Where Trees Grow, Expenditures Grow: Applying Spatial Matching to Evaluate Agroforestry's Household Welfare Impacts in Kenya

Seth Morgan, Department of Agricultural & Consumer Economics, The University of Illinois at Urbana-Champaign, smorgan9@illinois.edu

Katherine Baylis, Department of Agricultural & Consumer Economics, The University of Illinois at Urbana-Champaign, baylis@illinois.edu

Karl Hughes, The World Agroforestry Centre, k.hughes@cgiar.org

Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

Abstract

Agroforestry–perennials planted in association with annual field crops–has potential as a method for sequestering carbon while reportedly increasing agricultural yields and farmer income. However, measuring the downstream effect of agroforestry promotion on household welfare is difficult due to the long-term nature of the impact pathways and their dependence on local agro-ecological conditions. This paper demonstrates a method for using spatial matching methodology to select a household sample prior to data collection in a long-term impact evaluation of an agroforestry project in western Kenya. We find spatial matching to be an effective way to assemble an ex-post counterfactual for a quasi-experimental impact evaluation and present balance statistics, pre-treatment trends, and alternate specifications to validate the methodology. The agroforestry program is found to result in modest but significant welfare gains in terms of household assets and predicted expenditure.

1 Introduction

Agroforestry–defined as the "use of woody perennials on the same land management unit as agricultural crops, pastures, and animals" (Current, Lutz, and Scherr, 1995)–has been promoted as a land management strategy with potential for carbon sequestration, climate change adaptation and increased productivity and profitability, a so-called "triple-win" for smallholder agriculture (Bryan et al., 2011). Given smallholders' relative poverty, and pressure on existing forests due to increasing demand for income and agricultural land, agroforestry's promise to enhance environmental services and increase incomes makes it an attractive technology for promotion by organizations providing small farmer advisory services. But while agroforestry's potential as a means of increasing the carbon intensity of smallholder farming systems is relatively well established (Lorenz and Lal, 2014), the second two "wins"–climate adaptation and productivity gains–remain less well-evidenced. Even less well-understood is the magnitude of the household welfare impacts–if any–produced by these adaptation and productivity effects.

Estimating the effects of an intervention like agroforestry is a significant challenge. Like many interventions in the areas of climate change mitigation and sustainable agriculture, agroforestry is complex, affecting many elements of the farming system and farmers' livelihood strategies. Agroforestry promotion is also likely to require several interrelated interventions including germplasm provision and farmer advisory services. In addition, agroforestry's outcomes are likely to manifest themselves over a long time frame and to depend heavily on local agronomic conditions, increasing the variation in effect sizes likely to be observed.

This complexity of both the intervention and its causal pathways poses a particular challenge to measuring agroforestry's impact using the current gold-standard in impact evaluation design, the randomized control trial. To use experimental methods to study agroforestry, a program would need to be randomized, then tracked over a long period. The time lag, along with the already high variability of the phenomena, would magnify the possibility for variation in outcomes, decreasing the statistical power of the study and requiring a very high sample size. For this reason, considerations of time and expense might call for quasi-experimental methods. Additionally, given that a large number of agroforestry projects are already under way, demand for evidence in the near to medium-term may require the use of quasi-experimental methodology before a long-term RCT can be feasibly administered.

This paper presents a case study of a methodology which has potential for application in evaluations of complex, long-term and location-dependent interventions like agroforestry. We describe a spatial matching methodology which holds promise as a way to assemble a sample for the evaluation of complex interventions. We then demonstrate this methodology's use in an evaluation of an agroforestry-focused extension program in western Kenya and present a series of tests of its effectiveness.

Matching research designs are commonly used in the social sciences to help identify a comparison group by selecting comparison observations which are matched to treatment observations across observable variables (Abadie and Guido W Imbens, 2016). However, this approach requires a large pool of candidate observations from which to draw the matched sample. In the absence of suitable secondary data or the resources to survey a very large number of households, it may be difficult to assemble a dataset of comparison observations which are observably similar across the selected variables.

Along with using matching to identify similar controls from secondary data, matching can also be used prior to data collection to target control observations that are similar to those under treatment (Stuart, 2010). Most pre-data collection matching in the literature refers to a case when a round of household data has already been collected, and follow-up is being planned. The availability of geospatial datasets makes a new use-case possible: villages or administrative units can be matched prior to data collection, resulting in a geospatially matched sample whose variance has been reduced prior to survey administration. This approach is particularly attractive when the program being evaluated is assigned to particular geographic units, and where agro-ecological characteristics like altitude and rainfall are hypothesized to affect the outcomes under consideration.

The forest conservation literature offers a model for matching across geographic units using geospatial variables (Andam, Ferraro, Pfaff, et al., 2008; Andam, Ferraro, Sims, et al., 2010; Honey-Rosés, Baylis, and Ramírez, 2011) which has generated measures of the effect of protected areas and payments for ecosystem services on tree cover and on socio-economic outcomes. This paper adds a methodological contribution to this literature: we field-test a method for using spatial matching to match geospatial units (in this case villages) to be used for further detailed primary data collection to evaluate an agroforestry program. Our approach resembles that used by Alix-Garcia, Sims, and Yañez-Pagans (2015) to evaluate the efects of payments for ecosystem services in Mexico, though Alix-Garcia et al utilize household-level geospatial data for matching, while here we demonstrate a method for using village-level data to construct a sampling frame prior to collecting any data at the household level. To our knowledge this is the first application of this specific set of techniques in the environmental evaluation literature.

The intervention considered here is an agroforestry-focused extension program located in western Kenya implemented by the Swedish NGO Vi Agroforestry. The program is located in two counties in western Kenya, but due to funding restrictions it was only offered in a smaller geographic area, determined by sub-county geographic boundaries which have little administrative authority and which have since been superseded by post-2010 administrative redistricting. This geographic targeting allows for a quasi-experimental research design where villages outside Vi's program area can be compared with those within it.

In this paper we demonstrate the application of geospatial matching to a set of villages inside and outside the Vi program area, and show that (a) in this case village level matching results in a sample of households that is balanced across baseline wealth indicators, and (b) the matching method holds potential for reducing bias in the outcome estimation as demonstrated by the estimation of a subset of wealth outcome measures drawn from the wider impact evaluation study found in Hughes et al. (2017). We present intention to treat effects, including results from a doubly-robust estimator to account for outliers, as well as local average treatment effects derived from a two-stage least squares model using the program area as an instrument for program participation. Additionally we explore a number of robustness tests applied to these outcomes and compare the results using alternative models to account for the spatial distribution of treatment.

2 Background

The term agroforestry gained currency in the 1970s with the publication of John Bene's report "Trees, Food and People" for the Canadian International Development Research Centre, and was given further legitimacy by the founding of the International Council for Research in Agroforestry, also known as the World Agroforestry Centre (Nair, 1993). The techniques that fall under the category agroforestry include such disparate systems as improved fallows (seeding of perennial legumes into fallow fields), alley cropping (rows of perennials alternating with field crops), multi-layer tree gardens, and living hedges (ibid.).

Agroforestry systems have been associated with reductions in soil erosion (Otsuki, 2010), increases in soil organic carbon and soil microbe populations (Araujo et al., 2012), carbon sequestration (Lorenz and Lal, 2014) and other ecosystem services, including habitat, water quality and soil fertility (Tsonkova et al., 2014). Additionally, agroforestry is widely expected to result in household welfare impacts due to its purported potential to increase yields of important food crops, augment farm incomes for smallholders, and increase resilience to climate shocks.

Evidence for these outcomes exists, though rigorous identification of agroforestry as the causal mechanism is limited. A number of studies have in fact found agroforestry adoption to be positively associated with yields and productivity (Otsuki, 2010; Sjögren, Shepherd, and Karlsson, 2010), though the Sjogren results are conducted on experimental plots, and Otsuki's results are observational, with adjustment for endogenous selection using a treatment effects model. Current finds more diverse sources of income and improved cash flow in Latin America (Current, Lutz, and Scherr, 1995), in an observational study with a small sample size.

Impacts on total household income or consumption expenditure are even less clear. Haglund et al. (2011) find income increases of 18-24% in Niger associated with the adoption of Farmer Managed Natural Regeneration, a method for encouraging the regrowth of native tree species interspersed within cropland. Hegde and Bull (2011) find positive expenditure effects in Mozambique stemming from an agroforestry-related payment-for-ecosystem services program, though the contribution of agroforestry itself versus the payments themselves is unclear. Meanwhile Place et al. (2003) find no significant effects on household poverty or food security indicators in Western Kenya despite increases in crop yields–though Place makes no adjustment for selection effects, and with a sample size of 100 may be underpowered to detect an effect.

Many studies on the impacts of agroforestry suffer from the non-random nature of technology adoption. Hegde and Bull use a matching design to match adopters to non-adopters who are observably similar across the variables which are found to be correlated with adoption (2011). However, it must be assumed that unobserved variables do not account for the measured effect, and a very large sample may be required in order to draw a comparison group which is balanced across all relevant observable characteristics. It is possible to argue that Hegde and Bull's sample size of 290 could limit the number of variables which can be used for matching, and even if the matched sample is balanced, there may be unobservables correlated with the treatment.

The literature on the impacts of protected areas in forest conservation offers examples of matching research designs implemented at a landscape scale. In this case geographic units are matched across a vector of geospatial variables, providing a sample of similar pixels, polygons or counties (Andam, Ferraro, Sims, et al., 2010; Honey-Rosés, Baylis, and Ramírez, 2011; Robalino, Pfaff, and Villalobos, 2015). This methodology has been used to measure the impacts of protected areas on tree cover, as well as on poverty reduction and household welfare (Andam, Ferraro, Sims, et al., 2010). As with any matching, these evaluations may suffer from bias if the location and success of the protected areas is determined in part by unobservables.

This paper contributes to the environmental impact evaluation literature and to the literature on agroforestry more specifically by field-testing matching as a method for assembling a data frame to be used in an ex-post impact evaluation. The spatial matching method presented here holds promise as a way to feasibly conduct ex-post impact evaluation of long-term interventions using novel survey data when longitudinal data on treatment and control observations are not available. The demonstrated methodology may hold particular interest for fields such as agroforestry, sustainable agriculture, and forestry studies where long-term interventions are the norm and many studies in the current literature rely on secondary data in order to estimate effects.

3 Intervention

Vi Agroforestry (herein Vi) is a Swedish NGO founded in the 1980s with the goal of reforesting western Kenya. By the 1990s this objective was adapted to the realities of working in coordination with small-holder farmers, and Vi transformed itself into a provider of tree seeds and agricultural extension services. Vi's current program model works by promoting tree planting and providing other agricultural training through the mediation of farmer-organized groups.

Although Vi offers a range of different training topics through its extension programming, the primary focus of their program remains agroforestry, which in this context implies planting trees either around plot boundaries or in rows within fields. Vi received extensive training and support from the International Centre for Research in Agroforestry (ICRAF) throughout the 1990s and early 2000s, including a set of seminars for extensionists administered by ICRAF personnel, which incoming Vi staff and managers received from the mid 1990s up until 2005. Training from ICRAF during this period introduced a cropping system known as alley-cropping, which involves rows of leguminous shrubs planted periodically throughout annual crop fields. ICRAF also introduced Vi staff to the use of perennial legumes including *Calliandra callothyrsus* and *Sesbania sesbans* as fodder and improved fallow species (LePage Morgan, 2017). For this reason, Vi's program was identified by ICRAF as a promising site for an evaluation of agroforestry's potential as a household welfare-enhancing intervention. The data examined in this paper are the result of the ICRAF impact evaluation entitled "Assessing the Downstream Socioeconomic and Land Health Impacts of Agroforestry in Kenya," a broader study aimed at examining agroforestry's long-run impacts on household wealth, food security and land-health.

The agroforestry techniques primarily promoted by Vi can be described as variants of three separate planting patterns: alley-cropping, boundary planting, and tree planting along erosion control barriers. Alley-cropping is a technique developed in the 1970s by the International Institute of Tropical Agriculture (IITA). It consists of rows of leguminous perrennials planted within farmers' annual fields, with the intention of providing additional fertility through nitrogen fixation and the incorporation of leaf biomass (Douthwaite et al., 2002). Boundary planting is a common practice in the region which Vi has encouraged farmers to intensify and add more short-term leguminous shrubs to the spaces between long-term timber species which are typically planted in boundaries, creating a multi-story boundary planting system (LePage Morgan, 2017; Wachiye, 2008).

Erosion control structures can include simple grass strips, trash lines consisting of crop residue, small contour bunds, trenches, or terraces. All of these practices are common in the study area, and encouraged by Vi. Additionally, Vi trains farmers to incorporate multi-story perrennial plantings along these structures. This practice of planting trees for erosion control has been described by one promoter as a "slow terrace," and is reported to be less labor intensive than reliance solely on hand-built earthen terraces (LePage Morgan, 2017).

