International Development

ISSN 1470-2320

Prizewinning Dissertation 2023

No.23-BL

Competing for Land: A Spatial Investigation of Large-Scale Land Acquisitions, Their Target Context, and the Dynamics of Deforestation in Africa

Luc Bitterli

Published: June 2023

Department of International Development

London School of Economics and Political Science

Houghton Street

London

WC2A 2AE UK

Tel: +44 (020) 7955 7425/6252

Fax: +44 (020) 7955-6844

Email: d.daley@lse.ac.uk

Abstract

This paper empirically examines the frequency and extent to which large-scale land acquisitions (LSLAs) for agriculture in Africa target areas with pre-existing cropland and forests. It also analyses how these land acquisitions are associated with increased deforestation within and near the acquisition areas. Using a difference-in-differences method, the econometric results suggest that LSLAs are an important driver of deforestation. Furthermore, the remote sensing analysis of pre-acquisition land cover casts doubt on the claim that LSLAs mainly target "idle land". The findings support the need to adopt and implement comprehensive landscape plans for LSLAs to address potential trade-offs between economic, social, and environmental objectives.

Acronyms

ADM1	First-Order Administrative Division
ADM2	Second-Order Administrative Division
ATT	Average Treatment Effect on the Treated
CI	Confidence Interval
HCV	High Conservation Value
km	Kilometres
LL	Lower Limit
LM	Land Matrix
LSLA	Large-Scale Land Acquisition
OECM	Other Effective Area-Based Conservation Measures
RSPO	Roundtable on Sustainable Palm Oil
UL	Upper Limit
WGS-84	World Geodetic System
WKT	Well-Known Text

Table of Contents

1 Introduction

With the food price crisis in 2008 and the "global land rush", the number of large-scale land acquisitions (LSLAs) has risen sharply worldwide (Arezki et al., 2015). In this paper, the term "LSLAs" refers to transactions initiated since 2000 that involve the transfer of ownership, control or use of land by concession, lease or sale and cover an area for agricultural production exceeding 200 hectares (The Land Matrix, 2023). In light of recent developments such as geopolitical tensions, fears of food inflation and biofuel regulations to combat climate change, the global demand for land could rise again (Lay, Anseeuw, et al., 2021). This potential increase in demand continues to make LSLAs a salient issue, both as a feature of the ongoing academic discourse and in the broader context of land acquisition.

In theory, LSLAs can lead to a "positive sum" for both investors and host countries' societies by bringing unused land in "land-abundant" countries into agricultural production (Deininger & Byerlee, 2011). The underlying implication of the argument is that by targeting marginal land, LSLAs may not have a significantly negative impact on the agroecological environment (Borras et al., 2011) and the food production and supply of local communities (Borras & Franco, 2014). Moreover, spillover effects of LSLAs on smallholder farmers, such as improved access to inputs like fertilisers or seeds, as well as the provision of infrastructure, could help reduce the gap between potential and current agricultural yields (Collier, 2008; Deininger & Byerlee, 2011, 2012).

In contrast, numerous case studies on LSLAs contradict this rhetoric of "unused" or "idle" land and point out that LSLAs compete directly with polyvalent smallholder agriculture for land. This competition can, in turn, lead to the displacement of previous agricultural use, potentially triggering indirect land use changes, for example, if the displaced farmers clear new land. Moreover, Yang & He (2021) identify various case studies of LSLAs that targeted forests and led to deforestation. Those LSLAs directed towards forested regions may trigger direct land use changes with associated negative environmental impacts and lead to local communities losing access to natural resources, as documented in multiple cases (Oberlack et al., 2016).

Current knowledge about the targeting context and impacts of LSLAs comes largely from case studies, which may be affected by publication and selection biases and, therefore, may have limited external validity. While there is a rich body of qualitative case studies, empirical quantitative analyses of LSLAs are rare. Only a few quantitative studies are based on rigorous before-and-after analysis, mainly due to a lack of baseline data (Oya, 2013; Yang & He, 2021). Overall, there are limited systematic quantitative studies that transcend cases and contexts. This paper addresses this gap in the existing literature through a quantitative cross-country analysis of LSLAs in Africa. Africa is a notable example of this endeavour, as it is the continent where most land acquisitions have occurred. Furthermore, the prevalent issue of insecure land tenure combined with weak governance systems in many regions in

Africa are potential factors that could exacerbate potential adverse impacts of LSLAs (Cotula, 2009; Messerli et al., 2014), thereby warranting a cross-country analysis within Africa.

The aim of the paper is threefold: first, based on the general literature on land competition as well as the findings of LSLA case studies, the paper develops a conceptual framework that depicts the different mechanisms of potential direct and indirect land use changes caused by LSLAs. Moving from theory to practice, the paper subsequently analyses how frequently and to what extent large-scale land acquisitions in Africa target areas with pre-existing cropland and forests. Third, complementary to the existing case studies, the following research question is addressed:

To what extent are large-scale land acquisitions in Africa associated with land use change in the form of increased deforestation within and near LSLAs?

The paper's empirical aspiration calls for a quantitative methodology based on the analysis of satellite imagery. By analysing novel high-resolution time-series remote sensing data, this paper seeks to be the first to analyse the presence of pre-LSLA cropland in a cross-country analysis at this level of granularity. Furthermore, using Callaway & Sant'Anna's (2021) novel difference-in-differences method, which allows for treatment heterogeneity, I compared the temporal change of forest area within and near LSLAs with more distant control areas. In this sense, this paper contributes to the broader literature dealing with land competition and land use change in general and to the specific literature on LSLA target contexts (e.g. Dell'Angelo et al., 2017; Messerli et al., 2014) and environmental impacts (e.g Davis et al., 2023; Rulli et al., 2013).

The paper is structured as follows: section 2 reviews the existing literature and provides a conceptual framework for LSLA-related land use changes; section 3 describes the data and methodology used; section 4 discusses the results of the quantitative analysis; section 5 reflects on some of the limitations of the paper and points to avenues for future research; section 6 concludes.

2 Background and Conceptual Framework

2.1 Background

The drivers of the “global land rush” (Arezki et al., 2015) are manifold and range from the global to the national and sub-national levels. Drivers following a global dynamic include private sector expectations of higher commodity prices, increasing demand for agricultural commodities for food, biofuel and other commodities (Cotula, 2013; Zoomers, 2010), geopolitics (Oliveira, 2017), the potential future vulnerability of national food production to climate change (Davis et al., 2015), ecosystem protection efforts (Zoomers, 2010) and the financialization of land and food systems (Clapp & Isakson, 2018; Fairbairn, 2017). Following the Land Matrix (2023), LSLAs are hereafter understood as deals initiated since 2000 involving the transfer of ownership, control, or use of land by concession, lease, or sale and covering an area of more than 200 hectares.

Global investors originate from diverse geographical locations, including "food insecure" countries that rely on imports to feed their populations (e.g. the Gulf States) or countries that seek land for biofuel production (Zoomers, 2010). Investors' "real" origin often remains unknown, as many parent companies operate from financial centres and tax havens such as Cyprus, Singapore, the British Virgin Islands and Hong Kong. However, the global land rush has weakened recently, as visible in the decrease in documented deals concluded globally since 2015. Nonetheless, global demand for land could rise again due to enhanced biofuel regulations to tackle global warming, concerns about rising food prices, and geopolitical tensions (Lay, Anseeuw, et al., 2021). This potential future upward trend, combined with the current large number of already active LSLAs, must be seen against the background that the subject of LSLAs is highly contentious, evident in divergent (academic) perspectives on the different impacts of LSLAs (Lay, Nolte, et al., 2021).

Proponents of LSLAs argue that agricultural investments can increase agricultural productivity and reduce poverty. Spillover effects of LSLAs on smallholder farmers, for instance, by improving availability and access to technologies and inputs such as improved seeds, fertilisers, storage facilities, and explicit or implicit provision of credit, could positively impact smallholder yields. These spillover effects would help close the "yield gap" - the difference between potential agricultural yields and current yields – the largest in sub-Saharan Africa . Furthermore, LSLAs can generate wage employment and provide infrastructure such as roads, schools, or health centres (Collier, 2008; Deininger & Byerlee, 2011, 2012). Based on these arguments, foreign investments in agricultural land have been promoted, arguing that "idle land" would be put to productive use (R. Hall, 2011).

However, the classification of “idle land” is often based on perceived productivity rather than the existing use of land (Cotula, 2009). For example, Messerli et al. (2014) found that worldwide 35% of the georeferenced deals in the Land Matrix database contained land classes consisting of mixed mosaics of vegetation and rain-fed cropland, suggesting that part of the land was already used for farming. The authors argue that LSLAs can raise competition and conflicts over land. Local people's vulnerability can increase as the transfer of land rights to an investor may lead to people losing their access to and rights over land, often with little or no compensation (Schoneveld, 2014). Wage employment opportunities for displaced locals are limited and depend strongly on the type of crop grown and the associated labour intensity (Deininger & Byerlee, 2011; Deininger & Xia, 2016). Opponents have thus argued that LSLAs do not contribute significantly to poverty reduction and local development (Li, 2011) and refer to LSLAs as "land grabbing", which follows a process of "accumulation by dispossession" (D. Hall, 2013).

As Oya (2013) aptly puts it, the "land rush" was also accompanied by a "literature rush", which was, however, associated with major analytical gaps and methodological problems. As Edelman (2013, p. 489) notes about the epistemology of land acquisition data: "many researchers acknowledge the obvious – that land deal data are frequently problematic – but they then go on and analyse those data as if they were generated in a highly rigorous way." The criticism of "false precision" originally raised for the World Bank global poverty data by Pogge & Reddy (2005, p. 4) may also apply to certain LSLA studies in which limited data is incorrectly extrapolated, "thereby [creating] an appearance of precision that masks the high probable error of its estimates." For instance, studies such as the well-cited paper of Rulli et al. (2013) have been criticised for exuding a certain scientific rigour with their estimates of "water grabbing" concerning LSLAs, but this masks the inherent uncertainties in their data used (Oya, 2013; Schoneveld, 2014).

Most research on LSLAs relies on case studies, which are criticised for their context-specificity, difficulty in achieving external validity and potential publication and case selection biases (Yang & He, 2021). Quantitative, conclusive empirical evidence is rare. Out of some 128 LSLA case studies published since 2007, Yang & He (2021) find only eleven that provide quantitative insights into local livelihood changes, including four using specific quantitative analysis methods. A major challenge in the quantitative analysis of the impact of LSLAs is the lack of baseline data on land use, different groups and livelihoods - data that would be required for a rigorous before-and-after analysis (Oya, 2013).[1] While Yang & He (2021) identify 78 case studies worldwide that examine various environmental impacts of LSLAs, there is little quantitative analysis on whether the land use changes identified in these mostly qualitative case studies transcend cases and contexts. To fill this gap, I develop a simple conceptual framework in the following chapter. Based on this framework, the following two research questions are addressed:

(i) *How frequently and to what extent do large-scale land acquisitions in Africa target areas with pre-existing cropland and forests?*

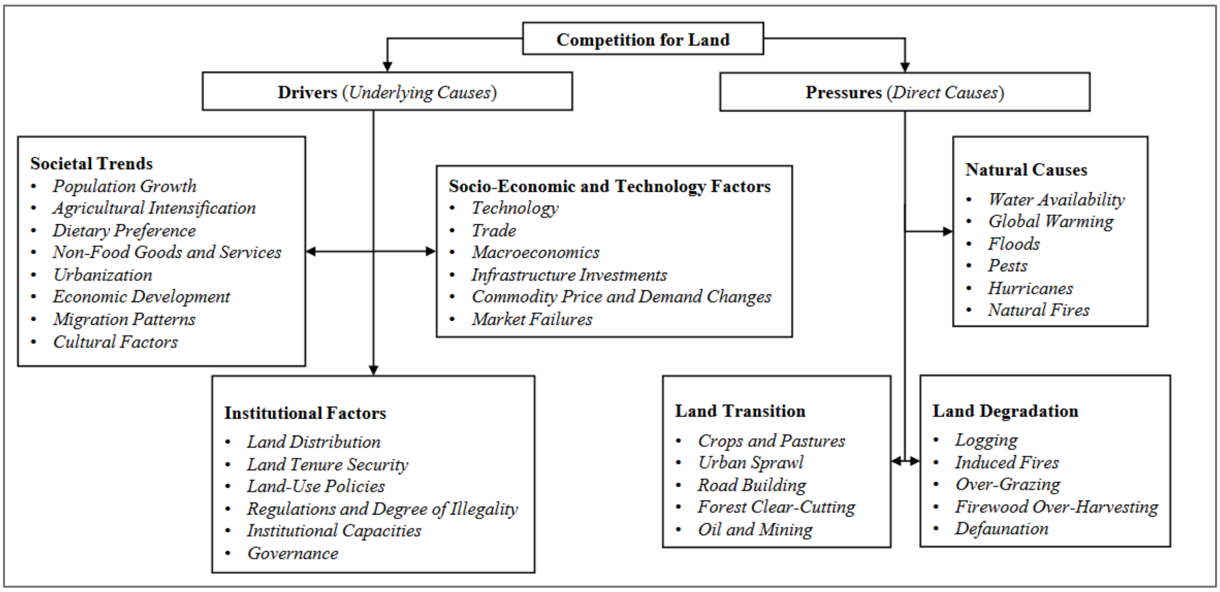
[1] For instance, studying the same LSLA in Sierra Leone, Yengoh & Armah (2015) found a decrease in perceived food security, Bottazzi et al. (2018) found a significant increase, and Hofman et al. (2019) found no significant effect at all.

(ii) To what extent are large-scale land acquisitions in Africa associated with land use change in the form of increased deforestation within and near LSLAs?

2.2 Conceptual Framework: Land Competition

In the following, I briefly discuss land use change in general and then in the context of LSLAs. Land conversions are a major driver of global environmental change (Turner et al., 2007). The most important form of land use change in terms of area is expanding crop and pastoral land into natural ecosystems. Land conversions, especially when tropical rainforests are involved, are often associated with significant losses of biodiversity (Drescher et al., 2016), so deforestation is not only a phenomenon of competition for land but also one of competition for ecosystem services (DeFries et al., 2010; Smith et al., 2010). Competition for land, although identified as a driver affecting land use, arises due to numerous other drivers and pressures. Following Smith et al. (2010), I define *pressures* as the direct causes motivating land competition. *Drivers*, on the other hand, represent the underlying causes for competition and are factors of higher causal order that determine the extent of the pressures (*see Figure 1*).

***Figure 1**: Drivers and Pressures*



Authors adaptation of Smith et al.'s (2010) framework

For simplicity, I assume that total land area Z in region m consists of the sum of land used for nature conservation (Z_C), agricultural land (Z_A), built-up areas (Z_B), water bodies (Z_W), and other designated land uses such as logging or mining (Z_O).

$$(1) \qquad Z_m = Z_{C,m} + Z_{A,m} + Z_{B,m} + Z_{W,m} + Z_{O,m}$$

Furthermore, $Z_{A,m}(0)$ and $Z_{A,m}(t)$ represent the agricultural land area at time t = 0 and t, respectively. Agricultural land is understood as land used for agricultural production, including cropland and pasture.

I model $Z_{A,m}(t)$ with the agricultural land growth rate r for $t \geq 1$ as follows:

$$(2) \qquad \mathbf{Z_{A,m}(t) = Z_{A,m}(t\text{-}1) \cdot (1+r(t))}$$

Land competition in the model is expressed in such a way that a change in the agricultural area simultaneously causes a change in non-agricultural areas:

$$(3) \qquad \frac{dZ_{A,m}(t)}{dt} = -(\alpha\frac{dZ_{C,m}(t)}{dt} + \beta\frac{dZ_{B,m}(t)}{dt} + \gamma\frac{dZ_{W,m}(t)}{dt} + \delta\frac{dZ_{O,m}(t)}{dt})$$

with α, β, γ, δ representing the influence of the corresponding land use changes on the rate of change of the agricultural area.

I assume that the growth rate of agricultural land at time t in region m depends on the previous growth rate and the changes in the underlying drivers $\Delta D_m=D_m(t)-D_m(t-1)$:

$$(4) \qquad r_m(t) = r_m(t\text{-}1) + \theta \cdot \Delta D_m$$

, with θ representing the influence of ΔD_m on $r_m(t)$.

The linear equation described is a strong simplification of the agricultural expansion process, as agricultural land expansion is typically a non-linear process (Lambin & Meyfroidt, 2010). Nevertheless, the equation might be helpful to understand that without substantial changes in the underlying drivers, the expansion rate of agricultural land might not alter significantly. For instance, due to the absence of agricultural inputs, it is common in most rural areas in sub-Saharan Africa for smallholders to try to increase their production and adapt to increased soil degradation by expanding their cropland (IAASTD, 2009). An in-depth description of all the potential drivers and pressures of land competition listed in *Figure 1* would go beyond the scope of this paper, so in the following I will focus on the most important drivers concerning LSLAs.

