil palms in the eastern Amazon has raised significant environmental concerns due to the existence of highly biodiverse forests in the region, which can be converted into plantation areas (Carvalho et al. 2015). Although the existing literature indicates that oil palm cultivation has been replacing forest areas in the Brazilian Amazon (Carvalho et al. 2015; Lameira, Vieira, and Toledo 2016; Furumo and Aide 2017; Benami et al. 2018; Almeida, Vieira, and Ferraz 2020), to the best of our knowledge, there are no articles that estimate the causal contribution of palm oil expansion to the trade-off between economic activity and deforestation in the region. That is, little is known about whether oil palm is a driver of deforestation or simply occupies areas that would be deforested anyway. This reflects the fact that the expansion of crops is normally associated with several economic, political, and social factors that make it difficult to measure their causal effects on deforestation (Edwards 2018; Kubitza and Gehrke 2018; Cisneros, Kis-Katos, and Nuryartono 2021), and because spatially disaggregated data covering the recent expansion of palm oil in Brazil are still scarce.

In other words, the identification of the causal effect of palm oil expansion on deforestation faces two important challenges: data scarcity and the endogeneity problem. To deal with the first challenge, we follow the pioneering work of Foster and Rosenzweig (2003) and Burgess et al. (2012) and use both optical spectral bands from Landsat-8 and radar backscatter data from Sentinel-1 in a 30 x 30 meter spatial resolution mapping process. We then employed Machine Learning algorithms to analyze satellite images and map oil palm expansion and deforestation in the Brazilian Eastern Amazon during the 2014-2020 period. A recent dataset

on oil palm plantations in Brazil was released by MapBiomas; however, our approach differs by integrating both optical and radar satellite data in the mapping process (Souza et al. 2020).[1] Studies such as Chong et al. 2017 and Xu et al. 2021 highlight the superior performance of this combined approach for oil palm mapping, as optical and radar sensors operate based on different physical and electromagnetic principles, capturing complementary information. Additionally, radar signals can penetrate cloud cover, a significant advantage for mapping in tropical regions where persistent cloudiness is common.

Next, we investigate the impact of oil palm on deforestation and the local economy. Because palm plantations are concentrated in only four large municipalities, our identification strategy concentrates on an indicator of economic impacts at the pixel level. More specifically, we follow Henderson, Storeygard, and Weil (2012) and use satellite data from the NASA/NOAA Visible Infrared Imaging Radiometer Suite (VIIRS) on nighttime lights as a proxy for more urban and less land-intensive economic activity. We then address the endogeneity problem by exploring the fact that each pixel differ exogenously in its productive potential for oil palm cultivation. In particular, we instrumentalize palm oil expansion using the maximum agro-climatically attainable oil palm yield from the Global Agro-Ecological Zoning (GAEZ) database calculated by the United Nation's Food and Agriculture Organization (FAO). Therefore, we compare pixels that were converted to oil palm with those that were not in order to estimate the causal impact of the crop on deforestation and local economic activity.

This effort is particularly relevant in light of recent developments in deforestation-free supply chain regulations, such as those introduced by the European Union and the United Kingdom, which require the implementation of robust and reliable monitoring systems for commodities such as palm oil. Our approach could be used for due diligence processes, allowing producers and suppliers to verify and demonstrate that their products are not sourced from deforested areas, thus ensuring compliance with global sustainability standards and contributing to environmental conservation efforts (Oliveira et al. 2024).

To summarize our results, we used a Random Forest algorithm that produced an overall classification accuracy of 94.53% and 95.53% for 2014 and 2020, respectively, which is much superior to the accuracy presented by the oil palm literature for the Amazon. From the land use and land cover transition analysis, we observed that oil palm cultivation in the region expanded significantly, increasing from 1,074 km^2 to 1,849 km^2—an overall growth of 72.16%. In particular, 156.88 km2 (20.24%) of this expansion directly replaced vegetation cover pixels. Our instrumental variable approach then indicated that oil palm plantations indeed expanded over forested land and their vicinity experienced a reduction in nightlight intensity, our proxy for more urban and less land-intensive economic activity.

Our work belongs to a growing body of economic literature that examines the causal impacts of palm oil expansion in tropical developing countries[2] and that leverages satellite-based remote sensing data to map deforestation and economic activity[3] or to analyze broader economic phenomena.[4] Additionally, this paper engages with the extensive body of research on the environmental consequences of oil palm expansion worldwide and in Brazil in particular.[5]

This paper is structured into six sections, in addition to this introduction. Section 2 provides background information on palm oil, while Section 3 details the dataset. Sections 4 and 5 outline

1. Lameira, Vieira, and Toledo (2016), Furumo and Aide (2017), and Almeida, Vieira, and Ferraz (2020) also contribute to oil palm mapping, but do not use the integrated approach we adopt here.

2. Edwards (2018), Kubitza and Gehrke (2018), Cisneros, Kis-Katos, and Nuryartono (2021), Krishna and Kubitza (2021), Hellmundt, Cisneros, and Kis-Katos (2024), Abman and Lundberg (2024), and Kraus, Heilmayr, and Koch (2024).

3. Foster and Rosenzweig (2003), Burgess et al. (2012), and Henderson, Storeygard, and Weil (2012).

4. Donaldson and Storeygard (2016).

5. See for example Koh and Wilcove (2008). For the Brazilian case, see Villela et al. (2014), Englund et al. (2015), Carvalho et al. (2015), Benami et al. (2018), and Brandão et al. (2021).

the methodological approach and present the results, respectively. Finally, Section 6 concludes with final considerations and policy implications.

2 Background

Originating from Africa, the oil palm is a perennial crop with a life cycle of approximately 25 years, and its trees can grow up to 20 meters in height. Due to these characteristics, it resembles a forest more than typical agricultural crops. Palm oil is typically cultivated in monoculture plantations that exhibit distinct and uniform geometric patterns, which facilitates its identification in satellite imagery. The crop thrives in humid tropical climates with abundant rainfall, high solar radiation, and temperatures ranging between 24°C and 32°C (Corley and Tinker 2008).

The production of vegetable oil from oil palm has significant potential, mainly due to its exceptionally high productivity, which can reach up to 368 tons per square kilometer. In comparison, soy has a much lower productive potential of just 42 tons per square kilometer (Carvalho et al. 2015; Englund et al. 2015). This high yield, combined with low production costs, has driven the exponential growth in global demand for palm oil (Xu et al. 2021). In Brazil, oil palm cultivation began in the 1970s and experienced rapid expansion in the 2000s, fueled by its growing demand in the food, cosmetics, and biofuel industries (Villela et al. 2014; Carvalho et al. 2015; Almeida, Vieira, and Ferraz 2020). The crop is particularly well-suited to the Amazon region, especially Pará, thanks to its favorable soil and climatic conditions, as well as the availability of suitable areas for further expansion.

The growth in palm oil production in Brazil was significantly driven from the 2000s onward by key government initiatives. These included the establishment of the PNPB in 2004, the ZAE under Decree No. 7.172/2010, and the PPSP in 2010 (Lameira, Vieira, and Toledo 2016; Nahum, Santos, and Santos 2020; Brandão et al. 2021). The PNPB, launched in 2004, aimed to increase biodiesel production, reduce greenhouse gas emissions, and promote regional development. Palm oil, with its high productivity, presents significant potential to become a key resource for the expansion of biodiesel production in Brazil (Carvalho et al. 2015; Nahum, Santos, and Santos 2020).

The ZAE and PPSP initiatives were established to regulate the expansion of oil palm cultivation while ensuring its social and environmental sustainability. The ZAE identified approximately 130,000 km^2 of deforested areas as suitable for cultivation—roughly 300 times the size of the current planted area. Its primary goals are to promote inclusive and sustainable regional economic development and to encourage the substitution of renewable energy sources for fossil fuels. Building on the ZAE framework, the PPSP was launched to further regulate palm oil expansion by restricting cultivation to degraded and deforested areas cleared before 2008. This approach incentivizes the recovery of degraded land, generating both social and environmental benefits (Carvalho et al. 2015; Benami et al. 2018; Brandão et al. 2021).

Driven by these institutional incentives, a new frontier of oil palm expansion emerged in Pará from 2010 onward, mainly replacing degraded pastures in the northeast of the state, a region with some of the most favorable conditions for palm cultivation in Brazil (Almeida, Vieira, and Ferraz 2020). Between 2010 and 2014, the area under cultivation increased by approximately 200%. However, it is important to note that about 60% of this expansion occurred near forested areas, raising environmental concerns. This is because the deforestation associated with oil palm expansion tends to disproportionately impact adjacent forest regions (Benami et al. 2018).

We concentrated our analysis in the municipalities of Acará, Moju, Tailândia, and Tomé-Açu, known as the oil palm pole in Brazil (Figure 1). Together, they cover an area of 23,014.36 km^2, of which a significant portion is located in the ZAE. These municipalities are the main producers of palm oil in Brazil, accounting for approximately 74.1% of the production in Pará and 72.5%

of the total palm oil production in the country in 2023 (Brazilian Institute of Geography and Statistics (IBGE) 2025). They are distinguished by their socioeconomic and agricultural dynamism, particularly in the Northeastern region of the state.