Vi takes a relatively neutral stance in regards to the relative value of the planting arrangements described above, but their self-described hallmark is the incorporation of leguminous perennials, especially *Sesbania sesban* and *Calliandra callothursus*, into the farming system. Several Vi staff reported that "if you see *Sesbania* you can guarantee that Vi has been there" (E. Wachiye, personal communication, March 2016). These trees are are quick-growing and nitrogen-fixing, and are used for firewood, fodder and "green manure"–incorporation into the soil as fertilizer.

In practice, Vi's program participants tend to pick and choose from among the available species and planting arrangements included in the Vi training and adapt them to their needs. The arrangement most commonly seen during scoping interviews at the initiation of this project was essentially an adapted form of alley-cropping, where lines of trees were incorporated into the fields. Some respondents with trees planted in this arrangement reported rotating these lines of trees through their fields as a way to enhance fertility, while others reported keeping them fixed in one place in order to control erosion. So while perennial alleys planted strictly according to the spacing recommendations developed by IITA were rare, there nonetheless seemed to be enthusiasm for an adapted version, with spacing and timing dictated by the farmers relative desire for firewood, erosion control, and nutrient cycling.

The training program Vi implemented in the study area was not standardized. The extensionists were given latitude to assess which topics were needed by the farmer led groups to which they were assigned. However, every group was expected to learn about the advantages of agroforestry, and Vi's activity calendar was coordinated around their twice-yearly seed distribution which corresponds with the arrival of the two rainy seasons in the study area. Vi distributes tree seeds free of charge to all its member groups. They provide seeds for direct-seeding in fields before the long rains, when maize is planted, and seeds intended for raising in small-scale tree nurseries before the short rains, when secondary bean crops and vegetables are planted.

The farmer-led groups through which Vi delivers its trainings and seed provision are common throughout western Kenya. They tend to have 10-15 members, and can be organized as women's groups, youth groups, farmer groups, or religiously focused groups though in practice these distinctions are not binding. Nearly all groups contain a mix of ages and genders, but women tend to predominate and most group members are older and slightly better off than the general population.

Vi's training services have been offered in the study area of Bungoma and Kakamega counties since 2008 through the operations of two different projects: the Kenya Agricultural Carbon Project (KACP) and the Farmer Organizations and Agroforestry (FOA) project. The two projects have separate staff and funding structures but share staff training on agroforestry and land management.

The KACP project, which is active in the Bumula and Sirisia divisions, focuses on increasing carbon sequestration in small-holder farming systems by encouraging tree planting and sustainable land management techniques. Tree planting in particular is incentivized by a small payment (equal to about $3.00 per person on average) disbursed to the farmer groups upon confirmation that trees have been planted and preserved on their farms.

The FOA project focuses more significantly on capacity building for farmer organizations, in addition to tree seed provision and land management training. FOA is active in Kimilili, Webuye East and Bungoma North sub-counties in Bungoma County, as well as Likuyani sub-county in Kakamega county. Vi staff describe the extension trainings provided by FOA as essentially identical to those in KACP, but a few key programmatic differences remain: FOA does not provide carbon payments, and does not perform the same level of monitoring on its tree planting activities. Additionally, since FOA is focused on empowering farmer organizations, the decision was made in 2014 to hand over its activities in Kimilili and Webuye East to partner Savings and Credit Cooperatives (SACCOs). This means that Vi trainings for farmer groups have been carried out by the SACCOs for 2 years out of the study period from 2014 to 2016.

Despite the differences in project design, the training and seed distribution regimen over all Vi's projects is largely the same, and Vi's expectation is that all their participants will become adopters of agroforestry. Project-level differences create opportunities for analysis of differential effects beyond the scope of this paper, but the projects are deemed similar enough to consider them essentially the same treatment.

The creation of these two projects in Bungoma and Kakamega, where Vi had never had operations previously, and no other organization was focusing significantly on agroforestry, makes this region attractive as a study area for evaluating Vi's program impact.

4 Study Area

The area included in this study comprises all of Bungoma County and the northern portion of Kakamega County in Western Province, Kenya. This area was considered to have high potential for an impact evaluation study because Vi Agroforestry had been promoting agroforestry practice in this area since 2008, Vi staff considered adoption to be significant enough for an estimation of impact to be feasible, and there remained a significant bordering area outside the Vi program area which could be used for comparison purposes.

The zones included in the study encompass the base of Mt. Elgon and the sugar cane-growing lower-midlands of Bungoma as well as the maize-growing upper-midlands of Eastern Bungoma county and Kakamega (Jätzold and Kutsch, 1982). Elevation varies from approximately 1200 to 2100 meters above sea level (Jarvis et al., 2008; Kruska and Kariuki, 2016).

Vi concentrated its agroforestry extension work within Sirisia, Malakisi, Bumula, Kimilili, Ndvisi, Tongaren and Likuyani Divisions of Bungoma and Kakamega Counties, as shown in figure 1. Within these divisions they further targeted by location–a smaller administrative unit–when they did not have enough resources to cover an entire division, which primarily occurred in Tongaren division.

Vi attempted to target their program at farmer groups in locations which were dominated by small-scale farmers, and where they perceived that farmers were somewhat less well off and farms were more subject to erosion. However, there is reason to believe that these criteria were not observed systematically, and that the distinction in terms of household wealth and landholding between locations included in the program and those outside was not large, and if anything would bias the impact evaluation results downward.

First, it should be noted that the organization intended to provide services in a broader area in Bungoma county, but could not expand beyond the area depicted in figure 1 until 2015 due to funding restrictions. Therefore it seems likely that the remaining area fits the profile of their target area. Second, on a village-to-village basis there are few discernible differences between those who are included in the program area and those who are not. According to staff, villages near the location borders are likely to be identical, so long as the metropolitan areas of Bungoma and Webuye towns are excluded.

It should also be noted that in 2010 Kenya underwent a rewrite of its constitution and reconfiguration of administrative boundaries which devolved a high degree of authority to the counties

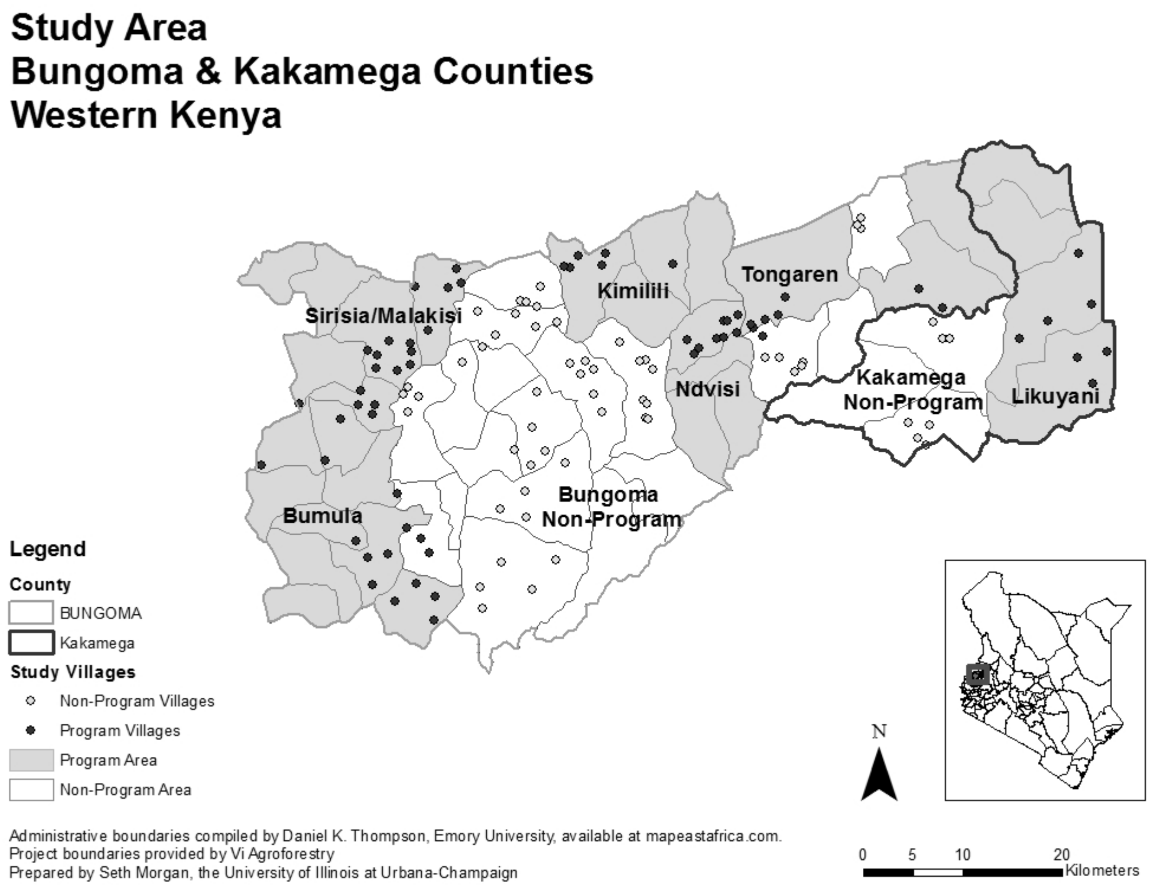
(Cornell and D'Arcy, 2014). As a part of this process the divisions and locations were reorganized by the counties into sub-counties and wards, which in some cases corresponded with the old administrative units and in other cases were the result of splitting or consolidating the older units. This means that Vi's geographic program targeting took place using administrative divisions that were superseded only two years into the program. The result is that it seems unlikely that the administrative boundaries which define Vi's program area would be correlated with unobservable differences in administrative quality or public goods provision which might confound our estimates of the treatment effects.

Vi utilized the location boundaries current at the time of the program's initiation to decide where to focus its limited resources. This often, but not always, included entire divisions. Crucially, many of these boundaries were reconsolidated during the devolution process following Kenya's 2010 constitution. Vi continued using the division and location boundaries from 2008, though administrative authority is now granted to ward administrators. Wards are primarily composed of the old locations, but they are smaller than the old divisions. To complicate matters further, many public goods investments are directed through constituency development funds corresponding to the constituencies of members of parliament–which are proportional to population and are not necessarily equivalent to wards (Barret, 2015). Furthermore, detailed maps showing the exact boundaries of the post-2010 wards and constituencies are not publicly available in Kenya, leading Vi's staff to rely on pre-2008 maps for their program planning and staff allocation.

The discussion which follows will use these 2008-era administrative division names, though many of these names and boundaries have changed, since these are the administrative units used to assign the program in the baseline period. Due to funding limitations Vi was unable to expand their operations into Bungoma Central, Bungoma South, or Webuye West Divisions until 2015. So this area—along with part of Tongaren in which they did not operate, and Lugari Division in Kakamega—was identified as the non-program area to be included in the study.

The background of shifting political and administrative units is important for the identification strategy used in this paper and the broader impact evaluation found in Hughes et al. (2017), which relies on the argument that once we account for the geospatial variables, which may vary non-randomly across the program area boundary, we can consider Vi's program area as being as good as randomly assigned

Figure 1: Study Area

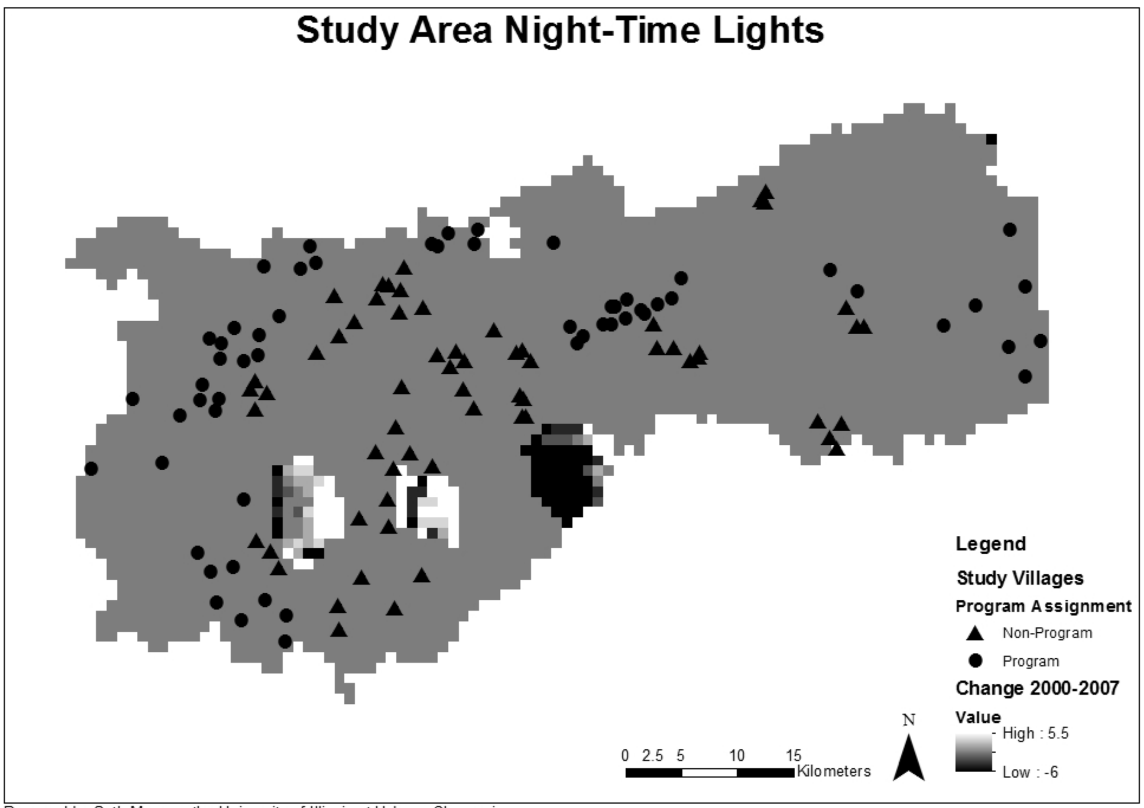


Despite the changes in administrative divisions it may still be possible that the overlap between pre-2010 divisions and post-2010 constituencies might imply a correlation with current administrative power and resource allocation, resulting in non-parallel trends in economic development between the program and non-program villages. To test this concern we examine night-time lights data for both the pre-study period from 2000 to 2007, and the first five years of the study period, from 2007 to 2013.