LSLAs, as one of the phenomena of globalisation, create new flows of goods, people, money and information, thus increasing the interconnectedness of land use across geographical distances (Liu et al., 2013; Oberlack et al., 2016). Due to global-scale spatial dynamics, land use changes can trigger direct and indirect land use changes on the global, regional and local scale (Lambin & Meyfroidt, 2011). The term indirect land use change was first proposed by Searchinger et al. (2008) to describe the unintended release of greenhouse gas emissions caused by global land use change due to cropland expansion for biofuel production. Indirect land use changes may occur far from the original displaced land use area (Andrade de Sá et al., 2013), but can also occur locally. In the following, only the local direct and indirect land use changes caused by an LSLA are modelled.

Let us assume that at time t=0 no LSLA was present, so that

$$(5) \qquad Z_{A,m}(0) = Z_{A,m,non\text{-}LSLA}(0)$$

and at time k an LSLA has occurred so that total agricultural land equals:

$$(6) \qquad Z_{A,m}(k) = Z_{A,m,non\text{-}LSLA}(k) + Z_{A,m,LSLA}(k)$$

The change in the growth rate of agricultural land (and thus a concomitant change in land use in region *m* at time *t*) caused by an LSLA corresponds to the following:

$$(7) \qquad \frac{\partial r_m(t)}{\partial LSLA} = \frac{\partial r_m(t)}{\partial \Delta D_m} \cdot \frac{\partial \Delta D_m}{\partial LSLA} = \overset{(i)}{\frac{dZ_{A,m,LSLA}(t)}{dt}} - \overset{(ii)}{\frac{dZ_{A,m,displaced}(t)}{dt}} + \overset{(iii)}{\frac{dZ_{A,m,indirect}(t)}{dt}}$$

with $Z_{A,m,displaced}(t) = Z_{A,m,non-LSLA}(t) \cap Z_{A,m,LSLA}(t)$

2.2.1 Direct Land Use Changes

The first element on the right-hand side of *equation (7)* comprises the expansion of the LSLA's agricultural area at time *t*. Corresponding changes in land use accompany the expansion of the LSLA. The only study I know of that examined the specific relationship between LSLAs and land use changes worldwide is Davis et al. (2023), who found a significant correlation between approximated areas of LSLAs and deforestation.

The second element $Z_{A,m,displaced}(t)$ represents the agricultural area at time *t* that the LSLA displaces. If, at time *t,* non-LSLA agricultural land is displaced by the LSLA expansion, the equation described above subtracts the displaced agricultural land from the LSLA area expansion. To my knowledge, the only cross-country study examining local pre-LSLA agricultural land cover is Messerli et al. (2014). However, this study has certain limitations, mainly because accurate data was unavailable. First, with a resolution of 300 metres, the dataset used by Messerli et al. is insufficient to detect smaller-scale land use patterns, so only mosaics between vegetation and cropland are reported, and the extent of affected cropland within an area cannot be determined. Second, the dataset used has limited temporal scope as it is only based on data from 2009. The limited temporal scope could result in production areas being misinterpreted as pre-LSLA land use for certain LSLAs that started production before 2009. At the same time, it is also conceivable that significant changes in land use occurred between 2009 and the LSLA implementation date. Third, Messerli et al. (2014) analysed the land use patterns in 2009 within 10 kilometres of the coordinate listed for each LSLA in the Land Matrix database. However, given the strong heterogeneity in the size of reported LSLAs, Messerli et al. thus potentially misestimate the areas directly and indirectly affected by LSLAs. I attempt to overcome these limitations as part of this dissertation by using a novel high-resolution time series dataset for cropland (see *section 3.1*).

2.2.2 *Local Indirect Land Use Changes*

The third element on the right-hand side of equation (7) describes the local indirect effects of the LSLA on the expansion of agricultural land within a region *m*. The insights derived from case studies such as Zaehringer et al. (2018) or von Maltitz et al. (2016) compel me to critically reconsider the assumption made in the theoretical framework of Kleemann & Thiele (2015) that farmland is constant and that displaced farmers automatically switch to wage labour on the LSLA and do not farm on the rest of the land. For example, the case study by Zaehringer et al. (2018) in Mozambique found that where the LSLA affected areas of existing cropland, a significant proportion of displaced smallholders cleared forest near the LSLA to create new cropland. A similar spillover effect, where former land users were displaced by an LSLA and cleared new farmland in the surrounding woodland, was also found in a case study by von Maltitz et al. (2016). Furthermore, Magliocca et al. (2020) analysed deforestation in and around land concessions in Cambodia and found that about 3-10.7% of all forest loss in Cambodia by 2016 was due to indirect land use change. These identified displacement effects depend on numerous factors, in particular, the existence and extent of $Z_{a,m,displaced}(t)$ and the opportunity costs of affected farmers, which in turn depend on factors such as the scarcity of land in the target region and the degree to which livelihoods depend on the displaced agricultural land (Oberlack et al., 2016).

In addition to displacement effects, the literature on land competition lists rebound, cascade and remittance effects. The latter two describe indirect effects that primarily affect entire regions or countries and are thus potentially negligible in the analysis of local direct and indirect effects of LSLAs. The rebound effect describes the reaction of actors or an economic system to the introduction of a new technology or other measures that reduce resource use. Thus, an increase in production efficiency lowers the cost of consuming a product, which in turn can affect the quantity produced (Lambin & Meyfroidt, 2011). Given the high intra-national trade costs in many developing countries (Atkin & Donaldson, 2015), road investments by large-scale land investors could increase the efficiency of surrounding farmers by lowering input costs and improving market access.

Furthermore, spillover effects of LSLAs on local farmers in the form of access to technologies, inputs, and credit (see *section 2.1*) could increase the efficiency of smallholder farmers, which in turn may affect the extensification process of agricultural land. Potential spillover effects depend on numerous factors, particularly whether the crop can be planted by both LSLA and local farmers, which in turn depends on the capital/labour intensity of the crop (Lay, Anseeuw, et al., 2021). Deininger & Xia (2016) and Lay, Nolte, et al. (2021) found evidence for technological spillover effects in Mozambique and Zambia, respectively. However, they found no or only marginal effects on the extensive margin of cultivated land. This suggests that the indirect effect of land use change for most LSLAs may depend primarily on the presence of the displacement effect. Moreover, the impact of a potential efficiency increase on agricultural expansion is unclear. It could be that an intensification of agriculture weakens the extensification trend of agricultural land. On the other hand, according to Jevon's paradox, higher efficiency could lead to higher profitability, which could provide an incentive to expand agricultural land (Lambin & Meyfroidt, 2011).

Another identified potential driver of indirect land use change, as found in Zaehringer et al. (2018), could be that an expansion of cropland can be caused by the immigration of people seeking work on the LSLA.

2.3 Hypotheses & Study Area

Considering these various possible drivers by which LSLAs could lead to land use changes, the conceptual framework advances the following testable hypotheses:

I. *LSLAs frequently target areas with existing cropland and forests.*
II. *LSLAs targeting forest areas are associated with increased deforestation.*
III. *Deforestation accelerates considerably more in the immediate vicinity of LSLA sites than in more distant areas after LSLA contract signing.*

My analysis focuses exclusively on Africa, as the continent is the most affected by LSLAs according to registered cases (Cotula, 2009; Messerli et al., 2014). Furthermore, the relatively high prevalence of tenure insecurity and weak governance systems in certain regions of Africa may exacerbate the potential negative impacts of LSLAs (German et al., 2013).

3 Methods and Data

3.1 Variables and Data Preparation

3.1.1 Large-Scale Land Acquisition Data

Data on LSLAs was obtained from the Land Matrix database, the most comprehensive and widely used database on land acquisition (The Land Matrix, 2023). The Land Matrix covers deals negotiated or concluded since 2000 that involve a transfer of ownership, control, or use of land by concession, lease, or sale and cover a contract area of more than 200 hectares. The database includes deals for carbon trading, industry, conservation, tourism, extraction, timber plantations and agricultural production (Lay, Anseeuw, et al., 2021). I restricted my analysis to land acquisitions within Africa for agricultural production. On the one hand, investments for agricultural production account for the vast majority of documented LSLAs in the Land Matrix; on the other hand, a separate analysis of agricultural LSLAs may be justified as the different types of land acquisitions have very heterogeneous impacts on land use due to their different investment purposes.

I considered agricultural LSLAs from both transnational and domestic investors. The original dataset contains information from 1012 transnational and 511 domestic deals from 38 African countries. However, most of these LSLAs have incomplete data, partly due to a lack of transparency driven by a high degree of secrecy in many of these transactions (Scoones et al., 2013). I proceed as follows: I kept only those LSLAs for which exact coordinates are available (163 LSLAs). Then, I filtered the LSLAs for which a post-2003 contract date is available. For all LSLAs that do not have a contract date, I tried to reconstruct it using the information and the original data source provided in the Land Matrix.

To increase the internal validity of this study, I used Google Earth Pro's timelapse to determine whether I can identify LSLA production patterns well before the contract date listed in the Land Matrix. While it is conceivable that large-scale land use changes (e.g. clearing of forest area) are already made in anticipation of the signing of the contract, it seems rather implausible to me that production starts before the contract is signed. Therefore, I filtered out those LSLAs for which production patterns do not match the contract date. However, this means that I also filtered out LSLAs where an ownership change (e.g. from a state-owned to a privately operated large-scale farm) has occurred, as according to satellite data inspection, the contract date of ownership change may not be verifiable. The exclusion of these LSLAs may limit the external validity of the results in that the pre-LSLA land use and the land use changes identified in this study may not apply to LSLAs located in areas with former large-scale agriculture.

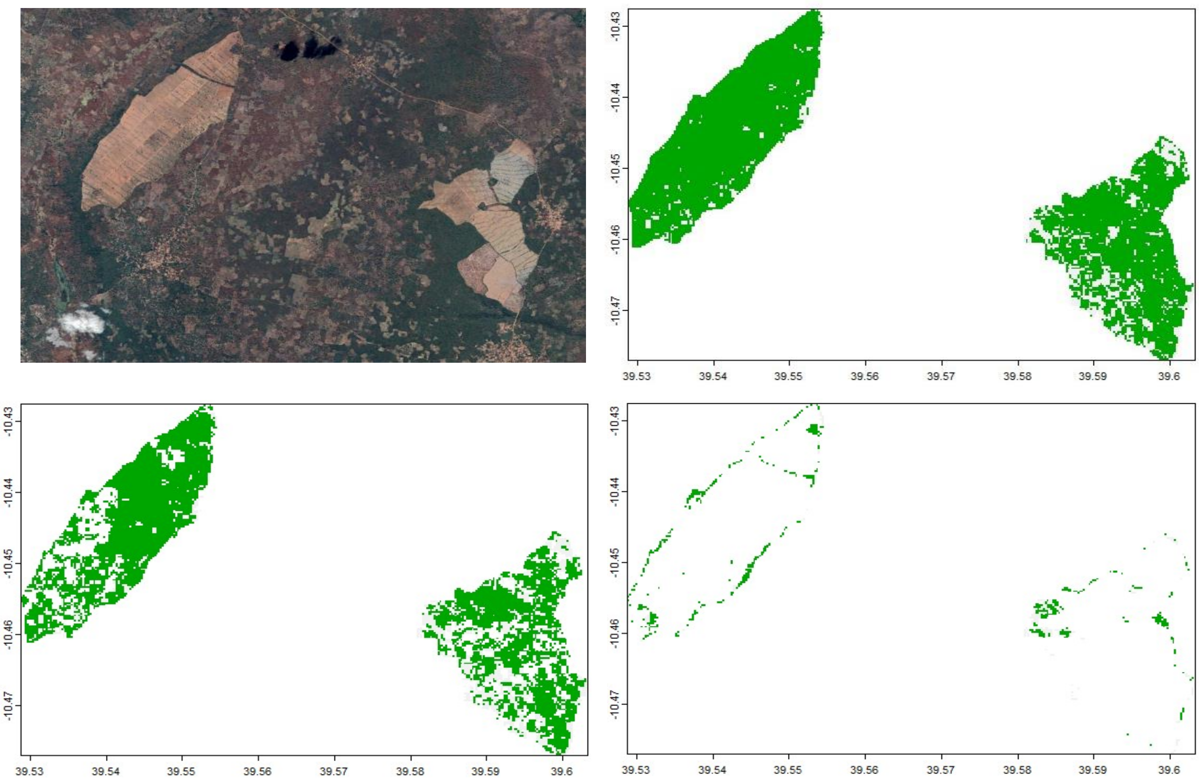
Since the Land Matrix only lists one coordinate per LSLA, often corresponding to the location of a main building within the LSLA or a road (Messerli et al., 2014), and thus does not list the exact boundaries of the acquired land, I approximated the LSLA contract area with a disc. Thus, for the remaining 75 LSLAs, I converted the reported coordinates into geospatial data and approximated the spatial extent by a disc with an area equal to the reported current contract size in the Land Matrix.

3.1.2 Data on Cropland and Forest Extent

To analyse the pre-LSLA land use, I calculated the proportion of cropland and forest within the disc before the LSLA contract date. The findings are reported in *section 4.1*. For the analysis of cropland, I relied on a novel dataset by Potapov et al. (2022), which is the first globally consistent time series of cropland at a 30-metre resolution to date. The dataset defines cropland as land used to cultivate annual and perennial herbaceous plants for human consumption, biofuel, and forage (including hay). Permanent pastures, perennial woody crops, and shifting cultivation are excluded from the definition. Potapov et al. (2022) mapped global cropland in five 4-year intervals from 2000 to 2019. To assess the pre-LSLA cropland extent within the approximated contract area of the LSLA, I analysed the cropland from the 4-year period prior to the LSLA contract date. For example, if the LSLA contract was signed between 2004 and 2007, I analysed cropland identified during the period 2000-2003. Assuming that no LSLA started production before the contract date (which I tried to verify with Google Earth Timelapse), this procedure minimises the risk of LSLA cropland being misclassified as pre-LSLA cropland.

To analyse pre-LSLA forest extent within the approximated contract area, I used the recently updated dataset of Hansen et al. (2013), which also has a resolution of 30 metres based on Landsat data. The dataset includes the tree cover extent from 2000 and raster data for annual tree cover loss until 2022. Following Hansen et al. (2013), I classified a pixel as originally forested in 2000 if the tree cover is above 50%. To analyse forest extent in the year prior to the signing of the LSLA agreement, I subtracted the annual loss of tree cover from the tree cover in 2000 for all years up to the year prior to the reported LSLA contract date. Afforestation may have also occurred within the approximated LSLA area between 2000 and the year before the LSLA contract date. However, data on afforestation are only available for forest increments between 2000 and 2012. As no annual forest growth data are available, the calculated annual forest extent should be regarded as a lower bound for the actual forest area.

***Figure 2**: From top left to bottom right: satellite image from 2019, forest extent 2001, 2015, and 2019 of LSLA ID7786*



Due to the exclusion of permanent pasture and shifting cultivation in Potapov et al.'s (2022) cropland mapping and the concurrent lack of other consistent time series data on such anthropogenic land uses, I cannot document all agricultural land use within and near the LSLA. This lack of data is a limitation of my analysis, as other land uses, such as pasture, can also trigger indirect land use change. For instance, studies on deforestation of the Amazon rainforest suggest that mechanised agriculture may cause significant indirect land use change by displacing old pastures, potentially spreading to forested areas (Barona et al., 2010; Lapola et al., 2010). Therefore, while I can measure deforestation trends within and near LSLA production areas, I cannot directly attribute them to alternative land uses, such as pasture expansion. The same challenge arises when analysing palm oil and other perennial woody crops, as they do not fall under the dataset's definition of cropland.

3.1.3 Creation of Shapefiles

The previously created disc serves only as an approximation of the actual LSLA contract area. The disc could include parts of the actual LSLA but also non-LSLA areas. Although non-LSLA areas within the approximated LSLA contract area could also be affected by the acquisition (see *section 2.2.2*), I need more precise information on the spatial extent of the LSLAs to better answer the question as of whether the LSLAs in the sample targeted pre-existing cropland and forest land. I addressed this challenge as follows: first, I used the spatial polygons created as part of the Land Matrix 2020 Mapathon (The Land Matrix, 2023). These polygons represent the spatial delineation of some LSLA areas based on visually identified production areas or contract areas identified using information from contract details (Lay, Anseeuw, et al., 2021). I analysed every single polygon for the previously filtered 75 LSLAs and kept only those polygons covering the full LSLA production or contract area from the contract date until 2020. For most LSLAs, polygons exist only for the production area but not for the contract area. I adjusted the polygons if I detected an extension of the production area beyond the contract area that I could assign to the LSLA based on the same production pattern visible on satellite images. For certain LSLAs, the production area may be approximately the same as the contract area. However, for other LSLAs, the contract area may exceed the production area. For example, the Roundtable on Sustainable Palm Oil (RSPO) certification requires that, according to Principle 7, land that contains or supports High Conservation Values (HCVs) may not be cleared for production (RSPO, 2018).