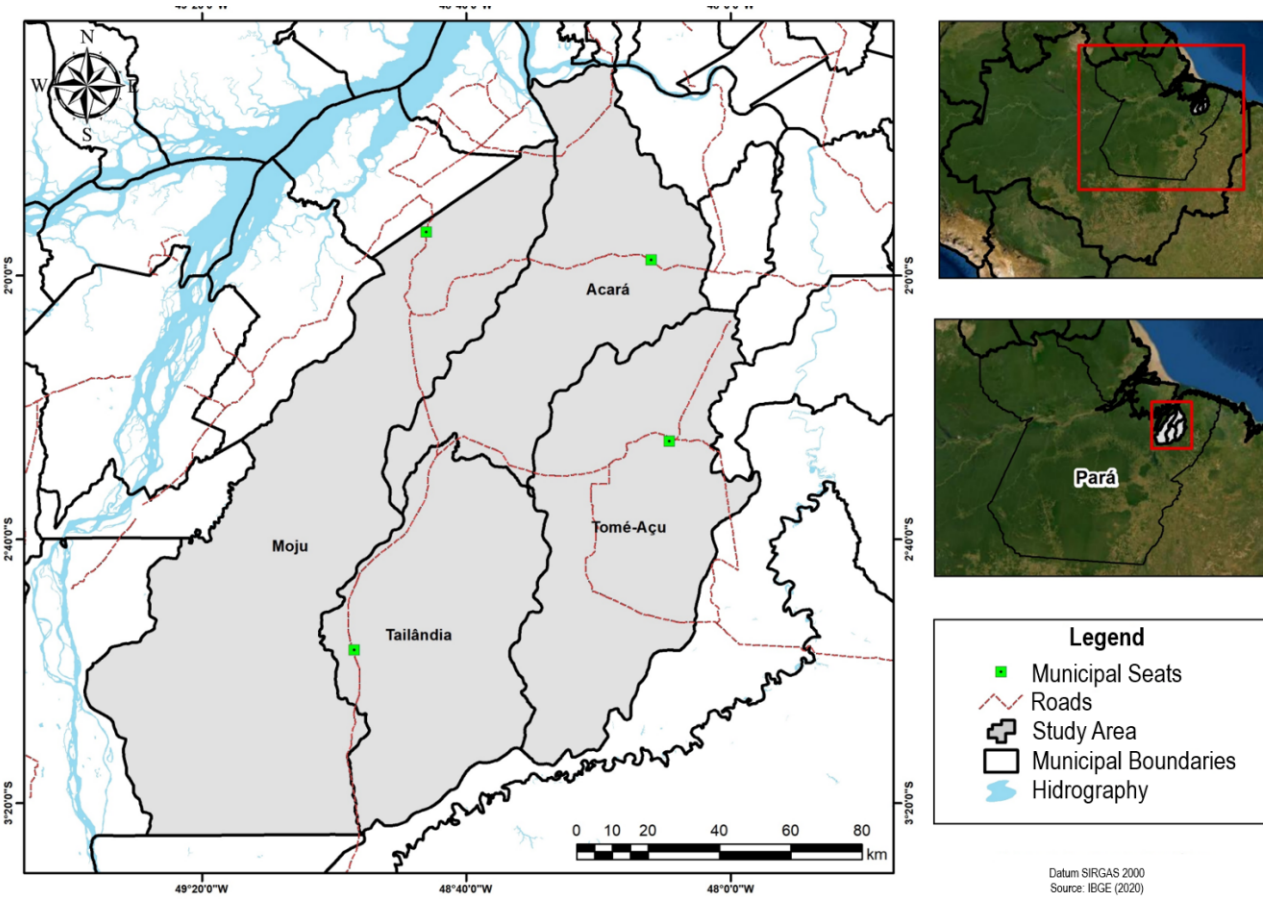

Figure 1: Land use and land cover maps.

The region receives an average annual rainfall of 2,500 millimeters, with a minimum monthly precipitation of 60 millimeters, which is sufficient to sustain oil palm cultivation without irrigation. The terrain is relatively flat, with altitudes ranging from 50 to 100 meters, and an average temperature of 26ºC. Historically, the landscape was predominantly used for cattle ranching, which expanded in the 1960s after the completion of the BR-010 highway (Belém-Braslia). Other economically significant agricultural activities include the cultivation of black pepper, açaí, eucalyptus for timber, cassava, and rice (Almeida, Vieira, and Ferraz 2020).

The region is also located near the Belém Endemism Center, an area recognized for its high species endemism, severe habitat fragmentation, and susceptibility to fires, making it one of the most deforested and ecologically threatened areas in the Amazon (Manhães, Rocha, Souza, et al. 2024). Despite still retaining over half of its area covered by forests, ongoing deforestation pressures pose significant environmental risks. During our study period, the forest area decreased from 13,281.43 km2 (57.45%) to 12,556.46 km2 (54.32%), highlighting the increasing threat to biodiversity and ecosystem stability (MapBiomas 2025).

3 Data

In order to conduct our analysis, we collect a number of pieces of information on optical imagery and radar data, land use, physical characteristics of the territory and potential socioe-

conomic drivers of land use and oil palm cultivation.

Remote sensing and physical characteristics of the territory

To map oil palm expansion and vegetation formation between 2014 and 2020, we used optical imagery from Landsat-8 and radar data from Sentinel-1, generating annual composites based on median values to address potential seasonality effects. Details of the dataset and methodology are provided in the following section and the Appendix.

We obtained elevation data from global satellite-derived Digital Elevation Models (DEMs), primarily sourced from the Shuttle Radar Topography Mission (SRTM). SRTM employs interferometric radar technology to map Earth's topography with high accuracy, offering elevation data at 90-meter (3 arc-seconds) resolutions. From this dataset, we derived a slope raster using terrain analysis functions, which calculate the rate of elevation change at each pixel, considering eight neighboring pixels to enhance local topographic representation. The data are expressed in radians and represent variations in terrain steepness.

Precipitation data were retrieved from the WorldClim database, which offers high-resolution gridded climate data derived from weather station observations, remote sensing, and interpolation techniques. This dataset provides monthly average precipitation (mm) at a spatial resolution of 10 arc-minutes ($\approx$ 18 km per pixel), which we aggregated into an annual value for the baseline year 2014.

We extract the maximum potential yield of oil palm from FAO-GAEZ under non-irrigated, low-input, and low-management conditions, reflecting extensive, minimally intensive cultivation practices. The yield estimates are based on average climatic conditions for the period 1961–1990, capturing long-term climatic trends that influence palm oil productivity while excluding potential impacts from recent climate change. This high-resolution (30 arc-seconds, 0.9 km x 0.9 km) spatial database is a globally recognized resource to assess the potential of agricultural productivity, integrating agronomic models, climatic conditions, soil properties and land suitability. In addition to oil palm, we also obtained yield estimates for soybean, maize, rice, and cassava to control for possible confounders.

To standardize the spatial resolution of our dataset to 30 meters, we applied bilinear resampling to all images with coarser resolutions. This interpolation method computes new pixel values as a weighted average of the four nearest neighboring pixels, ensuring smoother transitions and preserving spatial continuity in continuous variables. As a result, our final dataset consists of 39,946 observations (pixels) covering the study area.

Land use

We used data from MapBiomas Collection 9 to examine transitions in oil palm cultivation and forest area between 2014 and 2020, as well as the baseline pasture cover in 2014, all at a 30-meter pixel resolution. The MapBiomas dataset provides high-resolution annual land-use and land-cover maps for Brazil, derived from Landsat imagery, serving as a key resource for monitoring and analyzing land-use dynamics over time (Souza et al. 2020).

To check the robustness of our results, we also use the Hansen et al. (2013) dataset, widely known as the Global Forest Change (GFC) product. It offers high-resolution (30 meters) information on forest cover dynamics worldwide. This dataset integrates Landsat imagery with advanced data processing techniques to systematically map forest cover, loss, and gain from 2000 onwards.

In this study, we analyzed forest loss between 2014 and 2020, along with the forest cover in 2000, adjusted for net forest change (gain minus loss) between 2000 and 2014, to estimate the baseline forest cover in 2014.

Socioeconomic drivers

The transport infrastructure variables (roads, waterways and ports) were obtained from the MapBiomas Project, which processes and publicly provides data from official sources. These variables represent regional accessibility and can influence oil palm cultivation decisions by affecting transportation costs and market access. We also used the Urban Shapefile from the Brazilian Institute of Geography and Statistics (IBGE) to evaluate the impact of urban proximity on land use decisions.

We extracted mill geolocation data from the Universal Mill List (UML), a comprehensive global database of palm oil mills designed to improve supply chain transparency and support sustainable sourcing. Developed in collaboration with the World Resources Institute (WRI), Rainforest Alliance, Proforest, and Daemeter, the UML integrates data from certification bodies, industry reports, and company disclosures.

Data for conservation units and indigenous land are sourced from the Brazilian National System of Conservation Units (SNUC) and the National Indian Foundation (FUNAI), providing spatial data on official geographic boundaries.

The 2010 population density dataset used in this study is derived from WorldPop, which provides high-resolution global population maps (1 km^2 resolution). This data set integrates census data, satellite imagery (for example, Landsat, Sentinel), and machine learning techniques, offering detailed insights into population distribution patterns.

The Nighttime Light (NTL) data were obtained from VIIRS, onboard the Suomi NPP satellite, launched in 2011 as part of NASA's Earth observation program. This data set provides high-resolution measurements of Earth's nighttime brightness, using artificial illumination as a proxy for human activity and economic development. It measures absolute radiance in nanowatts per square centimeter per steradian ($nW/cm^2/sr$), capturing the intensity of emitted or reflected light at night at a 500-meter spatial resolution, including both stable sources (e.g., urban lighting) and transient sources (e.g., wildfires and gas flares). For this analysis, we computed changes in VIIRS-derived radiance between 2014 and 2020 to estimate shifts in economic activity over time.

