Yield gap analysis and variability of coffee yields in Mount Elgon, Uganda as affected by shade regime

Ivan Pontin Novotny

MSc thesis

Farming Systems Ecology Group

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Yield gap analysis and variability of coffee yields in Mount Elgon, Uganda as affected by shade regime

Ivan Pontin Novotny
Registration number 870909-610-220
Credits: 36
Code number/ name course: FSE80436
Period: May 2013 -May 2014
Supervisor: Johannes Scholberg
Professor/Examiner: Pablo Tittonell

i. Abstract

Coffee is an important commodity that is grown all around the world. In Mount Elgon, Uganda, *Coffea arabica* is the main source of income for farmers. The variation in climatic conditions, soil type, genotype and crop management all contribute to the inherently yield variability among and within regions, with Mount Elgon being no exception to that. Additionally, there is a lot of discussion in terms of what is the best shade regime condition for coffee. The current study aimed to unravel the drivers governing yield variability and to assess how shade affects production factors and to use this information to structure possible solutions to decrease the coffee yield gaps in Mount Elgon, Uganda. A total of 24 farms were selected during this study, and 345 coffee plants were chosen during a single production cycle of the coffee trees. The plants were selected based on their position as related to existing shade tree. Hemispherical pictures were also taken in order to determine shade density. Based on the shade density, the assessed plants could be classed into three groups being: low (31 – 50 %), medium (51 – 70 %) and high (71 – 90 %) shade density. Yield components, crop phytosanitary status, soil characteristics, plant nutrition, and crop management parameters were obtained from field measurements and interviews with farmers (e.g. management). A descriptive statistics analysis was then performed, to compare differences across the three groups. A boundary line analysis was conducted to determine the attainable yield of each group. Regression trees were generated (using the entire dataset) for a quick visualization of the underlying causes governing yield variability. Linear mixed effect models were also used to evaluate the effect of shade on coffee yield. Via use of regression trees and linear mixed effect models, groups with unique characteristics could be identified. These groups were then further explored using a mean comparison of selected variables. The group with medium shade density levels showed a higher attainable yield, followed by high and low shade densities. On average, low shade had also lower yields, the highest incidence of coffee leaf rust while branches were also affected by dieback. The regression tree analysis showed that leaf area was the main cause of yield variability while the number of stems and shade were the main drivers of leaf area variability. From the mixed effect model it may be concluded that there was no indication that shade affects coffee yield, although there was a tendency for leaf area to increase with shade density. Even though shade had no effect on yield for the dataset as whole, shade did show either positive or negative effect on yield, depending on the farm. The group of farms with positive response of yield to shade density tended to be less productive compared to plants belonging to farms that showed a negative response to shade. Based on results from the current study, it is concluded that yield limitations and/or yield gaps in the Mt Elgon Region are to a large extend related to phytosanitary and plant management aspects.

Table of Contents

1. Introduction

Globally, coffee is the most widely traded tropical agricultural commodity (International Coffee Organization n.d.), while in Uganda, it is the main export and cash crop (Food and Agriculture Organization of the United Nations 2014). Although green coffee only occupies the 18th position, based on cropped area in Uganda (Food and Agriculture Organization of the United Nations 2014), it contributes to 20 to 30% of total national foreign exchange earnings (Uganda Coffee Development Authority 2014). In Uganda, *Coffea Arabica* L. and *Coffea robusta* L. account for 20 and 80% of the planted area, respectively. Even though Robusta coffee occupies a much larger production area, the Arabica coffee generates, approximately 40% of the coffee export value in the country (Jassogne, Läderach et al. 2013). Coffee in Uganda is almost entirely produced by smallholder farmers with farm holdings ranging from less than 0,5 to 2.5 ha (Uganda Coffee Development Authority 2014).

There is a great variability in the occurring management practices and yield of coffee in Uganda (van Asten, Wairegi et al. 2011) which is related to different production regions featuring variable soil types, weather condition, regional management practices, etc. In spite of regional variability, there is also variability within the same region and /or farm (Tittonell, Vanlauwe et al. 2008). As consequence, yield variability is intrinsic to most production systems; although the degree of variability may change from place to place and from farm to farm, and may be partly reduced via enhanced management.

