Universitat de les Illes Balears

DOCTORAL THESIS
2016

TOURISM AND ENVIRONMENTAL ECONOMICS

QUANTITATIVE SPATIAL ANALYSIS OF DEFORESTATION IN LEGAL AMAZON: SELECTED TOPICS

Tomas Jusys

Thesis Supervisor: prof. William Nilsson
Thesis tutor: prof. Catalina Natividad Juaneda Sampol

Doctor by the Universitat de les Illes Balears

Abstract

The dissertation covers five topics on deforestation in Legal Amazon. The first study investigates spatial heterogeneity of deforestation determinants at municipality level. Spatial differences are assessed by geographically weighted regression. The distances between regression points are measured in travel time. The computation is done programmatically. Different drivers of deforestation emerge in different locations of Legal Amazon. For Pará and its surroundings, cattle market is an especially strong driver of deforestation. Crop cultivation leads to forest clearings only in a relatively small area, located in southeastern Pará and northeastern Mato Grosso. Rural credit constraints are effective in curbing deforestation in Pará. Here less deforestation happens where more forests are legally protected, where precipitation levels are favorable for agriculture and at lower altitudes. U-shaped environmental Kuznets curve is concluded for the entire region. However, significant links are found only in Amazonias, Roraima, Pará and its proximities. Timber value motivates deforestation in most parts of the Amazon biome. Official roads contribute to deforestation in Amazonias, Roraima and their surroundings. Adverse effect of unofficial roads on extant forests is especially evident in northern Rondônia and northeastern Pará. Links between rural population and deforestation are very strong for western parts of Rondônia and Mato Grosso, but are very weak in Pará. The implementation of economic distances relative to Euclidean distances changes the results significantly for some regions. The second article investigates whether sugarcane expansion in southern Brazil exports deforestation into the Amazon. This indirect land use change is captured using spatial Durbin model. The parameters are estimated by fixed-effects regression. The results indicate that sugarcane expansion exported 16.3 thousand km^2 (12.2%) of deforestation during period 2002-2012, which is equivalent of 189.4 Mg of carbon emissions. The third study contributes to the polemics of whether rural population is linked with deforestation on forest edges. Empirical strategy is as follows: Pará state is partitioned into 5x5 km grids, only cells that classify as forest frontier are retained, links between deforestation and its covariates (including rural population) are investigated both parametrically (fractional logistic regression) and non-parametrically (regression tree). The results confirm positive link between the size of rural communities and deforestation on forest frontiers. Both methods suggest that deforestation is positively linked with cattle herd size and distance to the most proximate river and negatively linked with forest cover and precipitation. Regression tree also reveals that deforestation within protected areas is substantially lower. The fourth paper quantifies avoided deforestation in Pará's protected areas, on their edges and in their peripheral areas (buffer zones) by matching. Location characteristics are converted into a single propensity score by the means of logistic regression. Pará avoided ~2900 km^2 of deforestation during 2000-2004. Space has huge implications: conservation units in remote regions do not avoid deforestation, whereas protected areas near deforestation hotspots save substantial areas of forests. Avoided deforestation is positive in buffer zones located to the west of highway BR-163 and on the banks of Amazon River, and negative in buffers located in eastern Pará. Boundaries of conservation units are well protected from edge effects. The last study maps deforestation at 5x5 km grids in selected territory in Rondônia from past and time-fixed factors. Eigenvector-based spatial filtering is applied to solve spatial autocorrelation problem and to improve mapping accuracy. Output values of trained artificial neural network satisfactory correlate with actual values (correlation coefficient is 0.79).

Resumen

La tesis cubre cinco tópicos distintos sobre la deforestación en Amazonia Legal. El primer estudio investiga la heterogeneidad espacial de determinantes de deforestación a nivel municipal. Las diferencias espaciales son evaluadas por regresión geográficamente ponderada. Las distancias entre puntos de regresión están medidas por tiempo de viaje. Según el estudio, la ganadería afecta a la deforestación más en Pará y sus alrededores. El cultivo de las cosechas aumenta la deforestación sólo en el sureste de Pará y nordeste de Mato Grosso. Las restricciones de crédito rural son una medida eficaz contra la deforestación en Pará. Aquí hay menos deforestación en las zonas con más bosques bajo protección legal, con niveles de precipitación favorables para la agricultura y en alturas más bajas. La relación entre PIB per cápita y deforestación sigue la curva en forma U. El valor de la madera explica la deforestación en la mayoría de las regiones del bioma Amazónico. En general, las carreteras contribuyen a la deforestación más en las regiones remotas. Los vínculos entre población rural y deforestación son más fuertes en el norte de Rondônia y norte de Mato Grosso. La implementación de las distancias por tiempo de viaje con respecto a las distancias Euclidianas cambia los resultados significativamente para algunas regiones. El segundo artículo investiga si la expansión de la caña de azúcar en el sur de Brazil exporta deforestación a la frontera. Los vínculos indirectos entre caña de azúcar y deforestación están capturados por modelo espacial de Durbin. Los parámetros están estimados por regresión de efectos fijos. La caña de azúcar exportó 16.3 miles de km^2 (12.2%) de deforestación durante el periodo 2002-2012, el equivalente de 189.4 Mg de las emisiones de carbono. El tercer estudio prueba empíricamente la declaración que la población rural está positivamente relacionada con la deforestación en los bordes del bosque. El estado de Pará se divide en cuadrículas de 5x5 km. Las relaciones entre deforestación y sus determinantes (incluyendo población rural) están investigados por dos métodos: regresión logística fraccional y árbol de regresión. Los resultados confirman que el tamaño de las comunidades rurales está relacionado con la deforestación. Además, ambos métodos sugieren que la deforestación está vinculada positivamente con el tamaño del rebaño bovino y la distancia al río más cercano, y negativamente con la cubierta forestal y la precipitación. El árbol de regresión revela que la deforestación es significativamente más baja dentro de las áreas protegidas. El cuarto artículo cuantifica la deforestación evitada en las áreas protegidas, en sus bordes y en sus zonas parachoques en Pará por método de pareamiento. Las características de localidades están convertidas en un único puntaje de propensión por regresión logística. Pará evitó ~2900 km^2 de deforestación durante 2000-2004. El espacio tiene implicaciones importantes: unidades de conservación en las regiones remotas no evitan la deforestación, mientras que áreas protegidas ubicadas cerca de los focos de deforestación salvan grandes áreas de bosques. La deforestación evitada es positiva en las zonas parachoques ubicadas hacia el oeste de la autopista BR-163 y en las orillas del río Amazonas, y negativa en las zonas parachoques situadas en el este de Pará. Los límites de las unidades de conservación están bien protegidos de los efectos de borde. El último estudio simula deforestación en cada cuadrícula de 5x5 km. Mediante filtrado espacial los vectores propios que solucionan el problema de autocorrelación espacial (y automáticamente mejoran la precisión de la predicción) están identificados. Los valores de deforestación están calculados por la red neuronal artificial. El coeficiente de correlación entre los valores reales y los simulados es 0.79).

Resum

La tesi cobreix cinc tòpics distints sobre la deforestació a l'Amazònia Legal. El primer estudi investiga la heterogeneïtat espacial de determinants de deforestació a nivell municipal. Les diferències espacials són avaluades per regressió geogràficament ponderada. Les distàncies entre punts de regressió estan mesurades per temps de viatge. Segons l'estudi, la ramaderia afecta la deforestació més a Pará i els seus voltants. El cultiu de les collites augmenta la deforestació només a la zona ubicada al sud-est de Pará i nord-est de Mato Grosso. Les restriccions de crèdit rural són una mesura eficaç contra la deforestació a Pará. Aquí hi ha menys deforestació a les zones amb més boscos amb protecció legal, amb nivells de precipitació favorables per a l'agricultura i a altures més baixes. La relació entre PIB per càpita i deforestació segueix la curva en forma d'U. El valor de la fusta explica la deforestació a la majoria de les regions del bioma Amazònic. En general, les carreteres contribueixen a la deforestació més a les regions remotes. Els vincles entre població rural i deforestació són més forts en el nord de Rondônia i nord de Mato Grosso. La implementació de les distàncies per temps de viatge respecte de les distàncies Euclidianes canvia els resultats significativament per a algunes regions. El segon article investiga si l'expansió de la canya de sucre a les regions d'Amazònia Legal fora del bioma Amazònic exporta deforestació al bioma. Els vincles distants entre canya de sucre i deforestació estan capturats per model espacial de Durbin. Els paràmetres estan estimats per regressió d'efectes fixos. La canya de sucre exportà 16.3 milers km^2 (12.2%) de deforestació durant el període 2002-2012. El tercer estudi prova empíricament la declaració que la població rural està positivament relacionada amb la deforestació a les voreres del bosc. L'estat de Pará se divideix en quadrícules de 5x5 km. Les relacions entre deforestació i els seus determinants (incloent població rural) estan investigats per dos mètodes: regressió logística fraccional i arbre de regressió. Els resultats confirmen que la mida de les comunitats rurals està relacionada amb la deforestació. A més, ambdós mètodes suggereixen que la deforestació està vinculada positivamente amb la mida del ramat boví i la distància al riu més pròxim, i negativament amb la coberta forestal i la precipitació. L'arbre de regressió revela que la deforestació és significativament més baixa dins les àrees protegides. El quart article quantifica la deforestació evitada a les àrees protegides, en els seus límits i a les seves zones para-xocs a Pará pel mètode d'aparellament. Les característiques de localitats estan convertides en una única puntuació de propensió per regressió logística. Pará evità ~2900 km^2 de deforestació durant 2000-2004. L'espai té implicacions importants: unitats de conservació a les regions remotes no eviten la deforestació, mentre que àrees protegides ubicades a prop dels focus de deforestació salven grans àrees de boscos. La deforestació evitada és positiva a les zones para-xocs ubicades cap a l'oest de l'autopista BR-163 i a les voreres del riu Amazonas, i negativa a les zones para-xocs situades a l'est de Pará. Els límits de les unitats de conservació estan ben protegits dels efectes de límit. El darrer estudi simula deforestació a cada quadrícula de 5x5 km. Mitjançant filtrat espacial els vectors propis que solucionen el problema d'auto correlació espacial (i automáticament milloren la precisió de la predicció) estan identificats. Els valores de deforestació estan calculats per la xarxa neuronal artificial. El coeficiente de correlació entre els valors reals i els simulats és de 0.79).

Table of contents

List of figures

List of tables

Acknowledgements

I would like to express my sincere gratitude to the University of the Balearic Islands for financial support that lasted 4 years. I greatly thank my supervisor prof. William Nilsson for continuous support and detailed comments and suggestions toward improving every manuscript, especially in the field of econometrics. I also thank William for organizing internal seminars, which gave me the opportunity to present my work to ampler audiences and receive feedback from other professors. Additionally, I give credit for prof. Andreu Sansó Roselló for detailed review of chapter 2, prof. José Luis Groizard for detailed review of chapter 5, prof. Tomás del Barrio Castro and Andrii Bodnar for useful overall comments. Finally, I thank my fellow Ernestas Zabarauskas for writing a computer script that collects Google distances and Stefânia Costa from IMAZON for providing the shapefile of Amazon's unofficial roads. All remaining errors and omissions are my own.

Chapter 1

Introduction

Massive rainforest deforestation takes place in a limited number of countries. Nonetheless, the consequences are global. Large amount of carbon stock is released into the atmosphere. Thus, forest clearings fuel the process of global warming, which is a widely discussed topic in today's political summits. Besides, continuous deforestation threatens a variety of endemic species and nature's genetic resources in general. Some reasons leading to deforestation are also global. Deforestation is often fueled by commercial agriculture, and the scope of agriculture is defined by the global demand of agricultural commodities. Global implications of deforestation underlie its importance and promote interest into the topic.

Brazil faces the largest annual deforestation in terms of total area cleared among all countries. Deforestation in Brazil came with colonization. By the end of the 19th century, most of the Atlantic forests in Brazil's northeast, south and center-south were cut. Agriculture was primarily responsible for further clearings in center-western and northern regions during the second half of the 20th century (Araujo et al., 2011). Early governments of Brazil favored colonization into the Brazilian Amazon. It wasn't until the eighth decade of the 20th century that deforestation raised serious concerns for governmental institutions. However, massive deforestation continued to soar, reaching its peak in 2004. Since then, combined efforts of the Brazilian government and NGOs coupled with global economic crisis led to a significant reduction in deforestation rates. Despite this, current level of forest loss remains a huge treat to environmental sustainability.

Deforestation is a complex phenomenon. It is influenced, among other factors, by economic activities, infrastructure layout, demographics, terrain characteristics and legal enforcement. The connections between deforestation and its determinants sometimes are bidirectional and often manifest indirectly through other phenomena. Occasionally, deforestation is affected by distal factors. Most importantly, causes of deforestation cannot be generalized to all locations. As a consequence, policy measures to mitigate deforestation can be effective only if local contexts are properly addressed. Spatial heterogeneity of the processes that affect deforestation remains incomprehensively researched. Therefore, the main objective of this dissertation is to empirically investigate the interactions between deforestation and various factors by addressing spatial differences.

This dissertation is comprised of five self-contained studies. The first study investigates direct and underlying causes of deforestation and how those factors affect deforestation in different locations across Legal Amazon. The second work answers the question whether sugarcane expansion in southern Brazil exports deforestation into the Amazon. The third investigation empirically tests the theory that rural communities are responsible for deforestation on forest frontiers. The fourth paper builds on researches into the influence of legal forest protection on deforestation by measuring avoided deforestation in conservation units, in buffer zones, and on the edges of conservation units. The last study aims to map deforestation from past and time-fixed variables by exploiting spatiotemporal contagion of deforestation.

Causes of deforestation, both direct and underlying, are widely analyzed in academic literature. Key direct causes of deforestation in Legal Amazon, as indicated by most investigators, are agriculture and infrastructure. Agriculture consists of livestock (mostly cattle) and crop cultivation (mostly soybean and sugarcane) businesses. Agriculture in Brazil relies heavily on the credit system. The official rural credit portfolio covers about a third of the annual financial needs of the agricultural sector in Brazil (Assunção et al., 2013b). Rural credit is loaned in accordance to rules and conditions issued by the Central Bank of Brazil. However, this money in the hands of farmers may fuel deforestation. In response to such fears, in 2008 Resolution 3545 was released, which conditioned the concession of rural credit for use in agricultural activities in the Amazon Biome upon presentation of proof of borrowers' compliance with environmental legislation, as well as of the legitimacy of their land claims and the regularity of their rural establishments. Among documents needed was the declaration stating the absence of current embargoes caused by economic use of illegally deforested areas (Assunção et al., 2013b). However, rural credit concessions do not necessarily increase deforestation. As Assunção et al. (2013b) notice, crop farmers are likely to invest a larger share of rural credit loans in the intensification of production, instead of expanding it by operating in the extensive margin as cattle ranchers do. Indeed, there are important differences between cattle ranching and crop cultivation in terms of pressure on forests. Geo-ecological barriers are in general more restrictive in the case of crop cultivation (Margulis, 2004), making cattle ranching the predominant industry. Numerically, ranching enterprises occupy roughly 75 percent of the deforested areas of Legal Amazon. The key restriction for plant cultivation is high rainfall (the other important constraint is steep slope), causing most problems during seeding and harvesting. Topographic and climatic characteristics vary across Legal Amazon, and, as a result, land suitability for crop cultivation. This is a strong argument in favor of models that capture spatial heterogeneity.

Another widely recognized determinant of deforestation is road network. A classic example of road-induced deforestation is the Trans-Amazonian highway, opened in the eighth decade of previous century. Thus, not constructing a road is a way to prevent deforestation. However, even bigger concern is the network of unofficial roads. These roads are generally built to open up forests to illegal logging, thus leading to new colonization, forest fragmentation, ecological degradation and increased fire risk (Barber et al., 2014).

Market growth and foreign trade are often named as contributors to deforestation. Both crop cultivation and cattle ranching satisfy both national and international markets. International demand of Brazilian products increases the need to deforest. Some efforts to mitigate international pressure on Brazilian rainforests have been made. For example, under the pressure of retailers and NGOs, major soybean traders signed Brazil's Soy Moratorium, which is a voluntary agreement not to purchase soy grown on lands, deforested after July of 2006. However, weaknesses in federal enforcement aggravate the potential of this initiative. Currently, for more than half of registered properties with embargoes producer identification is inconsistent with the system of Rural Environmental Registry of private properties (Gibbs et al., 2015b). This system is used by soy traders to check for embargoes. Since information is inconsistent, transactions with properties under embargoes continue to happen. Similarly, meatpacking companies in Pará began signing the legally binding Terms of Adjustment of Conduct (known as TAC), committing to purchase cattle only from ranchers registered with the Pará State Rural Environmental Register. Furthermore, in 2009 Brazil's largest slaughterhouses (JBS-Bertín, Marfrig and Minerva) signed an agreement with Greenpeace not to purchase meat production from ranches with deforestation (Gibbs et al., 2015a). The initiative only considers direct suppliers, thereby leaving plenty of room to circumvent commercialization restrictions. Trade of agricultural products, especially beef, continues to soar. The biggest importers of Brazilian beef are Russia, Hong Kong, Venezuela and Egypt (SECEX-MDIC, 2015). Arguably, ever increasing demand of agricultural products is a consequence of growing populations in foreign trade partners, because larger populations imply higher meat consumption.

Naturally, the size of population emerges as a potential direct cause of deforestation. Here the distinction is often made between rural and urban communities. Indeed, there is an ongoing debate among scholars, whether rural or urban population contributes to deforestation to a higher extent. Authors, who argue that urban population growth rather than rural population growth is a stronger accelerant of deforestation, defend their position by arguing that urbanization raises consumption levels and increases demand for agricultural products. DeFries et al. (2010) claim that urban consumers generally eat more processed foods and animal products than rural consumers, thereby stimulating commercial production of crops and livestock. On the contrary, Wright & Muller-Landau (2006) find that recent deforestation rates are positively related to local rural population density, and that the percent of the remaining forests is often negatively related to rural population density in the tropics. Key arguments in favor of rural-driven deforestation are immigration and high natality rate (Izquierdo et al., 2011).

The relationships between deforestation and its determinants are region-specific (Margulis, 2004), but most academic studies ignore this fact. To the best knowledge of the author, only Oliveira & Almeida (2011) investigated the causes of deforestation from local perspective in Legal Amazon at county level. This was achieved by applying geographically weighed regression (GWR). However, these authors were restricted to the limited capabilities of the software that estimates GWR, which led them to make two very restrictive assumptions. Specifically, those assumptions are: 1) all covariates are strictly exogenous, and 2) Euclidean distances reflect well the communications between municipalities.

As already discussed, deforestation is potentially explained by the size of populations. However, as Angelsen & Kaimowitz (1999) notice, growing populations affect labor market, technological progress and institutional changes. Thus, deforestation itself may attract new inhabitants as a result of these changes. Similar argument applies to national income. However, its effect on deforestation can be both positive and negative. Higher national income creates additional job opportunities outside agricultural sector, thus reducing pressure on forests, but higher income also increases the production of agricultural and timber products, thus stimulating deforestation. Furthermore, deforestation itself is a source of income. Therefore, both population and national income shall be treated as endogenous in deforestation modeling.

The application of GWR method requires distances between the regression points (municipality seats in this dissertation). Euclidean distances, most often used by the researchers, may not be a proper representation of communications inside Legal Amazon. Most logged woods are transported by roads in trucks. However, these roads in Legal Amazon often are winding due to geographical constraints, financial gains in building a single road that stretches through multiple towns, or unwillingness to pave a road that crosses intact forests to avoid potential deforestation. Moreover, roads are not homogenous in quality: some roads are highways and some are not paved. This has important implications on travel time. Furthermore, Legal Amazon hosts around 100 towns that do not have access by roads. These towns are located on the banks of the Amazon River or other major river in Legal Amazon. Connections between these towns are based on water transport. Rivers, especially smaller, meander to a high extent, thus extending the distance and time required to travel to a destination.

Chapter 2 investigates the main causes of deforestation in Legal Amazon. The findings of the linear model under endogenous treatment of population and GDP per capita are presented. Further, GWR based on economic (travel time) distances is applied to assess spatial heterogeneity of deforestation determinants. This is done programmatically. The code is written in Gauss platform, and is author's own elaboration. Also, a comparative analysis on how the implementation of economic distances instead of straight line distances changes the results is presented.

Complete analysis of deforestation determinants must consider the possibility that cause and effect are separated in space. It is widely theorized that in Brazil crop planters displace cattle ranchers into the Amazon, thereby indirectly contributing to deforestation. The displacement happens because the farmers residing in non-frontier regions sell their pastures to crop businesses and move to the frontier, where they purchase land from local smallholders. This phenomenon is known as indirect land use change (ILUC). To understand the reasons of ILUC, it is necessary to discuss the economic and geo-climatic context of Brazil. The territory of Brazil can be partitioned into two major zones that differ in conditions for agriculture. For simplicity I refer to those as agricultural and frontier zones. The former are located outside the Amazonian Biome and are characterized by high land prices and unavailability of productive land (especially, in closer proximities to the frontier). In most parts both crop cultivation and livestock farming can be practiced due to favorable climatic and topographic conditions. Conversely, the frontier is covered by vast areas of available and inexpensive land (forests to be cleared). However, due to high level of rainfall, crops cannot be readily cultivated. In economic terms this means that the frontier has a comparative advantage for livestock farming as opposed to crop cultivation. An economic factor that encourages farmers to sell land properties to crop planters in agricultural zones and move to the frontier is high differential in land prices (Sawyer, 2008).

Various crops are cultivated in Brazil, but the most widespread are soybean and sugarcane. Sugarcane is the main source of ethanol biofuel. Ethanol is widely used in Brazil in the transport industry. Its production is likely to expand further due to the potential size of the domestic market and to the opportunities for exporting (Walter et al., 2014). Due to the fact that growing sugarcane absorbs more carbon than is emitted when the ethanol is burned as fuel, ethanol is considered as a potential solution to global warming problem. The later statement relies on the assumption that sugarcane has no influence on deforestation. Empirical findings from the Amazon region generally show that sugarcane and deforestation are not directly linked. However, in case sugarcane planters displace livestock farmers into the Amazon, and, as a result, export deforestation, the superiority of ethanol over fossil fuel in terms of reducing CO_2 emissions may be overstated.

Much of the academic literature focuses on ILUC associated with soybean, the primary crop in Brazil. Several studies indirectly linked soybean with deforestation in the Amazon. However, empirical evidence of ILUC associated with sugarcane is scant. Therefore, chapter 3 empirically tests the hypothesis that sugarcane planters in southern Brazil contribute to deforestation in the Amazon Biome, and measures the magnitude of the effect. The study uses 2002-2012 panel dataset. Indirect linkages between sugarcane expansion in southern Brazil and deforestation are captured using spatial Durbin model. The parameters are estimated by fixed-effects regression.

Large producers who move to the frontier purchase lands from smaller producers. Questions, such as where those small producers come from and whether they contribute significantly to deforestation have to be answered. Rural family children in Legal Amazon follow one of the three alternatives in terms of migration. Some move to urban areas to seek off-farm employment, which enables to diversify risks and overcome credit constraints. Some new generation rural inhabitants remain on the farm. Finally, some move to forest edges. Even though those are relatively few, Carr & Burgdorfer (2013) theorize that rural farmers residing near forest edges have a disproportionally large adverse effect on extant forests, since newly arrived rural migrants often engage in expansive agriculture due to cheap family labor, scarce capital, low technology, high cost of transportation and insecure land tenure. The latter implies that as soon as new migrants arrive, they establish their farming systems and cut down trees to demonstrate land claims (Simmons et al., 2003). Later, these farmers seek official recognition of the land which they developed. As more rural migrants arrive, land availability decreases. As it happens, some farmers move to another undisturbed location on forest edge. Their previously developed lands are sold for large scale producers. Those lands are then consolidated to create large enough areas for cattle grazing or cultivation of commercial crops.

The relationship between deforestation and rural population may be different on forest frontiers and in long-settled rural areas. Long-settled rural areas are covered by large pastures or crop fields. The consumer of agricultural commodities that originate from those areas is often a foreign citizen. Therefore, distal demands of agricultural production is what control land use change in old rural settlements. This, in turn, implies that local rural population here is not a significant component of deforestation function or at least drives forest clearings to a lesser extent than on forest edges. Forest edges are characterized by abundant and undisturbed forest resources. As discussed, here migrant rural settlers engage in extensive deforestation, thereby opening large previously inaccessible areas for colonization. An additional rural migrant here creates a lot of pressure on standing forests. As a result, a separate analysis is needed to understand the associations between deforestation and rural demographics on forest edges.

Chapter 4 presents such a study, which empirically tests the argument of Carr & Burgdorfer (2013) that rural settlements are linked with deforestation on forest edges. Selected study area encompasses the state of Pará. To properly capture the associations between deforestation and its covariates on forest edges fine scale analysis is needed. Therefore, the territory of Pará is partitioned into 5x5 km grids. Besides rural population, all relevant and available covariates are included. The analysis is based both on parametric (fractional logistic regression) and non-parametric (regression tree) approaches. The findings also complement the analysis in chapter 2, as the links between deforestation and its explanatory factors in some instances may be scale-dependent (Pan & Carr, 2010).

As discussed, land insecurity in Brazil creates incentives to deforest (Araujo et al., 2011) and is a consequence of weaknesses in legislation. The Land Statute of Brazil states that squatters[I] who are developing a land during at least five consecutive years and are not in a conflict with landowners can claim formal property title over that land. Moreover, the Brazilian Constitution of 1988 states that unproductive establishments can be taken over and redistributed to other parties. Such legislation creates incentives to clear forests, since forests are considered as unused lands.

One of key measures in combating land insecurity in Brazil is the protected area system. The legal framework of the system was established in 2000 by the National Protected Areas System Law (SNUC). In 1998 the Amazon Region Protected Areas Program (ARPA) was formulated. The program foresaw the establishment of 15 conservation units under strict protection during four-year period (between 2000 and 2003). Later, in 2005 and 2006, the network of protected areas in Legal Amazon was expanded drastically, especially in the state of Pará. However, protected area itself is only a part of protection mechanism. Conservation units are surrounded by buffer zones. A buffer zone is a peripheral area around a conservation unit and is meant to benefit local populations by allowing low environmental impact activities. In this way it is expected that local inhabitants will be involved in the protection of a conservation unit near their residence. However, the establishment of protected areas may provoke displacement of deforestation locations (substitution effect). It is probable that loggers move to the surroundings of a newly created conservation unit instead of entering the protected territory, which they would have entered if that territory had remained unprotected. Therefore, avoided deforestation in buffer zones can be both positive and negative depending on whether effective buffer zone management or substitution effect dominates. Buffer zones also shed protected areas from edge effects. An edge effect in this dissertation is described as an excessive deforestation of park's edge relative to its internal area. There are at least two reasons why an edge of a conservation unit is subjected to higher deforestation risk: 1) it borders unprotected areas or, even worse, open areas without vegetation, and 2) deforested fields in nearby areas of a conservation unit are dry and can easily catch fire, affecting forests on the edge of a conservation unit. Therefore, buffer zones constitute a shield from those risks. Furthermore, buffer zone management has important implications in ecology, since buffer zones and ecological corridors ensure that species have sufficient habitat to survive and that those species can migrate through the forests. The importance of buffer zones is well understood by the Brazilian government. The Government of Brazil together with the Pilot Program for the Brazilian Rainforests allocated almost 23 million US dollars for buffer zone management under the four-year ARPA project. However, despite the immense importance of buffer zones, their implications on avoided deforestation did not receive sufficient attention from scientists. A meticulous study on avoided deforestation in conservation units of Legal Amazon is found in Nolte et al. (2013), but the study did not consider buffer zones.

[I] Squatters are individuals who invade lands and develop them, but hold no property rights over those lands.

To fill the gap of knowledge, chapter 5 investigates the implications of buffer zone management on avoided deforestation in Pará. The percentages of avoided deforestation are estimated by propensity score matching. Edge effects are tested by comparing avoided deforestation on park's edge and in its nuclear area (park's territory beyond its edge). Most importantly, buffer zones around protected areas are considered separately for each protected area or a group of protected areas that share borders. Location matters primarily due to the differences in deforestation pressure: some conservation units are located on deforestation frontiers and some are located in remote areas, where forest protection has only a cartographic meaning.

Knowing the factors responsible for deforestation and applying measures to counteract it is necessary, but does not suffice. Deforestation monitoring and prediction is equally important. Until 2004 the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) relied on voluntary reports of logging events. As a result, IBAMA could not locate deforestation at its roots. Finally, in 2004 a near real-time system of deforestation detection (DETER) was introduced. DETER is a satellite-based system that captures and processes georeferenced imagery on forest cover in 15-day intervals (Assunção et al., 2013a). The system uses LANDSAT imagery and is capable of detecting deforested areas that are larger than 25 hectares. Newer images are compared with older images to identify changes in forest cover. The imagery is prepared in a form of georeferenced digital maps. Deforestation identified by those maps is verified by field inspections. Some campaigns to prevent deforestation are initiated by NGOs. For example, Greenpeace uses an airplane to locate illegal log rafts and reports observed rafts to the authorities. In addition, Greenpeace developed a technique based on ultraviolet paint to track illegally chopped woods back to exporting companies.

Deforestation monitoring helps to detect deforestation in its initial stage and prevent forests from further exploitation. Chapter 6 offers a detailed methodology to map deforestation from past and time-fixed processes. Mapping is done at 5x5 km grids. To achieve successful results, several aspects have to be taken into consideration. Firstly, deforestation is spatially and temporarily contagious. Spatial contagion implies that deforestation in the neighboring locations increases the probability of deforestation in the reference location. Temporal contagion implies that deforestation that happened in recent past is likely to continue into the future. To account for this spatiotemporal contagion, a deforestation function includes past deforestation and its focal variables. However, contagion creates an econometrical problem known as spatial autocorrelation. Under the presence of spatial autocorrelation, linear model errors include both white noise and unobserved covariates. Filtering those important covariates from the errors of the initial model could greatly improve mapping accuracy. Several methods to accomplish this task have been developed, but the most modern and the most promising technique is eigenvector-based spatial filtering. Eigenvectors represent spatial patterns of different spatial autocorrelation levels. Therefore, filtered eigenvectors are artificial constructs of unobserved covariates and are simply added to the list of explanatory variables. To refine mapping accuracy, variables that control for legal protection, terrain and climatic characteristics, and infrastructure are also included. Deforestation values in chapter 6 are simulated using an artificial neural network. The method captures nonlinearities between deforestation and its determinants and, as a result, provides more precise estimates.

The main findings of this dissertation are synthesized and policy recommendations are given in chapter 7. Chapter 8 includes appendices. Chapter 9 presents programming codes.

1.1. References

Angelsen, A., & Kaimowitz, D. (1999). Rethinking the Causes of Deforestation: Lessons from Economic Models. *The World Bank Research Observer, 14*(1), 73–98.

Araujo, C.E., Araujo Bonjean, C., Combes, J.L., Combes, P.M., & Reis, E.J. (2010). *Does Land Tenure Insecurity Drive Deforestation in the Brazilian Amazon? Etudes et Documents*. Clermont-Ferrand, France: CERDI.

Assunção, J., Gandour, C., & Rocha, R. (2013a). *DETERring Deforestation in the Brazilian Amazon: Environmental Monitoring and Law Enforcement. Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Assunção, J., Gandour, C., Rocha, R., & Rocha, R. (2013b). *Does Credit Affect Deforestation? Evidence from a Rural Credit Policy in the Brazilian Amazon. Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Barber, C.P., Cochrane, M.A., Souza Jr., C.M., & Laurance, W.F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation, 177,* 203-209.

Carr, D., & Burgdorfer, J. (2013). Deforestation Drivers: Population, Migration, and Tropical Land Use. *Environment, 55*(1).

DeFries, R.S., Rudel, T., Uriarte, M., & Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience, 3,* 178-181.

Gibbs, H.K., Munger, J., L'Roe, J., Barreto, P., Pereira, R., Christie, M., … & Walker, N.F. (2015a). Did Ranchers and Slaughterhouses Respond to Zero-Deforestation Agreements in the Brazilian Amazon? *Conservation letters.* doi: 10.1111/conl.12175

Gibbs, H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., … & Walker, N.F. (2015b). Brazil's Soy Moratorium. Supply-chain governance is needed to avoid deforestation. *Science, 374,* 377-378.

Izquierdo, A.E., Grau, H.R., & Aide, T.M. (2011). Implications of Rural–Urban Migration for Conservation of the Atlantic Forest and Urban Growth in Misiones, Argentina (1970–2030). *Journal of Human Environment, 40*(3), 298-309.

Margulis, S. (2004). Causes of Deforestation of the Brazilian Amazon. In W. B. w. paper (Ed.), *World Bank Working Paper* (Vol. 22).

Nolte, C., Agrawal, A., Silvius, K.M., & Soares-Filho, B.S. (2013). Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. *Proceedings of the National Academy of Sciences, 110*(13), 4956-4961.

Oliveira, R.C., & Almeida, E. (2011). Deforestation in the Brazilian Amazonia and Spatial Heterogeneity: a Local Environmental Kuznets Curve Approach. In 57th Annual North American Meetings of the Regional Science Association International. Retrieved from http://www.poseconomia.ufv.br/docs/Seminario08-10-2010ProfEduardo.pdf

Pan, W.K., & Carr, D. (2010). Population, Multi-scale Processes, and Land Use Transitions in the Amazon. Proceedings of the European Population Conference. Retrieved from http://geog.ucsb.edu/~carr/PDFs_added_ Oct_31/pop_multiscale_processes.pdf

Sawyer, D. (2008). Climate change, biofuels and eco-social impacts in the Brazilian Amazon and Cerrado. *Philosophical Transactions of the Royal Society B, 363,* 1747–1752.

SECEX-MDIC (Foreign Trade Secretariat Database, Ministry of Development, Industry and Foreign Trade) (2015). *Brazilian Beef Exports.* Retrieved from http://www.mdic.gov.br/sitio/

Simmons, C.S., Perz, S., Pedlowski, M.A., & Silva, L.G.T. (2003). The changing dynamics of land conflict in the Brazilian Amazon: The rural-urban complex and its environmental implications. *Urban Ecosystems, 6,* 99-121.

Walter, A., Galdos, M.V., Scarpare, F.V., Leal, M.R.L.V., Seabra, J.E.A., Cunha, M.P., … & de Oliveira, C.O.F. (2014). Brazilian sugarcane ethanol: developments so far and challenges for the future. *Energy and Environment, 3*(1), 70-92.

Wright, S.J., & Muller-Landau, H.C. (2006). The Future of Tropical Forest Species. *Biotropica, 38*(3), 287-301.

Chapter 2

Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon[II]

Abstract. This study explores the main direct and underlying causes of deforestation in Brazil's Legal Amazon region by considering spatial differences. The computation of localized parameters is based on geographically weighted regression (GWR). The novelty of this paper lies in its incorporation of economic, rather than Euclidean, distances into the GWR. Economic distances are measured by travel time, sourced from Google Inc. A global approach revealed several important factors that affect deforestation, including: rural population, GDP (suggesting a U-shaped environmental Kuznets curve), forest stock, cattle ranching, timber value, and road networks (both official and unofficial). Local analysis uncovered patterns not seen under global models, especially in the state of Pará. Most notably, crop cultivation was found to accelerate deforestation in southeastern Pará and northeastern Mato Grosso, while in some regions (especially in the northeastern corner of Pará), the area covered by crop plantations was negatively associated with deforestation. For Pará, rural credit constraints, larger territories designated as sustainable use areas and indigenous lands, and higher levels of precipitation inhibit deforestation. Further, rural population has a very heterogeneous impact on deforestation across Legal Amazon: it is not a significant factor of deforestation in northern Pará and Amapá, but it has a relatively strong effect in the western parts of Mato Grosso and Rondônia. Also, official and illegal roads create significantly more pressure on forests in remote regions compared to developed areas. Finally, the use of economic distances, as opposed to Euclidean distances, leads to notably different GWR results.

Keywords. Deforestation, Legal Amazon, Google time distances, spatial heterogeneity, GWR.

[II] This artcile has been published. Publication details are: Jusys, T. (2016). Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon. *Applied Geography, 75*, 188-199.

2.1. Introduction

Reducing emissions from deforestation, a major source of CO_2, could be a highly cost-effective option for climate policy (Rametsteiner et al., 2009). Tropical deforestation also has other negative externalities, such as the loss of biodiversity, erosion, floods, and lowered water levels (Espindola et al., 2011). As such, research into causes of deforestation has a long history. Some studies investigate a specific cause of deforestation. For instance, Arima et al. (2011), Barona et al. (2010), Macedo et al. (2012), and Morton et al. (2006) study the effect of agriculture on deforestation; Barber et al. (2014) and Pfaff et al. (2007) investigate links between road networks and deforestation, and Carr & Burgdorfer (2013) discuss the implications of rural populations on forest clearing. Araujo et al. (2010), Bhattarai & Hamming (2001), Culas (2012), and Ehrhardt-Martinez et al. (2002) test for an environmental Kuznets curve. Assunção et al. (2013a) focus on causality between rural credit concessions and deforestation. Soares-Filho et al. (2006) assess the impact of protected areas. Araujo et al. (2010) investigate the effects of land insecurity on forests. Faria & Almeida (2016) explore the relation between openness to trade and deforestation. Other studies investigate the general causes of deforestation, including Aguiar et al. (2007), Hargrave & Kis-Katos (2013), Laurance et al. (2002), and Reis & Guzman (1993). The most relevant empirical findings on the drivers of deforestation were surveyed by Angelsen & Kaimowitz (1999) and Geist & Lambin (2002).

Evidence from empirical case studies that identify both proximate causes and underlying forces at work on tropical deforestation suggests that no universal link between cause and effect exists (Geist & Lambin, 2002). This is because policy is made at village, county, state, and national levels, rather than consistently over an area (Carr et al., 2012). The most popular technique to account for variability over such large land masses is called geographically weighted regression (GWR), developed by Brunsdon et al. (1998). Applications of GWR in deforestation and forest loss modeling can be found in Carr et al. (2012), Jaimes et al. (2010), Moon & Farmer (2012), Oliveira & Almeida (2011), and Witmer (2005). However, only Oliveira & Almeida (2011) applied GWR to the situation in Legal Amazon.

The objective of this study is to investigate the causes of deforestation in Legal Amazon, but with two important differences compared to Oliveira & Almeida (2011). Firstly, this study considers gross domestic product and demographic variables as endogenous, following recommendations by Angelsen & Kaimowitz (1999) and Kaimowitz & Angelsen (1998). Secondly, the weighting is based on economic, rather than geographical, distances, which are measured by travel time.

The extent of similarities between the results obtained by applying different distance measurement methods depends on the topographic characteristics of geographical regions, road networks, and an area's economic development, among other factors. If a territory is large, economically advanced, and has highly populated urban areas, straight lines are appropriate and represented by distances traveled by plane. However, in large, densely forested areas with numerous villages, plane connection is not cost-effective. Under such cases, ground transport is the only viable means of transportation. Here, the terrain itself is an important factor. For instance, in mountainous or densely forested areas, roads are winding (see Figure

A1), which leads to significant differences between road and straight line mileages. Moreover, considering only the mileage may be too restrictive, since roads are heterogeneous in quality and type. Undoubtedly, highways provide much faster access than dirt roads that cut through the landscape. Additionally, Legal Amazon has almost one hundred villages which can only be accessed via the river network, implying that access to these villages is slower than it would be in the presence of roads. Therefore, travel time is the most appropriate way to measure distances between locations in Legal Amazon.

2.2. Data and methods

The data covers 486 municipalities in Legal Amazon. Observations for which forest coverage was lower than 5% of the territory were omitted. Some municipalities were removed from the dataset because of a lack of information. Data available as shapefiles or in fine scale raster grids was aggregated to the level of municipalities in ArcGIS (Version 10.0, ESRI, Redlands, CA, US). Squares of GDP per capita were computed to test an environmental Kuznets curve. Past population and GDP per capita variables were used as instruments. This is a cross-sectional analysis, and the study year is 2010. Brief descriptions, units of measurements, and data sources are presented in Table 1. See Appendix B for municipal level descriptive statistics. Furthermore, it was verified that no severe multicollinearity between the covariates exists (Table C1). Nevertheless, the correlation coefficients reveal notable collinearity between rural credit per capita and GDP per capita, rural credit per capita and crop area, and official and unofficial roads.

As far as data regarding distance is concerned, the computation of straight line distances was based on decimal coordinates of municipality capitals, reported by the IBGE. Road and time distances used in the research are the property of Google Inc., located at 1600 Amphitheatre Parkway, Mountain View, CA 94043, United States. The distances are measured between municipality seats. The data was extracted using The Google Distance Matrix API service (Google Developers, 2014). However, almost one hundred locations in Legal Amazon do not offer road access. Therefore, distances by rivers between roadless municipality seats were computed in ArcGIS. The computations were based on a river shapefile, downloaded from GEOFABRIK, OpenStreetMap. Occasionally, the traveler, who aims to travel from one roadless location to another, may opt for a river journey from an initial village without a road to the nearest location with a road, then travel by road as far as is possible and make the final part of the trip on the river again. However, changing means of transport is not desirable and would pay off only when considering longe distances. Miscalculations of long distances have little to no effect on the results of GWR, because those distances are lightly weighted. Finally, to fill in the empty gaps made up of distances between locations without roads and locations with road networks, it was assumed that the traveler always prefers travelling by roads over travelling by rivers. Thus, any distance of this kind is measured as the distance by river between the initial roadless location to the nearest port accessible via roads, plus the distance by road between the port and the final destination. To calculate river travel times between ports, data on all fluvial routes, offered by the transportation company Cris Transporte Marítimo, was used to compute the average speed of passenger transport boats in the Amazon River and its tributaries. The speed proved to be relatively constant across routes (45 km/h). This figure was used to complete the time-disance matrix.

Table 1. Description of the variables

Abbreviation	Description	Unit	Source
DEF	Annual deforestation increments. Data on the municipal level is available on the INPE's website. It is aggregated from PRODES maps, which are distributed at a 60-meter spatial resolution and are created by digital image processing and visual interpretation of LANDSAT ™ imagery on computer screens.	km^2	INPE (2014)
POPURB	Number of urban inhabitants from 2010 census	count	IBGE (2014)
POPRUR	Number of rural inhabitants from 2010 census	count	IBGE (2014)
GDP	Gross domestic product per capita in 2010	R$ (BRL)	IBGE (2014)
FCOVER	Extant forests	%	INPE (2014)
ELEV	Average elevation over 90 square meter cells that fall within the borders of a municipality	meter	SRTM (2014)
CATTLE	Cattle (bovines)	count	IBGE (2014)
CROP	Planted acreage of temporal (yearly) crops	ha	IBGE (2014)
TIMBER	Value of all timber products	R$	IBGE (2014)
ROF	Total length of official roads, excluding urban streets and roads under construction	km	GEOFABRIK (2014)
RUNF	Total length of unofficial roads	km	IMAZON[III]
CREDIT	Sum of rural credit per capita, issued by official banks and credit cooperatives	R$	Central Bank of Brazil (2013)
TENURE	Percentage of private properties in total properties	%	IBGE (2014)
PREC	Annual precipitation over municipality seat (computed as in Arima et al., 2011).	mm	TRMM (2016)
STRICT	Territory designated as strict protection areas (IUCN categories I, II and III)[IV]	%	WDPA (2015)
SUST	Territory designated as sustainable use areas (IUCN categories IV, V and VI)	%	WDPA (2015)
INDIG	Territory designated as indigenous lands	%	WDPA (2015)
TERR	Territory of a municipality	km^2	IBGE (2014)
A	Autocovariate (normalized weighted sum of deforestation in the neighbors)	km^2	Compiled by the author
POPURBLG	Number of urban inhabitants from 2000 census	count	IBGE (2014)
POPRURLG	Number of rural inhabitants from 2000 census	count	IBGE (2014)
GDPLG	Gross domestic product per capita in 2009	R$ (BRL)	IBGE (2014)

INPE: Brazil's National Institute of Space Research; SRTM: Shuttle Radar Topography Mission; IBGE: Brazil's Institute of Geography and Statistics; IMAZON: Amazon's Institute of Human and Environment; TRMM: Tropical Rainfall Measuring Mission; WDPA: World Database on Protected Areas

[III] See acknowledgements.

[IV] For some records in the attribute table of the WDPA, a IUCN (International Union for Conservation of Nature) category of protected areas in the shapefile is not reported or is reported incorrectly. IUCN categories were taken from the cadastre of protected areas, managed by the Brazilian Ministry of Environment (MMA, 2016).