Next, in a time-consuming task, I created shapefiles for those LSLAs for which no spatial polygons exist. I only created shapefiles in Google Earth Pro for LSLAs that are clearly identifiable and whose extent of the production area can be clearly delineated using objective criteria. Suppose there are several large-scale agricultural production areas with similar visually recognisable structures (e.g. large-scale palm oil plantation) in the vicinity of the production area containing the coordinate listed in the Land Matrix. In that case, I only created a shapefile for the areas connected to the area containing the coordinate (e.g. by harvest road networks). This approach reduced the risk of incorrectly assigning production areas to an LSLA that does not belong to it. However, there is a risk that I did not identify the entire production area of an LSLA with this method. I tried to address this challenge by using Google Earth Timelapse and the temporal reconstruction of the visual emergence of the LSLA to analyse which production areas with similar production patterns emerged at about the same time as the area containing the coordinate. I acknowledge that this could still lead to incorrect allocation of production areas to LSLAs implemented in the same year or that LSLA production areas implemented later are not allocated correctly. Therefore, if the allocation of production areas to an LSLA is still unclear with the described approach, no shapefile for the respective LSLA was created.

To increase the objectivity of this spatial delineation work, I also relied on the dataset of Descals et al. (2021). This dataset identifies smallholder and industrial closed-canopy oil palm plantations for 2019 on a global scale based on Deep Learning with Sentinel-1 and Sentinel-2 imagery at a resolution of 10 metres. Based on the palm oil production pixels identified by this dataset, I converted these pixels into polygons, which I loaded into Google Earth Pro. There, I combined the polygons that form the production area into a shapefile and, if necessary, improved the exact extent of the plantation area.

With this described procedure, 52 polygons are available for pre-LSLA land use analysis. The proportions of pre-LSLA cropland and forest pixels concerning the area of the respective polygon are reported in *section 4.1*.

3.1.4 Treatment and Control Areas

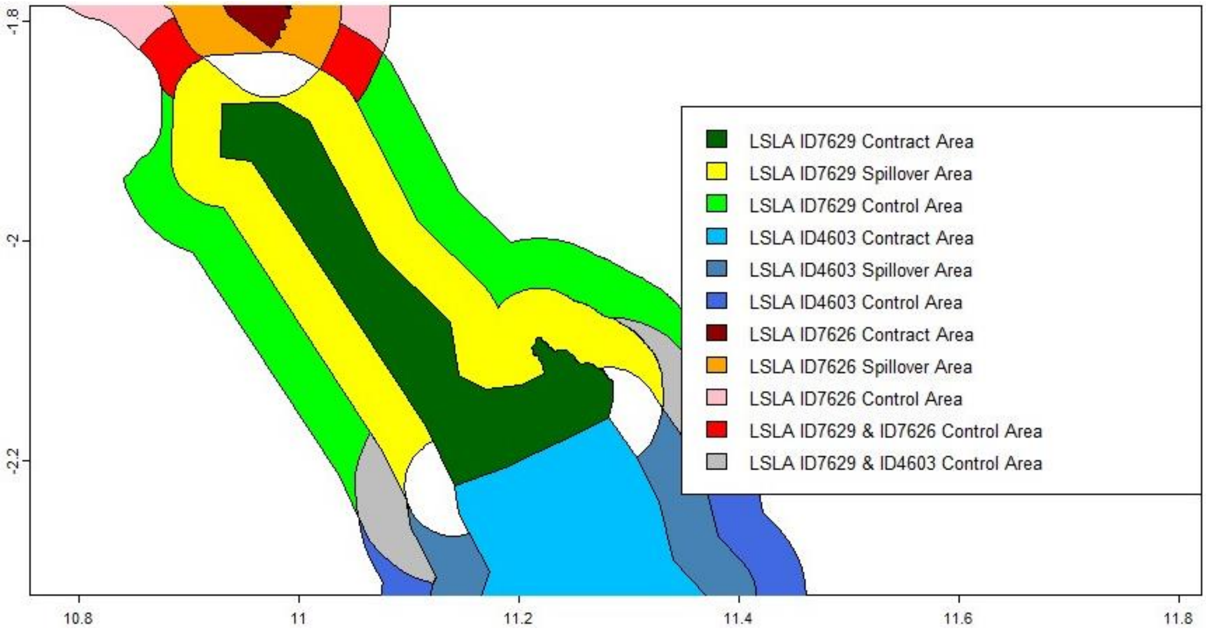
For the time series analysis of deforestation within and near LSLAs, I defined the control group as those areas between 5 and 10 km from the borders of the LSLA production/contract area. Based on case study findings, potential local indirect land use changes related to LSLAs tend to be more pronounced near LSLAs. Therefore, control areas should include areas that are not affected (directly or indirectly) by the LSLA. In light of the findings from a case study by Zaehringer et al. (2018), which identified indirect land-use changes within a 5-kilometre buffer zone of an LSLA in Mozambique, I, therefore, defined the area that may be indirectly affected by the LSLA as the area that lies within 5 kilometres of the LSLA boundaries. However, such a spatial division of the area into a "local spillover" area and a control area is arbitrary and case-specific, as the extent of potential local indirect land use change depends on numerous factors such as the pre-LSLA land use or the areal extent of the LSLA.

On the one hand, due to the physical proximity to the LSLA, this buffer-like division into spillover and control areas allows for better isolation of potential region-related confounding factors. On the other hand, incorrect spatial classification of the area where local indirect land use changes may occur leads to biased estimates of the "indirect" treatment effect. Suppose the 5-km "spillover" buffer zone is too narrow. In that case, the control area may also experience potential "indirect" treatment effects so that the change in the control area no longer reveals the counterfactual trend. If, on the other hand, the 5 km spillover buffer zone is too large, the estimates of the "indirect" treatment effect will be attenuated towards zero. I am aware that different buffer zones do not necessarily solve the described challenge, as different buffer zones can also lead to biased estimates and thus possibly falsely increase confidence in the main analysis with the spatial subdivision of 0-5-10 km. An alternative estimate that would avoid a manual spatial division into treatment and control areas would be one where a treatment effect curve would be measured as a distance function, using many rings rather than just one ring to measure the treatment effect (Butts, 2023). However, I did not use such a method due to limited time resources.

To reduce the risk that the estimated changes in forest area and cropland within the 0-5 and 5-10km buffer zone of an LSLA are partly driven by direct or indirect land use changes associated with other large-scale land acquisitions, I proceeded as follows:

First, based on the information in the Land Matrix, I grouped LSLAs that belong to the same land acquisition but are listed individually in the database. Second, for all LSLAs for which I have spatial polygons, I only considered the area within the 0-5km and 5-10km buffer zones that do not overlap with the other filtered agricultural LSLA areas and their 0-5km buffer zones. Assuming that potential local indirect land use changes only extend within 5 km of the LSLA production/contract area, I thus minimised the risk that the control area of one LSLA is also the “spillover area” of another LSLA and vice versa. An example of such a spatial subdivision for three LSLAs in Gabon, their “spillover” and control areas is shown in *Figure 3*. For tracking purposes, the LSLAs have the same IDs as reported in the Land Matrix.

Figure 3: *Example of Spatial Division into Contract, Spillover, and Control Areas for 3 Adjacent LSLAs*



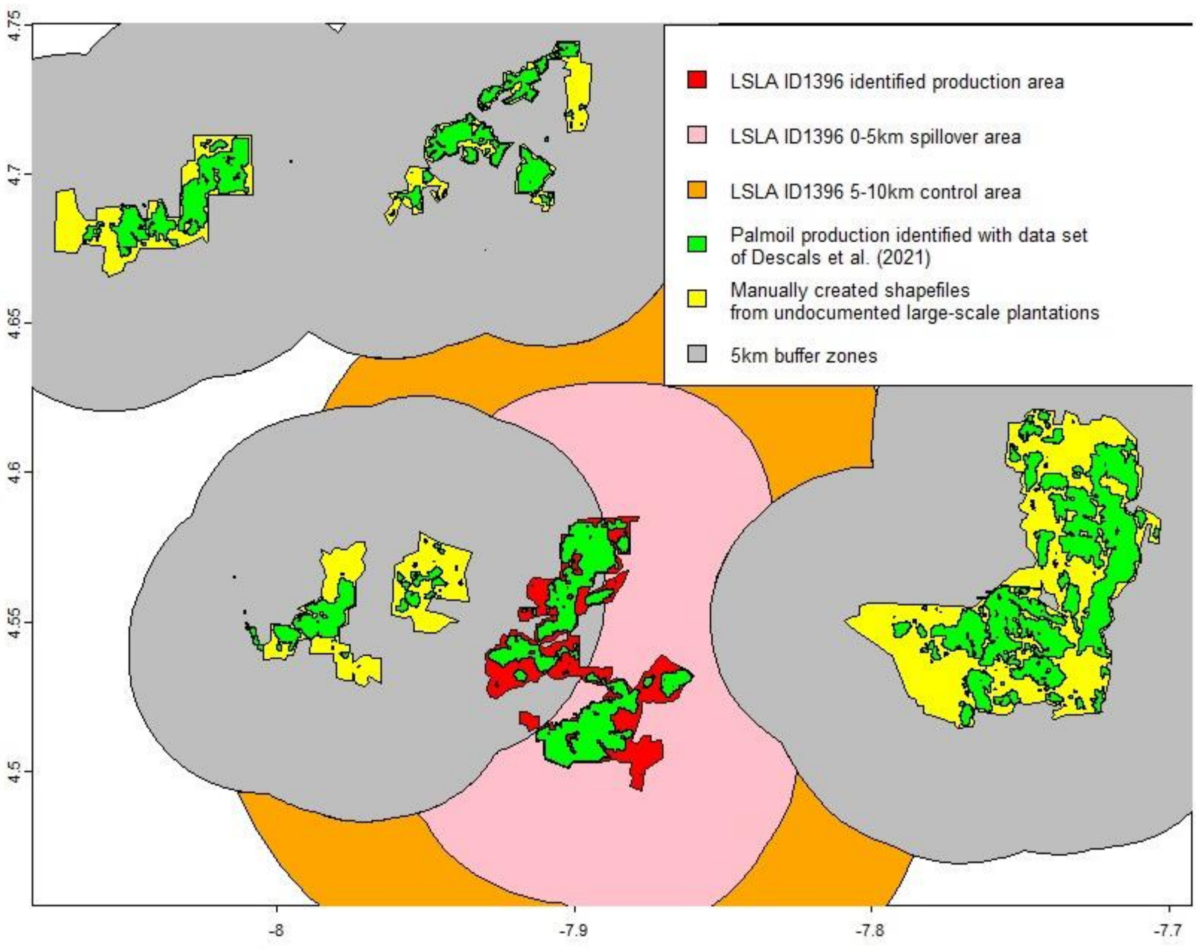
Notes: Areas are projected in longitude/latitude relative to the WGS-84 datum.

Third, all LSLAs with production/contract area shapefiles were removed from the time-series analysis if their 10-km buffer zones overlap with the contract area approximated by a disc of those agricultural LSLAs for which I could not create a production/contract area shapefile. I did this to minimise the risk of changes in forest and cropland areas near an LSLA being directly or indirectly driven by other LSLAs with unknown production/contract area extent.

Fourth, non-agricultural LSLAs, such as timber plantations, may also affect land use change within the 10km buffer zone of the agricultural LSLAs. I proceeded as explained in step two for those non-agricultural LSLAs for which I have shapefiles (345 LSLAs). For those non-agricultural LSLAs for which I only have coordinates (289 LSLAs), I created discs whose area corresponds to the reported contract area listed in the Land Matrix. If the 10km buffer around the agricultural LSLAs overlaps with these discs of the non-agricultural LSLAs, I removed the agricultural LSLAs from the time-series analysis.

Fifth, I assumed that LSLAs in the vicinity of other reported LSLAs have a relatively high probability of being discovered on field visits and subsequently included in the Land Matrix database. However, for LSLAs that are geodetically close to each other but may not be easily visible from ground level (e.g. due to the presence of forests), there may be a greater risk that these LSLAs will not be recorded during field visits. I, therefore, tried to identify potentially large-scale production areas in the vicinity for all filtered agricultural LSLAs using satellite imagery. The areas I identified are primarily large-scale palm oil plantations within a radius of 40 kilometres from the filtered LSLAs. I again used the dataset of Descals et al. (2021) to identify commercial large-scale palm oil plantations and then used the identified pixels to create shapefiles in Google Earth Pro for all palm oil plantations near the filtered LSLAs (see *Figure 4*). Subsequently, I created a 5km buffer zone around these palm oil plantations and erased parts of the 0-5 and 5-10km buffer zones of the filtered agricultural LSLAs that overlap with these palm oil plantations or their 5km buffer zones. I erased these areas as the counterfactual represents what would have happened to the treatment area without the LSLAs. While I do not know what would have happened in the treatment area without the LSLA, I assumed that at least no other large-scale agricultural production would be active in the treatment area. 4 of the remaining LSLAs that do not

Figure 4: *Identification of palm oil plantations and spatial division of LSLA ID1396 in Liberia into production, spillover, and control area (projected in longitude/latitude relative to the WGS-84 datum)*



target forest land were also removed. After the steps described above, 20 LSLAs, "spillover" and control areas each remain for the time series analysis of deforestation.

3.2 Empirical Strategy

An ideal experiment to estimate the causal effect of LSLAs on land use change within and near the acquisition would be to randomly allocate LSLAs to different areas, and the resulting land use change could then be compared to a counterfactual group of areas that did not receive LSLAs. In reality, however, the non-random assignment of LSLAs poses the problem of endogeneity, i.e. reverse causality and omitted variable bias. For example, it is conceivable that the cause-effect relationship is not only in the direction of LSLAs to deforestation but that logging or clearing prior to and independent of the acquisition attracts LSLAs. Furthermore, omitted variables such as socio-economic factors, government policies, availability and quality of infrastructure, or land tenure and property rights can influence both land use changes and the occurrence and location of LSLAs, making it difficult to obtain consistent and unbiased estimates of the effect of LSLAs on land use changes.

To better isolate possible location-related confounding factors, I compared the temporal changes in forest extent and cropland in the LSLA areas with control areas 5-10 km away from the LSLA areas (see *section 3.1.4*). I am aware that I cannot fully eliminate endogeneity with this strategy, as the chosen method of defining the control areas may not fully account for other confounding factors that may not be related to proximity. An alternative approach would be matching, which is regularly used in impact assessments of land use policies (Nelson & Chomitz, 2011; Nolte & Agrawal, 2013). However, I did not use such a method due to a lack of locally disaggregated data, especially for socio-economic factors. Furthermore, while remote sensing is an important tool in the analysis of land use changes (Ariti et al., 2015; Scharsich et al., 2017), it is difficult to identify causal relationships between LSLAs and land use changes without collecting more detailed information on the ground (Zaehringer et al., 2018). Triangulation with other data sources, including qualitative research such as interviews with local actors, would be necessary for a more comprehensive understanding of the drivers and impacts of land use changes caused by LSLAs (see *section 5*). Due to the unavailability of these data, I only analysed the association between LSLAs and land use changes without inferring causality.

3.3 Estimation Approach

3.3.1 Difference-in-Differences with Multiple Time Periods

In the panel dataset covering forest mapping for the period from 2000 to 2019, the areas receive the treatment - the LSLA - at different times. The leading approach for estimating the treatment effect for staggered treatment adoption has been a two-way fixed effects (TWFE) linear regression. However, several papers, such as Callaway & Sant'Anna (2021), have pointed out the methodological drawbacks of a TWFE linear regression in the case of multiple time periods, as TWFE is not robust to treatment effect heterogeneity. In the case of LSLAs, treatment effects may be dynamic, and the timing of LSLA implementation varies across regions. Therefore, I based the time-series analysis of deforestation on Callaway & Sant'Anna's novel difference-in-differences method with multiple time periods. I follow the notation of Callaway & Sant'Anna (2022) and define the outcome Y_{it} as the extent of forest at time t in area i, and the treatment period G_i as the period containing the LSLA contract signing date:

- $\boldsymbol{Y_{it}(0)}$ is the untreated potential outcome of area i. It represents the outcome for area i that it would experience in period t if it does not receive the treatment.
- $\boldsymbol{Y_{it}(g)}$ represents area i's potential outcome in period t if it receives treatment in period g.
- $\boldsymbol{G_i}$ represents the time period in which area i first becomes treated.
- $\boldsymbol{C_i}$ is a binary variable which takes the value 1 if area i never participates in the treatment.
- $\boldsymbol{D_{it}}$ is a binary variable that takes the value 1 when area i has been treated by time t.
- $\boldsymbol{Y_{it}}$ represents the observed outcome in time period t for area i. For those areas that never receive the treatment, $\mathbf{Y_{it} = Y_{it}(0)}$ in all time periods. For those areas that receive the treatment, the following applies: $\mathbf{Y_{it} = 1\{G_i > t\}Y_{it}(0) + 1\{G_i \leq t\}Y_{it}(G_i)}$. Assuming that there is no treatment anticipation, this notation implies that untreated potential outcomes are observed for areas that have not yet received treatment and treated potential outcomes are observed for areas that have received treatment.
- $\mathbf{X_i}$ includes the different regions as pre-treatment covariates, where a region is defined as the area that includes the LSLA, an adjacent "spillover" and a control area. It is assumed that only areas close to an LSLA would follow the same deforestation trend without the LSLA.

