FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

ASSESSING THE ECOLOGICAL IMPACT OF REFORESTATION ON FARMLAND IN NICARAGUA USING THE SATELLITE-DERIVED NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI)

Word count: 11.637

Pello Juan Pilar Múgica Gonzalez
Student number : 01206947

Supervisor: Prof. dr. Ilse Ruyssen
Co-supervisor: Prof. dr. Stijn Ronsse

Master's Dissertation submitted to obtain the degree of:

Master of Science in Business Engineering

Academic year: 2017 – 2018

UNIVERSITEIT GENT

“Development organizations almost always do all sorts of interventions without even knowing if those interventions actually help! I think that [this study] would be an amazing research contribution.” - *Dr. Kahlil Baker (2018)*

FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

ASSESSING THE ECOLOGICAL IMPACT OF REFORESTATION ON FARMLAND IN NICARAGUA USING THE SATELLITE-DERIVED NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI)

Word count: 11.637

Pello Juan Pilar Múgica Gonzalez
Studentennummer/ Student number : 01206947

Supervisor: Prof. dr. Ilse Ruyssen
Co-supervisor: Prof. dr. Stijn Ronsse

Master's Dissertation submitted to obtain the degree of:

Master of Science in Business Engineering

Academic year: 2017 – 2018

UNIVERSITEIT
GENT

PERMISSION

I declare that the content of this Master's Dissertation may be consulted and/or reproduced, provided that the source is referenced.

Pello Juan Pilar Mugica Gonzalez

:…………………………………………………………………………………………………

Handtekening/signature

Preface

Before you lies the Master's dissertation "Assessing the ecological impact of reforestation on farmland in Nicaragua using the satellite-derived normalized difference vegetation index (NDVI)", a case study with a strong statistical analysis that evaluates the ecological impact of a reforestation program by a Canadian NGO active in Northern Nicaragua using a "green-index". It has been written to fulfil the graduation requirements of the MSc degree in Business Engineering with main subject Data Analytics at the University of Ghent (UGent). I was engaged in researching and writing this dissertation from January to June 2018.

The project was undertaken in consultation with Taking Root, a Canadian not-for-profit organisation. I first got in touch with the organisations' co-founder Doctor Kahlil Baker via email in October 2017 in search of a development organisation willing to collaborate for this dissertation. My research subject was formulated together with my promotor, Dr. Prof. Ilse Ruyssen. Fulfilling a comprehensive research was of a great challenge, but conducting an extensive investigation, attending different congresses and meeting with the right people have allowed me to produce valuable insights. Fortunately, both Dr. Kahlil Baker and Taking Root's IT specialist Mr Newton Tse were always available to answer my queries.

I would like to thank my promotor for guiding me through this research, even though it was not clear at the beginning where this collaboration would lead to. Thank you very much for giving me this confidence. Thank you, Dr. Kahlil Baker for being so enthusiastic, fast and clear throughout all your communications, you kept me motivated. Also special thanks to Matthias Demuzere who assisted me to collect and geo-reference the vegetation indices on the Google Earth Engine. Because of all of you, my master's dissertation was a real exciting and educational adventure (as it should be). I also benefitted from debating issues with my dear friend and colleague Jakob Biebuyck. To my other friends and family, too many to mention, all have been a great support during the whole process. A particular note of thanks to my parents: your faith and unconditional support, as always, served me well.

I hope you enjoy your reading.

Pello Mugica Gonzalez
Ghent, June 5, 2018

Abstract

Successful reforestation is dependent on effective and efficient technical training and assistance. This study aims to gain understanding of what types of training and technical assistance to smallholder farmers are truly adding value to a reforestation project's success in the short term of 1 year. Does technical and sustainability training truly make a difference for smallholder farmers in a reforestation program? In this study, a geo-referenced dataset was extracted from a community development database of activities executed between September 2016 and September 2017 in Northern Nicaragua. This project dataset was then merged with a series of high-resolution satellite data in order to evaluate their impacts on vegetation cover. Robust sound conclusions are made. No significant positive effects are found in the short term in this study context. However, a negative effect of delivery of materials is found. The latter is discussed as a logical finding since materials are used to prepare the plot for planting, and thus resulting in a decrease of vegetation cover in the short term. The conclusion of this study implies that short-term impact evaluation is valuable to detect dysfunctionalities and opportunities that can have a systematic impact on the long-term.

Keywords: Reforestation, Poverty Alleviation, Smallholder Farming, Nicaragua, Data Analytics, Satellite Imagery

Table of contents

Glossary of terms

AIC	The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data (Wikipedia, 2018). The lower this value, the higher model quality.
Ba_{Ha}	Basal area per hectare is a term often used in forest management to define the area of a given section of land that is occupied by the cross-section of tree trunks and stems at the base.
Copernicus	Copernicus is a single earth observation programme and is directed by the European Commission in partnership with the European Space Agency (ESA)
DN	Digital Number; it is the generic term used for pixel values
ESA	European Space Agency
Fusion Tables	Fusion Tables is an experimental data visualization web application from Google to gather, visualize, and share data tables.
GIE	Geospatial Impact Evaluation is a term for describing impact evaluation that makes use of geospatial analysis
GIS	Geographic Information System
Drop-out	Drop-outs are data missing from a satellite image. Drop-outs can be caused by signal interference or sensor failure.
NDVI	Normalized Difference Vegetation Index; a satellite-based vegetation index that correlates strongly with aboveground net primary productivity.
NGO	Non-governmental organization
NIR	Near Infrared Reflectance
OECD	Organisation for Economic Cooperation and Development
RED	Red reflectance
ROI	Region of Interest
TOA	Top of the Atmosphere; by using Top of Atmosphere (TOA) reflectance values, instead of DNs, which are observed and measured at the sensor of the satellite, some distorting factors can be corrected.
TP_{Ha}	Trees per hectare is a term often used in forest management to define the density of a forest.
UInt16	Represents a 16-bit unsigned integer.
QA	Quality Assessment

List of Tables, Figures and Boxes

1 Introduction

While the concern has risen for development organizations to do impact evaluations on their projects (Ten Hoorn & Stubbe, 2013), it remains an on-going challenge because it is often seen as a very technical exercise that can only be carried out by external experts operating at some distance from the program (Perrin, 2012). Also identifying the appropriate data that can efficiently assess project impact is difficult (Perrin, 2012). Advances in geospatial analysis tools, from which numerous are open to the public, have enabled access to a wide variety of remote sensing[1] data that can function as evaluation indicators for a broad spectrum of development projects. In the words of BenYishay et al. (2017): "Geospatial impact evaluation (GIE) methods opened new opportunities to understand what works, what does not, and why at a substantially shorter time and lower financial cost". Examples of projects where remote sensing has been used for are health mapping (Sedda et al., 2015; Morikawa, 2014), evaluation of poverty alleviation projects (Morikawa, 2014), prediction of ecological effects of environmental change on ecosystems (Pettorelli et al., 2005), land titling (Buntaine, Hamilton & Millones, 2015), post-conflict environmental assessment (Steiner, 2007), and so on. Other examples of fields that make use of remote sensing are mineral exploration (Sabins, 1999), land cover/change classification (Rawat & Kumar, 2015), and many more. This study delivers empirical evidence of technical success drivers of a reforestation[2] project by using a remote sensing indicator.

The main aim of this study is to analyse to what extent forest performance is improved by technician visits and which type of visits have a stronger effect. It is possible to employ geospatial data as an indicator for forest performance since the four key ingredients for GIE (see Glossary) enumerated by BenYishay et al (2017) are available. Those ingredients are (a) a precisely defined and measured geographical scope of the intervention, (b) collection of spatially-explicit outcome and covariate data, (c) the possibility to fuse these geospatial data with in-situ measurement outcome and covariate data (d) and the access to econometric tools. Moreover, GIE is shown to be applicable to individual projects by (BenYishay, 2017;

[1] The following definition for remote sensing is used throughout this study: "The scanning of the earth by satellite or high-flying aircraft in order to obtain information about it."
[2] "Reforestation is an endeavor to improve the condition of land and to speed up natural succession by planting trees on old fields and cleared land. Benefits of forest reducing the effects of global warming, maintaining biodiversity, and providing recreation and educational places for people." (Shea, 1998).

Buchanan et al., 2016; Campbell & Hofmann, 2014; Dolan & Grepin, 2017), since remotely sensed data on forest cover and vegetation productivity are also accessible at fine spatial and temporal scales. These data are also proven to be feasible to measure agricultural productivity at the smallholder plot level in previous literature (Hansen et al., 2013). Another important reason why geospatial analysis is used for this study is that it is cheaper, saves a lot of time (BenYishay, Runfola & Buntaine, 2015) and it is believed by the researcher to be more generic for future research. The latter can be motivational because a researcher is able to assess the reforestation success through a remote sensing indicator, available at all times, independently from field observations.

Derived from satellites, a Normalized Difference Vegetation Index (NDVI) is a suitable indicator for monitoring overground vegetation because of its direct correlation with vegetation (Reed et al., 1994). NDVI is derived from the red reflectance (RED) and near-infrared reflectance (NIR) ratio, which are the amounts of near-infrared and red light, respectively, reflected by the vegetation and captured by the sensor of the satellite (Myneni et al., 1995). The NDVI formula *(Eq. 1)* is based on the fact that chlorophyll absorbs RED, whereas the mesophyll leaf structures scatter NIR. The NDVI values varies from -1 to +1, where negative values correspond to an absence of vegetation (Myneni et al., 1995), and 0.2 is a sign of living vegetation. That is for example how bare bottoms (rocks, sand, ploughed country) also can be identified.

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \qquad (1)$$

The photosynthesizing property of plants expressed in a NDVI value has been shown to reliably gauge vegetation dynamics and intensity over a wide range of climate and ecology conditions (Fensholt & Sandholt, 2005; Huete et al., 2002). The link between the NDVI index and the fraction of absorbed photosynthetic active radiation intercepted (fAPAR) has been well documented, both theoretically (Sellers et al., 1992) and empirically (Asrar, Fuchs, Kanemasu & Hatfield, 1984). NDVI is a dependable tool to measure a wide variety of vegetation-related characteristics including crop type (Vintrou et al., 2012), cropping area famine (Vintrou et al., 2012), early warning drought (Funk & Brown, 2006), land cover types (Combalicer et al., 2011), soil erosion (Butt, Waqas, Mahmood, & Groop, 2010), and hydrological modification (Poveda et al., 2001). Aside from measuring basic vegetation metrics, NDVI has been implemented as a proxy to reliably estimate various biodiversity indicators such as species richness (Gaitan et al., 2013; Parviainen, Luoto, & Heikkinen, 2009; Pau, Gillespie, & Wolkovich, 2012; Psomas et al., 2011), avian abundance (Mcfarland, Van Riper III, & Johnson, 2012), tree species richness (Bawa et al., 2002; Gillespie et al., 2009; Hernandez-Stefanoni, Gallardo-Cruz, Meave & Dupuy 2011; Levin et al., 2007; Mohammadi & Shataee, 2010) and vegetation species abundance (Virtanen et al., 2010).

There are several challenges for using remote-sensing data. First, the use of the data requires an advanced statistics and GIS. Secondly, since imagery is captured from space (or high in the sky) there is a lot of risk on atmospheric contamination (such as clouds, or birds disturbing the image). It makes daily monitoring impossible and it can also be a great challenge to filter the latter all out. Another problem can rise from low resolution of the images produced by public satellite-sensors when analysing small surfaces. Last but not least, satellites are in orbit and thus cannot constantly track the same place all the time.

This study employs the satellite-derived NDVI to assess the ecological impact of a reforestation project. The project is working with farmers in Northern Nicaragua and is led by a Canadian NGO called Taking Root. This organisation employs a payment for ecological services[3] (PES) system that is based on the performances and needs of the farmer. The latter implies that farmers who are participating in the program not only get paid based on the performance of their reforestation interventions, but also based on the needs and available budget of the farmer assessed by the employees of the organisation. Farmers participating in the project receive extra training and technical assistance to make their efforts on the field more effective. When farmers participate in the program they enter into an agreement that requires them to reach certain tree establishment and tree growth milestones. Taking Root employs extension agents (i.e. technicians) that work to help the farmers meet those targets. Because of that, it is of great interest to the organization and the farmers to know what current type of effective training is improving the survival rate of seedlings and growth of trees, thus creating a positive effect in vegetation growth and recovery. Perhaps it is even more critical to identify those interventions or trainings that have a counterproductive effect. Only by monitoring and

[3] (Rodrigo A. Arriagada, 2012) has proven that PES can increase participation in farm forest cover. Reforestation might be more attractive to landowners if they are paid for the ecological services. (Landell-Mills, 2012; Pagiola, 2002; Rietbergen-McCracken, 2007; Schuyt, 2005)

evaluating its efforts, the organization can tune its social and ecological services to deliver a truly positive impact in the long-term.

This study makes several contributions. First, a new dataset is collected and delivered on the vegetation indices of all the farms in the program from the beginning of the project (2010) until April 2018 which is useful for further research. Also, with the input of other perimeters, time-series of NDVI values are available on request. Second, this dataset is merged with remotely sensed activities of the organization itself. The latter can be used by the organization to conduct other research or as a framework for future evaluations. Third, the outcomes of this study can help the organization to improve their activities and trainings in the near future. Fourth, the methodology used in this study can be implemented in future research to investigate similar cases. A little extra effort makes it possible to implement and connect it to the organization's platform through an API so that remotely sensed vegetation index values of the parcels can be tracked and updated at all times. Finally, this research can expose the value of conducting qualitative impact evaluation for other development organisations. Furthermore, the researcher hopes this study can inspire others with the right skill-set to help other small organisations by conducting impact evaluation. In the words of Dr. Kahlil Baker, the co-founder of Taking Root, "Development organizations almost always do all sorts of interventions without even knowing if those interventions actually help! I think that [this study] would be an amazing research contribution."

The activities of interest of this study are different types of technical training sessions. Technicians (a team of 25 full-time Nicaraguan employees) of the organization trained and provided technical assistance to the farmers. The effect of individual trainings is examined and can be easily compared to the disaggregated overall effect of assistance to the farmers in the project. Also, the effects of the visits by different technicians have been taken into account. Examples of technical training are seeding, weed control, pruning trees, and tree nursery. Technicians also give advice on fertilization, harvesting and planting. The main goal for the majority of these technical trainings is to improve the seedlings survival rate and tree growth. Different technical training activities, such as weed control and pruning trees, have shown to contribute to the success of reforestation in previous literature (Le, Smith, Herbohn, & Harrison , 2012). A complete overview of all variables included in this study can be found in Appendix Table II. Only the technical drivers of reforestation are thus examined in this study. Le et al. (2011) have demonstrated that reforestation has many other significant drivers. Those drivers can be grouped into biophysical and technical drivers (e.g. tree species selection, site quality), institutional, management and policy drivers (e.g. forestry support programs, long-term management planning), socio-economic drivers (e.g. livelihood planning, corruption, socio-economic incentives) and project characteristics (e.g. goals and objectives, project implementers) (Le, Smith, Herbohn, & Harrison 2011).