Prior to the application of GWR, it must be verified that the errors of the global regression are independent and identically distributed. Failure to meet this condition leads to increased type I errors, thus aggravating hypothesis testing and prediction (Dormann et al., 2007). To check for spatial autocorrelation of model errors, Moran's I (Moran, 1950, Cliff & Ord, 1972) was calculated (see Appendix D). To correct for non-randomness in the spatial distribution of model errors, the initial list of covariates was supplemented by an autocovariate term (eq. 1), which was treated as an exogenous covariate (in the remainder of this study the global model without the autocovariate is referred to as 2SLS-IV, and the model with the autocovariate as SEM).

$$a_i = \frac{\sum_{j \in k_i} w_{ij} y_j}{\sum_{j \in k_i} w_{ij}} \quad (1)$$

Here, y_j is the response value at site *j* among site *i's* set of k_i neighbors and w_{ij} is the weight, which expresses site *j's* influence over site *i* (Dormann et al., 2007). The weights are calculated as inverse distances. The number of neighbors (*k*) is selected so that it minimizes the standard score. Further, to account for endogenous relations, the study implemented an instrumental variables approach (one instrument per endogenous covariate). After weighting the data, the GWR-IV estimator is:

$$\hat{\ }_i = \left[Z^T W_i X \right]^{-1} Z^T W_i y \quad (2)$$

Here, *X* is a matrix of all values of exogenous and endogenous variables with an implicitly written intercept, *Z* is a matrix of all values of exogenous and instrumental variables with an implicitly written intercept, *y* is a vector of response values, and W_i is a weighting matrix for observation *i*. The *T* symbol indicates the transpose. The weights were computed by applying a kernel function and then placing weighted distances from location *i* to all the other locations into the main diagonal of a weighting matrix, thus creating one weighting matrix per observation (eq. 3). The chosen kernel function is adaptive and Gaussian (eq. 4), and it was proposed by Brunsdon et al. (1998) as one of the options for GWR. Adaptive kernel functions are suggested when the spatial density of regression points (municipality seats) is not even. Figure A1 reveals that in some parts of Legal Amazon, municipality seats are concentrated in a particular area, whereas in other parts, municipality capitals are sparsely distributed.

$$W_i = \begin{pmatrix} k_{i1} & 0 & \dots & 0 \\ 0 & k_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & k_{in} \end{pmatrix}, \text{ where } k_{ii} = 1 \quad (3)$$

$$k_{ij} = \exp\left(-\frac{d_{ij}}{h_{i(p)}} \right), \; k_{ij} \in (0,1] \quad (4)$$

Scalar h in eq. 4 represents bandwidth. Here bandwidths are the distances to the p^{th} nearest neighbor. The optimal number of nearest neighbors was estimated by minimizing the cross-validated sum of squares (eq. 5). The computation, as developed by Brunsdon et al. (1998), begins by excluding the i^{th} observation from the model to avoid minimization at zero nearest neighbors. In computational terms, it means replacing the value of the k_{ii}^{th} element with zero. The optimal number of nearest neighbors gives a vector of bandwidths, which is plugged into eq. 2 to obtain local estimates of the coefficients. Then, it must be evaluated whether those estimates are sufficiently heterogeneous. There is adequate spatial variability if: 1) the variability in local coefficients of a covariate (eq. 6) exceeds the variance of the global coefficient corresponding to that covariate, or 2) the null hypothesis of a Monte Carlo simulation that all local coefficients are equal is rejected (for details see Brunsdon et al., 1998).

$$cvss(p) = \sum_{i=1}^{n} \left(y_i - x_i \,\breve{}_i (p) \right)^2 \quad (5)$$

$$v_j = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{}_{ij} - \overline{}_j \right)^2 \quad (6)$$

Vector x_i in eq. 5 is a row vector of the original values of exogenous and endogenous covariates of the i^{th} observation; y_i stands for the i^{th} element of the response vector. The ˘ symbol indicates the exclusion of the i^{th} observation from eq. 3; v_j in eq. 6 represents the variability of local coefficients of covariate j, and β_{ij} *hat* denotes the local coefficient of the i^{th} observation of covariate j. β_j *bar* is the average over local coefficients of covariate j. Local standard errors are computed by eq. 7 (see Fotheringham et al., 2002):

$$VCV\left(\hat{}_i \right) = C_i C_i^T \pm^2 \quad (7),$$

$$\text{where } C_i = \left[Z^T W_i X \right]^{-1} Z^T W_i \quad (8) \text{ and } \pm^2 = \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2}{n - tr(S)} \quad (9),$$

$$\text{where } \hat{y}_i = x_i \left[Z^T W_i X \right]^{-1} Z^T W_i y \quad (10)$$

VCV is the variance-covariance matrix; *tr(S)* is the trace of the hat matrix, and $\hat{y}_i$ is the predicted value of deforestation in the i^{th} location. The hat matrix S is formed by horizontally concatenating all s_is. The computations were based on the author's own programming code, written in *Gauss 10*.

2.3. Results

The errors of 2SLS-IV model (without the autocovariate term) are clustered (Table 2). After the inclusion of the autocovariate term, the null hypothesis that model errors are distributed randomly, cannot be rejected.

Table 2. Spatial autocorrelation and Moran's I

Model	# neighbors	Moran's I	E(MI)	10^3 var(MI)	z score	p value
2SLS-IV	-	0.0264	-0.0021	0.0348	4.8259	0
SEM	3	0.0065	-0.0021	0.0353	1.4421	0.1493

Table 3. Results of global regressions

	2SLS-IV	SEM
Cons	9392.4108	5909.0916
	(10288.145)	(9557.9127)
POPURB	-0.0058	-0.0028
	(0.0112)	(0.0104)
POPRUR	0.2167*	0.21**
	(0.1139)	(0.1057)
GDP	-1.947***	-1.5676***
	(0.4162)	(0.389)
GDPSQ	0.0232***	0.0183***
	(0.0062)	(0.0058)
FCOVER	129.1639***	72.1396*
	(45.0999)	(42.4398)
ELEV	-15.0342	-14.0207
	(12.6409)	(11.7349)
CATTLE	0.0484***	0.0474***
	(0.0093)	(0.0086)
CROP	0.0074	0.0055
	(0.0195)	(0.0181)
TIMBER	0.2778***	0.2012***
	(0.0785)	(0.0735)
ROF	23.952***	18.0965***
	(6.3932)	(5.9744)
RUNF	5.4575***	5.4181***
	(1.7038)	(1.5817)
CREDIT	-0.3701	-0.1069
	(0.9901)	(0.9194)
TENURE	45.1036	28.3457
	(81.4303)	(75.6246)
STRICT	98.0825	90.075
	(109.8015)	(101.9447)
SUST	-60.3399	-35.8517
	(42.3415)	(39.4135)
INDIG	-37.741	-74.6585
	(58.3051)	(54.3207)
PREC	-3.5709	-3.1728
	(3.002)	(2.7872)
TERR	0.216***	0.2482***
	(0.0772)	(0.0718)
A		405.8324***
		(49.3483)

Table 3 (continued)

R squared	0.5015	0.5712
Weak identification test	***F (1, 467)***	***F (1, 466)***
POPURB	5752.61	5859.83
POPRUR	1012.26	1018.2
GDP	299.28	296.21
GDPSQ	178.44	177.32
Cragg-Donald Wald F stat	40.35	40.1
Endogeneity test	Chi square: 34.43 (p=0)	Chi square: 29.24 (p=0)
Instruments	POPURBLG, POPRURLG, GDPLG, GDPSQLG	

Standard errors are in parentheses
Symbols '*', '**' and '***' imply statistical significance at 10%, 5% and 1% level respectively
Coefficients and standard errors are multiplied by 1000
Variable abbreviations are introduced in Table 1
GDPSQ: GDP per capita squared in 2010
GDPSQLG: GDP per capita squared in 2009

The percentage of explained variation in deforestation increased from 50% to 57% upon implementation of the SEM model (Table 3). A statistically significant coefficient on the autocovariate justifies the use of this model. The endogeneity test concludes that variables assumed to be endogenous should be treated as such, i.e., the null hypothesis that all variables treated as endogenous in the model could have been treated as exogenous is rejected. The weak identification test indicates that the selected set of instrumental variables is a strong set.

Global models reveal several contributors to deforestation, including rural population, forest stock, cattle ranching, the timber market, official and unofficial roads, and deforestation in the neighboring municipalities (captured by the autocovariate). Further, the relation between deforestation and GDP per capita is found to follow a U-shaped curve: more income is linked with less deforestation at lower levels of GDP per capita and with more deforestation at higher levels. The break-even sum was 3570 R$/month in 2010 (equivalent to ~1530 €/month per capita at 2010 prices). The break-even point was estimated by equating the first derivative of the deforestation function with respect to GDP per capita to zero and solving for GDP per capita. The global coefficient for territory is statistically significant, but the finding does not have economic meaning: the variable was only included in the analysis to control for heterogeneous sizes of municipalities. Insignificant factors under the global approach are: urban population, elevation, crop cultivation, rural credit schemes, land tenure, legal protection, and precipitation.

The cross-validated sum of squares under travel time distances is minimized at 15 nearest neighbors (the average bandwidth is 316.4 minutes). The quasi-global R^2 of the GWR regression is 89%, which constitutes a significant improvement over the global model fit of 57%. See

Figure E1 for local R^2 values (the formula for computing local R^2 can be found in Fotheringham & Brunsdon, 1999). The Monte Carlo simulation detected sufficient spatial variation in the local coefficients of forest cover, elevation, crop cultivation, strict legal protection, and territory (Table 4)[V]. In all cases however, variabilities of local coefficient estimates exceed the variances of corresponding global coefficients. This indicates that there is some justification for considering spatial variation patterns in all coefficients.

Table 4. Variability of local coefficients

	$\sqrt{v_j}\times 1000$	$ste(\ _j)\times 1000$	$\sqrt{v_j}\ /\ ste(\ _j)$	**p Monte Carlo**
Cons	14251.972	9557.9127	1.49	0.5926
POPURB	0.0335	0.0104	3.23	0.6399
POPRUR	0.3222	0.1057	3.05	0.323
GDP	0.5359	0.389	1.38	0.9815
GDPSQ	0.0086	0.0058	1.48	0.9568
FCOVER	111.3014	42.4398	2.62	0.0535*
ELEV	34.392	11.7349	2.93	0.0597*
CATTLE	0.0402	0.0086	4.66	0.2551
CROP	0.1751	0.0181	9.67	0.035**
TIMBER	0.3769	0.0735	5.13	0.4486
ROF	12.9506	5.9744	2.17	0.6132
RUNF	6.7584	1.5817	4.27	0.1584
CREDIT	2.6553	0.9194	2.89	0.2078
TENURE	89.34	75.6246	1.18	0.8683
STRICT	292.0364	101.9447	2.86	0.0412**
SUST	57.6968	39.4135	1.46	0.3416
INDIG	110.674	54.3207	2.04	0.4239
PREC	5.8732	2.7872	2.11	0.2737
TERR	1.063	0.0718	14.8	0.0062***
A	191.6645	49.3483	3.89	0.7984

Symbols '*' and '**' and '***' imply statistical significance at 10%, 5% and 1% level respectively
Variable abbreviations are introduced in Table 1

Urban population is significantly and negatively linked with deforestation only in Acre, the southeastern corner of Pará, and the northeastern corner of Mato Grosso. Elsewhere, the links are statistically insignificant. Rural population *per se* was found to contribute to deforestation the most in western Rondônia and western Mato Grosso (Figure 1). Also, the contribution is

[V] Note that statistically significant variation does not imply that all or some of the local coefficients are statistically significant. Sufficient variation in local coefficients is concluded if the null hypothesis of coefficient equality is rejected. Statistical significance of a coefficient is concluded if the null hypothesis stating that a coefficient is equal to zero is rejected. It is entirely possible that the latter hypothesis is not rejected (for all local coefficients), whereas the former hypothesis is rejected.

notable in Acre. The size of rural communities is not a significant determinant of deforestation in mid and eastern Rondônia, mid Mato Grosso, northwestern Pará and Amapá, and most parts of Maranhão. GDP per capita is linked with deforestation in Amazonias, Roraima, Pará, northern Amapá, and the northeastern corner of Mato Grosso. However, spatial variation is low, and the local relation between income and deforestation almost exclusively follows a U-shaped environmental Kuznets curve. Forest stock explains deforestation the most in the northeastern corner of Pará. The link is statistically insignificant in western and southeastern Pará, most parts of Roraima, Rondônia, and Mato Grosso, and in mid Amazonias. Elevation is significantly and negatively linked with deforestation almost exclusively in Pará and Amapá. The highest extent of cattle ranching contributing to deforestation occurs in Pará[VI]. The contribution is insignificant in Acre, Rondônia, and Mato Grosso.

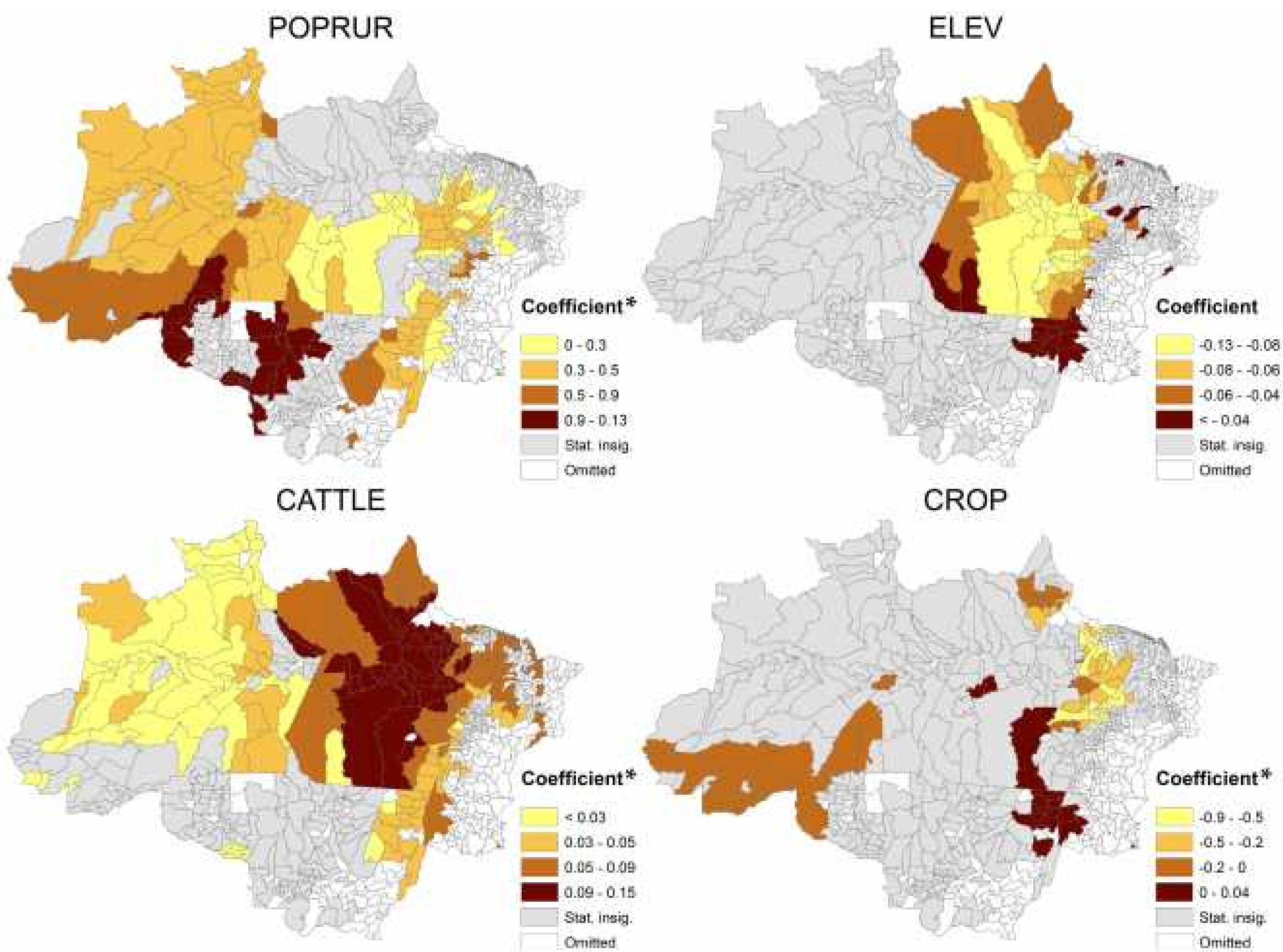

Figure 1. Spatial distribution of local coefficients of selected variables. Stars in the legends indicate that the coefficients are multiplied by 1000. *Stat. insig.* implies that a coefficient is statistically insignificant at 10% level (*p* value is greater than 0.1). Variable abbreviations that appear above the maps are introduced in Table 1. This 2010 municipality boundary map was downloaded from the IBGE's website.

[VI] Official statistics of cattle heads, which were used in the model, are flawed for some municipalities. For instance, Margulis (2004) showed that the IBGE significantly overestimates the density of animal units in Paragominas. Therefore, the results for particular municipalities should be interpreted with caution. Nevertheless, the general pattern is likely to be well represented by the map.

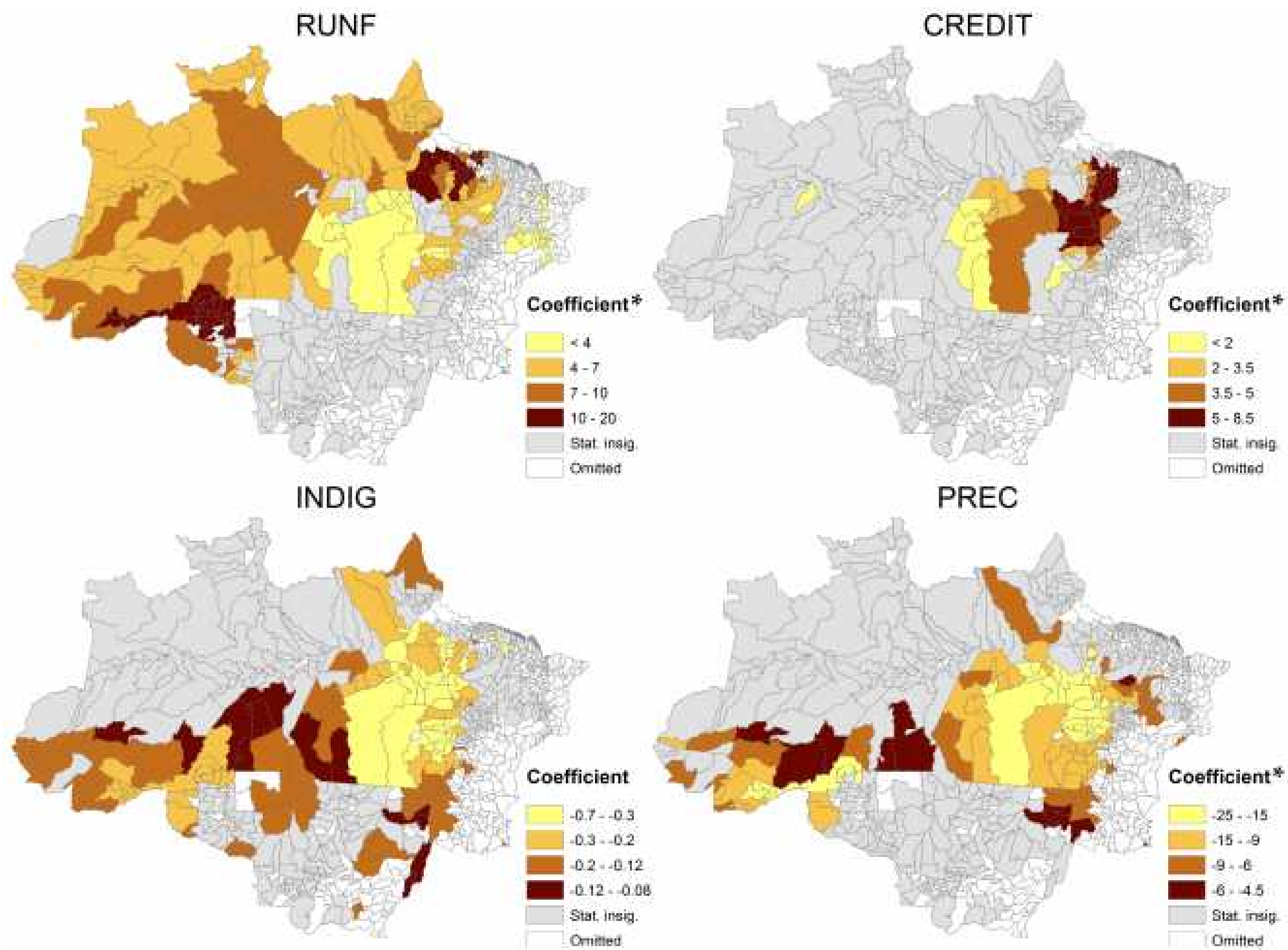

Figure 1 (continued)

Crop cultivation is negatively and significantly linked with deforestation in Acre, southern Amapá, and northeastern Pará and positively and significantly linked in a few municipalities in southeastern Pará and northeastern Mato Grosso. Local coefficients of the timber market show little spatial variation in areas where these coefficients are statistically significant. Timber value has an insignificant effect on deforestation in Acre, Rondônia, and Mato Grosso. Official roads drive deforestation in distant Amazon regions, mostly in Amazonias, Acre, the western edge of Pará, and most parts of Amapá. Spatial variation of local coefficients is relatively low. Unofficial roads significantly contribute to deforestation in all regions where such roads exist, except for Mato Grosso. However, the strongest impact was found in northern Rondônia and the northeastern corner of Pará. Rural credit schemes are significantly and positively linked with deforestation in Pará state (except for the northern regions), especially in the northeastern part, where deforestation for agriculture is rampant. Local coefficients of land tenure and strict legal protection are statistically insignificant. Furthermore, in Pará and Amapá, less deforestation happens where more forests are declared to be protected sustainable use areas. Indigenous lands are significantly and negatively linked with deforestation in most parts of the Amazon biome, except for in remote areas (mostly in Amazonias and Roraima). The links are especially strong in eastern Pará. Precipitation is significantly and negatively connected with deforestation in most parts of Pará, the southern edge of Amazonias, and in eastern Acre. The neighbor effect is strongest in Pará (especially, in the northeastern part) and Amapá.

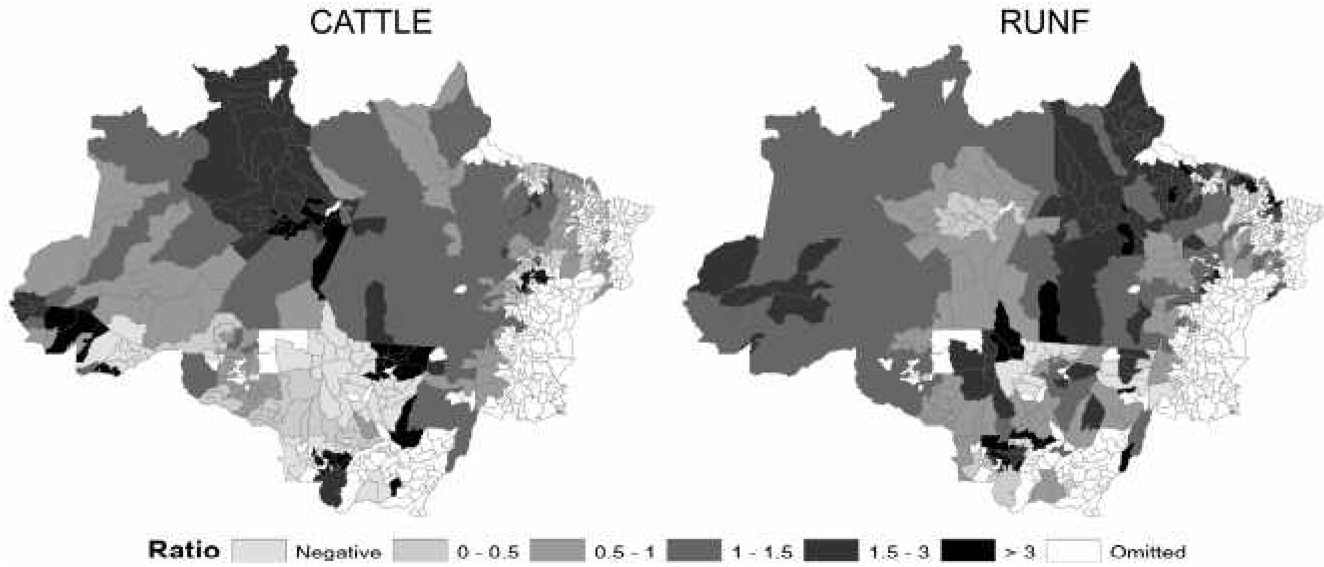

Figure 2. Comparison of GWR beta estimates under straight line and travel time distances for cattle ranching (left) and unofficial road network (right). Mapped figures are the ratios of local coefficients under straight line distances and corresponding coefficients under travel time distances. Therefore, darker colors indicate overestimation of the effect on deforestation, and lighter colors imply underestimation. This 2010 municipality boundary map was downloaded from the IBGE's website.

A comparative analysis reveals that the implementation of road/river distances does not dramatically alter the conclusions, but the results seen under straight line distances diverge significantly for some municipalities. Figure 2 contrasts those differences for cattle ranching and unofficial road networks. Among all regions where cattle ranching is a factor in explaining deforestation, straight line distances overestimate cattle contribution most evidently in Roraima and a few municipalities located in the middle of Legal Amazon, along the administrative borders of Amazonias and Pará. As for unofficial roads, the most notable overestimation is observed in Pará and Amapá, and the most evident underestimation is seen in the same region where the contribution of cattle farming is significantly overestimated. Significant differences exist for other factors as well, but these two examples suffice to illustrate that the method for measuring distance has important implications on results. These differences arise because relative distances change when using economic distances. For example, the geographical distance between Roraima and Pará is relatively short, but travel time between those locations is relatively long due to the presence of dense forests and the absence of road networks in some regions (a possible route between, say, Belém in Pará and Boa Vista in Roraima is via the Amazon river from Belém to Manaus and then by road BR-174 from Manaus to Boa Vista). Therefore, in the case of cattle ranching in Roraima, processes affecting deforestation in Pará have relatively little impact on deforestation in Roraima, and for this reason, the large adverse effect that the cattle market has on extant forests in Pará does not transfer to Roraima. However, the straight line approach cannot take into account factors such as travel by river or the absence of roads in dense forests; therefore, it overestimates the impact of cattle farming on deforestation in Roraima.

2.4. Discussion

Generally, the agricultural frontier advances at the expense of forests. The open economy and trade turns cattle ranching into a profitable business and puts severe pressure on tropical rainforests. As economies grow, individuals tend to spend more on meat products, thus increasing the demand for beef and other meats (Kaimowitz et al., 2004). As a consequence, such countries increase meat imports. Therefore, as Brazil's trading partners increase their demand for cattle production, more forests have to be cleared to supply cattle with necessary grazing areas. Cattaneo (2008) estimates that additional 18000 km^2 of land are needed annually in the Amazon to account for pasture expansion. Cattle farming is especially prevalent in eastern Pará. This region is highly urbanized, close to the main export centers, and has well developed road networks. Well capitalized farmers reside there (Margulis, 2004), and cattle ranches extensively replace forests. However, the effect of the cattle market on deforestation may have decreased since 2010, because individual meatpacking companies in Pará began signing legally binding Terms of Adjustment of Conduct (MPF-TAC) agreements in July 2009 to stop purchasing meat products from properties that partook in illegal deforestation (Gibbs et al., 2015a). Furthermore, later in 2009, Brazil's largest meatpacking companies signed the G4 agreement with Greenpeace to avoid being associated with deforestation. Gibbs et al. (2015a) showed that these agreements were effective in curbing deforestation. Nevertheless, both agreements currently govern only those properties selling directly to slaughterhouses, thereby leaving many opportunities to circumvent supply chain restrictions.

Crop cultivation is subject to climatic constraints, especially excessive rainfall (Margulis, 2004). As a result, in some areas crops cannot be grown. Therefore, crop cultivation was found to contribute to deforestation only in a certain region near the borders of Pará and Mato Grosso (Figure 1). This area lies on or near a deforestation frontier and has two favorable characteristics for crop cultivation: 1) it is located near trade centers, which increases commercialization opportunities and hence, increases profitability, and 2) average annual rainfall over that area is generally less than 2000 mm, which is favorable for crop cultivation. However, local coefficients on crop cultivation in Mato Grosso, where soybean growing prevails, are statistically insignificant. This seemingly counterintuitive finding has at least two potential explanations. Firstly, large areas in the state of Mato Grosso are either already deforested or covered by savannas. Thus, soybean cultivation does not require extensive deforestation of currently standing forests. Instead, crops are cultivated on lands previously occupied by cattle ranchers (Macedo et al., 2012). Soybean cultivation exports deforestation to other regions, especially northward, in the form of displaced cattle ranchers (Arima et al., 2011, Barona et al., 2010), thereby indirectly contributing to deforestation. Cattle ranchers have incentives to sell their lands in Mato Grosso to crop planters and purchase land near forest edges thanks to a strong differential in land prices (Sawyer, 2008). This indirect land use change may also explain significant and negative links between the area of crop plantations and deforestation in some regions, especially northeastern Pará. This area is a new frontier for soybean cultivation. Soy planters cannot advance further westward due to unfavorable climatic conditions (excessive rainfall). As a result, soybean plantations expand past cattle grazing areas, not directly causing deforestation, but at the same time displacing cattle ranchers to the nearby locations in the west where forests are cleared to accommodate grazing cattle herds. Therefore, more crop plantations are associated with less deforestation. Secondly, the decoupling of crop cultivation and deforestation is at least partially a consequence of Brazil's Soy Moratorium, which is a voluntary agreement signed by soybean traders not to purchase soy grown on lands deforested after July 2006. Gibbs et al. (2015b) showed that after the implementation of the agreement, deforestation due to soy production decreased dramatically.

Livestock farmers and crop planters in Legal Amazon rely heavily on a rural credit system. In 2008, resolution 3545 was issued. This policy set up conditions for the concession of rural credit for use in agricultural activities in the Amazon Biome upon presentation of proof of borrowers' compliance with environmental legislation, as well as proof of the legitimacy of their land claims and the regularity of their rural establishments (Assunção et al., 2013a). Also, a territorial performance approach to deforestation was adopted in which the geographical unit of intervention was the municipality instead of the individual farm (Nepstad et al., 2014). Access to agricultural credit was suspended for those farms located in priority municipalities (with the highest deforestation rates). The majority of those municipalities are located in Pará. Figure 1 suggests that rural credit restrictions are effective in curbing deforestation in Pará. Also, note that the global model (Table 3) found that rural credit is not a significant factor in explaining deforestation. This finding is simply an artifact of the failure to account for regional differences, and it underlies the importance of considering spatial heterogeneity.

Money is an important component in the deforestation function. However, the relationship between deforestation and income is often complex. The results suggest a U-shaped Kuznets curve for Legal Amazon. This result is backed by Araujo et al. (2010), but objected by Culas (2012), who finds an inverse U-shaped Kuznets curve in Latin America. The finding that richer communities stimulate deforestation is in line with Angelsen & Kaimowitz's (1999) argument that higher income increases pressure on forest resources. The conventional approach argues that the exploitation of environmental resources is the consequence of poverty. When a higher level of wealth is achieved, owners of land properties are expected to invest more in capital and labor intensive sectors. New employment opportunities should encourage individuals to shift their activities from the extraction of environmental resources to better paid jobs in the industrial, manufacturing, and tourism, etc. sectors. However, the predominant sectors in Brazil are land intensive. Therefore, higher incomes do not always lead to preventing extensive deforestation. Logging is costly, as it necessitates covering transportation and storage costs of cut down wood, as well as renting or buying machinery to log trees, paying wages for lumberjacks and other personnel involved in the process, or paying environmental fines when caught. However, agriculture may become very profitable if it is done on a large scale; this creates incentives to clear forests. Therefore, it can be argued that with lower incomes, farmers do not have sufficient funds to engage in mass deforestation. It could be speculated that money is invested in agriculture intensification, thus reducing the pressure on forests (this argument does not hold for forest edges though, where rural migrants are known to engage in expansive agriculture upon arrival). However, under a certain point of wealth (3570 R$/month) economic agents become involved in expansive agriculture, driven by the need to supply cattle with grazing lands. The average inhabitant in Legal Amazon contributes only 10000 R$ annually, to GDP (Table B1). Therefore, small scale farmers are in the downward sloping part of the GDP curve.

Agriculture is not the only sector that profits from deforestation. Economic agents also gain by selling cut wood and its derivative products. A significant and positive link between deforestation and the timber market is suggested by this investigation and is confirmed by other studies (Damette & Delacote (2011), Oliveira & Almeida, 2011, Reis & Guzman, 1993). This finding is also in line with the view expressed by Angelsen & Kaimowitz (1999). The economic explanation of why higher timber prices drive forest clearings is quite simple -– higher prices of wood mean that economic agents receive more income for the same quantity of wood sold. Opponents of this view argue that higher timber prices encourage more effective forest management. However, forests are subjected to market failures, as they are zero-priced in the market.

Large scale cash-orientated producers shape Amazonian forests by clearing land for pastures and crop fields. However, smallholders (those operating small, often subsistence, farms) in rural communities also play a role in deforestation, which is mostly manifested through rural migration. A portion of second generation rural farm children migrate to forest edges, where land is widely available. Newly arrived migrants engage in expansive land use change, which is motivated by cheap family labor, scarce capital, low technology, the high cost of transportation, and insecure land tenure (Carr & Burgdorfer, 2013). As more migrants arrive, land availability decreases. Eventually, smallholders sell their properties to larger producers, who consolidate the lands for large scale commercial agriculture. The smallholders move on to new previously untouched frontiers and begin a new round of deforestation (Izquierdo et al., 2011). The process of rural migration to forest frontiers may not be adequately captured in the municipality level analysis, but the results confirm the positive link between deforestation and rural populations in most parts of the Amazon biome.

The routes of smallholder migration are dictated by accessibility. Therefore, governments face tremendous pressure from scientists and environmentalists to halt road paving projects. Such projects commonly initiate a process of spontaneous colonization, logging, mining, land speculation, and investment (Fearnside, 2005, Fearnside & Graça, 2006, Laurance et al., 2001). Barber et al. (2014) estimate that almost 95% of Legal Amazon's deforestation happens within a distance of 5.5 km to the nearest road. New official roads often spur a large network of endogenous (unofficial) roads far into dense forests. These roads are the result of the interacting interests of migrant farmers and loggers who move to forest frontiers (Davidson et al., 2012). Therefore, the conclusion that both official and unofficial roads explain deforestation (Figure 1 and Table 3), especially in remote regions, is reasonable. In some regions with large open (deforested) areas, road networks were found to be an insignificant factor. Many roads located in these regions were built several decades ago and therefore drove previous deforestation until the lands became extensively deforested.

A popular policy measure to restrain forest frontier encroachment is the protected area system. Protected areas are managed under a wide range of legal regimes in order to achieve better ecological and social outcomes. A distinction is made between: strict protection areas that prohibit resource use, and often physical access; sustainable use areas that allow for controlled resource extraction, land use change, and in many instances, human settlements; and indigenous lands that are set up to protect the livelihood of indigenous people (Nolte et al., 2013). The three types of land restrictions were considered separately in the analysis. Sustainable use areas and indigenous lands were negatively linked with deforestation in Pará (section 3). The negative link may imply either that legal protection is an effective measure in curbing deforestation or that protected areas are located in low-deforestation regions. It is common knowledge that many protected areas are established far from deforestation hotspots and, therefore, legal protection here is only precautionary. These conservation units are expected to have a significant inhibitory effect in the future, when deforestation frontiers reach their proximities. Strict protection was not associated with less deforestation, but this finding by no means indicates that strict governance is ineffective in protecting forests. This is because the goal of protected areas is to protect a certain forest, not necessarily the whole municipality. Legal protection may be very effective in avoiding deforestation within the borders of a protected area, but concurrent deforestation in surrounding areas within the same municipality may be very extensive. Paradoxically, in the municipality level analysis, the link between legal protection and deforestation could be positive, even though legal protection effectively prevents deforestation inside protected lands. Another aspect that cannot be captured at the municipality level is the potential rearrangement of deforestation spots. Upon establishment of protected areas loggers may be displaced to other locations. This displacement should be most prevalent in young, expanding agricultural frontiers, where land tends to be cheap and abundant (Nepstad et al., 2006). However, even if deforestation is not decreased but only shifted from one place to another, the benefit of the legal protection system may lie in prioritizing the protection of ecosystems with higher ecological values.

Policies aimed at mitigating deforestation are also made at the property level. Forest Code (FC) is the central piece of legislation regulating land use and management on private properties. Created in 1965, it was transformed during the 90s into a *de facto* environmental law via a series of presidential decrees (Soares-Filho et al., 2014). The code requires landowners situated in the Amazon to set aside a legal reserve that occupies 80% of land property. Furthermore, a new FC, approved in 2012, introduced a mechanism known as the Environmental Reserve Quota. Implementing the mechanism could create a trading market for forested lands, adding monetary value to native vegetation (Soares-Filho et al., 2014). Historically however, the enforcement of FC legislation has proven to be largely ineffective (Nepstad et al., 2006). Soares-Filho et al. (2014) showed that even under Brazil's 2012 FC, most land properties located in Pará, Mato Grosso, and Rondônia do not comply with the requirements. Additional efforts to tackle illegal deforestation include The Action Plan for the Prevention and Control of Deforestation in the Legal Amazon, which introduced the use of a real-time deforestation detection system so the Brazilian Institute for the Environment and Renewable Natural Resources (IBAMA) can react in a timely manner and issue environmental fines upon verification of illegal deforestation. Even though the collection rate on environmental fines is low, they are often accompanied by other sanctions that are more binding, such as the seizure and destruction of production goods, tools, and materials, as well as embargoes on production areas (Assunção et al., 2013b).

2.5. Conclusions

Global results suggest that rural populations, forest stock, cattle business, the timber market, and road networks (both official and unofficial) are contributors to deforestation. The relation between GDP per capita and deforestation was found to follow a U-shaped environmental Kuznets curve. The cross-validated sum of squares of the GWR model, based on travel time distances, was minimized at 15 nearest neighbors. Using this method, the global R^2 increased from 57% to 89%.

Different drivers of deforestation emerge in different locations of Legal Amazon. A variety of factors explain deforestation in Pará and its surrounding areas. Here, cattle ranching has the strongest impact on deforestation among all regions in Legal Amazon. Crop cultivation contributes to deforestation only in the southeastern Pará as well as in northeastern Mato Grosso. Interestingly, the link between crop cultivation and deforestation is negative in northeastern Pará, which may be a consequence of indirect land use change associated with crop expansion past cattle grazing areas. High income is an accelerant of deforestation. Rural credit constraints curb deforestation. The clearing of forests is also motivated by timber value. The rural population in Pará has either an insignificant or a marginal effect on deforestation. Here, less clearing occurs in areas where more forests are under legal protection as well as on lands at higher altitudes and those in areas with unfavorable precipitation levels for agriculture. Roads, especially official ones, in Pará (except for in the northeastern part) and its proximities are generally weakly associated (in relative terms) with deforestation. Road networks here are mostly old, indicating that the adverse effect they have on forests has already taken place. Contrarily, in remote areas (Amazonias and Roraima) road networks strongly influence deforestation patterns. Forrest clearing here is also motivated by cattle farming, timber value, and high levels of income. More deforestation happens in regions with larger rural settlements. It is also worth noting that in the case of western and northern Rondônia, the rural population and networks of unofficial roads have the strongest contribution to deforestation across Legal Amazon. Comparative analysis revealed that the implementation of geographical distances instead of economic distances leads to notably different results for some regions, thereby highlighting the importance of the method used for measuring distances in this GWR application.

2.6. References

Aguiar, A.P.D., Câmara, G., & Escada, M.I.S. (2007). Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. *Ecological modelling, 209*, 169-188.

Angelsen, A., & Kaimowitz, D. (1999). Rethinking the Causes of Deforestation: Lessons from Economic Models. *The World Bank Research Observer, 14*(1), 73–98.

Araujo, C.E., Araujo Bonjean, C., Combes, J.L., Combes, P.M., & Reis, E.J. (2010). *Does Land Tenure Insecurity Drive Deforestation in the Brazilian Amazon? Etudes et Documents*. Clermont-Ferrand, France: CERDI.

Arima, E.Y., Richards, P., Walker, R., & Caldas, M.C. (2011). Statistical confirmation of indirect land use change in the Brazilian Amazon. *Environmental Research Letters, 6*(2), 7.

Assunção, J., Gandour, C., Rocha, R., & Rocha, R. (2013a). *Does Credit Affect Deforestation? Evidence from a Rural Credit Policy in the Brazilian Amazon. Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Assunção, J., Gandour, C., & Rocha, R. (2013b). *DETERring Deforestation in the Brazilian Amazon: Environmental Monitoring and Law Enforcement. Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Barber, C.P., Cochrane, M.A., Souza Jr., C.M., & Laurance, W.F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation, 177*, 203-209.

Barona, E., Ramankutty, N., Hyman, G., & Coomes, O.T. (2010). The role of pasture and soybean in deforestation of the Brazilian Amazon. *Environmental Research Letters, 5*(2), 9.

Bhattarai, M., & Hamming, M. (2001). Institutions and the Environmental Kuznets Curve for Deforestation: A Crosscountry Analysis for Latin America, Africa and Asia. *World Development, 29*(6), 995-1010.

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighed regression – modelling spatial non-stationarity. *Journal of the Royal Statistical Society. Series D (The Statistician), 47*(3), 431-443.

Carr, D., & Burgdorfer, J. (2013). Deforestation Drivers: Population, Migration, and Tropical Land Use. *Environment, 55*(1).

Carr, D., Davis, J., Jankowska, M.M., Grant, L., Carr, A.C., & Clark, M. (2012). Space versus place in complex human–natural systems: Spatial and multi-level models of tropical land use and cover change (LUCC) in Guatemala. *Ecological Modelling, 229*, 64-75.

Cattaneo, A. (2008). Regional Comparative Advantage, Location of Agriculture, and Deforestation in Brazil. *Journal of Sustainable Forestry, 27*(1-2), 25-42.

Central Bank of Brazil (2013). *Anuário estatístico do crédito rural (Até 2012)*. Retrieved from http://www.bcb.gov.br/?RELRURAL

Cliff, A., & Ord, K. (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis, 4*(3), 267-284.

Culas, R.J. (2012). REDD and forest transition: Tunneling through the environmental Kuznets curve. *Ecological Economics, 79*, 44-51.

Damette, O., & Delacote, P. (2011). Unsustainable timber harvesting, deforestation and the role of certification. *Ecological Economics, 70*(6), 1211-1219.

Davidson, E.A., Araújo, A.C., Artaxo, P., Balch, J.K., Brown, I.F., Bustamante, M.M.C., … & Wofsy, S.C. (2012). The Amazon basin in transition. *Nature, 481*, 321-328.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., … & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography, 30*(5), 609–628.

Ehrhardt-Martinez, K., Crenshaw, E., & Jenkins, J.C. (2002). Deforestation and the environmental Kuznets curve: A cross-national investigation of intervening mechanisms. *Social Science Quarterly, 83*(1), 226-243.

Espindola, G.M., de Aguiar, A.P.D., Pebesma, E., Câmara, G., & Fonseca, L. (2011). Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data. *Applied Geography, 32*, 240-252.

Faria, W.R., & Almeida, A.N. (2016). Relationship between openness to trade and deforestation: Empirical evidence from the Brazilian Amazon. *Ecological Modelling, 121*, 85-97.

Fearnside, P.M. (2005). Deforestation in Brazilian Amazonia: History, Rates and Consequences. *Conservation Biology, 19*(3), 680-688.

Fearnside, P.M., & Graça, P.M.L.A. (2006). BR-319: Brazil's Manaus-Porto Velho Highway and the potential impact of linking the arc of deforestation to central Amazonia. *Environmental Management, 38*(5), 705-716.

Fotheringham, A.S., & Brunsdon, C. (1999). Local Forms of Spatial Analysis. *Geographical Analysis, 31*(4), 340-358.

Fotheringham, S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. West Sussex, England: John Wiley & Sons, LTD.

Geist, H.J., & Lambin, E.F. (2002). Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *Bioscience, 52*(2), 143-150.

GEOFABRIK (2014). *OpenStreetMap Data Extracts*. Retrieved from http://download.geofabrik.de/

Gibbs, H.K., Munger, J., L'Roe, J., Barreto, P., Pereira, R., Christie, M., … & Walker, N.F. (2015a). Did Ranchers and Slaughterhouses Respond to Zero-Deforestation Agreements in the Brazilian Amazon? *Conservation letters*. doi: 10.1111/conl.12175

Gibbs, H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., … & Walker, N.F. (2015b). Brazil's Soy Moratorium. Supply-chain governance is needed to avoid deforestation. *Science, 374*, 377-378.

Google Developers (2014). *The Google Distance Matrix API*. Retrieved from https://developers.google.com/maps/documentation/distancematrix/

Hargrave, J., & Kis-Katos, K. (2013). Economic Causes of Deforestation in the Brazilian Amazon: A Panel Data Analysis for the 2000s. *Environmental and Resource Economics, 45*(4), 471-494.