Understanding the main sources of yield variability is of great importance. In theory, every crop has a potential yield, which is governed by genetics and climatic conditions and this determines how much a plant can produce when there is neither restriction of yield limiting (e.g. water and nutrients) and reducing factors (e.g. pests, weeds and diseases). More specifically, potential yield is affected by the genetic material and environmental factors such as latitude, altitude and cloud cover (Tittonell, Shepherd et al. 2008). Potential yields are therefore theoretical target values, which in reality are never attained under standard farming conditions. As consequence, when studying yield variability, attainable yield is a more relevant concept and a reference value. Following the train of thought proposed by Tittonell and Giller (2013), attainable yield in Africa is defined by comparing actual yield with the maximum local yield (Y_L) that can be found in a certain region. This is the yield that may be achieved by the use of locally available resources and it reflects crop performance in the most productive fields. Thus, the yield gap concept can be introduced and calculated as the difference between Y_L and the actual farmers yield levels (Y_A). This concept of yield gap was applied for food crops in Africa, but it can also be easily used for coffee production as well.

In the context of coffee, there is a lot of discussion pertaining to optimal shade regimes, especially in the context of sustainable coffee production and/or agroforestry based production systems (Wintgens 2003, DaMatta 2004, Aerts, Hundera et al. 2011). There is no agreement on what the optimum shade regime may be for coffee, for shade effects changes according the regional condition (e.g. altitude, temperature, soil fertility and photosynthetically active radiation) and the coffee variety (Estívariz, 1997 as quoted by Ricci, Costa et al. 2006). As a result, there are no universal recommendations and there is a lot of controversy regarding the matter (Beer, Muschler et al. 1998). This lack of understanding on the effect of shade on coffee yield is of great importance when trying to understand yield gap and yield variability in coffee systems.

A number of studies identified and explained the yield variability of different crops in specific areas (Fermont, van Asten et al. 2009, Wairegi, van Asten et al. 2010, Delmotte, Tittonell et al. 2011). However, there are virtually no studies exploring the yield variability of coffee and the effect of shade neither in Uganda nor in any other production region. Unraveling the main sources of yield variability, and identifying key practices that may decrease yield gap may provide valuable insights as to how best improve local coffee yields. This study aims to explore the factors that govern yield variability and thus productivity of *C. arabica* in Mount Elgon, which is an important coffee producer region in Uganda. The main research question of this study centers on how shade density and other factors contribute to yield variability and yield gap of coffee.

2. Materials and Methods

2.1. Study Area

Mount Elgon is located is situated in both Uganda and Kenya. The predominant coffee species in the area is *C. Arabica*. The region comprises a total of 1,121km^2 and its altitude reaches 4,321 masl (Uganda Wildlife Authority 2014). For this particular study, located in the Sironko district, the altitude, longitude and latitude ranged between 1,103 and 1,394 masl, 34°18'E and 034°31'E, and 01°14'N and 01°17'N, respectively. It should be noted that this combination of altitude and latitude is not the most suitable for the cultivation of *C. Arabica*. The climate in the region is dominated by seasonally alternating moist south-westerly and dry north-easterly air streams (Atlas of Uganda, 1967 as quoted by Hamilton and Perrott 1981). Annual rainfall in the region amounts to approximately 1200mm and it varies in function of the altitude, while the mean minimum and maximum temperature in Mount Elgon are 15 and 28°C, respectively (Bamutaze, Tenywa et al. 2010). The prevailing soil types in Mount Elgon are sandy clay loams soils (van Asten, Wairegi et al. 2011) and the parent rocks mainly consisting of Cenozoic volcanic outcrop (Schüter, 2008 as quoted by van Asten, Wairegi et al. 2011).

The human activities in the Mt. Elgon region are predominantly agricultural related, with most smallholder farmers producing subsistence and/or cash crops. The main staple crops include maize (*Zea mays* L.) and *matoke* (*Musa* sp.). Regarding cash crop, *C. Arabica* is the most important one. Coffee in this region is dominantly produced in association with other crops and/or shade trees, with Nyasaland being the most commonly cultivated coffee variety. Livestock production is mainly part of subsistence farming systems and animals are used for ploughing, and, in some occasions, as saving banks (Norgrove and Hulme 2006).