Reforestation has several ecological benefits such as prevention of soil erosion (Marden, 2012), quality improvement of degraded soils (Sauer et al., 2012) and water (Konijnendijk, Ricard, Kenney, & Randrup, 2006), balancing of global gas emission (Keles & Baskent, 2006), carbon sequestration (Ellis et al., 2012) and improvement of local air quality (Nowak, Crane, & Stevens 2006; Paolette, Bardelli, Giovanni & Pecchioli 2011). Reforestation also can have social and academic benefits. It can be used as an educational tool to teach people about sustainability and the environment. Because it also requires human interaction, it is an opportunity to work closely with the locals and a possibility to build community involvement.

In similar studies, time-series estimates of the treatment effect (which analyses differences over time) and supervised classification techniques are used in order to measure either the success of reforestation projects (Yu, 2017) or the effects of more infrastructural activities financed by public sector institutions (BenYishay et al., 2016). Because public databases are easy to access (the OECD's Creditor Reporting System, the International Aid Transparency Initiative), it is often easier to conduct research about public intervention cases. This study provides details about the effectiveness of the methods used in a reforestation project on the individual level of a development organisation. Impact evaluations on the level of a private development organisation are often not easy to access for three reasons: (1) private organisations do not traditionally undertake impact evaluations or (2) do not publish them because many evaluations may have addressed particular and far more narrow problems and finally, (3) because most development organisation evaluations have been undertaken more as an interval learning tool than as a mechanism to provide objective information to external stakeholders (Kruse et al., 1997).

Literature has identified different types of problems with reforestation programs in the past. Most complications can be grouped into two major groups. The first group is mostly caused by natural phenomena. For example, hypersalinization (Elset, 1998), too high sedimentation rates (Elset, 1998), lowering of the water level (Elset, 1998) or mammalian browsing (Keiffer, 1999). The latter can be reduced to tree shelters[4]. The second group is more of a social nature. This type of problem occurs when communities do not welcome new forestry in the first place. The phenomenon happens to reforestation projects that do not bother to find out what local people really want (Dudley, Mansourian, & Vallauri, 2005). Reforestation projects are thus often conducted without understanding why trees are disappearing in the first place (Eckholm, 1979). Sometimes, there has been a misunderstanding between social and economic needs without reference to ecological goals, and vice versa (Le, Smith, Herbohn, & Harrison, 2012). Ecological motivations for reforestation are either about restoring original forest cover or fighting global warming[5] (Ledig & Kitzmiller, 1992). For social development workers, the emphasis lies more in the field of establishing trees that are useful for fuel-wood, fruit, or as windbreaks and livestock enclosures (Le, Smith, Herbohn, & Harrison, 2012). Taking Root handles the socially natured problems by using a PES system which guarantees farmers that who are planting trees and have their own economic motivations to not reject the project goals. The organisation argues that the root cause of deforestation lies in the people's pursuit of better economic opportunities in the first place. Therefore, they argue that the only way to reverse deforestation is by designing reforestation initiatives that provide new economic opportunities for people. Because in this case people are smallholder farmers, the organisation can address deforestation and poverty simultaneously. Also note that the organisation's business model aligns their success with farmers' success. This way, contrary to most other non-profits that earn revenue from providing services regardless of the outcomes, both the farmers and the organisation are encouraged to increase productivity, decrease costs, and increase prices. The latter again motivates the value of this study. It is of great importance to all stakeholders of the organisation that the training sessions and technical assistance are effective to guarantee the long-term success of the whole system.

Several reforestation projects have nevertheless succeeded to reverse deforestation. Le, Smith, Herbohn, & Harrison (2012) have identified several successful reforestation projects and their success drivers in the Philippines. Also, infrastructural projects have proven they can produce significant development gains that can lead to a positive net impact on nearby forests (BenYishay et al., 2016). The impacts of these projects can indirectly reduce pressure on

[4] "Tree shelters are tubes or fencing that wraps around a sapling or seedling in order to make the tree unavailable to the deer" (Shea, 1998).

[5] "Many environmentalists agree that reforestation programs could play a crucial role in balancing our global gas emissions" (Keles, 2006).

forests. By way of illustration, if households have more access to reliable electricity their time collecting firewood for cooking and lightning is reduced (Foster, Lower, & Winkelman, 2011). Likewise, numerous studies have observed that roads can actually reduce forest infringement pressures by improving local development outcomes (Andersen, 2002; Deininger & Minten, 1997; Deng et al., 2011; Qiao & Rozelle, 1998).

Little quantitative academic research has been conducted on reforestation success drivers and potential success evaluation tools. This study delivers new empirical evidence of several technical drivers of a local reforestation program's success. A critical shortcoming of this study is that the relative importance of the different drivers is not known, nor is their impact on other potential indicators of ecological impact or reforestation success.

It is important to emphasize that parcels are being treated differently in the program according to their relative performance. Concretely this means that poorly performing farms (often new farms in the program) get more technical training and assistance than better performing farms. As a consequence, heterogeneous effects of the treatment variables are observed between different performing farms. These complications are handled in Section 4.4.

This study is organized as follows: In Section 2, the study context and the interventions that will be evaluated in this study are described. Section 3 describes data collection efforts. The empirical methodology is described in Section 4, results are presented and discussed in Section 5. Finally, conclusions are made in Section 6.

2 Study context

This study draws on information concerning the number of farmers that are participating in the Taking Roots' reforestation program. The organisation's database provided a sufficiently large number of high-precision parcel locations of the farmers (as described in Section 3.1) stored as perimeters. These perimeters are measured in-situ by the technicians with the use of GPS technologies in their smartphones.

The organisation that is involved in this study, Taking Root, is a Canadian based development organisation co-founded by Doctor Kahlil Baker. The team consists of 25 fulltime employees with up to 1,200 seasonal workers across two countries (Canada and Nicaragua). Taking Root uses a market-based approach to fight deforestation and poverty by developing fully traceable reforestation products with smallholder farmers. The organisation argues that people's pursuit of better economic opportunities is the root-cause of deforestation. That is why Taking Root fights deforestation by providing economic opportunities for people. By working with smallholder farmers deforestation and poverty can be simultaneously addressed. The organisation's business model aligns its success with farmers' success. This way, contrary to most other non-profits that earn revenue for providing services regardless of whether the outcomes are successful, both the management, technicians and the farmers are incentivized to increase productivity, decrease costs, and increase prices.

Taking Root identifies three main challenges that smallholder farmers face that are preventing them, economically, from growing trees on their farms. First, farmers often cannot access capital at an affordable interest rate, which prevents them from making long-term investments. Second, farmers in this area often lack the technical knowledge for advancing production techniques and meeting market requirements. Therefore, further assistance and training is a must to make their efforts more efficient. Third, smallholders receive low prices because they suffer from diseconomies and have to work together with intermediaries. Taking Root's non-profit business model addresses these production and market barriers by providing access to existing markets for reforestation products because it is an incentive for farmers to grow trees on their farms. Without this project, farms would end up being used for non-forested activities like raising livestock which is often more profitable in the short term.

To achieve all this, Taking Root enables and implements different technologies. They created a technological platform that helps them to train and guide the farmers more efficiently. It is because of this platform that they were able to build a data-driven strategy. This strategy helps them to increase their yields, reduce costs, and connect farmers to premium markets. The platform is called Farm-Trace. This platform automatically combines global environmental databases (to collect covariates) and field observations (registered via the local technician's smartphone). This research can also enrich the platform with behavioural insights to track production and make management recommendations in the future. This study is thus a first step to uncover the value of satellite-data for the project, so the platform can be enriched with satellite data in the near future for more efficient and accurate impact evaluation.

All farms that are participating in this program are situated in the Northwest of Nicaragua, a few hundred kilometres from the border with Honduras (see Figure 1). Farms are spread over four municipalities (Limay, Somoto, San Juan de Rio Coco and Palacaguina) and more than two hundred communities. The study area is primarily agricultural and typified by elaborate mosaics of trees and crops associated with multi-story agroforestry systems. A live interactive map of all parcels with key figures can be found on the Taking Root website[6].

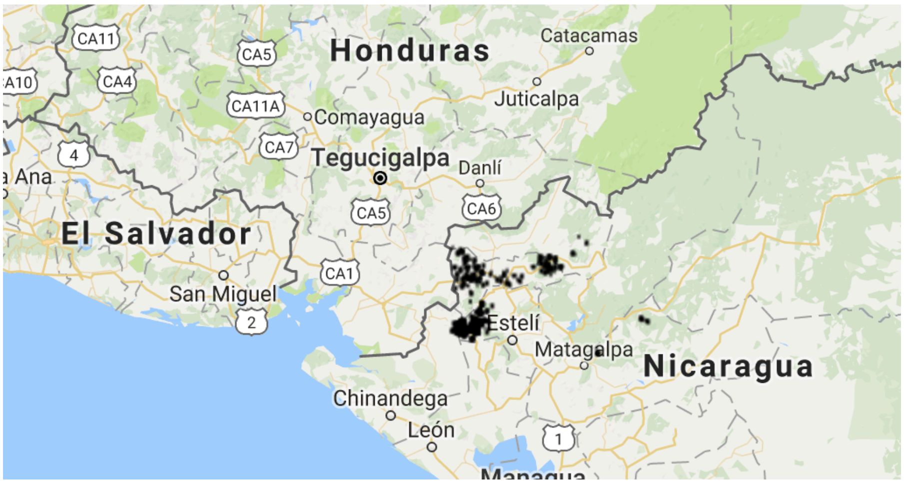
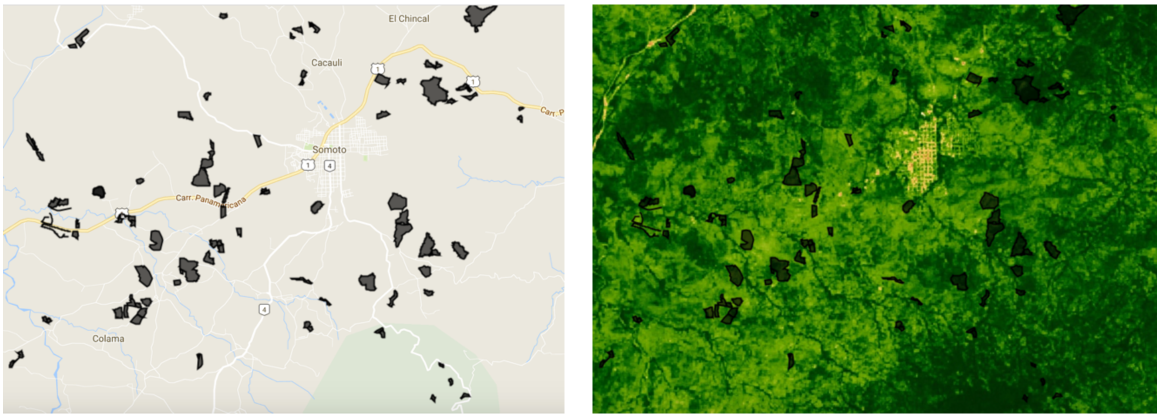


Figure 1 Visualisation of all the farms participating in the program

Figure 2 zooms in on the municipality of Somoto (Nicaragua) and visualizes its surrounding farms. In the left image (Figure 2.a) the farms are projected on a typical road map. In the right image (Figure 2.b) the same farms are projected on a map with a sequential colour palette of NDVI. In the latter, the level of greenness is an indication of the NDVI value where darker green indicates higher NDVI and lighter green indicates lower NDVI. The municipality Somoto can be identified as yellow pixels on Figure 2.b, which are negative NDVI values.

[6] http://www.farm-trace.com/en/Communitree/



(a) *(b)*

Figure 2 A visualisation of the farms around Somoto (Nicaragua) (a) shows a projection of the farms on a road map (b) shows a projection of the farms on a map with a sequential colour palette of NDVI

2.1 Timeframe

Taking Root's smallholder reforestation program collects large amounts of data on a continuous basis. While the first farmer entered the program in 2010, technician's activities were only being tracked since July 2016. Therefore, it is only possible to set a relatively short timeframe for this study (1 year). It is important to emphasize this because the true effects of reforestation are probably more relevant within a mid- to long-term period. Evaluation and analysis of the short-term activities are still of great value for a reforestation project. A lot of attention is dedicated within the first year due to vulnerable trees dying during this time period.

In this study, effects of training on the change in vegetation is measured between September 2016 and September 2017. It is important to choose the same period of the year since NDVI has a strong seasonal trend (Eastman et al., 2013). As a way of illustration, Figure 3 shows the time-series of the NDVI value of the first parcel that was active in the program since 1980. September is chosen because it was the month during the raining season with the most NDVI observations for unique parcels in both years. An advantage of picking a month during the raining season (May-October) for analysis is that vegetation is more healthy and visible. A disadvantage is the higher probability of less data observations because of atmospheric contamination (such as clouds). Farmers can have multiple parcels during this period. In total, NDVI values for 806 parcels from 512 farmers are measured in September for both years.

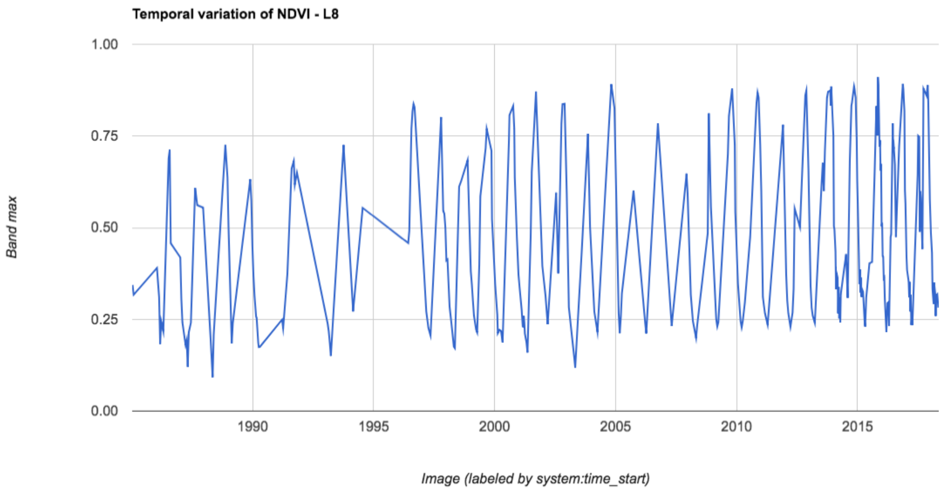


Figure 3 A visualisation of the time-series of NDVI from January 1980 until May 2018 of the first farm participating in the program since 2010

3 Data

3.1 Geographic Unit of Observation

As noted above, this study's analysis is focused on the parcels that are participating in the concerned reforestation program. Farms that are reforested (polygons) are extracted from the organisation's database. Technicians monitor the perimeter of a farm with their smartphone (used for financial data, geo-location data, tree measurements, etc.) after the farmer registers to the program. This dataset is pruned to only include farms that were active before September 2016 and did not leave the program before September 2017. This procedure yields an initial dataset of 768 parcels that are active in the program during the specified timeframe in Section 2.1.

3.2 Outcome data

The primary outcome measure, *Change in NDVI*, varies from -0,437 to 0,465 and reflects the change of the satellite-derived vegetation index from September 2016 to September 2017 ($NDVI_{i,2017} - NDVI_{i,2016}$). Positive values thus reflect an increase in vegetation, while negative values reflect a decrease in vegetation. Higher positive values reflect a more extensive increase in vegetation. Figure 4 shows the distribution of the primary outcome data. While some deviations can be observed in the details, *Change in NDVI* can be assumed to be normally distributed.