IBGE (Brazil's Institute of Geography and Statistics) (2014). *SIDRA database*. Retrieved from http://www.sidra.ibge.gov.br/

INPE (Brazil's National Institute for Space Research) (2014). *Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satélite*. Retrieved from http://www.obt.inpe.br/prodes/index.php

Izquierdo, A.E., Grau, H.R., & Aide, T.M. (2011). Implications of Rural–Urban Migration for Conservation of the Atlantic Forest and Urban Growth in Misiones, Argentina (1970–2030). *Journal of Human Environment, 40*(3), 298-309.

Jaimes, N.B.P., Sendra, J.B., Delgado, M.G., & Plata, R.F. (2010). Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. *Applied Geography, 30*(4), 576–591.

Kaimowitz, D., Mertens, B., Wunder, S., & Pacheco, P. (2004). *Hamburger Connection Fuels Amazon Destruction: Cattle ranching and deforestation in Brazil's Amazon*. Bogor, Indonesia: Center for International Forestry Research.

Kaimowitz, D., & Angelsen, A. (1998). *Economic Models of Tropical Deforestation: A Review*. Bogor, Indonesia: Center for International Forestry Research.

Laurance, W.F., Albernaz, A.K.M., Schroth, G., Fearnside, P.M., Bergen, S., Venticinque, E.M., … & da Costa, C. (2002). Predictors of deforestation in the Brazilian Amazon. *Journal of Biogeography, 29*, 737-748.

Laurance, W.F., Cochrane, M.A., Bergen, S., Fearnside, P.M., Delamônica, P., Barber, C., ... & Fernandes, T. (2001). The Future of the Brazilian Amazon. *Science, 291*, 438-439.

Macedo, M.N., DeFries, R.S., Morton, D.C., Stickler, C.M., Galford, G.L., & Shimabukuro, Y.E. (2012). Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. *Proceedings of the National Academy of Sciences, 109*(4), 1341-1346.

Margulis, S. (2004). Causes of Deforestation of the Brazilian Amazon. In W. B. w. paper (Ed.), *World Bank Working Paper* (Vol. 22).

MMA (Brazil's Ministry of Environment) (2016). *Cadastro Nacional de Unidades de Conservação*. Retrieved from http://www.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs

Moon, Z.K., & Farmer, F.L. (2012). Deforestation Near Public Lands: An Empirical Examination of Associated Processes. *Society and Natural Resources, 26*(5), 605-621.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*, 17-23.

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Arai, E., del Bon Espirito-Santo, F., … & Morisette, J. (2006). Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. *Proceedings of the National Academy of Sciences, 103*(39), 14637–14641.

Nepstad, D., McGrath, D., Stickler, C., Alencar, A., Azevedo, A., Swette, B., … & Hess, L. (2014). Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *Science, 344*, 1118-1123.

Nepstad, D., Schwartzman, S., Bamberger, B., Santilli, M., Ray, D., Schlesinger, P., … & Rolla, A. (2006). Inhibition of Amazon Deforestation and Fire by Parks and Indigenous Lands. *Conservation Biology, 20*, 65-73.

Nolte, C., Agrawal, A., Silvius, K.M., & Soares-Filho, B.S. (2013). Governance regime and location in-flu-ence avoided deforestation success of protected areas in the Brazilian Amazon. *Proceedings of the Na-tional Academy of Sciences, 110*(13), 4956-4961.

Oliveira, R.C., & Almeida, E. (2011). Deforestation in the Brazilian Amazonia and Spatial Heterogeneity: a Local Environmental Kuznets Curve Approach. In 57th Annual North American Meetings of the Regional Science Association International. Retrieved from http://www.poseconomia.ufv.br/docs/Seminario 08-10-2010ProfEduardo.pdf

Pfaff, A., Robalino, J., Walker, R., Aldrich, S., Caldas, M., Reis, E., … & Kirby, K. (2007). Road Investments, Spatial Spillovers, and Deforestation in the Brazilian Amazon. *Journal of Regional Science, 47*(1), 109–123.

Rametsteiner, E., Obersteiner, M., Kindermann, G., & Sohngen, B. (Eds.) (2009). *Avoided Deforestation: Prospects for Mitigating Climate Change. Economics of avoiding deforestation*. Routledge, London: International Institute for Applied Systems Analysis.

Reis, E.J., & Guzman, R.M. (1993). Um modelo econométrico do desflorestamento da Amazônia. *Pesquisa e Planejamento Economico, 1*(23), 33-64.

Sawyer, D. (2008). Climate change, biofuels and eco-social impacts in the Brazilian Amazon and Cerrado. *Philosophical Transactions of the Royal Society B, 363*, 1747–1752.

Soares-Filho, B.S., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., ... & Alencar, A. (2014). Cracking Brazil's Forest Code. *Science, 344*, 363-364.

Soares-Filho, B.S., Nepstad, D.C., Curran, L.M., Cerqueira, G.C., Garcia, R.A., Ramos, C.A., … & Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature, 440*, 520-523.

SRTM (Shuttle Radar Topography Mission) (2014). *90m Digital Elevation Model, version 2.1*. Retrieved from http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/

TRMM (Tropical Rainfall Measuring Mission) (2016). *Monthly product 3B43*. Retrieved from http://mirador.gsfc.nasa.gov/cgi-bin/mirador/

WDPA (World Database of Protected Areas) (2015). *WDPA dataset*. Retrieved from http://www.wdpa.org/

Witmer, F. (2005). *Simulating Future Global Deforestation Using Geographically Explicit Mode*. *Interim Report IR-05-010*. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Chapter 3

A confirmation of the indirect impact of sugarcane on deforestation in the Amazon

Abstract. A widely discussed assumption that the expansion of sugarcane indirectly contributes to deforestation in Brazil has been backed statistically by only a handful of studies. The present research measures the indirect effect of sugarcane in Brazil's frontier counties as a weighted summation of changes in sugarcane area in agricultural (non-frontier) counties, where weights are constructed using road distances and the bandwidth that minimizes overall model error. In addition to economic variables, indirect effect variables are employed to create a model that explains deforestation. Parameters are estimated following fixed-effects methodology. The results reveal that sugarcane indirectly contributed to deforestation in Brazil during the period from 2002-2012. The effect was estimated to be sizeable; in particular, 16.3 thousand km^2 of forest was cut by economic actors displaced by expanding sugarcane plantations. This figure constitutes 12.2% of deforestation in Brazil from 2002-2012 and is equivalent to 189.4 million Mg of carbon emissions.

Keywords. Deforestation, sugarcane, Brazil, indirect land use change.

3.1. Introduction

Recent trends in the decoupling of soybean and sugarcane expansion with deforestation in the Brazilian Amazon raised hopes that the dual goal of protecting the environment and satisfying an ever increasing demand for crops and derived crop products could be met (see Macedo et al., 2012). Indeed, Brown et al. (2005) and Galford et al. (2010) found that most of the recent increase in crop production is attributable to the slight expansion of already existing fields, conversion of already deforested land, and higher yields. Macedo et al. (2012) estimated that from 2001 to 2005, 74% of soybean expansion in Mato Grosso was into previously cleared pasture areas and that from 2006 to 2010 the figure reached 91%. Rudorff et al. (2010) found that sugarcane expands almost exclusively over pasture land and annual agricultural crops in the state of São Paulo. Some researchers, however, have theorized that cattle pastures are displaced by mechanized farming and, as a result, are reconstituted in distant regions on forest frontiers (see Barretto et al., 2013; Lima et al., 2011; Vera-Diaz et al. 2008; Walker et al., 2009; Walker, 2014). This phenomenon is known as indirect land use change (ILUC). ILUC associated with sugarcane in particular was discussed by Fargione et al. (2008), Lapola et al. (2010), Martinelli & Filoso (2008), Miccolis et al. (2014), Searchinger et al. (2008), and Walter et al. (2014).

Statistical evidence of ILUC associated with soy expansion in Legal Amazon during the first half of the 2000s was found by Arima et al. (2011). This study was later updated by Richards et al. (2014), who confirmed that soy expansion contributed to frontier deforestation during 2002-2011 in Brazil. Gollnow & Lakes (2014) verified ILUC associated with soy alongside the BR-163 highway during 2001-2004, but found no connection between soy expansion in Mato Grosso and deforestation alongside BR-163 during 2005-2012. de Sá et al. (2013) tested ILUC associated with sugarcane by creating an interactive variable using sugarcane area in São Paulo and cattle heads in the Legal Amazon, following the assumption that sugarcane's effect on deforestation is manifested by displacing ranching activities. The interaction terms, together with other explanatory variables, were regressed against deforestation. The study concluded that sugarcane production in São Paulo increased the impact of cattle ranching on land clearing in the Amazon.

The present study offers an alternative way to statistically test for and measure the indirect impact of sugarcane expansion in Brazil's non-frontier regions on deforestation. This indirect effect is measured as a weighted summation of changes in the area dedicated to sugarcane cultivation in non-frontier regions and is included in the model that explains deforestation. This methodology was implemented by Arima et al. (2011) and Richards et al. (2014) for testing ILUC associated with soy. A key challenge associated with this method is choosing a weighting scheme that reflects the relationship between the farmers' willingness to migrate and the distance from the original location of farmers' residence to the location of new pastures. The lack of information about this relationship has potentially led researchers to select sub-optimal weighting schemes, thereby leading to results that may be inaccurate. Arima et al. (2011) assumed a weighting scheme with a constant slope (weights and distances are linearly related), and Richards et al. (2014) used inverse distances (travel time) as weights. To overcome the lack of knowledge on how farmer migration depends on distance, I apply a computationally rigorous way of defining the weighting scheme. Specifically, the weighting function includes a parameter that defines how weights and distances are related. This parameter is estimated by minimizing model error. The results of this study have global importance, since the research tests the indirect impact of the Brazilian sugarcane industry on deforestation over more than a decade.

3.2. Research context

Sugarcane is used to produce sugar and ethanol and generate surplus electricity for the mills (Walter et al., 2014), and the sugarcane industry is rapidly expanding in Brazil. From 2002 to 2012 sugarcane plantations in southern Brazil (defined by non-frontier counties, see section 3.3) expanded by 45.9 thousand km². The biggest share of sugarcane expansion in southern Brazil was recorded in the state of São Paulo (56%), followed by Goiás (12.8%), Minas Gerais (12.3%), and Mato Grosso do Sul (10.4%). Figure 3 illustrates sugarcane expansion in the southern non-frontier regions during 2002-2012. The expansion can be explained by a combination of government policies aimed at the development of sustainable renewable energy, energy security, and rural development (Verdade et al., 2012; Walter et al., 2014).

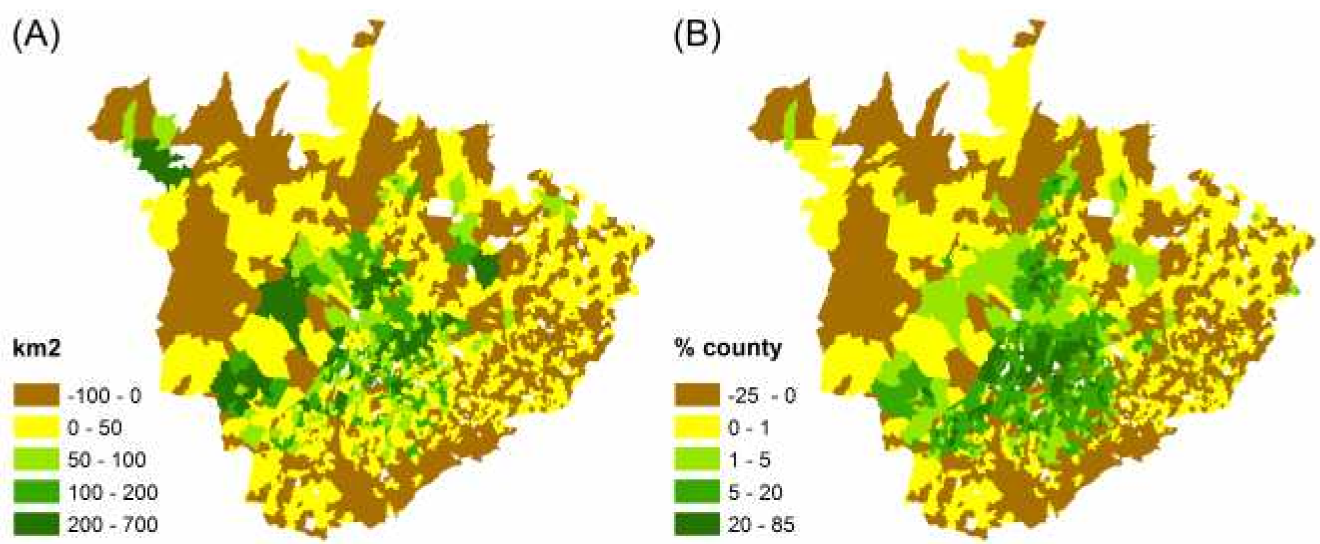

Figure 3. Changes in planted sugarcane area between 2002 and 2012 in agricultural counties (see Figure G1) in terms of square kilometers (left) and percent of county's territory (right). This municipality boundary map was downloaded from the IBGE's website. Municipality boundaries are not shown to improve visualization. Projection: Albers Equal Area Conic.

The production of first-generation ethanol using sugarcane is a conventional technology in Brazil. Local demand of ethanol was boosted after 2003 due to the introduction of flex-fuel cars that use either gasoline or pure ethanol into the market (Alkimim et al., 2015). Brazilian production reached 23.64 billion liters in the 2012/2013 season (Damaso et al., 2014). Ethanol is also the main biofuel used in the world, and its consumption is likely to rise in the future. Its energetic balance is generally positive, meaning that the growing sugarcane absorbs more carbon than is emitted when the ethanol is burned as fuel (Martinelli & Filoso, 2008). Therefore, recently, ethanol production started to be seen as a potential solution to the global warming problem.

A large share of Brazil's mitigation policy with regard to greenhouse gases is predicated on a belief that agronomically suited cattle pastures will be converted to cropland. The total area of pastures suitable for sugarcane production is still greater than the total area covered by sugarcane plantations in Brazil (Cohn et al., 2014). Besides, Morton et al. (2016) noted that few areas remain for the legal expansion of croplands, indicating that the major share of expansion has to be onto cattle pastures. Relative to cattle ranching, crop farming might promote regional economic development; cultivating crops on low-productivity pastures might also help to restore soil fertility, impel neighboring pastures into more intensive cultivation, and spare land from deforestation (Cohn et al., 2016). Moreover, the transition of former large cattle ranches and soybean fields to relatively smaller sugarcane fields in Brazil's non-frontier regions may be more beneficial to the environment, since it could regenerate vegetation (Redo et al., 2013). However, if sugarcane expansion moves cattle farmers to the frontier, forests are cleared to create pastures for displaced cattle and, as a result, a large volume of carbon stock is released. Searchinger et al. (2008) estimated that ethanol from Brazilian sugarcane could pay back the carbon emissions in four years if sugarcane only replaces tropical grazing land or in 45 years if displaced ranchers convert rainforest to grazing land.

Several economic explanations exist to support ILUC theory. Firstly, land is up to 10 times more expensive in southern Brazil than in frontier regions (Nepstad et al., 2006; Sawyer, 2008). As a result, cattle ranchers who own properties suitable for sugarcane production can sell their holdings with enormous capital gains and buy inexpensive land from smallholders in frontier regions. Land availability and insecurity in the Brazilian Amazon is another factor that may explain migration to the frontier. Cheap and abundant land often encourages cattle ranchers to engage in expansive agriculture, thereby leading to substantial deforestation. Land insecurity is a consequence of insufficient legal enforcement. For example, Brazil's Forest Code which requires keeping 80% of private properties in the Amazon as forests is often ignored by landowners, and environmental fines are rarely paid.

ILUC also relies on an assumption that beef demand is inelastic. This assumption is based on the argument that meat products have few substitutes. Converting pastures to crop fields reduces beef supplies, and in turn raises beef prices (Walker, 2011; Walker, 2014). If beef demand is inelastic, beef production must be reconstituted in another location, most likely, on forest frontiers.

3.3. Data and methods

Variable descriptions, units of measurement, and data sources are presented in Table 5. Annual changes in cattle herd size and planted areas of soy and sugarcane were calculated as differences between current and previous year data, as $t-(t-1)$. Data on environmental fines (*autuações ambientais*) on a county level was aggregated from the dataset of detailed records made available by Brazil's Institute of Environment and Natural Resources (IBAMA). In case of fine status changes, the same record appears more than once in the dataset. Therefore, repeating records were trimmed programmatically in *R* software (Version 3.2.2) (each record has a unique process number). The study used agricultural commodity prices recorded in the southern Brazilian state of Paraná; these prices are highly correlated with average local crop prices calculated for a sample of municipalities in Legal Amazon (Assunção et al., 2013). So, this price data only exhibits temporal variation. GDP, rural credit, environmental fines, and commodity prices were deflated to year 2012 using Brazil's consumer price index (*Índice Nacional de Preços ao Consumidor Amplo;* IPCA).

Table 5. Description of the variables

Abbreviation	Description	Unit	Source
DEF	Annual deforestation increments. Data on the municipal level is available on the INPE's website. It is aggregated from PRODES maps, which are distributed at a 60-meter spatial resolution and are created by digital image processing and visual interpretation of LANDSAT ™ imagery on computer screens.	km^2	INPE (2016)
CATTLE	Annual changes in size of cattle (*bovines*) herd	count	IBGE (2016), SIDRA database
SOY	Annual changes in planted area of soy	ha	IBGE (2016), SIDRA database
SUGAR	Annual changes in planted area of sugarcane	ha	IBGE (2016), SIDRA database
CREDIT	Sums of rural credit for agriculture (both cattle farming and crop cultivation) in real terms, issued by official banks and credit cooperatives	1000 R$ (BRL)	Central Bank of Brazil (2016)
GDP	Gross domestic product in real terms	1000 R$ (BRL)	IBGE (2016), SIDRA database
FINES	Sums of issued environmental fines in real terms	1000 R$ (BRL)	IBAMA (2016)
PCATTLE	Real price of fat cattle (*boi gordo*) received by the producers	R$ per arroba[VII]	SEAB/PR (2016)
PSOY	Real price of soy received by the producers	R$ per 60 kg	SEAB/PR (2016)
PSUGAR	Real price of sugarcane received by the producers	R$ per tone	SEAB/PR (2016)

INPE: Brazil's National Institute of Space Research; IBGE: Brazil's Institute of Geography and Statistics; SIDRA: IBGE's system of Automatic Recovery; IBAMA: Brazil's Institute of Environment and Natural Resources; SEAB/PR: Paraná State Agriculture and Supplies Bureau

Brazil is undergoing a very dynamic process of county administrative border adjustments. Also, new municipalities are being created. The study used Brazil's administrative boundary maps from 2000 and 2014, which are available from Brazil's Institute of Geography and Statis-

[VII] Arroba is a weight measure used by the Brazilian farmers. It is equivalent to 15 kilograms of beef carcass or 30 kilograms of live weight (Borras et al., 2011).

tics (IBGE), to calculate the territory of each county in both years. Administrative border adjustments that resulted in at least a 2% change in territory were considered significant and were addressed. Counties that saw their borders changed as well as newly established counties were highlighted on the map (refer to Appendix F for the map of territorial changes). Furthermore, the maps from 2000 and 2014 were overlaid in ArcGIS (Version 10.0, ESRI, Relands, CA, USA) to identify the smallest time-constant geographical units. The data for each geographical unit was obtained by aggregating the data of the counties within each unit. The seat of the largest municipality in a geographical unit was considered to be the seat of the whole unit.

Municipalities in the study area were partitioned into deforestation and agricultural counties. The former are those located within the Amazonian Biome (the shapefile of the Amazonian Biome is made available by IBGE). The latter are the counties of the states of Goiás, Minas Gerais, Mato Grosso (except for those counties that are located within the Amazonian Biome), Mato Grosso do Sul, Paraná, and São Paulo. Those six states host the majority of Brazil's sugarcane plantations. The study period encompasses years 2002-2012; it was constrained by data availability and the methodological approach.

To estimate the indirect effects of crop expansion the study used road distances between municipality seats. The shapefile of roads was obtained from Brazil's Ministry of Environment (MMA, 2016). The coordinates of county seats were obtained from IBGE, and they were used to create a point shapefile. Both shapefiles were loaded into ArcGIS and projected to the same Albers Equal Area Conic projection. Planned roads were removed from the attribute table of the road shapefile. Next, the polylines (road segments) in the road shapefile were split by the points (counties) in the seat shapefile. The ArcGIS Network Analyst extension was employed to locate municipality seats on the road network using a 10km search radius. Counties without a road connection were omitted from the analysis. Roadless locations face zero or little deforestation and have marginally sized cattle herds. Thus, it is safe to assume that displaced deforestation in remote areas is close to zero. Refer to Figure G1 for the illustration of deforestation and agricultural counties that remained in the sample after excluding roadless locations.

The dimensions of the distance matrix are 2105x302 (2105 agricultural counties and 302 deforestation counties). The weights were calculated from road distances using the kernel function presented in eq. 11 (subscripts f and a represent deforestation and agricultural counties, respectively, h is bandwidth).

$$w_{f,a} = \exp\left(\frac{-d_{f,a}}{h}\right) \quad (11)$$

The weighting scheme assumes that the farmers are unwilling to move long distances as high migration costs associated with long distances may encourage the farmers to switch to crop cultivation or move to nearby metropolitan areas in southern Brazil instead of selling their pastures and relocating cattle ranching to Legal Amazon. However, the exact relationship between distances, migration costs, and willingness to migrate are unknown. Therefore, an accurate weighting scheme cannot be determined *a priori*. The weighting in eq. 11 is controlled by a bandwidth, which determines how fast the weights decrease as distance increases. The computation of optimal bandwidth is explained later in this section.

The weighted indirect effect (*WIE*) in county *f* at time *t* is calculated as the weighted sum of cropland area (*c*) in agricultural counties at time *t*, normalized by the sum of weights. The idea of using distances to compute the indirect effects of cropland expansion was originally proposed by Arima et al. (2011).

$$WIE_{f,t} = \frac{\sum_a c_{a,t} w_{f,a}}{\sum_a w_{f,a}} \quad (12)$$

Eq. 12 formally illustrates the concept of ILUC. Crop expansion in all agricultural counties is assumed to heterogeneously affect deforestation in each county located in the Amazonian Biome. This heterogeneity is created by assigning higher weights to nearer agricultural counties and lower weights to distant agricultural counties, thereby introducing cross sectional variation. Weighting is motivated by the assumption that migrant cattle farmers are much more willing to move short distances due to lower migration costs. See Figure H1 for the estimates of the WIE of sugarcane when using eq. 12.

The model in this study explains deforestation as a function of economic variables, listed in Table 5. Additionally, it includes the one-year lag of deforestation, the square of GDP (to account for environmental Kuznets curve theory), the weighted indirect effects of soy and sugarcane, and one-year lags of these two variables due to the possibility that changes in agricultural counties may have a delayed impact on land use changes in the deforestation counties. Even though testing for ILUC associated with soy is not the goal of this study, it is essential that this variable is included in the modeling to avoid omitted variable bias.

The model parameters were estimated by fixed-effects (FE) panel methodology. Under FE, both observed (elevation, soil type, etc.) and unobserved county fixed effects are removed by demeaning. This procedure also eliminates the need to model deforestation determinants that change slowly over time, such as the road network, population, and areas that receive special protection. A programming code that estimates FE regressions was integrated with the code that computes the weighted indirect effects from the dataset (programming was done in *Gauss*, the code in available in section 9.2). This code was used to loop through bandwidths with increments of 1 km and save the overall mean squared errors (MSEs) under each bandwidth. Then, the optimal bandwidth is the one that minimizes overall MSE (model error). The formula to estimate the overall MSE is given by eq. 13. As for notation, $y_{f,t}$ is the deforestation increment in county *f* at time *t*; $X_{f,t}$ is a row vector of covariate values in county *f* at time *t*, and is a column vector of coefficient estimates (all data is considered before demeaning).

$$MSE = \sum_t \sum_f \left(y_{f,t} - X_{f,t} \quad \right)^2 \quad (13)$$

3.4. Results and discussion

The dependency of model error on bandwidth size for the interval [10, 250] km is graphically illustrated in Figure 4A. Model error was minimized at h=99 km. The weighting in eq. 11 with optimal bandwidth leads to a 21.6% reduction in model error as compared to inverse distance weighting. Figure 4B shows the dependence between weight and distance when optimal bandwidth is used.

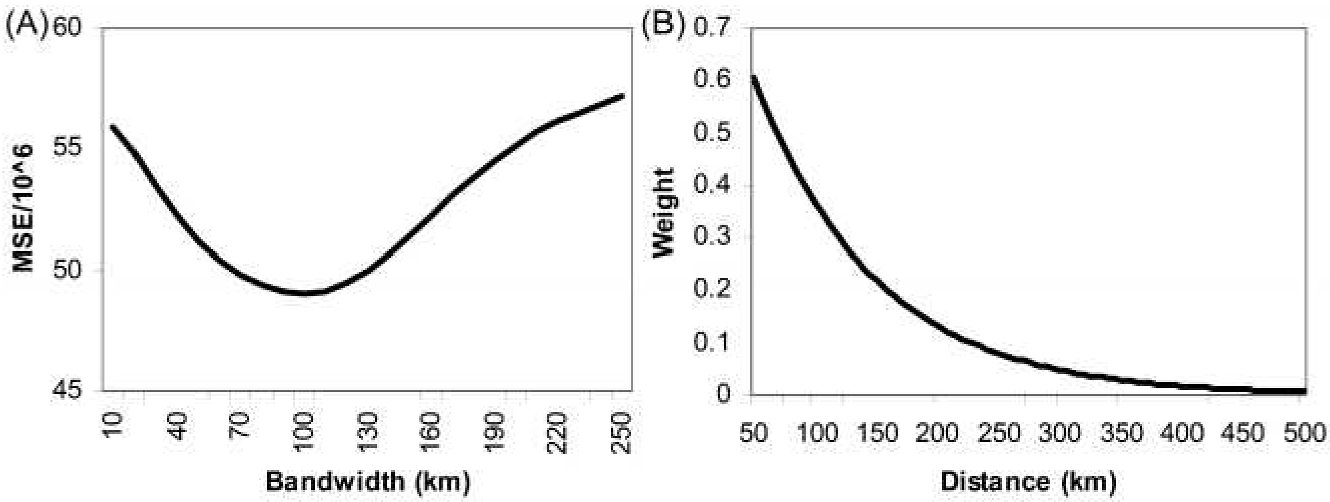

Figure 4. Overall mean squared error for various bandwidths (left) and weighting mechanism with optimal bandwith, by road distance (right)

Regression results are presented in Table 6. Collinearity statistics are available in Table I1, which reveal that the covariates are weakly correlated. The key finding for this research is that sugarcane expansion in agricultural counties contributed to deforestation in the Amazonian Biome (i.e., the coefficients of the weighted indirect effect of sugarcane are statistically significant). From the results, it can be suggested that sugarcane expansion during the study period (2002-2012) indirectly explained 12.2%, or 16.3 thousand km^2 in absolute terms, of deforestation that happened in Legal Amazon during the same time. Given that the average forest carbon density in Brazil is 116 Mg ha^{-1} (Harris et al., 2012), sugarcane expansion in southern Brazil was responsible for 189.4 million Mg of carbon emissions during the study period. Following Goldemberg & Guardabassi (2010), 6617 liters of ethanol can be produced from one hectare of sugarcane per year. Combining this figure with the findings of this study, it can be concluded that the yearly production of every one billion liters of ethanol from sugarcane in Brazil during the 2002-2012 period indirectly released 17.5 million Mg of carbon into the atmosphere.

For comparison, the indirect contribution of soy expansion to deforestation was found to be 17.6%, or 23.6 thousand km^2 in absolute terms. The results suggest that carbon emissions related to soy expansion generated 273.6 million Mg. These figures are much more modest estimates compared to previous studies: Arima et al. (2011) suggested that a 10% reduction of soybean plantations in the Amazonian savannas would have reduced deforestation by around 26 thousand km^2 during the 2003-2008 period (40% of the deforestation that took place during that period), and Richards et al. (2014) found that one third of deforestation during the 2002-2011 period was indirectly caused by soy expansion in the Brazilian regions outside the Amazon. The lower estimates of soy's indirect effect can be partially explained by the spatial extent of this study, which covers only six states, where a major share of Brazil's sugarcane is cultivated. The differences in the magnitude may also be due to methodological differences (weighting scheme, variable selection, etc.).

Table 6. Regression results

Abbreviation	Fixed-effects
DEF (-1)	106.2492 (45.7455)*
CATTLE	0.1447 (0.0654)*
SOY	1.4061 (0.5357)**
SUGAR	-1.5515 (2.2226)
CREDIT	-0.6418 (0.1896)**
GDP	-0.0118 (0.00766)
GDP^2	0.000176 (0.0000948)‡
FINES	-0.6559 (0.2168)**
PCATTLE	-359.9817 (118.0482)**
PSOY	677.6511 (142.065)**
PSUGAR	-957.1251 (155.7607)**
WIESOY	1.014 (0.1571)**
WIESUGAR	24.6213 (5.9131)**
WIESOY (-1)	0.9343 (0.2121)**
WIESUGAR (-1)	14.026 (7.9228)‡
# obs	3322
# groups	302
Sigma u	79.31
Sigma e	58.19

** p<0.01 * p<0.05 ‡ p<0.1
Periodization: 2002-2012
Parameters are multiplied by 1000
Cluster-robust standard errors are in parentheses
Variable abbreviations are introduced in Table 5
GDP^2 is GDP squared, divided by 10^6
WIE stands for weighted indirect effect
(-1) indicates one-year lag

The findings only indicate that sugarcane indirectly contributed to deforestation during the period from 2002 to 2012. However, it is unclear whether this tendency will continue into the future. If it does, two types of policy measures could be applied to mitigate this problem: 1) policies that target cattle ranching, because the indirect effect of sugarcane on deforestation is manifested through displaced cattle farming, and 2) policies that directly target sugarcane and its derivative products.

Popular policy agreements to control cattle-driven deforestation are known as MPF-TAC (Terms of Adjustment of Conduct) and G4. The former was signed in 2009 by individual meatpacking companies in order to stop purchases from properties where illegal deforestation took place. The latter is a zero-deforestation agreement with Greenpeace signed in 2009 by Brazil's largest meatpacking companies. Both agreements currently govern only those properties selling directly to slaughterhouses (Gibbs et al., 2015), thereby leaving access to commercialization channels for calving or breeding ranches that illegally expand their properties at the expense of forests. Nevertheless, the two agreements are effective in curbing deforestation associated with cattle farming (Gibbs et al., 2015) and create disincentives for cattle ranchers located in agricultural counties to migrate.

A way to limit pressure on forest frontiers from the cattle industry is through intensification. The Brazilian cattle stocking rate grew from 0.47 head/ha in 1960 to 1.2 head/ha in 2010 (McManus et al., 2016). Cohn et al. (2014) suggested two strategies for how to promote intensification further: a tax on cattle from conventional pastures and a subsidy for cattle from semi-intensive pastures. A tax would raise agricultural commodity prices and would lead to higher productivity agriculture. However, higher prices may also increase production. A subsidy would increase the output due to higher productivity systems as well as reduce the prices of agricultural commodities. However, lower prices would increase consumption. A study conducted by Cohn et al. (2014) found that both cattle ranching intensification policies in Brazil can sucessfully limit deforestation and carbon emissions.

Alternatively, the adverse effect of sugarcane expansion on forests can be mitigated through the price of biofuels. The price could be adjusted to compensate for indirect environmental costs, especially, the loss of carbon stock. However, raising the price of biofuels may not be desirable, because it reduces the attractiveness of ethanol biofuel as compared to conventional fossil fuels.

Other results presented in Table 6 are of secondary importance for the purpose of this study, but have great importance overall. Most coefficients have the expected signs. The results indicate that: 1) deforestation last year is a valid predictor of deforestation this year, 2) changes in cattle herd size and acreage of soy plantations directly contribute to deforestation, 3) changes in sugarcane area does not directly affect forest clearing, 4) higher GDP is related to more deforestation, and 5) environmental fines constrain forest clearing.

Two findings, however, defy prior expectations. Firstly, rural credit is negatively linked with deforestation (with a statistically significant coefficient). Rural credit constraints are one of the policy measures applied by the Brazilian government to curb deforestation. Therefore, a positive link between the two variables should exist. However, Jusys (2016) showed that rural credit is positively linked with deforestation (with statistically significant local coefficients) in Pará when the spatial heterogeneity of deforestation determinants is considered, even though the sign is negative in the conventional global model. The other unexpected result is that beef and sugarcane prices are negatively associated with deforestation. The fact that only temporal variation in prices was captured by the model could have led to this result.

3.5. Conclusions

This study tested the assumption that the sugarcane industry indirectly contributes to deforestation in the Brazilian Amazon. The indirect effect was measured as a weighted summation of annual changes in sugarcane area, divided by the sum of weights. The methodological novelty proposed by this study is to select the weighting scheme by estimating the bandwidth (a parameter that controls the relationship between weights and distances) that leads to the lowest model error. A model that explains deforestation as a function of economic variables, including indirect effects, was built, and the parameters were estimated by fixed-effects regression.

The research concludes that sugarcane expansion in Brazil's non-frontier regions during 2002-2012 is indirectly associated with deforestation. Specifically, the findings suggest that the sugarcane industry in southern Brazil is indirectly responsible for 16.3 thousand km^2 of cleared land, which corresponds to 12.2% of total deforestation during the study period. Those figures translate into 189.4 million Mg of carbon emissions. Also, a yearly production of every one billion liters of ethanol from sugarcane resulted in 7.5 million Mg carbon emissions. The research also confirms ILUC associated with soy expansion. The estimates suggest that 23.6 thousand km² (corresponding to 17.6% of deforestation) was attributable to the soy industry in agricultural counties during the study period. Therefore, policies for mitigating the effects of ILUC must be integral. That is, both soy and sugarcane expansion must be considered.

3.6. References

Alkimim, A., Sparovek, G., & Clarke, K.C. (2015). Converting Brazil's pastures to cropland: An alternative way to meet sugarcane demand and to spare forestlands. *Applied Geography, 62*, 75-84.

Assunção, J., Gandour, C., Rocha, R., & Rocha, R. (2013). *Does Credit Affect Deforestation? Evidence from a Rural Credit Policy in the Brazilian Amazon. Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Barretto, A.G., Berndes, G., Sparovek, G., & Wirsenius, S. (2013). Agricultural intensification in Brazil and its effects on land use patterns: An analysis of the 1975-2006 period. *Global Change Biology, 19*, 1804-1815.

Borras, S.M., McMichael, P., & Scoones, I. (2010). The politics of biofuels, land and agrarian change: editors' introduction. *The Journal of Peasant Studies, 37*, 575-592.

Brown, J.C., Koeppe, M., Coles, B., & Price, K.P. (2005). Soybean Production and Conversion of Tropical Forest in the Brazilian Amazon: The Case of Vilhena, Rondônia. *Journal of the Human Environment, 34*, 462-469.

Central Bank of Brazil (2016). *Anuário estatístico do crédito rural (Até 2012)*. Retrieved from https://www.bcb.gov.br/?RELRURAL

Cohn, A.S., Gil, J., Berger, T., Pellegrina, H., & Toledo, C. (2016). Patterns and processes of pasture to crop conversion in Brazil: Evidence from Mato Grosso State. *Land Use Policy, 55*, 108-120.

Cohn, A.S., Mosnier, A., Havlík, P., Valin, H., Herrero, M., Schmid, E., ... & Obersteiner, M. (2014). Cattle ranching intensification in Brazil can reduce global greenhouse gas emissions by sparing land from deforestation. *Proceedings of the National Academy of Sciences, 111*(20), 7236-7241.

Damaso, M.C.T., Machado, C.M.M., Rodrigues, D.D.S., Belem, S.G., & Salum, T.F.C. (2014). Bioprocesses for biofuels: an overview of the Brazilian case. *Chemical and Biological Technologies in Agriculture, 1*(6), 8.

de Sá, S.A., Palmer, C., & di Falco, S. (2013). Dynamics of Indirect Land-Use Change: Empirical Evidence from Brazil. *Journal of Environmental Economics and Management, 65*, 377-393.

Fargione, J., Hill, J., Tilman, D., Polasky, S., & Hawthorne, P. (2008). Land clearing and the biofuel carbon debt. *Science, 319*, 1235–1238.

Galford, G.L., Melillo, J., Mustard, J.F., Cerri, C.E.P., & Cerri, C.C. (2010). The Amazon Frontier of Land-Use Change: Croplands and Consequences for Greenhouse Gas Emissions. *Earth Interactions, 14*(15), 1-24.

Gibbs, H.K., Munger, J., L'Roe, J., Barreto, P., Pereira, R., Christie, M., … & Walker, N.F. (2015). Did Ranchers and Slaughterhouses Respond to Zero-Deforestation Agreements in the Brazilian Amazon? *Conservation letters*. doi: 10.1111/conl.12175.

Goldemberg, J., & Guardabassi, P. (2010). The potential for first-generation ethanol production from sugarcane. *Biofuels, Bioproducts and Biorefining, 4*(1), 17-24.

Gollnow, F., & Lakes, T. (2014). Policy change, land use, and agriculture: The case of soy production and cattle ranching in Brazil, 2001-2012. *Applied Geography, 55*, 203-211.

Jusys, T. (2016). Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon. *Applied Geography, 75*, 188-199.

Harris, N.L., Brown, S., Hagen, S.C., Saatchi, S.S., Petrova, S., Salas, W., … & Lotsch, A. (2012). Baseline Map of Carbon Emissions from Deforestation in Tropical Regions. *Science, 336*, 1573-1576.

IBAMA (Brazil's Institute of Environment and Natural Resources) (2016). *Consulta de Autuações Ambientais e Embargos*. Retrieved from https://servicos.ibama.gov.br/ctf/publico/areasembargadas/ConsultaPublicaAreasEmbargadas.php

IBGE (Brazil's Institute of Geography and Statistics) (2016). *SIDRA database*. Retrieved from http://www.sidra.ibge.gov.br/

INPE (Brazil's National Institute of Space Research) (2016). *Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satellite*. Retrieved from http://www.obt.inpe.br/prodes/index.php

Lapola, D.M., Schaldach, R., Alcamo, J., Bondeau, A., Koch, J., Koelking, C., … & Priess, J.A. (2010). Indirect land-use changes can overcome carbon savings from biofuels in Brazil. *Proceedings of the National Academy of Science, 107*, 3388-3393.

Lima, M., Skutsch, M., & Costa, G.D.M. (2011). Deforestation and the Social Impacts of Soy for Biodiesel: Perspectives of Farmers in the South Brazilian Amazon. *Ecology and Society, 16*(4), article 4.

Macedo, M.N., DeFries, R.S., Morton, D.C., Stickler, C.M., Galford, G.L., & Shimabukuro, Y.E. (2012). Decoupling of deforestation and soy production in the southern Amazon during the late 2000s. *Proceedings of the National Academy of Sciences, 109*, 1341-1346.

Martinelli, L.A., & Filoso, S. (2008). Expansion of sugarcane ethanol production in Brazil: environmental and social challenges. *Ecological Applications, 18*, 885–898.

McManus, C., Barcellos, J.O.J., Formenton, B.K., Hermuche, P.M., de Carvalho Jr., O.A., Guimarães, R., ... & Neto, J.B. (2016). Dynamics of Cattle Production in Brazil. *PloS one, 11*(1), e0147138.

Miccolis, A., de Andrade, R.M.T., & Pacheco, P. (2014). Land-use trends and environmental governance policies in Brazil: Paths forward for sustainability. *CIFOR Working Paper* (no. 171). Bogor, Indonesia: CIFOR.

MMA (Brazil's Ministry of Environment) (2016). *Download de dados geográficos*. Retrieved from http://mapas.mma.gov.br/i3geo/datadownload.htm#

Morton, D.C., Noojipady, P., Macedo, M.N., Gibbs, H., Victoria, D.C., & Bolfe, E.L. (2016). Reevaluating suitability estimates based on dynamics of cropland expansion in the Brazilian Amazon. *Global Environmental Change, 37*, 92-101.

Nepstad, D., Stickler, C.M., & Almeida, O.T. (2006). Globalization of the Amazon soy and beef industries: opportunities for conservation. *Conservation Biology, 20*, 1595–1603.

Redo, D., Aide, T.M., & Clark, M.L. (2013). Vegetation change in Brazil's dryland ecoregions and the relationship to crop production and environmental factors: Cerrado, Caatinga, and Mato Grosso, 2001–2009. *Journal of Land Use Science, 8*(2), 123-153.

Richards, P., Walker, R., & Arima, E. (2014). Spatially Complex Land Change: The Indirect Effect of Brazil's Agricultural Sector on Land Use in Amazonia. *Global Environmental Change, 29*, 1-9.

Rudorff, B.F.T., Aguiar, D.A., Silva, W.F., Sugawara, L.M., Adami, M., & Moreira, M.A. (2010). Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. *Remote Sensing, 2*, 1057-1076.

Sawyer, D. (2008). Climate change, biofuels and eco-social impacts in the Brazilian Amazon and Cerrado. *Philosophical Transactions of the Royal Society B, 363*, 1747–1752.

SEAB-PR (Paraná State Agriculture and Supplies Bureau) (2016). *Preços*. Retrieved from http://www.agricultura.pr.gov.br/

Searchinger, T., Heimlich, R., Houghton, R.A., Dong, F., Elobeid, A., Fabiosa, J., … & Yu, T.H. (2008). Use of U.S. Croplands for Biofuels Increases Greenhouse Gases Through Emissions from Land-Use Change. *Science, 319*, 1238-1240.

Vera-Diaz, M.D.C., Kaufmann, R., Nepstad, D., & Schlesinger, P. (2008). An interdisciplinary model of soybean yield in the Amazon basin: The climatic, edaphic, and economic determinants. *Ecological Economics, 65*, 420–431.

Verdade, L.M., Gheler-Costa, C., Penteado, M., & Dotta, G. (2012). The Impacts of Sugarcane Expansion on Wildlife in the State of São Paulo, Brazil. *Journal of Sustainable Bioenergy Systems, 2*, 138-144.

Walker, R. (2011). The Impact of Brazilian Biofuel Production on Amazonia. *Annals of the Association of American Geographers, 101*(4), 929-938.

Walker, R. (2014). Saving Land for Nature in the Brazilian Amazon: Implications from Location Rent Theory. *Geographical Analysis, 46*(1), 18-36.

Walker, R., DeFries, R., Vera-Diaz, M.D.C., Shimabukuro, Y., & Venturieri, A. (2009). The Expansion of Intensive Agriculture and Ranching in Brazilian Amazonia. *Geophysical Monograph Series, 186*, 61-81.

Walter, A., Galdos, M.V., Scarpare, F.V., Leal, M.R.L.V., Seabra, J.E.A., Cunha, M.P., … & Oliveira, C.O.F. (2014). Brazilian sugarcane ethanol: developments so far and challenges for the future. *Energy and Environment, 3*, 70-92.

Chapter 4

Associations between deforestation and population on forest frontiers in Pará: empirical study

Abstract. The presence of connection between deforestation and population is subjected to academic debates. While in general it is agreed that the trends shifted towards urban-induced deforestation, it is theorized that rural communities directly contribute to deforestation on forest frontiers. Recent release of population estimates at high spatial resolution enables to zero in on frontier forests. This study covers the state of Pará in Brazil, where deforestation represent a serious concern. Study period is 2010-2014. The analysis is done at 5 km spatial resolution. The study interprets the results of fractional logistic regression and the regression tree. The results reveal that higher rural population density is associated with more deforestation on forest frontiers in Pará. Weak links between distance covariates reflecting the access to urban markets and deforestation likely imply the presence of significant subsistence component of agriculture, further strengthening the conclusion that rural inhabitants directly contribute to deforestation.

Keywords. Deforestation, rural population, Pará, forest frontier, high spatial resolution.

4.1. Introduction

Historically, growing rural communities and rural insurgencies prompted Brazilian authorities to support smallholder colonization of rainforest regions (Rudel et al., 2009). Road construction provided easy access to forest resources and, as a result, led to massive deforestation (Pan & Carr, 2010). In particular, rainforest colonization stretched alongside newly constructed Trans-Amazonian highway. Slash and burn agriculture is believed to have caused two-thirds of past tropical deforestation (Wright & Muller-Landau, 2006). However, private agricultural enterprises gradually increased their production for international markets (Rudel et al., 2009). Whether rural communities still directly induce a significant share of deforestation or the trends shifted towards deforestation driven by urban and international demands is an ongoing debate. In any case, the relationship between rural population, urban population and deforestation is likely to be complex (Brook et al., 2006).