2.2. Data Collection

The study was conducted between late June and late September of 2013, coinciding with the late flowering, and ripening period of coffee in this region. A total of 24 farms were selected. Since shade trees are commonly present in most coffee farms around Mount Elgon, and these trees have a clear impact on coffee morphology and production, coffee plants exposed to different shade regimes were selected. These plants were carefully measured in order to study the effect of shade regime on coffee yield variability. Out of the 24 studied farms, 60 plots were selected. Each plot contained a shade tree, being either Albizia (*Albizia africana*) or Cordia (*Cordia. Africana*), which are the two dominant shade trees of the studied region. In total, 31 plots were designated to the shade tree Albizia and 29 to Cordia. For most of the plots, 6 coffee plants around each tree were chosen for measurements. Coffee plants close to the shade tree's trunk, on the verge of the shade tree's canopy and in the open field were selected for obtaining different shade density levels (Figure 1). A total of 345 coffee plants were included in this study, during the measurement period. The number of sub-plots and coffee plants assessed per farm was not the same.

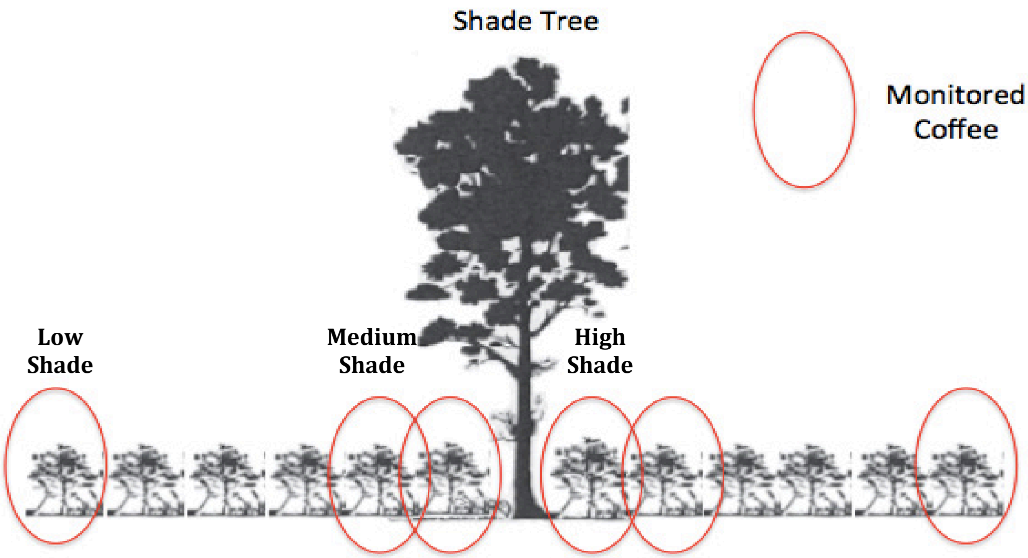

Figure 1: Outline of design used for on-farm selection of the coffee plants used for measurements, according to the shade regime they were exposed.

Individual farm interviews were performed with a family member of each farm, in order to collect information on the coffee age and variety, fertilizer application, commonly occurring pest and disease control, and pruning practices and stumping management. Moreover, yield-related information (e.g. coffee berry yield and leaf area index), disease infestation and field characteristics (e.g. soil analysis and slope) were obtained from on-farm measurements.

The yield of coffee was estimated according to the methodology described by Cilas and Descroix (2008). To estimate the yield, first we predicted the average number of fruits per coffee tree by the use of the following formula:

Average number of fruits = number of stems per coffee plant X number of fruit-bearing branch per stem X number of glomerules per fruit-bearing branch X number of cherries per glomerule

Then, to estimate the yield, we simply multiplied the average number of fruits by the weight of one cherry. The weight of one cherry was calculated by averaging the weight of 100 cherries of each measured coffee tree.

The leaf area per plant (m^2) was estimated using a non-destructive method developed by Barros (as quoted by Martinez, Bragança et al. 2005). In summary, plants were visually divided in four sections and one branch was chosen for each of those sections. From those branches, one leaf from the mid-part was then selected and used to measure the leaf width and length. To calculate the leaf area, the following formula was applied:

Leaf area = width X length / 0.677

The estimated leaf area per leaf was then multiplied by the total amount of leaves from that branch we selected the leaf for measuring. Then the averaged values were taken for the four selected branches; thereby defining the average leaf area per branch for the entire coffee tree. The total leaf area per plant was calculated by multiplying the average leaf area per branch by the total number of branches.