3.3.2 Conditional Parallel Trends Assumption

As an extension of the parallel trends assumption in a simple difference-in-differences with two periods and two groups, the parallel trends assumption for the difference-in-differences with multiple periods based on never-treated areas states:

For all $\boldsymbol{g} = \mathbf{2,\ldots, T}$, $\boldsymbol{t} = \mathbf{2,\ldots, T}$ with $\boldsymbol{t \geq g}$:

$$\mathbf{E[Y_t(0) - Y_{t-1}(0) \mid X, G=g] = E[Y_t(0) - Y_{t-1}(0) \mid X, C = 1]}$$

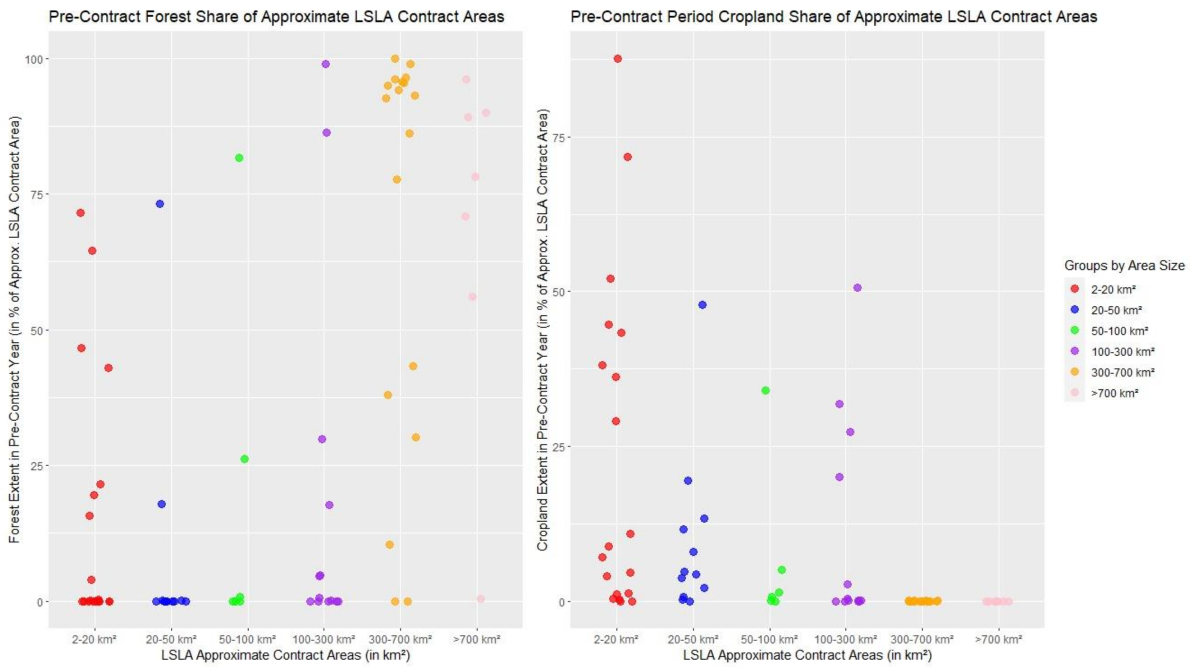
Thus, conditional on covariates, it is assumed that in the absence of treatment, the average untreated potential outcomes for the areas receiving treatment in time g and the areas never receiving treatment would have followed parallel trends in all post-treatment periods $t \geq g$.

4 Results & Discussion

4.1 Pre-LSLA Land Cover Analysis

The left-hand plot in *Figure 5* represents the percentage of identified forest pixels in the approximated documented contract area of the LSLAs in the year before the LSLA contract was signed. The 75 approximated contract areas are divided into different groups according to their area size. For the LSLAs with a documented contract area of 2-20 square kilometres, about 26% of all the approximated contract areas have a forest share of more than 20%, and just under 10% have a forest share of more than 50%. The proportion of pre-LSLA forest in the approximated contract area is

Figure 5: *Proportion of forest and cropland within approximated pre-LSLA contract areas*



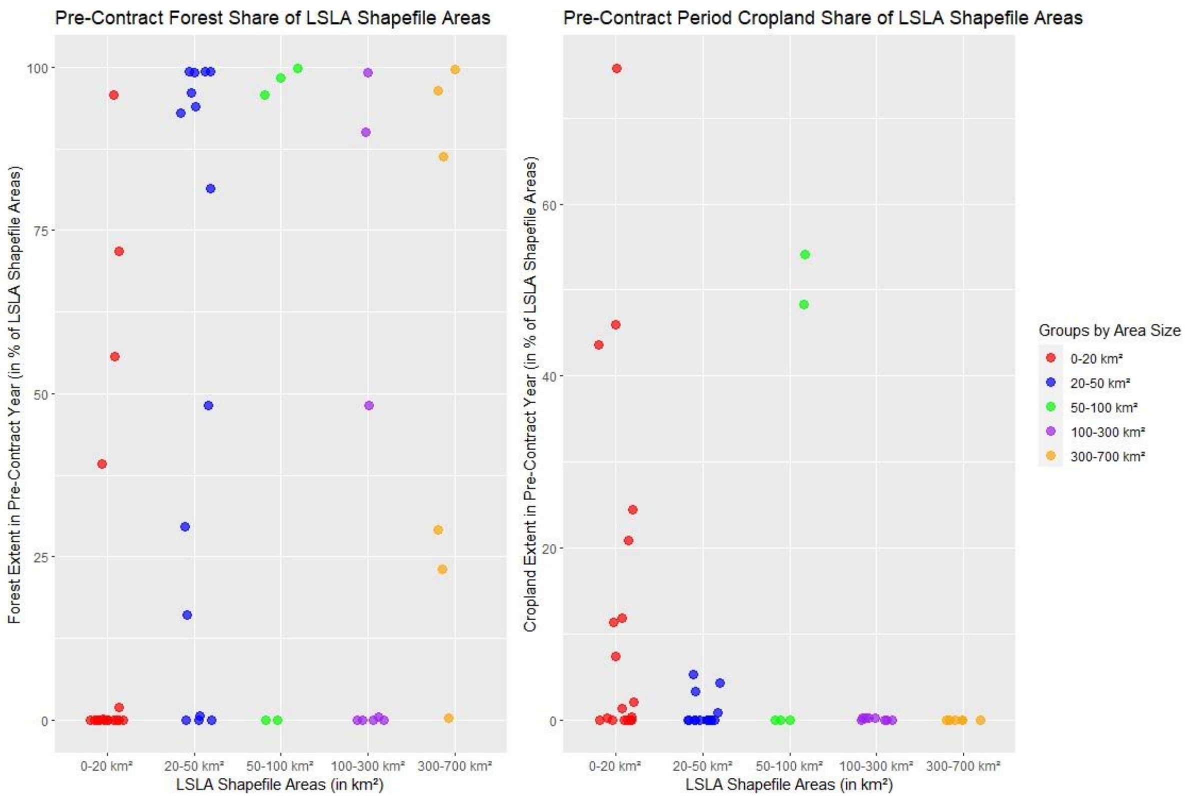
particularly high in the group with a contract area of 300-700 square kilometres, where around 66% of the areas have more than 50% forest cover. The relative share of pre-LSLA forest in the approximated contract area tends to increase with the size of the contract area, which can be partly explained by the fact that large plantations such as palm oil are often located in forest areas.

The right-hand plot in *Figure 5* represents the percentage of identified cropland pixels in the approximated contract area of LSLAs in the 4-year period prior to the LSLA contract date. For the LSLAs with a documented contract area of 2-20 square kilometres, I found that about 42% of all the contract areas have a pre-LSLA cropland share of more than 20%, and around 16% of the areas have a cropland share of more than 50%. The proportion of identified pre-LSLA cropland pixels decreases with the approximate LSLA contract area size.

Figure 6 again shows the relative proportion of pre-LSLA forest and cropland pixels within the identified LSLA production/contract area. The different distribution in the groups is because only 52 LSLAs are analysed, whose areas may differ in extent and spatial form from the contract areas approximated by a buffer. The graph on the left shows that 50% of all identified areas have a forest share of more than 10%, and about 36.5% have a forest share of more than 50%. The right plot of *Figure 6* shows that smaller LSLAs tend to have a higher pre-LSLA cropland share. This phenomenon can again be partly explained by the fact that larger LSLAs, such as palm oil production, more often target forest areas where it is less likely that cropland already exists.

Overall, the results indicate that a high proportion of LSLAs affect areas with pre-existing forest and cropland. My findings are thus consistent with the findings of Messerli et al. (2014) and various case studies that challenge the rhetoric that LSLAs mainly target "unused land".

Figure 6: *Proportion of pre-LSLA forest and cropland within shapefiles*



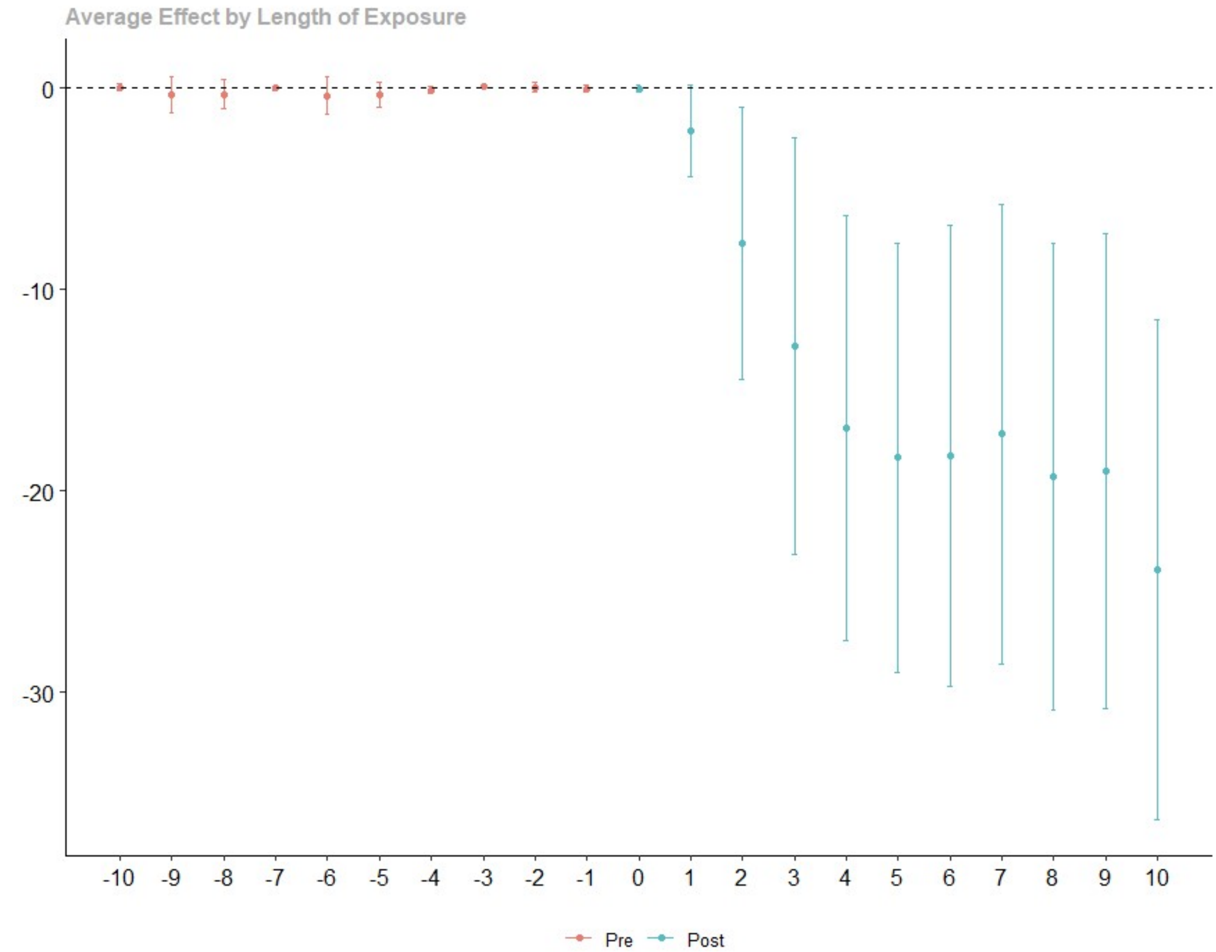
4.2 Difference-in-Differences Results

4.2.1 Direct Land Use Change

In the following, the results of the difference-in-differences analysis for the relationship between LSLA and deforestation are discussed. In the conditional parallel trend assumption, it is assumed that only areas with the same characteristics would follow the same trend in deforestation in the absence of the LSLA. In this case, I assumed that LSLAs and their control area nearby have the same characteristics. However, factors that influence deforestation in the LSLA but are independent of the LSLA itself may not be comparable to those in the control area, but this would be necessary for a causal interpretation. Thus, even in the vicinity of an LSLA, environmental regulations, land tenure systems, socio-economic conditions or market access may be different. As robustness checks, I, therefore, compared changes in forest extent in the treatment area with changes in forest extent in control areas that are in the same ADM1 or ADM2 region as the LSLA and do not contain protected areas or built-up expansion areas (see *Robustness Checks 8.1.1-8.1.5)*.

Figure 7 shows the average effect of participation in the treatment *t* time periods after receiving the treatment for all areas that have experienced LSLA presence for exactly *t* time periods. The red lines represent the point estimates and 95% confidence intervals for the pre-treatment periods. The blue lines are the point estimates and 95% confidence intervals for the treatment effect. Clustered bootstrapped standard errors at the area level were used in all inference procedures, and autocorrelation of the data was considered.

Figure 7: *Event Study Based on Table 1*



I fail to reject the null hypothesis of no relationship between LSLA and forest area in the year of contract signing and the following year. From the third year after the event, the relationship between LSLA and forest extent appears negative, increases in magnitude, and is statistically significant at the 5% significance level. In the third year in which an LSLA is present, forest area is estimated to decrease by 7.7 percentage points more than in the control area; by 12.8 percentage points in the fourth year; by 16.9 percentage points in the fifth year; and by 18.4 percentage points in the sixth year. I find an overall treatment effect parameter - an average across groups over the average treatment effect for each group - of -14.6 percentage points, significant at the 5% level (see *Table 1*).

I fail to reject the null hypothesis for the conditional parallel trends assumption that there are, on average, no significant differences in forest extent trends between the treatment and control group prior to the treatment. Thus, the results suggest that, on average, the groups followed parallel conditional trends in forest extent before treatment so that post-treatment differences in outcomes between the treatment and control groups can be attributed more confidently to the treatment itself and not to pre-existing trends.

Table 1: *LSLA Aggregated Treatment Effect Estimates*

			95% CI	
Type	**ATT**	**Std. Error**	**LL**	**UL**
Simple	-14.5738	3.678	-21.7826	-7.365*
Dynamic	-14.1507	3.5639	-21.1358	-7.1657*
Event Time	**Estimate**	**Std. Error**	**LL**	**UL**
-5	-0.3230	0.2751	-0.9460	0.3000
-4	-0.1002	0.0808	-0.2831	0.0827
-3	0.0633	0.0349	-0.0156	0.1422
-2	0.0319	0.1010	-0.1969	0.2606
-1	-0.0266	0.0839	-0.2166	0.1635
0	-0.0255	0.0842	-0.2161	0.1651
1	-2.1416	1.0025	-4.4116	0.1283
2	-7.7038	2.9817	-14.4554	-0.9523*
3	-12.8323	4.5619	-23.1617	-2.5029*
4	-16.8748	4.6645	-27.4366	-6.3130*
5	-18.3638	4.7042	-29.0154	-7.7122*

Number of observations: *920*
Units of observations: *40*

*Notes: The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the ATT in a two-group two-period difference-in-differences. The row "Dynamic" reports a cross-group average of the average treatment effect for each group, where groups are defined by the length of exposure. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend, with clustering at the area level and using the doubly robust estimator (*see *Callaway & Sant'Anna, 2021). The control group consists of never treated units. 0 anticipation periods were assumed. '*' indicates that the 95% confidence band does not cover 0. The complete event study can be found in Appendix 8.3.*

The finding of a significant negative association between LSLA and forest area extent only from the third year onwards is consistent with the literature, which states that LSLA implementation in Africa is, on average, significantly slower than in Central Asia or Latin America, for example. This circumstance is partly explained by the fact that the land in the latter areas is often already "ready for cultivation", as the necessary infrastructure may already be in place (Lay, Anseeuw, et al., 2021).