Some scientists argue that rural communities no longer constitute a primary threat to forests. Rudel et al. (2009) stress that well-capitalized ranchers, farmers, and loggers producing for consumers in distant markets became more prominent and that this globalization weakened historically strong relationship between local population growth and forest cover. Browder et al. (2008) conducted a study in Rondônia (Brazil) that revealed a dramatic reduction in the portion of production consumed on the farm. More specifically, in 1991 the shadow priced value of farm production consumed by households was equivalent to 2843 Brazilian reals, dropping to 825 Brazilian reals in 2001. These findings suggest that farmers are shifting from subsistence to commercial agriculture. DeFries et al. (2010) argue that urbanization raises consumption levels and increases demand for agricultural products. They also stress that urban citizens consume more processed food than rural inhabitants. If urban populations constitute a major driver of deforestation, this is a growing concern, since urban growth is vastly outpacing rural growth (DeFries et al., 2010). Furthermore, Carr (2004) notes that small-scale frontier farmers have many incentives to sell their properties to cattle ranchers, who continuously clear forests to supply their cattle with land. Those incentives arise, because small producers do not have easy access to markets and credit, and face other economic and political constraints (Carr, 2004, Izquierdo et al., 2011). This can be an argument in favor of urban-induced deforestation, since cattle ranchers mostly produce for urban and international markets. Even though cattle ranching requires solid investment, the direct return on cattle ranching itself (excluding profits from the sale of timber) consistently exceeds 10% (Margulis, 2004).

The advocates of rural-induced deforestation emphasize that land cover change remains extraordinarily expansive per capita on forest edges. There are several reasons for that. Firstly, it is the rational investment in land expansion given cheap available family labor, scarce capital, low technology, high cost of transportation, and insecure land tenure (Carr & Burgdorfer, 2013, Carr, 2009). The latter characteristic implies that new forest frontier settlers are encouraged to clear the land as quickly as possible to establish their claim and deter potential squatters (Carr, 2004). Further, it is important to discuss what happens with second generation frontier farm children. Barbieri et al. (2009) point out that out-migration tends to dominate second-generation settler household demographic dynamics. This out-migration is directed towards urban areas, since second generation of original settlers sees the urban environment and off-farm employment as increasingly attractive livelihood options (VanWey et al., 2012). This economic migration serves as a hedge against lost harvest or an injured family member and helps to raise income, diversify risk and overcome credit constraints (Barbieri & Carr, 2005, Barbieri et al., 2009, Davis & Carr, 2014, VanWey et al., 2012). However, many second generation settlers remain in original lands, and this leads to land fragmentation. Côrtes et al. (2013) conducted a study in Santarém (Pará, Brazil) to reveal that 44 out of 311 properties in 2003 were fragmented and that those 44 properties in 2003 became 113 in 2009. The portion of second generation settlers who migrate to another location on forest frontier is small. Despite this fact, Carr (2009), Carr (2012) and Carr & Burgdorfer (2013) argue that this small portion has a disproportionally large adverse effect on extant forests. Another argument in favor of expansive agriculture on forest edges is that intensification represents an unnecessary labor burden, is uneconomical, inefficient and too risky for small semi-subsistence producers (Carr, 2004). Moreover, farmers are encouraged to expand their farms due to economic reasons, as those farms are later sold to larger producers (Izquierdo et al., 2011). Afterwards, the farmers move to another location to initiate a new round of deforestation (Brondizio et al., 2013).

The lack of agreement on whether and how rural and urban inhabitants spur deforestation in 21st century partly arises from the scarcity of detailed local studies that specifically investigate the relationship between population dynamics and deforestation (Izquierdo et al., 2011). Among such studies is Wright & Muller-Landau (2006), who find that proportion of forest cover remaining is closely correlated with human population density. Another study is by Jorgenson & Burns (2007), who conclude that deforestation is positively linked with rural population growth and negatively correlated with urban population growth. On the contrary, an empirical study by DeFries et al. (2010) positively linked urban citizens with deforestation and found no correlation between the size of rural communities and deforestation. The mentioned studies are country level analyses. Carr & Burgdorfer (2013) criticized DeFries et al. (2010) by arguing that if rural population growth data is used at the national scale, the demographic situation at the local level on the external frontier may be inappropriately reflected. To investigate what happens on forest frontier, it is necessary to use sufficiently small scale. In general, a positive correlation between population growth and deforestation is found at temporal and spatial macro-scales, but evidence for associations between population and deforestation at micro-scales is scant (Carr, 2004). An early attempt to model deforestation and rural population in Legal Amazon at relatively high spatial resolution is found in Laurance et al. (2002). The best population data available at that time was from census 2000, carried out by IBGE. As a result, within municipality variations could not be observed. The lack of population data at high spatial resolutions could explain the paucity of studies at micro-scales. In 2015 ~100x100 m grid population raster was released by WorldPop. This data constitutes an opportunity to empirically test the arguments made in theoretical studies. WorldPop data has already been used by Ryan et al. (2015) in modeling relationships between population and forest loss in Africa.

Most importantly, Pan & Carr (2010) conclude that the results of land use studies can be scale-dependent: inferences and relationships observed at one scale or level are not necessarily the same for larger or smaller scales. This was confirmed by Ryan et al. (2015), who modeled forest cover change in selected territory in Africa and found that different drivers of forest cover loss emerge at local and national scales. Interestingly, the findings suggest that rural communities are associated with more deforestation at national scale, but with less deforestation at micro scale. Relationships between deforestation and rural population in Legal Amazon are researched by many authors, but mostly at national or district level. Besides, few empirical analyses concentrate on forest edges. Therefore, the contribution of this research is a micro-scale (5x5 kilometer grid) investigation into the linkages between deforestation and the size of rural communities on forest edges in Pará state in Brazil. The study empirically tests the main argument made by Carr & Burgdorfer (2013) that rural inhabitants are directly linked with deforestation on forest frontiers.

It is important to mention, however, that the results of this study cannot be generalized to all locations with tropical rainforests both inside and outside Legal Amazon, because site-specific variations do exist. Such variations stem from differences in market access and history, infrastructure, social support or natural resource bases, etc. (Oestreicher et al., 2014). For instance, there is evidence that forest attrition in Ecuador is positively linked with population on the frontier, but negatively linked with overall rural population (de Sherbinin et al., 2007). Although the findings can be different across areas, the priority is always to understand what happens in deforestation hotspots, such as Pará.

Linkages between deforestation and rural population were evaluated by two distinct methods: fractional logistic regression and regression tree. The regression is non-linear, since deforestation was measured as a fraction of deforested land. It is crucial to understand that the findings do not prove causality and can only be interpreted in terms of associations. This is because complete understanding of demography and environment cannot privilege a single causal direction (Vanwey et al., 2012). That is, not only population influences deforestation, but deforestation can also initiate changes in population. For instance, a new deforestation frontier may attract settlers from other areas. To account for such complex relationship, valid instrumental variables are needed. However, such data is not available at high spatial resolution. Nevertheless, associations between deforestation and rural population do answer the question whether or not deforestation happens within rural settlements.

4.2. Materials and methods

The analysis includes eleven covariates to explain deforestation. Those consist of cattle (coded as CTL later in the text), rural population (POP), altitude (ALT), slope (SLOPE), surface flatness (FLAT), forest cover (FC), the shortest travel time to the nearest city (TIME), the shortest Euclidean distance to the nearest road, river and protected area (ROAD, RIVER and PA), and precipitation (PREC). The study period covers 5 years, from 2010 to 2014 inclusive.

Deforestation was measured as a fraction of cleared land in total land during 2010-2014 inclusive. The data source is Brazil's National Institute of Space Research (INPE, 2015). Grid level estimates of cattle heads were based on municipality level statistics from Brazil's Institute of Geography and Statistics (IBGE, 2015). The number of cattle animals (2009-2014 average) was distributed evenly over available land within each municipality. Available land was defined as total area net of forests. Forest cover estimates (2009-2014 average) were obtained from

MODIS VCF product, version 051 (NASA, 2014). Population data comes from the WorldPop Americas dataset (WorldPop, 2015). It is reported at ~100 m spatial resolution at the equator. Each cell contains an estimated number of persons. The data can be used both to support applications for planning interventions, measuring progress, and to predict response variables intrinsically dependent on the population distribution (Sorichetta et al., 2015). For detailed explanations how the population values were estimated the reader is directed to Sorichetta et al. (2015). The original figures were adjusted to match municipality totals. This was achieved by estimating correction factors for each municipality (see Appendix J). The final figures, used in the research, are the averages of adjusted population estimates corresponding to years 2010 and 2015, expressed in persons per square kilometer. It is obvious that population data accuracy decreases as spatial resolution increases. Therefore, a too high spatial resolution may invalidate the results. As a result, the research was done at 5 km spatial resolution. In this way each 5x5 kilometer raster cell is defined by a sum of estimated persons from 2500 original cells. Altitude (in meters) comes from the Shuttle Radar Topography Mission 90m Digital Elevation Model, version 2.1 (SRTM, 2015). One of its derivatives, slope, was computed in ArcGIS. It was estimated from 8 adjacent cells as described in Burrough & McDonell (1998). The resulting figures are in degrees. Surface flatness is the sum of altitude differences between each 90x90 meter parcel and the lowest altitude within each aggregated (5x5 km) parcel of land. Thus, lower figures indicate flatter surfaces. Forest cover is a percentage of forests within each parcel of land in 2010 (at the beginning of the study period). The data is found in MODIS VCF product. Time distances are estimated as travel time in minutes to the nearest city of 50000 or more inhabitants in 2000. The data was retrieved from the Global Environment Monitoring Unit - Joint Research Centre of the European Commission (GEMU-JRC, 2014). I acknowledge that some cities could have grown and that new roads were constructed after 2000, but this is the best data of this kind available at high spatial resolution. Euclidean distances to the nearest road (excluding urban streets and roads under construction), river and protected areas were computed in ArcGIS. Road and river shapefiles were obtained from GEOFABRIK, OpenStreetMap (GEOFABRIK, 2014). Shapefiles of protected areas are from the World Database on Protected Areas (WDPA, 2014). The distance data was expressed in kilometers. Precipitation was measured as a multiyear average of rainfall in millimeters. The data source is World-Clim (World-Clim, 2015). A shapefile of forest frontiers is provided by the World Resource Institute (WRI, 2015). Frontier forests are defined as being primarily forested, of sufficient size to support viable populations of the full range of indigenous species associated with that particular forest ecosystem given periodic natural disturbance episodes, and exhibiting a structure and composition shaped largely by natural events, as well as by limited human disturbance from traditional activities (WRI, 2015). Administrative borders of Pará and its municipalities were obtained from IBGE. All GIS datasets were projected to Albers Conic Equal Area projection, which is an appropriate projection for Brazil (IBGE, 2015). Refer to Appendix K for descriptive statistics of the variables.

Only populated cells (≥1 person per km^2) with relatively dense forest cover (≥40%) were retained. Additionally, the cells contaminated by severe cloud coverage (≥20%) were removed. The WordPop dataset includes both rural and urban populations. The distinction between rural and urban settlements is usually based on population density. Since forest frontiers are sparsely populated, all communities residing near forest edges are rural. In fact, the highest population density in the sample is less than 75 persons per km² (see Appendix K). National census authorities in most advanced countries use 400 persons per km² as a cut-off mark.

To estimate regressions, it is essential to consider that the response variable (deforestation) is a fraction. Early works (as well as some recent studies, see DeFries et al. (2010) for example) applied the arc-sine transformation on the dependent variable. However, it does not perform well at the extreme ends of a distribution and does not confine the data between 0 and 1 (Hardy, 2002). To account for fractional response, fractional logistic regression was applied. However, heteroskedasticity is likely to be present, since the variance of the dependent variable conditioned on its covariates is unlikely to be constant when response values are bounded between 0 and 1 (Papke & Wooldridge, 1996). As a result, robust standard errors were reported. The computations were done using *Stata's fracreg* command.

Additionally, a regression tree was created for complementary analysis. This was done in *R* software using the library *tree*. The *tree* routine is based on classification and regression tree (CART) algorithm. However, high correlation between any of the covariates may invalidate the results (Sutton, 2005). Therefore, the covariates were tested for multicollinearity by computing variance inflation factors. No significant collinearity was detected. Next, the data was randomly divided into training and testing sets, each containing half of the data.

The working principle of CART is to select the split that leads to the greatest reduction in the sum of the squared differences between the response values for the training sample cases corresponding to a particular node and their sample mean (Sutton, 2005). Formally, it minimizes the following expression:

$$\sum_{i:x_i \in A} \left(y_i - \overline{y_A} \right)^2 + \sum_{i:x_i \in B} \left(y_i - \overline{y_B} \right)^2 \quad (14)$$

Here x_i is a vector of covariate values of i^{th} observation, y_i is response value of i^{th} observation, y_A *bar* and y_B *bar* are the training sample means of the response values corresponding to nodes *A* and *B* respectively. The partitioning is recursive. That is, each child node, in turn, becomes a parent node, unless it is a terminal node (Moisen, 2008). The number of terminal nodes (*leaves*) defines the size of a tree. Finding the optimal size requires two steps. Firstly, an overly large tree is grown. However, some stopping rule must be applied. Otherwise, the tree grows until the number of terminal nodes equals the number of observations in the training sample, which leads to a training prediction error equal to zero (the predicted response value in each terminal node is the actual response value). Undoubtedly, such a tree would not be applicable outside the training sample. Therefore, the tree continued to be grown until the child nodes contained at least 10 observations or a pair of child nodes contained at least 20 observations. The motive to grow an oversized tree instead of looking for the optimal size by sequentially increasing the tree size is that some early trivial split may lead to an important split in the subsequent step. The second step involves tree pruning. Pruning is based on the procedure called cross-validation. Specifically, mean squared errors (*MSEs*) on the testing sample are recorded for different sizes of a tree, ranging from 1 to the size of non-pruned tree. The optimal size is the size that minimizes MSE on the testing sample. MSE is computed as follows:

$$MSE = \frac{1}{n}\sum_{m=1}^{l} SS_m \quad (15), \text{ where } SS_m = \sum_{i\in m}\left(y_i - \overline{y_m}\right)^2 \quad (16)$$

Here n represents the testing sample size and SS_m is the within-node sum of squared differences from the mean in terminal node m, and l is the number of terminal nodes. The predicted response value in each terminal node is simply the mean of within-node response values. For further reading on decision trees refer to Ripley (2008).

4.3. Results

The main research interest is in the relationship between deforestation and rural population. The link could be both positive and negative. A negative association could imply that most deforestation occurs outside rural settlements. This, in turn, could indicate that commercial agriculture prevails, and that the demand of agricultural products and the pressure on standing forests is dictated by urban and international consumers. Alternatively, a negative link could arise due to the fact that small producers often sell their land properties to cattle ranchers, whose farms are generally larger than the farms dedicated to semi-subsistence food production. As a result, low population density is associated with more deforestation (Carr, 2004). Since cattle ranchers mostly produce for urban and international markets, it would be suggestive that deforestation is more responsive to urban meat consumption levels rather than subsistence farming. A positive association could indicate that rural farmers actively cut forests to open up land for agriculture.

Forest frontier deforestation is concentrated in two major quadrants (Figure 5). One is located in the western Pará. Here most deforestation happens alongside the Trans-Amazonian highway and adjacent to (and even within) protected areas. The other quadrant is on the eastern side, where deforestation stretches mostly alongside the borders of protected areas (see Appendix L for locations of the Trans-Amazonian highway and the network of protected areas). Spatial comparison of the arrangement of deforestation hotspots and rural settlements on forest frontiers tends to suggest that positive link between the two variables may exist.

Regression results (see Table 7) indicate that the size of rural settlements is positively linked with deforestation (refer to Appendix M for marginal effects, OLS results are reported for comparison). The link is statistically significant. The regression tree (Figure 6), which offers a complementary analysis, also suggests a positive association between deforestation and rural population. Specifically, the results indicate that in areas with 130-847 cattle heads per parcel and relatively sparse vegetation (<58.7%) deforestation is 4.6% if rural population density is less than 2.9 persons per km² and 7.6% otherwise. Also, in parcels with more than 847 cattle animals and favorable levels of precipitation (<2099 mm) deforestation is 6.8% if rural population density is lower than 1.8 persons per km² and 11% otherwise. Therefore, the findings empirically confirm the theoretical argument made by David Carr that forest frontier settlers directly contribute to deforestation. However, its importance is not as overwhelming as suggested by David Carr. The coefficients of a logistic regression do not reveal the relative importance of each determinant, but a regression tree provides such information. The first two splits are based on cattle heads, meaning that cattle variable is the most powerful discriminator between areas with relatively low and relatively high deforestation (0.7% in parcels with less than 130 cattle heads, 4.3% in parcels with 130-847 cattle animals and 8.5% in the remaining parcels). Population appears only in the fourth split.

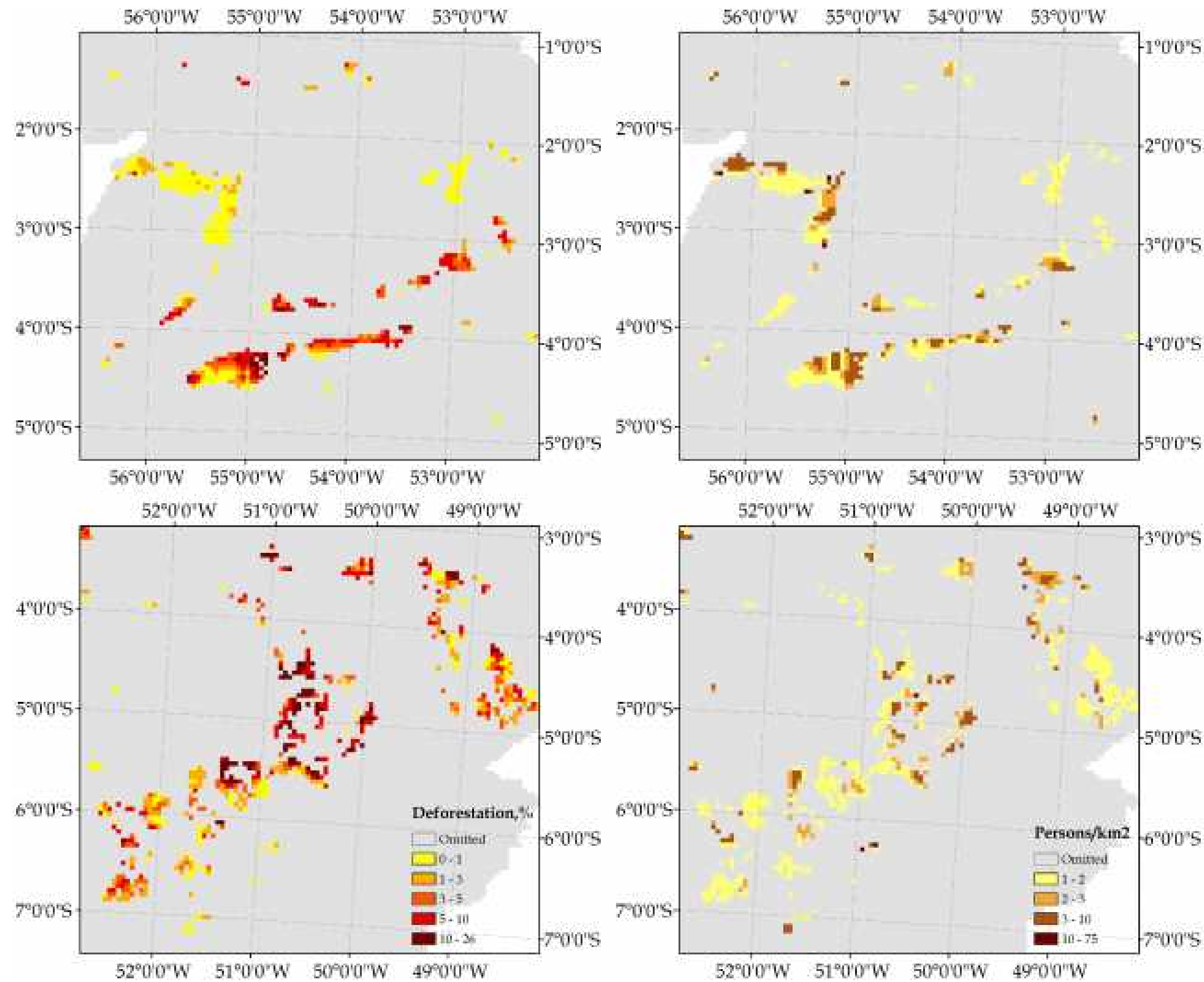

Figure 5. Spatial distribution of deforestation and population on forest frontiers in Pará during 2010-2014. Light grey areas indicate out-of-sample territories.

The next question that naturally rises is how deforestation is related to urban population. Unfortunately, to assess the impact of urban population at high spatial resolution is difficult, since deforestation and urban consumption of agricultural products happen at different and often distant locations is space. The influence of urban citizens on extant forests may be best proxied by the distance to the nearest road or the distance to the nearest urban area. If agriculture is urban-orientated, farmers are likely to be situated in close proximities to fast-access roads or cities, because it reduces transportation costs of agricultural products. Indeed, Rudel et al. (2009) point out that landowners intensified agriculture close to newly constructed or improved roads. The evidence on linkages between deforestation and distance variables in this study is contradictory. The regression suggests that both time and road distance covariates are negatively linked with deforestation. The latter association is found to be statistically significant. These findings are contradicted by the results of the regression tree, which indicates a positive link between road distance and deforestation. However, the link is weak, as it appears in the last split, meaning that many conditions have to be met for road covariate to have an effect. Weak relationships are not so surprising owing to the fact that forest frontiers are such regions, which are characterized by limited accessibility. In any case, the relationship between distance covariates and deforestation should not be taken as evidence in favor of or against urban-induced deforestation. This is primarily because choices of farming locations by cash-orientated producers also depend on other factors, such as legal enforcement, soil quality, climatic conditions and land price (lands are often purchased from smaller producers and then consolidated).

Table 7. Regression results

	Coefficients and standard errors	
	OLS	**FRACLOG**
Cons	97.4716***	-1224.635***
	(14.8377)	(350.906)‡
CTL	0.0412***	0.9232***
	(0.00531)	(0.1582)‡
POP	0.7781**	18.9974**
	(0.3921)	(9.2255)‡
ALT	-0.00831	0.1951
	(0.017)	(0.4617)‡
SLOPE	-9.3328	-327.6202**
	(6.2091)	(140.4977)‡
FLAT	-0.0189	-0.3791
	(0.018)	(0.4265)‡
FC	-0.6271***	-20.7258***
	(0.14)	(3.7508)‡
TIME	0.00025	-0.0495
	(0.00121)	(0.0452)‡
ROAD	-0.0479	-2.6065*
	(0.0512)	(1.4755) ‡
RIVER	0.139**	3.7232**
	(0.058)	(1.5331)‡
PREC	-0.0191***	-0.62***
	(0.00564)	(0.1525)‡
PA	-0.0717	-2.5044
	(0.0587)	(1.5633)‡
# obs	1172	1172
R squared	0.274	-

Standard errors are in parentheses
Symbol '‡' indicates robust standard errors
*** p<0.01 ** p<0.05 * p<0.1
Parameters are multiplied by 1000

Terrain variables do not seem to influence deforestation on forests frontiers in Pará. Terrain is much more of a constraint for crop cultivation. Crop fields constitute a relatively minor agricultural use in the study area. Here the overwhelmingly predominant type of land use is pastures, which is not so responsive to surface characteristics. Interestingly, forest cover is negatively associated with deforestation. This finding possibly indicates that most farmers expand the existing fields until the forests adjacent to those fields are cleared, while denser forests remain relatively untouched until the existing pastures become unviable or are sold to large producers. It is well-known that excessive precipitation constrains agriculture. This fact is reflected in the results. The fractional logistic regression suggests a negative and statistically significant association between deforestation and the level of rainfall. The regression tree indicates that the threshold value of rainfall is 2099 millimeters. Conditional on relatively large cattle herds, deforestation is 9.5% in dryer areas compared with 4.2% in areas with excessive rainfall. Another important finding is that deforestation in protected areas is almost three times lower than in unprotected lands (1.4% versus 4%). This finding holds for land parcels with middle-sized cattle herds and dense forests. Note that cut-off mark in the regression tree is 2.5 km. Since distances were calculated from the centers of each cell, the actual cut-off mark is zero, that is, the borders of protected areas. Despite this finding, it cannot be concluded that lower deforestation is attributable to effective protection, since climatic, economic and other characteristics inside and outside protected areas often are very different. It is a common knowledge that conservation units are located in areas that face lower deforestation pressure.

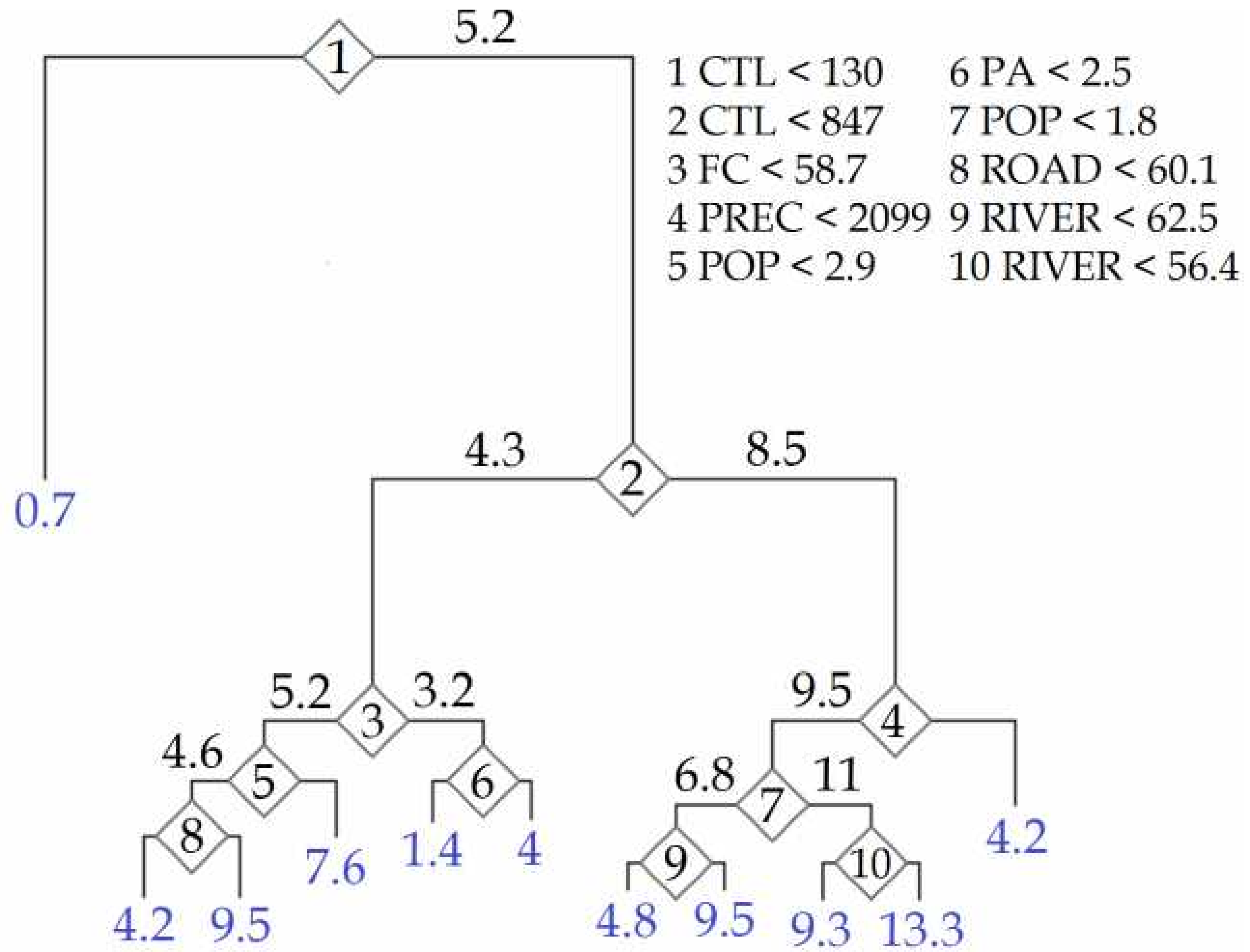

Figure 6. Pruned regression tree. The numbers in rhombuses correspond to the numbers in the legend. In case a condition in the legend is correct, the results to interpret are on the left branch. The figures in the tree are within-node means of deforestation. Figures at the bottom of branches are deforestation means in the terminal nodes. Refer to Appendix N for detailed statistics.

4.4. Discussion and conclusions

The findings have two important implications. Firstly, a policy that fosters rural-to-urban migration should lead to a reduction in deforestation pressure on forest frontiers in Pará. This is line with forest transition theory. As predicted by the theory, industrialization, urbanization and abandonment of rural settlements should lead to forest regeneration. Nevertheless, Pará is a source of large volume of agricultural products, especially beef, which satisfy the demands of ample international markets. Unless consumption trends change, cattle ranching will remain a leading primary business in Pará. Therefore, rural-to-urban migration may only redistribute the relative contribution to deforestation by rural and urban settlements. That is, the gain in avoided deforestation attributable to shrinking rural communities due to rural-to-urban migration may be lost, beacuse those migrants may increase pressure on forests indirectly through meat consumption. Nonetheless, urban-induced deforestation is easier to control. This is primarily because commercial production must be purchased and transported. Currently, a package of measures is being applied by the Brazilian government and NGOs to mitigate the adverse impact of commercial agriculture on forests. Those include, but are not limited to, credit constraints, embargoes and voluntary agreements by the retailers and traders not to purchase agricultural production from illegally deforested lands. Additionally, trucks can be inspected along the roads. Unlike commercial agriculture, subsistence farming and its implications on deforestation are difficult to observe and control, because the products are not purchased by the companies and are consumed on site. Arguably, the best measure against rural-driven deforestation is the network of areas under legal protection. Its importance will grow in the future as frontiers approach the borders of conservation units. Therefore, the second important policy implication is to ensure effective buffer zone management. On its effectiveness depends whether or not small-scale farmers will invade protected territories, thereby starting another round of deforestation, or they will be forced to consider alternative options.

The key message from the research is that rural communities are directly linked with deforestation on forest frontiers in Pará, as theorized by David Carr. The regression tree suggests that the link is evident in areas with middle-sized cattle herds and sparse vegetation, and in areas with relatively large cattle herds and favorable precipitation level for agriculture. Other important findings are: 1) cattle heads is by far the most powerful discriminator between regions with high and low deforestation, 2) distance covariates are weakly linked with deforestation, 3) terrain characteristics have no significant influence on deforestation, 4) territories with less dense forest cover are more likely to be deforested if those territories host middle-sized cattle herds, 5) the level of rainfall is an important factor in farmers' decision where to cut forests if cattle farming is prevalent, 6) deforestation in conservation units is significantly lower than in unprotected forest frontiers in areas with dense vegetation and middle-sized cattle herds.

4.5. References

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

Barbieri, A.F., Carr, D., & Bilsborrow, R.E. (2009). Migration Within the Frontier: The Second Generation Colonization in the Ecuadorian Amazon. *Population Research and Policy Review, 28*(3), 291-320.

Brondizio, E.S., Cak, A., Caldas, M.M., Mena, C., Bilsborrow, R., Futemma, C.T., … & Batistella, M. (2013). Small Farmers and Deforestation in Amazonia. *International Journal or Remote Sensing, 30*(10), 2547-2577.

Brook, B.W., Bradshaw, C.J.A., Koh, L.P., & Sodhi, N.S. (2006). Momentum Drives the Crash: Mass Extinction in the Tropics. *Biotropica, 38*(3), 302-305.

Browder, J.O., Pedlowski, M.A., Walker, R., Wynne, R.H., Summers, P.M., Abad, A., … & Mil-Homens, J. (2008). Revisiting Theories of Frontier Expansion in the Brazilian Amazon: A Survey of the Colonist Farming Population in Rondônia's Post-frontier, 1992–2002. *World Development, 36*(8), 1469-1492.

Burrough, P.A., & McDonell, R.A. (1998). *Principles of Geographical Information Systems*. Oxford, United States: Oxford University Press.

Carr, D. (2004). Proximate Population Factors and Deforestation in Tropical Agricultural Frontiers. *Population and Environment, 25*(6), 585-612.

Carr, D. (2009). Rural migration: The driving force behind tropical deforestation on the settlement frontier. *Progress in Human Geography, 33*(3), 355-378.

Carr, D. (2012). Agro-ecological drivers of rural out-migration to the Maya Biosphere Reserve, Guatemala. *Environmental Research Letters, 7*(4), 7.

Carr, D., & Burgdorfer, J. (2013). Deforestation Drivers: Population, Migration, and Tropical Land Use. *Environment, 55*(1).

Côrtes, J.C., Suter, L.K., D'Antona, A.O., & Carr, D. (2013). Population mobility and land fragmentation: land use-cover change in Brazil and Guatemala. In Proceedings of the International Union for the Scientific Study of Population (IUSSP). Retrieved from http://iussp.org/sites/default/files/event_call _for_papers/cortes_etal_IUSSP2012_0.pdf

Davis, J., & Carr, D. (2014). Migration, remittances and smallholder decision-making: implications for land use and livelihood change in Central America. *Land Use Policy, 36,* 319-329.

DeFries, R.S., Rudel, T., Uriarte, M., & Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience, 3,* 178-181.

de Sherbinin, A., Carr, D., Cassels, S., & Jiang, L. (2007). Population and Environment. *Annual Review of Environment and Resources, 32,* 345-373.

GEMU-JRC (Global Environment Monitoring Unit - Joint Research Centre of the European Commission) (2014). *Travel time to major cities: A global map of Accessibility*. Retrieved from http://bioval.jrc.ec.europa.eu/products/gam/index.htm

GEOFABRIK (2014). *OpenStreetMap Data Extracts*. Retrieved from http://download.geofabrik.de/

Hardy, I.C.W. (2002). *Sex ratios: Concepts and research methods.* Cambridge, U.K.: Cambridge University Press.

IBGE (Brazil's Institute of Geography and Statistics) (2015). *Bases and references, cartographic bases, digital maps*. Retrieved from http://mapas.ibge.gov.br/

INPE (Brazil's National Institute for Space Research) (2015). *Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satélite*. Retrieved from http://www.obt.inpe.br/prodes/index.php

Izquierdo, A.E., Grau, H.R., & Aide, T.M. (2011). Implications of Rural–Urban Migration for Conservation of the Atlantic Forest and Urban Growth in Misiones, Argentina (1970–2030). *Journal of Human Environment, 40*(3), 298-309.

Jorgenson, A.K., & Burns, T.J. (2007). Effects of Rural and Urban Population Dynamics and National Development on Deforestation in Less-Developed Countries, 1990–2000. *Sociological Inquiry, 77*(3), 460-482.

Laurance, W.F., Albernaz, A.K.M., Schroth, G., Fearnside, P.M., Bergen, S., Venticinque, E.M., … & da Costa, C. (2002). Predictors of deforestation in the Brazilian Amazon. *Journal of Biogeography, 29,* 737-748.

Margulis, S. (2004). Causes of Deforestation of the Brazilian Amazon. In W. B. w. paper (Eds.)*, World Bank Working Paper* (Vol. 22).

Moisen, G.G. (2008). Classification and regression trees. In: Ecological Informatics (Eds S. E. Jorgensen & B. D. Fath), Vol. 1, pp. 582–588. Amsterdam, the Netherlands: Elsevier Science.

NASA (National Aeronautics and Space Administration) (2014). *MODIS data*. Retrieved from http://reverb.echo.nasa.gov/reverb

Oestreicher, J.S., Farella, N., Paquet, S., Davidson, R., Lucotte, M., Mertens, F., … & Saint-Charles, J. (2014). Livelihood activities and land-use at a riparian frontier of the Brazilian Amazon: quantitative characterization and qualitative insights into the influence of knowledge, values, and beliefs. *Human Ecology, 42*(4), 521-540.

Pan, W.K., & Carr, D. (2010). Population, Multi-scale Processes, and Land Use Transitions in the Amazon. Proceedings of the European Population Conference. Retrieved from http://geog.ucsb.edu/~carr/PDFs_added_ Oct_31/pop_multiscale_processes.pdf

Papke, L.E., & Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics, 11,* 619-632.

Ripley, B.D. (2008). *Pattern Recognition and Neural Networks.* Cambridge, U.K.: Cambridge University Press.

Rudel, T.K., DeFries, R.S., Asner, G.P., & Laurance, W.F. (2009). Changing Drivers of Deforestation and New Opportunities for Conservation. *Conservation Biology, 23*(6), 1396-1405.

Ryan, S.J., Palace, M., Hartter, J., Diem, J.E., Chapman, C.A., & Southworth, J. (2015). Population pressure and global markets drove a decade of deforestation in Africa's Albertine Rift. *Quantitative Biology,* 1-21.

Sorichetta, A., Hornby, G.M., Stevens, F.R., Gaughan, A.E., Linard, C., & Tatem, A.J. (2015). High-resolution gridded population datasets for Latin America and the Caribbean in 2010, 2015 and 2020. *Scientific data, 2.* doi: 10.1038/sdata.2015.45

SRTM (Shuttle Radar Topography Mission) (2015). *90m Digital Elevation Model, version 2.1*. Retrieved from http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/

Sutton, C.D. (2005). Classification and Regression Trees, Bagging, and Boosting. *Handbook of Statistics, 24,* 303-329.

VanWey, L.K., Guedes, G.R., & D'Antona, Á.O. (2012). Out-migration and land-use change in agricultural frontiers: insights from Altamira settlement project. *Population and Environment, 34*(1), 44-68.

WDPA (World Database of Protected Areas) (2014). *WDPA dataset*. Retrieved from http://www.wdpa.org/

World-Clim (2015). *Data for current conditions (~1950-2000)*. Retrieved from http://www.worldclim.org/current

WorldPop (2015). *WorldPop Americas dataset*. Retrieved from http://www.worldpop.org.uk/

WRI (World Resources Institute) (2015). *Global Forest Watch Frontier Forests*. Retrieved from www.globalforestwatch.org

Wright, S.J., & Muller-Landau, H.C. (2006). The Future of Tropical Forest Species. *Biotropica, 38*(3), 287-301.

Chapter 5

Quantifying avoided deforestation in Pará: protected areas, buffer zones and edge effects[VIII]

Abstract. Percentage of forests saved due to the establishment of protected areas is an important piece of information for government institutions and, therefore, is the goal of this study. However, non random location selection bias makes such information directly unobservable. To overcome this problem, propensity score matching was applied. Unlike in previous studies, implications of buffer zone management were assessed by estimating avoided deforestation in buffer zones and park edges. The study area is the state of Pará. Overall results achieved satisfactory mean absolute bias of 7.8% and revealed that park protection saved 0.72% of protected surface from deforestation during period 2000-2004 (~2900 km² of forests). The highest percentage of avoided deforestation was recorded in protected areas, situated near deforestation hotspots: central part of eastern Pará, alongside Trans-Amazonian highway and on the banks of Amazon River. The findings also suggest that buffer zones tend to reduce deforestation where deforestation pressure is lower, but the substitution effect takes over in areas of high deforestation pressure (since loggers are prevented from deforestation within conservation units, deforestation in surrounding areas increases). Finally, the study does not find evidence for edge effects in the state of Pará.

Keywords. Avoided deforestation, protected areas, buffer zones, edge effects, Pará, propensity score matching.

[VIII] This article has been published. Publication details are: Jusys, T. (2016). Quantifying avoided deforestation in Pará: Protected areas, buffer zones and edge effects. *Journal for Nature Conservation, 33*, 10-17.

5.1. Introduction

The protected area system is among popular measures to combat deforestation and save genetic resources in selected territories. Historically, the aim of protected areas was the protection of ecosystems of significant esthetic and cultural value. Later, the primary goal became the protection of ecosystems with threatened species and/or with commercial stocks in decline (ISA, 2015). The process of identification of areas for conservation began with Radam Project in the seventies. However, the criteria were based on a singular geological and geomorphologic phenomenon. Shortly afterwards, a new proposal arose suggesting to prioritize areas with high concentration of endemism (ISA, 2015). In 1990 project Workshop 90 was adopted. Under this project, the criteria for the selection of areas for conservation considered bio-geographical analyses of endemism and richness of species, taking into account the occurrence of rare or threatened species, the presence of special geological phenomenon, and the degree of vulnerability of ecosystems (ISA, 2015). However, due to technical difficulties in assessing the richness of biodiversity, the focus switched to the distribution of ecosystems and landscapes. Also, the possibility of the area being defensible and protected was taken into consideration. Key event was the seminar held in September of 1999, which had the objective to define priority areas and actions for conservation. As a result, a new map of priority areas of Legal Amazon was created, which guides the establishment of protected areas (ISA, 2015).

The network of protected areas in Legal Amazon is ever being expanded. As of 2010, more than 680 thousand km^2 in Pará were declared as protected areas, constituting around 55% of its territory.

Quantifying the percentage of forests saved due to the establishment of protected areas is of key importance for governments aiming at reducing the intensity of deforestation. This information can be used in assessing whether the establishment of a particular protected area is a financially sound decision, in deciding which protected areas require better management and surveillance, or in understanding how park characteristics influence the percentage of avoided deforestation.

However, this seemingly simple task is hampered by the fact that locations for protected areas are not selected randomly. For this reason land characteristics inside and outside protected areas are not similar, thus making direct comparisons between the two types of territories misleading. Since usually protected areas are located in lower deforestation pressure areas, the area of forests saved is exaggerated if non random location selection bias is ignored (Joppa & Pfaff, 2011).

Recent academic literature uses matching methods to overcome non-randomness of location selection. Among studies, which compute avoided deforestation by matching methods are Joppa & Pfaff (2011), Andam et al. (2008), Nelson & Chomitz (2009) and Nolte et al. (2013).

However, those studies do not consider avoided deforestation in or due to buffer zones, thus providing an incomplete picture of protection of conservation units. The Amazon Region Protected Areas Program (ARPA) was a four-year (2000-2003) project that described the protection of conservation units in Brazil. ARPA revealed that the Government of Brazil contributed to the creation of the protected area system by supplying US$ 18 million as direct support to conservation units and US$ 6.5 for buffer zone management. In addition, the Pilot Program for the Brazilian Rainforests (PPG7) supplied US$ 10 million as direct support to conservation units and US$ 16.4 million for buffer zone management. The fact that more than 20 million US dollars were allocated to buffer zone management and that the sum is almost as large as for direct conservation unit protection illustrates the importance of buffer zone management.

The ARPA documentation describes buffer zones as being of benefit to local populations by allowing low environmental impact activities. The participation of local communities is increased by promoting the sustainable use of natural resources in buffer zones through income generating activities such as tourism, sustainable use of genetic resources, and environmental services. In this way local communities are expected to understand the benefits of protected areas and become involved in their protection.

Thus, two seemingly contradictory goals of buffer areas exist: save the environment and allow exploitation to benefit local inhabitants at the same time. Indeed, Martino (2001) raises the question *do buffer zones serve as extensions of national parks or integrate parks and people*. Some authors assume the former (Martino, 2001). However, this is a misconception. If the goal of a buffer zone is to save forests to the same extent as within a conservation unit, buffer zones as such would not be necessary, since a protected area itself could be extended. In 2000, a law governing the protected area system in Brazil was issued (law № 9985, Conservation Units National System, SNUC). Under article 2 of respective law, buffer zone is defined as the environment around a protected area, where human activities are subject to specific norms and restrictions in order to minimize negative impacts on that protected area. Also, where only those human activities are allowed that do not cause damage to the nuclear area (article 41, paragraph 1). Therefore, a buffer zone is meant to protect its conservation unit from deforestation rather than itself.

It is important to understand that buffer zones are multifunctional. Among their functions is to reduce edge effects (Martino, 2001). If no buffer zones are established, park edges would be exposed to potential deforestation or, at least, deforestation could reach to the very edges of protected areas. Logged fields near forests are dry and prone to fire (Cochrane & Laurance, 2002), thus posing fire risk to park edges. Some other functions of buffer zones involve protection from gold mining, drug cultivation, poaching, and maintaining viable population of species. However, those aspects are outside the scope of this research.

The research design is illustrated in Figure 7. Three treatment groups were identified: internal protected area (core), external protected area (edge) and buffer zone. Edge effects were tested by comparing avoided deforestation on park's edge and in its core. Thus, the presence of excess deforestation on park edges can be detected. Avoided deforestation in buffer zones can be both positive and negative. Positive result would indicate that forest resources in a buffer zone are used in a sustainable way that helps to avoid deforestation relative to not having any buffer zone. However, in areas with high deforestation pressure, often characterized by land intensive agricultural businesses and relatively little forests remaining, the creation of a protected area could increase deforestation in the surrounding areas of that protected area (substitution effect). In this case, avoided deforestation is negative.