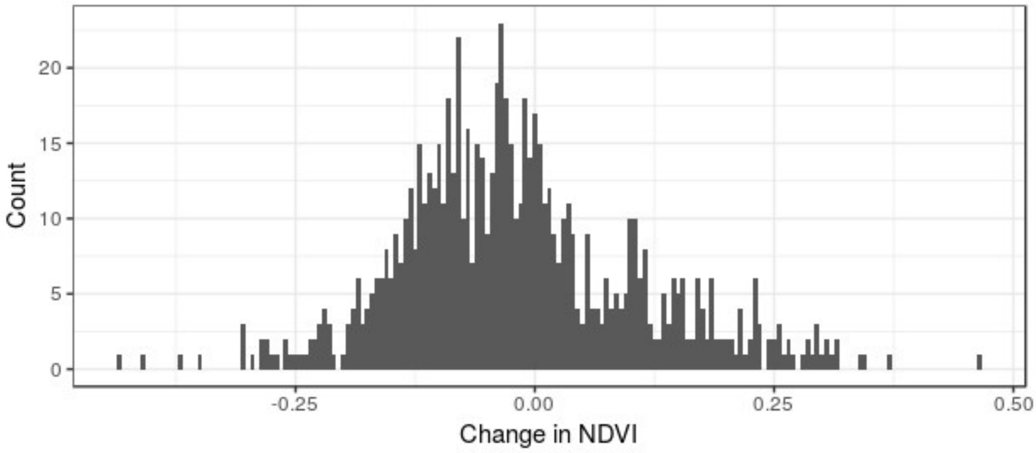


Figure 4 The distribution of *Change in NDVI*

Alternative, non-geospatial, forest performance indicators are Ba_{Ha} (basal area per hectare is the section of the land that is occupied by the cross-section of tree trunks) and TP_{Ha} (trees per hectare is an indicator of tree density). These indicators are currently used to measure the performance of the reforestation efforts but are less accurate since they require extrapolating techniques to make estimations about the entire parcel. Extrapolation is a technique to estimate the value of a variable, beyond the original observation range, based on the basis of its relationship with another variable (Muhammad, 2017). As explained in Box 1, the process thus allows to estimate the Ba_{Ha} and TP_{Ha} values for an entire parcel, by only measuring a smaller part of the parcel. For only 52 parcels historical values of these indicators are available to make in-time comparisons.

Box 1 A review of the Ba_{Ha} and TP_{Ha} performance indicators.

- Farms that are reforested (polygons) consist of many monitoring points representing about 10% of the farm by area.
- Approximately every 12 months, a farm is monitored and all trees within monitoring points are measured (species, height, diameter). These data are then extrapolated to get estimated values at the farm level for indicators as Ba_{Ha} and TP_{Ha}.

Previous literature has shown a significant positive relationship between *NDVI*, Ba_{Ha} and TP_{Ha} (A.A. Souza, 2010; Hadi Fadaei, 2009). In this study weak positive correlation is found between $NDVI_{baseline}$ and Ba_{Ha} $(r = 0.079, df = 662, p = 0.041)$ and the $NDVI_{baseline}$ and TP_{Ha} $(r = 0.0566, df = 662, p = 0.132)$. No evidence of a relationship between *Change in NDVI* and *Change in* TP_{Ha} $(r = 0.135, df = 40, p = 0.3939)$ and between *Change in NDVI* and *Change in* Ba_{Ha} $(r = -0.112, df = 40, p = 0.4795)$ can be found in this study. Unfortunately, the latter is hard to further investigate due to the low number of observations, thus no more attention is paid to these alternative outcome data. If sufficient data would have been available, it could have been useful to compare the results of the models with different reforestation indicators.

3.3 Treatment Data

The biggest challenge for doing impact evaluation in a development organisation is often the gathering of treatment data that are tracking the activities accurately (Ten Hoorn & Stubbe, 2013). Taking Root's smallholder reforestation program collects large amounts of data on a continuous basis and over a long period of time. As mentioned above, the organisation has a full-time staff of 25 people that record almost everything that they do with their smartphones (financial data, geo-location data, tree measurements, etc.), which are synchronized to a central platform.

Box 2 A review of the treatment data.

- On a periodic basis (1 to 29 times per year), technicians visit the farms to help the farmers improve their tree growth. The data records the visits and what exactly the technician did at location.

Per parcel, a few treatment variables are computed from these recorded data. Because the technician not only tracks his/her visits but also the purpose of his/her visit, it is possible to count the number of different training sessions a farmer has received in a period in time. Notice that a farmer receives training and can have multiple parcels (see distribution in Table 1). Thus, the number of trainings a farmer receives is the same for all his/her parcels. This potential source of bias is handled further down. Appendix Table II summarizes the distribution of all treatment variables.

$$\sum_{October\ 2016}^{August\ 2017} Training_i \qquad for\ trainings\ i = 1 \dots n \quad (2)$$

As mentioned in Section 3.1, only farms that were active in the program before September 2016 and which did not quit the program before September 2017 are included in this analysis. In total, 339 unique farmers are identified for this timeframe (with available outcome data), good for a total of 768 parcels which are included in further analysis.

Table 1 Distribution of parcels per farmer

Min	*1st Qu.*	*Median*	*Mean*	*3rd Qu.*	*Max*
1	1	1	2,265	2	24

3.4 Covariate Data

Some covariate data are hosted in this analysis that are known to influence forest cover change (Andam et al, 2008; Hao et al. 2012). The covariates in this study include mean annual precipitation and elevation (extracted from the WorldClim [7] database) that already are integrated in the organisation's database. Table II in Appendix provides summary statistics for these covariates. In this dataset *Elevation* varies from 154 m to 1311 m and has a mean value of 479.8 m. *Mean Annual Precipitation* varies from 967 mm to 1596 mm and has a mean value of 1321 mm. It can be assumed that these natural phenomena also will have a significant and robust explanatory power for the defined response variable in this study. The latter can thus be used for controlling the relations of our variables of interest with the change in vegetation.

4 Methods

4.1 Data collection methods

To obtain the final necessary data for the analysis, different pre-processing methods are executed. The pre-processing methodology can be summarized as the steps in Box 3.

[7] http://www.worldclim.org/

Box 3 A summarization of the pre-processing methodology to obtain NDVI values per parcel.

1. Collect NDVI value per parcel on day t (Google Earth Engine)
2. Compute *Change in NDVI* (R Studio)
3. Extract and compute independent variables from organisation database (R studio)
4. Merge two datasets to a base table ready for analysis (R studio)

Step 1) Collect NDVI value per parcel on day t (Google Earth Engine)

The Sentinel-2 dataset[8] is used to compute the NDVI values (seasonal time-series) of all parcels (polygons). The dataset characterizes spectral bands holding different wavelength values for a pixel on a Sentinel-2 satellite-image. This dataset has been consulted and pre-processed on the Google Earth Engine (GEE) platform. Using the satellite-derived data results in challenges with detecting clouds, cloud shadows and other atmospheric contamination. The data are only valid after being atmospherically corrected[9]. The GEE platform allows researchers to pre-process their geospatial data extractions in a Java-based code-editor and provides in library functions to handle the mentioned challenges. Using the GEE browser-based platform and code-editor has the following advantages: (a) fast debugging, (b) library of specific functions for pre-processing georeferenced-data, (c) fast computation time and (d) easy visualization on geographical maps. The pre-processing methodology on the GEE platform is visualized in Box 1. Note that the Landsat 8 dataset (NASA, 2018) is one of the alternatives to the Sentinel-2 dataset (ESA, 2018) which could be used for this study, but Sentinel-2 is chosen because of its better resolution and thus more accurate observations, which is very important in this study with small surface polygons. While Sentinel-2 has a 10 m resolution for the used bands, Landsat 8 has a resolution of 30 m. Appendix Table I is a summary of all bands that represent a wavelength in the Sentinel-2 database.

[8] The Sentinel-2 dataset contains 13 UINT16 spectral bands representing TOA reflectance scaled by 10000. Each band holds a different wavelength value for a pixel on an image. Also, three QA bands are present where one (QA60) is a bitmask band with cloud mask information (ESA, 2015).

[9] Atmospheric correction involves removing the effects of clouds (and other things) and aerosols. The result is an apparent surface reflectance image. The path of light can change as the radiance/light travels through the atmosphere, suffering wavelength-dependent scattering (UNOOSA, 2017). The reflectance thus must be calibrated. The Sentinel-2 database allows to calibrate its sensors to top-of-atmosphere (TOA) reflectance.

Box 4 Steps in Google Earth Engine

1. Import polygons from Fusion Tables
2. Initialize ROI[10] & timeframe
3. Collection of Sentinel-2 dataset for all pixels in ROI
4. Cloud masking all pixels using the Sentinel-2 QA60 bitmask band
5. Compute the NDVI per pixel on day t
6. Compute NDVI per imported polygon on time t (mean of all pixels)
7. Export table with mean NDVI value per polygon per day t in CSV format

The NDVI value per pixel was calculated as shown in *Eq. 1*. In total, 204 Sentinel-2 images from September 2016 and September 2017 were acquired. Finally, for every parcel, which is in mathematical words a polygon and thus a demarcated group of pixels on an image, an aggregate was computed by taking the mean of the vegetation index of all pixels in that polygon per available[11] day.

Step 2) Compute Change in NDVI (R Studio)

The primary outcome measured reflects the change of vegetation in each parcel[12]. To compute the NDVI value in September 2016 (resp. September 2017), another aggregate was calculated by taking the mean of the vegetation index of the parcels of all days available in September 2016 (resp. September 2017). Then, the difference ($NDVI_{i,2017} - NDVI_{i,2016}$) between the two values is taken as the indicator of change in vegetation.

[10] The ROI in this case is Nicaragua

[11] It happens that on cloudy days, or due to technical defects or other atmospheric contamination, an image becomes unusable or unavailable.

[12] A parcel is a plot of land owned by a farmer that is participating in the project. A farmer can have more than one parcel.

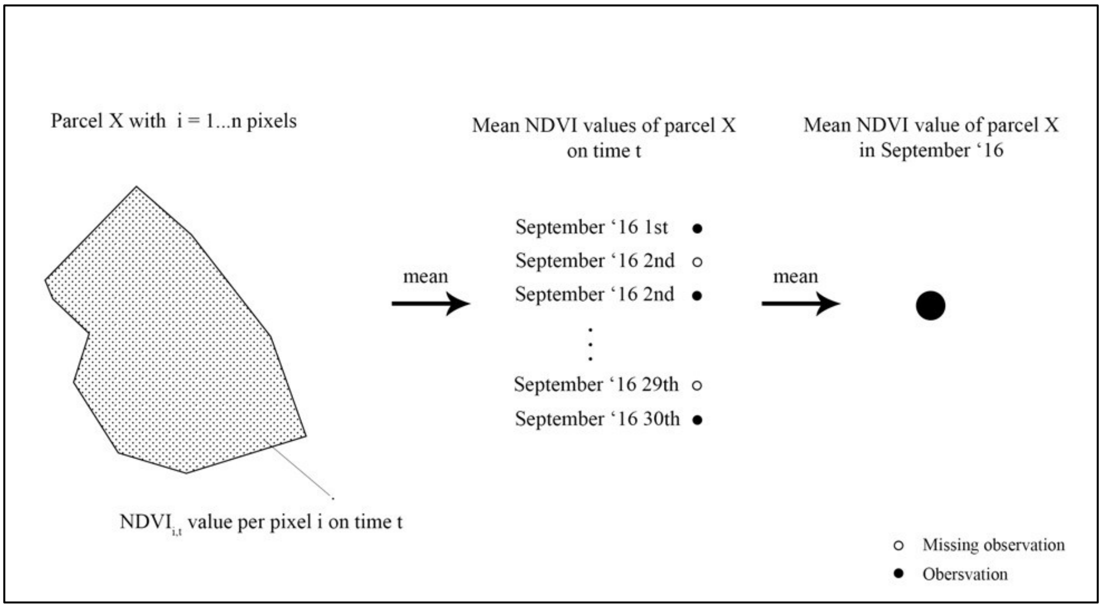


Figure 5 A visualisation of the computation of the average NDVI value in September 2016 of a random parcel.

Figure 5 visualizes the complete process for how the average value of NDVI was computed in September 2016 for a random parcel in the program. The same method was applied to other parcels in the program that year and for all parcels in September 2017. Note that not all days have available data due to atmospheric contamination as explained above.

Step 3) Extract and compute independent variables from organisation database (R studio)

The treatment variables are extracted from the organisation's PostgreSQL database by setting up a database connection in R studio. Tables with information about the parcels, farmers and activities ("activity log") of the technicians were extracted. The most important features that are used in this study are explained in as follows. The parcel table contains technical information about the farmer plot (such as registration date, owner of the parcel, perimeter, program type, etc.). The parcel identifier and perimeter from this table are also used from Step 1 to extract outcome data. The activity log tracks every technician visit, to which farmers on a certain day and for what reason (i.e. type of training). From the latter, the frequency per type of received training per farmer is computed (which are the variables of interest in this study).

Many independent variables can be extracted from this rich database. Yet, is has been suggested to keep models as simple as possible to make them less demanding (Zuur, 2009). The reason is that for every variable added to the model the complexity increases, which makes identification of the impact the variables of interest more challenging. For that reason, some variables are aggregated and others are left out. In this study, model simplification is done by (1) filtering out or aggregating variables that are highly correlated (because they explain the same and thus add not much explanatory power, but require more model complexity), (2) grouping variables that can be grouped intuitively, (3) leaving out variables that cannot function as control variables, that are not relevant to the study and that are not often used in academic research. Arbitrary choices are avoided, by comparing the AIC of GLM models (with control variables) with and without the handled variable to pursue the highest model quality with available data. This process is discussed in detail in the pre-processing Section 4.2 and 4.3 below.

Step 4) Merge the two datasets to a base table ready for analysis (R studio)

Eventually the two datasets resulting from step 2 (outcome data) and step 3 (treatment data) are merged using the parcel identifier. The latter results in a base table that is ready for pre-processing and will be covered in more detail in Section 4.2 and 4.3. A complete overview of all variables available in this interim base table can be found in Table II in Appendix.

4.2 Pre-analysis: defining high quality control variables

The first relations between different control variables are examined. An example of a "null" GLM model *(Eq. 3)* (a model that we used for comparison) for this quality test is specified as

$$y_i = Age_i + \epsilon_i \qquad with\ parcels\ i = 1 \dots n \quad (3)$$

and is then being compared with models that also include (1) *Elevation* and (2) *Mean Annual Precipitation*. The covariate *Elevation* did show in previous literature a positive relationship with the vegetation index (Zhong, 2012), but in this context an inverse relationship is found with the $NDVI_{baseline}$ *($r = -0.500$, $df = 766$, p-value = 0.000)*. In this study, *Elevation* is also inversely correlated with the *Mean Annual Precipitation ($r = -0.658$, $df = 766$, p-value = 0.000)* which asks for an intervention. Only the *Mean Annual Precipitation* is left in the model which proves to be increasing the quality of the model significantly *($AIC_{null} = -993.31$, $AIC_{Mean\ Annual\ Precipitation} = -1008.99$)*, when compared to *Elevation* that is decreasing model quality *(AIC_{null} =*

-993.31, $AIC_{Elevation}$ *= -991.01)*. If both would have improved model quality, this would have been a possibility to make one index of both variables.