The National Oceanic and Atmospheric Administration's DMSP-OLS Nighttime Lights dataset represents annualized nighttime light images which have been found to be highly correlated with localized economic development and infrastructure (Doll, Muller, and Morley, 2006). Figure 2 shows the change in nighttime lights from 2000 to 2007 within the study area. These data indicate infrastructure growth in the urban areas surrounding the major towns in the study area: Bungoma, Webuye and Kimilili, while the villages included in the study show night-time lights change near zero both inside and outside the program area. This demonstrates that by excluding villages near Bungoma town, we effectively excluded almost all of the areas which experienced improvements in electricity infrastructure and–by proxy–significant economic development in the pre-study period.

The pre-treatment trends in the night-time lights data provides evidence for the hypothesis that non-parallel trends in economic development in the regions primarily center around patterns of urban expansion. To adjust for this pattern we restricted our sampling frame to villages within 10 km of the C42 roadway–the major tarmac road leading to Bungoma town–and outside of a 2 km buffer surrounding Bungoma town itself. This is intended to reduce the chance that differential trends in economic development resulting from proximity or access to Bungoma town–which did experience significant population growth both before and during the study period–might bias the results.

Figure 2: Study Area Nighttime Lights Change 2007-2013



Prepared by Seth Morgan, the University of Illinois at Urbana-Champaign
Image and Data Processing by by NOAA's National Geophysical Data Center.
DMSP data collected by the US Air Force Weather Agency.
Administrative boundaries compiled by Daniel K. Thompson, Emory University, available at mapeastafrica.com

Since the study was targeted primarily at rural villages, and the location surrounding Bungoma town was specifically excluded from the sampling frame, most study villages are within cells which show no change in nighttime lights from 2000 to 2007. Of those which do show positive change, the majority are in the non-program area, since the peri-urban areas surrounding Bungoma and Webuye are outside the program area. This means that if nighttime lights data are a reasonable proxy for economic growth, and if economic growth has a measurable effect on the relative asset wealth of the households in our sample, then the sampling frame has–if anything–biased any effect size downward by including a few villages within the non-program area which experienced growth in nighttime lights during the period when data is available. As a robustness test, outcome estimates restricted to villages who experienced no change in nighttime lights are presented in the appendix. The magnitude of the estimated treatment effect is slightly larger for some outcomes, but the difference between the estimates is not statistically significant.

It thus seems probable that villages on either side of Vi's program area boundary are highly comparable to one another in terms of their baseline wealth and infrastructure investment. One might still be worried that uneven investment in infrastructure from 2007 onward might create bias in the treatment effect, but the available nighttime lights data from 2007 to 2013 also confirms that the villages continue on parallel trends for at least the first half of the study period. Since nighttime lights in the area seems to be primarily driven by urban infrastructure, this conforms with our expectation and provides confirmatory evidence that uneven patterns of infrastructure development in the study area cannot explain the observed treatment effect. Therefore we argue that if we match villages on either side of the 2008 administrative boundaries using their geospatial characteristics, we may assume that outcomes in the matched villages outside the control area will form a valid counterfactual for the villages within the program area.

5 Methods

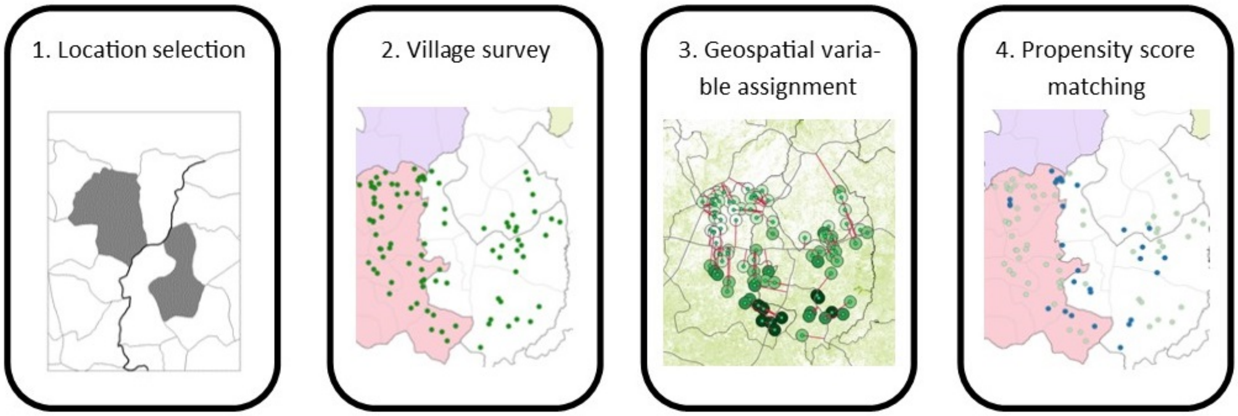
5.1 Village Selection Procedure

The execution of the village selection process described in this paper followed a four-step procedure: (1) sub-location selection, (2) scoping survey administration, (3) geospatial variable assignment, and (4) propensity score matching.

First, qualitative interviews were conducted to guide a purposive selection of sub-locations (the smallest administrative unit in Kenya) for the study. These sublocations were selected from Vi's program area and the non-program area based on their similarity in terms of relative wealth and agro-ecological characteristics. These interviews were carried out with Vi field staff, Kenya Ministry of Agriculture field officers, and farmer group leaders. Once these interviews were carried out the sample was restricted to these purposively matched sublocations.

Second, a scoping survey was administered within villages in the matched sublocations. Local consultants were hired to enter villages in the non-program areas, administer a short survey with a key informant, and take the geocode of the village center. The scoping survey instrument recorded the number of households, determined if there were active farmer groups which had been formed at or before the study period, and took note of the activities and outside NGO or government support

Figure 3: Village Selection Procedure



each of these groups received.

Third, the geocodes collected in the scoping survey were used to determine the values of the following geospatial variables. A 1 km buffer was generated around each village's central geocode, then the Zonal Statistics tool in ArcGIS was used to calculate the average value across the raster grids containing each variable's values. The village average values were then assigned to the dataset of village names using the Extract Multi Values by Points tool in ArcGIS, resulting in a table of geospatial variables as described in the data section below.

Finally, the geospatial variables described in the data section below were used to assign a propensity score and the villages were matched using this score. Due to the limited time before enumeration was to start, the scoping process was completed in four phases: the study area was split into four zones, and the matching process was applied to each one separately as the scoping data were collected. In each zone, the team selected fifteen villages in the program area and fifteen villages in the comparison area for a matched sample of thirty villages in each zone. At each stage the 30 best-matched villages were chosen by gradually reducing the caliper width using Stata's psmatch2 command, then balance was tested on the measured variables across the entire sample of villages in the completed zones. Then, as the team of enumerators began collecting data in the matched villages, the scoping team moved forward, matching the villages in each zone in advance of data collection. The details of this process are described in full in the appendix.

5.2 Geospatial Data

Villages which contained farmer-led groups which had been active since the beginning of the study period (2008-2016) were considered candidates for inclusion in the study. Using geocodes from the village centers, geospatial variables were assigned to each village for input into the matching model. The variables used to match villages include agro-ecological characteristics, as well as socio-economic indicators such as population density and distance from major roads. These variables were chosen because they are likely to effect agricultural yields and market access, and as such they would likely be significant confounders of the measured treatment effect if the selected villages were unbalanced across them. The following variables were chosen for inclusion in the propensity score model:

- Number of Households
- Average Soil Sand Content (Vågen et al., 2016)
- Average Soil pH (ibid.)
- Average Soil Organic Carbon in 2007 (ibid.)
- Average Tree Cover in 2005 (Sexton et al., 2013)
- Elevation (Jarvis et al., 2008; Kruska and Kariuki, 2016)
- Average Population Density in 2010 and 2015 (Stevens et al., 2015)
- Average Rainfall (Funk et al., 2015)
- Distance to Tarmac Road
- Binary for Villages 0.25 m from Tarmac Road ("on road")
- Binary for presence of microfinance activities

Elevation, tree cover, population density and soil variables were measured as an average value calculated across a circle 1 km in radius extending from a central point in the village. Rainfall was measured as the value of the raster cell in which the village center was found. The cells for the Climate Hazards Infrared Precipitation with Stations (CHIRPS) rainfall dataset measure approximately 5.5 km across (ibid.).

Household numbers were taken from the village-level scoping survey. The consultants requested the number of households from leaders of farmer groups, and if they were unable to estimate this number they requested it from a village elder. The tarmac road network was taken from OpenStreetMap data, and ground-truthed by travel in the region. The binary for the presence of microfinance activities was taken from Vi's records on their participating groups and from the scoping survey in the case of comparisons. Activities listed as "table banking" or "merry-go-round" were counted as microfinance activities.

5.3 Propensity Score Matching

The identification strategy for the impact evaluation of Vi's program assumes that if the geospatial characteristics which may vary non-randomly across the boundary of the program area are accounted for, then the program area is as good as randomly assigned (Hughes et al., 2017). So a method is needed to account for geospatial variables that might effect both selection and outcome such as soil type, average rainfall, or distance to major roads. The chosen methodology described here implements a matching design across villages in the study area.

To select comparison villages, we utilize a matching estimator used in the social sciences to match individuals across observable variables (Abadie and Guido W Imbens, 2016). Applying this method to villages allows us to compare households in villages within the program area to household in villages outside the program area which are similar across the range of relevant covariates. Matching assumes that average treatment effects τ can be estimated by taking the average of the difference between the expected outcome Y of untreated observations–conditional on a vector of covariates X–from the expected outcome of the treated observations conditional on the same covariates, as seen in equation one, where W = 1 denotes treatment assignment and W = 0 represents non-treatment.

$$\tau = \mathbb{E}[\mathbb{E}[Y|W=1, X=x] - \mathbb{E}[Y|W=0, X=x]] \quad (1)$$

(ibid.)

This methodology has been used to estimate the effects of protected areas (Andam, Ferraro, Pfaff, et al., 2008; Andam, Ferraro, Sims, et al., 2010; Honey-Rosés, Baylis, and Ramírez, 2011), a similar application in that national parks and other protected areas are designated non-randomly, but a particular geographic unit's assignment to the area inside or outside the park may be considered as good as random, conditional on its being matched across relevant geospatial values. This literature includes examples of matching on observational units at multiple scales, including pixels in a raster grid (Andam, Ferraro, Pfaff, et al., 2008; Robalino, Pfaff, and Villalobos, 2015), polygons corresponding to land management units (Honey-Rosés, Baylis, and Ramírez, 2011), and census tracts (Andam, Ferraro, Sims, et al., 2010). Our analysis focuses on a 1 km circular buffer drawn around each village considered for the study.

Matching designs in the literature use two primary methodologies for assessing observations' similarity across covariates: nearest neighbor and propensity score matching (Joppa and Pfaff, 2010). Nearest neighbor matching calculates the multi-dimensional distance between two observations given the vector of covariates. Propensity score matching condenses the covariates to a single score using a regression model to calculate each observation's conditional probability of receiving treatment given the covariate values (Rosenbaum and Rubin, 1983). Propensity score matching was identified as the most appropriate method, since the objective was to identify comparison villages with a high conditional probability of being included in the program area given the measured geospatial variables.