4.2.2 Land Use Change Near LSLAs

To examine potential indirect land use change, I compared changes in forest extent between the 0-5km and 5-10km buffer zones around the LSLA. I found an overall treatment effect parameter of -0.13 percentage points (see *Appendix 8.3*). However, the null hypothesis that deforestation accelerates significantly more near LSLA sites than in more distant areas cannot be rejected, as the described relationship is not statistically significant.

As discussed in *section 2.2.2*, the presence of certain factors such as the displacement effect is needed to trigger indirect land use change. However, it remains unclear to what extent these elements are present in the LSLAs I have analysed. Subsequently, deforestation trends are examined exclusively for LSLAs, in which information on displacement is ascertainable. However, given the lack of information in the Land Matrix on the presence of displacement (let alone its temporal and spatial extent), the number of LSLAs available for analysis is so small that the difference-in-differences method reached its limitations, so no results are reported.

5 Limitations and Avenues for Future Research

The main limitation encountered in this dissertation is the significant lack of reliable data on land acquisitions, their spatial and temporal extent, and the different local land uses coupled with detailed local information required for attributing (indirect) land use changes to their causes. Several inconsistencies were found in the systematic probing of parts of the data listed in the Land Matrix, especially the contract dates. I tried to reduce uncertainty about the actual size of LSLA areas by analysing only the changes in forest extent areas of LSLAs whose production areas I could identify from satellite data or whose contract area coordinates are listed in verifiable sources. However, this method is also subject to corresponding uncertainties. Certain control and “spillover” areas of identified LSLA production areas could, for example, include parts of the contract area of the respective LSLA. At the same time, changes in control and “spillover” areas might be influenced by unobserved spatial confounders such as unidentified LSLAs or other factors. Given these uncertainties, largely due to the lack of transparency in many of these land acquisitions and the associated limitations in data collection, the findings of this dissertation should be viewed with appropriate caution.

My method only analysed LSLAs that could be spatially delineated from their surroundings by visual inspection or information from contracts. Therefore, it remains to be seen to what extent the identified land use changes also apply to the documented LSLAs that could not be delineated. The same applies to the LSLAs that could be delineated, but their 10-km buffer zones overlapped with a buffer of an LSLA of unknown spatial extent. In addition, I am aware of the possible selection bias affecting the

Land Matrix database itself. For example, media and research interest and the presence or absence of Land Matrix reporting networks in certain regions can bias reporting. For example, deals by foreign investors generally create more attention than domestic investors and are, therefore, reported more frequently in relative terms (The Land Matrix, 2023). Thus, Oya (2013) correctly argues that the Land Matrix acts as a "biased sampling frame" of the total unknown population of LSLAs, which may lead to a focus in the analysis only on certain land acquisitions while excluding potentially similarly relevant ones. These circumstances must be taken into account when considering the external validity of my results, as potentially the sample I analysed is not representative of the total unknown population of LSLAs.

In the dataset of Potapov et al. (2022), the heterogeneous landscapes in regions such as Africa, Europe and Asia result in lower regional accuracy of cropland mapping compared to the Americas, where large-scale industrial agriculture dominates. In particular, Potapov et al.'s (2022) machine learning has some difficulties mapping changes in cropland. Local classification and change analysis would be required to increase the accuracy of cropland mapping. Local classification would also be required to identify other land covers and land uses such as pastoralism, shifting cultivation and forestry use. However, remote sensing-based land use classifications may only be of limited help, as it is difficult to identify such land use types based on satellite imagery alone. Due to my inexperience with geospatial data analysis, I had to spend significant time within the limited time frame to acquire the relevant analytical skills. Therefore, learning additional skills for local land classification, for example, with Google Earth Engine cloud computing, and travelling to the respective locations, would have been beyond an MSc dissertation's time and financial scope.

Assumptions about the spatial extent of potential indirect land use changes were made based on a single case study of Zaehringer et al. (2018), as research about local indirect land use changes of LSLAs is scarce. The arbitrary spatial allocation of "spillover" and control areas is associated with corresponding problems (see *section 3.1.4*). As little information is available on local impacts on communities and their livelihoods, such as whether displacement has occurred, it is difficult to analyse land use changes near LSLAs separately for those areas where such local impacts may have occurred and for those areas where they have not.

Furthermore, in-depth local information would also help identify possible treatment anticipations and subsequently include them in the analysis. The difference-in-differences with multiple time periods discussed in section *3.3.1* was conducted under the assumption that there is no treatment anticipation, an assumption that should be critically questioned. For example, villagers might clear land in anticipation of future LSLAs to better assert possible compensation or land claims, as Lamb et al. (2017) briefly describe in the case of a region in Cambodia. However, the lack of information makes it difficult to make a plausible assumption about the temporal extent of such anticipatory behaviour, so no anticipation was assumed.

Future research may follow a mixed-method approach and combine remote sensing with more detailed information on the ground, for instance, through interviews with local actors. Furthermore, as this dissertation has primarily focused on local changes, future research could also analyse possible general equilibrium effects of LSLAs, such as the impact of LSLA-induced out-migration in urban and rural areas, instead of only examining changes in areas adjacent to LSLAs.

6 Conclusion and Policy Implications

Research on LSLAs faces significant challenges, notably the lack of reliable and complete data. Therefore, most research on LSLAs relies on case studies, which have been criticised for their specificity and difficulty in achieving external validity. Concurrently, some cross-sectional analyses of LSLAs have faced critique due to their analytical and methodological limitations. This study tried to mitigate some of these analytical and methodological challenges by intensively checking the data listed in the Land Matrix.

In a cross-sectional analysis of LSLAs in Africa, I used forest and cropland mapping datasets to investigate how often and to what extent LSLAs target areas with pre-existing cropland and forests. For the LSLA contract areas, which I approximated by buffer zones, I found that around 47% of these buffers, irrespective of the size of the contract area, had a forest cover of more than 25% in the year before the contract date. In contrast, the share of identified pre-LSLA cropland in the approximate contract area appears to be higher for smaller LSLAs than for larger LSLAs. I found a similar distribution of the proportion of cropland and forest across the different area sizes of the LSLAs when I instead used shapefiles derived either from contract data or by visual identification of the production areas. The results align with the findings of Messerli et al. (2014) and various case studies that have challenged the rhetoric that LSLAs mainly target "idle land".

Next, using the novel difference-in-differences method of Callaway & Sant'Anna (2021), I compared the temporal change in forest extent inside and near LSLAs with a control area 5-10 km from the identified LSLA area. The results indicate that, on average, deforestation rates inside LSLAs have increased significantly more than in control areas. However, I only find a significant effect, on average, from the third year after signing the contract. This lagged treatment effect is consistent with the literature, which notes that actual implementation often occurs with a time lag to contract signing. Furthermore, to gain insight into possible indirect land use changes, I have compared the areas between 0- and 5 kilometres from the LSLAs with the control areas before and after the LSLA contracts were signed. However, I do not find a significant effect, which could be because there may be strong but undocumented heterogeneity in the LSLAs studied concerning the drivers of indirect land use change.

The findings have important policy implications. First, the Land Matrix Initiative and its open data approach are important sources for increasing land investment transparency and accountability.

However, the limited data shows that more transparency is needed concerning LSLAs and the stakeholders involved.

Second, the results support the adoption and implementation of sustainable land management practices in these land acquisitions to mitigate potential negative environmental impacts. Land conversions in the form of deforestation are associated with significant biodiversity losses, especially when tropical rainforests are affected (Davis et al., 2023; Drescher et al., 2016). Furthermore, forest conversion could directly exacerbate climate change by releasing larger amounts of carbon into the atmosphere (Liao et al., 2020). Following Lay, Anseeuw, et al. (2021), this paper argues that more transparency is needed in LSLA's environmental impact studies and concession contracts for environmental protection.

Third, the pre-LSLA land cover analysis showed that cropland has already pre-existed in many land acquisitions. Admittedly, my study is limited to cropland and forest mapping data, so alternative land uses, such as pastoralism, were not considered. Nevertheless, these findings highlight the need to consider competing land uses and interests in addition to cost-benefit factors in policy debates on LSLAs to mitigate potential negative socio-economic impacts.

Based on the above, this paper follows the recommendation of Lay, Anseeuw, et al. (2021), who highlight the need for comprehensive landscape plans to address potential trade-offs between economic, social and environmental objectives. Publicly accessible landscape plans that experts and stakeholders could independently review would reduce the lack of transparency of many of these LSLAs and increase accountability to all stakeholders involved.

7 References

Andrade de Sá, S., Palmer, C., & Di Falco, S. (2013). Dynamics of indirect land-use change: Empirical evidence from Brazil. *Journal of Environmental Economics and Management*, *65*(3), 377–393.

Arezki, R., Deininger, K., & Selod, H. (2015). What drives the global "land rush"? *The World Bank Economic Review*, *29*(2), 207–233.

Ariti, A. T., van Vliet, J., & Verburg, P. H. (2015). Land-use and land-cover changes in the Central Rift Valley of Ethiopia: Assessment of perception and adaptation of stakeholders. *Applied Geography*, *65*, 28–37.

Atkin, D., & Donaldson, D. (2015). *Who's getting globalized? The size and implications of intra-national trade costs*. National Bureau of Economic Research.

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), 024002.

Borras, S. M., Fig, D., & Suárez, S. M. (2011). The politics of agrofuels and mega-land and water deals: Insights from the ProCana case, Mozambique. *Review of African Political Economy*, *38*(128), 215–234.

Borras, S. M., & Franco, J. C. (2014). From threat to opportunity? Problems with codes of conduct for land grabbing. *Rethinking Food Systems: Structural Challenges, New Strategies and the Law*, 147–162.

Bottazzi, P., Crespo, D., Bangura, L. O., & Rist, S. (2018). Evaluating the livelihood impacts of a large-scale agricultural investment: Lessons from the case of a biofuel production company in northern Sierra Leone. *Land Use Policy*, *73*, 128–137.

Butts, K. (2023). JUE Insight: Difference-in-differences with geocoded microdata. *Journal of Urban Economics*, *133*, 103493.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230.

Callaway, B., & Sant'Anna, P. H. C. (2022, Juli 19). *Introduction to DiD with Multiple Time Periods*. https://bcallaway11.github.io/did/articles/multi-period-did.html

Clapp, J., & Isakson, S. R. (2018). Risky returns: The implications of financialization in the food system. *Development and Change*, *49*(2), 437–460.

Collier, P. (2008). The politics of hunger: How illusion and greed fan the food crisis. *Foreign Affairs*, 67–79.

Cotula, L. (2009). *Land grab or development opportunity?: Agricultural investment and international land deals in Africa*. Iied.

Cotula, L. (2013). The international political economy of the global land rush A critical appraisal of trends, scale, geography and drivers. In *The new enclosures: Critical perspectives on corporate land deals* (S. 31–62). Routledge.

Davis, K. F., Müller, M. F., Rulli, M. C., Tatlhego, M., Ali, S., Baggio, J. A., Dell'Angelo, J., Jung, S., Kehoe, L., & Niles, M. T. (2023). Transnational agricultural land acquisitions threaten biodiversity in the Global South. *Environmental Research Letters*, *18*(2), 024014.

Davis, K. F., Rulli, M. C., & D'Odorico, P. (2015). The global land rush and climate change. *Earth's Future*, *3*(8), 298–311.

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*(3), 178–181.

Deininger, K., & Byerlee, D. (2011). *Rising global interest in farmland: Can it yield sustainable and equitable benefits?* World Bank Publications.

Deininger, K., & Byerlee, D. (2012). The rise of large farms in land abundant countries: Do they have a future? *World development*, *40*(4), 701–714.

Deininger, K., & Xia, F. (2016). Quantifying spillover effects from large land-based investment: The case of Mozambique. *World Development*, *87*, 227–241.

Dell'Angelo, J., D'odorico, P., Rulli, M. C., & Marchand, P. (2017). The tragedy of the grabbed commons: Coercion and dispossession in the global land rush. *World Development*, *92*, 1–12.

Descals, A., Wich, S., Meijaard, E., Gaveau, D. L., Peedell, S., & Szantoi, Z. (2021). High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth System Science Data*, *13*(3), 1211–1231.

Drescher, J., Rembold, K., Allen, K., Beckschäfer, P., Buchori, D., Clough, Y., Faust, H., Fauzi, A. M., Gunawan, D., & Hertel, D. (2016). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1694), 20150275.

Edelman, M. (2013). Messy hectares: Questions about the epistemology of land grabbing data. *Journal of Peasant Studies*, *40*(3), 485–501.

Fairbairn, M. (2017). 'Like gold with yield': Evolving intersections between farmland and finance. In *New Directions in Agrarian Political Economy* (pp. 137–156). Routledge.

German, L., Schoneveld, G., & Mwangi, E. (2013). Contemporary processes of large-scale land acquisition in Sub-Saharan Africa: Legal deficiency or elite capture of the rule of law? *World Development*, *48*, 1–18.

Hall, D. (2013). Primitive accumulation, accumulation by dispossession and the global land grab. *Third World Quarterly*, *34*(9), 1582–1604.

Hall, R. (2011). Land grabbing in Southern Africa: The many faces of the investor rush. *Review of African political economy*, *38*(128), 193–214.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., & Loveland, T. R. (2013). High-resolution global maps of 21st-century forest cover change. *science*, *342*(6160), 850–853.

Hofman, P., Mokuwa, E., Richards, P., & Voors, M. (2019). Local economy effects of large-scale agricultural investments. *Biodiversity and Economics for Conservation (BIOECON)*.

IAASTD. (2009). *Agriculture at a Crossroads. International assessment of agricultural knowledge, science and technology for development (IAASTD). Sub-Saharan Africa (SSA) Report*. Washington.

https://www.globalagriculture.org/fileadmin/files/weltagrarbericht/IAASTDBerichte/SubglobalReportSubSaharanAfrica.pdf

ICPAC. (2017). *Africa—Admin Level 0* [dataset]. https://geoportal.icpac.net/layers/geonode%3Aafr_g2014_2013_0

Kleemann, L., & Thiele, R. (2015). Rural welfare implications of large-scale land acquisitions in Africa: A theoretical framework. *Economic Modelling*, *51*, 269–279.

Lamb, V., Schoenberger, L., Middleton, C., & Un, B. (2017). Gendered eviction, protest and recovery: A feminist political ecology engagement with land grabbing in rural Cambodia. *The Journal of Peasant Studies*, *44*(6), 1215–1234.

Lambin, E. F., & Meyfroidt, P. (2010). Land use transitions: Socio-ecological feedback versus socio-economic change. *Land use policy*, *27*(2), 108–118.

Lambin, E. F., & Meyfroidt, P. (2011). Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences*, *108*(9), 3465–3472.

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 Sciences*, *107*(8), 3388–3393.

Lay, J., Anseeuw, W., Eckert, S., Flachsbarth, I., Kubitza, C., Nolte, K., & Giger, M. (2021). *Taking stock of the global land rush: Few development benefits, many human and environmental risks. Analytical Report III*.

Lay, J., Nolte, K., & Sipangule, K. (2021). Large-scale farms in Zambia: Locational patterns and spillovers to smallholder agriculture. *World Development*, *140*, 105277.

Li, T. M. (2011). Centering labor in the land grab debate. *The Journal of Peasant Studies*, *38*(2), 281–298.

Liao, C., Jung, S., Brown, D. G., & Agrawal, A. (2020). Spatial patterns of large-scale land transactions and their potential socio-environmental outcomes in Cambodia, Ethiopia, Liberia, and Peru. *Land Degradation and Development*, *31*(10), 1241–1251. Scopus. https://doi.org/10.1002/ldr.3544

Liu, J., Hull, V., Batistella, M., DeFries, R., Dietz, T., Fu, F., Hertel, T. W., Izaurralde, R. C., Lambin, E. F., & Li, S. (2013). Framing sustainability in a telecoupled world. *Ecology and Society*, *18*(2).