Parcels longer in the program are assumed to make less marginal improvement than newcomers *(r = -0.058, df = 766, p-value = 0.1039),* but this assumed weak relationship in this case. Therefore, the same quality test is done for the *Age* variable (compared to a null model with *Mean Annual Precipitation* only). The results show no motivation to keep the variable in the model *(*AIC_{null} *= -1020.473,* AIC_{Age} *= -1008.999)*. Furthermore, intuitively it is expected that the $NDVI_{baseline}$ should have a positive relationship, as an indication of reforestation success on the long-term *(r = 0.398, df = 766, p-value = 0.000)*.

Also, better performing parcels leave less room for improvement than bad performing parcels. A strong indication of this relationship is found in this study by testing correlation between *Change in NDVI* and the *Relative Performance Score* (*r = -0.708, df = 766, p-valve = 0.000).* It goes without saying that the *Relative Performance Score* strongly contributed to the quality of the model (AIC_{null} *= -1020.473,* $AIC_{Relative\ Performance\ Score}$ *= -1535.246)*. The *Relative Performance Score* is computed out of the $NDVI_{baseline}$ but contains more info about the benchmark group of a parcel and is therefore more preferred ($AIC_{Relative\ Performance\ Score}$ *= -1535.246,* $AIC_{NDVI\ baseline}$ *= -1493.793)*. The calculation of the *Relative Performance Score* and more info on benchmarks groups can be found in Section 4.4.

Regional identifiers (*Municipality* and *Community*) and *Agroforestry Type* are left out for the same reason. We observe a significant quality drop for *Community* (AIC_{null} *= -1535.246,* $AIC_{Community}$*= -1311.494)* that is probably due to the 200+ communities and a small decrease for *Municipality* (AIC_{null} *= -1535.246,* $AIC_{Community}$*= -1519.993)*. Although a strong significant difference is found for the relation between *Change in NDVI* and the different types of *Agroforestry type (F(5,762) = 3.279, p-value = 0.00616)* it does not add enough quality to the model (AIC_{null} *= -1535.246,* $AIC_{Agroforestry\ type}$*= -1511.251)*.

The *Relative Benchmark Group*[13] is a categorical variable and is assumed to be an important source of bias and is therefore added to the base model with control variables. By adding it to the model another slight increase is observed in model quality (AIC_{null} *(4)= -1535.246,* $AIC_{Relative\ Benchmark\ Group}$ *(7) = -1539.685).*

As a result from the pre-analysis *Mean Annual Precipitation*, the *Relative Performance Score*, and *Relative Performance Group* are selected as qualitative control variables that will be used for further research. The distribution of these variables can be found in Table II in Appendix.

4.3 Pre-analysis: defining the variables of interest

Nine relevant training types received by the farmers in the specified timeframe are extracted from the organisation's database. To reduce further model complexity the variables are grouped intuitively. *Weeding, Pruning, Planting and Tree Nursery* are aggregated into *Technical Training. Introduction*, *Diagnose* and *Certificate* are aggregated into *Sustainability Training*. *Other work* is left as a separate category because it is already an aggregation of unclassified tasks. Also, *Materials Delivery* remains untouched. This transformation further improves the model quality ($AIC_{No\ Transformation}$ *= -1460.838,* $AIC_{Transformation}$*= -1503.110).*

The variables of interest used in the further analysis are thus $Technical\ Training$, $Sustainability\ Training$, $Other\ work$ and $Materials\ Delivery$. A complete overview of all variables used in further analysis can be found in Table II in Appendix.

4.4 Empirical methodology

R studio (2018) was used for data analysis. The process for calculating the outcome data (see Section 4.1) used to measure reforestation success is described in Table 3. This study follows the strategy that is recommended by Gelman & Hill (2007) and Verbeek (2008) for building a model with random effects. First, bivariate analysis is conducted to identify association between the independent and dependent variables (see Appendix Table II), and simple linear models are built to explore associations between the different independent

[13] More info and calculations can be found in Section 4.4

variables. Then, a more complex multilevel model with varying-intercept is set up to better understand the by-farmer variability in the variables of interest. Each of the simpler models can be informative in its own right, and they help to understand the partial pooling[14] in a multilevel model (Gelman & Hill, 2007).

First, the interactions between the control variables and the dependent variables are being studied. Then, these control variables are held constant and two other models are built to understand the effect of the different treatment variables of interest. The first model assumes a homogeneous effect of the different types of training by aggregating - or simply taking the sum of - all the number of training sessions received. The second model considers a heterogeneous effect of the different training sessions by allowing the coefficient to vary over different types of sessions. Also, because it can be expected that the person who gives the training can have a significant influence on the training effect, the robustness of the latest model is tested by adding the visits of the different technicians. Eventually, the interactions between the different *Relative Performance Groups* and the types of training are being studied, with respect to the design of the study context, these interactions should give a deeper understanding of the effect of the variables of interest. Before conducting linear regressions, a preliminary analysis was conducted to ensure no violation of the assumptions of normality, linearity, multi-collinearity, and homoscedasticity among the variables. A summary of this preliminary analysis can be found in the Appendix. The simple linear regression models can be written mathematically as

$$y_{ij} = \alpha + X_{ij}\beta + \varepsilon_{ij} \qquad for\ parcels\ i = 1 \dots n \quad (4)$$

$$y_{ij} = \beta_1 X_{j1} + \cdots + \beta_k X_{jk} + \varepsilon_i \ \ for\ farmers\ J = 1 \dots j \ (5)$$

where α denotes a constant term, the errors ε_i have independent normal distributions with mean 0 and standard deviations σ . The vector X_i denotes the variables *Age, Elevation, Mean Annual Precipitation* and *Relative Performance Score* which, can function as control variables since they all are expected to have a logical relationship with the response variable. All included control variables are held constant during the analysis to further understand the relationship between the dependent variable and the variables of interest. An overview of all models can be found in Table 3.

---

[14] Partial pooling is a synonym for hierarchical or multilevel models.

As introduced above, heterogeneous effects are expected to exist for different reasons. First, the response variable is defined in a lower dimension than the variables of interest[15]. Second, heterogeneous groups are also expected to exist because the organisation reports that different performing farms are being treated differently by the technicians. This last statement can affirmed for *Sustainability training* as determined by one-way ANOVA *(F(3,764) 5.477, p = 0.001), Material delivery (F(3,764) = 7.055, p = 0.000)* and *Other work (F(3,764) = 2.987, p = 0.030)*. The groups are treated in the same way for *Technical training (F(3,764) = 1.086, p = 0.354)*. Relatively worse performing farms receive on average more sustainable support and materials, and vice versa for relatively better performing farms. To be able to catch this potential source of bias, a relative performance score (% of benchmark value) for each farm was computed by comparing its $NDVI_{baseline}$ performance to a benchmark (see Eq. 6). This variable captures the effect of bad farms continuing to perform poorly. It is thus also an alternative for the $NDVI_{baseline}$ *(r = 0.755, df = 766, p = 0.000),* but it is believed by the researcher to be more comprehensive because it also takes the group-related benchmark into account. If a farm has a NDVI value of 0,20 in the previous time period and the benchmark is 0.5, it would have (0.20/0.5) 40% of the benchmark value. A benchmark value was computed for each group of farms with the same technical specifications (type of agroforestry, program type, and rounded years since registration to the program). Eventually, four groups were extracted from this relative performance score. Quantile 1 are the relatively worst performing farms up to quantile 4 which are relatively the best performing farms as shown in Table 2.

$$Relative\ Performance\ Score_i = \frac{NDVI_{i,Sept2016}}{Benchmark_k} \qquad (6)$$

Table 2 Distribution of the Relative Performance Score

Min		*1st Qu.*		*Median*		*3rd Qu.*		*Max*
0.2569		0.8719		1.0205		1.1065		1.6311
	Relative Performance Group 1		Relative Performance Group 2		Relative Performance Group 3		Relative Performance Group 4	

[15] While the dependent variable is observed per $parcel_i$ from $farmer_j$, the variables of interest are observed per $farmer_j$.

4.4.1 Pooled model

After only estimating the effects of the control variables, a pooled model introducing the variable of interest (see Table 3 (2)) was fitted. A pooled model ignores the fact that there are heterogeneous effects for different types of training and that the farmers own several parcels. It gives an idea of the overall mean effect of all the treatment variables on the *Change in NDVI*. The variable is called *Frequency of training,* which is the total sum of all training sessions a farmer of a parcel received in the defined timeframe. The specification for each parcel that was estimated can be found in Table 3.

4.4.2 Pooled model with heterogeneous effects

The pooled model with heterogeneous effects (see Table 3 (3)) examines the individual effect of separate trainings which are the variables of interest in this study. Furthermore, because it can be assumed that technicians who provide these training sessions have an effect on the different variables of interest, another model is built to check robustness of the previous findings by adding the *$Technician_m$ visit* variables. These dichotomous variables are thus added to the base model in order to capture the effect of the technicians. If a $technician_m$ visited a $farmer_j$ at least once in the defined timeframe, the variable $technician_m$ visit will be 1, and 0 otherwise. The full specification and results of this model can be found in Table 3.

Finally, as mentioned above, an interaction model (Table 3 (5)) is built with an interaction term between the different variables of interest and the *Relative Performance Group*. Worse performing farms should receive more assistance by the technicians and therefore other effects are expected across groups.

4.4.3 Random effect model

In all the models created above, the response variable (per parcel of a farmer) is measured in a different dimension than the variables of interest (per farmer). Because no farmer indicator was implemented in the previous models, average results for different parcels of farmers were obtained. Yet, we may expect to have similar effects that will be obtained for parcels owned by the same farmer, which might form another potential source of bias. Only by including this effect, the above models and findings can be tested to be robust. Unfortunately, it is not possible to include a dummy identifying farmers as a fixed effect because this dummy would have the same dimension as the variables of interest, in which case, the latter would be dropped from the model. If adding farmers as fixed effect would have been a possibility, a Hausman specification test could have been used to weigh the fixed effects model against a model with a farmer-specific random effect and select the most efficient one. In this case, the only option is to build a model with the farms as random effect. This implies that it is assumed that farmer-specific random effects are not correlated with independent variables. Only if this assumption holds in reality, the random effect model can be called consistent (Verbeek, 2008).

Thus, in order to further model unexplained variance and to better understand the by-farmer variability in the treatment variables, the *Farmer* variable is included as a random effect. A random effects model can be thought of as a method for compromising between the two extremes of excluding farmers from a complete pooling model, or estimating separate models per farmer (Gelman, 2007). In this method only varying-intercepts for the different are assumed to catch the random effect. Varying intercepts can be interpreted as interactions between an individual level intercept (parcel) and farmer intercept (Gelman & Hill, 2007). Adding the farmer as a random effect improves the model quality significantly *(AIC_{null} = -1539.685, $AIC_{Random\ Effect}$ = -1740.165).*

A model with random effects leads to higher model complexity. Compared to the pooling model *(Eq. 3)* has four times as many vectors of second-level coefficients *(a,b)* and potential correlation between the random effect and the predictors. Assume that x_{ij} and the varying intercepts α_j correlate. If this correlation is not modelled, it will be absorbed into the error term η_{j1}, which results in the violation of a key Gauss-Markov assumption (Verbeek, 2008). If *Eq. 8* is substituted in *Eq. 7* the error terms combine to create a new error term. This new regression error now correlates with the predictor in the model. This violation may result in poor estimates of parameter uncertainty (Verbeek, 2008). These complications are handled in Appendix. The notation *i* used for individual parcels and *ij* for the Farmer *j* that owns parcel *i* (Verbeek, 2008).

$$y_{ij} = \alpha_j + \sum \beta\, x_{ij} + \varepsilon_{ij}, \qquad for\ parcels\ i = 1 \dots n \qquad (7)$$

$$\alpha_j = a_0 + \sum b_0\, u_j + \eta_j \qquad for\ farmers\ j = 1\ , \dots , J. \qquad (8)$$

Here, x_i and u_j represent predictors at the parcel and group levels, respectively, and are independent error terms at each of the two levels. The number of farmers *(J = 339)* are less than *n (= 768)*, the sample size of the lower-level model. As shown in Figure 4, the response variable is assumed to be normally distributed which is an assumption for the mixed model that is built.

Table 3 Summary of all linear regression models

	Dependent variable:					
	Change in NDVI					
			generalized least squares			*linear random effects*
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.482*** (0.035)	0.477*** (0.036)	0.454*** (0.037)	0.417*** (0.044)	0.470*** (0.039)	0.420*** (0.042)
Control variables						
Mean Annual Precipitation	-0.0001*** (0.00002)	-0.0001*** (0.00002)	-0.0001*** (0.00002)	-0.00003 (0.00003)	-0.0001*** (0.00002)	-0.00003 (0.00003)
Relative Performance Score	-0.359*** (0.036)	-0.360*** (0.036)	-0.358*** (0.036)	-0.358*** (0.036)	-0.362*** (0.037)	-0.358*** (0.032)
Relative Performance Group 2	-0.050*** (0.012)	-0.050*** (0.012)	-0.050*** (0.012)	-0.048*** (0.012)	-0.065*** (0.016)	-0.048*** (0.010)
Relative Performance Group 3	-0.043** (0.020)	-0.043** (0.020)	-0.045** (0.020)	-0.042** (0.020)	-0.046** (0.021)	-0.047*** (0.017)
Relative Performance Group 4	-0.064*** (0.015)	-0.064*** (0.015)	-0.061*** (0.015)	-0.061*** (0.015)	-0.080*** (0.018)	-0.055*** (0.013)
Variables of interest						
Frequency of training		0.0004 (0.001)				
Sustainability training			0.003 (0.002)	0.002 (0.002)	0.003 (0.004)	0.003 (0.002)
Technical training			0.0005 (0.002)	0.001 (0.002)	0.0001 (0.004)	0.001 (0.003)
Materials delivery			-0.009*** (0.003)	-0.013*** (0.003)	-0.016** (0.007)	-0.010** (0.005)
Other work			0.002 (0.002)	0.0001 (0.002)	-0.002 (0.004)	0.001 (0.002)
Technician visits						
Technician 1 visit				-0.049*** (0.014)		
Technician 2 visit				0.089 (0.063)		
Technician 3 visit				0.004 (0.025)		
Technician 4 visit				0.063*** (0.020)		
Technician 5 visit				-0.012 (0.008)		
Technician 6 visit				0.024 (0.016)		
Technician 7 visit				0.045*** (0.014)		
Technician 9 visit				-0.012 (0.012)		
Technician 11 visit				0.013 (0.008)		

Table 3 *Continued*

	Dependent variable:

	Change in NDVI					
			generalized least squares			*linear mixed effects*
	(1)	(2)	(3)	(4)	(5)	(6)
Technician 12 visit				-0.025*** (0.009)		
Technician 15 visit				-0.018 (0.022)		
Technician 16 visit				-0.018 (0.026)		
Technician 17 visit				0.005 (0.018)		
Technician 19 visit				0.013 (0.022)		
Interaction effects						
Sustainability training * Relative Performance Group 2					-0.005 (0.005)	
Sustainability training * Relative Performance Group 3					0.00004 (0.005)	
Sustainability training * Relative Performance Group 4					0.003 (0.006)	
Technical training * Relative Performance Group 2					0.009 (0.006)	
Technical training * Relative Performance Group 3					-0.007 (0.005)	
Technical training * Relative Performance Group 4					0.005 (0.006)	
Materials delivery * Relative Performance Group 2					0.009 (0.009)	
Materials delivery * Relative Performance Group 3					0.009 (0.010)	
Materials delivery * Relative Performance Group 4					0.008 (0.009)	
Other work * Relative Performance Group 2					0.004 (0.005)	
Other work * Relative Performance Group 3					0.006 (0.005)	
Other work * Relative Performance Group 4					0.005 (0.006)	
Observations	768	768	768	768	768	768
Log Likelihood	776.843	770.633	762.586	747.911	718.344	861.829
Akaike Inf. Crit.	-1,539.685	-1,525.267	-1,503.171	-1,445.821	-1,390.689	-1,699.659
Bayesian Inf. Crit.	-1,507.234	-1,488.190	-1,452.234	-1,330.520	-1,284.550	-1,644.091

Significance codes: *** p-value < 0.01 (2-tailed), ** p-value < 0.05 (2-tailed), * p-value < 0.10 (2-tailed)

5 Results and discussion

5.1 Relationships between different types of training and the change in the Normalized Difference Vegetation Index (NDVI)

While previous literature has observed technical support (Borglandan et al., 2001) and sustainable management (Nawir & Rumboko, 2007) to be important drivers for the success of reforestation programs in different developing countries, no significant training support drivers of reforestation success in terms of *Change in NDVI* can be found in this context (Table 3). The beta weights suggest that *Materials Delivery* explained most of the variance of all variables of interest showing a negative relationship (Table 3) with the change in vegetation. This relationship is visualised in Figure 6. Plots of other variables can be found in Appendix. Findings are proven to be robust to the effect of technicians, relative performance and farmers. These features are proven to be three important sources of variance that have no effect on the relationships between the variables of interest and the response variable. Interactions effects between the different *Relative Performance Groups* and training frequencies showed no evidence of heterogeneous effects for the types of training between the groups.