The propensity score is defined formally as the probability of treatment, conditional on a vector of covariates. The propensity score model can be expressed by equation two below:

$$e(X_i) = Pr(T_i = 1|X_i) \quad (2)$$

Where $e(X_i)$ is the probability of being included in the treatment group, and X_i is the vector of covariates listed above. The propensity score was generated using a probit model estimated within Stata by the psmatch2 command. The probit model takes the form:

$$z = X\beta + \varepsilon \tag{3}$$

Where z is an unobserved variable and y is the observed binary corresponding to treatment assignment such that:

$$y_i = \begin{cases} 1 \text{ if } \mathrm{z}_i \geq 0 \\ 0 \text{ if } \mathrm{z}_i \leq 0 \end{cases} \quad (4)$$

Following propensity score estimation, villages were matched based on their propensity scores, providing a sample of villages with similar predicted probabilities of receiving treatment. After matching, covariate balance was checked to confirm that the process generated a sample of villages for which treatment and comparison villages are balanced across all selected covariates. This covariate balance is shown in table 1.

5.4 Village-Level Balance

After sampling, a team of enumerators was sent to each village to interview the selected respondents. The enumerators collected geo-codes at each household. Once data collection was complete, a village center point was generated by calculating the mean center of these household points. In some cases this location was somewhat different from the original village point collected by the scoping team, so post-data collection balance was tested in order to check that balance remained consistent in the actual villages where sampling took place.

The table below includes balance statistics for the treatment and comparison villages across the selected set of geospatial variables. None of the variables are significant at the 5% level, indicating that it is not possible to reject the hypothesis that these villages are identical across the given variables. We therefore argue that the program area villages are indistinguishable from the non-program area villages across the measured variables.

Table 1: Village Balance Statistics

	Sample Mean	Program Mean	Non-Program Mean	Normalized Difference	Difference/se
Soil Sand Content	19.96	20.57	19.36	0.11	1.21
					(1.42)
Soil pH	5.95	5.97	5.94	0.20	0.03
					(0.02)
Tree Cover 2005	6.07	5.97	6.17	-0.06	-0.21
					(0.45)
Elevation	1570.52	1575.63	1565.49	0.05	10.13
					(26.18)
Population Density 2010	4.41	4.40	4.43	-0.02	-0.03
					(0.22)
Soil Organic Carbon 2007	25.57	24.71	26.43	-0.19	-1.72
					(1.16)
Rainfall	136.71	133.97	139.40	-0.23	-5.42
					(2.90)
Distance to Road	0.03	0.03	0.03	0.07	0.00
					(0.00)
On Road	0.02	0.02	0.03	-0.07	-0.02
					(0.03)
Observations	121				

5.5 Household Dataset

Having shown that the sets of villages inside and outside the program area are balanced across village-level geospatial variables, it remains to be seen that the sample of households within these villages are indistinguishable across socio-economic variables. We expected the households to be balanced on socio-economic characteristics in as much as they would be correlated with geospatial variables. However it remains an empirical question.

The data used in this analysis are the result of household surveys administered to randomly selected members of Self-Help Groups (SHGs) existing in the villages through the matching process described above. After village selection was complete, an advance team of enumerators was sent to each village ahead of the data collection team to prepare lists of farmer group members. The project's sampling frame required that all respondents in both treatment and comparison areas pass the following screening criteria:

- Must be a member of a group that was formed in the year 2008/09 or before
- Must have been an active member of that group since 2008/09 or before

- Household must have existed in 2007 or before
- Household must have been farming the same main parcel of land from 2007 to the present

These screening criteria guarantee that all respondents in the non-program area have been active members of SHGs for the entire study period, ruling out the possibility that the treatment effect produced by the program is due to Vi's model of service delivery, which operates through the recruitment of pre-existing SHGs.

The sampling team contacted farmer group leaders in the selected villages and requested lists of members who met the above screening criteria. Once the lists from each group in the village were assembled they followed a random selection procedure to select 12 female respondents and 12 male respondents from each village, when equal numbers of males and females were available in the SHGs' membership. When male participation was not sufficient, a sample of females was substituted. These respondents were then informed and mobilized by the group leaders so that they would be present and available for interviews at the time chosen for data collection.

The survey was administered to 2,860 households from 121 villages, of which 63 households were dropped from the sample due to data quality issues or violation of the screening criteria, leaving 2,797 observations in the dataset. Summary statistics for this sample are shown in table 2.

Table 2: Summary Statistics for Baseline Variables

	Sample Mean	Program Mean	Non-Program Mean
Distance to Road	3.02	3.08	2.96
Respondent Luhya	0.97	0.97	0.98
number of productive adults in HH	3.51	3.52	3.50
number of children in HH	3.64	3.56	3.71
number of adults in HH	2.96	2.97	2.94
Respondent Literate	0.89	0.89	0.89
Years of education head	9.14	9.08	9.20
female headed HH	0.22	0.22	0.22
Formal Land Title	0.33	0.35	0.32
Respondent Holds Land Title	0.58	0.55	0.62
HH Member Formally Employed	0.15	0.16	0.13
HH Member Owns Business	0.30	0.30	0.31
HH Member Farmer	0.98	0.99	0.98
Land Size at Baseline	2.33	2.29	2.37
HH Owns Livestock	0.59	0.62	0.57
PCA-Weighted Assets 2007	1.51	1.54	1.49
Predicted Per Capita Expenditure 2007	3.97	4.02	3.93
Observations	2797		

5.6 Balance at Household Level

Given this sample of households which was matched at the village level, two empirical questions remain: is the sample well-balanced at the household level, and did the matching process improve on a simple selection of random villages on either side of the program boundary? The balance statistics in table 3 demonstrate that the resulting sample is indeed balanced across baseline socio-economic variables. While we are not able to reproduce a sample using a simple random selection across the boundary, we are able to provide evidence which may address the second question by splitting the sample into the closest matched villages and the farthest, based on propensity score, and compare balance in these split samples.

Table 4 below displays balance statistics for a sub-sample of the population chosen by using a caliper of 0.1 on the original village matching model, implemented in Stata using the psmatch2 command. Table 5 reports balance statistics for the rest of the sample, those villages not within the 0.1 caliper. Observe that the observations presented in table 3, where the village matching is restricted to a 0.1 caliper, are much better matched than the other specifications, including the whole sample shown in table 3, with no variables for which we can rule out the null hypothesis that they are identical with greater than 95% certainty.

Table 3: Balance Statistics for Whole Sample

	Difference	Normalized Difference
Distance to Road	0.12*	0.04
Respondent Luhya	-0.01	-0.04
Formal Land Title	0.03*	0.04
female headed HH	-0.00	-0.01
HH Member Formally Employed	0.03**	0.06
HH Member Owns Business	-0.00	-0.00
HH Member Farmer	0.00	0.03
Land Size at Baseline	-0.08	-0.02
HH Owns Livestock	0.06***	0.08
PCA-Weighted Assets 2007	0.06	0.04
Predicted Per Capita Expenditure 2007	0.09	0.03
Observations	2797	

* p<0.1, ** p<0.05, *** p<0.01

It should also be noted that the sample shown in table 5 is already likely to be better matched than a sample using a random selection of villages in the study area, since it still only contains villages which were matched across geospatial variables. So if it is true, as indicated by the improved match within the 0.1 caliper, that the village matching process results in a closer matched sample, then it is likely true that even the worst-matched half of the presented sample would match closer than a sample which included the other villages discarded by the PSM model. Moreover the fact that a more restrictive village match results in a sub-sample better matched across household variables can be considered evidence that the village matching methodology is an improvement over an unmatched sample.

Table 4: Summary Statistics for Sample Inside 0.1 Caliper

	Difference	Normalized Difference
Distance to Road	0.06	0.02
Respondent Luhya	0.01	0.03
Formal Land Title	0.03	0.05
female headed HH	-0.02	-0.03
HH Member Formally Employed	-0.02	-0.04
HH Member Owns Business	0.00	0.01
HH Member Farmer	0.00	0.02
Land Size at Baseline	0.08	0.03
HH Owns Livestock	0.04	0.06
PCA-Weighted Assets 2007	-0.07	-0.05
Predicted Per Capita Expenditure 2007	-0.13	-0.05
Observations	1365	

* p<0.1, ** p<0.05, *** p<0.01

Table 5: Summary Statistics for Sample Outside 0.1 Caliper

	Difference	Normalized Difference
Distance to Road	0.18*	0.06
Respondent Luhya	-0.03***	-0.12
Formal Land Title	0.03	0.04
female headed HH	0.01	0.01
HH Member Formally Employed	0.08***	0.16
HH Member Owns Business	-0.01	-0.01
HH Member Farmer	0.01	0.03
Land Size at Baseline	-0.23*	-0.07
HH Owns Livestock	0.07***	0.10
PCA-Weighted Assets 2007	0.17***	0.11
Predicted Per Capita Expenditure 2007	0.31***	0.10
Observations	1432	

* p<0.1, ** p<0.05, *** p<0.01

By way of comparison, restricting the sample to a subset of villages which cluster closely around the boundary between the program and non-program areas also results in a matched sample with smaller differences between program and non-program average values. Table 6 displays balance statistics for roughly half of the sample which is closest to the boundary between program and non-program areas. Table 7 shows the balance statistics for the other half of the sample, taken from villages farther away from the program boundary.

Table 6: Summary Statistics for Closest Half of Sample

	Difference	Normalized Difference
PCA-Weighted Assets 2007	0.07	0.04
Predicted Per Capita Expenditure 2007	-0.02	-0.01
Distance to Road	0.52***	0.19
On Road	-0.02*	-0.07
Formal Land Title	0.01	0.01
Land Size at Baseline	-0.17	-0.05
HH Owns Livestock	0.04	0.06
Respondent Received Technical Education	0.02*	0.07
Max Years of Education in HH	-0.17	-0.04
Observations	1350	

* p<0.1, ** p<0.05, *** p<0.01

Table 7: Summary Statistics for More Distant Half of Sample

	Difference	Normalized Difference
PCA-Weighted Assets 2007	0.05	0.03
Predicted Per Capita Expenditure 2007	0.20*	0.07
Distance to Road	-0.25**	-0.09
On Road	-0.00	-0.01
Formal Land Title	0.05**	0.08
Land Size at Baseline	0.01	0.00
HH Owns Livestock	0.07***	0.10
Respondent Received Technical Education	0.02	0.06
Max Years of Education in HH	-0.28	-0.06
Observations	1447	

* p<0.1, ** p<0.05, *** p<0.01

Note that the balance statistics for the villages which are closest to the program boundary are indeed somewhat better than the balance statistics for the more distant villages. But one variable in particular has a larger difference for the closer sample: distance to road. Restricting the sample to those which are simply closest to the program boundary resulted in forcing some matches with varying distances to the nearest tarmac road. This would create a sample where the program area villages have a higher degree of access to markets, potentially biasing the sample.

In this case the the distance to road variable highlights the greater flexibility of the village-level propensity score matching methodology. This illustrates a strength of the propensity score matching methodology over a simple spatial discontinuity: it allows us to assemble a sample with an explicit goal of balancing over a vector of relevant covariates. It might often be the case that infrastructure or town centers may be sited in such a way that a simple comparison within and without an area might lead to bias. In this case matching may well outperform a simple comparison across borders, since it enables us to utilize a wider set of villages across space while narrowing the candidates down though matching across a set of variables.

6 Empirical Approach

6.1 Dependent Variable

In order to examine the sensitivity of the outcomes to the match balance, we will present results for the primary household wealth outcomes used in the Vi Agroforestry impact evaluation. The primary dependent variable presented here is a measure of asset wealth: an asset index constructed by performing Principal Component Analysis on the set of assets included in the household survey. The household survey recorded ownership as well as quality characteristics of over a hundred different household assets in five categories: household items, transportation, agricultural equipment, livestock, and housing characteristics. Of these, we created 91 binary asset variables by categorizing the quality as above or below the median quality for each asset, and in some cases combining assets such as sheep and goats into a single indicator. The result is a matrix of binary variables indicating asset ownership in 91 categories.

Principal Component Analysis (PCA) reduces this matrix into a single component by using the leading eigenvectors of the covariance matrix to construct a component which maximizes the over-all variance within the matrix (StataCorp, 2013). This component is assumed to represent the underlying construct represented by the 91 asset variables. Therefore the asset score generated by the predict command following PCA in Stata is taken to be a measure of the over-all asset wealth of the household.

The asset index score given to each household in this methodology is a weighted sum of the assets which they own, with the weights derived from the standardized first principle component of matrix described above. The asset index score can be represented by equation three:

$$A_i = \hat{\gamma}_1 a_{i1} + ... + \hat{\gamma}_k a_{ik} \tag{5}$$

Where A_i is the asset score given to each observation, $a_1 - a_k$ represent the list of assets included from the survey data, and $\gamma_1 - \gamma_k$ represent the weights derived from the first principal component assigned to each asset respectively (Sahn and Stifel, 2003).

The PCA method for deriving an asset index has been found to correlate well with consumption expenditure (Filmer and Pritchett, 2001) and to outperform consumption expenditure as a predictor of household nutrition indicators (Sahn and Stifel, 2003).