Magliocca, N. R., Van Khuc, Q., de Bremond, A., & Ellicott, E. A. (2020). Direct and indirect land-use change caused by large-scale land acquisitions in Cambodia. *Environmental Research Letters*, *15*(2), 024010.

Messerli, P., Giger, M., Dwyer, M. B., Breu, T., & Eckert, S. (2014). The geography of large-scale land acquisitions: Analysing socio-ecological patterns of target contexts in the global South. *Applied Geography*, *53*, 449–459.

Nelson, A., & Chomitz, K. M. (2011). Effectiveness of strict vs. multiple use protected areas in reducing tropical forest fires: A global analysis using matching methods. *PloS one*, *6*(8), e22722.

Nolte, C., & Agrawal, A. (2013). Linking management effectiveness indicators to observed effects of protected areas on fire occurrence in the Amazon rainforest. *Conservation Biology*, *27*(1), 155–165.

Oberlack, C., Tejada, L., Messerli, P., Rist, S., & Giger, M. (2016). Sustainable livelihoods in the global land rush? Archetypes of livelihood vulnerability and sustainability potentials. *Global environmental change*, *41*, 153–171.

Oliveira, G. D. L. (2016). The geopolitics of Brazilian soybeans. In *Soy, Globalization, and Environmental Politics in South America* (pp. 98–122). Routledge.

Oya, C. (2013). The Land Rush and Classic Agrarian Questions of Capital and Labour: A systematic scoping review of the socioeconomic impact of land grabs in Africa. *Third World Quarterly*, *34*(9), 1532–1557.

Pogge, T., & Reddy, S. G. (2005). How not to count the poor. *Available at SSRN 893159*.

Potapov, P., Hansen, M. C., Pickens, A., Hernandez-Serna, A., Tyukavina, A., Turubanova, S., Zalles, V., Li, X., Khan, A., & Stolle, F. (2022). The global 2000-2020 land cover and land use change dataset derived from the Landsat archive: First results. *Frontiers in Remote Sensing*, *3*, 856903.

Potapov, P., Turubanova, S., Hansen, M. C., Tyukavina, A., Zalles, V., Khan, A., Song, X.-P., Pickens, A., Shen, Q., & Cortez, J. (2022). Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food*, *3*(1), 19–28.

RSPO. (2018). *RSPO Principles and Criteria*. Roundtable on Sustainable Palm Oil. https://rspo.org/wp-content/uploads/rspo-principles-criteria-for-production-of-sustainable-palm-oil-2018revised-01-february-2020-with-updated-supply-chain-requirements-for-mills.pdf

Rulli, M. C., Saviori, A., & D'Odorico, P. (2013). Global land and water grabbing. *Proceedings of the National Academy of Sciences*, *110*(3), 892–897.

Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., & Melville, G. (2020). geoBoundaries: A global database of political administrative boundaries. *PLoS One*, *15*(4), e0231866.

Scharsich, V., Mtata, K., Hauhs, M., Lange, H., & Bogner, C. (2017). Analysing land cover and land use change in the Matobo National Park and surroundings in Zimbabwe. *Remote sensing of environment*, *194*, 278–286.

Schoneveld, G. C. (2014). The geographic and sectoral patterns of large-scale farmland investments in sub-Saharan Africa. *Food Policy*, *48*, 34–50.

Scoones, I., Hall, R., Borras Jr, S. M., White, B., & Wolford, W. (2013). The politics of evidence: Methodologies for understanding the global land rush. In *Journal of Peasant Studies* (Bd. 40, Nummer 3, pp. 469–483). Taylor & Francis.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., & Yu, T.-H. (2008). Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science*, *319*(5867), 1238–1240.

Smaller, C., Speller, W., Mirza, H., Bernasconi-Osterwalder, N., & Dixie, G. (2015). *Investment contracts for agriculture: Maximizing gains and minimizing risks*.

Smith, P., Gregory, P. J., Van Vuuren, D., Obersteiner, M., Havlík, P., Rounsevell, M., Woods, J., Stehfest, E., & Bellarby, J. (2010). Competition for land. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2941–2957.

The Land Matrix. (2023). https://landmatrix.org/faq#what-is-a-land-deal

Turner, B. L., Lambin, E. F., & Reenberg, A. (2007). The emergence of land change science for global environmental change and sustainability. *Proceedings of the National Academy of Sciences*, *104*(52), 20666–20671.

UNEP-WCMC and IUCN. (2023). *Protected Planet: The World Database on Protected Areas (WDPA) [Online]* [dataset]. www.protectedplanet.net

von Maltitz, G. P., Gasparatos, A., Fabricius, C., Morris, A., & Willis, K. J. (2016). Jatropha cultivation in Malawi and Mozambique: Impact on ecosystem services, local human well-being, and poverty alleviation. *Ecology and Society*, *21*(3).

Yang, B., & He, J. (2021). Global land grabbing: A critical review of case studies across the world. *Land*, *10*(3), 324.

Yengoh, G. T., & Armah, F. A. (2015). Effects of large-scale acquisition on food insecurity in Sierra Leone. *Sustainability*, *7*(7), 9505–9539.

Zaehringer, J. G., Atumane, A., Berger, S., & Eckert, S. (2018). Large-scale agricultural investments trigger direct and indirect land use change: New evidence from the Nacala corridor, Mozambique. *Journal of land use science*, *13*(3), 325–343.

Zoomers, A. (2010). Globalisation and the foreignisation of space: Seven processes driving the current global land grab. *The Journal of Peasant Studies*, *37*(2), 429–447.

8 Appendix

8.1 Robustness Checks

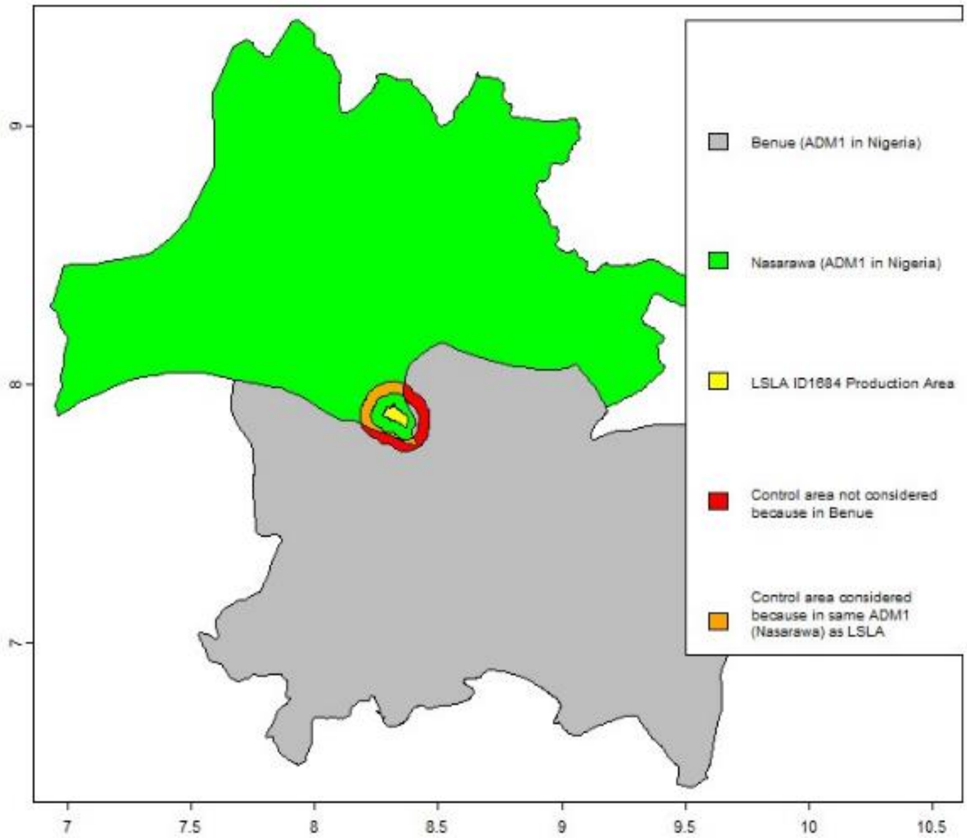
8.1.1 Robustness Check I: Within Country Analysis

It is conceivable that, for instance, institutional factors differ between the LSLA treatment area and control area, which may have a different impact on the deforestation rate in the two areas. Institutional factors such as regulations or land-use policies (see *Figure 1*) may differ between countries. Therefore, I only analysed those areas within the control area of an LSLA that are in the same country as the LSLA. According to my analysis, the entire area of the control area *i* is already in the same country as the LSLA *i* for all LSLAs studied.

8.1.2 Robustness Check II: Within First-Order Administrative Division(s) Analysis

Institutional factors, however, may differ within a country between first-order administrative divisions. I, therefore, only analysed those areas within a control area of an LSLA if the areas are in the same ADM1(s) as the LSLA. A visual example of this robustness check is shown in *Figure 8*. The production area of LSLA ID1684 is located in the ADM1 Nasarawa in Nigeria. Therefore, only the area that lies between 5-10km from the LSLA and lies in Nasawara (coloured in orange) is used as the control area for LSLA ID1684. The part of the control area that lies in Benue (coloured in red) is not analysed.

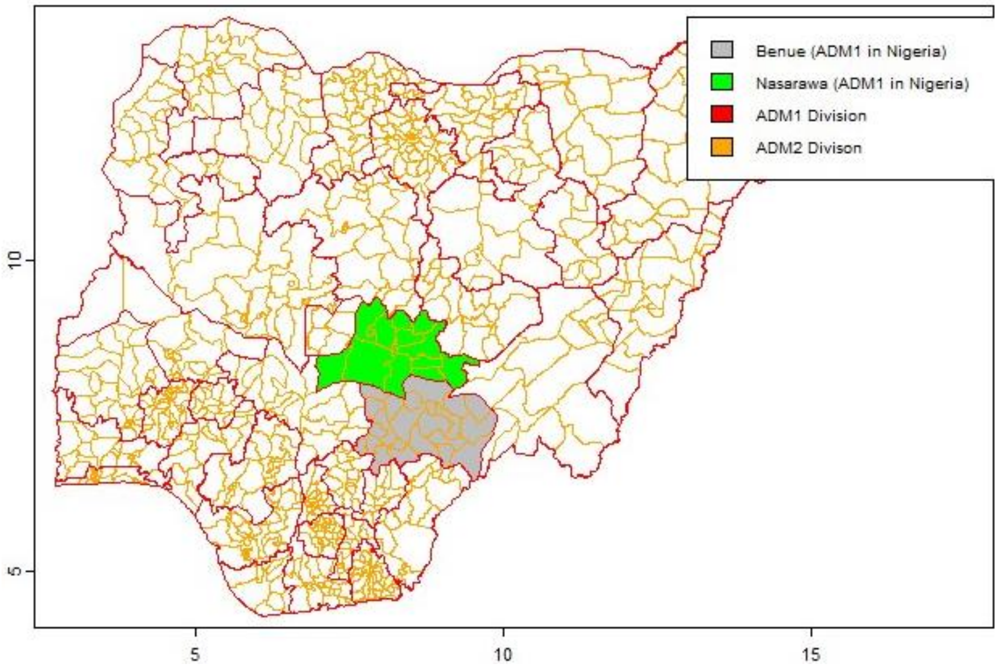
***Figure 8**: Example of Robustness Check II: LSLA ID1684 in Nigeria*



Notes: The areas are projected in longitude/latitude relative to the WGS-84 datum.

Figure 9 shows the location of ADM1 Nasarawa and Benue within Nigeria. All ADM1 borders are coloured red, all ADM2 borders are coloured orange.

***Figure 9**: ADM1 & ADM2 in Nigeria*



Notes: Nigeria's ADM1 and ADM2 borders are projected in longitude/latitude relative to the WGS-84 datum.

Table 2 presents the aggregated treatment effect estimates for the analysis with control areas in the same ADM1 region as the LSLAs.

Table 2

LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas in the Same First-Order Administrative Division (ADM1) Region as LSLA

			95% CI	
Type	ATT	Std. Error	LL	UL
Simple Aggregation	-14.1611	3.6865	-21.3866	-6.9357*
Dynamic Aggregation	-13.7986	3.5316	-20.7204	-6.8768*

Number of observations: 920
Units of observations: 40

Notes: The table reports simple and dynamic aggregated treatment effect parameters under the conditional parallel trend assumption and with clustering at the area level. The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the Average Treatment Effect on the Treated (ATT) in a simple difference-in-differences analysis with two groups and two periods. The row "Dynamic" reports a cross-group average on the average treatment effect for each group, where groups are defined by length of exposure. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that are in the same first-order administrative division as the LSLA. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

Table 3 presents the estimated effects for the event study analysis with control areas in the same ADM1 regions as the LSLAs. The estimates do not deviate much from the estimates in *Table 1*, which could be explained by the circumstance that deforestation trends within the control areas may not be very different between the ADM1 regions studied. Interestingly, at event time -3 the coefficient is statistically

Table 3

Robustness Check: LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas in the Same First-Order Administrative Division (ADM1) as LSLA

			95% CI	
Event Time	Estimate	Std. Error	LL	UL
-10	0.0424	0.0835	-0.1555	0.2404
-9	-0.273	0.392	-1.2027	0.6568
-8	-0.2554	0.3411	-1.0644	0.5536
-7	-0.0041	0.0548	-0.1341	0.126
-6	-0.3627	0.42	-1.3588	0.6334
-5	-0.2898	0.309	-1.0226	0.4429
-4	-0.0761	0.0739	-0.2512	0.0991
-3	0.0877	0.0337	0.0077	0.1677*
-2	0.0816	0.0974	-0.1494	0.3126
-1	0.0037	0.0745	-0.1729	0.1804
0	-0.0046	0.0769	-0.1871	0.1778
1	-2.1056	1.0194	-4.5231	0.312
2	-7.6109	2.9265	-14.5513	-0.6706*
3	-12.6923	4.2772	-22.8361	-2.5486*
4	-16.6521	4.4526	-27.2117	-6.0925*
5	-18.0739	4.6971	-29.2134	-6.9344*
6	-17.926	4.8161	-29.3477	-6.5042*
7	-16.7654	4.8592	-28.2894	-5.2415*
8	-18.6667	5.1295	-30.8316	-6.5017*
9	-18.3341	4.7887	-29.6909	-6.9773*
10	-22.9528	5.4347	-35.8416	-10.064*

Number of observations: 920
Units of observations: 40

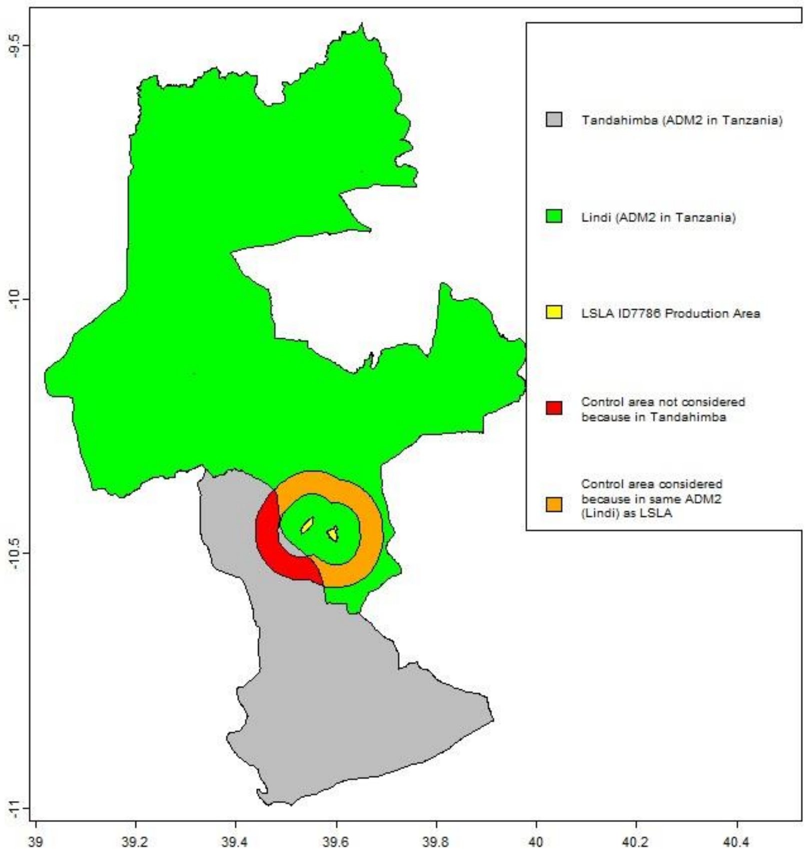
Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that are in the same first-order administrative division as the LSLA. '' indicates that the 95% confidence band does not cover 0.*

significant at the 5% significance level. This means that I no longer fail to reject the null hypothesis of no difference in deforestation trends between treatment and control areas for all pre-treatment periods. However, given that the significant coefficient at event time -3 is close to 0, it is probably less of a concern for the conditional parallel trends assumption.