Foliar chlorophyll content was measured using a chlorophyllmeter (SPAD 502, Konica-Minolta Inc., Osaka, Japan). The readings were taken from 15 pair of leaves (3rd or 4th pair of leaf, counting from the tip of the branch), all around the plant, from fruit-bearing branches located in the mid-section of the coffee, as described by Rodrigues Junior, Vieira et al. (2011). Later, the 30 readings were averaged to give the overall SPAD value for each coffee tree.

The height of coffee was measured with the help of a measuring stick, as well as the distance between the shade tree and the coffee plant. The percent of slope of each plot was measured directly with the clinometer SUUNTO PM-5. To measure the shade density, first hemispherical pictures were taken with a Canon EOS 5D camera, and the lens used was Sigma 8mm f/3.5 EX DG Circular Fisheye. The pictures were shot right above the coffee plant. Later, the pictures were analyzed with ImageJ 1.47v for Mac, and the percentage of shade was obtained, based on a calculation of the grey scale.

Soil samples were collected for half of the monitored coffee plant. For each soil sample analyzed, 4 sub-samples were collected around the coffee plant, at a depth of 30cm, and mixed to form a composite sample. The samples were dried in the oven at 40°C for 48 to 96 hours, ground and sieved (<2mm). The soil pH was measured using a 1:2.5 sediment: water extract. Soil organic carbon was obtained by the Walkley-Black procedure (Nelson and Sommers 1996). Total N was determined using the Kjeldahl digestion with sulphuric acid and selenium as a catalyst and measured using a spectrophotometer. Available P and exchangeable cations (Ca, Mg and K) by the Mehlich-3 extraction solution (Mehlich 1984). K was measured using a flame photometer, while the other cations were determined using an atomic absorption spectrophotometer.

For the foliar analysis, 4 plagiotropic branches, located halfway up the tree, and oriented in opposite directions, were selected, during the fruit maturation period. From each branch, the 3rd and 4th pair of leaves were picked for the analysis, as proposed by Snoeck and Lambot (2004) (Figure 2). The samples were then oven dried at 72°C for 47 to 96 hours, then ground to less than 2mm particle size, and later digested in a sulfuric and selenium acid mixture. Regarding Zn and Fe analysis, tissue samples were dry-ashed at 500°C for 2 hours, the ash was diluted in dil. aqua regia solution. The samples were then calorimetrically analyzed for N and P; while K, Ca, Mg, Zn and Fe were analyzed using an atomic absorption spectrophotometer. The used analytical methods are described by Okalebo, Gathua et al. (1993).

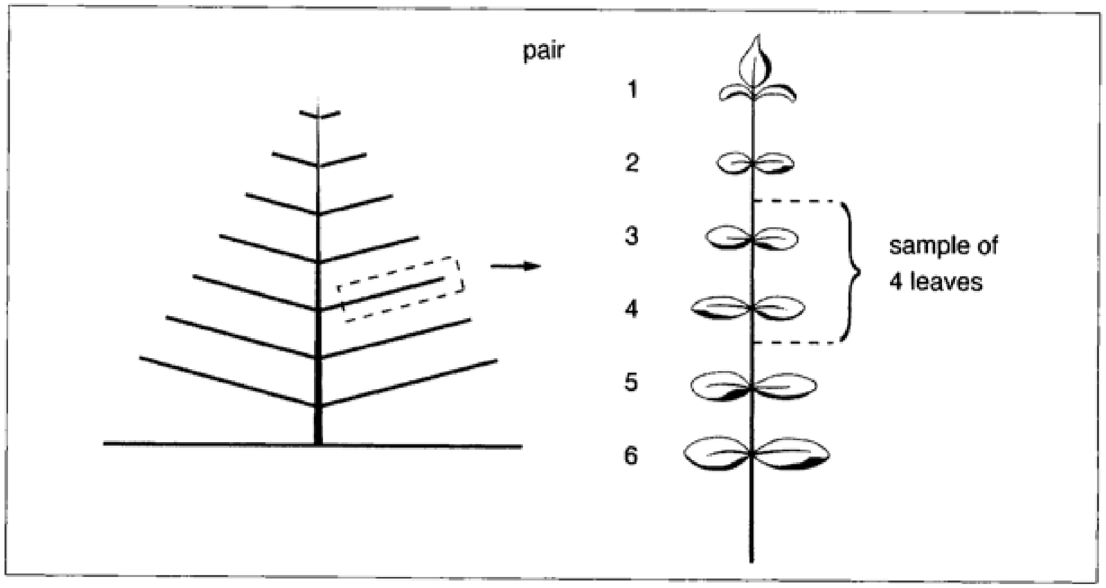

Figure 2: Leaf sampling scheme. The 4 leaves sampled are indicated in the 3rd and 4th pair of leaves, from tip to main stem direction (Wintgens 2003).