In this study the asset index estimated for 2016 asset holdings has a correlation coefficient of 0.52 with 2016 daily consumption expenditure, and the differenced asset index measure has a correlation coefficient of 0.74 with a differenced measure of predicted consumption expenditure. Additionally, following Sahn and Stifel (2003) we also find that the PCA asset index is more highly correlated with dietary measures than is the level of consumption expenditure, with a correlation of 0.23 between assets and the number of dietary categories consumed in the last 24 hours, as compared to a correlation of 0.15 for consumption expenditure.

In addition to the PCA-constructed asset index for 2016, we also construct a measure of principal component-weighted growth in asset wealth. This second asset measure consists of a set of assets presumed to be commonly owned in both 2016 and 2007 (excluding items such as computers and cell phones, which were not widespread in rural Kenya in 2008), which are differenced, and PCA analysis performed on the positive differenced binaries. Thus the underlying construct is assumed to be positive growth in asset wealth over the period from 2007 to 2016.

An alternate way to generate this outcome variable would be to generate a separate score in each time period, then the two asset indices are differenced to reflect change over time. This procedure has the theoretical advantage of accounting for differences in asset value during the two periods, i.e. if a radio was relatively more expensive and therefore represented a higher level of over-all wealth in 2007 than in 2016 this procedure would account for that difference (Moser and Felton, 2007). However, this has the disadvantage of not explicitly measuring which assets are most correlated with growth in asset wealth over time, which is an important concern of the study.

The two methods of principal component analysis are highly correlated with a coefficient of 0.84. This paper focuses on the PCA-weighted growth in assets, with weights assigned according to the principal component associated with the differenced binaries. But the outcome regression coefficient if using the variable constructed by separate weights from 2007 and 2016 is qualitatively similar, with the same sign, but wider standard errors. See appendix for results using this alternate outcome variable.

6.2 Empirical Models

To estimate the effects of the Vi Agroforestry extension program on household assets and expenditure, we use three models. Intention to Treat effects are given by the OLS model shown in equation 6, where P represents residence in a village within the Vi program area, the coefficient τ represents the ITT effect of the program, and $x_1 - x_k$ is a vector of covariates with coefficients $\beta_1 - \beta_k$ respectively.

$$Y = \beta_0 + \tau P + \beta_1 x_1 + ... + \beta_k x_k + \varepsilon \tag{6}$$

As a robustness test we also present ITT effects using an inverse-probability weighted, regression-adjusted estimator. This doubly-robust model controls outliers by weighting observations by their probability of being included in treatment, as well as estimating a linear model of the outcome in both the treatment and control groups and taking the treatment effect from the differenced expected values in each group. In equation 7 below, τ represents the estimated treatment effect found by taking the average of the difference between the predicted outcome $\hat{\mu}$ given by linear model g_k, estimated for the treatment population ($t_i = 1$) and the control population ($t_i = 0$). Observations in this model are weighted by their inverse probability of receiving treatment, given by w_i in equation 9.

$$\tau = 1/n \sum_{i=1}^{n} \{\hat{\mu}_{i,k,wreg}(X_i, t_i = 1) - \hat{\mu}_{i,k,wreg}(X_i, t_i = 0)\} \tag{7}$$

$$g_k(\mu_{i,k}) = \gamma_k t_i + x_i \beta_k; \tag{8}$$

with inverse probability weights given by:

$$w_i = t_i/\hat{\rho}_i + (1 - t_i)/(1 - \hat{\rho}_i) \tag{9}$$

where ρ = the estimated probability of residing in the program area give the vector of covariates X. (Kreif et al., 2013)

This model is called "doubly-robust" because if either of the treatment model which provides the inverse-probability weights, or the outcome model which provides the regression adjustment, is misspecified, the estimate remains unbiased. The fact that the ITT effects remain qualitatively the same in both the OLS and doubly-robust models is an indication that the treatment effect is unlikely to be entirely driven by outliers.

I also present Local Average Treatment Effects estimated by a Two-Stage Least Squares model. In this model, residence in a village within the Vi Agroforestry program area is used as an instrument for participation in Vi's program. The first stage, shown by equation 10, fits a model where v represents participation in the Vi program, given residence in the program area P and a vector of covariates $x_1 - x_k$ with coefficients $\delta_1 - \delta_k$ respectively.

$$v = \delta_0 + \tau P + \delta_1 x_1 + ... + \delta_k x_k + \varepsilon \tag{10}$$

In the second stage, the local average treatment effect is given by the coefficient ρ corresponding to the fitted value $\hat{v}$ generated by the first stage equation.

$$y = \gamma_0 + \rho\hat{v} + \gamma_1 x_1 + ... + \gamma_k x_k + \eta \tag{11}$$

To conclude that ρ in the equation above represents an unbiased estimate of the effect of participation in Vi Agroforestry's program we must rely on two primary assumptions: that the outcome is uncorrelated with the instrument, the program area, and that the program area effects the outcome only by way of increasing the probability of participating in Vi's program (Angrist and Guido W. Imbens, 1995).

Given the discussion above concerning the boundaries of the program area, both assumptions are likely to hold in this case. If we observe that the program area is defined by outdated administrative boundaries, and contains rural villages which experience little to no change in infrastructure or economic development as measured by nighttime lights during the pre-study period, it seems reasonable to assume that no other intervention or investment which might effect the outcome is significantly correlated with Vi's program area.

7 Results

7.1 Wealth Outcomes

Rows 1-3 of table 8 report the coefficients of interest corresponding to the OLS model, the doubly-robust IPWRA model and the two-stage least squares (2SLS) model respectively. Column 1 displays the treatment effect where the principal component-weighted asset index is the outcome. Column 2 contains the coefficients where differenced predicted household expenditure is the outcome, and column 3 contains values for the coefficients associated with differenced predicted per capita expenditure.

All equations use fixed effects at the zone level, corresponding to Vi's program zones, which have separate staff. All specifications presented in table 8 use a vector of covariates which a model of the household production function would predict might have an impact on the outcome, along with covariates which were correlated with the program area at the 10% significance level. Results from specifications using a more parsimonious set of covariates including only those which correlate with the program area can be found in the appendix, as well as specifications with county fixed effects.

The local average treatment effect estimation for the PCA-derived asset index is 0.09. This corresponds to approximately 12% of the mean in the differenced variable. In other words, this treatment effect indicates that program participation is associated with 12% more growth in assets over the course of the study period compared to non-participation. An additional note which may help in interpreting this coefficient: 0.09 is similar in scale to the component weights assigned to kitchen utensils or pots and pans. This means that participation in the program has a similar correlation with the underlying construct measured by the first principal component as, say, a set of pots and pans, while it has a lower correlation than large assets such as improved walls or flooring material.

In terms of growth in predicted consumption expenditure, the participation in Vi's program is associated with a $0.39 PPP increase compared to the non-participants. As you might expect, since both measures are based on the same underlying assets, just as in the PCA asset index this is equivalent to 12% of the mean difference. Relating it to expenditure in purchasing power parity-adjusted dollars allows us to more straightforwardly compare this change over time to the baseline mean. The effect size is approximately 3% of mean baseline predicted consumption expenditure per household, which is about $13.00 per household, per day. The per capita numbers, adjusted for household size and household returns to scale, retain the same ratio. In terms of commonly consumed food items in the local market, the effect on expenditure scales to almost 2 chapatis (a flatbread commonly consumed in East Africa) per household per day, or approximately one small avocado per person per day.

Table 8: Wealth Outcomes

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
ITT (OLS)	0.07***	0.31*	0.10*
	(0.02)	(0.18)	(0.06)
ITT (IPWRA)	0.07***	0.30*	0.09
	(0.02)	(0.18)	(0.06)
LATE (2SLS)	0.09***	0.39*	0.12*
	(0.03)	(0.23)	(0.07)
Observations	2785	2785	2785

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

7.2 Outcome Sensitivity to Match Balance

As with the sample balance, the wealth outcome estimates are somewhat effected by the quality of the geospatial match. Tables 9, 10 and 11 display ITT estimates for each of the primary outcome variables using a nested model to compare the outcomes using the worst-matched sample (villages with propensity scores outside 0.1 caliper), and outcomes using the best-matched sample (inside the 0.1 caliper). Column 1 shows results for the villages outside the 0.1 caliper, column 2 shows results for villages inside the 0.1 caliper, and column 3 replicates the results above from the entire sample.

Note that the point estimates are indeed effected by the match quality, and the estimates in the sample outside the 0.1 caliper are higher than those from the sample inside the caliper. This indicates the possibility of upward bias if no matching used to select the sample. However, a test for equality of coefficients cannot rule out the null hypothesis that they are equal. So the upward trend as match balance worsens seems to indicate that the matching process may have prevented upward bias, though it is impossible fully account for variation which might occur out of sample in the absence of outcome data from villages not included in our sampling frame.

Table 9: PCA Assets: ITT Estimates with Samples Inside and Outside 0.1 Caliper

	(1) Outside Caliper	(2) Inside Caliper	(3) Full Sample
ITT (OLS)	0.077**	0.060*	0.069***
	(0.035)	(0.036)	(0.024)
Observations	2785	2785	2785

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 10: Differenced Predicted Household Expenditure: ITT Estimates with Samples Inside and Outside 0.1 Caliper

	(1) Outside Caliper	(2) Inside Caliper	(3) Full Sample
ITT (OLS)	0.359	0.248	0.306*
	(0.259)	(0.253)	(0.179)
Observations	2785	2785	2785

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 11: Differenced Predicted Household Expenditure: ITT Estimates with Samples Inside and Outside 0.1 Caliper

	(1) Outside Caliper	(2) Inside Caliper	(3) Full Sample
ITT (OLS)	0.359	0.248	0.306*
	(0.259)	(0.253)	(0.179)
Observations	2785	2785	2785

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

7.3 Alternate Spatial Models

In addition to the matching model which is described above, there are two other estimators which might be considered for use in a design like this one in which spatial variation plays a strong role: spatial regression discontinuity and spatial fixed effects. Spatial regression discontinuity treats distance from the boundary between the treatment area and the untreated area as a running variable, with a sharp discontinuity at the boundary which is taken to be the effect of treatment (Keele and Titiunik, 2015).

This design assumes continuity across space–i.e. that all covariates vary smoothly across space so that it is possibly to treat the distance from the boundary as a continuous running variable. This assumption may be violated in the case of the Vi program area, in which features such as soil types and tarmac roads cut across the area creating important variations which do not vary smoothly with distance from the program area boundary.

Imagine a patchwork of spatial variables overlaid non-continuously with boundaries which do not smoothly vary across space. If the corridor around the program area happened to include some pixels with–say–high soil organic carbon, the spatial RD estimator would suffer from omitted variable bias. The spatial RD also relies on a mass of data near the boundary line, so if the data points are more dispersed there might be increasing possiblity of such bias. The matching estimator on the other hand, which allows greater heterogeneity in terms of distance, explicitly controls for the variables in the model, solving this problem if the correct covariates are specified.

Nevertheless we present the spatial RD results in table 12, as produced by the rdrobust command in Stata 14. Note that for the PCA-weighted asset measure the results are almost double the coefficient from the matching design presented above. It is difficult to rule out the hypothesis that this estimation is biased upward by the way in which the covariates are distributed across space.