8.1.3 Robustness Check III: Within Second-Order Administrative Division(s) Analysis

In *section 8.1.2*, only those areas within the control areas of the LSLAs that are located in the same ADM1 region as the LSLA itself were analysed. This method reduces the risk that deforestation trends are influenced by differences in ADM1 institutional factors between treatment and control areas. However, it is conceivable that there is further heterogeneity in institutional factors within an ADM1 region. Therefore, exclusively those areas within a control area of an LSLA were analysed if the areas are located in the same ADM2 region(s) as the LSLA. An example of this robustness check is shown in Figure 10. The production area of the LSLA ID7786 is located within the ADM2 region Lindi in Tanzania. Therefore, deforestation trends were only analysed within the control area in Lindi, but not in Tandahimba.

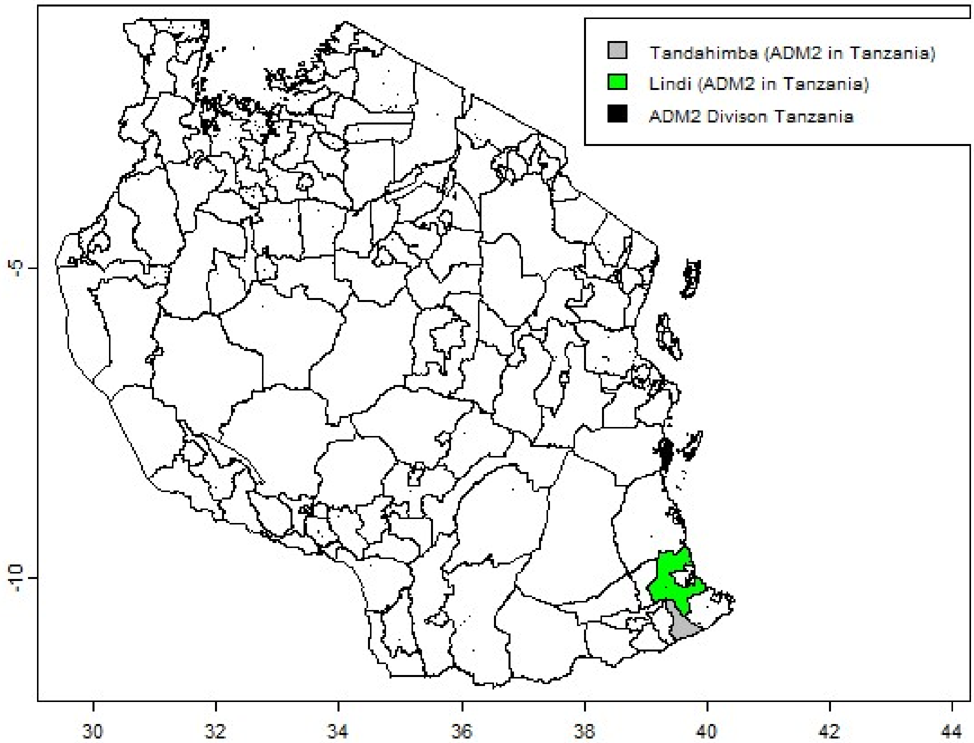
***Figure 10**: Example of Robustness Check II: LSLA ID1684 in Nigeria*



Notes: The areas are projected in longitude/latitude relative to the WGS-84 datum

Figure 11 shows the location of the ADM2 regions Tandahimba and Lindi in Tanzania. The ADM2 boundaries are coloured black.

***Figure 11**: ADM2 in Tanzania*



Notes: Tanzania's ADM2 boundaries are projected in longitude/latitude relative to the WGS-84 datum

Table 4 presents the aggregated treatment effect estimates for the analysis with control areas in the same ADM2 region as the LSLAs.

Table 4

LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas in the Same Second-Order Administrative Division (ADM2) as LSLA

			95% CI	
Type	ATT	Std. Error	LL	UL
Simple Aggregation	-14.3108	3.802	-21.7626	-6.859*
Dynamic Aggregation	-14.0355	3.6177	-21.1261	-6.945*

Number of observations: 920
Units of observations: 40

Notes: The table reports simple and dynamic aggregated treatment effect parameters under the conditional parallel trend assumption and with clustering at the area level. The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the Average Treatment Effect on the Treated (ATT) in a simple difference-in-differences analysis with two groups and two periods. The row "Dynamic" reports a cross-group average on the average treatment effect for each group, where groups are defined by length of exposure. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that are in the same second-order administrative division as the LSLA. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

Table 5 presents the estimated effects for the event study analysis with control areas in the same ADM2 as the LSLAs. Again, the estimates are similar to those of the main analysis presented in *Table 1*.

Table 5

Robustness Check: LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas in the Same Second-Order Administrative Division (ADM2) as LSLA

			95% CI	
Event Time	Estimate	Std. Error	LL	UL
-10	0.0841	0.1005	-0.1504	0.3186
-9	-0.282	0.3841	-1.1782	0.6142
-8	-0.2786	0.3378	-1.0669	0.5097
-7	-0.017	0.0579	-0.1522	0.1181
-6	-0.3535	0.4204	-1.3345	0.6275
-5	-0.3074	0.3176	-1.0486	0.4338
-4	-0.0658	0.0698	-0.2286	0.0971
-3	0.0754	0.038	-0.0133	0.164
-2	0.0854	0.1016	-0.1517	0.3225
-1	-0.0102	0.0713	-0.1767	0.1563
0	-0.0041	0.0899	-0.2139	0.2058
1	-2.1207	0.98	-4.4076	0.1662
2	-7.6786	2.9271	-14.5089	-0.8482*
3	-12.8126	4.6171	-23.5866	-2.0386*
4	-16.7551	4.6702	-27.653	-5.8573*
5	-18.1819	4.5382	-28.7718	-7.592*
6	-18.092	4.7244	-29.1162	-7.0677*
7	-17.0678	4.789	-28.2428	-5.8928*
8	-19.1512	5.0605	-30.9597	-7.3426*
9	-18.9221	5.1406	-30.9175	-6.9266*
10	-23.6049	5.1884	-35.7121	-11.4977*

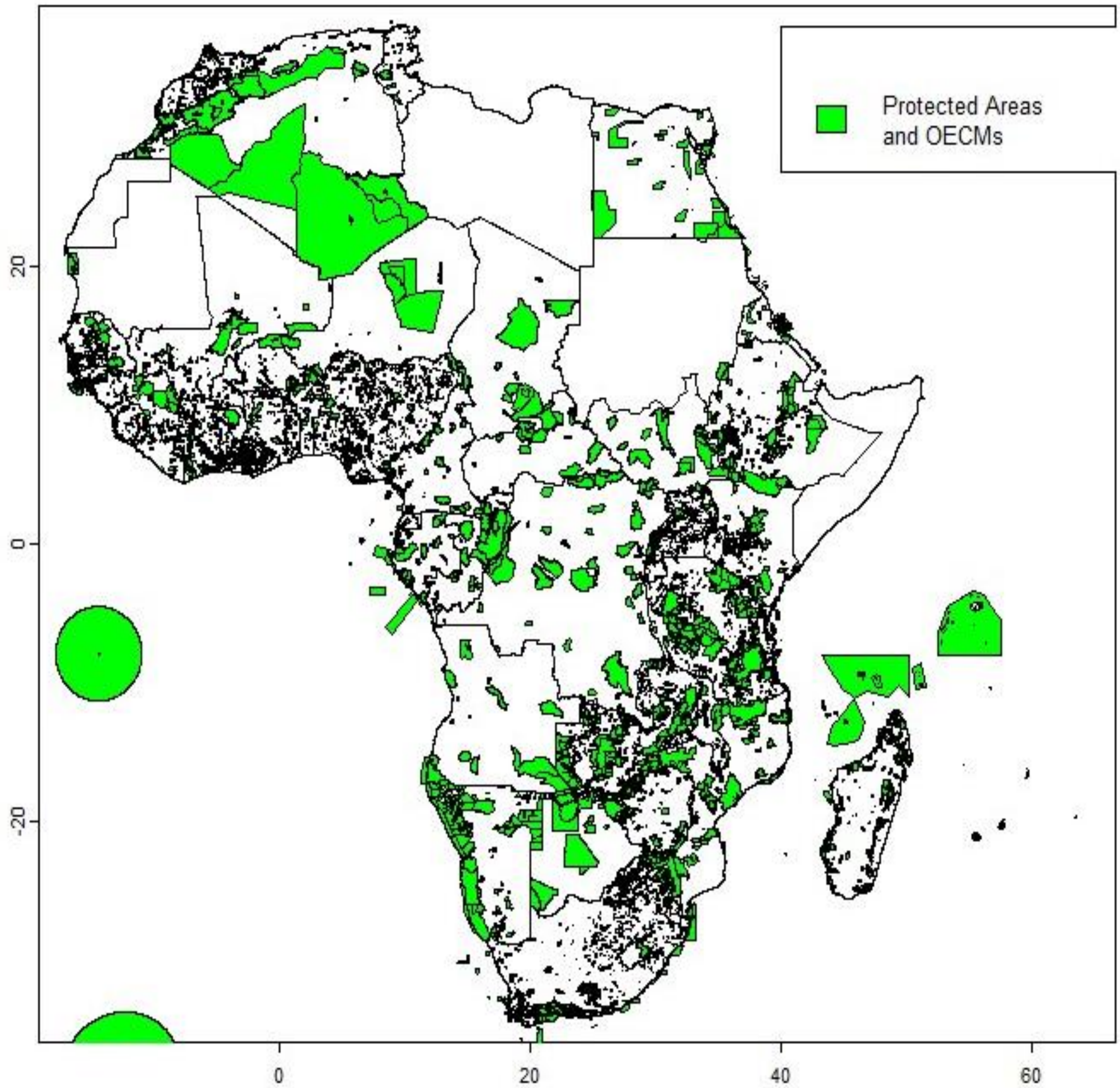
Number of observations: 920
Units of observations: 40

Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that are in the same second-order administrative division (ADM2) as the LSLA. '' indicates that the 95% confidence band does not cover 0.*

8.1.4 Robustness Check IV: Control Areas Excluding Protected Areas and OECMs

Institutional factors may differ within ADM2 regions, for example due to the presence of protected areas. If the control area or part of it is located in a protected area, the deforestation trends could be different than in the LSLA area without treatment. In other words, the control area would not reveal the counterfactual trend. This circumstance could lead to a misestimation of the treatment effect. To analyse to what extent protected areas and other effective area-based conservation measures (OECMs) influence the estimated treatment effect, I only analysed deforestation trends in those areas within the control areas that do not comprise protected areas and OECMs. I used data from the most comprehensive global database of protected areas from UNEP-WCMC and IUCN (2023). *Figure 12* shows the 8290 protection zones and OECMs for Africa listed in the dataset.

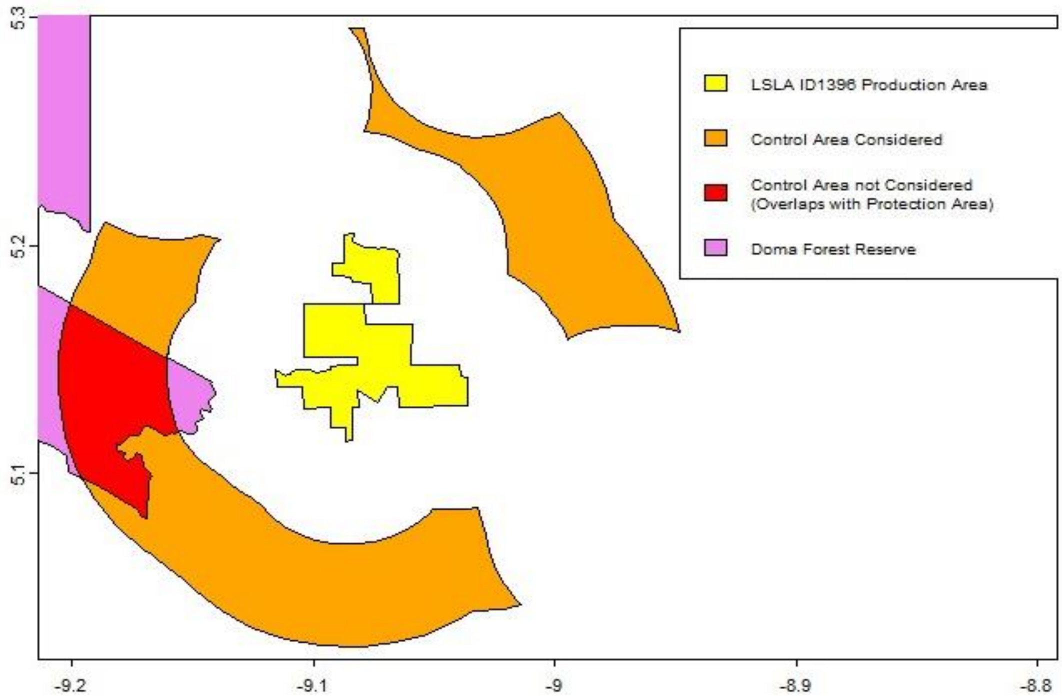
***Figure 12**: Protected Areas and OECMs in Africa*



Notes: Africa is projected in longitude/latitude relative to the WGS-84 datum.

Figure 13 shows as an example the production and control area for LSLA ID1396 in Liberia. Part of the control area overlaps with the Doma Forest Reserve (coloured red). Deforestation trends in the overlapping area are not analysed. The adjusted control area thus corresponds to the orange-coloured area.

Figure 13*: Production & Adjusted Control Area for LSLA ID13396*



Notes: The areas are projected in longitude/latitude relative to the WGS-84 datum.

Regulations and their implementation within the protected areas may vary between countries and regions. In fact, the production areas of two LSLAs are located in protected areas. LSLA ID4521 in Nigeria overlaps with the Usonigbe Forest Reserve (*Figure 14*), and LLSA ID4623 in Ethiopia lies within the Omo National Park and Tama Wildlife Reserve (*Figure 15*). These findings prompt the question of the extent to which protected zones in the control areas actually influence deforestation trends.

Figure 14*: LSLA ID4521*

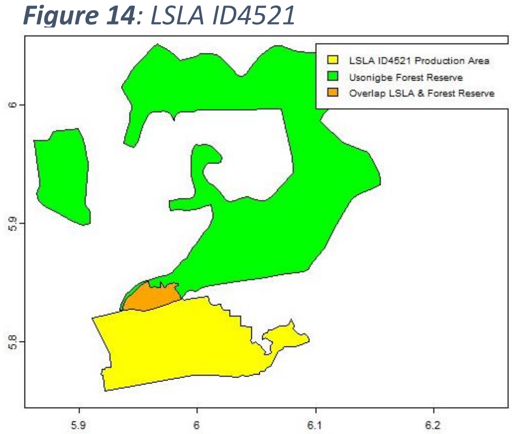
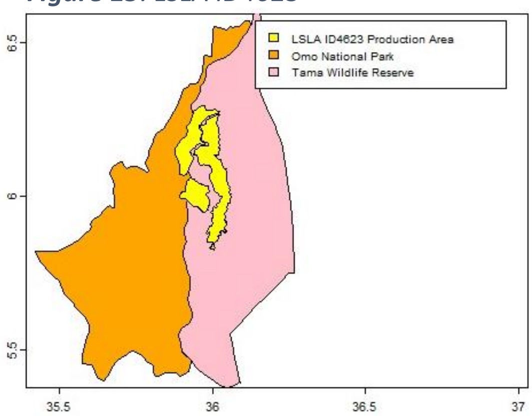


Figure 15*: LSLA ID4623*



Notes: The areas are projected in longitude/latitude relative to the WGS-84 datum.

Table 6 presents the aggregated treatment effect estimates for the analysis with control areas that do not contain protected areas and OECMs.

Table 6

LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas Not in Protected Areas and Other Effective Area-Based Conservation Measures

			95% CI	
Type	ATT	Std. Error	LL	UL
Simple Aggregation	-14.6595	3.7219	-21.9543	-7.3647*
Dynamic Aggregation	-14.1941	3.6805	-21.4078	-6.9805*

Number of observations: 920
Units of observations: 40

Notes: The table reports simple and dynamic aggregated treatment effect parameters under the conditional parallel trend assumption and with clustering at the area level. The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the Average Treatment Effect on the Treated (ATT) in a simple difference-in-differences analysis with two groups and two periods. The row "Dynamic" reports a cross-group average on the average treatment effect for each group, where groups are defined by length of exposure. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that do not comprise protected areas and other effective area-based conservation measures (OECMs). 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

Table 7 shows the estimated effects for the event study analysis with control areas that do not contain protected areas and OECMs. The estimated coefficients are similar to those of the main analysis.