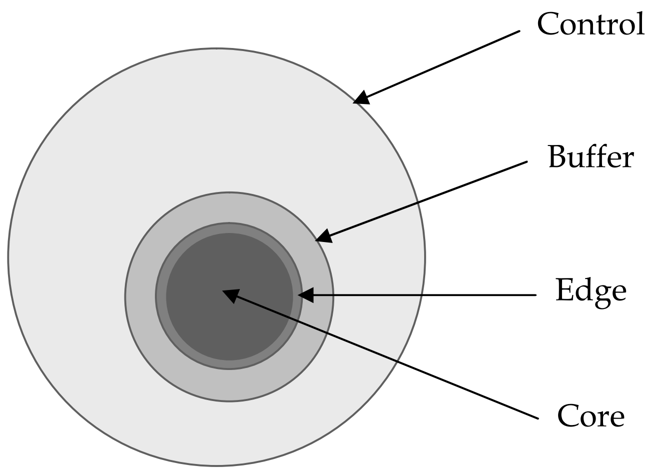

Figure 7. Layers of park protection

The goal of this research is to quantify the percentages of avoided deforestation for each protected area (including on its edge and in its surrounding areas), located in the state of Pará, during period 2000-2004 inclusive. Short study period is motivated by the fact that in 2005 and 2006 the majority of remaining territories of Pará were declared as protected areas, thus leading to an insufficient number of control samples to perform propensity score matching. The contribution of this research lies in analyzing what repercussions have the existence of buffer zones on avoided deforestation and whether edge effects are present. However, the results shall not be used for interpretations of how effective buffer zone management is for two reasons. Firstly, some very well managed buffer zones may not avoid deforestation simply because a protected area it surrounds is located far from deforestation hotspots. Secondly, even if avoided deforestation in a buffer zone is significantly lower than in a corresponding protected area, the benefits of a buffer zone may lie in protecting the edge of that protected area or protecting park's recourses and habitat from poaching and illegal mining. However, the results shall be used in understanding how much deforestation was avoided depending on parks' characteristics and their location.

Buffer zone analysis is relatively rare in academic literature (Martino, 2001). Alexandre et al. (2010) focus on determining optimal widths of buffer zones in Brazil. However, elaborate studies that relate buffer zones and avoided deforestation are missing. This research fills in this gap of knowledge.

5.2. Materials and methods

For this research the data on eleven variables was collected. The outcome variable is deforestation. The ten covariates used in building a propensity score include altitude, slope, surface flatness, agricultural suitability, number of fires, percentage of forest cover, the nearest distances to major city, major road and river, and precipitation.

Deforestation estimates were based on deforestation maps, provided by Brazil's National Institute of Space Research under PRODES project (INPE, 2014a). However, some deforestation was observed later than it actually occurred due to cloud coverage. Therefore, it is likely that some part of deforestation included in the analysis happened before 2000. In those cases the data was still classified as deforestation during 2000-2004 year period. Likewise, certain parcels were cleared during the study period, but could not be identified due to cloud coverage. To address the problem of cloud coverage, deforestation map that includes next year deforestation was downloaded. In case areas marked as clouds in the original deforestation map were classified as forests in the latter map, clouds were replaced by forests in the original map. The move is justifiable, since it is highly likely that areas, which are currently forested, were forests last year. Since most cloud coverage was observed over intact, impenetrable forests, this procedure substantially reduced the number of land parcels, marked as clouds in the original deforestation map. Given that observations are cloud free, each area represented by a pixel is classified as cleared, forested, non-forested or hydrographic. The figures are available at 120 meter spatial resolution. The data was aggregated by forming 8x8 squares. Deforestation was computed as a percentage of deforested 120x120 meter parcels in each square kilometer during 2000-2004. Finally, the data was resampled to 1 km spatial resolution using methods implemented in ArcGIS. See Figure O1 for spatial patterns of all variables.

The data on altitude was retrieved from the Shuttle Radar Topography Mission (SRTM) 90m Digital Elevation Model (DEM), version 2.1 (SRTM, 2014). It is measured as an average altitude in meters over 90x90 meter parcels in each square kilometer. Slope was computed from 8 adjacent parcels, as described in Burrough & McDonell (1998). The figures are in degrees. Surface flatness is the sum of height differences between each 90x90 meter parcel and the lowest altitude in square kilometer under consideration. Agricultural suitability in this research is a measure of climate, soil and terrain constraints for agriculture in 2002, where 1 refers to no constraints and 7 – unsuitability for agriculture. The variable is a coarsened continuous variable with values ranging from 0 to 100. The data was downloaded from the International Institute for Applied System Analysis, Global Agro-Ecological Zones dataset, Plate 28 (IIASA, 2014). The shapefiles containing coordinates of fires were obtained from Brazil's National Institute of Space Research (INPE, 2014b). The data was converted to raster format in ArcGIS, where each pixel contains the number of fire events during the study period. The data on forest cover comes from MODIS Vegetation Continuous Fields (MOD44B, version 5) product (NASA, 2014). The figures are five-year averages of forest cover percentages. Time distances are travel time in minutes to the nearest city of 50000 or more inhabitants in 2000. The data was retrieved from the Global Environment Monitoring Unit - Joint Research Centre of the European Commission (GEMU-JRC, 2014). Euclidean distances to the nearest river (in meters) were computed in ArcGIS. The computations were based on the river shapefile, downloaded from GEOFABRIK, OpenStreetMap (GEOFABRIK, 2014). The road shapefile was obtained from the same source. Road distances are Euclidean distances in meters to the nearest highway, primary, secondary, or tertiary road. I acknowledge that new roads were built in the last ten years. However, this shapefile is the best information that could be accessed. Precipitation data in millimeters was obtained from World-Clim (World-Clim, 2015). The figures are multiyear averages of precipitation. Since the study period encompasses five years, those figures should reflect actual precipitation relatively well.

Territories covered by clouds were removed from the analysis. Also, deforestation data was not available for a square-shaped area in the northeastern Pará due to satellite technical problems. As a result, this area was not considered. Additionally, all observations with sparse coverage of forests in 2000 (less than 20%) were discarded. This automatically eliminated hydrographic areas, depleted lands, urban areas, savannas, etc. Finally, out of the remaining observations 20% were chosen randomly for matching.

The shapefile of administrative borders of Pará was downloaded from GADM database of Global Administrative Areas, version 2.0 (GADM, 2014). The shapefiles of protected areas were obtained from the World Database of Protected Areas (WDPA, 2014). Additionally, missing information on attributes of protected areas was filled in using the National Cadastre of Protected Areas, provided by Brazil's Ministry of Environment (MMA, 2014).

A protected area was excluded from the analysis if: 1) it was established in 2005 or later, 2) has an area of less than 200 km² or contains less than 200 suitable observations for matching within its boundaries, 3) is embedded in another protected area and hence has no buffer zone as such, and 4) is classified either as an environmental protection area or as a particular patrimonial reserve (except the protected area of Arquipélago do Marajó, which is covered by dense forests and is interesting due to its size). The last criterion follows from SNUC (article 25), which states that environmental protection areas and particular patrimonial reserves are not required to be surrounded by buffer zones.

Thus, the analysis concerns 36 protected areas in Pará (see Appendix P for geographical location and park characteristics). However, buffer zones and edges were not constructed separately for each protected area, because some parks share borders. Instead, buffers and edges were formed on a basis of geographical units. The justification for this approach lies in the fact that assigning a buffer for each protected area that borders other protected areas is complicated, since the management of a buffer zone near one park may be affected by the management of a buffer zone near another park in the same geographical unit. Moreover, building a buffer around a group of protected areas increases the number of observations that fall within that buffer and raises the chance that reliable matching results are achieved. Therefore, 17 buffer zones and 17 edges were formed.

Finally, the widths of both buffer zones and park edges have to be selected. SNUC does not specify the exact width of buffer zones, the law only states that the limits of buffer zones may be defined either under the establishment of protected area or afterwards (article 25, paragraph 2). However, the resolution by Brazil's National Council of Environment (CONAMA) (№ 013, article 2) states that any activity within a 10 kilometer distance from conservation units that may affect the biota must be licensed by the competent environmental authority. Therefore, the chosen width of a buffer zone is 10 kilometers. As for park edges, too narrow width may lead to the lack of observations for reliable matching, too wide width may conceal (average out) potential edge effects. One of the main threats to park edges is fires. Cochrane & Laurance (2002) studied edge effects caused by fires in the state of Pará and found that fires may penetrate up to 2.4 kilometers inside forests destroying 40% of all standing stems. Therefore, a selected width of a park edge in this study is 3 kilometers.

The most similar pairs of observations from inside and outside of each park's territory were identified by propensity score matching (the computations were done using *psmatch2* command in *Stata 11.0*). The idea of the method is to compare a deforestation estimate of each land parcel inside a protected area (the treatment sample) with a deforestation estimate in the most similar land parcel from the control sample. To obtain a single score from ten variables a simple logistic regression, as implemented in *Stata 11.0*, was applied. Propensity scores were constructed by computing probabilities of success for each observation (latent dependent variable was assigned value 1 for observations in the treatment group). Formally:

$$ps_i = \frac{\exp\left(x_i^T \quad\right)}{1+\exp\left(x_i^T \quad\right)} \quad (17)$$

Here PS_i is the propensity score of i^{th} observation, x_i represents values of all ten covariates of i^{th} observation, β is the vector of maximum likelihood estimates of coefficients on covariates, and letter T denotes transpose. Next, each observation in the treatment sample is matched with one observation from the control sample in a way that absolute difference of propensity scores is minimal among all possible pairs in the control sample. This method is known as the nearest neighbor matching and is described formally as follows (Becker & Ichino, 2002):

$$C\left(i^0\right) = \min_{i^1} \left\| ps_{i^0} - ps_{i^1} \right\| \quad (18)$$

$C(i^0)$ is the set of control units matched to the treated unit i^1 with an estimated value of the propensity score of matched control unit (Becker & Ichino, 2002). Zero difference between propensity scores of a cell from the treatment sample and its nearest neighbor from the control sample would imply perfect matching given observed characteristics. In many cases the minimal difference proved to be close to zero. However, to enhance the quality of avoided deforestation estimates, caliper of 0.01 was imposed. That is, observations from the treatment group were only considered paired if the absolute difference between propensity scores did not exceed 0.01. This procedure ensures that poor matches are removed from the analysis. If multiple observations from the control sample with exactly the same propensity score were identified to be the closest matches, all those observations were considered as matched pairs of a corresponding treatment observation. A control estimate of deforestation under such cases is an average of deforestation estimates over all matched pairs. Also, matching was with replacement – the same control cell could have been paired with more than one treatment cell.

Average treatment effects on the treated (ATTs) are calculated as differences between deforestation values of matched control and corresponding treatment observations. The resulting figures were averaged to obtain results on protected area level. Formally:

$$ATT = E\left(Y_i^1 - Y_i^{0m}\right) \quad (19)$$

Here Y_i^1 denotes deforestation value of observation i in the treatment sample and Y_i^{0m} stands for deforestation value of matched pair of observation i.

Further, the quality (balance) of matching has to be assessed. One way is to perform *t-test* comparisons. However, this procedure is highly controversial due to its sensitivity to sample size. That is, under large number of observations the test tends to suggest that data is imbalanced even if differences in the means are negligible. Therefore, an alternative measure, standardized differences, was selected. Following Rosenbaum & Rubin (1985), standardized percentage bias of covariate j is estimated as follows:

$$bias_j = \left(\frac{\overline{x_j^1} - \overline{x_j^{0m}}}{\sqrt{0.5\left(\mathrm{var}\left(x_j^1\right) + \mathrm{var}\left(x_j^0\right)\right)}} \right) \times 100 \quad (20)$$

Here $\overline{x_j^1}$ and $\overline{x_j^{0m}}$ are average values of covariate j in the treatment and matched control samples respectively, $\mathrm{var}\left(x_j^1\right)$ and $\mathrm{var}\left(x_j^0\right)$ are variances of covariate j values in the treatment and control groups respectively.

Overall bias was then computed as an arithmetic average of 10 biases in absolute terms:

$$\overline{bias} = 0.1\sum_{j=1}^{10}\left|bias_j\right| \quad (21)$$

To compute percentages of bias reduction, standardized percentage biases were estimated both before and after matching. Then, the latter was divided by the former. Finally, obtained proportions were subtracted from one and multiplied by 100 to express reduction in percentages. If bias reduction of covariate j is 100 percent, it implies that the mean value of covariate j in matched control sample is identical to the mean value of covariate j in corresponding treatment sample. Hence, the bias resulting from observed covariates is fully eliminated. However, this does not guarantee that pairs are well matched, since, frequently, not all important factors are observed. In theory, it is possible that two locations in the treatment and control samples may have similar observed characteristics, but at the same time they may differ in important characteristics, which are unobserved. However, if a sufficient number of propensity score covariates are included, such a possibility becomes unlikely. Bias reduction of zero percent would imply that the bias before and after the matching remained unchanged. Negative values in bias reduction would imply that matching degraded inferences relative to not matching. Poor matches are a consequence of lacking relevant control observations. Therefore, it is important that the control sample includes locations with characteristics similar to characteristics of treatment locations.

Further, approximate standard errors of treatment effects were used to assess whether or not estimates of avoided deforestation are statistical zeros. Those figures are computed assuming independent observations, homoskedasticity of the outcome variable within the treated and within the control groups and that the variance of the outcome does not depend on the propensity score (Leuven & Sianesi, 2003). The formula under the nearest neighbor matching simplifies to (n is the number of matched pairs):

$$SE(ATT) = \sqrt{\frac{\mathrm{var}(Y^1) + \mathrm{var}(Y^{0m})}{n}} \quad (22)$$

Finally, edge effects were tested by comparing ATTs in internal (core) and external (edge) areas of protected areas. The *t* statistic for the test of mean equality assuming unequal variances of ATT estimates of internal and external park areas was computed using *ttest* command in *Stata 11.0*[IX]. The null hypothesis of the test is that ATT in core area of a conservation unit is equal to ATT on its edge.

5.3. Results and discussion

Table 8. Determinants of propensity score

Variable	Coefficients and standard errors
Cons	-4361.662 (70.8596)*
Altitude	0.8638 (0.0616)*
Slope	114.1634 (8.3334)*
Surface flatness	-0. 1336 (0.004021)*
Agricultural suitability	-33.7718 (6.1986)*
Number of fires	-1081.213 (19.5347)*
Forest cover	9.5959 (0.4949)*
Time to city	-0.2872 (0.004072)*
Distance to river	0.007019 (0.0003034)*
Distance to road	0.018 (0.0001032)*
Precipitation	1.1179 (0.0224)*
McFadden R^2	0.2039
# obs	196494
Treatment (%)	29.5
Control (%)	70.5

Standard errors are in parentheses
* p<0.01
Parameters are multiplied by 1000
Buffer zones and park edges were not included in the treatment group
Only 20% of samples were used in the estimation (chosen randomly)

The fact that locations of conservation units are highly selective is reflected in Table 8[X]. The results reveal that all ten covariates are statistically significant at 1% level. Protected areas, as compared with areas under no protection, are situated in higher altitudes with steeper slopes and flatter terrains, in closer proximities to urbanized areas and further away from the nearest roads and rivers. Conservation units also face fewer fires, are covered by larger percentage of forests, and are located in regions with fewer constraints for agriculture and higher amount of precipitation.

To perform a quality assessment of matching, standardized percentage biases shall be analyzed. Table 9 provides statistics for such analysis. The values reported in Table 9 are weighted averages (weighted by matched pairs). The highest percentage of bias after matching remains in covariates of altitude, surface flatness and distance to the nearest road. The least percentage of the remaining bias is in fire covariate. Overall results achieved satisfactory mean absolute standardized percentage bias of 7.8%. There is no clear consensus over what percentage of bias should imply proper matching, but 10% mark has been taken to indicate a negligible difference in the means (Austin, 2011). Most bias was eliminated in fire covariate; also, more than 90% of bias was removed in precipitation and distance to nearest road covariates. The lowest percentage reduction in absolute bias

[IX] For technical note see Stata Base Reference Manual, Volume 1, A-H, Release 11, page 2002.

[X] Here a single logistic regression was estimated. However, matching was done independently for each geographical unit (park core, its edge, and buffer zone).

was achieved in slope and surface flatness variables. Overall, propensity score matching was successful, as it reduced the initial bias by almost 88%.

Table 9. Balance statistics: mean absolute standardized percentage bias (MASPB) after matching and percentage of bias elimination (BE)

MASPB	alt	slope	flat	agr	fires	fc	time	rivers	roads	prec	mean
Buffer	11.96	7.66	11.2	8.26	1.05	8.76	9.79	7.89	12.37	9.19	8.84
Edge	6.66	5.29	6.91	4.16	2.29	4.67	4.47	4.9	4.19	4.97	4.86
Core	8.14	7.56	8.02	7.28	2.48	4.1	6.15	6.89	5.7	8.68	6.53
All	10.58	7.15	10.08	7.33	1.41	7.57	8.42	7.19	10.18	8.26	7.84
BE, %	**alt**	**slope**	**flat**	**agr**	**fires**	**fc**	**time**	**rivers**	**roads**	**prec**	**mean**
Buffer	89.38	81.54	74.42	83.51	98.05	87.42	80.95	79.62	91.21	91.27	87.52
Edge	91.85	78.61	75.48	90.2	93.37	90.53	90.01	88.99	93.89	92.7	90.07
Core	90.52	75.02	74.71	82.03	93.56	92.01	85.77	86.26	92.91	87.83	87.57
All	89.85	80.76	74.6	84.67	97.1	88.18	82.97	82.35	91.61	91.3	87.93

The results revealed that 0.72% of conservation units' surface avoided deforestation during 2000 and 2004 inclusive due to the existence of protected areas (see Table 10). This figure translates into ~2900 km^2 of forests. Avoided deforestation in internal areas of conservation units is 1.12%, in external areas (edges) the result is 1.52%. However, excessive 0.89% of surface was deforested around conservation units as compared with no protection. Higher percentage of forests saved on park edges relative to core areas may be explained by arguing that in some instances edges of protected areas are subject to higher deforestation pressure than their internal areas, since the former are directly exposed to open (scarcely forested) areas, which facilitates access to forest resources. The figures also tend to suggest that no edge effects are present, but the full picture can be drawn only after the analysis on conservation unit level. Negative avoided deforestation in buffer zones suggests that the substitution effect dominates. That is, some deforestation events are directed towards the proximities of a protected area rather than occurring inside it. However, this implication is just a general result and does not necessarily hold for each buffer zone separately.

Further, the findings reveal that protected areas established prior to 1990 shielded a higher percentage of forests from deforestation than areas established during the succeeding decade. The result is as expected, as longer period of protection may imply more effective management. Unexpectedly, protected areas established during the study period avoided very similar percentage of deforestation as conservation units, present since at least 1989. Possibly, locations of newer conservation units are subject to much higher deforestation pressure. Protected areas governed by ministries or agencies avoided more deforestation than areas governed by indigenous people. The result is intuitive, since indigenous people in Pará live further away from urbanized areas, that is, were deforestation pressure is lower. It is also evident that strict protection areas (IUCN categories Ia, Ib and II)[XI] avoided roughly 1.5 times more deforestation than sustainable use areas (IUCN categories V and VI). Finally, large con-

[XI] Refer to IUCN (2015) for the description of IUCN categories.

servation units can be linked with lower avoided deforestation. This has two potential explanations. Firstly, large intact forests and their protected areas are usually located further away from urbanized areas (lower deforestation pressure). Secondly, larger areas are more difficult to manage.

Table 10. Average treatment effects on the treated by park characteristics

	ATT	№ protected areas	№ matched pairs
Overall	-0.72	70	80167
Buffer	0.89	17	16910
Edge	-1.52	17	5545
Core	-1.12	36	57712
Period:			
1961-1989	-1.45	15	24811
1990-1999	-0.65	14	23383
2000-2004	-1.41	7	9518
Government:			
indigenous people	-0.77	19	38723
agency/ministry	-1.82	17	18989
IUCN category:			
I-II	-2.48	4	2728
V-VI	-1.71	13	16261
Area (square km):			
200 – 3000	-1.29	11	2747
3000 – 10000	-2.01	15	12541
More than 10000	-0.84	10	42424

ATTs here are weighted averages of ATTs of geographical units
Avoided deforestation on edges of conservation units is not considered in the breakdown of ATTs by park characteristics (year of establishment, governing type, IUCN category, and area)

For comparison, Nolte et al. (2013), who investigated avoided deforestation in Legal Amazon, find that strict protection and sustainable use areas established before 2000 saved about 2% of forests during 2001 and 2005 year period. This figure seems to corroborate the findings of this study (approximately 1.8% forests saved over the five-year study period). However, Nolte et al. (2013) find that conservation units under the management of indigenous people avoided more than 4% of deforestation over a five-year period in Legal Amazon, whereas the findings of this research suggest only 0.8% of avoided deforestation in protected indigenous lands of Pará. Andam et al. (2008) concluded that the network of protected areas in Costa Rica saved around 10% of forests over almost four decades. This result is similar to the result for Pará, which is reasonable, because Costa Rica and Pará are similar in terms of climatic and economic conditions. Joppa & Pfaff (2011) find that on global scale around 0.5% of forests were salvaged from deforestation between 2000 and 2005.

It is important to understand that low figures of avoided deforestation do not imply that the management of these particular protected areas is ineffective in reducing deforestation. Some protected areas are precautionary, that is, are created in remote areas, where little deforestation happens. Therefore, low figures of avoided deforestation may reflect the degree of deforestation pressure rather than effectiveness. Since different locations face very different rates of deforestation, results of avoided deforestation vary greatly between different protected areas. This observation was made both by Nolte et al. (2013) and Joppa & Pfaff (2011), and is confirmed for Pará by analyzing Figure 9. The analysis shall be complemented by the analysis of balance and statistical significance of ATT estimates (Figure 8).

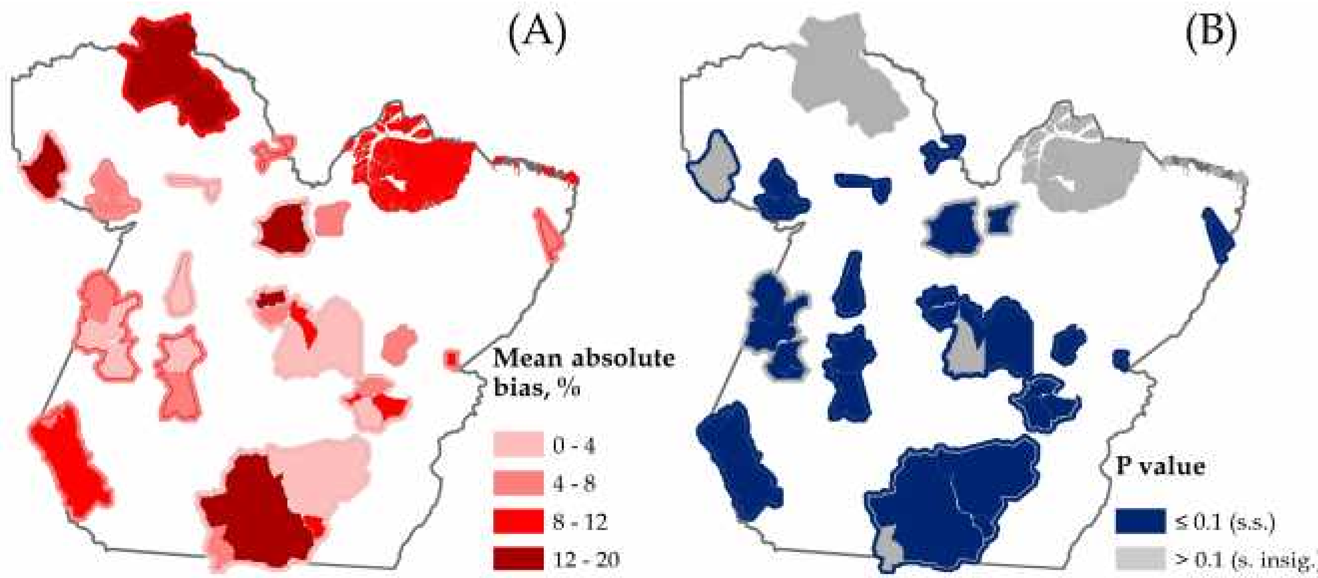

Figure 8. Mean absolute standardized percentage biases (A) and statistical significance of avoided deforestation (B)

As can be inferred from Figure 8A, characteristics of major part of small conservation units and buffer zones were well matched with characteristics of respective control samples. Larger territory of a conservation unit seems to negatively affect the quality of matching. Avoided deforestation estimates of conservation units that appear in dark colors in Figure 8A shall be interpreted with caution. Figure 8B reveals that avoided deforestation is a statistical zero in conservation units of Araweté Igarapé Ipixuna (5)[XII], Arquipélago do Marajó (6), Nhamundá/Mapuera (21), Panará (22), Rio Paru D'Este (25) and its buffer zone, and Parque do Tumucumaque (34) and its buffer zone, as well as in buffer zones around conservation units of Amazônia (2), Andirá-Marau (3), Caxiuanã (9), Itaituba I (11), Itaituba II (12), and Verde para Sempre (35). Zero avoided deforestation in those conservation units is an intuitive result, since forests located in northwestern Pará are remote and intact, and Arquipélago do Marajó is primarily aimed to protect marine resources.

Figure 9 reveals that the highest percentage of avoided deforestation happened in the central part of eastern Pará (located in so-called *Arc of Deforestation*), alongside the Trans-Amazonian highway and on the banks of Amazon River, that is, in deforestation hotspots. The most pro-

[XII] Figures in parentheses are spatial references of protected areas. These numbers correspond to the numbers in Figure P1. The names of conservation units can be found in Table 11.

tected conservation units include indigenous lands of Arara (4) and Parakanã (23), and national forest of Tapajós (30). Specifically, avoided deforestation in core areas of those conservation units are 9.5, 8.2 and 6.1 percent respectively over the five-year study period. Interestingly, two protected areas – Itacaiunas (10) and Alto Rio Guamá (24) – faced excessive deforestation relative to no protection at all. This is not a new finding. Nolte et al. (2013) stress that global protected area assessments have identified countries where protected areas exhibit higher rates of land use change than counterfactual territories. The two potential explanations provided by these authors are methodological weaknesses and protests against protection. Protests may arise because local communities no longer have access to natural resources upon which they depend.

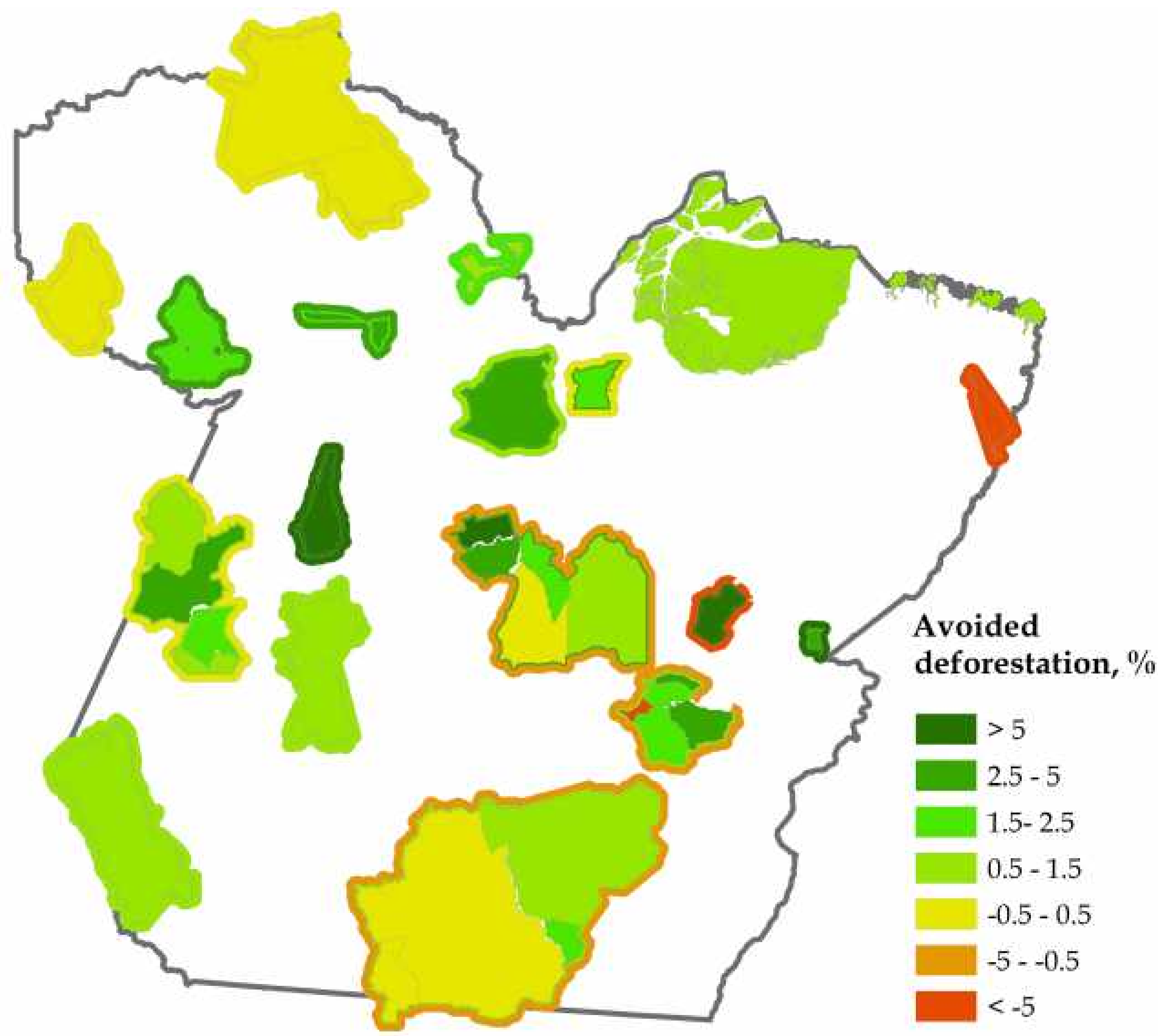

Figure 9. Avoided deforestation in percentages during 2000-2004 year period. Avoided deforestation on edges of protected areas can hardly be analyzed from the map. Therefore, the reader is directed to Table 11 for edge effect analysis. In cases where a buffer zone or park edge share a border (or overlap) with a protected area, excluded from the analysis, territories along that border were cut from buffer or edge zone. Only those observations that fall within territory of Pará were used in computations of ATT. However, on the map full territories of protected areas are depicted (both territories inside and outside the state of Pará).

An interesting observation can be made by analyzing geographical arrangement of protected areas in Pará. Some areas have borders alongside rivers or mountains, thus forming natural buffer zones. In Pará, many protected areas share borders, implying that the total perimeter of conservation units is reduced. This decreases the area exposed to edge effects and reduces the area needed for buffer zones, thus facilitating buffer zone management. Also, many protected areas are round, which further shortens the perimeter of conservation units.

As can be inferred from Figure 9, buffer zones of conservation units located to the west of highway BR-163 and on the banks of Amazon River contributed to avoiding deforestation. Most notably, those are the buffers around conservation units of Mulata (19), Rio Trombetas (26), Saracá Taquera (29), Tapajós (30), and, unexpectedly, Mãe Maria (17). In those territories legal protection during the study period was achieved without damaging surrounding areas, and even protecting them from deforestation. However, in the eastern part of Pará, known for its high rates of deforestation, the substitution effect takes over. Buffer zones that experienced the most additional deforestation due to the establishment of conservation units are the buffers of Parakanã (23) and Alto Rio Guamá (24). The main conclusion stemming from the findings is that buffer zones in Pará do avoid deforestation, but only up to a certain extent of deforestation pressure.

Table 11. Edge effect analysis

GU	Name of a protected area	Difference	$P(T < t)$	$P(\lvert T \rvert > \lvert t \rvert)$
1	Altamira, NF (1) Riozinho do Anfrísio, ER (27)	0.0098	0.5079	0.9842
2	Amazônia, NP (2) Andirá-Marau, IL (3) Itaituba I, NF (11) Itaituba II, NF (12)	-0.5451	0.1691	0.3383
3	Arara, IL (4) Araweté Igarapé Ipixuna, IL (5) Kararaô, IL (15) Koatinemo, IL (16) Trincheira Bacaja, IL (33)	1.0555	0.9809	0.0383
4	Badjonkore, IL (7) Kayapó, IL (14) Menkragnotí, IL (18) Panará, IL (22)	0.0504	0.5361	0.9278
5	Carajás, NF (8) Itacaiunas, NF (10) Tapirapé, BR (31) Tapirapé-Aquiri, NF (32) Xikrin do Rio Catete, IL (36)	-0.8739	0.2391	0.4783
6	Caxiuaná, NF (9)	1.1873	0.8394	0.3212
7	Jari, ES (13)	1.113	0.9487	0.1026

Table 11 (continued)

8	Mãe Maria, IL (17)	-0.4672	0.4469	0.8938
9	Mulata, NF (19)	-1.5833	0.047	0.094
10	Munduruku, IL (20) Sai-Cinza, IL (28)	0.4762	0.8648	0.2705
11	Nhamundá/Mapuera, IL (21)		Not reported	
12	Parakanã, IL (23)	3.3291	0.9511	0.0978
13	Alto Rio Guamá, IL (24)	-5.0842	0.0485	0.0969
14	Rio Paru D'Este, IL (25) Parque do Tumucumaque, IL (34)		Not reported	
15	Rio Trombetas, BR (26) Saracá Taquera, NF (29)	0.5177	0.7881	0.4239
16	Tapajós, NF (30)	0.2512	0.5723	0.8553
17	Verde para Sempre, ER (35)	0.1758	0.5506	0.8987

Abbreviations: BR – biological reserve, ES – ecological station, ER – extractive reserve, IL – indigenous land, NF – national forest, NP – national park; environmental protection area of Arquipélago do Marajó (marked as 6 in Figure P1) is not included, since it is not required to be surrounded by a buffer zone by law; difference here is calculated as $DIFF = ATT^{core} - ATT^{edge}$; some results are not reported, since avoided deforestation in both core and edge were found to be statistical zeros

It shall not be forgotten that buffer zones are multifunctional and one of their goals is to protect conservation units from edge effects. *P* values of two alternative hypotheses are reported in Table 11: one stating that ATTs are not equal and the other that ATT on park edge is lower than in its core. Generally, the results suggest no evidence of edge effects in Pará. This implies that park edges avoid statistically equal percentage of deforestation as compared to core areas. Possibly, this is due to effective buffer zone management.

Finally, it is interesting to visually inspect maps to see if the existence of protected areas can be visually observed. Figure 10 is an image of an area with high deforestation pressure, where protected areas of Parakanã and Mãe Maria are located. The contours of both protected areas can be seen clearly in Figure 10, indicating that legal protection is a very effective measure in avoiding deforestation within parks' territories. The picture also seems to support the conclusion of no edge effects in those two conservation units, since areas along the edges and core areas seem to be equally intact. However, lands in the immediate proximities of both protected areas are extensively deforested, suggesting a strong substitution effect. Also, both conservation units can be dubbed as *green islands*[XIII]. The fact that only protected lands remain forested in territories with high deforestation pressure underlies the importance of legal protection.

[XIII] Isolation of vegetation is a huge problem in ecology. For implications of *green island* problem refer to DeFries et al. (2005).

Figure 10. Parakanã and Mãe Maria (picture taken on 2014 04 30). Source: Map data © 2014 Google.

5.4. Conclusions

All ten covariates proved to be important factors to consider in constructing a propensity score. Propensity score matching as a technique proved to be successful, since overall bias after matching dropped to satisfactory 7.8%, thus eliminating 88% of the initial bias. However, matching in few larger protected areas featured some balance problems.

The findings suggest that Pará's network of protected lands avoided 0.72% of deforestation during 2000 and 2004 inclusive as a consequence of protection. Avoided deforestation in cores and edges of protected areas was 1.12% and 1.52% respectively. However, buffer zones experienced excessive deforestation equal to 0.89% as a result of displaced deforestation. Conservation units that avoided the most deforestation are those established prior to 1990, governed by ministries or agencies, classified under categories I and II under IUCN scheme, and smaller in territory. The results also indicate that the most deforestation was avoided in protected areas located on a deforestation frontier in mid-eastern Pará, alongside the Trans-Amazonian highway, and on the banks of Amazon River. More than 5% of deforestation was avoided in Arara, Parakanã and Tapajós. In protected areas, isolated by natural obstacles, avoided deforestation was concluded to be a statistical zero. Those include Arquipélago do Marajó, Nhamundá/Mapuera, Rio Paru D'Este, and Parque do Tumucumaque. Additionally, legal protection had no effect on deforestation patterns in Araweté Igarapé Ipixuna and Panará. Avoided deforestation in Alto Rio Guamá and Itacaiunas was found to be negative.

Buffer zones around conservation units located to the west of highway BR-163 and on the banks of Amazon River avoided deforestation. The most deforestation was avoided in buffer zones around Mulata, Rio Trombetas, Saracá Taquera, and Tapajós. On the contrary, buffer zones located in eastern Pará were overwhelmed by the substitution effect (except for the buffer around Mãe Maria). The most severe excessive deforestation was found in buffer zones of Alto Rio Guamá and Parakanã. However, the results offer no evidence of edge effects in spite of geographical location of a conservation unit.

The findings call for stronger buffer zone management on deforestation frontiers. However, the results cover time period 2000-2004. Analyses based on recent observations are needed.

Unfortunately, prospects of such analyses are hampered by the lack of counterfactual samples resulting from a continuous creation of new conservation units. Future research should focus on finding ways to perform matching under the shortage of adequate control pairs.

5.5. References

Alexandre, B., Crouzeilles, R., & Grelle, C.E.V. (2010). How Can We Estimate Buffer Zones of Protected Areas? A Proposal Using Biological Data. *Natureza and Conservação, 8*(2), 165-170.

Andam, K.S., Ferraro, P.J., Pfaff, A., Sanchez-Azofeifa, G.A., & Robalino, J.A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences, 105*(42), 16089-16094.

Austin, P.C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*(3), 399-424.

Becker, S.O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *The Stata Journal, 2*(4), 358-377.

Burrough, P.A., & McDonell, R.A. (1998). *Principles of Geographical Information Systems*. Oxford: Clarendon Press.

Cochrane, M., & Laurance, W.F. (2002). Fire as a large-scale edge effect in Amazonian forests. *Journal of Tropical Ecology, 18,* 311-325.

DeFries, R., Hansen, A., Newton, A.C., & Hansen, M.C. (2005). Increasing isolation of protected areas in tropical forests over the past twenty years. *Ecological Applications, 15*(1), 19-26.

GADM (Database of Global Administrative Areas) (2014). *Version 2.0*. Retrieved from http://www.gadm.org/country

GEMU-JRC (Global Environment Monitoring Unit - Joint Research Centre of the European Commission) (2014). *Travel time to major cities: A global map of Accessibility*. Retrieved from http://bioval.jrc.ec.europa.eu/products/gam/index.htm

GEOFABRIK (2014). *OpenStreetMap Data Extracts*. Retrieved from http://download.geofabrik.de/

Google Maps (2014). *Parakanã and Mãe Maria*. Retrieved from https://maps.google.com/

IIASA (International Institute for Applied System Analysis) (2014). *Global Agro-Ecological Zones dataset. Plate 28: Climate, soil and terrain slope constraints combined*. Retrieved from http://webarchive.iiasa.ac.at/Research/LUC/GAEZ/

INPE (Brazil's National Institute for Space Research) (2014a). *Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satélite*. Retrieved from http://www.obt.inpe.br/prodes/index.php

INPE (Brazil's National Institute for Space Research) (2014b). *Monitoramento de Queimadas*. Retrieved from http://www.inpe.br/queimadas/

ISA (Brazil's Socio-Environmental Institute) (2015). *Protected Areas Monitoring Program and Geoprocessing Laboratory. Selection of areas in Brazil*. Retrieved from http://uc.socioambiental.org/

IUCN (International Union for Conservation of Nature) (2015). *IUCN Protected Areas Categories System*. Retrieved from http://www.iucn.org/

Joppa, N.L., & Pfaff, A. (2011). Global protected area impacts. *Proceedings of the Royal Society, 278,* 1633-1638.

Leuven, E., & Sianesi, B. (2003). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.*

Martino, D. (2001). Buffer Zones Around Protected Areas: A Brief Literature Review. *Electronic Green Journal, 1*(15).

MMA (Brazil's Ministry of Environment) (2014). *Consulta - Relatórios de UC*. Retrieved from http://www.mma.gov.br/areas-protegidas/cadastro-nacional-de-ucs/consulta-gerar-relatorio-de-uc

NASA (National Aeronautics and Space Administration) (2014). *MODIS data*. Retrieved from http://reverb.echo.nasa.gov/reverb

Nelson, A., & Chomitz, K.M. (2009). *Protected Area Effectiveness in Reducing Tropical Deforestation. A Global Analysis of the Impact of Protection Status. Evaluation Brief 7*. Washington, D.C.: the World Bank.

Nolte, C., Agrawal, A., Silvius, K.M., & Soares-Filho, B.S. (2013). Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. *Proceedings of the National Academy of Sciences, 110*(13), 4956-4961.

Rosenbaum, P., & Rubin, D.B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician, 39*(1), 33-38.

SRTM (Shuttle Radar Topography Mission) (2014). *90m Digital Elevation Model, version 2.1*. Retrieved from http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/

WDPA (World Database of Protected Areas) (2014). *WDPA dataset*. Retrieved from http://www.wdpa.org/

World-Clim (2015). *Data for current conditions (~1950-2000)*. Retrieved from http://www.worldclim.org/current

Chapter 6

Using artificial neural networks and eigenvectors to predict deforestation at 5x5 kilometer grids

Abstract. This article aims to map deforestation at 5x5 km grids from lagged and time-fixed data. Important innovation lies in inclusion of eigenvectors to solve spatial autocorrelation problem and to improve prediction accuracy. Initially, data on 20 covariates is collected, mostly consisting of past deforestation in a cell and in its nearest neighbors, distance covariates, protection covariates and terrain characteristics. Using backward elimination procedure 10 key covariates are identified. The residuals of OLS model with selected variables are severely clustered. However, after adding eigenvectors to the covariate list OLS model errors are randomly distributed, and adjusted R squared increases by more than 50%. Further, artificial neural network is applied to increase generalization ability. Testing sample RMSE of trained network is lower than the corresponding measure of OLS by more than 10%. Correlation coefficient between actual deforestation and simulated values from the final artificial neural network with eigenvectors is 0.79.

Keywords. Deforestation, prediction, eigenvectors, artificial neural networks.

6.1. Introduction

Anthropogenic land conversion is a huge threat for environmental sustainability. The conversion from forests to pastures and, to a lesser extent, to croplands continues in the tropical rainforests of the Amazon. In order to properly respond to deforestation, the authorities need fine resolution maps of deforested areas. Academic researches on mapping and predicting deforestation could be grouped into four distinct types. Firstly, deforestation maps are produced for deforestation monitoring. In Brazil deforestation activity is being traced by DETER, which is a satellite-based system that captures and processes georeferenced imagery on forest cover in 15-day intervals. These images are used to identify deforestation hotspots and issue alerts signaling areas in need of immediate attention (Assunção et al., 2013). The system can detect deforestation events larger than 25 hectares. Recent work by Diniz et al. (2015) describes the upgraded variant of DETER, called DETER-B, which is capable of detecting logging spots smaller than 25 hectares. Kehl et al. (2012) use MODIS imagery to develop a tool capable of detecting daily deforestation in the Amazon. Jacobson et al. (2015) offer a method to detect near real-time deforestation using Google imagery. Some studies map deforestation that already happened. Those studies usually offer some refinements over existing methods. Examples of such researches are Arai et al. (2011), Morton et al. (2005) and Zhan et al. (2002). Asner et al. (2006) map selective logging undetectable by DETERS system. Another sort of studies aims to project deforestation into far future. Among such investigations are Lapola et al. (2011), Laurance et al. (2001), Moreira et al. (2009), Rosa et al. (2013), Soares-Filho et al. (2006) and Wassenaar et al. (2007). Those studies offer deforestation projections under different policy scenarios. Finally, some academic works aim to predict next period deforestation. Such studies are scant. One example is the research by Mas et al. (2004), who predict next period deforestation in Mexico. Such investigations offer important benefits, since they 1) provide a better understanding on how various processes are linked with deforestation, 2) generate future scenarios of deforestation, 3) predict the locations of forest clearings and 4) are important piece of information for policy making (Mas et al., 2004). This paper contributes to the literature by offering a method to map deforestation from past and time-fixed variables.