Table 12: Spatial Regression Discontinuity Results

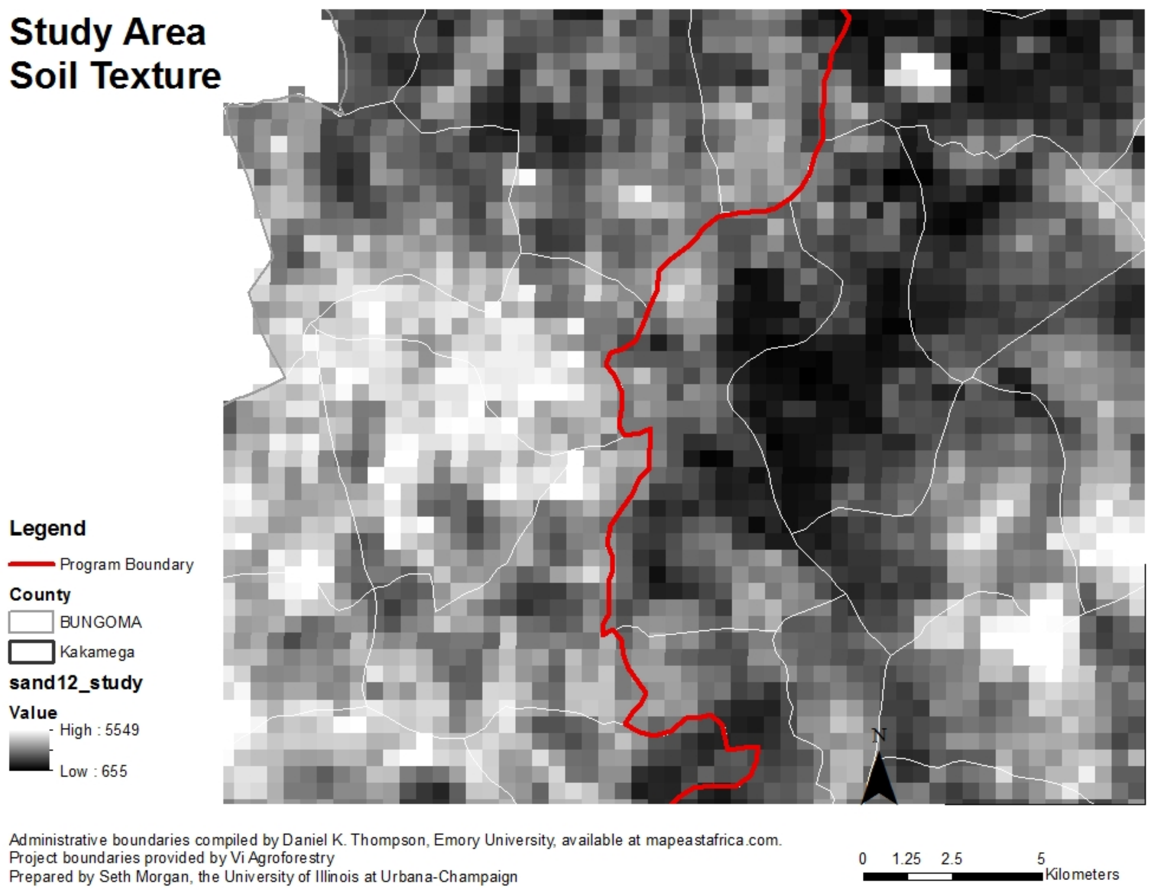
	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
RD_Estimate	0.169*	2.359***	0.358
	(2.11)	(3.55)	(1.81)
N	1489	1207	1489

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 5 demonstrates one reason why this methodology might introduce bias given the study area for this project. Note that soil sand content–an important measure of soil texture with considerable agronomic significance–is not distributed in an even gradient across the program boundary. In fact, if you matched villages solely based on nearness to the boundary you would likely compare villages with high sand content (the light colored region in the left of the figure) to villages with low sand content (the dark colored region in the center). If there is concern that sand content might effect the outcome being measured–which is a consideration in the present study given its agricultural nature–then it may be more desirable to match villages with high sand content with other villages on the other side of the peninsula of low sand content, creating a match which is more similar across observable soil characteristics though the villages are farther apart.

Figure 4: Sand Content on Either Side of Program Boundary



Another way to account for variation across space in the treatment effects estimation would be spatial fixed effects. In order to implement a spatial fixed effects model we restrict the observations to those which are near to the boundaries between program and non-program areas. We then add a variable indicating to which of the five lines dividing program from non-program areas each observation is nearest. The combination of this spatial restriction and the variables for boundary lines provides a fixed effect for the neighborhood each observation occupies, allowing a comparison within a narrow spatial extent, and controlling for whatever unobservable differences might vary across wider spaces in the study area (Magruder, 2010).

Table 13: Wealth Outcomes with Spatial Fixed Effects

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
OLS	0.07**	0.37	0.11
	(0.03)	(0.25)	(0.07)
2SLS	0.09**	0.48	0.15
	(0.04)	(0.33)	(0.10)
Observations	2340	2340	2340

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

This specification is essentially a more localized set of fixed effects than the models presented in table 8, which included fixed effects for the program zone–a set of 4 zones determined by Vi Agroforestry's staffing allocation. The spatial fixed effect controls for variation across the five boundary lines demarcating the program area. The results from this specification are qualitatively the same as the results produced by the models in table 8. Results with spatial fixed effects and a restriction to within 0.03 arc degrees (about 3.3 km at this latitude) from each boundary line are shown in table 13. The results are robust to variation in the distance restriction, though increased standard errors reflect the decrease in sample size due to the distance restriction.

Table 14: Summary Statistics for Sample Inside 0.1 Caliper

	Difference	Normalized Difference
PCA-Weighted Assets 2007	0.01	0.01
pre_ce_pc_07	-0.03	-0.01
dist_tarma	0.53***	0.20
hh_on_road	-0.03***	-0.13
lnd_title_07	0.02	0.03
Land Size at Baseline	-0.12	-0.04
hh_lstock_07	0.04**	0.06
r_tech_skill	0.02**	0.06
Highest years of education of any adult in HH	-0.24*	-0.05
Observations	2351	

* p<0.1, ** p<0.05, *** p<0.01

It should also be noted that the tightened distance restriction called for by the spatial fixed effects specification presented here does worsen the balance across observable variables. Balance statistics for this restricted sample are presented in table 14. Note especially the worsened balance across the distance to major roads. Given the importance of the major tarmac roads to market access it seems reasonable to allow more flexibility on distance from the program boundary in order to achieve better balance across this variable and others which do not vary smoothly across the program area boundary.

8 Discussion

The results presented in the outcome tables below represent a modest welfare gain for participants in Vi Agroforestry's program. Of particular interest for this paper's purpose is the degree to which geospatial balance mapped to household-level balance, and the relative sensitivity of the estimates to the geospatial match quality. The results above indicate that village-level balance across geospatial variables did, in this case, translate into a sample of households matched across baseline socio-economic variables. They also demonstrate the value of the matching process by illustrating the possibility for bias created by a poorly matched sample.

Results estimated from the least-well matched half of the sample exhibit higher point estimates compared to both the entire sample and the subset of best-matched villages. This gestures toward the possibility that the matching process utilized was prevented upward bias, though it was not possible to collect out-of-sample data which would confirm this trend.

Outcome estimates using alternate spatial specifications largely confirm the primary conclusions. The estimates from a spatial regression discontinuity exhibit what appears to be significant upward bias, since this specification is not robust to covariates which do not vary smoothly across the program boundary. Results from a spatial fixed effects utilizing proximity to program boundary lines to restrict the sample largely mirror the primary results, with somewhat higher point estimates, but wider standard errors thanks to the reduced sample size.

It is of course impossible to compare the methodology demonstrated here to all other possible methods for assembling an appropriate sampling frame, but it should be noted that the village-level matching methodology was implemented with a small team in a limited time frame on a limited budget.

As an example of how this pre-matching added value to the over-all research design, a simple power calculation using the standard deviation of the baseline wealth variable in the best-matched half of the sample vs. the worst-matched half of the sample reveals that the sample size necessary to detect the observed effect size of 0.08 with 95% confidence grows from 1704 to 1874 when moving from the well-matched sample to the less-well matched sample–an increase of 170 households, or 10%. Given that even the less-well-matched sample in this data is still drawn from villages which were geospatially fairly well matched, it is reasonable to assume that the study would lose even more statistical power if the sample had not been matched at all.

9 Conclusion

Agroforestry is an intervention that poses significant challenges to researchers analyzing its impacts. Because it is long-term and highly dependent on local agro-ecological context, it is not easily amenable to evaluation by randomized control trial. This paper has presented a methodology whereby a counterfactual can be generated in the context of a quasi-experiment where the treatment is assigned by geography. The method demonstrated uses village matching across geospatial variables to create a credible counterfactual comparison group in a relatively cost-effective manner, making impact evaluation feasible even in the absence of reliable baseline data. This approach lends itself well to estimating the effects of agroforestry, but would also have value in other similar interventions whose effects might take a long time to manifest such as protected areas, farmer-managed natural regeneration, conservation agriculture, terracing, or use of organic inputs.

The positive gains in asset wealth associated with agroforestry in these results speak into a growing debate on the relative importance of outside finance for the implementation of agricultural techniques linked to climate change mitigation. These finding suggest that agroforestry may be sustainable for farmers even in the absence of payments for ecosystem services. But their modest size also indicates that programs targeted at increasing the reach of agroforestry are unlikely to be financed by the private sector. If then agroforestry is seen as a cost-effective way to sequester carbon in agricultural systems, it may be necessary to subsidize the program costs of organizations such as Vi who are tasked with implementing training and distributing tree seeds to farmers.

The magnitude and time-frame of the agroforestry impacts examined in this paper provide the opportunity to demonstrate a method for evaluating interventions which may present challenges to experimental methods. The significance of this work is its demonstration of geospatial matching as a way to construct a counterfactual of a long-term intervention in the absence of baseline data. Further study may be needed to determine the external validity of this finding. It is not guaranteed that matching across the geospatial variables presented here would achieve the same degree of sample balance in a context with different degrees of agro-ecological or socio-economic variability than found in western Kenya.

In many respects the sample presented here has a relatively high degree of homogeneity: nearly all small-holder agriculturalists, mostly within the upper-midlands area of western Kenya, with no urban areas or pastoralists communities included. Nevertheless, this sample well represents the target population for agroforestry, and may reflect the target population for a number of other interventions related to sustainable agriculture and climate change mitigation as well. We therefore argue that geospatial matching merits consideration for ex-post impact evaluations of a wide variety of interventions of interest in the environmental and agricultural literature, and represents a potentially cost effective methodology with broad applicability.

A Appendix A: Village Selection Process

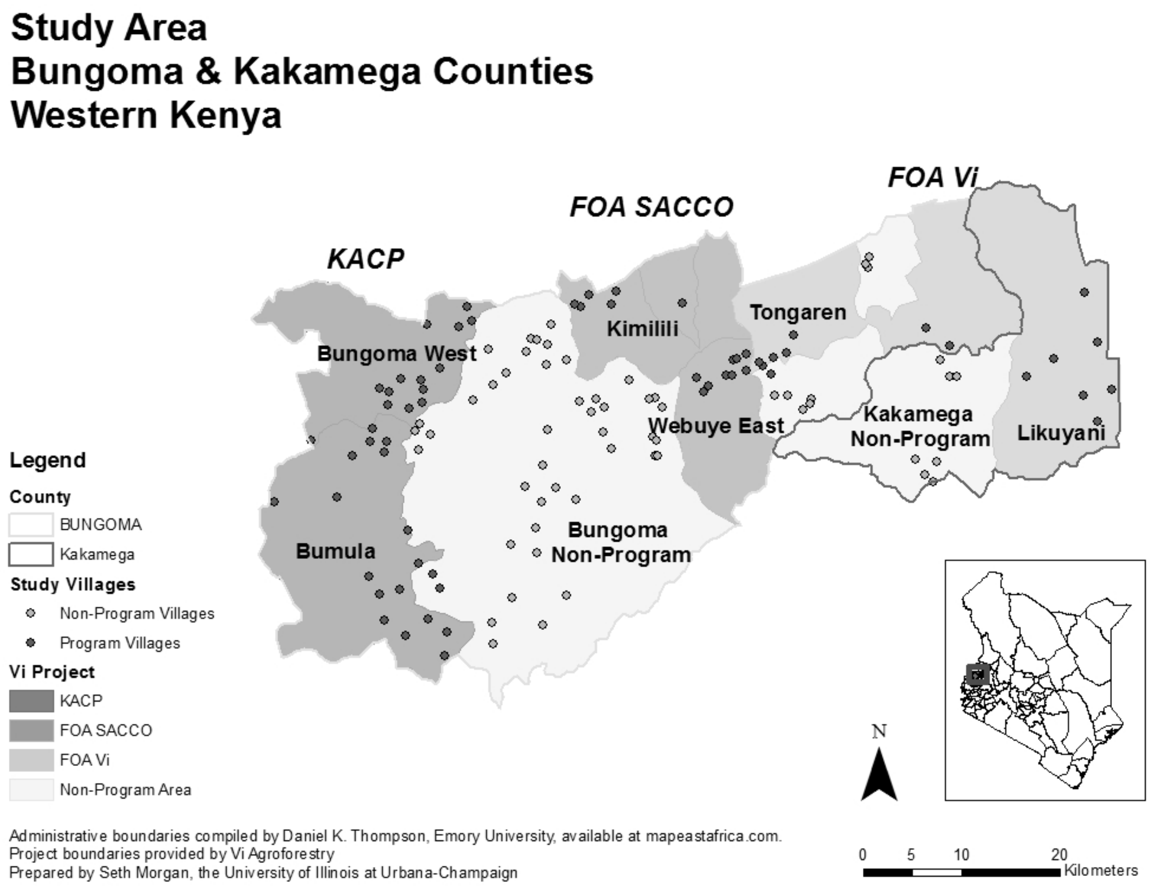
The methods section of this paper reports the village selection procedure in the abstract. However, due to time constraints scoping was not completed before data collection was scheduled to proceed, so this procedure was implemented separately within four defined geographic zones, then balance was tested across the whole study area in an iterative fashion as data was collected in each zone.

For planning and logistics purposes the study area was broken into four zones, roughly corresponding to the zones defined by Vi itself when planning its operations. The study area zones, with names drawn from the administrative divisions which they contain, are as follows:

- Sirisia/Malakisi
- Bumula
- Kimilili/Ndvisi
- Tongaren/Likuyani

Each of the above zones has its own team of Vi staff, or in some cases, its own affiliation with local Savings and Credit Cooperatives (SACCOs) which carry out training and extension as Vi's partners. Given the different staff and different geographic areas each of these zones represents, they each represented differing challenges for the scoping team as the selection procedure was implemented. The zones are depicted below in Figure 4.