Table 7

Robustness Check: LSLA Aggregated Treatment Effect Estimates: Event Study with Control Areas Not Comprising Protected Areas and Other Effective Area-Based Conservation Measures (OECMs)

			95% CI	
Event Time	Estimate	Std. Error	LL	UL
-10	0.0291	0.0749	-0.146	0.2042
-9	-0.3293	0.3968	-1.2563	0.5978
-8	-0.3031	0.3324	-1.0798	0.4735
-7	-0.0201	0.0542	-0.1468	0.1067
-6	-0.3688	0.422	-1.3549	0.6172
-5	-0.3274	0.28	-0.9818	0.3269
-4	-0.1061	0.0845	-0.3036	0.0915
-3	0.0652	0.035	-0.0167	0.147
-2	0.0171	0.1135	-0.2482	0.2823
-1	-0.0193	0.0769	-0.199	0.1603
0	0.0033	0.1086	-0.2503	0.257
1	-2.1174	0.9556	-4.3501	0.1153
2	-7.685	3.0483	-14.8074	-0.5625*
3	-12.835	4.3882	-23.088	-2.582*
4	-16.939	4.7325	-27.9965	-5.8815*
5	-18.4498	4.7598	-29.5711	-7.3285*
6	-18.3946	4.7031	-29.3833	-7.4058*
7	-17.29	4.6524	-28.1603	-6.4197*
8	-19.3903	5.4468	-32.1166	-6.6639*
9	-19.1509	4.7741	-30.3055	-7.9964*
10	-23.8869	5.2287	-36.1037	-11.6701*

Number of observations: 920
Units of observations: 40

Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that are not within protected areas and other effective area-based conservation measures (OECMs). '' indicates that the 95% confidence band does not cover 0.*

8.1.5 Robustness Check V: Control Areas Excluding Areas of Built-Up Expansion

In *section 2.2 equation 3,* I modelled land competition between agricultural and non-agricultural land use as follows:

$$\frac{d\boldsymbol{Z}_{A,m}(\boldsymbol{t})}{d\boldsymbol{t}} = -(\boldsymbol{\alpha}\frac{d\boldsymbol{Z}_{C,m}(\boldsymbol{t})}{d\boldsymbol{t}} + \boldsymbol{\beta}\frac{d\boldsymbol{Z}_{B,m}(\boldsymbol{t})}{d\boldsymbol{t}} + \boldsymbol{\gamma}\frac{d\boldsymbol{Z}_{W,m}(\boldsymbol{t})}{d\boldsymbol{t}} + \boldsymbol{\delta}\frac{d\boldsymbol{Z}_{O,m}(\boldsymbol{t})}{d\boldsymbol{t}})$$

, where Z_C denotes land used for nature conservation, Z_A agricultural land, Z_B built-up areas, Z_W water bodies, and Z_O other designated land uses such as logging or mining.

However, land competition exists not only between agricultural and non-agricultural land use but also within non-agricultural land use. If in the absence of the LSLA

$$\frac{d\boldsymbol{Z}_{C,TA}}{d\boldsymbol{t}} \neq \frac{d\boldsymbol{Z}_{C,CA}}{d\boldsymbol{t}}$$ [3]

, where *TA* and *CA* denote treatment and control area, respectively, and $TA, CA \in m$,

this could bias the treatment estimates, as the control area would not reveal the counterfactual trend.

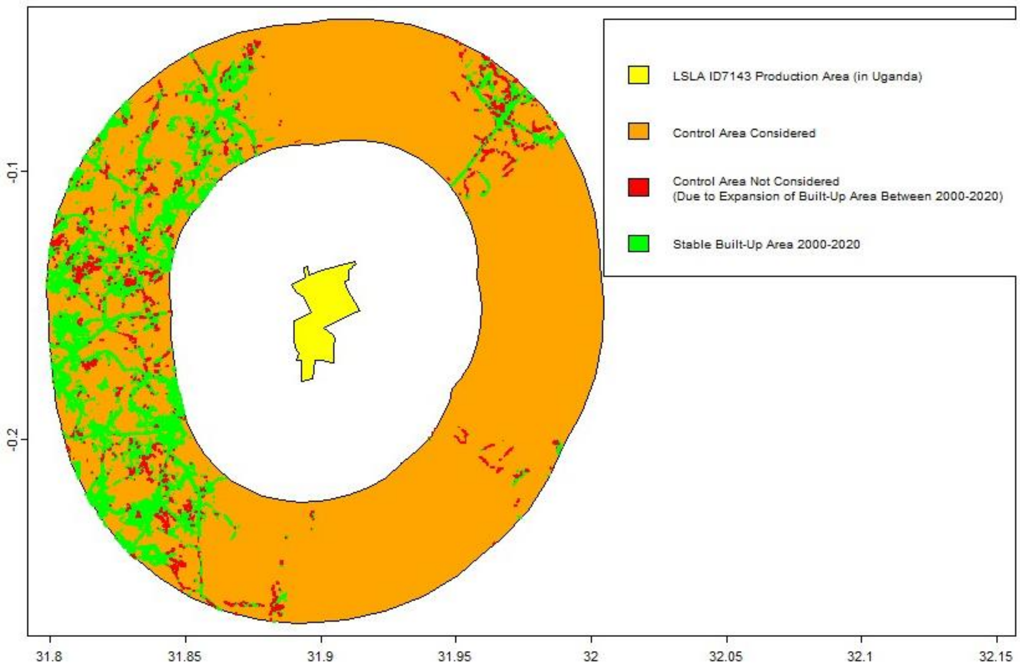
For example, a larger expansion of the built-up area in the control area could lead to higher deforestation rates than a smaller hypothetical expansion in the treatment area without the treatment. The hypothetical built-up area expansion in the treatment area cannot be observed by definition. However, it is possible to analyse the deforestation trend in the control area only for those areas that have not experienced an expansion of built-up area. The analysis of deforestation trends only in areas that have not experienced an expansion of built-up area is not necessarily counterfactual to the treatment area in the absence of the treatment. Nevertheless, it does provide information on direct land competition between $Z_{C,CA}$ and $Z_{B,CA}$. For the mapping of stable built-up and built-up expansion between 2000-2020, I relied on the dataset of Potapov, Hansen, et al. (2022).

Figure 16 shows as an example the production area of LSLA ID7143 and the control area, which does not include areas of built-up expansion. The red coloured area represents the area that was built on between 2000 and 2020. I analysed the change in deforestation in the control area outside these red coloured areas, as the inclusion of these areas in the analysis of deforestation could bias the estimate of the treatment effect.

---

[2] For the equation to hold when CA is the perfect counterfactual of TA, I assume here for simplicity that CA and TA have the same spatial extent, since the derivative with respect to time is usually expressed in absolute values.

***Figure 14**: LSLA ID7143*



Notes: The areas are projected in longitude/latitude relative to the WGS-84 datum.

Tables 8 and *9* show the aggregate estimates of the treatment effect for the analysis with control areas not comprising built-up expansion areas.

Table 8

LSLA Aggregated Treatment Effect Estimates: Event Study Excluding Built-Up Expansion Areas in Control Areas

			95% CI	
Type	ATT	Std. Error	LL	UL
Simple Aggregation	-14.5836	3.7227	-21.8799	-7.2873*
Dynamic Aggregation	-14.1618	3.529	-21.0786	-7.245*

Number of observations: 920
Units of observations: 40

Notes: The table reports simple and dynamic aggregated treatment effect parameters under the conditional parallel trend assumption and with clustering at the area level. The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the Average Treatment Effect on the Treated (ATT) in a simple difference-in-differences analysis with two groups and two periods. The row "Dynamic" reports a cross-group average on the average treatment effect for each group, where groups are defined by length of exposure. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that do not include areas of built-up expansion. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

Table 9

Robustness Check: LSLA Aggregated Treatment Effect Estimates: Event Study Without Area of Built-Up Area Expansion in Control Areas

			95% CI	
Event Time	Estimate	Std. Error	LL	UL
-10	0.0317	0.0777	-0.1516	0.215
-9	-0.3323	0.3953	-1.265	0.6004
-8	-0.3025	0.3435	-1.1129	0.508
-7	-0.0212	0.0543	-0.1492	0.1069
-6	-0.3737	0.4099	-1.3409	0.5935
-5	-0.3243	0.3239	-1.0887	0.4401
-4	-0.1008	0.0791	-0.2875	0.0859
-3	0.0627	0.0362	-0.0226	0.148
-2	0.0292	0.098	-0.2022	0.2605
-1	-0.0292	0.077	-0.211	0.1526
0	-0.0263	0.0823	-0.2204	0.1679
1	-2.1432	0.9403	-4.362	0.0757
2	-7.7069	2.8845	-14.5135	-0.9004*
3	-12.8382	4.4741	-23.3956	-2.2808*
4	-16.88	4.8946	-28.4297	-5.3303*
5	-18.3712	4.8327	-29.7749	-6.9675*
6	-18.2922	4.7824	-29.5773	-7.0072*
7	-17.2059	4.4765	-27.7692	-6.6427*
8	-19.3115	5.2454	-31.6891	-6.9339*
9	-19.0743	4.9051	-30.6489	-7.4997*
10	-23.93	5.1855	-36.1663	-11.6937*

Number of observations: *920*
Units of observations: *40*

Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units that do not include areas of built-up expansion. '' indicates that the 95% confidence band does not cover 0.*

The aggregated treatment effect estimates in *Table 8* as well as the estimates in *Table 9* again do not differ strongly from the estimates in the main analysis.

Overall, I have shown that the findings of a significant positive relationship between LSLA presence and deforestation within the LSLA production/contract area are robust when comparing changes in forest extent in the treatment area with:

a) changes in forest extent in control areas in the same ADM1 region as the LSLA.

b) changes in forest extent in control areas in the same ADM2 regions as the LSLA.

c) changes in forest extent in control areas that do not comprise protected areas and OECMs.

d) changes in forest extent in control areas that do not include built-up expansion areas.

Although causal inference may still not be possible due to the potential presence of other unobservable confounding factors, the checks performed nevertheless strengthen my confidence in the robustness of my results.

8.2 Data Sources

The data analysis was carried out using R. Due to the significant length of the code (more than 7000 lines), the code is divided into 9 different files. R code is available upon request. The created spatial polygons of the LSLAs and other identified plantations are stored in well-known text (WKT) format and are available on request.

Data	Source & Description
Data on Large-Scale Land Acquisitions	The Land Matrix, 2023 Downloaded from: https://landmatrix.org/list/deals/
Cropland Mapping	Potapov, Turubanova, et al., 2022 Downloaded from: https://glad.umd.edu/dataset/croplands Compressed GeoTiff files aggregated to quadrant mosaics for NW, NE, SW,SE for 2003, 2007, 2011, 2015, 2019, respectively, were downloaded.
Palm Oil Production Mapping	Descals et al., 2021 Downloaded from: https://zenodo.org/record/4473715 Compressed GeoTiff files aggregated to 100x100km tiles were downloaded
Forest Mapping	Hansen et al., 2013 Downloaded from: https://storage.googleapis.com/earthenginepartners-hansen/GFC-2022-v1.10/download.html Compressed GeoTiff files for tree canopy for the year 2000, and the year of gross forest area loss for the period 2000-2022 were downloaded for tiles 00N_010E,00N_030E, 10N_000E, 10N_010E, 10N_010W, 10N_020W, 10N_030E, 10S_020E, 10S_030E, 20N_020W, 20S_030E and 30N_030E.
Built-Up Mapping	Potapov, Hansen, et al., 2022 Downloaded from: https://glad.umd.edu/users/Potapov/GLCLUC2020/Built-up_change_2000_2020/ Compressed GeoTiff files for stable built-up and built-up change were downloaded for tiles 00N_010E,00N_030E, 10N_000E, 10N_010E, 10N_010W, 10N_020W, 10N_030E, 10S_020E, 10S_030E, 20N_020W, 20S_030E and 30N_030E.
Protected Areas and Other Effective Area-Based Conservation Measures	UNEP-WCMC and IUCN, 2023 Downloaded from: https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA
Shapefiles Africa	ICPAC, 2017 Downloaded from: https://geoportal.icpac.net/layers/geonode%3Aafr_g2014_2013_0
Shapefiles ADM1 & ADM2	Runfola et al., 2020 Downloaded within R with package rgeoboundaries

8.3 Regression Tables

***Table 10**: Complete Event Study for All Event Times for Table 1*

LSLA Aggregated Treatment Effect Estimates: Event Study

			95% CI	
Event Time	Estimate	Std. Error	LL	UL
-10	0.0318	0.0764	-0.1412	0.2049
-9	-0.3307	0.3899	-1.2136	0.5522
-8	-0.3018	0.3316	-1.0528	0.4491
-7	-0.0204	0.0544	-0.1436	0.1029
-6	-0.3723	0.4072	-1.2943	0.5496
-5	-0.3230	0.2751	-0.9460	0.3000
-4	-0.1002	0.0808	-0.2831	0.0827
-3	0.0633	0.0349	-0.0156	0.1422
-2	0.0319	0.1010	-0.1969	0.2606
-1	-0.0266	0.0839	-0.2166	0.1635
0	-0.0255	0.0842	-0.2161	0.1651
1	-2.1416	1.0025	-4.4116	0.1283
2	-7.7038	2.9817	-14.4554	-0.9523*
3	-12.8323	4.5619	-23.1617	-2.5029*
4	-16.8748	4.6645	-27.4366	-6.3130*
5	-18.3638	4.7042	-29.0154	-7.7122*
6	-18.2825	5.0541	-29.7265	-6.8385*
7	-17.1905	5.0528	-28.6316	-5.7495*
8	-19.2920	5.1284	-30.9041	-7.6799*
9	-19.0518	5.2067	-30.8412	-7.2623*
10	-23.8993	5.4826	-36.3133	-11.4852*

Number of observations: 920
Units of observations: 40

Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

Due to the staggered adoption of treatment, the composition of these average treatment effects is different for different time points, so they should not be directly compared.

Table 11: *Event Study Results on Changes in Forest Extent Near the LSLA*

LSLA Aggregated Treatment Effect Estimates: Event Study for Spillover Area

Event Time	Estimate	Std. Error	95% CI	
			LL	UL
-10	0.039	0.0234	-0.0159	0.0939
-9	-0.0063	0.0136	-0.0381	0.0256
-8	-0.0963	0.0988	-0.3278	0.1352
-7	0.0202	0.0344	-0.0603	0.1007
-6	-0.0142	0.0241	-0.0706	0.0422
-5	-0.0054	0.0488	-0.1196	0.1088
-4	0.0117	0.0219	-0.0396	0.063
-3	0.0201	0.0214	-0.03	0.0702
-2	-0.0529	0.0611	-0.196	0.0902
-1	0.002	0.0425	-0.0975	0.1016
0	-0.0425	0.0395	-0.1349	0.0499
1	-0.1215	0.1118	-0.3832	0.1403
2	-0.0671	0.1929	-0.5188	0.3846
3	-0.0791	0.2282	-0.6137	0.4555
4	-0.234	0.3738	-1.1096	0.6416
5	-0.2041	0.4108	-1.1664	0.7581
6	-0.1407	0.4897	-1.2877	1.0062
7	-0.0218	0.5755	-1.3698	1.3263
8	-0.1179	0.6992	-1.7555	1.5197
9	-0.1306	0.7716	-1.9379	1.6768
10	-0.2917	1.0112	-2.6602	2.0769

Number of observations: *920*
Units of observations: *40*

Notes: The table reports average treatment effects by the length of exposure to the LSLA. The column "Event Time" indexes the length of exposure to the treatment. Estimates are made assuming a conditional parallel trend and with clustering at the area level. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*

***Table 12**: Aggregated Treatment Effect Estimates for Forest Extent Near the LSLA*

LSLA Aggregated Treatment Effect Estimates: Simple & Dynamic Aggregation for Spillover Area

			95% CI	
Type	ATT	Std. Error	LL	UL
Simple Aggregation	-0.5093	0.7071	-1.8952	0.8765
Dynamic Aggregation	-0.1319	0.4347	-0.984	0.7201

Number of observations: 920
Units of observations: 40

Notes: The table reports simple and dynamic aggregated treatment effect parameters under the conditional parallel trend assumption and with clustering at the area level. The row "Simple" reports an average treatment effect parameter that has a very similar interpretation to the Average Treatment Effect on the Treated (ATT) in a simple difference-in-differences analysis with two groups and two periods. The row "Dynamic" reports a cross-group average on the average treatment effect for each group, where groups are defined by length of exposure. The estimates use the doubly robust estimator (see Callaway & Sant'Anna, 2021). The control group consists of never treated units. 0 anticipation periods were assumed. '' indicates that the 95% confidence band does not cover 0.*