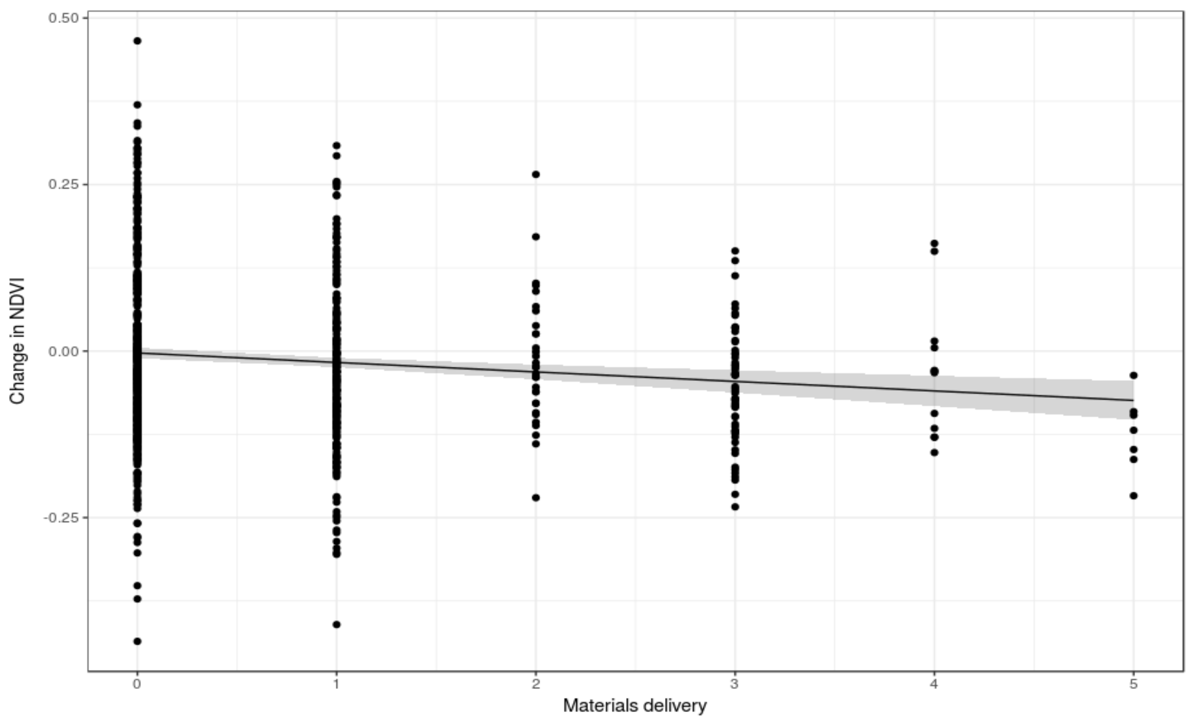


Figure 6 Visualisation of the negative relationship between *Change in NDVI* and *Materials Delivery*

Taking Root explains that *Materials Delivery* is usually related to nurseries. Only farms that are in need to plant new trees require new materials. These materials are then used for clearing the land to be able to plant the trees. Finding a negative relationship between *Materials Delivery* and *Change in NDVI* in short term is thus a logical finding.

5.2 Implications for reforestation planning and management

This study revealed that reviewing the ecological impact of a reforestation project in the short-term can provide useful insights. The risk of focusing on this ecological indicator in isolation to measure the impact runs the risk of focusing on the symptoms of poor performance rather than the underlying causes. Nonetheless, understanding the true meanings of these results allows an organization to identify points of leverage where change can have a broad systematic effect.

The next step could be to build a management assessment framework to allow for more efficient and effective management on the level of an individual technician. The framework result can be optimized thus identifying and reducing unintended consequences.

However, in developing a management assessment framework based on the results created, it is important to recognize some of the limitations of this study. First, the understanding that was gained in this study of the driver/indicator relationships only extends to those variables for which data were collected and the researcher was aware that the actual working methodology contains many soft variables that are difficult to measure but are important to the working of the project. Soft variables that are related to the farmer were included by adding the random effect of the farmer. Second, understanding the relationships and reasons for significant relationships between the independent and the dependent variable may not be obvious and the results do not necessarily fully explain why variables are related. Luckily, in this case for *Materials Delivery*, a fairly certain explanation for the relation can be assumed as mentioned above. Third, this static analysis does not take the continuous character of the project into account. Technicians are making different decisions every day for different reasons that cannot be tracked, so it cannot be used to simulate continuous dynamic behaviour over time. Fourth, this study is conducted on a short-term basis. Sometimes, after an area is cleared for reforestation, there is thus a period of decrease in vegetation so that the trees can be planted. In the short-term period, this is thus captured as a decrease in vegetation and thus an important observation given the short period of this study. Also, reforestation programs, and development programs in general, often aim to deliver positive results in terms of forest expansion in the long-term (Perrin, 2012). Last but not least, only if the assumption that farmer-specific random effects are not correlated with control variables holds in reality, the random effects model can be called consistent.

It is also important to note that in this study and also in previous literature reported that only a varying-intercept random effects model (Table 3 (6)) is constructed. It makes much more sense to assume varying-slopes for all farmers as well (Winter, 2013). Researchers in different fields have shown through simulations that mixed-models without varying-slopes are often conservative (Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009) which leads to a high type I error rate (easier to find significant results that are actually due to chance). (Schielzeth & Forstmeier, 2009) therefore suspect that many published findings have too narrow confidence intervals. It is thus recommended for further research to build a mixed-model with varying-slopes to reduce type I errors and reduce residual variance by accounting for between-parcel variation in slopes. The latter makes it easier to detect treatment effects that are applied between parcels, hence reducing type II errors as well (Schielzeth & Forstmeier, 2009). Given these limitations, it is recommended to the concerned organization to pay particular attention to the following when adjusting their reforestation methods.

By maintaining a long-term commitment from both the farmers and technicians, the information generated from their interaction will be a significant source of bias for vegetation changes. As mentioned earlier, by implementing the PES system and a performance–based business plan, Taking Root is expecting farmers to establish and maintain their own economic motivations in the long-term. Through the reforestation program, farmers can take out payments in advance to cover expensive establishment costs. In the literature, past reforestation projects success has been linked to similar profit-sharing arrangements and planting payments.

Developing countries with large population and poor economy, like Nicaragua would not truly benefit from a good reforestation program that is only successful years later. Through implementation of various integrated production systems (i.e. agroforestry, reforestation, woodcrafts, timber production, coffee production, etc.) can lead to long-term success of reforestation resulting in several income generating opportunities on both a short and long-term basis. Farmer participating the Taking Roots' program already profit from different income streams from other production systems (to connect farmers to buyer, trees are selectively harvested to make woodcrafts). New incentives produced by these streams of income, motivate farmers to grow more trees. Another key to the success of the reforestation efforts that is heavily depending on the species of the trees and the ability of the forest that will satisfy the demand of the local people and their livelihoods.

5.3 *Recommendations to the organization*

Taking the limitations into account that were described in Section 5.2 several recommendations can be made for the organisation. First, the use of geo-spatial data has proven to be an efficient tool for evaluating the project success in terms of ecological impact. It is believed by the researcher that it is worth the investment to enrich the Farm-Trace platform with these techniques. NDVI might be a more accurate and cheaper performance indicator than the current Ba_{Ha} and TP_{Ha} forest measures. There are several services that allow the extraction of the NDVI for a polygon. Google Earth Engine has a well-documented Python API that allows more complex analysis and access to different datasets, but an easier option might be to use a cURL-request[16] from the Proba Vegetation programme (ESA) that can be directly connected to the organisation's database as a first step. The latter allows you to retrieve a time series for a given point or polygon with simple parameters. Second, it is strongly advised to conduct periodically impact evaluations. Even short-term analysis, like in this study conducted, might give insights to hidden inefficiencies or opportunities that enables greater value on the (mid) long-term. It also can be used to build prescriptive analysis to give management recommendations based on behavioural insights. For example, it is possible to predict when new materials need to be delivered to a specific farmer. Prescriptive analysis tools can thus increase management efficiency significantly by providing, to individual technicians, guidelines on which farmer, needs what at what time and what type of training; taking to account different time- and space-related constraints. Of course, it is important to also stress the value of mid- and long-term analysis, which will deliver totally different insights that are at least as important to consider. Publishing results of impact evaluations can be an effective way to promote the organisation's innovative effort and efficiency (Zhong, 2018). The researcher arguments that the credibility of the analysis can be higher if it is outsourced to an independent organisation and if it is done periodically. Third, the data-driven approach of the organisation demonstrates their beliefs about the value of gathering data, but it is important to stress that working on techniques to improve the quality and quantity of the data is always a good idea.

[16] https://proba-v-mep.esa.int/sentinel-web-services

6 Conclusions

Using satellite-derived data is shown to be a convenient tool to do ecological impact evaluation of a reforestation program on the level of a development organization. Robust models are built to evaluate different types of training that smallholder farmers receive to improve their reforestation efforts. This study cannot provide any effective evidence of this type of technical support in order to improve reforestation efforts in the short-term. On the other hand, a negative relationship between the delivery of materials and the change in vegetation is found. The relationship shows that when a farmer needs to plant new trees, materials are delivered by the organization must first clear the land to be able to plant the new trees. Based on these findings and discussion, it can be concluded that also short-term impact evaluations of reforestation programs are useful to identify dysfunctionalities and opportunities that can have a systematic impact on the system on the long-term. The methodology of this study also can be used for evaluations on the medium to long-term. The significant negative relationship of the delivery of materials also shows that interpretation of these models needs to be conducted carefully, with full respect to the study context and in dialogue with concerning organisation's management.

References

Andam, K. S., Ferraro, P. J., Pfaff, A., Sanchez-Azofeifa, G. A., & Robalino, J. A. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences, 105*(42), 16089-16094. doi:10.1073/pnas.0800437105

Andersen, L. E. (2002). *The dynamics of deforestation and economic growth in the Brazilian Amazon*: Cambridge University Press.

Arriagada, R. A., Ferraro, P. J., Sills, E. O., Pattanayak, S. K., & Cordero-Sancho, S. (2012). Do payments for environmental services affect forest cover? A farm-level evaluation from Costa Rica. *Land Economics, 88*(2), 382-399.

Asrar, G., Fuchs, M., Kanemasu, E., & Hatfield, J. (1984). Estimating Absorbed Photosynthetic Radiation and Leaf Area Index from Spectral Reflectance in Wheat 1. *Agronomy journal, 76*(2), 300-306.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bawa, K., Rose, J., Ganeshaiah, K., Barve, N., Kiran, M., & Umashaanker, R. (2002). Assessing biodiversity from space: an example from the Western Ghats, India. *Conservation Ecology, 6*(2), 7.

BenYishay, A., Parks, B., Runfola, D., & Trichler, R. (2016). *Forest cover impacts of Chinese development projects in ecologically sensitive areas.* Paper presented at the SAIS CARI 2016 Conference. October.

BenYishay, A., Rotberg, R., Wells, J., Lv, Z., Goodman, S., Kovacevic, L., & Runfola, D. (2017). Geocoding Afrobarometer rounds 1–6: methodology & data quality. *AidData*.

BenYishay, A., Runfola, D., & Buntaine, M. (2015). *Going Geospatial with Impact Evaluations*. Retrieved from https://usaidlearninglab.org/events/geospatial-impact-evaluations:

BenYishay, A., Runfola, D., Trichler, R., Dolan, C., Goodman, S., Parks, B., . . . Anand, A. (2017). A Primer on Geospatial Impact Evaluation Methods, Tools, and Applications. *AidData Working Paper #44.*

Bishop, J., & Pagiola, S. (2012). *Selling forest environmental services: market-based mechanisms for conservation and development*: Taylor & Francis.

Borlagdan, S.B., Guiang, E.S., Pulhin, J.M., 2001. *Community-Based Forest Management in the Philippines: A Preliminary Assessment.* Institute of Philippine Culture, Ateneo de Manila University, Philippines.

Buchanan, G. M., Parks, B. C., Donald, P. F., O'Donnell, B. F., Runfola, D., Swaddle, J. P., & Butchart, S. (2016). The Impacts of World Bank development projects on sites of high biodiversity importance. *url: http://aiddata. org/sites/default/files/wps20_world_bank_ biodiversity_0. pdf (visited on 11/21/2016)*.

Buntaine, M. T., Hamilton, S. E., & Millones, M. (2015). Titling community land to prevent deforestation: An evaluation of a best-case program in Morona-Santiago, Ecuador. *Global Environmental Change, 33*, 32-43.

Butt, M. J., Waqas, A., Mahmood, R., & Group, H. R. (2010). The combined effect of vegetation and soil erosion in the water resource management. *Water resources management, 24*(13), 3701-3714.

Campbell, S., & Hofmann, S. C. e. a. (2014). Independent external evaluation UN Peacebuilding Fund Project portfolio in Burundi 2007–2013. 85 p.