Figure 5: Study Area with Zone Labels



A.1 Sirisia/Malakisi

The first zone included in the matching algorithm was the zone comprising the Sirisia and Malakisi divisions of Bungoma County, along with the comparison area in Bungoma Central which was regarded as similar by key informants. This area included 53 villages, 29 treatment and 24 comparison. The sample was restricted to villages within 10 km of the C42 roadway, a major road that runs more or less through the middle of the study area. This restriction prevented matches that were too far away from Bungoma town or the major roads to be a credible match with the comparison areas. Balance statistics for the propensity score model estimated on this restricted sample are found in Table 1.

Table 15: Balance Statistics for PSM Matched Sample

	Sirisia	
households	39.81	(1.23)
sand	1.535	(0.86)
ph	0.0138	(0.36)
soc07	-2.943	(-1.01)
tree05	-0.745	(-0.77)
elev	10.19	(0.32)
pop10	-0.200	(-0.47)
pop15	-0.231	(-0.47)
AvgRainBun	0.935	(0.31)
dist_tarma	0.000481	(0.07)
on_road25	-0.125	(-1.05)
vilmicro_yn	0	(.)
N	32	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Data were collected in 30 of the 32 villages displayed in this model. Due to a misspecification during the time of village selection, which took place during enumerator training, one day before data collection was scheduled to begin, two villages were left out. Household size was left out of the model when villages were being assigned for data collection. Including it actually expands the sample of villages selected by the model with good balance statistics. This means that two villages were not selected.

One of these two, Ngalasia, was a neighbor to a program village called Lutaso. Once the advance sampling team began their work, it was found that there were too few farmer groups in Lutaso itself and that these groups shared members with the neighboring village of Ngalasia. So the group members in Ngalasia were sampled, in a cluster with Lutaso. The remaining village, Teremi B, was not sampled at all in this zone, but it is located in a sublocation which was identified as a potential match for both Sirisia/Malakisi and the Kimilili zone. So in the end it was included in that sample though it was not included in the Sirisia/Malakisi matched sample. Balance statistics using two-way t-tests on the villages actually included in the Sirisia/Malakisi sample are below in Table 2.

Table 16: Balance Statistics for Actual Sample

	Sirisia	
households	29.03	(0.92)
sand	1.001	(0.57)
ph	0.00810	(0.20)
soc07	-2.547	(-0.85)
tree05	-0.790	(-0.80)
elev	13.21	(0.40)
pop10	-0.202	(-0.46)
pop15	-0.232	(-0.46)
AvgRainBun	0.331	(0.11)
dist_tarma	-0.000956	(-0.14)
on_road25	-0.0708	(-0.65)
vilmicro_yn	0	(.)
N	31	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As the above table demonstrates, balance over key variables was not significantly affected by the village substitutions made due to logistical complications.

A.2 Bumula

The next treatment zone to be included was to be Bumula. This zone encompasses the Bumula division on the Western side of Bungoma County as well as much of Bungoma Central. The Bumula zone is near to Bungoma township, the largest town and county seat of Bungoma county. In order to reduce the risk that proximity to Bungoma—which according to key informants has experienced significant growth in the past few years—would violate the parallel trends assumption, a buffer of 2 km was generated surrounding the edge of the location which contains Bungoma town. Any villages within this buffer were dropped from the sample.

The resulting list of villages included 91 villages; 54 treatment, 37 comparison. The same propensity score model as above was applied to this list of candidate villages. Table 3 displays balance statistics for this propensity score match.

Table 17: Balance Statistics for PSM Matched Sample

	Bumula	
households	55.93	(1.37)
sand	-2.151	(-1.04)
ph	0.0000327	(0.00)
soc07	-0.800	(-0.57)
tree05	0.276	(0.34)
elev	-63.36**	(-3.20)
pop10	-0.158	(-0.59)
pop15	-0.184	(-0.60)
AvgRainBun	1.496	(0.62)
dist_tarma	0.00545	(1.08)
on_road	0	(0.00)
vilmicro_yn	-0.133	(-1.47)
N	30	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Here too the demands of the data collection schedule forced a departure from the desired sample. As the enumeration team entered Bumula, some scoping data were still coming in, including geocodes for some comparison areas which would be used to acquire geospatial variable values. So a preliminary model using available data was estimated, and the enumerator team was sent to areas perceived to be highly likely to be sampled once the entire sample was complete.

The discrepancy between the list of matched villages generated by the full propensity score model and the list of villages actually sampled was limited to two comparison villages and four treatment villages. In the comparison areas the village of Khakula in the E. Bukusu Location was included in the actual sample despite not being chosen by the PSM model, while Kimoi in Bukembe location, though chosen by the PSM model, was not included in the actual sample. In the treatment area, Netima in Siboti and Kibachenje in S. Bukusu were not included in the actual sample, though they were in the PSM match, and Lunakwe and Mateka in S. Bukusu were included though they were outside the PSM match. These discrepancies and substitutions are summarized in the table below.

Table 18: Village Sample Inclusion Summary

Village	Treatment Status	PSM Model	Actual Sample
Khakula	Comparison	Not Included	Included
Kimoi	Comparison	Included	Not Included
Netima	Treatment	Included	Not Included
Kibachenje	Treatment	Included	Not Included
Lunakwe	Treatment	Not Included	Included
Mateka	Treatment	Not Included	Included

In addition to these substitutions, it was found after data collection that several of the geocodes provided by Vi for the chosen villages were inaccurate. In consequence, the enumerators ended up collecting data in areas some distance from the desired sampling zone. The discrepancies between the original geocodes and the clusters of geocodes taken by enumerators at the time of data collection are depicted in Figure 2.

Despite these discrepancies in location, when balance was tested across the village-level covariates at the actual village locations for the Sirisia/Malakisi and Bumula zones, the balance remained acceptable, as shown below in Table 5. The only variable for which the difference between treatment and comparison villages was statistically significant was elevation, which as is shown in the following sections is balanced out by the inclusion of higher-altitude treatment villages in the Kimilili/Ndvisi zone.

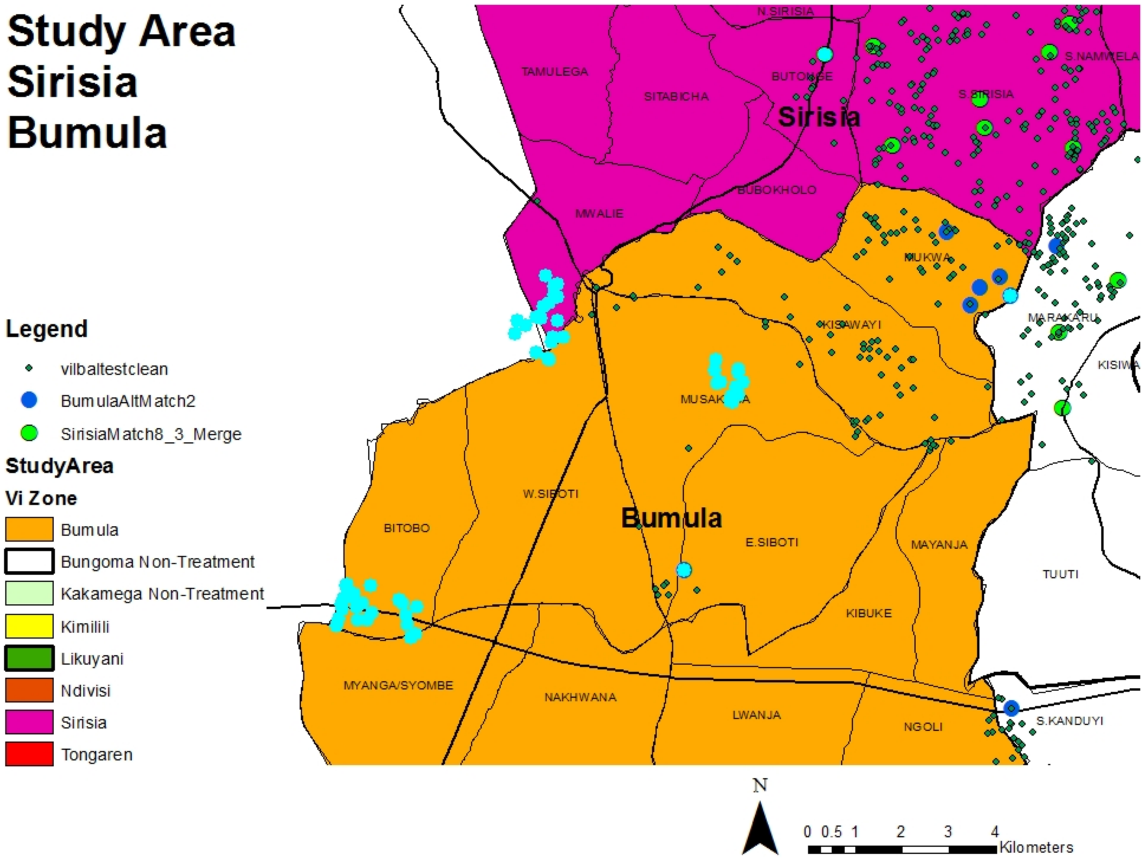
Table 19: Balance Statistics for Surveyed Villages

pop10	0.158	(0.60)
pop15	0.198	(0.60)
soc07	-2.383	(-1.57)
sand	3.012	(1.61)
ph	0.0348	(1.64)
elev	-50.10^{*}	(-2.15)
avgrainbun	-1.923	(-1.07)
tree05	-0.378	(-0.57)
dist_tarma	0.000751	(0.17)
on_road	0.00285	(0.07)
N	65	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 6: Village Geocode Discrepancies



A.3 Kimilili/Ndvisi

The Kimilili/Ndivisi zone lies on the opposite side of Bungoma Central from the Sirisia/Malakisi and Bumula zones. This means that some of the centrally located sublocations which were matched with Sirisia/Malakisi or Bumula were also matched with Kimilili/Ndvisi. So the villages already chosen during the matching exercise in the earlier zones were discarded before completing the matching exercise in this zone.

Kimilili and Ndvisi are distinct from the previous zones in Vi's organizational structure. While Sirisia/Malakisi and Bumula fall under the Kenya Agricultural Carbon Project (KACP)—a carbon credit scheme which provides token payments to farmers in addition to extension services—Kimilili and Ndivisi are managed under the Farmer Organizations and Agroforestry (FOA) project, which provides similar services but does not receive carbon credit funding. It should also be noted that in Kimilili and Ndvisi two SACCOs have taken over the direct implementation of Vi's training program with farmer groups. Vi still provides funding and support, but the SACCOs' staff provide the services directly. This challenged the scoping process because the records for groups and their locations were less well-kept than in the more data-rich KACP project. The consultants were forced to collect most of the information on the location of the groups in these zones themselves, which took additional time. However, this allowed the research team to have more confidence in the location of the chosen villages since the geocodes were taken more recently than in the KACP areas.

In seeking to assemble a balanced sample in the Kimilili/Ndivisi zone, elevation and rainfall were particular challenges. In order to get within as close a range as possible, villages outside of a designated rainfall range were dropped from the sample. Then the PSM model was estimated, returning a sample whose balance statistics are depicted below.

Table 20: Balance Statistics for Kimilili/Ndvisi Zone

households	-32.87	(-0.88)
sand	0.494	(0.35)
ph	0.00600	(0.31)
tree05	0.500	(0.68)
elev	79.91***	(4.77)
pop10	-0.129	(-0.38)
soc	-0.272	(-0.17)
AvgRainBun	-8.619***	(-6.95)
dist_tarma	-0.000143	(-0.02)
on_road	-0.0667	(-1.00)
N	30	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The difference between treatment and control in terms of rainfall and elevation is significant, but this imbalance becomes statistically insignificant when pooled with the other zones. As noted above, the Bumula zone was imbalanced across elevation due to a high number of low-elevation treatment villages. The introduction of the higher-elevation treatment villages in Kimilili/Ndvisi balanced out this issue. Balance statistics for the pooled samples from Kimilili/Ndvisi, Bumula and Sirisia/Malakisi are shown below.

Note that in this pooled sample rainfall is still statistically significant, but with a smaller difference in means than Kimilili/Ndivisi by itself. This difference in mean rainfall also lessens and becomes statistically insignificant when the final zone, Tongaren/Likuyani is included.

Table 21: Balance Statistics for Kimilili/Ndvisi, Bumula and Sirisia/Malakisi

households	-32.87	(-0.88)
sand	0.494	(0.35)
ph	0.00600	(0.31)
tree05	0.500	(0.68)
elev	79.91***	(4.77)
pop10	-0.129	(-0.38)
soc07	-0.272	(-0.17)
avgrainbun	-8.619***	(-6.95)
dist_tarma	-0.000143	(-0.02)
on_road	-0.0667	(-1.00)
N	30	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.4 Tongaren/Likuyani

The Tongaren and Likuyani zone actually comprises two distinct program areas in two different counties. Tongaren in a sub-county in Bungoma county, while Likuyani is in Kakamega county. Both areas fall under Vi's Farmer Organizations and Agroforestry Project, but they have separate staff implementing the training activities.