4 Methods

4.1 Mapping and Classification

We mapped palm oil plantations in 2014, 2017, and 2020 by integrating optical imagery from Landsat-8 and radar imagery from Sentinel-1, extracting relevant features for classification using machine learning algorithms. We chose 2014 as the baseline year, because Sentinel-1 was launched and became operational only in that year.

Since Sentinel-1 offers 10-meter spatial resolution imagery, we resampled it to 30 meters to match Landsat-8's resolution and ensure spatial consistency in data processing. The combination of spectral and backscatter information from these satellite sources provides critical insights into biophysical properties, such as biomass, vegetation health, and phenological stages, enabling accurate land-cover classification.

We collected training and testing samples using a random sampling approach, resulting in the following pixel distributions for each class: (i) Vegetal Formation (3,330 pixels), encompassing primary and secondary forests; (ii) Oil Palm (1,564 pixels); and (iii) Bare Soil (1,300 pixels), which included exposed soil, cropland, and pasture.

For classification, we implemented multiple machine learning algorithms, including K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector

Machines (SVM), and Random Forests (RF). During the training phase, we identified the most relevant variables, validated the models using the test dataset, and then applied the best-performing algorithm for classification.

In the post-classification phase, we applied a spatial mode filter and performed a manual reclassification of palm oil areas to improve the accuracy of the results, using the 2017 map to identify and correct inconsistencies in land-use changes. To track palm oil expansion, we conducted a pixel-by-pixel land-use transition analysis between 2014 and 2020. A detailed description of each methodological step in the classification and mapping process is provided in the Appendix.

4.2 Empirical Strategy for Causal Inference

Identifying the causal impacts of palm oil on deforestation and economic activity is difficult due to the endogeneity of its expansion. The high potential costs and benefits of investments in oil palm plantations make their location often endogenous with regional characteristics. Estimating the causal effects of palm oil expansion involves two main challenges: (i) unobservable variables that may be correlated with palm oil expansion, economic activity and deforestation; (ii) reverse causality, as a higher rate of deforestation and/or economic activity could also encourage the expansion of palm oil.

This paper explores an exogenous variation, the maximum potential agro-climatically attainable palm oil yield, to instrumentalize its expansion in the Eastern Amazon and thus estimate its causal impact on deforestation and economic activity. The instrument is measured as the agro-climatically attainable palm oil yield at the pixel level. It is calculated by FAO-GAEZ based on agronomic models and provides data on agroclimatic potential yield for different crops and levels of input and management at 30 arc-second (0.9 x 0.9 km) resolution. In this paper, we use the maximum potential yield of palm oil under non-irrigated, low-input, and low-management conditions for an average climate for the period 1981-2010. We selected the low-input level as it has been shown in the literature to have the strongest correlation with the actual expansion of oil palm plantations (Kubitza and Gehrke 2018).

The first-stage in our instrumental variables (IV) estimation procedure is given by the following equation:

$$\Delta Oil\_Palm_i = \alpha + \gamma \text{Yield}_i + \delta Controls_i + u_i \tag{1}$$

where ΔOil_Palm_i is a binary variable with value 1 when the pixel i is converted to the oil palm class between 2014 and 2020. By concentrating on changes between 2014 and 2020 we hope to capture a more consolidated impact of oil palm plantations on forests and economic activity over a longer time span, which accommodates for both direct impacts and indirect effects via displacement of other activities, for example. $Yield_i$ is the maximum agro-climatically attainable palm oil yield for pixel i, $Controls_i$ represents a vector of pixel-level control variables that include physical attributes (altitude, slope, precipitation, latitude, longitude), 2014 baseline population density and land use (forest, pasture and night light); maximum attainable yields (for soybean, maize, rice and cassava), infrastructure (distance to roads, waterways, ports, urban areas, mills) and distance to conservation units and indigenous land; and u_i is an error term. The first stage intuition is that higher potential yields increase the probability of crop expansion. given that palm oil firms decide to expand production based on their expected productivity in addition to potential profitability (Edwards 2018).

Then, we estimate the following two second-stage equations to measure the impacts of oil palm expansion,

$$\Delta Deforestation_i = \alpha + \beta \Delta \widehat{Oil\_Palm}_i + \delta Controls_i + \varepsilon_i \tag{2}$$

$$\Delta Nightlight_i = \alpha + \beta \Delta \widehat{Oil\_Palm}_i + \delta Controls_i + \varepsilon_i \quad (3)$$

where $\Delta Deforestation_i$ is a binary variable that is equal to 1 when pixel i changes from the vegetation class to any other class between 2014 and 2020; $\Delta Nightlight_i$ is the change in night lights within a 500 meters radius from pixel i during the same period; $\Delta \widehat{Oil\_Palm}_i$ is the instrumentalized variable from the first stage; $Controls_i$ are the same controls as in the first stage equation; and ε_i is the error term.

We first estimate equations (1) and (2) using a logit specification. For nightlights, we estimate (3) with a linear model. However, the distribution of our database is not uniformly distributed with only 11% and 3,36% of the pixels indicating deforestation and palm oil expansion, respectively. Therefore, it may not be reasonable to expect that pixels with a 50% deforestation probability are more sensitive than other pixels to changes in independent variables, as assumed in the logit model. In this context, to further check the robustness of our results, we re-estimated equations (1) – (2) using the more general and flexible Skewed logistic regression approach (scobit) proposed by Nagler 1994, and which includes the logit model as a special case. The scobit model is better suited to outcomes characterized as rare events, such as deforestation and oil palm expansion in this paper (Alix-García and Millimet 2023). For comparison, we also estimate equation (2) using a binary measure of deforestation from Hansen et al. (2013) (logit and scobit in both stages) and a non-binary measure of deforestation that also includes reforestation from MapBiomas (linear model in the second stage).

The crucial identification hypothesis in our context is that the maximum agro-climatically attainable palm oil yield affects deforestation and economic activity only through the oil palm expansion channel. According to Kubitza and Gehrke (2018), the instrument is highly correlated with the expansion of oil palm, as, together with access to land and markets, yield potential is the main determinant of land use patterns. Despite this, there are still some threats to the identification strategy that need to be addressed.

First, other crops have agroclimatic conditions and expansion patterns that may be similar to those for palm oil. Therefore, the instrument may be capturing potential agricultural productivity in general, which would not allow causal interpretations for the estimated effects. For example, an important input for GAEZ palm oil productivity is precipitation, which in addition to affecting the outcome for alternative tropical crops, also impacts deforestation and economic activity (Chomitz and Thomas 2003). Although the instrument is specific for oil palm, it is important to exclude this threat from the identification strategy. Therefore, we control for the potential yields for other competing agricultural activities such as soybean, maize, rice, and cassava in our estimations. We also accounted for whether the pixel was classified as pasture in the baseline by the MapBiomas project, as this classification could increase the likelihood of palm oil expansion due to regulations introduced by the PPSP.

Second, another possible threat to the identification strategy is that the instrument may be capturing general geographic features at the pixel level that may be correlated with deforestation and economic activity. Since our empirical strategy relies on changes in oil palm, deforestation and night lights for the 2014-2020 period, unobserved time-invariant variables at the pixel level such as geographic features are eliminated from the estimated equations. In addition, we include latitude and longitude variables in our equations to control for potential geographically distributed omitted variables.

Finally, it is also plausible that initial differences in infrastructure, economic scale and conservation status could induce different trends in crop expansion, forest loss, and economic activity. Therefore, we control for the distance of pixel i to the nearest road, waterways, ports, cities,mills, conservation units and indigenous land to capture the availability of infrastructure, access to markets and protected areas.

5 Results

5.1 Classification

We selected the preferred oil palm mapping result by first comparing each machine learning algorithm using its overall accuracy and Kappa coefficient estimated in the test sample. In addition, we validated all models with the 5-fold cross-validation method for both feature selection and final classification. The results appear in Table A19 in the Appendix and reflect the statistics after applying the recursive feature elimination method. To check whether the algorithms performed well, we considered the overall accuracy and the Kappa coefficient for the years mapped. The weight of the evidence points at the Random Forest (RF) algorithm as the preferred choice.[6] Therefore, we present only the classification results for the RF algorithm due to its better performance and to maintain the consistency of the following analyses. Finally, we calculate the confusion matrices based on our test samples and present the results in Figure 4 in the Appendix.

Next, we calculated the user's accuracy (UA) and producer's accuracy (PA)[7] to assess the robustness of our results. UA represents the proportion of correctly classified pixels within a given class relative to all pixels assigned to it, indicating the reliability of the classification. PA evaluates the proportion of reference pixels for a given class that are correctly classified, reflecting how well the classification captures the actual extent of that class. For oil palm, UA and PA values of 0.865 and 0.866 in 2014, and 0.93 and 0.929 in 2020, respectively, demonstrate the effectiveness of the methodology in distinguishing oil palm plantations from other land-use classes and ensuring reliable monitoring of their expansion. Vegetal formations consistently achieved the highest accuracies, with UA and PA exceeding 0.97 across all years, while bare soil presented slightly lower values, with UA and PA reaching 0.917 and 0.913 in 2020.