Combalicer, M. S., Kim, D., Lee, D. K., Combalicer, E. A., Cruz, R. V. O., & Im, S. (2011). Changes in the forest landscape of Mt. Makiling Forest Reserve, Philippines. *Forest Science and Technology, 7*(2), 60-67. doi:10.1080/21580103.2011.572615

Deininger, K., & Minten, B. (2002). Determinants of deforestation and the economics of protection: an application to Mexico. *American Journal of Agricultural Economics, 84*(4), 943-960.
Deng, X., Huang, J., Uchida, E., Rozelle, S., & Gibson, J. (2011). Pressure cookers or pressure valves: do roads lead to deforestation in China? *Journal of Environmental Economics and Management, 61*(1), 79-94.
Dolan, C., Grepin, K., G., M., & J., T. (2017). *The Impact of an Insecticide Treated Bednet Campaign on All-Cause Child Mortality: A Geospatial Impact Evaluation from the Democratic Republic of Congo.* Retrieved from
Dudley, N., Mansourian, S., & Vallauri, D. (2005). Forest landscape restoration in context. In *Forest Restoration in Landscapes* (pp. 3-7): Springer.
Eastman, J. R., Sangermano, F., Machado, E. A., Rogan, J., & Anyamba, A. (2013). Global trends in seasonality of normalized difference vegetation index (NDVI), 1982–2011. *Remote Sensing, 5*(10), 4799-4818.
Eckholm, E. (1979). *Planting for the future: forestry for human needs*.
Ellis, S., Holahan, C., Klynstra, J., Madison, J., & Robison, R. (2013). The feasibility of a local reforestation project at Colgate University. In.
Elster, C. (2000). Reasons for reforestation success and failure with three mangrove species in Colombia. *Forest Ecology and Management, 131*(1-3), 201-214.
ESA. (2015). *Sentinel-2 User Handbook*.
Fadaei, H., Sakai, T., Yoshimura, T., Moriya, K., & Torii, K. (2009). *Relationship between tree density and vegetation index of juniper forest in the northeast of Iran.* Paper presented at the Proceedings of the 30th Asian Conference on Remote Sensing (ACRS), CD-ROM.
Fensholt, R., & Sandholt, I. (2005). Evaluation of MODIS and NOAA AVHRR vegetation indices with in situ measurements in a semi-arid environment. *International Journal of Remote Sensing, 26*(12), 2561-2594.
Foster, J., Lowe, A., & Winkelman, S. (2011). The value of green infrastructure for urban climate adaptation. *Center for Clean Air Policy, 750*, 1-52.
Funk, C. C., & Brown, M. E. (2006). Intra-seasonal NDVI change projections in semi-arid Africa. *Remote Sensing of Environment, 101*(2), 249-256.
Gaitán, J. J., Bran, D., Oliva, G., Ciari, G., Nakamatsu, V., Salomone, J., . . . Humano, G. (2013). Evaluating the performance of multiple remote sensing indices to predict the spatial variability of ecosystem structure and functioning in Patagonian steppes. *Ecological indicators, 34*, 181-191.
Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*: Cambridge university press.
Gillespie, T., Saatchi, S., Pau, S., Bohlman, S., Giorgi, A., & Lewis, S. (2009). Towards quantifying tropical tree species richness in tropical forests. *International Journal of Remote Sensing, 30*(6), 1629-1634.
Guiang, E. S., Borlagdan, S. B., & Pulhin, J. M. (2001). Community-based forest management in the Philippines: a preliminary assessment. *Institute of Philippine Culture, Ateneo de Manila University*.
Günter, S., Gonzalez, P., Álvarez, G., Aguirre, N., Palomeque, X., Haubrich, F., & Weber, M. (2009). Determinants for successful reforestation of abandoned pastures in the Andes: soil conditions and vegetation cover. *Forest Ecology and Management, 258*(2), 81-91.
Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., . . . Loveland, T. (2013). High-resolution global maps of 21st-century forest cover change. *science, 342*(6160), 850-853.

Hao, F., Zhang, X., Ouyang, W., Skidmore, A. K., & Toxopeus, A. (2012). Vegetation NDVI linked to temperature and precipitation in the upper catchments of Yellow River. *Environmental Modeling & Assessment, 17*(4), 389-398.

Hernández-Stefanoni, J. L., Gallardo-Cruz, J. A., Meave, J. A., & Dupuy, J. M. (2011). Combining geostatistical models and remotely sensed data to improve tropical tree richness mapping. *Ecological indicators, 11*(5), 1046-1056.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment, 83*(1-2), 195-213.

Keiffer, C. H. (1999). *Implications of Reforestation: Controlling Deer Browsing And Competing Vegetation.* Miami University,

Keleş, S., & Başkent, E. (2007). Modelling and Analyzing Timber Production and Carbon Sequestration Values of Forest Ecosystems: A Case Study. *Polish Journal of Environmental Studies, 16*(3).

Konijnendijk, C. C., Ricard, R. M., Kenney, A., & Randrup, T. B. (2006). Defining urban forestry–A comparative perspective of North America and Europe. *Urban forestry & urban greening, 4*(3-4), 93-103.

Kruse, S. E., Kyllönen, T., Ojanperä, S., Riddell, R., & Vielajus, J. (1997). Searching for Impact and Methods: NGO Evaluation Synthesis Study, Report Prepared for the OECD/DAC Expert Group on Evaluation. *Helsinki: University of Helsinki (Institute of Development Studies); online: www. valt. helsinki. fi/ids/ngo*.

Landell-Mills, N., & Porras, I. T. (2002). Silver bullet or fools' gold?: a global review of markets for forest environmental services and their impact on the poor.

Le, H. D., Smith, C., & Herbohn, J. (2014). What drives the success of reforestation projects in tropical developing countries? The case of the Philippines. *Global Environmental Change, 24*, 334-348.

Le, H. D., Smith, C., Herbohn, J., & Harrison, S. (2011). More than just trees: Assessing reforestation success in tropical developing countries. *Journal of Rural Studies*, 1-15.

Le, H. D., Smith, C., Herbohn, J., & Harrison, S. (2012). More than just trees: assessing reforestation success in tropical developing countries. *Journal of Rural Studies, 28*(1), 5-19.

Ledig, F. T., & Kitzmiller, J. H. (1992). Genetic strategies for reforestation in the face of global climate change. *Forest Ecology and Management, 50*(1-2), 153-169.

Levin, N., Shmida, A., Levanoni, O., Tamari, H., & Kark, S. (2007). Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Diversity and Distributions, 13*(6), 692-703.

Maginnis, S., Rietbergen-McCracken, J., & Sarre, A. (2012). *The forest landscape restoration handbook*: Routledge.

Mansourian, S., & Vallauri, D. (2005). *Forest restoration in landscapes: beyond planting trees*: Springer Science & Business Media.

Marden, M. (2012). Effectiveness of reforestation in erosion mitigation and implications for future sediment yields, East Coast catchments, New Zealand: a review. *New Zealand Geographer, 68*(1), 24-35.

Mcfarland, T., Van Riper III, C., & Johnson, G. (2012). Evaluation of NDVI to assess avian abundance and richness along the upper San Pedro River. *Journal of Arid Environments, 77*, 45-53.

Mohammadi, J., & Shataee, S. (2010). Possibility investigation of tree diversity mapping using Landsat ETM+ data in the Hyrcanian forests of Iran. *Remote Sensing of Environment, 114*(7), 1504-1512.

Morikawa, R. (2014). Remote sensing tools for evaluating poverty alleviation projects: A case study in Tanzania. *Procedia Engineering, 78*, 178-187.
Muhammed, A. W. (2017). *Interpolation and Extrapolation*. Retrieved from https://www.researchgate.net/publication/313359516_Interpolation_and_Extrapolation:
Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing, 33*(2), 481-486.
Nawir, A. A., & Rumboko, L. (2007). *Forest rehabilitation in Indonesia: where to after more than three decades?* : Center for International Forestry Research (CIFOR).
Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban forestry & urban greening, 4*(3-4), 115-123.
Oindo, B. O. (2008). Understanding the relationship between environmental energy availability and bird species richness in Kenya using remote sensing and ancillary data. *The Open Remote Sensing Journal, 1*, 1-6.
Paoletti, E., Bardelli, T., Giovannini, G., & Pecchioli, L. (2011). Air quality impact of an urban park over time. *Procedia Environmental Sciences, 4*, 10-16.
Parviainen, M., Luoto, M., & Heikkinen, R. K. (2009). The role of local and landscape level measures of greenness in modelling boreal plant species richness. *Ecological Modelling, 220*(20), 2690-2701.
Pau, S., Gillespie, T. W., & Wolkovich, E. M. (2012). Dissecting NDVI–species richness relationships in Hawaiian dry forests. *Journal of Biogeography, 39*(9), 1678-1686.
Perrin, B. (2012). Linking monitoring and evaluation to impact evaluation. *Impact Evaluation Notes (2)*.
Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in ecology & evolution, 20*(9), 503-510.
Poveda, G., Jaramillo, A., Gil, M. M., Quiceno, N., & Mantilla, R. I. (2001). Seasonally in ENSO-related precipitation, river discharges, soil moisture, and vegetation index in Colombia. *Water resources research, 37*(8), 2169-2178.
Psomas, A., Kneubühler, M., Huber, S., Itten, K., & Zimmermann, N. (2011). Hyperspectral remote sensing for estimating aboveground biomass and for exploring species richness patterns of grassland habitats. *International Journal of Remote Sensing, 32*(24), 9007-9031.
Qiao, F., & Rozelle, S. (1998). Tenure of forest land and the development of forestry sector. *Problems of Agricultural Economy, 7*(5), 23-29.
Rawat, J., & Kumar, M. (2015). Monitoring land use/cover change using remote sensing and GIS techniques: A case study of Hawalbagh block, district Almora, Uttarakhand, India. *The Egyptian Journal of Remote Sensing and Space Science, 18*(1), 77-84.
Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. (1994). Measuring phenological variability from satellite imagery. *Journal of vegetation science, 5*(5), 703-714.
Running, S. W. (1990). Estimating terrestrial primary productivity by combining remote sensing and ecosystem simulation. In *Remote sensing of biosphere functioning* (pp. 65-86): Springer.
Sabins, F. F. (1999). Remote sensing for mineral exploration. *Ore Geology Reviews, 14*(3-4), 157-183.
Sauer, T. J., James, D. E., Cambardella, C. A., & Hernandez-Ramirez, G. (2012). Soil properties following reforestation or afforestation of marginal cropland. *Plant and soil, 360*(1-2), 375-390.

Schielzeth, H., & Forstmeier, W. (2008). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416-420.

Sedda, L., Tatem, A. J., Morley, D. W., Atkinson, P. M., Wardrop, N. A., Pezzulo, C., . . . Rogers, D. J. (2015). Poverty, health and satellite-derived vegetation indices: their inter-spatial relationship in West Africa. *International health, 7*(2), 99-106.

Sellers, P., Berry, J., Collatz, G., Field, C., & Hall, F. (1992). Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sensing of Environment, 42*(3), 187-216.

Shea, K. L., & Stange, E. E. (1998). Effects of deer browsing, fabric mats, and tree shelters on Quercus rubra seedlings. *Restoration Ecology, 6*(1), 29-34.

Souza, A. A., Galvão, L. S., & Santos, J. R. (2010). Relationships between Hyperion-derived vegetation indices, biophysical parameters, and elevation data in a Brazilian savannah environment. *Remote Sensing Letters, 1*(1), 55-64. doi:10.1080/01431160903329364

Ten Hoorn, E.M., Stubbe W.M. (2013). Resultaat- en impactmeting voor goede doelen. Retrieved from https://www.cbf.nl/Uploaded_files/CBFinteractief/files/assets/common/downloads/publication.pdf

Steiner, A. (2007). Sudan: Post-Conflict Environmental Assessment. In: UN Environment Program.

Verbeek, M. (2008). *A guide to modern econometrics*: John Wiley & Sons.

Vintrou, E., Desbrosse, A., Bégué, A., Traoré, S., Baron, C., & Seen, D. L. (2012). Crop area mapping in West Africa using landscape stratification of MODIS time series and comparison with existing global land products. *International Journal of Applied Earth Observation and Geoinformation, 14*(1), 83-93.

Virtanen, R., Luoto, M., Rämä, T., Mikkola, K., Hjort, J., Grytnes, J. A., & Birks, H. J. B. (2010). Recent vegetation changes at the high-latitude tree line ecotone are controlled by geomorphological disturbance, productivity and diversity. *Global Ecology and Biogeography, 19*(6), 810-821.

Weyerhaeuser, H., Wilkes, A., & Kahrl, F. (2005). Local impacts and responses to regional forest conservation and rehabilitation programs in China's northwest Yunnan province. *Agricultural Systems, 85*(3), 234-253.

Winter, B. (2013). A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499*.

Yu, J. (2017). *Successes and Failures of China's Grain-For-Green Program.* Tufts University,

Zhan, Z.-Z., Liu, H.-B., Li, H.-M., Wu, W., & Zhong, B. (2012). The Relationship between NDVI and Terrain Factors--A Case Study of Chongqing. *Procedia Environmental Sciences, 12*, 765-771.

Zhong, R. I. (2018). Transparency and firm innovation. *Journal of Accounting and Economics*.

Zuur, A. F. (2009). *Mixed effects models and extensions in ecology with R.*: New York: Springer.

Appendix

Table I Overview of all bands in the Sentinel-2 database

Sentinel-2 Bands	*Central Wavelength (µm)*	*Resolution (m)*	*Bandwidth (nm)*
Band 1 – Coastal aerosol	0.443	60	20
Band 2 – Blue	0.490	10	65
Band 3 – Green	0.560	10	35
Band 4 – Red	0.665	10	30
Band 5 – Vegetation Red Edge	0.705	20	15
Band 6 – Vegetation Red Edge	0.740	20	15
Band 7 – Vegetation Red Edge	0.783	20	20
Band 8 – NIR	0.842	10	115
Band 8A – Narrow NIR	0.865	20	20
Band 9 – Water vapour	0.945	60	20
Band 10 – SWIR – Cirrus	1.375	60	20
Band 11 – SWIR	1.610	20	90
Band 12 – SWIR	2.190	20	180

Table II Overview of all variables extracted from datasets before pre-processing

Variable	**Description**	***Min***	***1st Qu.***	***Median***	***Mean***	***3rd Qu.***	***Max***	***Corr. with Change in NDVI***
Dependent variables								
Change in *NDVI*	Difference in NDVI between September '17 and September '16	-0.437	-0.099	-0.033	-0.018	0.04104	0.465	1***
Change in Ba_{Ha}	Difference in Ba_{Ha} between '17 and '16	-2.342	0.2122	0.753	1.197	1.875	6.896	0.086 (52 obs.)
Change in TP_{Ha}	Difference in TPH between '17 and '16	-875.43	-153.43	5.89	-65.71	71.62	526.19	-0.141 (52 obs.)
Baselines								
$NDVI_{baseline}$	The baseline greenness of parcel in September 2016.	0.158	0.533	0.655	0.623	0.740	0.801	-0.683***
TP_{Ha2016}	The baseline trees per hectare of parcel in 2016.	127.4	284.9	395.0	684.1	1220.5	1859.9	0.064
Ba_{Ha2016}	The baseline basal area per hectare of a parcel in 2016.	0.001	0.165	0.898	1.280	2.204	4.792	-0.196
Parcel specifications								
Age *(years)*	Time in years since the parcel is registered to the program.	1.668	2.668	4.668	4.260	5.671	7.671	-0.058