In Tongaren, the non-program areas are found in the Kiminini-Bungoma, Kabuyefwe and Mbakalo locations. These alternate in a checker-board fashion with the Vi program areas. In Likuyani, the Vi program area is separated from the non-program area by a national forest. The candidate villages were chosen during the scoping process so that they would be situated symmetrically on both sides of this forest.

Since two counties are represented in this zone, exact matching was implemented within each county so that villages were only matched with villages in their own county. The resulting balance statistics are shown below.

A.5 Discussion

As this appendix should make clear, the application of the village-level spatial matching methodology discussed in this paper was far from logistically straightforward. However, it should also be apparent that this is a methodology with enough in-built flexibility that it is possible to operationalize it in a broad array of contexts. Importantly, instead of gathering a data in a broad array of villages, one researcher with two assistants was able to narrow down the list of candidates using a short key informant survey and satellite data. This relatively inexpensive process resulted in a sample which was well-balanced across geospatial variables, and this balance translated in the end into sample balance across household socio-economic variables to a satisfactory degree. This despite a process constrained by time and labor. The end result tends to validate the methodology both on academic and operational grounds.

B Appendix B: Alternate Specifications

Table 22: Wealth Outcomes with County Fixed Effects

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
OLS	0.07***	0.30*	0.09
	(0.02)	(0.18)	(0.06)
IPWRA	0.07***	0.29*	0.09
	(0.02)	(0.18)	(0.06)
2SLS	0.09***	0.38*	0.11
	(0.03)	(0.22)	(0.07)
Observations	2785	2785	2785

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 23: Wealth Outcomes with County Fixed Effects, Limited Covariates

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
OLS	0.06***	0.29	0.08
	(0.02)	(0.18)	(0.06)
IPWRA	0.06***	0.29*	0.09
	(0.02)	(0.18)	(0.06)
2SLS	0.08***	0.37	0.10
	(0.03)	(0.23)	(0.07)
Observations	2790	2790	2790

Standard errors in parentheses

Covariates include only variables which were correlated with treatment at the 10% level

* p<0.10, ** p<0.05, *** p<0.01

Table 24: Wealth Outcomes with Zone Fixed Effects, Limited Covariates

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
OLS	0.07***	0.30*	0.09
	(0.02)	(0.18)	(0.06)
IPWRA	0.06**	0.27	0.07
	(0.02)	(0.18)	(0.06)
2SLS	0.08***	0.38*	0.11
	(0.03)	(0.23)	(0.07)
Observations	2790	2790	2790

Standard errors in parentheses

Covariates include only variables which were correlated with treatment at the 10% level

* p<0.10, ** p<0.05, *** p<0.01

Table 25: Wealth Outcomes Excluding Villages with Nighttime Lights Pre-Trend Change

	(1) PCA-Weighted Asset Growth	(2) Predicted Household Expenditure	(3) Predicted Per Capita Expenditure
OLS	0.07***	0.28	0.09
	(0.02)	(0.18)	(0.06)
IPWRA	0.06**	0.28	0.08
	(0.03)	(0.18)	(0.06)
2SLS	0.08***	0.36	0.11
	(0.03)	(0.23)	(0.07)
Observations	2737	2737	2737

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 26: PCA-Weighted Assets Using 2007 First Principal Component Weights

	(1)	(2)	(3)
OLS	0.05*		
	(0.03)		
IPWRA		0.05*	
		(0.03)	
2SLS			0.06*
			(0.03)
Observations	2785	2785	2785

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

References

Abadie, B Y Alberto and Guido W Imbens (2016). "Large Sample Properties of Matching Estimators for Average Treatment Effects Author (s): Alberto Abadie and Guido W . Imbens Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/3598929 Accessed : 04-06-2016 09 : 55 UTC Your u". In: 74.1, pp. 235–267.

Alix-Garcia, Jennifer M., Katharine R E Sims, and Patricia Yañez-Pagans (2015). "Only one tree from each seed? Environmental effectiveness and poverty alleviation in Mexico's Payments for Ecosystem Services Program". In: *American Economic Journal: Economic Policy* 7.4, pp. 1–40. ISSN: 1945774X. DOI: 10.1257/pol.20130139.

Andam, Kwaw S, Paul J Ferraro, Alexander Pfaff, et al. (2008). "Measuring the effectiveness of protected area networks in reducing deforestation." In: *Proceedings of the National Academy of Sciences of the United States of America* 105.42, pp. 16089–16094. ISSN: 0027-8424. DOI: 10.1073/pnas.0800437105.

Andam, Kwaw S, Paul J Ferraro, Katharine R E Sims, et al. (2010). "Protected areas reduced poverty in Costa Rica and Thailand." In: *Proceedings of the National Academy of Sciences of the United States of America* 107.22, pp. 9996–10001. ISSN: 0027-8424. DOI: 10.1073/pnas.0914177107.

Angrist, Joshua D. and Guido W. Imbens (1995). "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensit". In: *Journal of the American Statistical Association* 90.430, pp. 431–442. ISSN: 01621459. DOI: 10.2307/2291054.

Araujo, Ademir Sérgio Ferreira et al. (2012). "Microbiological process in agroforestry systems. A review". In: *Agronomy for Sustainable Development* 32.1, pp. 215–226. ISSN: 1774-0746. DOI: 10.1007/s13593-011-0026-0.

Barret, Sam (2015). "Subnational Adaptation Finance Allocation: Comparing Decentralized and Devolved Political Institutions in Kenya". In: *Global Environmental Politics* 15.3, pp. 118–139. ISSN: 1526-3800. DOI: 10.1162/GLEP.

Bryan, Elizabeth et al. (2011). "Agricultural Management for Climate Change Adaptation, Greenhouse Gas Mitigation, and Agricultural Productivity". In: June, p. 52.

Cornell, Agnes and Michelle D'Arcy (2014). "Plus ça change ? County-level politics in Kenya after devolution". In: *Journal of Eastern African Studies* 8.November, pp. 173–191. ISSN: 1753-1055. DOI: 10.1080/17531055.2013.869073.

Current, Dean, Ernst Lutz, and Sara J Scherr (1995). "The Costs and Benefits of Agroforestry to Farmers". In: *The World Bank Research Observer* 10.2, pp. 151–180.

Doll, Christopher N H, Jan Peter Muller, and Jeremy G. Morley (2006). "Mapping regional economic activity from night-time light satellite imagery". In: *Ecological Economics* 57.1, pp. 75–92. ISSN: 09218009. DOI: 10.1016/j.ecolecon.2005.03.007.

Douthwaite, B. et al. (2002). "The adoption of alley farming and Mucuna: Lessons for research, development and extension". In: *Agroforestry Systems* 56.3, pp. 193–202. ISSN: 01674366. DOI: 10.1023/A:1021319028117.

Filmer, Deon and Lant H Pritchett (2001). "Estimating Wealth Effects Without Expenditure Data fffdfffdfffd or Tears: An Application to Educational Enrollments in States of India". In: *Demography* 38.1, pp. 115–132. ISSN: 1533-7790. DOI: 10.1353/dem.2001.0003.

Funk, Chris et al. (2015). "The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes". In: *Scientific Data* 2. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.66.

Haglund, Eric et al. (2011). "Dry land tree management for improved household livelihoods: Farmer managed natural regeneration in Niger". In: *Journal of Environmental Management* 92.7, pp. 1696–1705. ISSN: 03014797. DOI: 10.1016/j.jenvman.2011.01.027.

Hegde, Ravi and Gary Q. Bull (2011). "Performance of an agro-forestry based Payments-for-Environmental-Services project in Mozambique: A household level analysis". In: *Ecological Economics* 71.1, pp. 122–130. ISSN: 09218009. DOI: 10.1016/j.ecolecon.2011.08.014.

Honey-Rosés, Jordi, Kathy Baylis, and M. Isabel Ramírez (2011). "A Spatially Explicit Estimate of Avoided Forest Loss". In: *Conservation Biology* 25.5, pp. 1032–1043. ISSN: 08888892. DOI: 10.1111/j.1523-1739.2011.01729.x.

Hughes, Karl et al. (2017). "Assessing the Downstream Socioeconomic and Land Health Impacts of Agroforestry Promotion in Western Kenya".

Jarvis, A et al. (2008). *Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database*.

Jätzold, Ralph and Horst Kutsch (1982). "Agro-Ecological Zones of the Tropics , with a Sample from Kenya". In: *Journal of Agriculture and Rural Development in the Tropics and Subtropics* April 1982, pp. 15–32.

Joppa, Lucas and Alexander Pfaff (2010). "Reassessing the forest impacts of protection: The challenge of nonrandom location and a corrective method". In: *Annals of the New York Academy of Sciences* 1185, pp. 135–149. ISSN: 00778923. DOI: 10.1111/j.1749-6632.2009.05162.x.

Keele, Luke J. and Rocío Titiunik (2015). "Geographic boundaries as regression discontinuities". In: *Political Analysis* 23.1, pp. 127–155. ISSN: 14764989. DOI: 10.1093/pan/mpu014.

Kreif, Noémi et al. (2013). "Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation". In: *Health Services and Outcomes Research Methodology* 13.2-4, pp. 174–202. ISSN: 13873741. DOI: 10.1007/s10742-013-0109-2.

Kruska, Russell and Patrick Chege Kariuki (2016). *ILRI - GIS Services Digital Elevation Model*.

LePage Morgan, Claire (2017). *From the Lab to the Farm: Using Process-Tracing Methods to Evaluate Impact Pathways in Agroforestry Research*. Tech. rep.

Lorenz, Klaus and Rattan Lal (2014). "Soil organic carbon sequestration in agroforestry systems. A review". In: *Agronomy for Sustainable Development* 34.2, pp. 443–454. ISSN: 17730155. DOI: 10.1007/s13593-014-0212-y.

Magruder, Jeremy R (2010). "High Unemployment Yet Few Small Firms : The Role of South African Labor Regulations". In: 4.March, pp. 1–45.

Moser, Caroline and Andrew Felton (2007). "The Construction of an Asset Index Measuring Asset Accumulation in Ecuador".

Nair, P. K. Ramachandran (1993). *An Introduction to Agroforestry*. ISBN: 0792321340. DOI: 10.1016/0378-1127(95)90008-X.

Otsuki, Tsunehiro (2010). "Estimating Agroforestry's Effect on Productivity in Kenya : An Application of a Treatment Effects Model". In: *Africa* C, pp. 2007–2009.

Place, Frank et al. (2003). "The Impact of Agroforestry-Based Soil Fertility Replenishment Practices on the Poor in Western Kenya".

Robalino, Juan, Alexander Pfaff, and Laura Villalobos (2015). "Deforestation Spillovers from Costa Rican Protected Areas". In: *Documentos de trabajo U. Costa Rica*.

Rosenbaum, B Y Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1083, pp. 41–55.

Sahn, David E. and David Stifel (2003). "Exploring alternative measures of welfare in the absence of expenditure data". In: *Review of Income and Wealth* 49.4, pp. 463–489. ISSN: 00346586. DOI: 10.1111/j.0034-6586.2003.00100.x.

Sexton, J.O. et al. (2013). *Landsat Tree Cover Continuous Fields*. DOI: 10.1080/17538947.2013.786146.

Sjögren, Hans, Keith D. Shepherd, and Anders Karlsson (2010). "Effects of improved fallow with Sesbania sesban on maize productivity and Striga hermonthica infestation in Western Kenya". In: *Journal of Forestry Research* 21.3, pp. 379–386. ISSN: 1007-662X. DOI: 10.1007/s11676-010-0085-0.

StataCorp (2013). *Stata 13 Base Reference Manual*. College Station, TX: Stata Press.

Stevens, Forrest R. et al. (2015). "Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data". In: *PLoS ONE* 10.2, pp. 1–22. ISSN: 19326203. DOI: 10.1371/journal.pone.0107042.

Stuart, Elizabeth a (2010). "Matching methods for causal inference: A review and a look forward". In: *Stat Sci* 25.1, pp. 1–21. DOI: 10.1214/09-STS313.Matching.

Tsonkova, Penka et al. (2014). "Ecosystem services assessment tool for agroforestry (ESAT-A): An approach to assess selected ecosystem services provided by alley cropping systems". In: *Ecological Indicators* 45, pp. 285–299. ISSN: 1470160X. DOI: 10.1016/j.ecolind.2014.04.024.

Vågen, Tor-G. et al. (2016). "Mapping of soil properties and land degradation risk in Africa using MODIS reflectance". In: *Geoderma* 263, pp. 216–225. ISSN: 0016-7061. DOI: http://dx.doi.org/10.1016/j.geoderma.2015.06.023.

Wachiye, Emmanuel (2008). *AF AND SALMS Presentation - in use since 2008*.