Our results is in line with the recent literature. Xu et al. (2021) also compared several algorithms and obtained the best classification performance with the RF algorithm. Although Shaharum et al. (2020) obtain a better overall accuracy and Kappa for the SVM, they emphasize that RF dominated SVM when the criterion is delimitation of the oil palm areas – our goal here. Specifically for the Eastern Amazon, Almeida, Vieira, and Ferraz (2020) also performed a classification exercise using RF and obtained an overall accuracy of 88.06% and a Kappa of 0.85, using Landsat images. The superior results we present here may be due to the combination of optical and radar images. Radar images capture additional geometric information, which allows for better discrimination of targets with unique geometric characteristics such as oil palm.

Next, Figure 5 in the Appendix shows the selected features with the recursive feature elimination method for the Random Forest algorithm. In practice, this method selects the best combination of features for the final classification. The relevant characteristics for oil palm mapping differ depending on the year and may reflect different climatic, phenological, and stage of maturation conditions. For 2014, the method selected 12 features: Landsat.4, Landsat.2, EVI, Endmember_2, DVI, Landsat.3, Landsat.1, Entropy, Setinel_VV, GI and Sentinel_VH. For 2017, 15 features stand out: EVI, Landsat.2, Landsat.4, Sentinel_VH—VV, DVI, Endmember_1, Sentinel_NDI, Landsat.1, Sentinel_VH-VV, Sentinel_VV, Sentinel_VH, Entropy, GI, Endmember_2

6. In 2014, RF was the best-performing algorithm (accuracy of 94.45% and Kappa of 0.0975) followed by the Support Vector Machine (SVM) (accuracy of 94.03% and Kappa of 0.8993). In 2017, on the other hand, K-Nearest Neighbor (KNN) was the best algorithm (Accuracy of 94.61% and Kappa of 0.9088) followed by RF (Accuracy of 94.28% and Kappa of 0.9031). Finally, in 2020, RF was again the best algorithm (Accuracy of 95.53% and Kappa of 0.9239) followed by KNN (Accuracy of 93.87% and Kappa of 0.8956).

7.

$$\text{UA} = \frac{\text{Number of correctly classified pixels for a class}}{\text{Total pixels classified as that class}} ; \text{PA} = \frac{\text{Number of correctly classified pixels for a class}}{\text{Total reference pixels for that class}}$$

and Contrast. Finally, for 2020, 14 features were selected: EVI, Sentinel_VH—VV, Landsat.2, Landsat.3, Sentinel_NDI, Sentinel_VH-VV, Landsat.4, Sentinel_VH, Landsat.1, DVI, GI, Endmember_2, Endmember_1 and Sentinel_VV[8]. The final classification retained most variables across all years, with RFE primarily eliminating certain texture metrics that contributed little to accuracy or introduced redundancy.

Features related to Landsat-8 reflectance surface information were more important for classifying oil palm areas, especially the enhanced vegetation index (EVI). However, Sentinel-1 features were also important, with the difference and the ratio between VV and VH appearing as the most relevant, respectively. The order of importance differs from those obtained by Xu et al. (2021), which used Top of Atmospheric (TOA) data from Landsat-8 and Sentinel-1.

5.2 Land Use and Land Cover Transition

Figure 2 presents land use and land cover maps for the study area and years 2014, 2017, and 2020, classified with the Random Forest algorithm. Visual inspection of the maps suggests that oil palm expansion occurred along roads and in both degraded areas (bare soil), as stimulated by oil palm policies in the country, and areas covered with vegetation formation. We obtain additional insight on the expansion of oil palm by inspecting changes in land classes over time. The areas covered by oil palm and bare soil increased faster between 2014 and 2017 and more slowly between 2017 and 2020. These changes translated into a faster rate of decrease in vegetation formation in the first period and a slower contraction in the second period.[9]

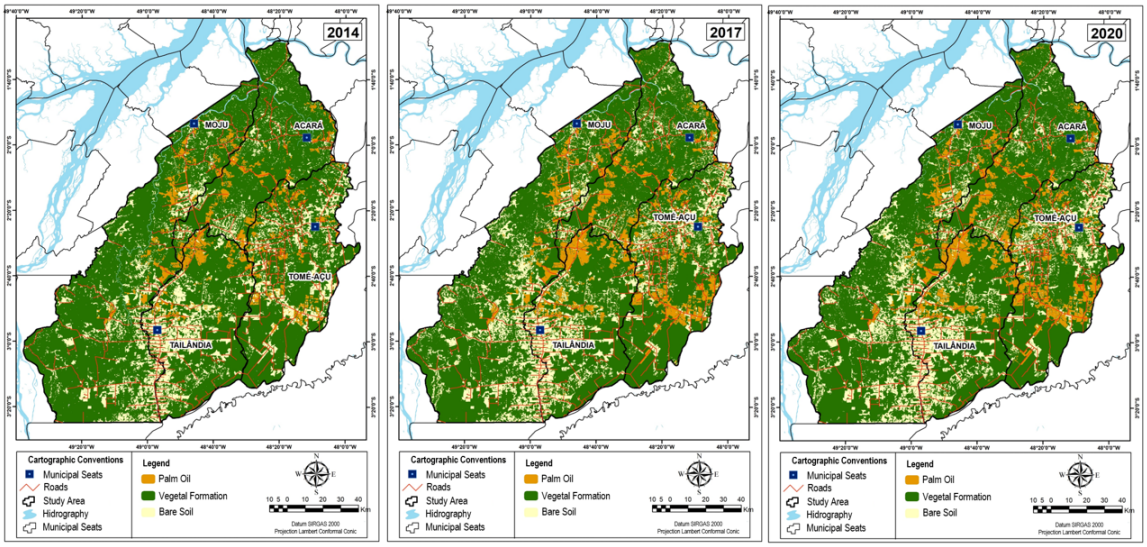

Figure 2: Land use and land cover maps.

Comparing our results with those of MapBiomas, they estimated a total of 877.07 km² of oil palm plantations in 2014, while our study reports an area of 1,074 km². Similarly, MapBiomas calculated 1,510.3 km² of plantations in 2020, while our study estimates 1,849 km² of oil palm

8. The features for 2017 follows the same pattern (Figure 4 in the Appendix), reinforcing the results.

9. The vegetation formation class went from an area of 17,529.57 km² (76.37% of the total) in 2014 to 16,436.8 km² (71.61%) and in 2017 and 16,289.13 km² in 2020 (70,96%). The oil palm class, in turn, had an area of 1,074.93 km² (4.68%) in 2014, increasing to 1,703.27 km² (7.42%) in 2017 and 1,849.89 km² (8.06%) in 2020, an overall growth of 72.16%. Finally, the Bare Soil class had an area of 4,349.72 km² (18.95%) in 2014, growing to 4,814.15 km² (20.97%) in 2017 and to 4,815.2 km² in 2020 (20 .98%). The oil palm area expanded over time by 58.45% (2014-2017) and 8.61% (2017-2020), while the bare soil areas increased by 10.68% (2014-2017) and 0.02% (2017-2020). Meanwhile, the area covered by vegetal formation decreased by 6.23% between 2014 and 2017, and by 0.9% between 2017 and 2020. See Figure 6 in the Appendix.

farms. This represents an approximate 22.5% increase over the MapBiomas estimate, suggesting that the integration of radar imagery with traditional optical data enhances the detection and classification of oil palm plantations in the Amazon. In addition, oil palm plantations from MapBiomas expanded in both pastures (574.94 km^2) and natural vegetation (52.52 km^2) between 2014 and 2020. That is, most of the growth occurred within the scope of public policies for palm oil in the Brazilian Amazon, but more than 8% of the crop expansion occurred in forest areas. Furthermore, pastures occupied an additional 1346.08 km^2 of forest areas during the same period, and some of this expansion could be indirectly attributed to oil palm through the displacement of consolidated pastures.

From the perspective of forests and natural ecosystems, our results indicate that the area lost in the vegetation formation class was occupied by both oil palm and bare soil, although the latter responded to the lion's share of the cleared land. This, in turn, can be a consequence of the fact that the palm oil industry is still relatively small. Therefore, it is important to ask whether sustained growth of the palm oil industry is likely to have a significant impact on deforestation in the Brazilian Amazon region. To explore this question, we first need to recognize that Figure 3 represents an accounting identity for land class changes in the region, but does not inform us whether oil palm is indeed a potentially significant driver of local deforestation. Our mapping effort is a vital element of the assessment of the impact of the palm oil industry on deforestation, but does not clarify whether the deforestation associated with oil palm was caused by the crop itself or whether it would have happened anyway due to other socioeconomic drivers that cause deforestation (and possibly also the expansion of the palm oil industry). In the next section, we present our instrumental variable results for the causal impacts of palm oil expansion on deforestation and on an indicator of economic activity. Throughout the analysis, we concentrate on land use changes between 2014 and 2020. We chose the longer time frame to increase the chance of capturing both direct and indirect impacts of oil palm in forested areas.