Performance indicators								
Relative Performance Score	A score that captures the relative performance of a parcel. It is the performance compared to a benchmark (see Section 4.4).	0.2569	0.8719	1.0205	0.9895	1.1065	1.6311	-0.735***
Relative Performance Group *(categorical)*	Parcels are divided into 4 groups, which are the quantiles of the Relative Performance Score.	*na*	*na*	*na*	*na*	*na*	*na*	*one-way ANOVA* F*(3,764) = 223.5****
Covariates								
Elevation		154	310	366.5	479.8	657	1311	0.166***
Mean Annual Precipitation		967	1187	1372	1321	1460	1596	-0.222***
Training type								
Weeding	Technician is assisting in weeding with a machete around the trees.	0	0	0	0.468	1	3	0.092**
Pruning	Pruning the trees so that they grow straight and make quality trees.	0	0	0	0.287	0	9	-0.048
Planting	Planting trees session.	0	0	0	0.404	0	6	0.053
Diagnose	Checking in on a farm to see how it is doing.	0	0	1	1.147	1	17	0.130***
Materials delivery	Handing over materials needed for the managing the plantations. Includes materials for tree nurseries, pruning scissors, etc.	0	0	0	0.704	1	5	-0.139***
Introduction	Presentation about to program.	0	0	0	0.125	0	4	0.092**
Certificate	Working to register the plantation with some certification like with the government or some standard.	0	0	0	0.013	0	5	-0.027
Tree nursery	Working with the tree nursery.	0	0	0	0.350	0	7	0.038
Other work	Default option for when other things aren't available.	0	0	1	1.788	3	11	-0.025
Frequency of training	Disaggregated count of all trainings received in timeframe.	1	2	5	5.971	7	29	0.040
Technician visit								
Technician *(dichotomous)*	Dummy for each technician if he/she visited the farmer in the timeframe (0 = no visit, 1 = visit)	*na*	*na*	*na*	*na*	*na*	*na*	*na*
Agroforestry type								

Agroforestry category *(categorical)*	Type of agroforestry the parcel has. The program includes Living Fence (old), Mixed Species, Shade Coffee and Silvopastoral.	na	na	na	na	na	na	*one-way ANOVA F(5,762) = 3.279****
Significance codes: *** *p*-value < 0.01 (2-tailed), ** *p*-value < 0.05 (2-tailed), * *p*-value < 0.10 (2-tailed)								

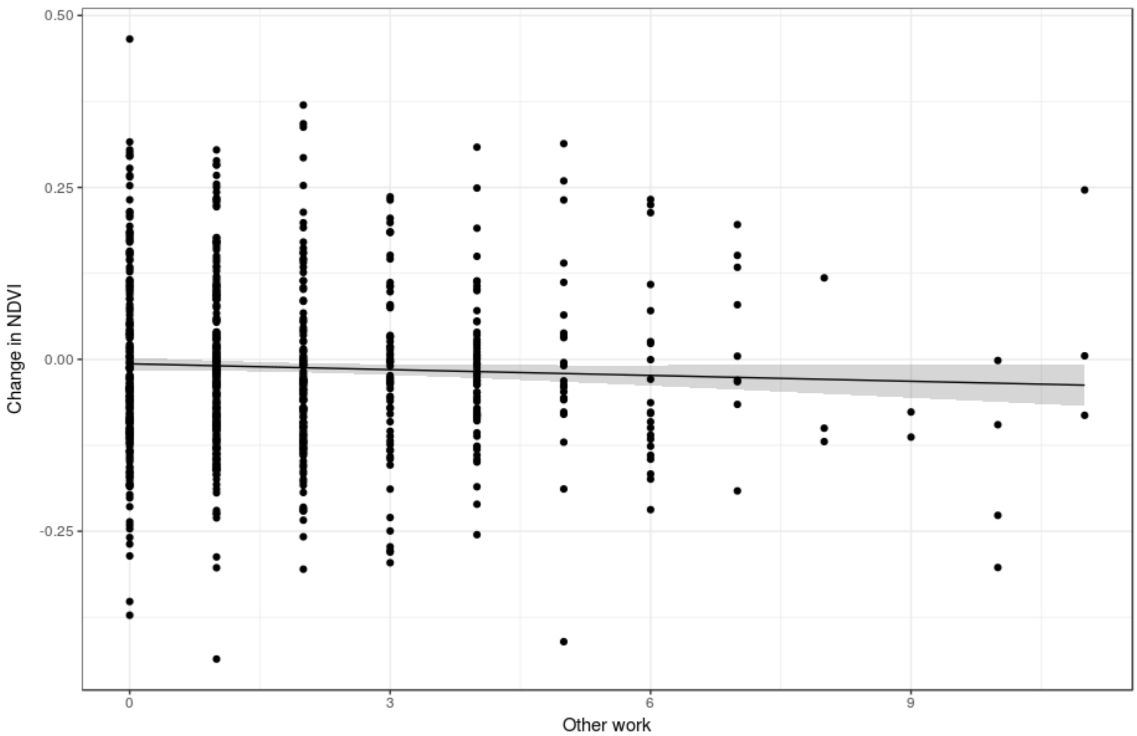


Figure i Visualisation of the relationship between *Change in NDVI* and *Other work*

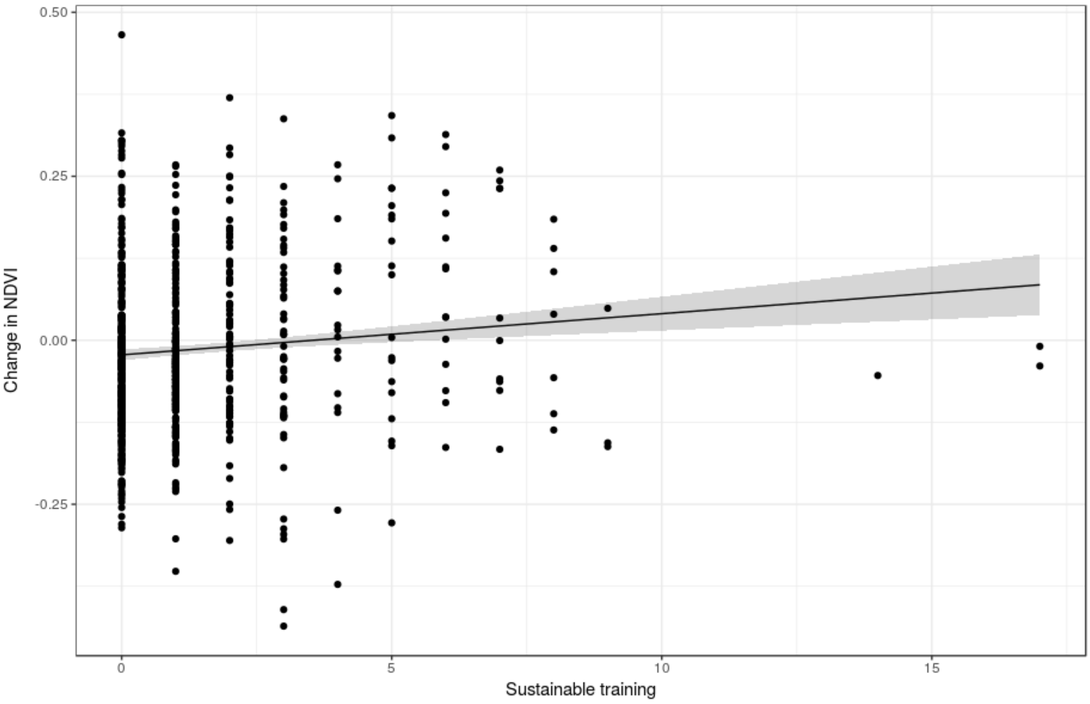


Figure ii Visualisation of the relationship between
Change in NDVI and *Sustainability Training*

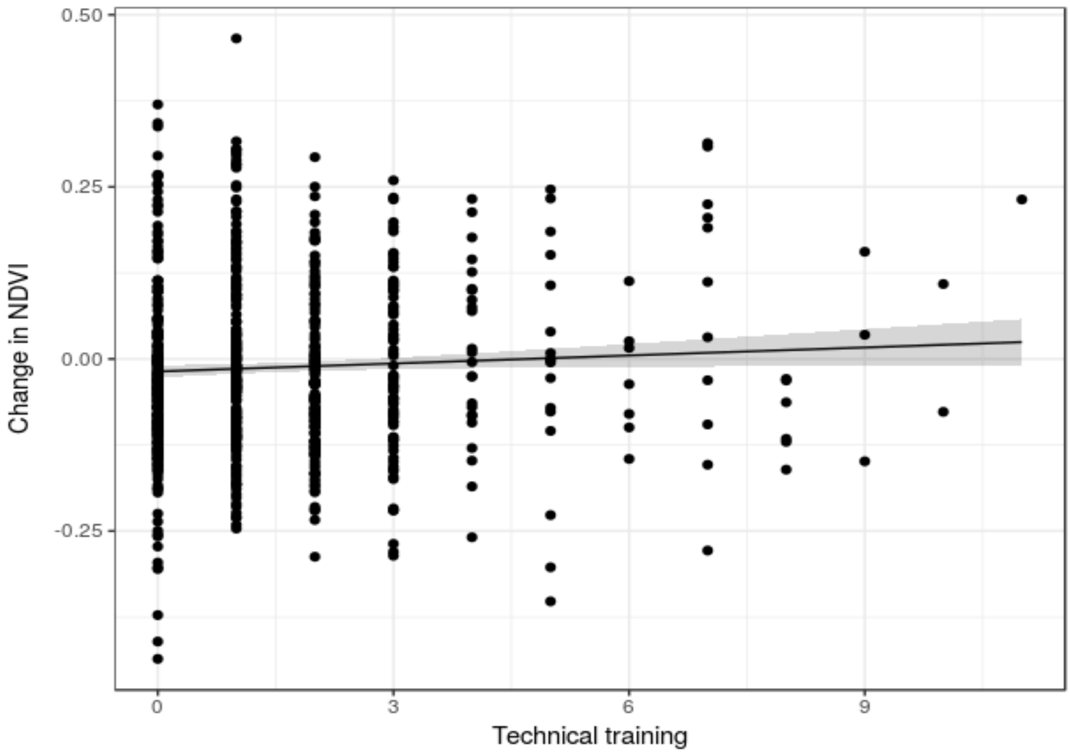


Figure iii Visualisation of the relationship between
Change in NDVI and *Technical Training*

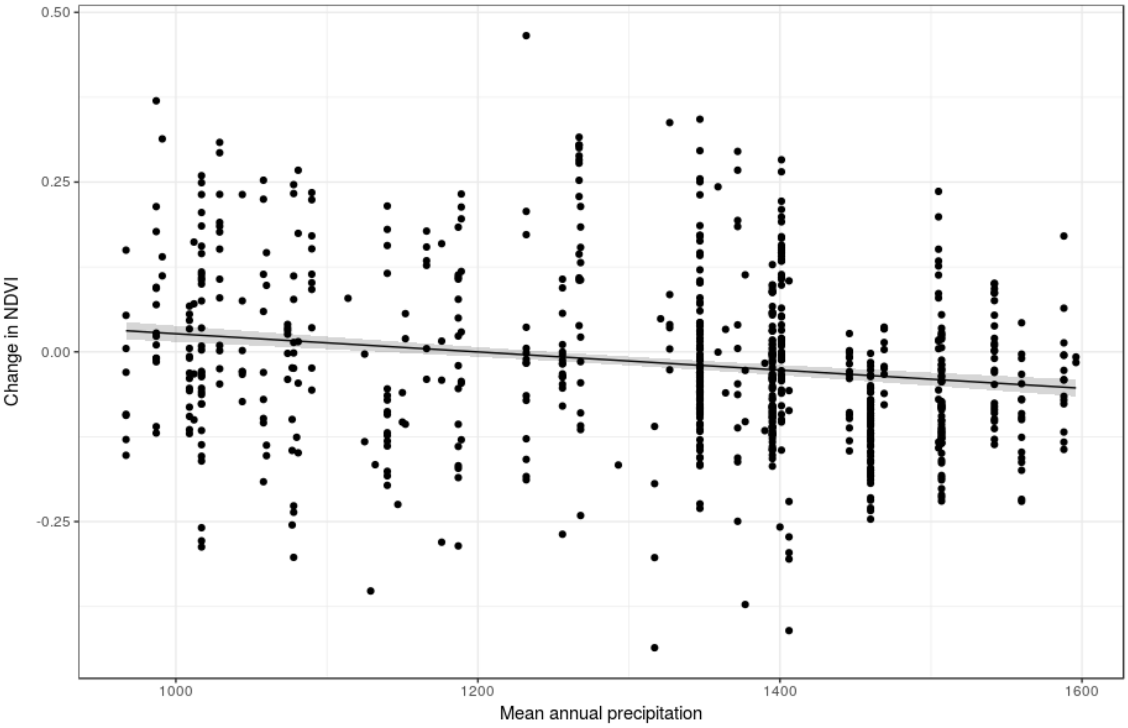


Figure iv Visualisation of the relationship between
Change in NDVI and *Mean Annual Precipitation*

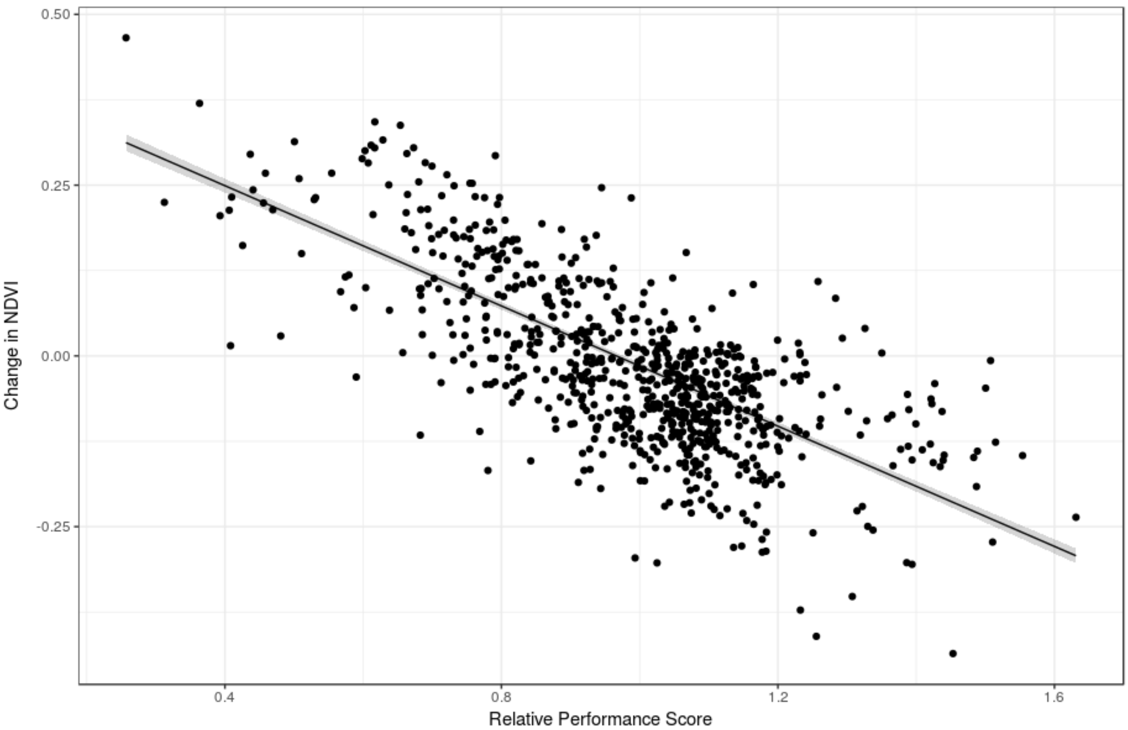


Figure v Visualisation of the relationship between
Change in NDVI and *Relative Performance Score*

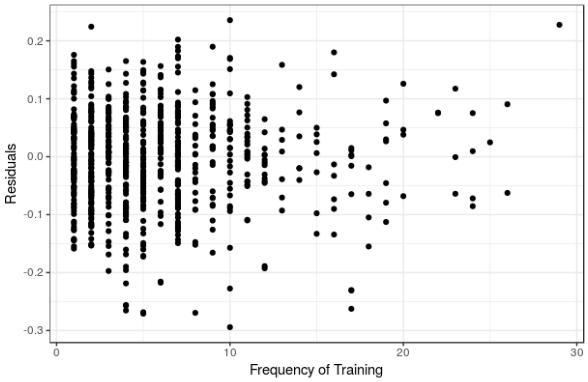
Assumptions of the regression model

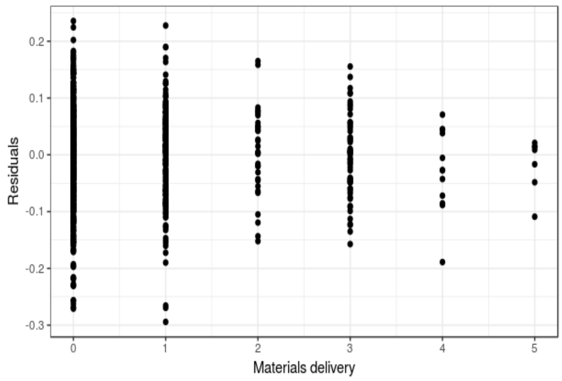
Validity

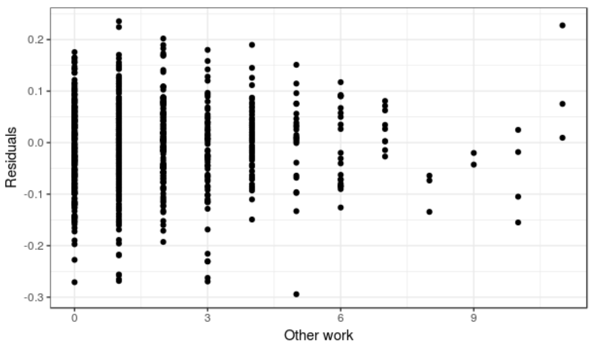
The validity of the panel data used in this study is discussed in Section 3. This assumption implies that the data used in the model should map to the research question.