The study exploits the fact that deforestation is spatially and temporarily contagious. Contagion implies that locations surrounded by recently deforested areas are more likely to suffer deforestation themselves (Rosa et al., 2013). This phenomenon is widely discussed in academic literature (see, for example, Aguiar et al., 2007, Alves, 2002, Asner et al., 2006, Espindola et al., 2011, Robalino and Pfaff, 2012, Wassenaar et al., 2007). Figure Q1 illustrates deforestation contagion in the study area. It reveals that deforestation are not only spatially, but also temporally contagious. The fact that deforestation is contagious is promising for prediction. However, positive spatial correlation of observations leads to clustered model errors when deforestation is predicted from its determinants by OLS regression. What is important, model prediction accuracy could be improved greatly by eliminating non-randomness in spatial distribution of model errors. An effective way to avoid spatial autocorrelation is to select cells from every k^{th} row and column, as is done by Arima et al. (2007). Some authors apply methods that eliminate spatial dependencies without sacrificing observations. For example, Aguiar et al. (2007) implement spatial lag model. This research uses the technique known as spatial filtering. The main source of spatial autocorrelation in the dependent variable is unobserved covariates (Tiefelsdorf and Griffith, 2007), whereas spatial filtering is a modern and the most promising method in filtering unobserved covariates from the initial model errors.

The study uses OLS regression to obtain eigenvectors. However, predicted deforestation is an output of trained artificial neural network (ANN). The method has strong capability to capture nonlinearities between the dependent variable and its covariates, is non-parametric and is able to learn data relationships that are not otherwise known. Examples of ANN applications in deforestation and forest fire modeling are found in Kehl et al. (2012), Maeda et al. (2009) and Mas et al. (2004).

6.2. Study area

The study area is a regular square tessellation. It is located in the state of Rondônia between approximately 9° 23′ and 10° 33′ South and 63° 32′ and 64° 48′ West (Figure 11). It encompasses ~18000 square kilometers and contains 728 5x5 kilometer parcels. The northern half of the study area lies in the municipality of Porto Velho, southwestern quadrant is located in Nova Mamoré, Buritis is located in the mid section of the study area, and southeastern corner lies in Campo Novo de Rondônia. Additionally, the study region includes western edges of Alto Paraíso, Ariquemes and Monte Negro.

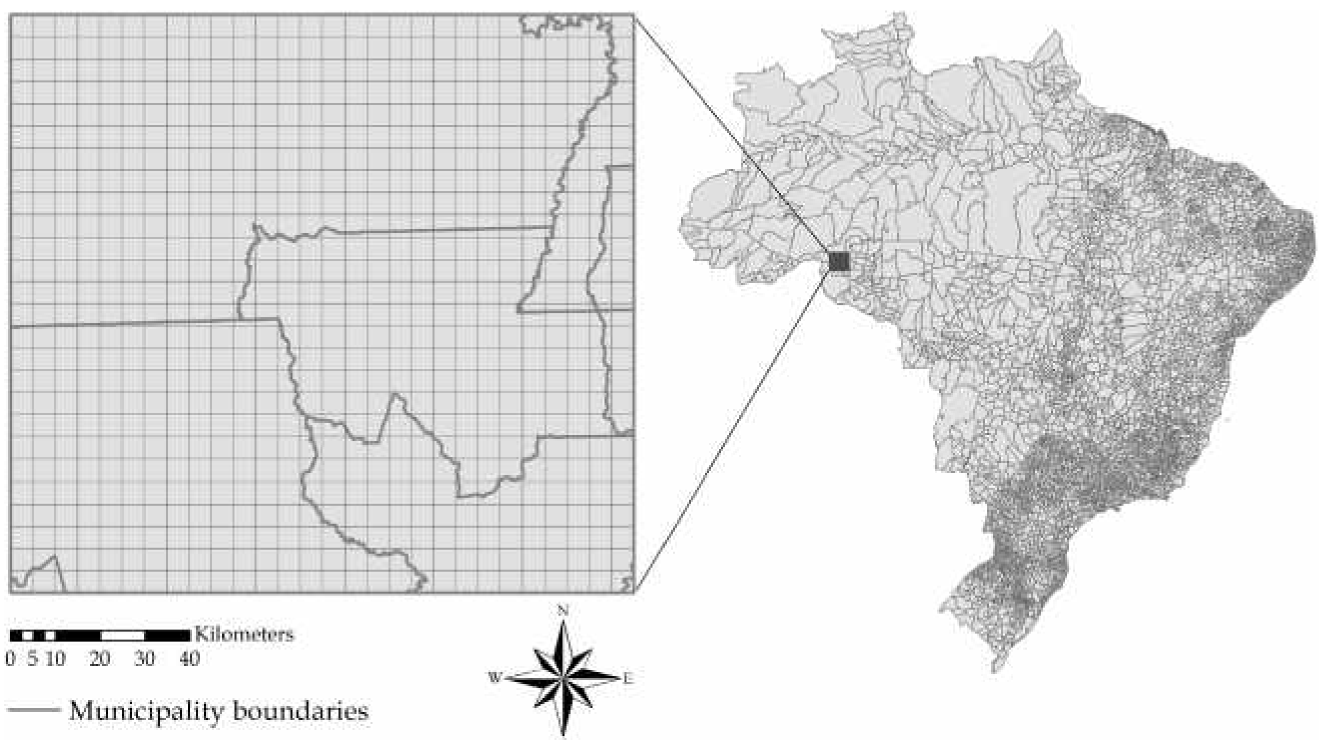

Figure 11. Geographical location of the study area

The study area is densely covered by unofficial road network. It totals more than 6500 kilometers (see Figure 12). The only city with more than 25000 inhabitants is Buritis. Some forests within selected study area are legally protected. Specifically, conservation units include indigenous lands of Karipuna and Karitiana (southern edge), national park of Guajará-Mirim (northern part), extractive reserve of Jaci-Paraná and national forest of Bom Futuro. The study area can be characterized by intensive deforestation. In 2012 the area of cleared forests totaled 230 km². The following year deforestation reached 431 km². In 2014 additional 378 km² were deforested. At the end of 2014 8636 km² of forests remained.

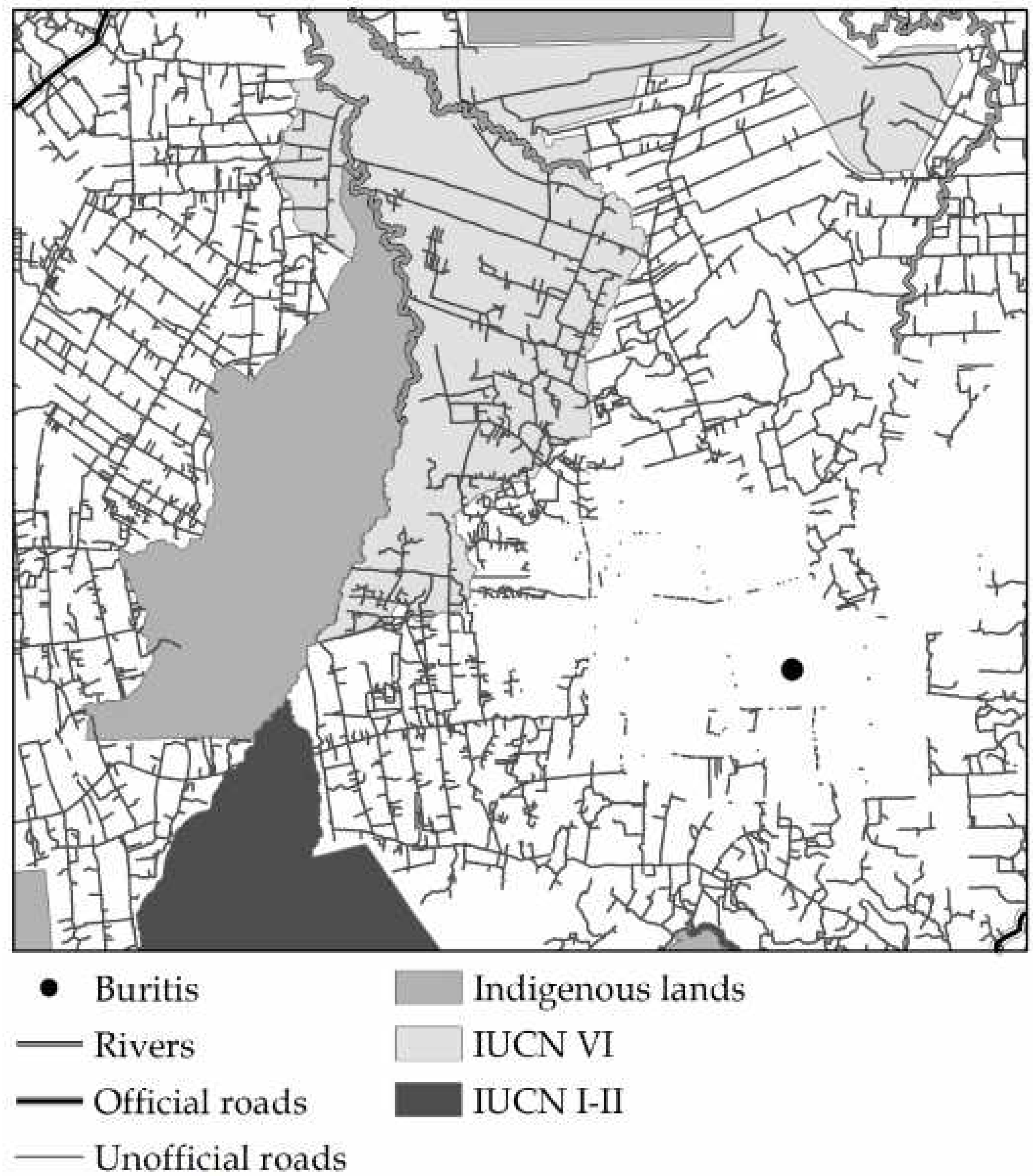

Figure 12. Location of roads, rivers, city of Buritis and conservation units. Those are classified into indigenous lands, sustainable use areas (IUCN category VI) and strict protection areas (IUCN categories I and II).

6.3. Materials and methods

This section is comprised of three parts. Firstly, data sources and main processing steps are revealed. Second part explains how observed covariates and eigenvectors are selected. Lastly, the calibration of the artificial neural network is concisely described.

6.3.1. Data sources and processing

Initially, data on 21 variables was collected. Firstly, PRODES map of 2014 in raster format was downloaded from Brazil's National Institute of Space Research (INPE, 2016a). The cells are classified into forested, non-forested and deforested by the year of deforestation (some cells are classified as hydrography and clouds, but those are few). The data is distributed at ~60 m spatial resolution. Deforestation in 2014 (target variable) was measured as the percentage of cleared pixels in each 5x5 km parcel. Studies that model deforestation in Legal Amazon at square grids (those are Kirby et al., 2006 and Rosa et al., 2013) guided which covariates to select. This study is based on the hypothesis that deforestation is spatially and temporally contagious. Therefore, four covariates, consisting of past deforestation (one-year and two-year lags, coded as *d12* and *d13* later in the text) in a parcel and in its adjacent parcels (*d12f* and *d13f*), were included in the analysis. I adhere to Queen's definition of adjacency[XIV]. Likewise, previous period forest coverage (*for13*) was estimated as a percentage of forested pixels within each 5x5 km parcel. Lagged fire covariate (*f13*) and its focal variable (fire in adjacent cells, *f13f*), measured as the total number of fire events, were constructed from decimal coordinates of fire events, available on INPE's website (INPE, 2016b). The motivation to include fires is that selective logging usually precedes the burning. Since the moisture content is initially high, the vegetation is left to dry until the next dry season, when it is finally burned in order to clear the area (Cochrane and Laurance, 2002). Thus, fires reflect past deforestation and potentially can predict future clearings if the hypothesis of temporal contagion holds. Further, five distance covariates were included: Euclidean distances to the nearest official road (*rof*), unofficial road (*runf*), river (*riv*), city (*cit*) and open area (*op13*). Road shapefiles (both official and unofficial roads) were obtained from IMAZON with help of Stefânia Costa. River shapefile was retrieved from GEOFABRIK, OpenStreetMap (GEOFABRIK, 2016). Decimal coordinates of Amazon's cities are distributed by Brazil's Institute of Geography and Statistics (IBGE, 2015). Distance to open area is a measure of accessibility. It is measured as Euclidean distance to the nearest non-forested cell. Thus, cells in open areas have value 0 and isolated cells in the middle of forests have high values. All distance covariates were estimated at 60 meter spatial resolution and later aggregated by averaging the values to form 5x5 kilometer grids. This average provides an effective index of density (Laurance et al., 2002). The research further includes three protection covariates: one for indigenous lands (*indig*), one for sustainable use areas (*sust*) and one for strict protection areas (*strict*). The polygons of conservation units are freely accessible at the World Database of Protected Areas (WPDA, 2016). The polygons were merged onto the polygon of Rondônia. The combined shapefile was converted into 60x60 meter raster grids. In this way, the borders of conservation units can be adequately reflected. Each cell was classified either as protected (discriminating between the three types of protection) or unprotected. Afterwards the data was aggregated by calculating the mean of binary values within each 5x5 km^2 quadrant. Next, the list of covariates was supplemented by terrain and climatic characteristics. Elevation in meters (*elev*) comes from the Shuttle Radar Topography Mission 90m Digital Elevation Model, version 2.1 (SRTM, 2016). Slope in degrees (*slope*) was calculated from adjacent cells. A measure of climate, soil and terrain constraints

---

[XIV] According to Queen's definition a cell is adjacent if it either shares a border or a corner with a neighboring cell.

for agriculture (*soil*) was obtained from the International Institute for Applied System Analysis, Global Agro-Ecological Zones dataset, Plate 28 (IIASA, 2016). Precipitation in millimeters (multiyear average, *prec*) is found on World-Clim's website (World-Clim, 2016). Lastly, rural population in 2015 (*pop*) was retrieved from WorldPop (WorldPop, 2015). All covariates were grouped into time-varying (dynamic) and time-consistent (static). Past deforestation and fires and their focal variables, forest cover, distance to open area and protection covariates were treated as dynamic. Some variables that are dynamic were treated as static due to data constrains. Those include rural population, distances to roads, distance to cities and precipitation. However, as those variables change slowly over time, static treatment does not pose a serious concern.

6.3.2. Covariate selection and spatial filtering

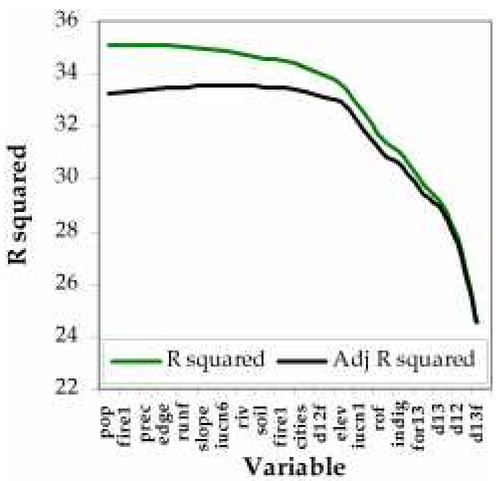

Figure 13. Covariate selection. Horizontal axis shows the variable to be eliminated.

Initially, the covariates were assumed to be linearly connected with deforestation and were selected by backward elimination (the variable with the highest p value is discarded and the model is re-estimated). The initial OLS model included all 20 covariates, which continued to be discarded as long as all covariates become statistically significant at 5 percent level. Figure 13 graphically illustrates the procedure step-by-step. In this way, 10 predictors of deforestation were identified. Those include past deforestation (both one-year and two-year lags) and corresponding focal variables, lagged forest cover, elevation, strict protection and indigenous governance covariates, distance to the nearest official road and distance to the most proximate city. Further, it was verified that no substantial multicollinearity between the picked variables exists (Table R1).

Deforestation is also explained by unobserved covariates. Some of the missing covariates can be effectively replaced by eigenvectors, constructed by the means of eigenvector-based spatial filtering. Under presence of spatial autocorrelation, model residuals can be viewed as composites of unobserved covariates and white noise (Griffith and Peres-Neto, 2006). Following Tiefelsdorf and Griffith (2007) and Griffith (2000), the rationale of spatial filtering is that eigenvectors that are extracted from a transformed spatial link matrix exhibit distinctive spatial patterns with associated spatial autocorrelation levels. Formally, the transformed spatial link matrix is given by:

$$\left(I - \frac{jj^T}{n}\right) C \left(I - \frac{jj^T}{n}\right) \quad (23)$$

Here I is an identity matrix, j is a vector of ones, n is a number of observations, C is a connectivity matrix and letter T denotes transpose. In this research the connectivity matrix is binary coded with value 1 if a cell is the nearest neighbor and 0 otherwise. The transformation in eq.

23 guarantees that the extracted eigenvectors are orthogonal and linearly independent (Griffith, 2000). Key implication of orthogonality is that the eigenvectors follow a strict sequence whereby each eigenvector explains a specific proportion of the variance in the residuals, with the first selected eigenvector capturing the largest amount of variation, the second selected eigenvector the second largest proportion, and so on (Tiefelsdorf and Griffith, 2007). The approach is semiparamatric. Eigenvectors are estimated from the dataset (non-parametric part) and then are additively coupled with the set of covariates whose coefficients need to be estimated (parametric part). Therefore, the covariates are used in assessment of which eigenvectors to choose, but are not used in constructing the eigenvectors.

Spatial filtering was done in *R* software, using library *spdep*. The algorithm implemented in *R* finds the single eigenvector reducing the standard variate of Moran's I for regression residuals most, and continuing until no candidate eigenvector reduces the values by more than predefined value. For further interest in spatial filtering refer to Griffith (2013). See supplementary material for programming codes used to extract eigenvectors.

6.3.3. Network calibration

Deforestation values were simulated from trained artificial neural network. It is worth noting that the obtained set of eigenvectors is sub-optimal for ANN. However, statistical packages cannot estimate eigenvectors that nullify spatial autocorrelation in ANN model errors directly. Such a procedure would be computationally cumbersome.

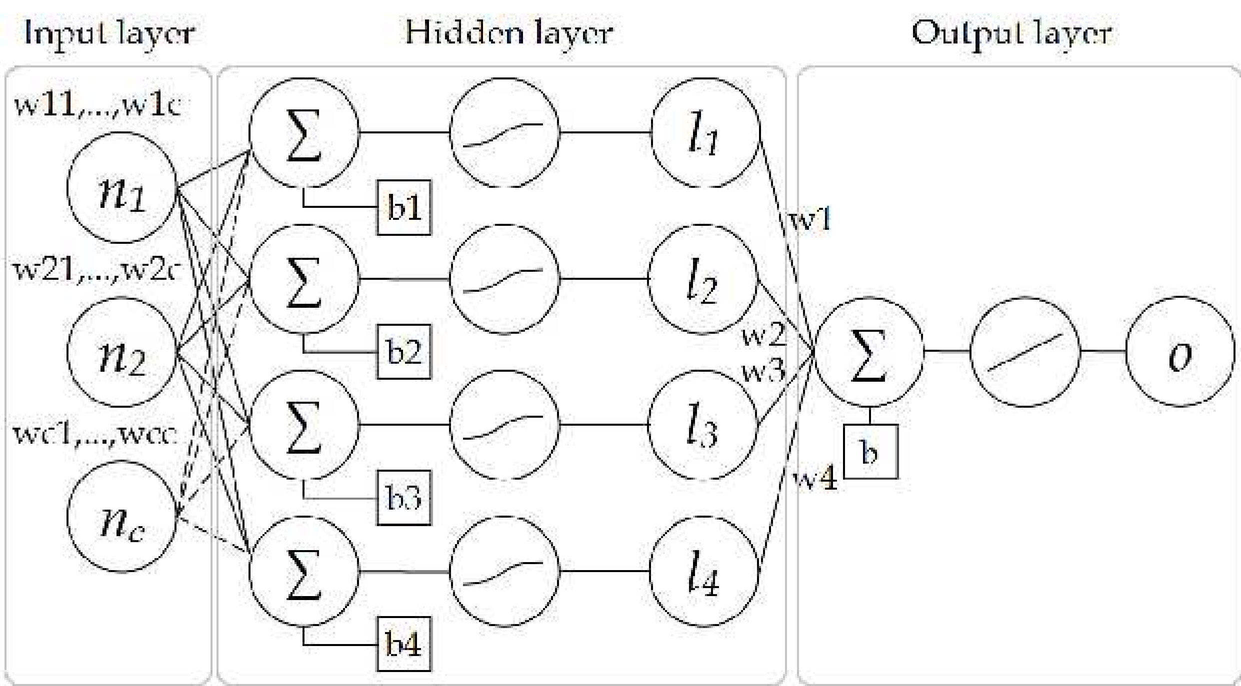

Figure 14. Graphical illustration of artificial neural network with *c* inputs and 4 hidden neurons

ANN training was done in MATLAB. Before network training, the dataset was randomly partitioned into training, validation and testing sets by the ratio 7:2:1. Training set is used to calibrate the network, validation set serves as a hedge against overfitting and testing set is used to evaluate generalization ability of the trained network. Figure 14 can be used to explain the derivation of output functions. Summation signs in the graph indicate that the products of inputs (denoted by letter *n*) and input layer weights are added together. Biases (additional weights to enhance flexibility) are also added (in the graph biases are embedded in the squares). The expressions are then passed through a transfer function (illustrated by the bent lines). In this study log-sigmoid function was used. Next, the expressions (denoted by letter *l*) are weighted by the hidden layer weights and summed (also adding the bias). This final expression is the output function (*o*). The training process error of observation *i* is the difference between target (actual) value of observation *i* and its output function.

Network training was based on Levenberg–Marquardt algorithm. Let's define a Jacobian matrix (*J*) as a matrix whose elements are first order derivatives of training process errors with respect to weights and biases. Matrix *J* is written as follows (Yu and Wilamowski, 2011):

$$J = \begin{pmatrix} \frac{\partial e_1}{\partial w_1} & \frac{\partial e_1}{\partial w_2} & \cdots & \frac{\partial e_1}{\partial w_c} & \frac{\partial e_1}{\partial b_1} & \cdots & \frac{\partial e_1}{\partial b_d} \\ \frac{\partial e_2}{\partial w_1} & \frac{\partial e_2}{\partial w_2} & \cdots & \frac{\partial e_2}{\partial w_c} & \frac{\partial e_2}{\partial b_1} & \cdots & \frac{\partial e_2}{\partial b_d} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial e_p}{\partial w_1} & \frac{\partial e_p}{\partial w_2} & \cdots & \frac{\partial e_p}{\partial w_c} & \frac{\partial e_p}{\partial b_1} & \cdots & \frac{\partial e_p}{\partial b_d} \end{pmatrix} \quad (24)$$

The Jacobian matrix has *p* rows and *c*+*d* columns. Letter *p* is the size of the training set, *c* is the number of weights and *d* is the number of biases. At the beginning of calibration, all weights and biases are generated randomly. Those figures are used to obtain numerical values of the elements in matrix *J*. Under Levenberg–Marquardt's method, initial weights and biases are iteratively updated according to the following rule:

$$W_{(t+1)} = W_{(t)} - \left(J_{(t)}^T J_{(t)} + \text{«} I \right)^{-1} J_{(t)}^T e_{(t)} \quad (25)$$

Index *t* marks iterations, *W* is a composite vector of weights and biases, *e* is a vector of errors, $\text{«} I$ is the product of combination coefficient and identity matrix. The latter term ensures that the expression within the brackets in eq. 25 (approximation of Hessian matrix) is invertible. The procedure in eq. 25 continues until validation error starts to increase. In this study root mean squared error (RMSE) was used. Increasing validation error signals that the network is starting to overfit and, therefore, the calibration has to be terminated. Optimal set of weights and biases is then used to simulate target values of the testing sample. Those values are then compared with actual values to assess generalization ability of the trained network. Specifically, the study used RMSE as a measure of generalization. Formal description of Levenberg–Marquardt algorithm is found in Marquardt (1963) and Levenberg (1944). For detailed theory and applications of neural computing refer to Ripley (2008).

The calibration of the final neural network was accomplished in three steps. Firstly, variables that may have non-linear relationships with deforestation were identified by graphical analysis. Those include precipitation, lagged distance to open area, distance to unofficial roads and rural population. However, due to multicollinearity (see Table R1), only the latter two were retained. Next, 10 networks with identical (MATLAB's default) settings were run on each dataset. Note that the validation error reaches only local minimum, implying that different sets of initial weights and biases lead to different results. A preferable dataset is the one that leads to a lower testing sample RMSE. In the second step optimal network configuration was empirically determined by testing different numbers of hidden neurons. Specifically, neural networks with 2, 4, 6, 8, 10, 12, 14 and 16 neurons in the hidden layer were calibrated 10 times each. Lastly, the best configuration was run 100 more times to select the final network.

6.4. Results

Spatial distribution of the initial OLS model residuals is illustrated in Figure 15A. It suggests that the residuals are highly clustered. Clustering is verified by the Moran's statistic: the null hypothesis that errors are randomly distributed in space is strongly rejected (see Table 12).

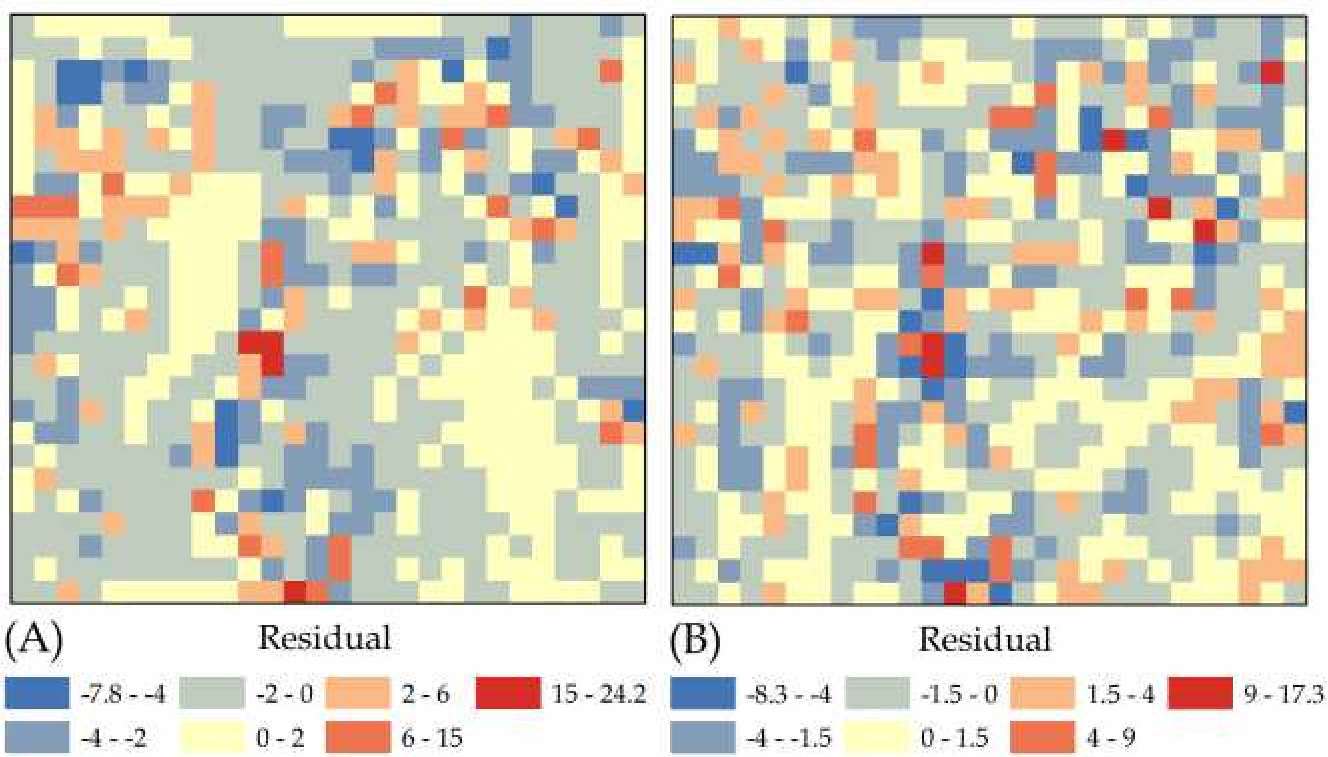

Figure 15. Spatial distribution of OLS (A) and OLS-E (B) residuals

Table 12. Moran's I of model residuals

Model	Moran's I	E(MI)	10^3 var(MI)	z score	p value
OLS	0.2235	0.0014	0.691	8.5552	0
OLS-E	0.0188	0.0014	0.7	0.7632	0.4454

Spatial filtering identified 39 eigenvectors that capture spatial autocorrelation levels. Appendix S visualizes the map patterns of the first six eigenvectors. The values of the first eigenvector are highly spatially clustered, indicating that it captures a large share of spatial autocorrelation. The maps become increasingly fragmented as more eigenvectors are included. Figure 15B portrays the residuals of OLS model with eigenvectors (further as OLS-E), which appear to be randomly distributed. Moran's statistic confirms the randomness in the residuals (see Table 12). Nevertheless, the most important question for this research is to what extent the filtered eigenvectors improve the prediction power. Table 13 presents the key statistics. Even though the inclusion of 39 eigenvectors leads to a sizeable loss in the degrees of freedom, adjusted *R* squared increases by more than 50%.

Table 13. Prediction power of OLS and OLS-E models

Model	# variables	R squared	Adj R squared	Correlation
OLS	10	34.26	33.34	58.5
OLS-E	49	53.42	50.05	73.1

ANN calibration yielded the following results: 1) additional covariates that may have non-linear relationships with deforestation do not improve generalization ability, 2) optimal number of hidden neurons is 4, 3) testing sample RMSE of the final model is 2.84. The latter figure means that the generalization was improved by more than 10% compared to OLS–based model (testing sample RMSE of OLS-E model is 3.17). The correspondence between actual deforestation in 2014 and the simulated values for that year, measured in terms of Pearson correlation coefficient, is 0.79. Figure 16 visually contrasts actual and simulated values.

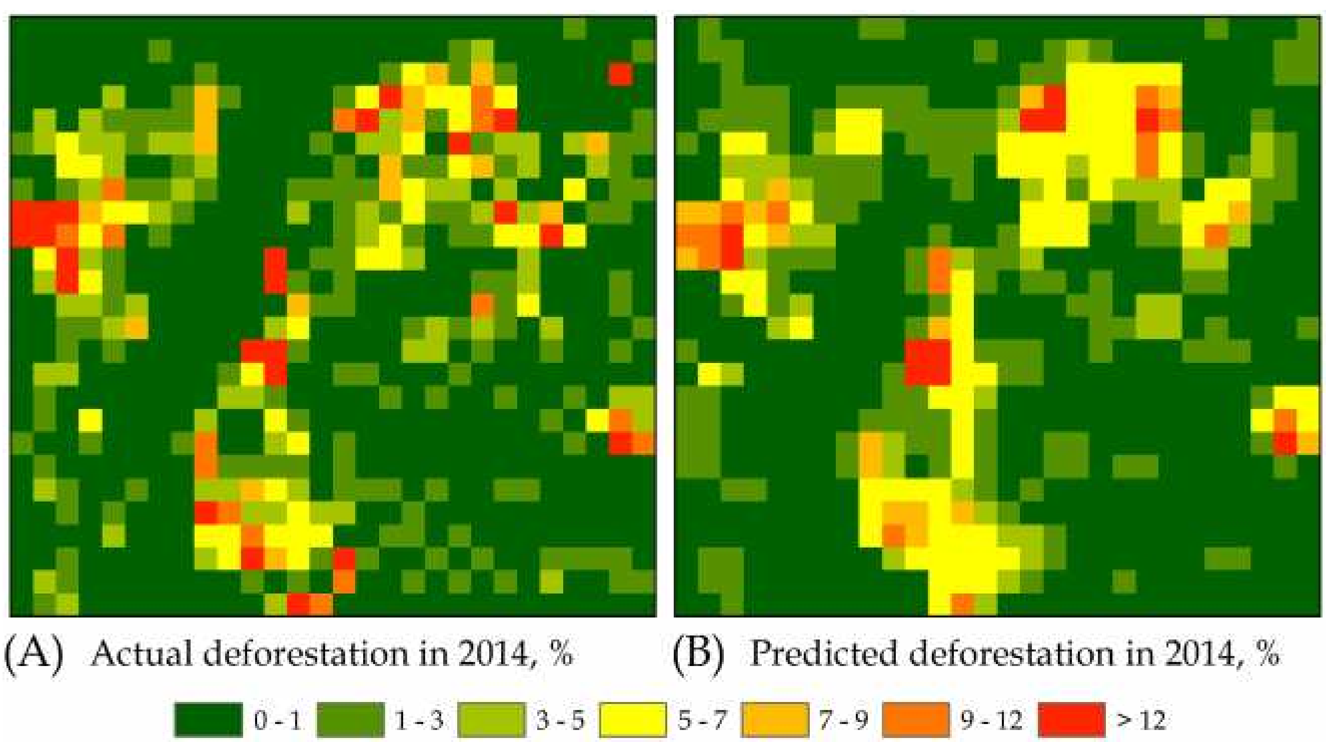

Figure 16. Actual and simulated deforestation in 2014

Simulated map captures very well general patterns of deforestation, implying that the generalization would be higher at coarser scales. Cumulative deforestation for the whole region is predicted with very high accuracy. Actual deforestation in 2014 was 382.6 km^2, whereas predicted deforestation for that year is 382.9 km^2. The accuracy at the cell level seems fair, but some obvious limitations of the model emerge. Firstly, the model cannot identify new deforestation events (events that appear in areas previously untouched by deforestation). This is due to the fact that deforestation in adjacent cells is one of the key explanatory variables. If no deforestation happened in previous years in the surrounding areas, the model predicts zero or close to zero deforestation. Secondly, the model cannot foresee occasions when forest clearings suddenly cease to continue. Again, this is because the model heavily relies on the hypothesis of contagion. That is, if deforestation happened in previous year, it is likely to continue this year. However, those two kinds of situations are rare. Lastly, the model underpredicts the extent of deforestation in the hotspots. This is best seen by analyzing Figures 15B and 16B: the cells with the largest positive errors spatially correspond to the hotspot cells. Model's ability to identify hotspots is analyzed next. This is done by computing two metrics: matching rate (the ratio between the count of correctly identified hotspots and the count of actual hotspots) and commission rate (the ratio between the count of incorrectly identified hotspots and the count of actual non-hotspots). Figure 17 graphs both metrics for different thresholds.

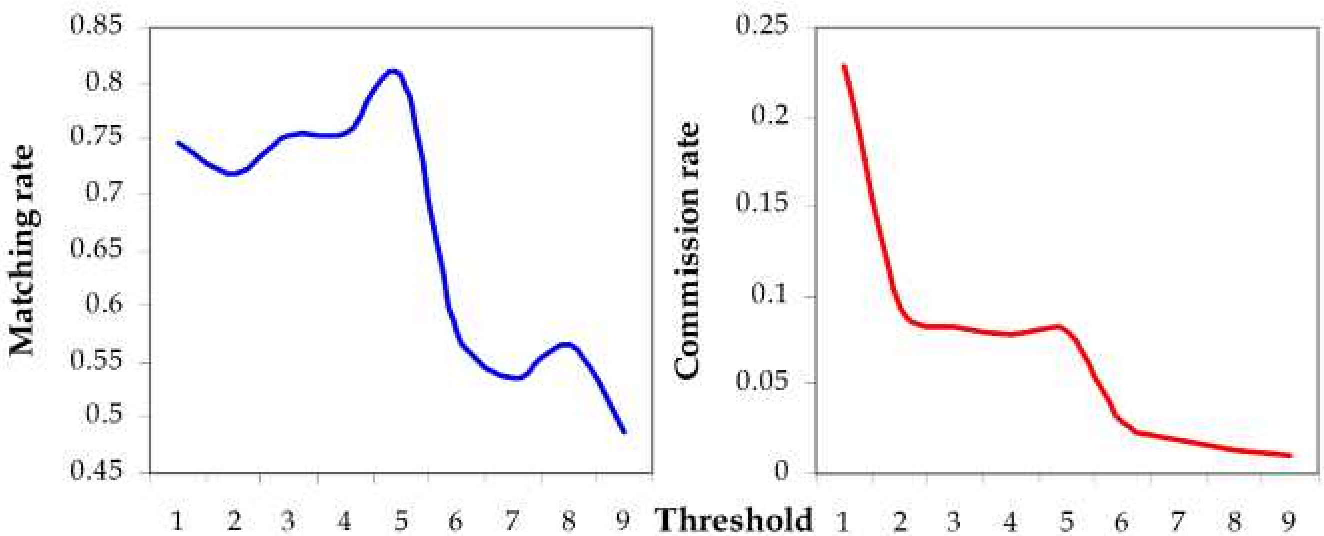

Figure 17. Matching and commission rates for different thresholds

A threshold is a cut-off mark, which is used to classify all land parcels into deforestation hotspots and non-hotspots. For instance, if a threshold is set at 5%, all cells with deforestation higher than 5% are treated as hotspots and the rest of the cells are identified as non-hotspots. Figure 17 reveals that the matching rates fall within the interval from 0.7 to 0.8 for thresholds from 1 to 5. The highest matching rate, exceeding 0.8, is reached at threshold equal to 5. However, if a hotspot cell is defined as facing at least 6% of annual deforestation, matching rate is substantially lower, below 0.6. If a threshold is set at 9%, less than half of deforestation hotspots are correctly identified by the model. Commission rate is very high (> 0.2) for threshold equal to 1. The percentage of incorrectly identified hotspots is much lower at thresholds from 2 to 5, ranging between 7 and 9 percent. In case a threshold is set at 6 to 9 percent, commission percentage is only 1 to 2 percent. The latter result implies that for higher thresholds the model is unlikely to predict large scale deforestation where it does not happen.

6.5. Conclusions

To explain deforestation, the following set of covariates is needed: 1) covariates that reflect spatiotemporal contagion of deforestation, 2) additional factors, consisting of altitude, presence of legal protection (indigenous governance and strict protection) and proximities to the nearest official road and city, 3) eigenvectors that capture different spatial autocorrelation levels. Residual analysis showed that eigenvector-based spatial filtering nullified spatial autocorrelation. What is key, however, is that the linear model that additionally includes eigenvectors as covariates is able to explain a substantially larger share of variation in deforestation. Additional generalization is gained by calibrating an artificial neural network that uses the observed covariates and eigenvectors as inputs. The output values of trained network fairly correlate with actual deforestation (Pearson correlation coefficient is 0.79). Those simulated values portray very well the general patterns of actual deforestation. As a result, prediction accuracy would be greater at coarser scales. The model could be applied to identify deforestation hotspots. If cells with at least 5% of annual deforestation are defined as hotspots, the model is able to correctly identify 80% of actual cases, while commission error is a satisfactory 8%. Thus, provided that spatiotemporal contagion is taken into consideration and artificial covariates are created to capture spatial autocorrelation patterns deforestation can be mapped at 5x5 km grids with fair precision. Notwithstanding, spatial filtering is computationally expensive. The time required to filter eigenvectors exponentially increases as more observations are added. This poses a challenge if the intension is to predict deforestation for large areas. However, the problem can be circumvented by partitioning large territories into a number of tiles and produce eigenvectors separately for each tile. This would allow mapping deforestation in large areas within a reasonable timeframe. The model is an important step towards predicting future deforestation more accurately. However, in its current form it cannot be successfully applied for mapping future deforestation patterns. Even though model's observed predictors reflect the past or are time-fixed, unobserved processes that are captured by eigenvectors are often dynamic. As a result, outdated set of eigenvectors becomes redundant in the model that aims to map future deforestation patterns.

6.6. References

Aguiar, A.P.D., Câmara, G., & Escada, M.I.S. (2007). Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. *Ecological Modelling, 209*(2-4), 169-188.

Alves, D.S. (2002). Space-time dynamics of deforestation in Brazilian Amazônia. *International Journal of remote sensing, 23*(14), 2903-2908.

Arai, E., Shimabukuro, Y.E., Pereira, G., & Vijaykumar, N.L. (2011). A Multi-Resolution Multi-Temporal Technique for Detecting and Mapping Deforestation in the Brazilian Amazon Rainforest. *Remote Sensing, 3*(9), 1943-1956.

Arima, E.Y., Simmons, C.S., Walker, R.T., & Cochrane, M.A. (2007). Fire in the Brazilian Amazon: a Spatially Explicit Model for Policy Impact Analysis. *Journal of Regional Science, 47*(3), 541-567.

Asner, G.P., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Knapp, D.E., & Silva, J.N.M. (2006). Condition and fate of logged forests in the Brazilian Amazon. *Proceedings of the National Academy of Sciences, 103*(34), 12947-12950.

Assunção, J., Gandour, C., & Rocha, R. (2013). *DETERring Deforestation in the Brazilian Amazon: Environmental Monitoring and Law Enforcement*. *Climate policy initiative*. Rio de Janeiro, Brazil: Climate Policy Institute.

Cochrane, M.A., & Laurance, W.F. (2002). Fire as a large-scale edge effect in Amazonian forests. *Journal of Tropical Ecology, 18*(3), 311-325.

Diniz, C.G., Souza, A.A.A., Santos, D.C., Dias, M.C., da Luz, N.C., Moraes, D.R.F., ... & Adami, M. (2015). DETER-B: The New Amazon Near Real-Time Deforestation Detection System. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8*(7), 3619-3628.

Espindola, G.M., Aguiar, A.P.D., Pebesma, E., Câmara, G., & Fonseca, L. (2011). Agricultural land use dynamics in the Brazilian Amazon based on remote sensing and census data. *Applied Geography, 32,* 240-252.

GEOFABRIK (2016). *OpenStreetMap Data Extracts*. Retrieved from http://download.geofabrik.de/

Griffith, D.A. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems, 2*(2), 141-156.

Griffith, D.A. (2013). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization.* Springer Science & Business Media.

Griffith, D.A., & Peres-Neto, P.R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology, 87*(10), 2603-2613.

IBGE (Brazil's Institute of Geography and Statistics) (2015). *Cadastro de Localidades*. Retrieved from http://www.ibge.gov.br/home/geociencias/cartografia/territ_localidades.shtm

IIASA (International Institute for Applied System Analysis) (2016). *Global Agro-Ecological Zones dataset. Plate 28: Climate, soil and terrain slope constraints combined*. Retrieved from http://webarchive.iiasa.bac.at/Research/LUC/GAEZ/

INPE (Brazil's National Institute for Space Research) (2016a). *Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satélite*. Retrieved from http://www.obt.inpe.br/prodes/index.php

INPE (Brazil's National Institute for Space Research) (2016b). *Monitoramento de Queimadas*. Retrieved from http://www.inpe.br/queimadas/

Jacobson, A., Dhanota, J., Godfrey, J., Jacobson, H., Rossman, Z., Stanish, A., … & Riggio, J. (2015). A novel approach to mapping land conversion using Google Earth with an application to East Africa. *Environmental Modelling & Software, 72,* 1-9.

Kehl, T.N., Todt, V., Veronez, M.R., & Cazella, S.C. (2012). Amazon Rainforest Deforestation Daily Detection Tool Using Artificial Neural Networks and Satellite Images. *Sustainability, 4,* 2566-2573.

Kirby, K.R., Laurance, W.F., Albernaz, A.K.M., Schroth, G., Fearnside, P.M., Bergen, S., … & da Costa, C. (2006). The future of deforestation in the Brazilian Amazon. *Futures, 38*(4), 432-453.

Lapola, D.M., Schaldach, R., Alcamo, J., Bondeau, A., Msangi, S., Priess, J.A., … & Soares-Filho, B.S. (2011). Impacts of Climate Change and the End of Deforestation on Land Use in the Brazilian Legal Amazon. *Earth Interactions, 15*(16), 1-29.

Laurance, W.F., Albernaz, A.K.M., Schroth, G., Fearnside, P.M., Bergen, S., Venticinque, E.M., … & da Costa, C. (2002). Predictors of deforestation in the Brazilian Amazon. *Journal of Biogeography, 29*, 737-748.

Laurance, W.F., Cochrane, M.A., Bergen, S., Fearnside, P.M., Delamônica, P., Barber, C., ... & Fernandes, T. (2001). The Future of the Brazilian Amazon. *Science, 291*, 438-439.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics, 2*(2), 164-168.

Maeda, E.E., Formaggio, A.R., Shimabukuro, Y.E., Arcoverde, G.F.B., & Hansen, M.C. (2009). Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation, 11*(4), 265-272.

Marquardt, D.W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics, 11*(2), 431-441.

Mas, J.F., Puig, H., Palacio, J.L., & Sosa-López, A. (2004). Modelling deforestation using GIS and artificial neural networks. *Environmental Modelling & Software, 19*(5), 461-471.

Moreira, E., Costa, S., Aguiar, A.P.D., Câmara, G., & Carneiro, T. (2009). Dynamical coupling of multiscale land change models. *Landscape Ecology, 24,* 1183-1194.

Morton, D.C., DeFries, R.S., Shimabukuro, Y.E., Anderson, L.O., Espírito-Santo, D.B.F., Hansen, M., … & Carroll, M. (2005). Rapid Assessment of Annual Deforestation in the Brazilian Amazon Using MODIS Data. *Earth Interactions, 9*(8), 1-22.

Ripley, B.D. (2008). *Pattern Recognition and Neural Networks.* Cambridge, U.K.: Cambridge University Press.

Robalino, J.A., & Pfaff, A. (2012). Contagious development: Neighbor interactions in deforestation. *Journal of Development Economics, 97*(2), 427-436.