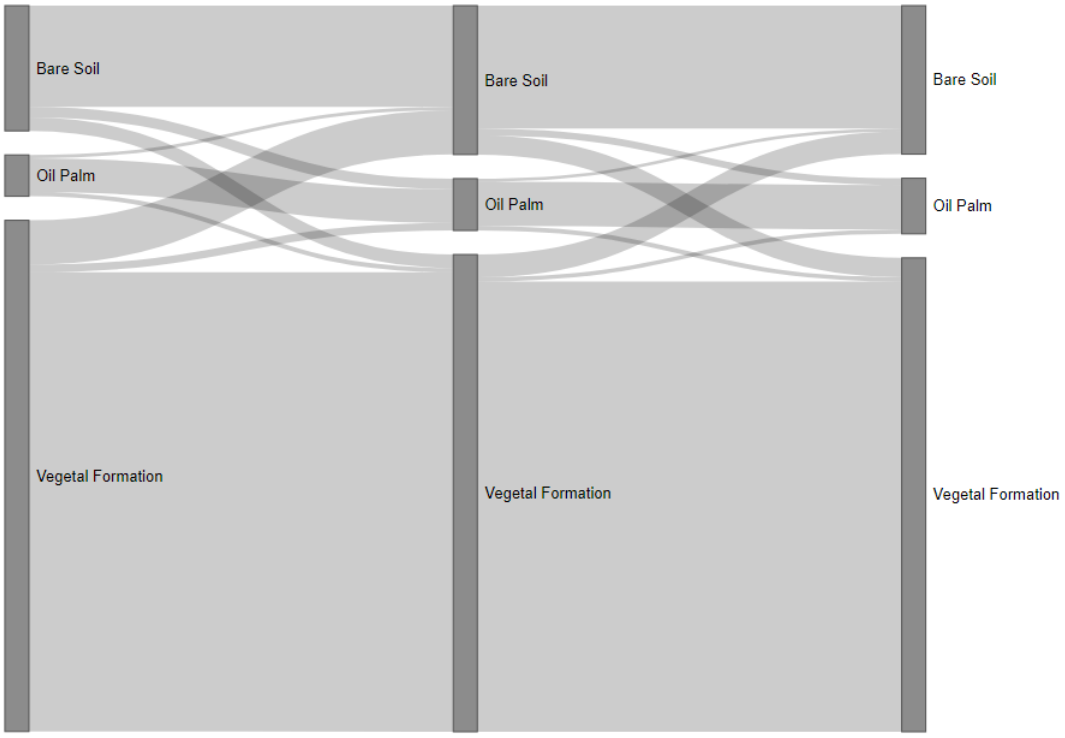

Figure 3: Land use and land cover transitions between 2014-2017 and 2017-2020.

5.3 Causal Effects Of Palm Oil Expansion

Table 4 in the appendix reports first-stage estimates from our IV strategy. The oil palm yield potential is positively and significantly correlated with the expansion of oil palm in the region. Furthermore, the size and significance of the yield coefficients are robust to different specifications and the inclusion of a rich set of controls. These results suggest that our instrument does not simply capture broader geographic, agricultural, or socioeconomic determinants of palm oil expansion, thus supporting both the relevance and exclusion restrictions required for the validity of our instrumental variable.

Turning to the main results of interest, Table 1 reports our estimates of oil palm expansion on deforestation (Panel A) and nightlights (Panel B). Column (1) reports naïve correlations that do not account for the endogeneity of oil palm expansion with a logit regression for our binary measure of deforestation and ordinary least squares (OLS) regression for the continuous nightlight variable. In Panel A, both stages rely on either a logit (columns (2) and (3)) or a scobit (columns (4) and (5)) specification, whereas in Panel B, the first stages uses the same strategy as in Panel A, but the second stage relies on a linear regression model.

Although the relationship between oil palm and deforestation is positive and significant in all specifications in Panel A, the estimates in columns (2)–(5) suggest a significant endogeneity bias in the simple correlation estimate from column (1). Adding controls to our models reduces the size of the coefficients, but they remain significant. Panel A suggests that oil palm indeed contributes to deforestation in the region.

In Panel B, the estimated coefficients go from insignificant in column (1) to negative and marginally significant in the IV specifications with controls (p-value of 0.058 in column (3) and 0.068 in column (5)). A sizable and significant increase in light intensity near oil palm plantations could suggest higher returns to complementary activities that are more urban and less land-intensive, which could make investments in deforestation less attractive (Hanusch 2023). However, this does not appear to be the case, at least in the vicinity of the area recently occupied by oil palm. Because our results stem from relatively high spatial resolution data (up to 500 m from a pixel converted to oil palm), they need to be interpreted with caution. For example, Ferreira et al. (2023) use a different methodological approach to examine more aggregated municipal data for a longer period of time (2002-2017) and identify positive spillover effects of oil palm expansion on broader economic activity in the same region.

One potential concern is that our deforestation proxy may not exclusively capture forest clearings, but could also include other land use changes misclassified as vegetation formation. This could put an additional burden of proof on the causal interpretation of our results. To further assess the robustness of our findings, we re-estimate the specifications in Figure 1, Panel A, using the Hansen et al. (2013) dataset. The magnitudes of the coefficients are not directly comparable because Hansen et al. (2013) specifically tracks primary forest loss, whereas we account for both primary and secondary forest loss, as well as degradation of other natural ecosystems. The results align with our main findings, providing further evidence that oil palm expansion is contributing to deforestation in the eastern Amazon.

6 Conclusion

Palm oil is the most consumed vegetable oil in the world and a potentially important element of a transition to low-carbon energy. At the same time, it has been associated with large-scale deforestation, ecosystem degradation, and loss of biodiversity in tropical forests, especially in Southeast Asia. The strategic importance of palm oil and ecosystem preservation for a cleaner

Table 1: Impact of Oil Palm Expansion on Deforestation and Nightlights

	Logit (1)	**Logit-IV** (2)	**Logit-IV** (3)	**Scobit-IV** (4)	**Scobit-IV** (5)
Panel A: Deforestation					
Oil Palm Expansion	0.9190***	21.6495***	8.0818***	20.5510***	8.4074***
	(0.0035)	(1.6870)	(0.2990)	(1.6018)	(0.5404)
Marginal Effect	0.9190***	2.1147***	0.6495***	2.1318***	0.6325***
	(0.0035)	(0.1654)	(0.0233)	(0.1699)	(0.0281)
	OLS (1)	**IV** (2)	**IV** (3)	**IV** (4)	**IV** (5)
Panel B: Nightlights					
Oil Palm Expansion	-0.0056	0.8813***	-0.0412*	0.8782***	-0.0353*
	(0.0041)	(0.0754)	(0.0217)	(0.0751)	(0.0193)
Geographic	Yes	No	Yes	No	Yes
Yield potential	Yes	No	Yes	No	Yes
Socioeconomic	Yes	No	Yes	No	Yes
Observations	39,497	39,497	39,497	39,497	39,497

Notes: Panel A shows results for deforestation, while Panel B focuses on nightlight. Columns (1) to (5) represent different model specifications as described in the table. Standard errors, shown in parentheses, are clustered at the pixel level.

Table 2: Robustness Checks Using Hansen et al. (2013) Dataset

	Logit (1)	**Logit-IV** (2)	**Logit-IV** (3)	**Scobit-IV** (4)	**Scobit-IV** (5)
Panel A: Hansen					
Oil Palm Expansion	0.5093***	4.6325***	0.7319***	4.2754***	0.4666**
	(0.0753)	(1.1728)	(0.2990)	(1.0852)	(0.2368)
Marginal Effect	0.0754***	0.7046***	0.1085***	0.7043***	0.0775**
	(0.0111)	(0.1783)	(0.0410)	(0.1782)	(0.0393)
Geographic	Yes	No	Yes	No	Yes
Yield potential	Yes	No	Yes	No	Yes
Socioeconomic	Yes	No	Yes	No	Yes
Observations	39,497	39,497	39,497	39,497	39,497

Notes: Panel A shows results for deforestation from Hansen et al. (2013), while Panel B focuses on X. Columns (1) to (5) represent different model specifications as described in the table. Standard errors, shown in parentheses, are clustered at the pixel level.

economy has led the Brazilian government to design policies to promote oil palm plantations on already degraded lands and without causing deforestation. This is an especially sensitive issue in the Brazilian Amazon, where development opportunities for the local population are much needed and biodiversity along with ecosystem services can hold an important key to a modern economy.

As a consequence of national palm oil policies, the eastern Amazon region has consolidated itself as the main and fastest growing producer of oil palm in Brazil in recent years. This paper assesses the broader impacts of palm oil promotion in an area concentrating more than 70% of Brazilian oil palm plantations. More specifically, we investigate the impact of oil palm on local deforestation and economic activity. To do so, we tackled two challenges for policy evaluation in this specific context: the scarcity of high-quality data on the location of oil palm plantations; and endogeneity bias likely to plague the estimation of causal effects of oil palm on deforestation and the economy. We address the data scarcity problem by integrating optical imagery from Landsat-8 and radar imagery from Sentinel-1 in machine learning models to produce a more accurate map for oil palm plantations. The accuracy of palm oil classification reached 94.53% in 2014 and 95.53% in 2020, with Kappa coefficients of 0.9075 and 0.9239, respectively. These results surpass those of similar studies that mapped oil palms in the eastern Amazon without integrating optical and radar data. Our results indicate that most of the oil palms planted between 2014 and 2020 occurred in already occupied area and therefore within the scope of national policies. However, more than 20% of the crop expansion occurred in forest areas. In addition, we also observed significant deforestation in the region, which could also be indirectly attributed to oil palm growth through the displacement of consolidated pastures.

The land use change figures are nothing more than an accounting identity and our next step was to estimate the impact of oil palm plantations on local deforestation and nightlights, a measure of more urban and less land-intensive economic activity. We estimated an instrumental variable strategy that uses the potential for crop yield as an instrument, and our results suggest that oil palm contributes indeed to local deforestation. In addition, we do not find any evidence that it also promotes other less land-intensive activities that could contribute to slowing down deforestation in the vicinity of the plantations.