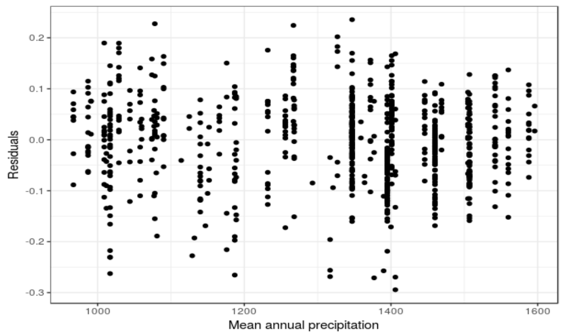
Additivity and linearity

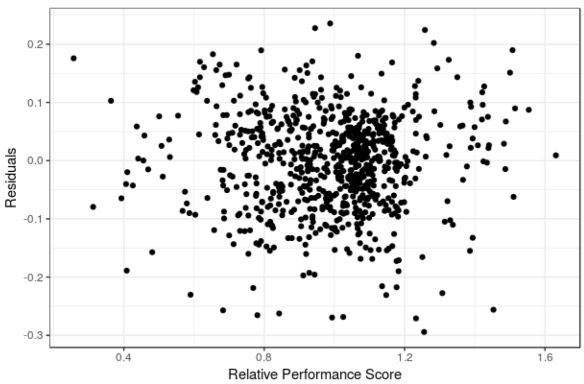
This assumption is checked by plotting the residuals to the independent variables. If additivity is violated (a non-linear pattern in the plot), it might make sense to transform the data.

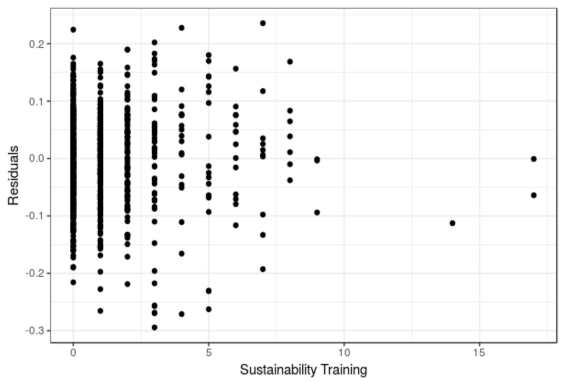












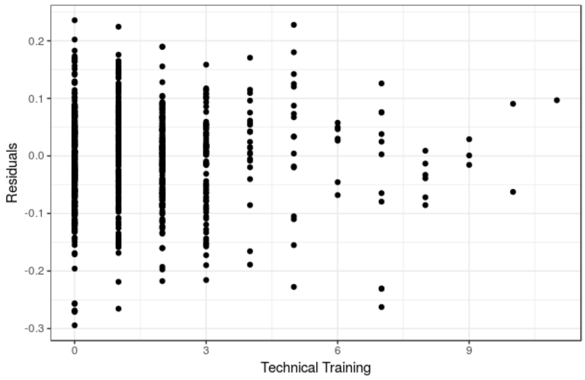


Figure vi Checking for non-linear relationships between the independent variables and the residuals

Independence of errors

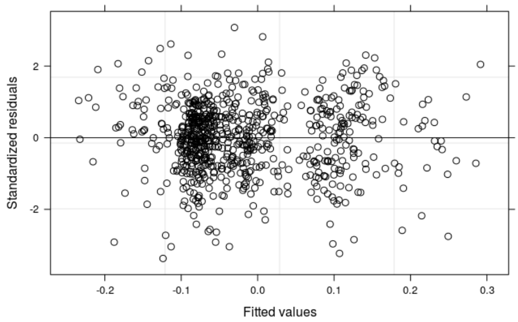
The regression model assumes that the errors of the prediction line are independent. Since Generalized Least Squares models are built in this study no high autocorrelation problems are assumed. Also for the linear random effects model, no violation is found.

Table III *Independence of errors*

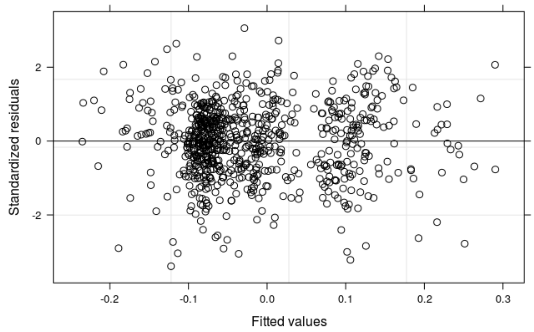
	Dependent variable:						
	Change in NDVI						
	generalized least squares					*linear random effects*	
	(1)	(2)	(3)	(4)	(5)	(6)	
						independent	*random*
Durbin-Watson statistic	1.872	1.876	1.875	1.932	1.898	1.862	1.972

Equal variance of errors

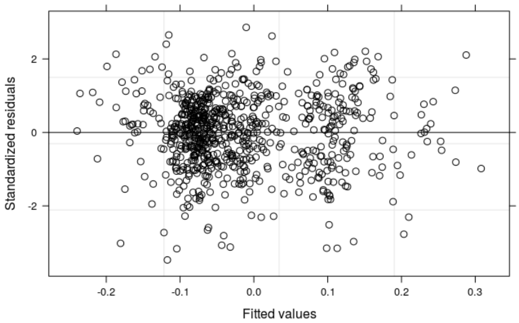
Variance over all errors should be equally distributed and show no pattern.



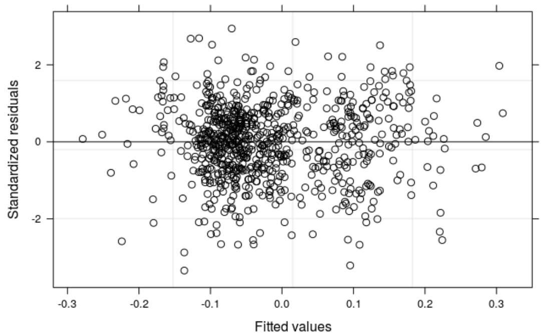
(1)



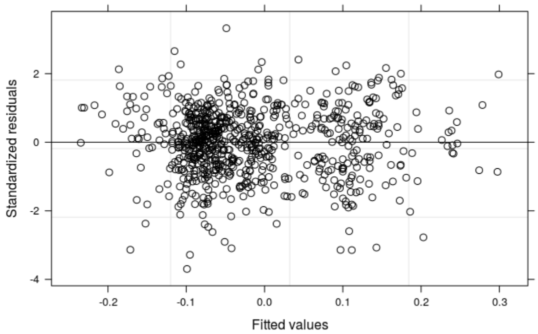
(2)



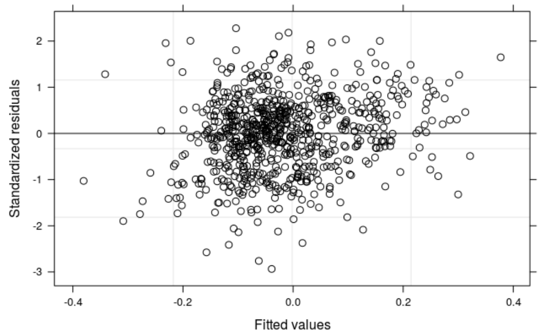
(3)



(4)



(5)

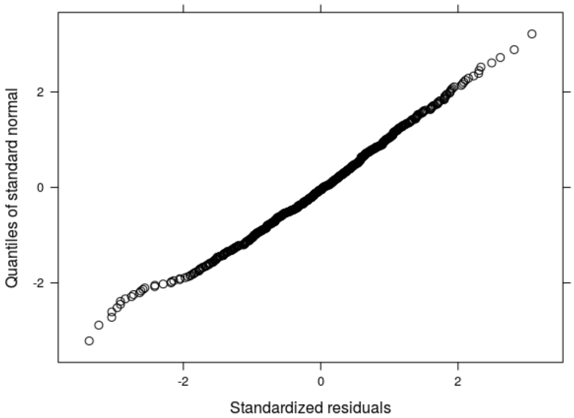


(6)

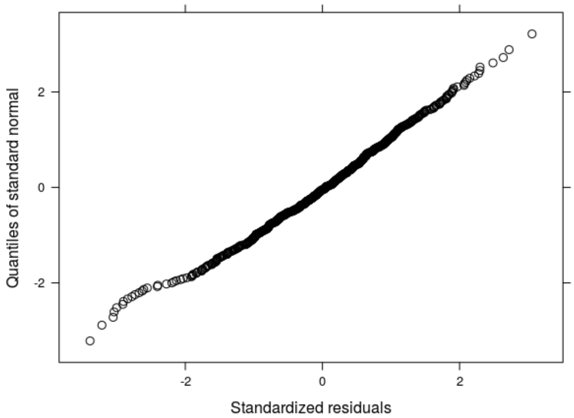
Figure vii Inspection of the error terms of all models from Table 3

Normality of errors

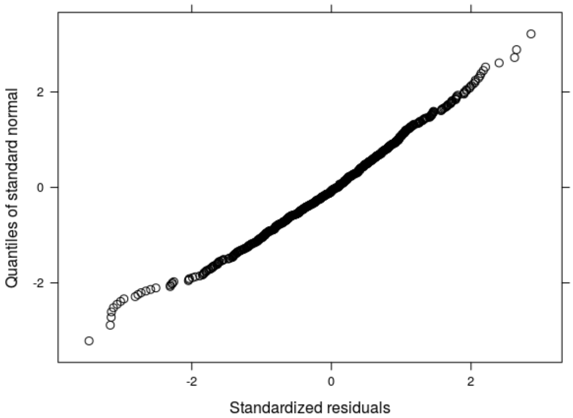
In a regression model errors are assumed to be normally distributed. This assumption is checked visually.



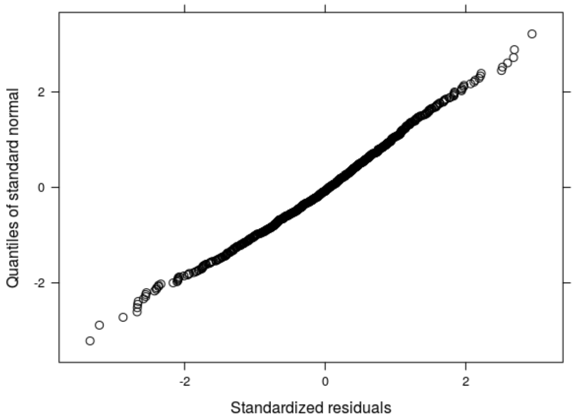
(1)



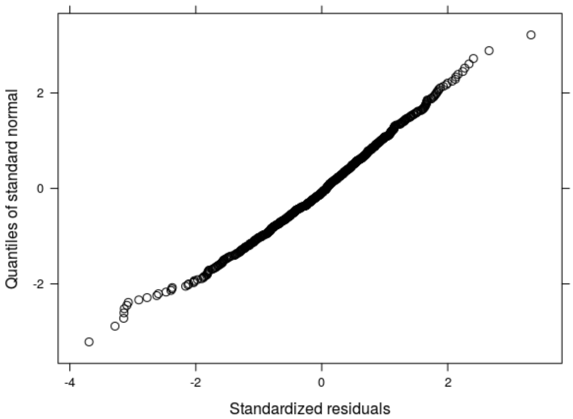
(2)



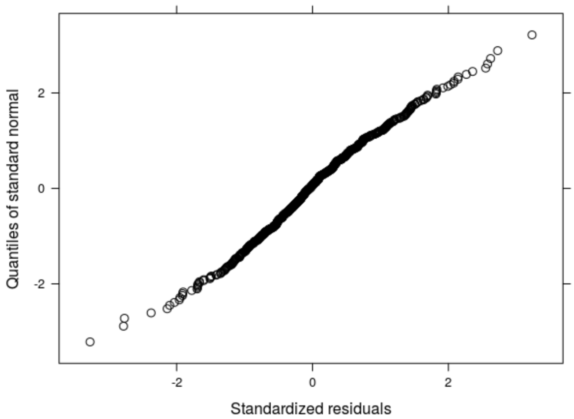
(3)



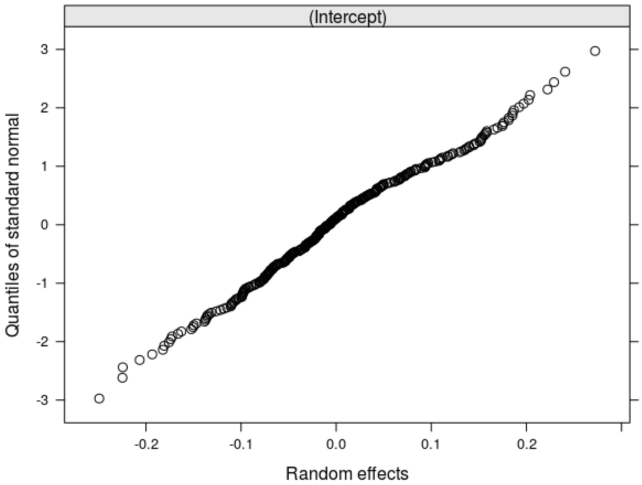
(4)



(5)



(6)



(6 – Random Effect)

Figure viii Distributions of the error terms of all models in Table 3