Rosa, I.M.D., Purves, D., Souza Jr., C., & Ewers, R.M. (2013). Predictive Modelling of Contagious Deforestation in the Brazilian Amazon. *PLOS ONE, 8*(10).

Soares-Filho, B.S., Nepstad, D.C., Curran, L.M., Cerqueira, G.C., Garcia, R.A., Ramos, C.A., … & Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature, 440,* 520-523.

SRTM (Shuttle Radar Topography Mission) (2016). *90m Digital Elevation Model, version 2.1*. Retrieved from http://dds.cr.usgs.gov/srtm/version2_1/SRTM3/South_America/

Tiefelsdorf, M., & Griffith, D.A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A, 39*(5), 1193-1221.

Wassenaar, T., Gerber, P., Verburg, P.H., Rosales, M., Ibrahim, M., & Steinfeld, H. (2007). Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. *Global Environmental Change, 17*(1), 86-104.

WDPA (World Database of Protected Areas) (2016). *WDPA dataset*. Retrieved from http://www.wdpa .org/

World-Clim (2016). *Data for current conditions (~1950-2000)*. Retrieved from http://www.worldclim. org/current

WorldPop (2015). *WorldPop Americas dataset*. Retrieved from http://www.worldpop.org.uk/

Yu, H., & Wilamowski, B.M. (2011). In *Levenberg-Marquardt Training*. *Chapter 12, Intelligent Systems*. Boca Raton, Florida: CRS press.

Zhan, X., Sohlberg, R.A., Townshend, J.R.G., DiMiceli, C., Carroll, M.L., Eastman, J.C., … & DeFries, R.S. (2002). Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment, 83*, 336-350.

Chapter 7

Conclusions

This dissertation answered the following questions: what are the main drivers of deforestation in Legal Amazon and how the results depend on space, does the choice of travel time instead of Euclidean distances lead to different findings of GWR, does sugarcane expansion in southern Brazil is related to deforestation in the Brazilian Amazon, is rural population linked with deforestation on forest frontiers in Pará, to what extent conservation units protect forests from deforestation in Pará and how space influences the results, is buffer zone management effective in protecting the edges of Pará's conservation units, can deforestation be accurately mapped using only past and time-fixed variables.

Commonly, the researchers rely on various packages, where GWR is implemented, thus automatically accepting their limitations. Geographical (Euclidean) distances are accepted as appropriate representations of access between locations. However, for the case of Legal Amazon, distance approximation to Euclidean is very crude due to substantial differences in road quality, spatial arrangement of the road network, and absence of roads in certain regions. A comparative analysis in chapter 2 confirmed that the implementation of travel time distances into GWR notably alters the findings for some regions. For example, it was shown that the model based on Euclidean distances significantly overestimates the impact of cattle market on deforestation in Roraima and the impact of unofficial roads in Pará and Amapá, and underestimates the effect that unofficial roads have on deforestation in the region located to the south of the Brazilian port of Manaus. Therefore, using economic (travel time) distances was an important step towards obtaining more accurate results.

Space is important in deforestation modeling. Chapter 2 revealed that deforestation determinants in Legal Amazon are location-specific. It was concluded that cattle ranching contributes to deforestation most evidently in Pará (especially, eastern side) and its surroundings. Those regions lie on a deforestation frontier, where professional and well-capitalized ranchers reside. Livestock farms in eastern Pará are located in highly urbanized areas with dense road network, thereby facilitating the transportation of beef and other cattle products to trade centers. Those favorable economic conditions make cattle business profitable. Furthermore, chapter 4 concluded that cattle ranching is an important economic activity of rural migrants on forest edges in Pará. The findings indicate that deforestation during period 2010-2014 was 0.7% on forest edges with fewer cattle animals (less than 130 per 25 km^2), 5.2% in parcels with middle-sized cattle herds (from 130 to 847 animals), and 8.5% in parcels with relatively large cattle herds (more than 847 animals). Also, cattle farming was found to be the most powerful discriminator between high and low deforestation areas. On the contrary, forests in remote regions, such as Amazonas and Roraima, are not subjected to high deforestation pressure from cattle business partially due to the fact that poor accessibility and long distances to trade centers negatively affect its profitability. Those economic constraints also apply to crop cultivation. However, additionally, it is constrained by climatic (excessive rainfall) and terrain (steep slopes) characteristics. As a result, crop cultivation was found to induce deforestation only in a certain region, located in southeastern Pará and northeastern Mato Grosso, where average annual rainfall is generally lower than 2000 mm. The results also suggest that deforestation is negatively linked with precipitation in most parts of Pará, thereby confirming that rainfall shapes the patterns of deforestation.

GWR results suggest that crop cultivation is not linked with deforestation in most parts of Mato Grosso, where soybean and sugarcane (the two most commonly cultivated crops in Brazil) are widely grown. However, it would be incautious to conclude that crop cultivation here is decoupled from deforestation. This is because cause and effect may be separated in space. Specifically, crop growing in non-frontier Brazilian states may stimulate deforestation is Legal Amazon. Since crops almost exclusively expand over past grazing areas, new crop fields often appear without forest clearing. However, livestock farmers who sell their pastures to crop planters may migrate to frontier regions, where cattle production is reconstituted. Chapter 3 empirically investigated this indirect land use change, but with the emphasis on sugarcane (ILUC associated with soybean expansion is already well researched). The findings revealed that sugarcane was not directly linked with deforestation in Legal Amazon during 2002-2012. The adverse effect of sugarcane expansion in southern Brazil on deforestation manifested indirectly through displaced cattle ranchers. Specifically, expanding sugarcane plantations in southern Brazil was indirectly associated with 12.2% of Legal Amazon's deforestation during 2002-2012, which constitutes 16.3 thousand km^2 of forests. These figures translate into 189.4 Mg of carbon emissions. Moreover, the production of every one billion liters of ethanol from sugarcane in Brazil led to 17.5 Mg loss of carbon stock. For comparison, indirect contribution of soy industry to deforestation was found to be 17.6 %, or 23.6 thousand km^2 of forests. The findings do not reveal whether ILUC associated with crop expansion continues into this day, but if it does, then policies aimed at mitigating deforestation from ILUC must be integral. That is, both sugarcane and soy expansion must be addressed.

Chapter 2 revealed that deforestation is not linked with urban population. The finding most likely underlies the importance of international markets. However, these aspects are outside the scope of this dissertation. Generally, the results suggest that the size of rural communities *per se* is linked with deforestation, but the relationship is very complex. GWR identified significant spatial differences across the regions. Rural communities have the strongest direct impact on extant forests in western parts of Rondônia and Mato Grosso. Interestingly, the results indicate that rural population is not a significant factor is explaining deforestation in northern Pará, Amapá, Maranhão and regions with sparse vegetation in Rondônia and Mato Grosso. However, it should not be concluded that rural population is not linked with deforestation in those areas due to the fact that rural migrants are known to engage in expansive agriculture on forest edges. Forest edges and long-settled rural regions are very different environments. The latter is covered by extensive cash-oriented pastures and crop fields. Here the size of rural communities may not be linked or even negatively linked with deforestation, because cattle ranches are usually larger than subsistence-oriented farms. Economic agents responsible for deforestation on forest edges are newly-arrived rural settlers. Forest edges attract rural migrants from other rural areas. This migration is motivated by land scarcity in old rural settlements and land insecurity on the frontier. As discussed by Carr & Burgdorfer (2013), those migrants get involved in expansive agriculture due to cheap family labor, low transportation costs and lax legal enforcement. Chapter 4 provides empirical evidence supporting this view. A parametric approach (fractional logistic regression) confirms a positive link between rural population and deforestation. However, causal claims cannot be made, because rural population was treated exogenously. A non-parametric method (regression tree) indicates that larger rural population is linked with more deforestation in land parcels with sizeable cattle herds, relatively sparse vegetation and favorable levels of precipitation (less than ~2100 mm).

Migrant farmers tend to sell their land properties to larger producers, as cattle business is too risky for smallholders. Indeed, financial aspect is very important in explaining deforestation, as can be learned from chapter 2. The relationship between deforestation and GDP per capita follows a U-shaped curve, meaning that increasing incomes curb deforestation at lower levels of GDP per capita and boost it at higher levels. The result is generally applicable to all locations in Legal Amazon, as little spatial variation was observed. Financial resources turn into an accelerant of deforestation only if they are sufficient, because agriculture has to be large-scale to be profitable. As economies grow, urban citizens increase the consumption of meat products, and rural dwellers receive increasingly greater remittances from migrant family members. Therefore, growing levels of income may become a concern in the future. Financial resources for agriculture are often obtained in a form of rural credit. Therefore, to restrain deforestation, the Brazilian government restricts the access to rural credit for land properties located in high deforestation municipalities. Rural credit restrictions proved to be effective in curbing deforestation in Pará (chapter 2). This is especially true for northeastern Pará, characterized by extensive cattle ranching and relatively dense forests. This finding is a strong argument in favor of GWR approach: if spatial differences are ignored, rural credit scheme is concluded to be an inefficient policy.

Chapter 2 further revealed that the road network shapes deforestation patterns across Legal Amazon. Accessibility is a necessary prerequisite for deforestation. Roads are necessary to access logging spots and to transport harvested woods. As those spots become more easily accessed, squatters invade the lands and get involved in extensive deforestation. As lands become more developed, they are purchased by large producers and further deforested. Therefore, government plans to build or pave a road always provoke resistance from environmentalists. Fast road connection is important for Brazil's economy. However, roads should be paved only if mechanisms that ensure forest integrity alongside planned roads are well prepared and could be effectively implemented. Adverse effects of road network on forests shall not be underestimated. Newly paved roads commonly spur large networks of endogenous (unofficial) roads deep into the forests, thus exposing vast forest areas to potential deforestation. Chapter 2 found that both official and unofficial roads are contributors to deforestation. A general observation is that a road network leads to much more deforestation in remote regions (Amazonias and Roraima) as opposed to deforestation frontiers. This finding does not indicate that roads on forest frontiers are weaker contributors to deforestation, but instead it reflects the fact that most forest frontier roads are old and, therefore, their contribution to deforestation took place in the past until areas alongside those roads became deforested.

The protected area system is among policy measures applied by the Brazilian government to mitigate deforestation. The effect of legal protection on deforestation was researched from few different angles in this dissertation. Global municipality level analysis presented in chapter 2 did not find any connection between legal protection and deforestation, but the links emerged under the local approach. It was concluded that in Pará and nearby regions less deforestation happens where more forests are declared as sustainable use areas and indigenous lands. The background reason for this finding, however, cannot be revealed by the model. It could be that the negative link simply reflects the fact that many protected areas are established far from deforestation hotspots (precautionary protected areas). Alternatively, it could mean that legal protection is effective in curbing deforestation. The results also found no link between strict protection and deforestation. This counterintuitive finding underlies the caveats of analyzing the effect of legal protection on deforestation at municipality scale. Strictly protected forests may avoid deforestation completely, whereas areas nearby, located within the borders of the same municipality, may be subjected to extensive deforestation. In this way the effectiveness of strict legal protection cannot be seen in municipality level analyses. Chapter 4 presents a finer scale grid level analysis and reveals that deforestation is threefold lower within protected areas as compared to unprotected forests (the conclusion holds for parcels with middle-sized cattle herds and dense vegetation). However, even this finding cannot be taken as evidence that legal protection inhibits deforestation, because characteristics inside and outside protected areas are often dissimilar. Therefore, a separate study in chapter 5 was devoted to isolate the protection effect by matching.

Chapter 5 quantified avoided deforestation due to legal protection in Pará state. The results revealed that 0.72% of forests (~2900 km^2) avoided deforestation during period 2000-2004 inclusive. If displaced deforestation in buffer zones is ignored, avoided deforestation is overestimated by 60%. The most forest clearings were avoided in conservation units established until 1990, governed by a ministry or an agency, classified by the IUCN into categories Ia, Ib or II, and small in territory. However, the finding that indigenous lands saved less forests than protected areas managed by the governmental institutions is not backed by Nolte et al. (2013), who conclude the opposite. For the case of Pará, it is intuitive that indigenous lands avoid less deforestation, because those are located further away from deforestation hotspots and have relatively large areas. Thus, low avoidance of deforestation is a consequence of low deforestation pressure and challenging management (territory per environmental inspector is large).

The overall figure of avoided deforestation in Pará seems low. However, it is important to realize that large areas of protected lands lie in remote territories where deforestation pressure is low and, therefore, little deforestation can be avoided. The boundaries of some protected areas, located on deforestation frontiers, can be discerned clearly by looking at the maps that show forest cover. For instance, such conservation units are Parakanã and Mãe Maria, where avoided deforestation was computed to be 8.2% and 4.8% respectively during 2000-2004. Other conservation units with high percentage of avoided deforestation include Arara and Tapajós (9.5% and 6.1% of saved forests respectively). For the case of Arara, under no legal protection its forests would become completely deforested in slightly more than 50 years if conditions affecting deforestation remained unchanged. In general, protected areas that avoided the most deforestation are located in close proximities to deforestation hotspots: alongside the Trans-Amazonian highway, on the banks of Amazon River, and in the central part of eastern Pará. The majority of conservation units where avoided deforestation is a statistical zero are located in remote areas or areas surrounded by natural obstacles (northwestern corner of Pará and archipelago of Marajó). The protection of those lands may become important in the future if deforestation pressure builds. Avoided deforestation is negative in Alto Rio Guamá and Itacaiunas. This may be a consequence of protests by local communities who are ceased unlimited access to forest resources that were available prior to the establishment of a protected area.

Protecting the edges of conservation units and their surroundings is a challenging task. Since the establishment of a protected area implies that the local communities no longer can exploit forest resources without restrictions in that area, deforestation may be displaced to the surrounding areas. Buffer zones are meant to protect park edges from deforestation and to encourage local communities to participate in the preservation of conservation units. The findings suggest that buffer zones around protected areas located in eastern Pará experienced excessive deforestation relative to no legal protection. Deforestation in eastern Pará is intensive and forest cover is declining. Potential deforestation spots are substantially narrowed by the establishment of protected areas. Deforestation pressure is therefore transferred to unprotected forests. However, the result is different where deforestation is not so intensive (to the west of highway BR-163 and on the banks of Amazon River). Here buffer zones were able to absorb the pressure from deforestation displacement and even avoid deforestation relative to counterfactual territories.

Chapter 5 found no evidence of edge effects in Pará. Moreover, in some instances park edges avoided more deforestation than nuclear areas (the difference proved to be statistically significant). Those conservation units are located near deforestation hotspots. A possible explanation of this seemingly counterintuitive finding is that park edges face higher deforestation pressure than nuclear areas and, as a result, avoid more deforestation. Forest edges are directly exposed to unprotected lands, and this may be a source of excessive pressure on park edges relative to the cores. However, edge effects were tested assuming equal deforestation pressure both in the nuclear of each conservation unit and on its edge. If park edges are subjected to significantly higher deforestation pressure than the nuclear areas, this could invalidate the results. Some conservation units situated in eastern Pará can be discerned in Google Maps. Visual inspection of the map suggests that forests on park edges and in core areas are equally intact. This observation backs the conclusion that edge effects are not present in Pará.

The research framework in chapter 5 has some other caveats. To start with, deforestation displacement is assumed to occur only within 10 kilometers from protected areas' borders. Some loggers may migrate longer distances, where forest resources are abundant or economic conditions are more favorable. Further, the methodology in chapter 5 has its shortcomings. One problem is that not all factors affecting deforestation can be observed at fine scales. Therefore, even if a control cell with identical observed characteristics is found, the differences in deforestation between treatment and control cells may be due to the differences in unobserved factors rather than legal protection. The other problem is that propensity score matching cannot be done under the lack of relevant control observations. This is the reason why this research considered only period 2000-2004. In subsequent years vast previously unprotected forests became legally protected. If up-to-date matching study was attempted, most forests in Pará would fall under the treatment group and most deforested areas would be in the control group. Matching then is not balanced, because no pairs of treatment and control observations with similar characteristics can be found. Solving this problem is left for future.

Chapter 2 revealed that deforestation is linked with deforestation in the neighboring counties (neighborhood effect was captured by the autocovariate). Thus, deforestation is contagious in space. What is more, it is well-known that past deforestation to a certain extent explains deforestation patterns at current times. This spatiotemporal contagion can be used for prediction. Chapter 6 raises and tests a hypothesis that deforestation can be relatively precisely mapped exclusively from past and time-fixed data. An artificial neural network was trained to learn patterns of current deforestation at regular 5x5 km grids from past dynamic processes, factors that do not change over time, and eigenvectors that partially capture unobserved covariates. The correlation coefficient between simulated and actual deforestation is 0.79. If all grids with deforestation higher than 5% are defined as hotspots, the model is able to identify 80% of actual cases with commission error equal to 7.8%. Even though the results are promising, predicting next period deforestation remains challenging. This is because unobserved covariates that are replaced by eigenvectors must be assumed to be static to be able to predict deforestation in the following year. However, most unobserved processes affecting deforestation are dynamic. As a result, the initial eigenvectors become redundant if the model is used to foresee deforestation, thereby reducing the prediction accuracy significantly.

Economic actors involved in deforestation can be separated into large scale producers and small scale farmers. The influence on deforestation by large scale commercial farmers manifests through extensive agriculture (mostly cattle farming and soybean cultivation). To limit deforestation due to expansion of pastures and soy fields, NGOs initiated few important agreements, such as Soy Moratorium and Beef Moratorium. Gibbs et al. (2015a) and Gibbs et al. (2015b) showed that those measures are effective in curbing deforestation, but have some limitations (see chapter 1). Further enforcement and continuation of those initiatives is important. Sugarcane cultivation also contributes to deforestation, but indirectly through displaced cattle ranchers. Therefore, policies that aim to curb deforestation resulting from pasture expansion should also be effective in reducing indirect land use change associated with sugarcane (and soy) expansion. Also, adverse environmental effects associated with sugarcane cultivation (deforestation and consequential contribution to CO_2 emissions) may be included in the price of biofuels.

Small scale farmers occasionally migrate to forest edges where they initiate deforestation. The protected area system may be an effective hedge against such colonization. It is especially important to protect the perimeter of protected lands by ensuring effective buffer zone management, which becomes increasingly important as deforestation approaches the borders of protected areas. More efforts and financial resources should be allocated to protect the perimeters of conservation units located in high deforestation regions, such as eastern Pará. Also, it is strongly advisable to postpone the paving of roads that stretch through intact and remote Amazonian forests until measures that can effectively prevent smallholder colonization and subsequent deforestation alongside newly paved roads are in place.

7.1. References

Carr, D., & Burgdorfer, J. (2013). Deforestation Drivers: Population, Migration, and Tropical Land Use. *Environment*, *55*(1).

Gibbs, H.K., Munger, J., L'Roe, J., Barreto, P., Pereira, R., Christie, M., … & Walker, N.F. (2015a). Did Ranchers and Slaughterhouses Respond to Zero-Deforestation Agreements in the Brazilian Amazon? *Conservation letters.* doi: 10.1111/conl.12175

Gibbs, H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., … & Walker, N.F. (2015b). Brazil's Soy Moratorium. Supply-chain governance is needed to avoid deforestation. *Science, 374,* 377-378.

Nolte, C., Agrawal, A., Silvius, K.M., & Soares-Filho, B.S. (2013). Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. *Proceedings of the National Academy of Sciences, 110*(13), 4956-4961.

Chapter 8

Appendices

8.1. Appendix A

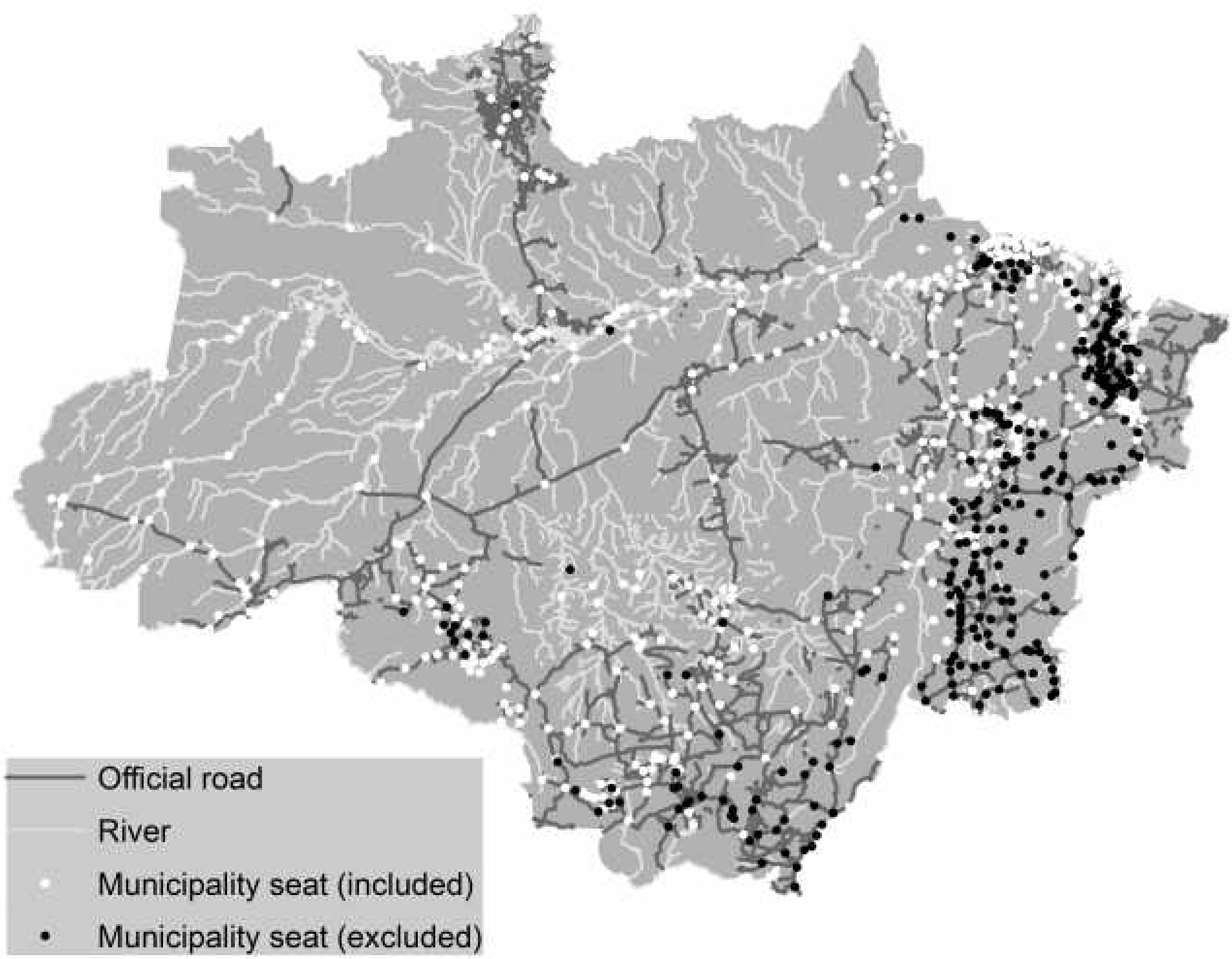

Figure A1. Road and river network in Legal Amazon. Road and river shapefiles were downloaded from GEOFABRIK in 2014. White dots indicate seats of municipalities that were included in the analysis, whereas black dots mark the rest of municipality seats in Legal Amazon.

8.2. Appendix B

Table B1. Descriptive statistics for municipalities

Variable	Mean	Standard deviation	Minimum	Maximum
DEF	12.51	29.05	0	353.7
POPURB	24570.23	98526.72	836	1792881
POPRUR	10565.43	11548.68	195	125336
GDP	9971.64	9166.79	2269.82	103404
FCOVER	38.35	27.48	5	98.74
ELEV	178.06	123.19	5.32	760.94
CATTLE	121573.7	176922.8	15	2022366
CROP	17874.78	69311.28	7	875839
TIMBER	4789.89	13591.55	1	162906
ROF	190.78	227.58	0	1733.31
RUNF	683.49	1143.53	0	10607.14
CREDIT	760.83	1571.31	0	15970.07
TENURE	89.61	13.15	18.57	100
STRICT	3.55	9.96	0	71.87
SUST	13.46	25.61	0	100
INDIG	10.38	19.45	0	99.55
PREC	1861.16	393.77	902.02	3250.17
TERR	9027.19	16477.56	66.28	159533.3
POPURBLG	18772.53	78281.08	518	1396768
POPRURLG	9761.46	10136.72	458	80139
GDPLG	9216.02	9868.97	1946.36	119560.1

Variable abbreviations are introduced in Table 1

8.3. Appendix C

Table C1. Correlation matrix for municipalities

	DEF	POPURB	POPRUR	GDP	FCOVER	ELEV	CATTLE	CROP	TIMBER	ROF	RUNF	CREDIT	TENURE	STRICT	SUST	INDIG	PREC	TERR
DEF	1																	
POPURB	0.066	1																
POPRUR	0.345	0.215	1															
GDP	-0.07	0.135	-0.177	1														
FCOVER	0.247	0.046	0.097	-0.011	1													
ELEV	0.007	-0.079	-0.332	0.501	-0.064	1												
CATTLE	0.497	0.041	0.084	0.189	-0.002	0.364	1											
CROP	-0.017	0.011	-0.063	0.523	-0.068	0.358	0.014	1										
TIMBER	0.330	0.022	0.165	0.007	0.130	-0.022	0.187	0.006	1									
ROF	0.450	0.270	0.178	0.307	0.090	0.320	0.493	0.318	0.171	1								
RUNF	0.545	0.091	0.107	0.298	0.147	0.351	0.732	0.294	0.361	0.664	1							
CREDIT	-0.075	-0.072	-0.248	0.629	-0.165	0.514	0.158	0.604	-0.023	0.225	0.328	1						
TENURE	0.033	-0.084	0.152	-0.304	0.078	-0.342	-0.126	-0.284	-0.021	-0.173	-0.235	-0.368	1					
STRICT	0.064	0.070	-0.053	0.096	0.305	0.028	0.030	-0.047	-0.027	0.090	0.014	-0.071	0.004	1				
SUST	-0.035	0.046	0.122	-0.154	0.129	-0.377	-0.197	-0.107	-0.044	-0.119	-0.166	-0.186	0.085	0.041	1			
INDIG	0.105	-0.046	-0.047	0.053	0.334	0.243	0.062	0.019	-0.045	0.158	0.102	-0.019	-0.064	0.152	-0.154	1		
PREC	0.025	-0.024	0.060	-0.112	0.414	-0.345	-0.196	-0.115	0.074	-0.058	-0.065	-0.198	0.122	0.157	0.092	0.149	1	
TERR	0.388	0.039	0.134	-0.002	0.471	0.033	0.200	0.012	0.107	0.373	0.339	-0.058	0.002	0.323	0.041	0.354	0.296	1

Variable abbreviations are introduced in Table 1

8.4. Appendix D

Formulae of Moran's I in terms of matrix algebra are presented below. The notation is as follows: MI – Moran's I, n – number of observations, I – identity matrix, j – vector of ones, £ – vector of model errors, W – matrix of weights (inverse distances). Symbol '$\circ$' in formulas D6, D7 and D9 denotes element by element multiplication.

$$MI = \frac{n}{S_0} \times \frac{£^T \left(I - \frac{jj^T}{n} \right) W \left(I - \frac{jj^T}{n} \right) £}{S_1} \quad \text{(D1)}$$

$$\operatorname{var}(MI) = \frac{n\left(S_2\left(n^2-3n+3\right)-nS_3+3\left(S_0\right)^2\right)-S_5\left(S_2\left(n^2-n\right)-2nS_3+6\left(S_0\right)^2\right)}{(n-1)(n-2)(n-3)\left(S_0\right)^2}-\left(\frac{1}{n-1}\right)^2 \quad \text{(D2)}$$

$$E(MI) = -\frac{1}{n-1} \quad \text{(D3)}$$

$$S_0 = j^T W j \quad \text{(D4)}$$

$$S_1 = £^T \left(I - \frac{jj^T}{n} \right) £ \quad \text{(D5)}$$

$$S_2 = \frac{1}{2} \times j^T \left(\left(W + W^T \right) \circ \left(W + W^T \right) \right) j \quad \text{(D6)}$$

$$S_3 = j^T \left(\left(\left(j^T W \right)^T + Wj \right) \circ \left(\left(j^T W \right)^T + Wj \right) \right) \quad \text{(D7)}$$

$$S_4 = £ - \frac{j£^T j}{n} \quad \text{(D8)}$$

$$S_5 = \frac{n}{\left(S_1\right)^2} \times \left(S_4 \circ S_4 \right)^T \left(S_4 \circ S_4 \right) \quad \text{(D9)}$$

8.5. Appendix E

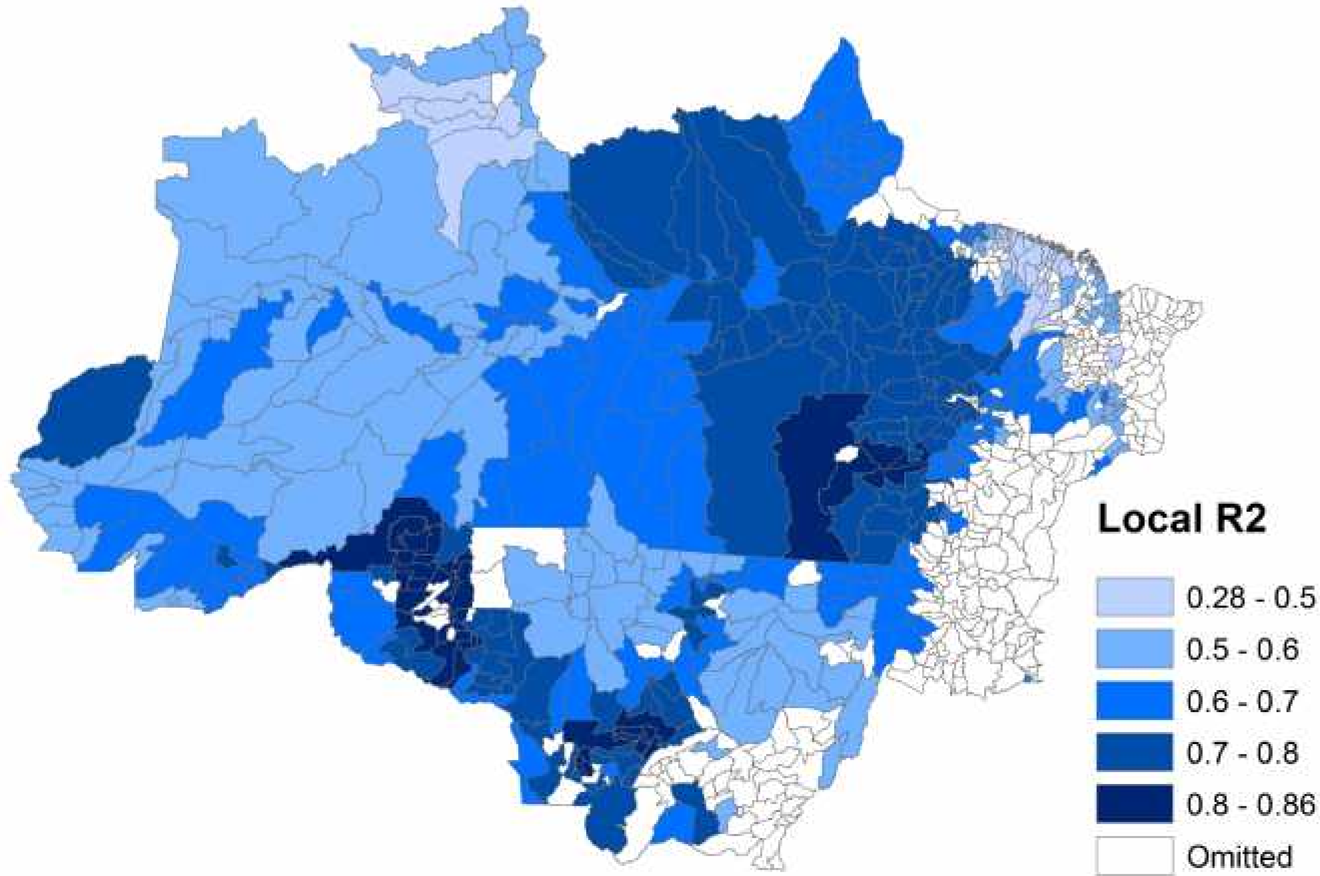

Figure E1. Local determination coefficients. This 2010 municipality boundary map was downloaded from the IBGE's website.

8.6. Appendix F

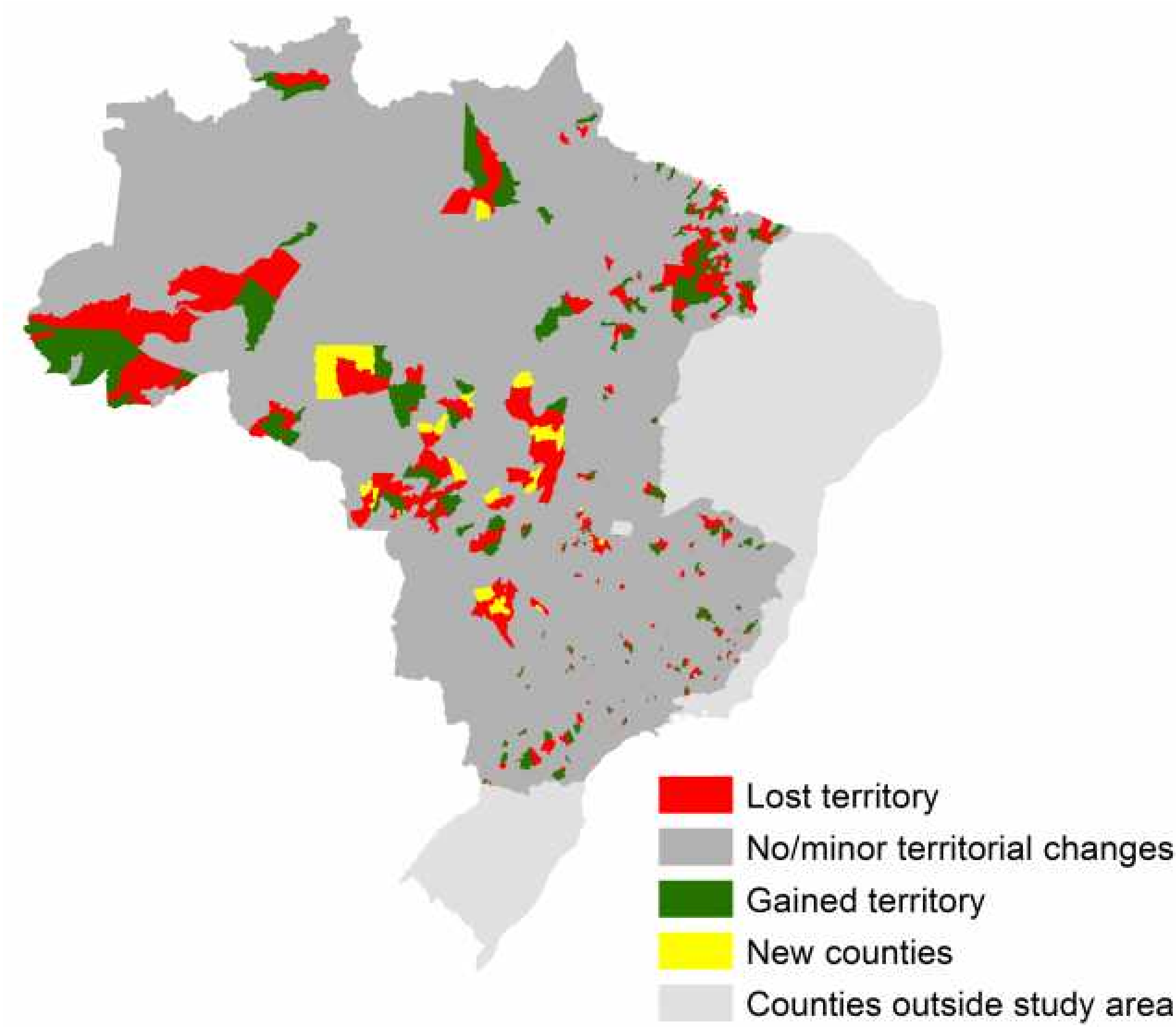

Figure F1. Territorial changes of Brazilian counties between 2000 and 2014. This municipality boundary map was downloaded from the IBGE's website. Municipality boundaries are not shown to improve visualization. Projection: Albers Equal Area Conic.

8.7. Appendix G

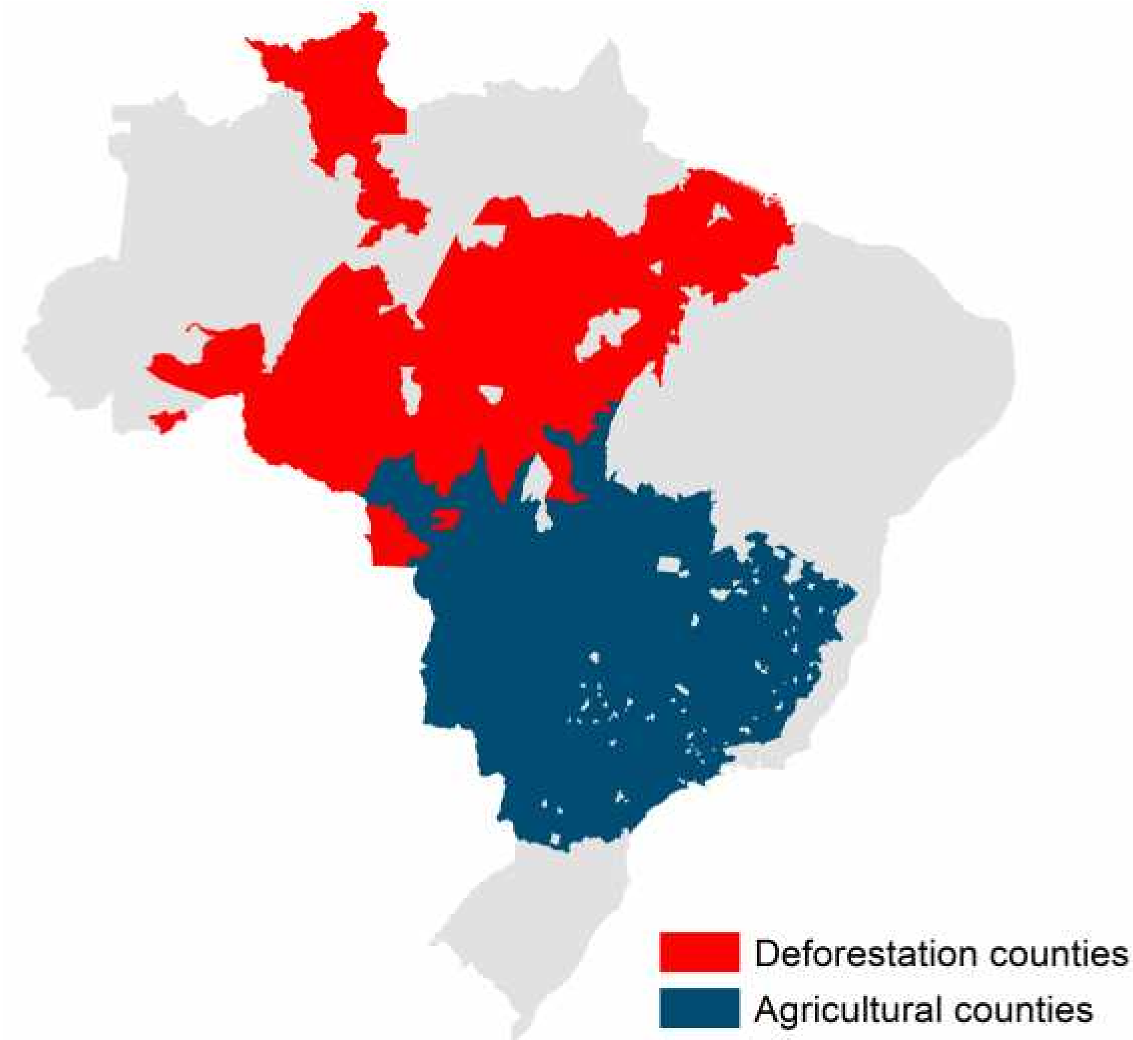

Figure G1. Deforestation and agricultural counties with road connection. This municipality boundary map was downloaded from the IBGE's website. Municipality boundaries are not shown to improve visualization. Projection: Albers Equal Area Conic.

8.8. Appendix H

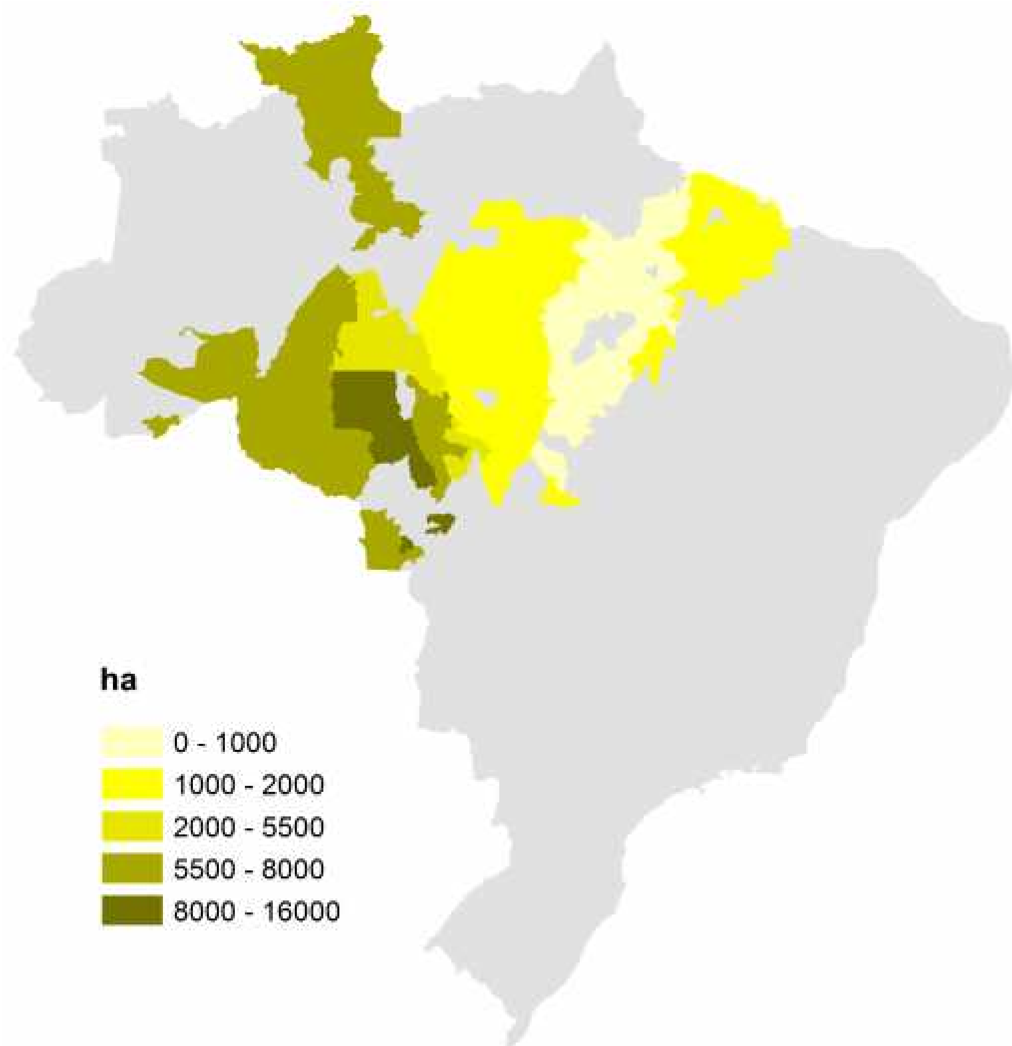

Figure H1. Total weighted indirect effect (the sum of current and lagged weighted indirect effects) of sugarcane measured in hectares during 2002-2012. This municipality boundary map was downloaded from the IBGE's website. Municipality boundaries are not shown to improve visualization. Projection: Albers Equal Area Conic.

8.9. Appendix I

Table I1. Collinearity statistics

Abbreviation	VIF	R-squared
DEF (-1)	1.18	0.1496
CATTLE	1.14	0.1197
SOY	1.32	0.2445
SUGAR	1	0.0023
CREDIT	1.36	0.264
GDP	1.01	0.0085
FINES	1.09	0.0856
PCATTLE	1.62	0.3814
PSOY	1.49	0.3286
PSUGAR	1.15	0.1292
WIESOY	1.52	0.3437
WIESUGAR	1.65	0.3949
WIESOY (-1)	1.61	0.3796
WIESUGAR (-1)	1.59	0.3727

VIF stands for variance inflation factor
VIF>5 is considered as a sign of severe multicollinearity
Variable abbreviations are introduced in Table 5
WIE stands for weighted indirect effect
(-1) indicates one-year lag

8.10. Appendix J

Brazil's Institute of Geography and Statistics (IBGE) recently (in 2010) conducted a meticulous survey of all households in Brazil. This reliable source of information is a perfect way to assess the accuracy of population estimates by WorldPop for Pará and to adjust the data to match municipality totals. The computations were done in ArcGIS. Firstly, it was verified that GIS estimates of territory closely match the figures reported by IBGE (see Figure J1A).