Although the evidence suggests that policies to promote oil palm plantations in already degraded areas of the Brazilian Amazon were in great part successful, the fact that a significant portion of the crop occupies areas previously covered by natural ecosystems and that palm oil production in the country is still small, but growing, calls for targeted monitoring of oil palm farms and deforestation in their vicinity. Monitoring the expansion of the palm oil production chain and enforcement of forest protection at an early stage is strategic both for the industry, as deforestation might affect their access to international markets, and for the region, which may benefit from opportunities from the industry and a broader development pattern centered on the preservation of its natural capital.

References

Abman, Ryan, and Clark Lundberg. 2024. "Contracting, market access and deforestation." *Journal of Development Economics* 168:103269. https://doi.org/https://doi.org/10.1016/j.jdeveco.2024.103269.

Alix-García, Jennifer, and Daniel L. Millimet. 2023. "Remotely Incorrect? Accounting for Nonclassical Measurement Error in Satellite Data on Deforestation." *Journal of the Association of Environmental and Resource Economists* 10 (5): 1335–1367. https://doi.org/10.1086/723723.

Almeida, Arlete Silva, Ima Célia Guimarães Vieira, and Silvio F.B. Ferraz. 2020. "Long-term assessment of oil palm expansion and landscape change in the eastern Brazilian Amazon." *Land Use Policy* 90:104321. https://doi.org/10.1016/j.landusepol.2019.104321.

Benami, E, Lisa Curran, M Cochrane, Adriano Venturieri, R Franco, J Kneipp, and A Swartos. 2018. "Oil palm land conversion in Pará, Brazil, from 2006-2014: Evaluating the 2010 Brazilian Sustainable Palm Oil Production Program." *Environmental Research Letters* 13:034037. https://doi.org/10.1088/1748-9326/aaa270.

Brandão, Frederico, George Schoneveld, Pablo Pacheco, Ima Vieira, Marc Piraux, and Dalva Mota. 2021. "The challenge of reconciling conservation and development in the tropics: Lessons from Brazil's oil palm governance model." *World Development* 139:105268. https://doi.org/https://doi.org/10.1016/j.worlddev.2020.105268.

Brazilian Institute of Geography and Statistics (IBGE). 2025. *IBGE System for Automatic Retrieval (SIDRA).* Accessed: February 2025. https://sidra.ibge.gov.br/home/ipp/brasil.

Burger, S. V. 2018. *Introduction to machine learning with R.* 1st edition. p.200. Beijing, [China].

Burgess, Robin, Matthew Hansen, Benjamin Olken, Peter Potapov, and Stefanie Onder. 2012. "The Political Economy of Deforestation in the Tropics." *The Quarterly Journal of Economics* 127:1707–1754. https://doi.org/10.2307/41812147.

Carvalho, Carolina Monteiro de, Semida Silveira, Emilio Lèbre La Rovere, and Allan Yu Iwama. 2015. "Deforested and degraded land available for the expansion of palm oil for biodiesel in the state of Pará in the Brazilian Amazon." *Renewable and Sustainable Energy Reviews* 44:867–876.

Chomitz, Kenneth, and Timothy Thomas. 2003. "Determinants of Land Use in Amazônia: A Fine-Scale Spatial Analysis." *American Journal of Agricultural Economics* 85:1016–1028.

Chong, Khai Loong, Kasturi Devi Kanniah, Christine Pohl, and Kian Pang Tan. 2017. "A review of remote sensing applications for oil palm studies." *Geo-spatial Information Science* 20 (2): 184–200. https://doi.org/10.1080/10095020.2017.1337317.

Cisneros, Elías, Krisztina Kis-Katos, and Nunung Nuryartono. 2021. "Palm oil and the politics of deforestation in Indonesia." *Journal of Environmental Economics and Management* 108:102453. https://doi.org/10.1016/j.jeem.2021.102453.

Corley, R. H. V., and P. B. H Tinker. 2008. *The Oil Palm.* p.592. John Wiley Sons.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98. https://doi.org/10.1257/jep.30.4.171.

Edwards, Ryan B. 2018. "Export agriculture and regional development: evidence from Indonesia."

Englund, Oskar, Göran Berndes, U Martin Persson, and Gerd Sparovek. 2015. "Oil palm for biodiesel in Brazil—risks and opportunities." *Environmental Research Letters* 10 (4): 044002. https://doi.org/10.1088/1748-9326/10/4/044002.

Ferreira, Susane Cristini Gomes, Claudia Azevedo-Ramos, Hilder André Bezerra Farias, and Pedro Mota. 2023. "Spillover effect of the oil palm boom on the growth of surrounding towns in the eastern Amazon." *Land Use Policy* 133:106867. https://doi.org/https://doi.org/10.1016/j.landusepol.2023.106867.

Foster, Andrew D., and Mark R. Rosenzweig. 2003. "Economic Growth and the Rise of Forests*." *The Quarterly Journal of Economics* 118 (2): 601–637. https://doi.org/10.1162/003355303321675464.

Furumo, Paul, and T. Mitchell Aide. 2017. "Characterizing commercial oil palm expansion in Latin America: Land use change and trade." *Environmental Research Letters* 12. https://doi.org/10.1088/1748-9326/aa5892.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-scale geospatial analysis for everyone." Big Remotely Sensed Data: tools, applications and experiences, *Remote Sensing of Environment* 202:18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (6160): 850–853. https://doi.org/10.1126/science.1244693.

Hanusch, Marek. 2023. *A balancing act for Brazil's Amazonian states: An economic memorandum.* World Bank Publications.

Hellmundt, Tobias, Elías Cisneros, and Krisztina Kis-Katos. 2024. "Land-use transformation and conflict: The effects of oil palm expansion in Indonesia." *SSRN Electronic Journal* (February). https://doi.org/10.2139/ssrn.4728074.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028. https://doi.org/10.1257/aer.102.2.994.

Holloway, Jacinta, and Kerrie Mengersen. 2018. "Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review." *Remote Sensing* 10 (9). https://doi.org/10.3390/rs10091365.

Kamusoko, C. 2019. *Remote Sensing Image Classification in R.* p.189. Springer Geography.

Koh, Lian Pin, and David S. Wilcove. 2008. "Is oil palm agriculture really destroying tropical biodiversity?" *Conservation Letters* 1 (2): 60–64. https://doi.org/10.1111/j.1755-263X.2008.00011.x.

Kraus, Sebastian, Robert Heilmayr, and Nicolas Koch. 2024. "Spillovers to Manufacturing Plants from Multimillion Dollar Plantations: Evidence from the Indonesian Palm Oil Boom." *Journal of the Association of Environmental and Resource Economists* 11 (3): 613–656. https://doi.org/10.1086/727196.

Krishna, Vijesh V., and Christoph Kubitza. 2021. "Impact of oil palm expansion on the provision of private and community goods in rural Indonesia." *Ecological Economics* 179:106829. https://doi.org/https://doi.org/10.1016/j.ecolecon.2020.106829.

Kubitza, Christoph, and Esther Gehrke. 2018. *Why does a labor-saving technology decrease fertility rates? Evidence from the oil palm boom in Indonesia.* EFForTS Discussion Paper Series 22. University of Goettingen, Collaborative Research Centre 990 "EFForTS, Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)".

Lameira, Wanja Janayna Miranda, Ima Célia Guimarães Vieira, and Peter Mann Mann de Toledo. 2016. "Expansão da dendeicultura em relação às zonas agroecológiacas de Tomé-Açu, Pará." *Revista Brasileira de Cartografia* 68 (10).

Liang, Hu, Na Li, and Shengrong Zhao. 2021. "Salt and Pepper Noise Removal Method Based on a Detail-Aware Filter." *Symmetry* 13 (3). https://doi.org/10.3390/sym13030515.

Manhães, A.P., F. Rocha, T. Souza, et al. 2024. "Social and biological impact of oil palm (Elaeis guineensis) plantations in the Eastern Brazilian Amazon." *Biodiversity and Conservation* 33:3295–3310. https://doi.org/10.1007/s10531-024-02913-x. https://doi.org/10.1007/s10531-024-02913-x.

MapBiomas. 2025. *MapBiomas - Collection of Annual Land Cover and Land Use Maps of Brazil.* Accessed: February 2025. https://brasil.mapbiomas.org/en/.

Meroni, Michele, Raphaël d'Andrimont, Anton Vrieling, Dominique Fasbender, Guido Lemoine, Felix Rembold, Lorenzo Seguini, and Astrid Verhegghen. 2021. "Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and -2." *Remote Sensing of Environment* 253:112232. https://doi.org/https://doi.org/10.1016/j.rse.2020.112232.

Nagler, Jonathan. 1994. "Scobit: An Alternative Estimator to Logit and Probit." *American Journal of Political Science* 38 (1): 230–255.

Nahum, João, Leonardo Santos, and Cleison Santos. 2020. "Formation of palm oil cultivation in Para's Amazon." *Mercator* 19. https://doi.org/10.4215/rm2020.e19007.

Oliveira, Susan E.M., Louise Nakagawa, Gabriela Russo Lopes, Jaqueline C. Visentin, Matheus Couto, Daniel E. Silva, Francisco d'Albertas, Bruna F. Pavani, Rafael Loyola, and Chris West. 2024. "The European Union and United Kingdom's deforestation-free supply chains regulations: Implications for Brazil." *Ecological Economics* 217:108053. https://doi.org/https://doi.org/10.1016/j.ecolecon.2023.108053.

Sano, Edson Eyji, Eraldo Aparecido Trondoli Matricardi, and Flávio Fortes Camargo. 2020. "State-of-the-art of Radar Remote Sensing: Fundamentals, Sensors, Image Processing, and Applications." *Revista Brasileira de Cartografia* 72. https://doi.org/10.14393/rbcv72nespecial50anos-56568.

Shaharum, Nur Shafira Nisa, Helmi Zulhaidi Mohd Shafri, Wan Azlina Wan Ab Karim Ghani, Sheila Samsatli, Mohammed Mustafa Abdulrahman Al-Habshi, and Badronnisa Yusuf. 2020. "Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms." *Remote Sensing Applications: Society and Environment* 17:100287. https://doi.org/https://doi.org/10.1016/j.rsase.2020.100287.

Somers, Ben, Gregory P. Asner, Laurent Tits, and Pol Coppin. 2011. "Endmember variability in Spectral Mixture Analysis: A review." *Remote Sensing of Environment* 115 (7): 1603–1616. https://doi.org/10.1016/j.rse.2011.03.003.

Souza, Carlos M., Julia Z. Shimbo, Marcos R. Rosa, Leandro L. Parente, Ane A. Alencar, Bernardo F. T. Rudorff, Heinrich Hasenack, et al. 2020. "Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine." *Remote Sensing* 12 (17). https://doi.org/10.3390/rs12172735.

Villela, Alberto A., D'Alembert B. Jaccoud, Luiz P. Rosa, and Marcos V. Freitas. 2014. "Status and prospects of oil palm in the Brazilian Amazon." *Biomass and Bioenergy* 67:270–278. https://doi.org/10.1016/j.biombioe.2014.05.005.

Weiss, M., F. Jacob, and G. Duveiller. 2020. "Remote sensing for agricultural applications: A meta-review." *Remote Sensing of Environment* 236:111402. https://doi.org/10.1016/j.rse.2019.111402.

Xu, Kaibin, Jing Qian, Zengyun Hu, Zheng Duan, Chaoliang Chen, Jun Liu, Jiayu Sun, Shujie Wei, and Xiuwei Xing. 2021. "A New Machine Learning Approach in Detecting the Oil Palm Plantations Using Remote Sensing Data." *Remote Sensing* 13 (2). https://doi.org/10.3390/rs13020236.

A Appendix

A.1 Remote Sensing

The increasing availability of satellite images with high spatial, temporal, and spectral resolution, and in low-cost monitoring techniques, have facilitated their application in economic and environmental analysis (Donaldson and Storeygard 2016; Weiss, Jacob, and Duveiller 2020). Remote sensing allows you to analyze objects on the Earth's surface without physical contact, being basically of two types: active and passive. Passive sensors are normally optical, capturing the electromagnetic reflectance of targets while active sensors, such as radar and LIDAR, emit their signals and capture the backscatter effects. Remote sensing data has four types of resolution: (i) – spatial (pixel size); (ii) – temporal (image frequency); and (iii) – spectral (electromagnetic spectrum bands); (iv) radiometric (sensitivity).

In this paper, we combined optical images from Landsat-8 and radar images from Sentinel-1 from which we extract features to perform the classification. We then collect training and testing samples, with the most relevant variables selected in the training stage. Finally, we validated our model on the test sample and used it for classification and mapping.

A.1.1 Images Selection and Composition from Landsat-8 and Sentinel-1

First, we selected Landsat-8 and Sentinel-1 images to compose the database and then mask the clouds based on the "pixel_qa" quality band, whose pixel value (322) has no such interference. Next, we composite the annual images with the median pixels from thecollection on the Google Earth Engine (GEE) platform. Using annual composites and cloud-free images, we integrated Landsat-8 (30 m) and Sentinel-1 (10 m) data, standardizing the resolution at 30 m to prevent artificial spatial autocorrelation. The images were pre-processed through GEE's proprietary repositories, ensuring suitability for analysis (Gorelick et al. 2017).

A.1.2 Feature Selection and Extraction

From Landsat-8, we used the following spectral bands: (i) – blue; (ii) – green; (iii) – red; (iv) – infra-red. These spectral information capture important biophysical relationships. For example, the vegetation biomass is related to the red band - which captures photosynthetic efficiency - and the near-infrared band - identifies the accumulation of biomass. In this context, it is common to use vegetation indices that relate the spectral information captured by the sensors to the health, development, and phenological stages of the vegetation.

From Sentinel-1[10] we use the dual-polarized C-band: (i) – VV single co-polarization (vertical transmit/vertical receive) and (ii) – dual cross polarization VH (vertical transmit/horizontal receive). Sentinel-1 captures information from the Earth's surface at a time-frequency of 6 days considering the equator and at decreasing intervals as it moves towards the poles.

Vegetation backscatter values are primarily influenced by leaf angle, size, and water content, while soil backscatter reflects moisture levels and surface roughness. Additionally, these indices capture aboveground biomass and the three-dimensional structure of the soil-canopy complex.

10. For more information about the state of the art on radar images and their applications, see Sano, Matricardi, and Camargo (2020).

Beyond its ability to penetrate cloud cover, Sentinel-1's Synthetic Aperture Radar (SAR) provides complementary electromagnetic information, enabling the identification of distinct crop development characteristics. Consequently, integrating optical and radar sensors significantly enhances classification accuracy (Meroni et al. 2021).

We also generate a linear spectral model of the bare soil fraction and the oil palm fraction with the Landsat images to reduce the confusion between these two classes, and between oil palm and forest areas. The spectral mixing model separates the spectral signatures of different materials contained in a pixel, resulting in components that are called *endmembers*, which in practice are pure pixels of a target on the earth's surface (Somers et al. 2011).

Next, we calculated vegetation and texture indices for the bands of the Landsat-8 images in R, using the Rtoolbox and GLCM packages. Vegetation indices are obtained from arithmetic operations between the different bands of remote sensing images to capture vegetation growth and structure, as well as soil characteristics and other information related to vegetation development and health. In particular, we calculated the following vegetation indices: (i) – Difference Vegetation Index (DVI); (ii) – Ratio Vegetation Index (RVI); (iii) – Greenness Index (GI); (iv) – Normalized Difference Vegetation Index (EVI); (v) – Enhanced Vegetation Index (EVI); (vi) – Soil-Adjusted Vegetation Index (SAVI). Among them, it is worth mentioning: (i) – NDVI, combines information from the red and infrared bands to capture the biomass, however, it presents saturation from a certain level; (ii) – EVI, combines blue, red, and infrared to correct atmospheric interference and thus reduce NDVI saturation; (iii) – GCVI, near-infrared and green to capture chlorophyll concentration to identify nutritional deficit.

Finally, we also computed texture indices from the spectral bands of Landsat-8 using the gray-level co-occurrence matrix (GLCM), which captures key geometric patterns essential for accurate class discrimination. In particular, we calculated the following characteristics of GLCM with a 7x7 moving window: (i) – Contrast; (ii) – Angular Second Moment (ASM); (iii) – Correlation; (iv) – Entropy. According to Xu et al. (2021), these additional information is essential to improve the classification of oil palm due to its unique geometric characteristics. We also used Sentinel-1 images to calculate three indicators from the SAR backscattering values : (i) – ratio (VV/H); (ii) – difference (VV-VH); (iii) normalized difference (NDI) – (VV-VH)/(VV+VH). The effort resulted in 19 variables that was used to train the classification algorithms.

A.1.3 Sampling Process

To detect and map oil palm plantations, we collected a training sample of 6,194 observations through visual interpretation of Landsat-8 images. Samples were categorized into three classes: Vegetal Formation, Bare Soil, and Oil Palm. Vegetal Formation includes primary forests and secondary vegetation, while Bare Soil encompasses areas altered by human activity, such as urban infrastructure, agriculture, pasture and roads.

Sampling followed a random selection approach, yielding the following distribution (in pixels): Vegetal Formation (3,330), Oil Palm (1,564), and Bare Soil (1,300). The dataset was validated using high-resolution Google Earth imagery. Finally, we randomly split the sample into 80% for training and 20% for testing to ensure model robustness.

A.2 Machine Learning

The primary goal of Machine Learning[11] is to develop models for prediction or classification. Unlike traditional statistical methods, which emphasize asymptotic theory and causal

11. See Burger (2018) for a general overview of Machine Learning algorithms, Kamusoko (2019) for remote sensing applications, and Holloway and Mengersen (2018) for agricultural and environmental studies.

inference, Machine Learning prioritizes predictive and classification accuracy. There are two main approaches: supervised and unsupervised learning. In this study, we employed supervised algorithms to classify remote sensing data, assigning land cover classes to each pixel. To determine the most effective method, we tested five complementary algorithms: K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), and Random Forests (RF).

To implement this approach, the model is first trained to minimize misclassification while avoiding overfitting. This requires splitting the dataset into two subsets: one for training and another for testing. The testing sample is then used to evaluate the model's classification accuracy. To reduce potential bias, we applied sampling techniques, allocating 80% of the data for training and 20% for testing.

To check the robustness of the results, we use a k-fold cross-validation method, with 5-fold, to ensure that the testing data represents our data sample. This technique splits the data sample into k chunks and creates, for each chunk, a training and testing sample and estimates the model. Then, it takes the average of the k predicted errors from each chunk. In other words, the k-fold cross-validation enables to access the potential classification variation due to sampling. Next, we adopted the Recursive Feature Elimination (RFE) method, a feature selection technique that iteratively removes the least important variables to enhance model performance. RFE trains a model, ranks feature importance, eliminates the weakest predictors, and repeats this process until the optimal subset is identified. This approach improves accuracy, reduces dimensionality, and mitigates overfitting.