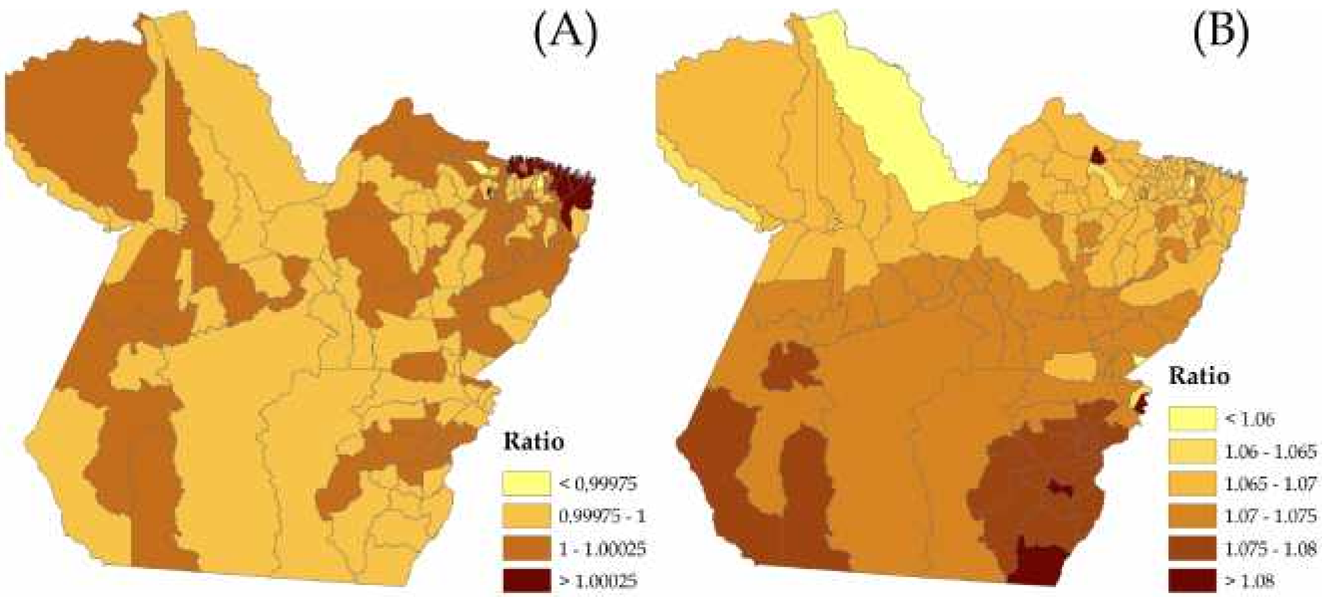

Figure J1. Population data (WorldPop) accuracy assessment: the ratio between territory reported by IBGE and territory computed in ArcGIS (A) and the ratio between population reported by IBGE (census 2010) and 2010 population estimates by WorldPop (B). Computations in ArcGIS were based on Albers Conic projection, the same projection used by IBGE. Selected standards match those used by IBGE and are as follows: longitude of origin is -54 degrees, latitude of origin is -12 degrees, standard parallel 1 is -2 degrees and standard parallel 2 is -22 degrees.

Estimated population is the sum of estimated persons over all pixels that fall within administrative borders of a municipality. Correction factor is the ratio between census 2010 data, reported by IBGE, and ArcGIS estimates. Figure J1B suggests that WorldPop underestimates population in Pará, but with almost homogenous factor, ranging between 1.06 and 1.08. The adjustment was done at the original spatial resolution (~0.096 km).

8.11. Appendix K

Table K1. Descriptive statistics (grid level)

	Mean	**Standard deviation**	**Minimum**	**Maximum**
DEF	0.04	0.04	0	0.26
CTL	502	345	4	1748
POP	2.37	2.77	1	72.63
ALT	190.74	94.52	16.5	532.71
SLOPE	0.26	0.22	0	1.57
FLAT	138.19	84.91	27.07	683.8
FC	57.6	10.8	40	81.9
TIME	1021	1041	88	6816
ROAD	33.8	23.2	0	147.1
RIVER	28	22.3	0	95.5
PREC	2000	201.31	1430.6	2606.03
PA	22.3	22.1	0	97.1

8.12. Appendix L

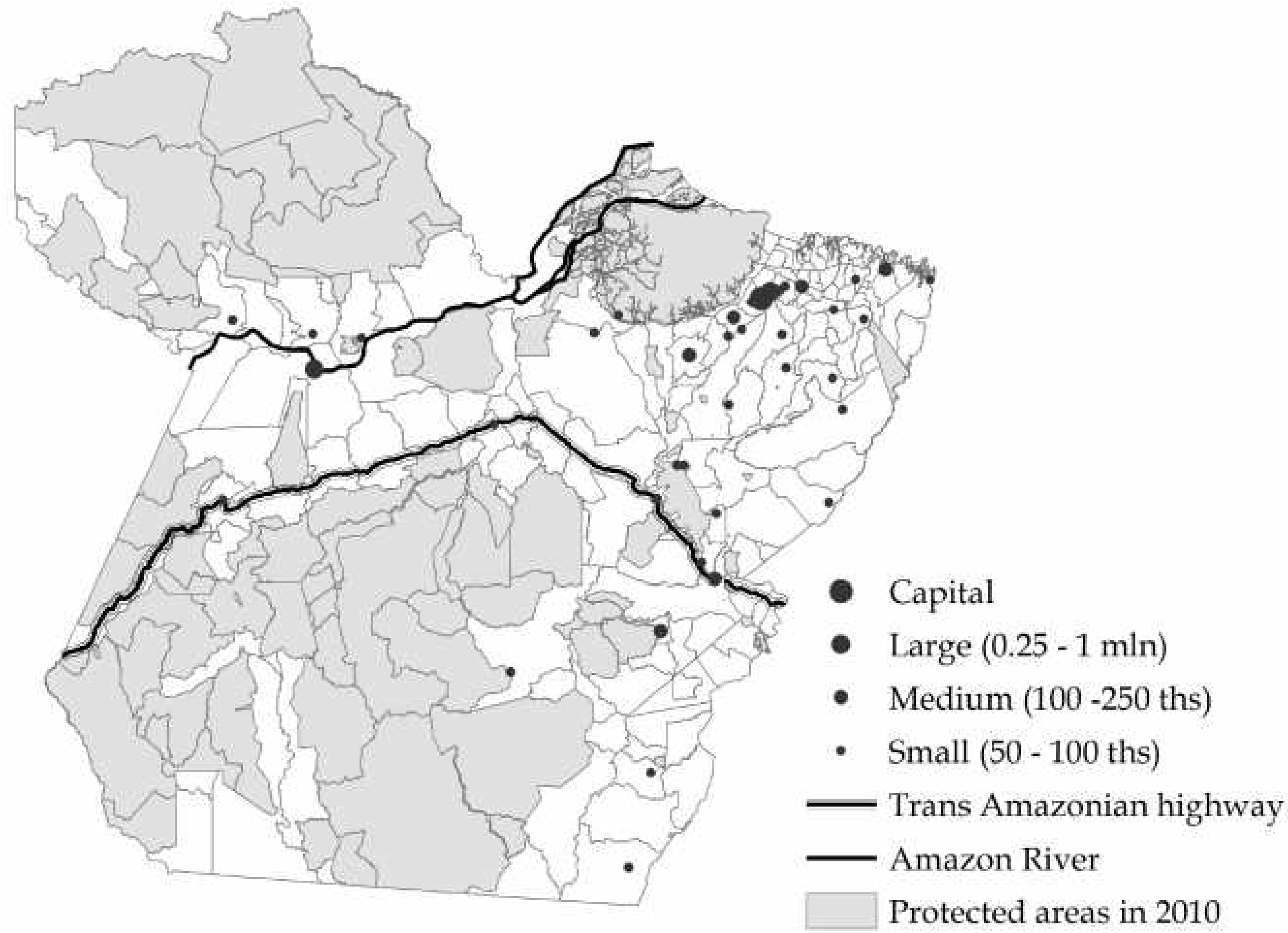

Figure L1. Map of Pará. Circles represent cities. The size of a circle is determined by the number of inhabitants in 2010. Protected areas include indigenous lands.

8.13. Appendix M

Table M1. Marginal effects of variables

	$ME_{FRACLOG}$	$ME_{FRACLOG}/ME_{OLS}$
CTL	0.035	0.85
POP	0.7211	0.93
ALT	0.0074	-0.89
SLOPE	-12.4357	1.33
FLAT	-0.0144	0.76
FC	-0.7867	1.25
TIME	-0.0019	-7.53
ROAD	-0.099	2.06
RIVER	0.1413	1.01
PREC	-0.0235	1.23
PA	-0.0951	1.33

8.14. Appendix N

Table N1. Regression tree: detailed statistics

Condition	# observations	Sum of squares	Deforestation mean
CTL < 130*	141	0.0148	0.0068
CTL > 130	445	0.8127	0.052
CTL < 847	351	0.4493	0.0433
FC < 58.7	200	0.2683	0.0515
POP < 2.9	161	0.1888	0.0457
ROAD < 60.1*	149	0.1276	0.0417
ROAD > 60.1*	12	0.03	0.0948
POP > 2.9*	39	0.0513	0.0757
FC > 58.7	151	0.1496	0.0324
PA < 2.5*	43	0.0112	0.0139
PA > 2.5*	108	0.1178	0.0398
CTL > 847	94	0.2367	0.0846
PREC < 2099	75	0.1807	0.0954
POP < 1.8	26	0.0491	0.0678
RIVER < 62.5*	15	0.0274	0.0476
RIVER > 62.5*	11	0.0072	0.0953
POP > 1.8	49	0.1012	0.1101
RIVER < 56.5*	28	0.0566	0.0932
RIVER > 56.5*	21	0.0261	0.1325
PREC > 2099*	19	0.013	0.0421

Stars indicate terminal nodes

8.15. Appendix O

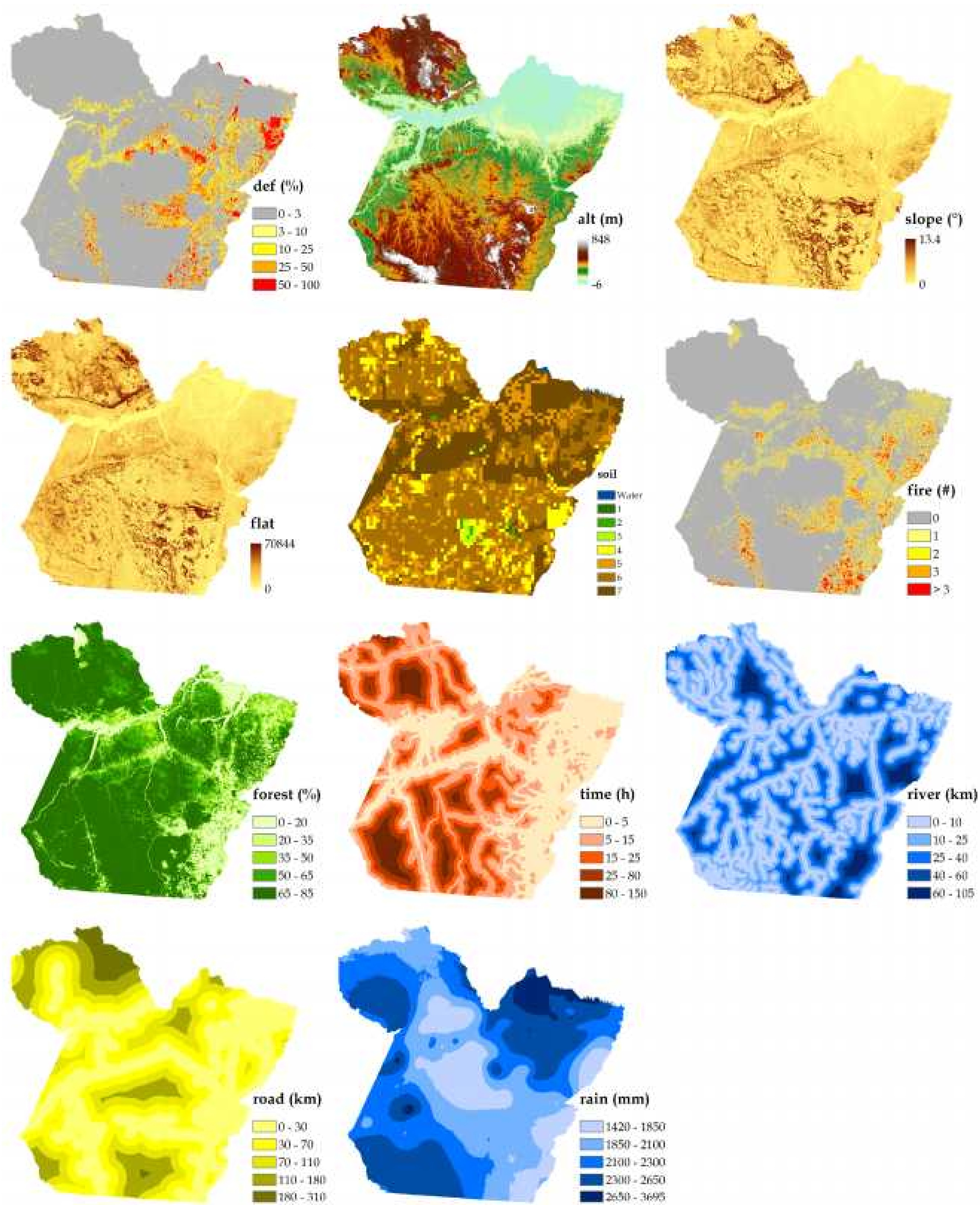

Figure O1. Spatial patterns of characteristics

8.16. Appendix P

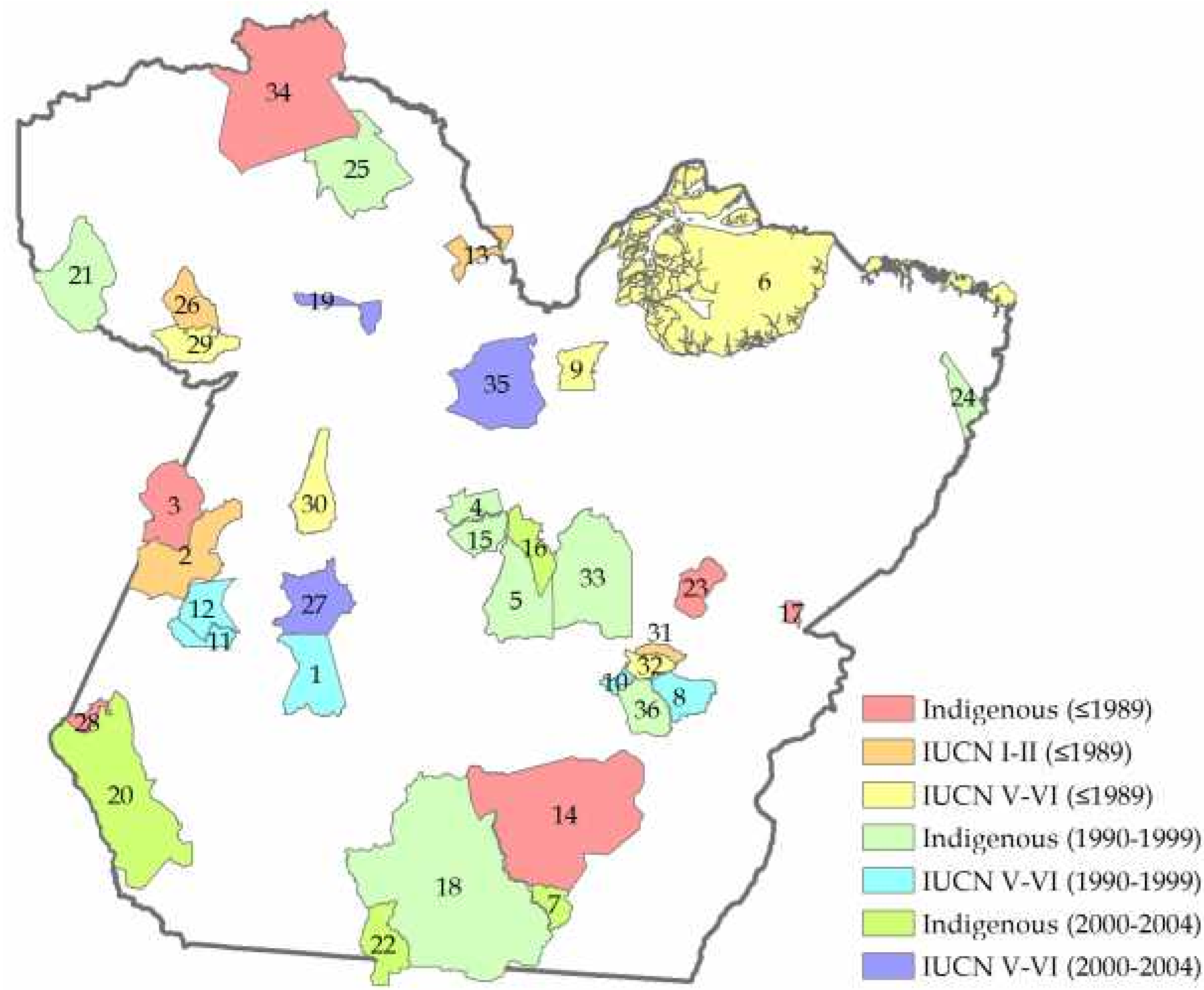

Figure P1. Protected areas in Pará by characteristics (year of establishment, type of governance, and IUCN category). The map shows only those protected areas, which were included in the analysis.

8.17. Appendix Q

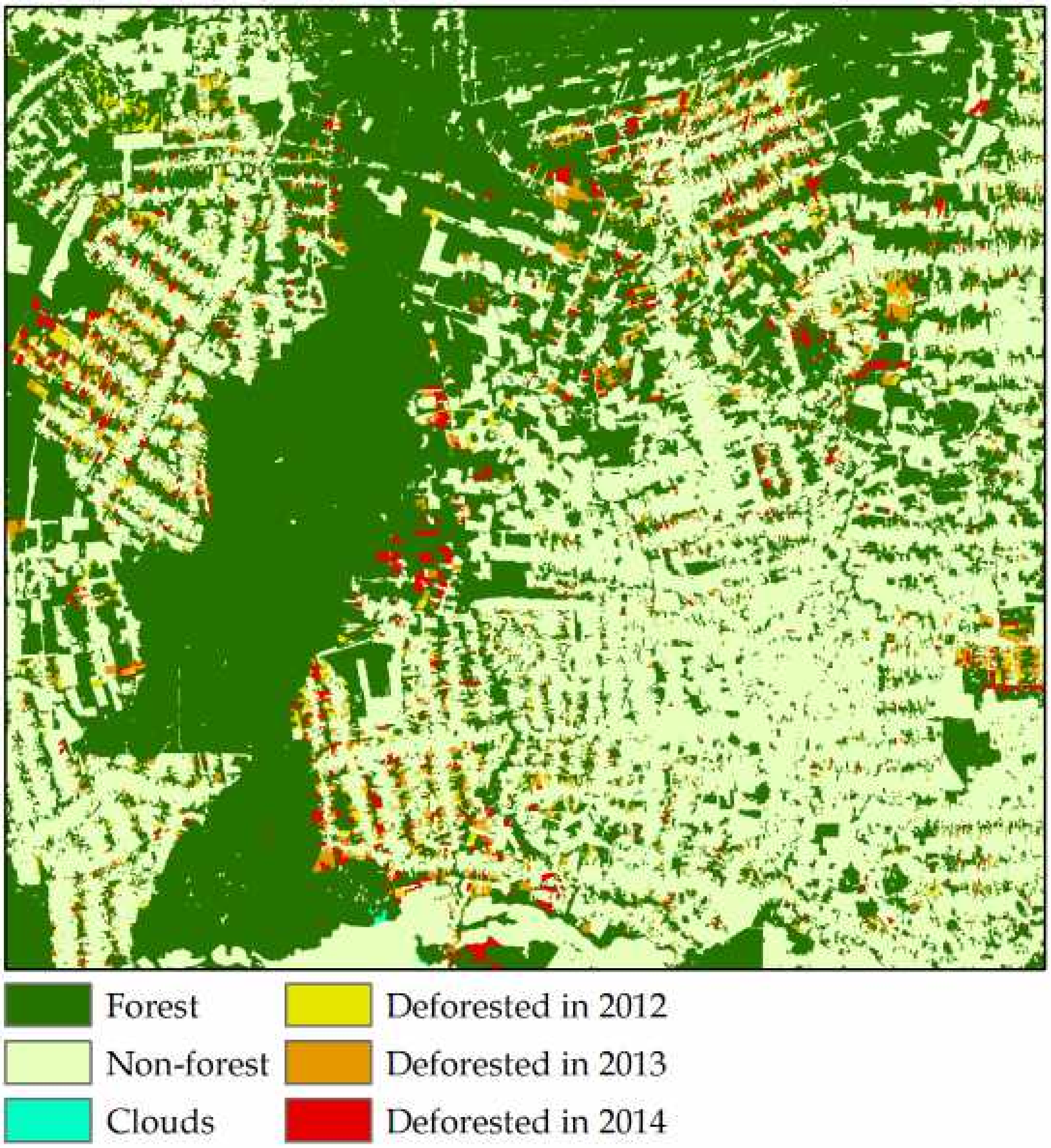

Figure Q1. Contagious deforestation. This map is adapted from INPE.

8.18. Appendix R

Table R1. Multicollinearity statistics: variance inflation factors (VIFs), condition numbers, and determinants of correlation matrices

	VIF		
	OLS 10	**OLS 12**	**OLS 14**
d13	1.9	1.91	1.91
d13f	2.47	2.52	2.83
for13	2.72	2.87	3.14
d12	1.85	1.85	1.85
d12f	2.74	2.78	2.9
elev	1.67	1.68	2.01
strict	1.6	2.05	2.12
indig	2	2.58	3.64
rof	1.31	1.33	2.45
cit	2.24	2.26	5.24
prec	-	-	4.94
op13	-	-	4.56
runf	-	2.35	3.44
pop	-	1.17	1.19
Condition no	27.52	28.96	130.35
Det(CorrMat)	0.0246	0.0089	0.0004

8.19. Appendix S

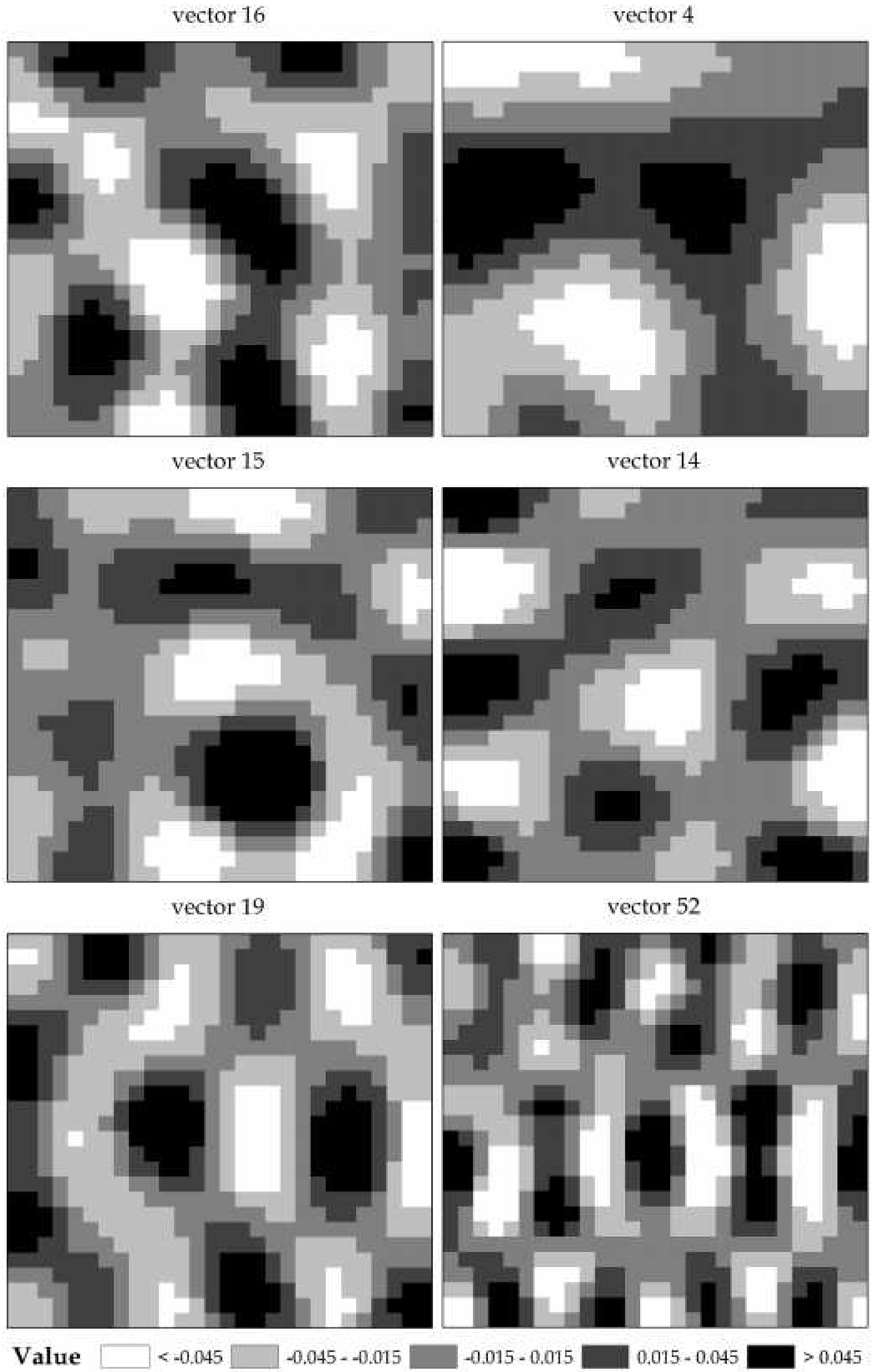

Figure S1. Map patterns of the first six eigenvectors

Chapter 9

Supplementary material

9.1. GWR-2SLS codes

This section presents a collection of Gauss codes that: 1) generate autocovariate term, 2) compute 2SLS parameters, 3) calculate Moran's statistic, 4) search for optimal number of the nearest neighbors, 5) estimate local coefficients for selected number of the nearest neighbors, 6) estimate local determination coefficients, 7) estimate variabilities in local coefficients and run Monte Carlo simulations, 8) estimate local standard errors and *p* values. Here *d* is the distance matrix (may be Euclidean, travel time, binary, etc.), *x0* is the matrix of endogenous and exogenous variables, *i0* is the matrix of exogenous and instrumental variables, *y* is the vector of response values.

```
» n=rows(y); /*the number of observations*/
» j=ones(n,1); /*creates a vector of ones*/
» id=eye(n); /*creates an identity matrix*/
» w0=1/d; /*here weights are calculated as inverse distances*/
» z=zeros(n,1); /*creates a vector of zeros*/
» w=diagrv(w0,z); /*replaces diagonal elements of matrix w0 with zeros*/
» proc (1)=sortallc(w); /*procedure to sort each column of a matrix separately*/
for i(1,cols(w),1);
w[.,i]=sortc(w[.,i],1);
endfor;
retp(w);
endp;
» ws=sortallc(w); /*sorts weights*/
» wn={};
» for i(1,cols(w),1); wn=wn~sortind(w[.,i]); endfor; /*creates a matrix of original row indices*/
» kn=3; /*selected number of neighbors for autocovariate*/
» nb=1;
» sum=0;
» sum1=0;
» do while (nb<=kn);
        a=(diag(ws[n+1-nb,.].*y[wn[n+1-nb,.],1])); /*numerator of eq. 1*/
        a1=(ws[n+1-nb,.])'; /*denumerator of eq. 1*/
```

```
        sum=sum+a;
        sum1=a1+sum1;
        a2=sum./sum1; /*creates a vector of autocovariate values*/
        nb=nb+1;
        endo;
» x=j~x0~a2; /*adds autocovariate to the list of covariates*/
» i=j~i0~a2; /*adds autocovariate to the list of exogenous covariates and instruments*/
» ii=inv(i'*i);
» b=inv(x'*i*ii*i'*x)*x'*i*ii*i'*y; /*2SLS estimator*/
» yp=x*b; /*predicted values of 2SLS model*/
» e=y-yp; /*errors of 2SLS model*/
» e2=(y-yp)^2;
» ss=sumc(e2); /*residual sum of squares*/
» ybar=meanc(y); /*average value of the dependent variable*/
» sstot=sumc((y-ybar)^2); /*total sum of squares*/
» r2=1-ss/sstot; /*determination coefficient of 2SLS regression*/
» cl=cols(x); /*number of coefficients to be estimated*/
» mse=sumc(e2)/(n-cl); /*mean squared error*/
» vcv=(inv(x'*i*ii*i'*x))*mse; /*variance-covariance matrix of 2SLS regression*/
» v=diag(vcv); /*variances of 2SLS coefficients*/
» ste=sqrt(v); /*standard errors of 2SLS coefficients*/
» tstat=b./ste; /*t statistics of 2SLS coefficients*/
» s0=j'*w*j; /*eq. D4*/
» s1=e'*(id-(j*j')/n)*e; /*eq. D5*/
» moran=(n/s0)*(e'*(id-(j*j')/n)*w*(id-(j*j')/n)*e)/s1; /*Moran's I*/
» s2=0.5*j'*((w+w').*(w+w'))*j; /*eq. D6*/
» s3=j'*(((j'*w)'+w*j).*((j'*w)'+w*j)); /*eq. D7*/
» s4=e-(j*e'*j)/n; /*eq. D8*/
» s5=(n*(s4.*s4)'*(s4.*s4))/(s1^2); /*eq. D9*/
» var1=n*((s2*(n^2-3*n+3)-n*s3+3*s0*s0))/((n-1)*(n-2)*(n-3)*s0*s0); /*see eq. D2*/
```

```
» var2=s5*(s2*(n^2-n)-2*n*s3+6*s0*s0)/((n-1)*(n-2)*(n-3)*s0*s0); /*see eq. D2*/
» emoran=-1/(n-1); /*eq. D3*/
» var=(var1-var2-emoran^2); /*eq. D2*/
» zscore=(moran-emoran)/sqrt(var); /*z score of Moran's statistic*/
» wsg=sortallc(d); /*sorts each column of the distance matrix separately*/
» g=1; /*number of nearest neighbors*/
» do while (g<=100); /*computes cvss (eq. 5) for gs from 1 to 100 with increments of 1*/
        loop=1;
        sum=0;
        do while (loop<=n);
                restart: if(loop<=n);
                h=wsg[g+1,.]; /*bandwidths are the distances to the pth nearest neighbor*/
                k=exp(-d./h); /*Gaussian type kernel*/
                k[loop,loop]=0; /*replaces k_ii^th entry with zero*/
                betaw=inv(x'*(k[.,loop].*id)*i*inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id)*x)*x'*
                (k[.,loop].*id)*i*inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id)*y; /*local coefficients es-
                timated using modified weighting matrices to avoid minimization at 0 nearest
                neighbors; the formula applies both to the case when the number of instruments
                (i) equals the number of endogenous covariates (z) and to the case when i>z;
                therefore, the formula is different from eq. 2, which is applicable only under
                i=z*/
                cv=(y-(x*betaw))^2;
                cvss=cv[loop,1];
                sum=sum+cvss; /*cross validated sum of squares*/
                loop=loop+1;
                goto restart;
                endif;
                endo;
        print sum;
        g=g+1;
```

```
        endo;
» g=15; /*optimal number of neighbors*/
» loop=1;
» coef={};
» s={};
» do while (loop<=n); /*the following lines estimate local coefficients and hat matrix*/
        restart: if(loop<=n);
        h=wsg[g+1,.];
        k=exp(-d./h);
        c0=inv(x'*(k[.,loop].*id)*i*inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id)*x)*x'*(k[.,loop].*id)*i*
        inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id);  /*eq. 8*/
        coef0=c0*y;
        coef=coef~coef0; /*local coefficients*/
        s0=x[loop,.]*c0;
        s=s|s0; /*hat matrix*/
        loop=loop+1;
        goto restart;
        endif;
        endo;
» i=1;
» do while (i<=n); /*the following lines estimate local determination coefficients*/
        yp=x*coef[.,i];
        nom=sumc(k[.,i].*(yp-meanc(yp)).*(y-meanc(y)))/n;
        denom1=sqrt(sumc(k[.,i].*((yp-meanc(yp))^2))/n);
        denom2=sqrt(sumc(k[.,i].*((y-meanc(y))^2))/n);
        r=nom/(denom1*denom2);
        r2=r^2;
        print r2; /*prints local determination coefficients*/
        i=i+1;
        endo;
```

```
» i=1;
» ss={};
» do while (i<=n); /*the following lines estimate residual sum of squares of GWR*/
        yp=x*coef[.,i];
        e2=(y[i,1]-yp[i,1])^2;
        ss=ss|e2;
        i=i+1;
        endo;
» r2gl=1-sumc(ss)/sstot; /*pseudo-global determination coefficient of GWR*/
» vbar=meanc(coef'); /*averages of local coefficients by covariate*/
» v0=sumc((coef'-vbar')^2)/n; /*variabilities*/
» se0=sqrt(v0);
» ratio=se0./ste;
» ll=1;
» varm={};
» do while (ll<=n-1); /*the following lines run Monte Carlo simulations*/
        rr=rndi(n,1); /*randomly generates a vector of values from univariate distribution*/
        rr1=rr/maxc(rr); /*normalizes the values; this step is not necessary*/
        dr=rr1~d; /*adds the column of randomly generated values to the distance matrix*/
        drs=sortc(dr,1); /*sorts the distance matrix by randomly generated values*/
        drs1=drs[1:n,2:n+1]; /*sorted distance matrix (by row)*/
         drt=rr1~drs1'; /*adds the same vector of randomly generated values to the transposed
        distance matrix*/
        drts=sortc(drt,1); /*sorts the transposed distance matrix by randomly generated values
        */
        drts1=drts[1:n,2:n+1]; /*sorted distance matrix (both by row and column)*/
        wsg1=sortallc(drts1); /*sorts each column separately*/
        loop=1;
        coefr={};
        do while (loop<=n);
```

```
            restart: if(loop<=n);
            h=wsg1[g+1,.];
            k=exp(-drts1./h);
            betaw=inv(x'*(k[.,loop].*id)*i*inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id)*x)*x'*
            (k[.,loop].*id)*i*inv(i'*(k[.,loop].*id)*i)*i'*(k[.,loop].*id)*y;
            coefr=coefr~betaw;
            loop=loop+1;
            goto restart;
            endif;
            endo;
        vbar=meanc(coefr');
        v=sumc((coefr'-vbar')^2)/n;
        varm=varm~v;
        ll=ll+1;
        endo;
» rank0=varm[.,.].>v0; /*returns value 1 if the condition is correct and 0 otherwise*/
» rank=sumr(rank0); /*sums by row*/
» p=rank/n; /*approximate p values of Monte Carlo simulation*/
» yhat=s*y; /*GWR model predictions, eq. 10*/
» v1=sumc(diag(s)); /*trace of the hat matrix*/
» sigma=sumc((y-yhat)^2)/(n-v1); /*eq. 9*/
» loop1=1;
» varcoef={};
» do while (loop1<=n); /*the following lines estimate the variances of local coefficients*/
        restart: if(loop1<=n);
        h=wsg[g+1,.];
        k=exp(-d./h);
        c0=inv(x'*(k[.,loop1].*id)*i*inv(i'*(k[.,loop1].*id)*i)*i'*(k[.,loop1].*id)*x)*x'*(k[.,loop1].*
        id)*i*inv(i'*(k[.,loop1].*id)*i)*i'*(k[.,loop1].*id);
        varcoef0=diag(c0*c0'*sigma); /*eq. 7*/
```

```
        varcoef=varcoef~varcoef0; /*local variances*/
        loop1=loop1+1;
        goto restart;
        endif;
        endo;
» stecoef=sqrt(varcoef); /*local standard errors*/
» tcoef=coef./stecoef; /*local t statistics*/
» pv0=cdfTc(tcoef,n-cl);
» t1=tcoef[.,.].>0;
» pv1=pv0.*t1*2;
» t2=abs(t1-1);
» pv2=(1-pv0).*t2*2;
» pv=pv1+pv2; /*local p values*/
» pvt=pv';
```

9.2. Fixed-effects code with computation of ILUC variables

A programming code displayed in this section iteratively plugs in different bandwidths starting from 10 and ending at 1000 with increments of 10 in eq. 11 and prints MSE of fixed-effects regression under each bandwidth.

```
» h=10;
» do while (h<=1000);
    load data[2105,302]=D:\PhD\road_dist.txt; /*loads the matrix of road distances*/
    droad=data;
    load data[2105,25]=D:\PhD\crops.txt; /*loads the matrix of crop values, the first column is a state code, columns 2:13 include soy values of each year (from 2002 to 2012), columns 14:25 store sugarcane values of each year*/
    s=data;
    droad1=s[.,1]~droad; /*the following 5 lines are not necessary, those only offer a possibility to compute ILUC from a specific state*/
    m=droad1[.,1].<7;
    droad2=selif(droad1,m);
    c=cols(droad2);
    droad3=droad2[.,2:c]/1000;
    w=exp(-droad3/h); /*the kernel function given by eq. 11*/
    m1=s[.,1].<7; /*those two lines are not needed if the whole study area is considered*/
    s1=selif(s,m1);
    cc1={2,14}; /*vectors cc1 and cc2 are used to estimate indirect effects of soy and sugarcane separately*/
    cc2={13,25};
    k=1;
    crop3={};
    do while (k<=2);
        s2=s1[.,cc1[k,.]:cc2[k,.]];
        s3=s2';
        w1=w';
```

```
        n=rows(w1); /*# frontier counties*/
        p=1;
        crop2={};
        do while (p<=12);
                q=1;
                crop1={};
                do while (q<=n);
                        nom=sumr(s3[p,.].*w1[q,.]); /*nominator of eq. 12*/
                        denom=sumr(w1[q,.]); /*denominator of eq. 12*/
                        crop=nom/denom; /*eq. 12*/
                        crop1=crop1|crop;
                        q=q+1;
                        endo;
                crop2=crop2|crop1;
                p=p+1;
                endo;
        crop3=crop3~crop2; /*contains indirect effect estimates of soy and sugarcane*/
        k=k+1;
        endo;
crop4=trimr(crop3,0,302); /*the following 3 lines create lagged indirect effects*/
z=zeros(302,2);
crop5=z|crop4;
crop6=crop3~crop5; /*contains current and lagged indirect effect estimates (one-year
lags) of soy and sugarcane*/
i2=1; /*the following procedure creates an id column*/
c2={};
do while (i2<=12);
        i1=1;
        c1={};
        do while (i1<=302);
```

```
            c1=c1|i1;
            i1=i1+1;
            endo;
        c2=c2|c1;
        i2=i2+1;
        endo;
i3=1; /*the following procedure creates a t (time) column*/
jj=ones(302,1);
c3={};
do while (i3<=12);
        jj1=jj*i3;
        c3=c3|jj1;
        i3=i3+1;
        endo;
crop7=c2~c3~crop6;
crop8=sortmc(crop7,1|2); /*indirect effect data is sorted by id and by t*/
load data[3624,12]=D:\PhD\data_for_code.txt; /*loads the matrix of variables, already
sorted by id and by t*/
d00=data;
d0=crop8[.,1]~d00~crop8[.,3:6]; /*adds an id column and indirect effect variables to the
dataset */
d=packr(d0); /*deletes rows with missing data (observations of 2001)*/
t=11; /*# time periods (years)*/
obs=302; /*# observations (frontier counties)*/
ii=1; /*the following lines do demeaning*/
d3={};
do while (ii<=obs);
        mask=d[.,1]./=ii;
        d1=delif(d,mask);
        mnc=meanc(d1);
```

```
        mnct=mnc';
        jj=ones(t,1);
        mncj=mnct.*jj;
        d2=d1-mncj;
        d3=d3|d2;
        ii=ii+1;
        endo;
 cl=cols(d3);
dfe=d3[.,2:cl]; /*demeaned variables*/
yfe=dfe[.,1]; /*demeaned dependent variable*/
xfe=dfe[.,2:cl-1]; /*demeaned independent variables*/
bfe=inv(xfe'*xfe)*xfe'*yfe; /*fixed-effects (FE) estimator*/
sefe=sqrt(diag(inv(xfe'*xfe)*(sumc((yfe-xfe*bfe)^2)/(obs*(t-1)-cols(xfe))))); /*FE stan-
dard errors*/
tfe=bfe./sefe; /*FE t statistics*/
d4=d[.,1]~d3[.,2:cl];
j=1; /*the following lines calculate cluster-robust standard errors of FE regression*/
sum=0;
do while (j<=obs);
        mask1=d4[.,1].==j;
        ds=selif(d4,mask1);
        dsy=ds[.,2];
        dsx=ds[.,3:cl];
        dse=dsy-dsx*bfe;
        see=dse*dse';
        sw=dsx'*see*dsx;
        sum=sum+sw;
        j=j+1;
        endo;
secfe0=inv(xfe'*xfe)*sum*inv(xfe'*xfe)*(obs/(obs-1))*(obs*t-1)/(obs*t-cols(xfe));
```

```
secfe=sqrt(diag(secfe0));
tcfe=bfe./secfe; /*FE t statistics under cluster-robust standard errors*/
o=bfe~secfe~tcfe; /*output (coefficients, cluster-robust standard errors and t statistics*/
y=d[.,2]; /*original values of the dependent variable*/
x=d[.,3:cols(d)]; /*original values of the independent variables*/
p=x*bfe; /*predicted values*/
mse=sumc((y-p)^2); /*eq. 13*/
print mse;
h=h+10;
endo;
```

9.3. Spatial filtering codes

This section presents programming codes used to obtain eigenvectors. Nearest neighbors were identified in ArcGIS using *Generate Near Table* tool.

Gauss 10 script for obtaining the connectivity matrix

```
» load data[5504,4]=D:\PhD_D\P3\Data728\near_nb.txt, /*.txt file was created directly from ArcGIS, necessary information is stored in the second column (id of a cell) and in the third column (id of one of its nearest neighbors)*/
» d=data,
» d1=d[.,2:3],
» d2=d1+1, /*original ids range from 0 to n-1, n is a number of observations*/
» n=rows(d2),
» l=1,
» w={}, /*stores the connectivity matrix*/
» do while (l<=n),
        mask0=d2[.,1]./=l,
        d3=delif(d2,mask0),
        i=1,
        k={},
        do while (i<=n), k=k|i, i=i+1, endo, /*creates a 1,2,…,n vector*/
        c=rows(d3),
        j=1,
        s={},
        do while (j<=c),
                mask=k[.,1].==d3[j,2],  /*creates a vector of n rows which contains value 1 if the element is a c^th nearest neighbor of a cell to which a vector corresponds and 0 otherwise*/
                s=s~mask,
                j=j+1,
        endo,
```

```
        v=sumr(s), /*creates a vector that codes all nearest neighbors as ones and non-neigh-
        bors as zeros*/
        w=w~v,
        l=l+1,
        endo,
» output file = D:\PhD_D\P3\Data728\conn_nb.txt reset,
» w,
» output off, /*Gauss writes data from left to right with 4 elements in each row*/
```

R 3.2.2 script for obtaining eigenvectors

```
> library(maptools)
> library(spdep)
> setwd("D:/PhD_D/P3/Data728")
> d<-read.table("D:/PhD_D/P3/Data728/d.txt") /*contains values of dependent and inde-
pendent variables*/
> c0<-as.matrix(read.table("D:/PhD_D/P3/Data728/conn_nb.txt")) /*works only if n squared
is divisible by 4, otherwise c0 cannot be read as a matrix*/
> c<-matrix(t(c0),728,728) /*converts c0 into original connectivity matrix*/
> x <- mat2listw(c) /*converts matrix c to a weights list object*/
> lw <- nb2listw(x$neighbours, style="W") /*W coding implies row standardization*/
> nSFE <- SpatialFiltering(V1 ~ ., data = d, nb = x$neighbours, style = "W", ExactEV = T)
/*ExactEV uses exact expectations and variances rather than the expectation and variance of
Moran's I from the previous iteration*/
> nlmSFE <- lm(V1 ~ . + fitted(nSFE), data = d) /*OLS with eigenvectors*/
> summary(nlmSFE) /*reports OLS results*/
> write.csv(fitted(nSFE), "eigen_nb.csv") /*writes values of eigenvectors into .csv file*/
```