Then, we compare the classification results by applying the trained model to the test sample to obtain the performance of the estimations in a sample that was not used in the training. Finally, we compared each classification algorithm based on the classification's overall accuracy level and the Kappa coefficient.

A.3 Post-classification processing

After classification, we applied a 7×7 spatial mode filter to each pixel to reduce the salt-and-pepper effect, a common issue in pixel-based classifications where isolated misclassified pixels introduce noise and fragmentation in the image. This effect occurs because pixel-wise classification ignores spatial dependencies, leading to inconsistencies in neighboring pixels (Liang, Li, and Zhao 2021). To address this, we used the spatial mode filter from the *raster* package in R, which assigns each pixel the most frequent class within its surrounding window. This process enhances spatial coherence and improves the overall accuracy of the classification.

Then, we performed a manual reclassification of palm oil areas based on visual interpretation, considering their geometric characteristics, color, texture, and spatial patterns. This step was necessary to correct residual misclassified pixels within palm oil cultivation areas that were not fully addressed by the spatial mode filter. The validation and adjustment process was conducted using high-resolution imagery from Google Earth, allowing for a more detailed comparison between the classified map and actual land cover. Common misclassifications included small patches of bare soil or vegetation being labeled as oil palm and vice versa, often due to spectral similarities or transitional land cover conditions. By refining the classification through expert visual assessment, we minimized classification noise, improved spatial coherence, and enhanced the overall accuracy of the final land cover map, ensuring a more precise and reliable delineation of oil palm plantations.

Finally, we identify land use and land cover transitions between our classes at the pixel level using the *OpenLand* package in R, which quantifies temporal changes in landscape composition. This method compares categorical land cover maps from different periods, generating transition matrices that measure the area or percentage of land shifting from one class to another (e.g.,

forest to palm oil). By assessing the extent, direction, and patterns of change, this approach provides a detailed understanding of land use dynamic.

A.4 Results

Table 3: Machine Learning Algorithm Performance

	Overall Accuracy		
Algorithm	**2014**	**2017**	**2020**
K-Nearest Neighbor	92.54%	94.61%	93.87%
Artificial Neural Network	89.84%	83.51%	78.44%
Decision Tree	86.50%	84.34%	86.74%
Support Vector Machine	94.03%	94.20%	93.29%
Random Forest	94.53%	94.28%	95.53%
	Kappa Coefficient		
Algorithm	**2014**	**2017**	**2020**
K-Nearest Neighbor	0.8737	0.9088	0.8956
Artificial Neural Network	0.8284	0.7201	0.6313
Decision Tree	0.7678	0.7305	0.7707
Support Vector Machine	0.8993	0.9022	0.8867
Random Forest	0.9075	0.9031	0.9239

Source: Prepared by the authors.

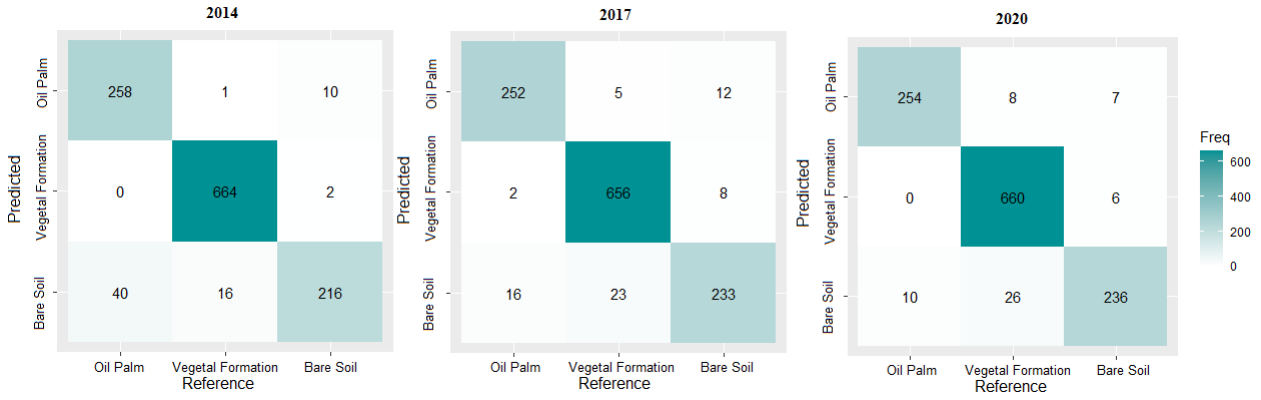

Figure 4: Confusion matrices for the test sample.

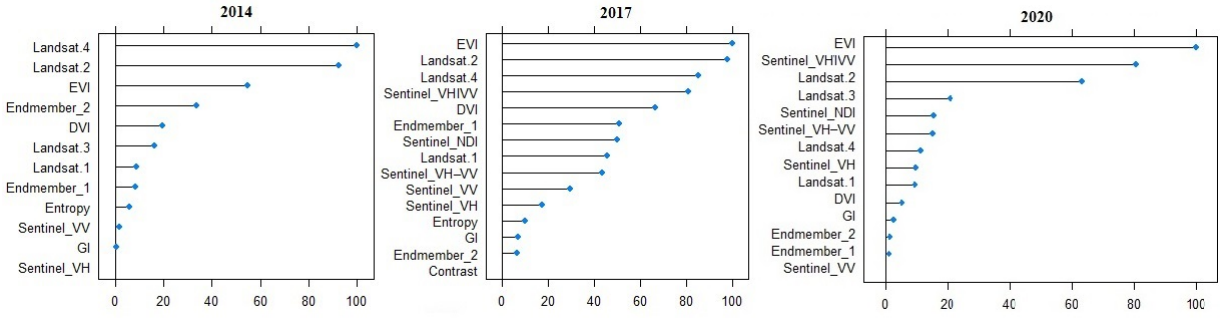

Figure 5: Features – importance for the RF classification.

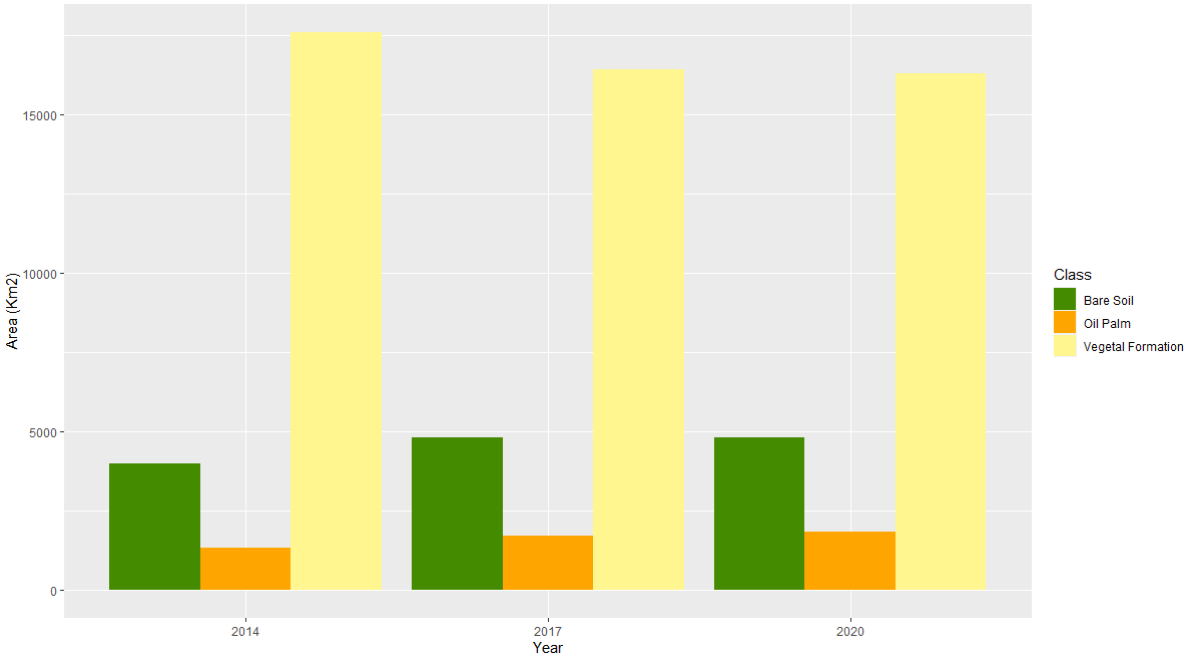

Figure 6: Classes total area in km².

796

797

Table 4: First-Stage Results: Predicting Palm Oil Expansion

	Logit (1)	**Logit (2)**	**Scobit (3)**	**Scobit (4)**
Palm Oil Yield	0.0029^{***}	0.0032^{***}	0.0029^{***}	0.0035^{***}
	(0.0004)	(0.0009)	(0.0004)	(0.0008)
Geographic	No	Yes	No	Yes
Yield potential	No	Yes	No	Yes
Socioeconomic	No	Yes	No	Yes
Observations	39,497	39,497	39,497	39,497

Notes: This table presents first-stage regression results predicting palm oil expansion. The instrument is the palm oil potential yield. Columns (1) and (2) present first-stage estimates based on logit models, while Columns (3) and (4) are estimated using scobit models. Coefficients marked with *** are significant at the 1% level. Standard errors, shown in parentheses, are clustered at the pixel level.