The percentage of leaves infected by coffee leaf rust (CLR) was obtained by visually separating the coffee plant in three sections, lower, mid and upper section. From each section, one branch was picked at random for two sides of the plant and a total of six branches were selected per plant. Later, the number of leaves infected by CLR was counted and then divided by the total number of leaves for that branch. Subsequently, values for the two branches per section were averaged, and finally the average of each section was summed together and divided by three to give the average percentage of leaves infected by CLR.

The number of branches afflicted by dieback was counted on two opposite sides of the coffee plant. Later, the number of twigs affected by dieback on each side was summed together and multiplied by two, in order to obtain an estimation of twigs affected by dieback on all the cardinal points of the plant. Finally, that number was divided by total number of branches of the plant to obtain the percentage of branches disturbed by dieback. The equation is presented below:

*Dieback (% of affected branches) = (number of affected branches (south oriented) + number of affected branches (north oriented)) * 2 / total number of branches*

2.3. Data Analysis

2.3.1. Correlation Analysis

To verify the association between yield and all the other independent variables, a correlation test was performed. Given the non normal distribution of the data and the presence of outliers (Chok 2010), the Kendall's correlation test was performed on R 3.0.2 GUI 1.62 for Mac, using the package "stats" (R Core Team 2013).

2.3.2. Boundary Line Analysis

Boundary line analysis consists in identifying local maximums (upper boundary points) in a bi-dimensional scatter plots, and then fitting a boundary line through those points. The upper boundary points are the highest observed response of a dependent variable for a specific level or range of an independent variable (e.g. highest yield found for a certain pH value). In other words, the boundary line, in theory, expresses the response of the plant to a certain independent factor if no other variable is limiting the crop production. (van Asten, Wopereis et al. 2003)

To perform the boundary line analysis, the whole dataset was subdivided into three groups, based on their shade density, which were low (31 – 50 %), medium (51 – 70 %) and high (71 - 90 %) shade levels. Yield was placed as the response variable for all the shade regimes brackets. Explanatory variables included: SPAD readings, distance to the tree, shade density, number of stems, pH, organic matter in the soil, P content in the soil, CLR frequency, plant density, coffee age, slope, die-back incidence, and application of manure (in the form of farm yard manure), NPK (17-17-17) and CAN (calcium ammonium nitrate) as fertilizer.

The yield was plotted in separate scatter plots, one for each of the explanatory variables, and corresponding lines were drawn following the steps for boundary lines analysis, as described below (adated from Shatar and McBratney 2004):

1st: plotting a certain independent against yield for a specific light regime,

2nd: removal of outliers,

3rd: identification of the upper points by conditional restriction in excel

4th: fitting of a curve through the upper points.

Since many models can be used to fit a curve, many models were used and compared with one another to find the best fit. The comparison was done by calculating the root mean squared error (RMSE) of the models, with the lowest value being the best for fitting (Fermont, van Asten et al. 2009). The models that were used for comparison included: linear, Weibull, log-logistic, Brain-Cousens, Michaelis-Menten and Asymptotic regression models. All the plots and boundary lines were done on R 3.0.2 GUI 1.62 for Mac, using the packages "drc" (Ritz and Streibig 2005) and "lattice" (Sarkar 2008)

2.3.3. Regression Tree

Regression trees were used to predict or explain the response of a continuous target variable (e.g. yield) from a group of explanatory variables that may be either continuous or discrete. As described by Lewis (2000), such method uses binary recursive partitioning rules, where "binary" means that each group (node) is split in two groups (child nodes); "recursive" means that the splitting rules can be applied over and over again for each node; "portioning" sampling means that each group can be split based on a threshold in categorical or continuous variables. The final outcome of regression trees is the terminal nodes (clusters).

According to Tittonell et al. (2008), regression trees have some advantages over more conventional statistical methods. Those advantages are: "(i) there are no assumptions regarding the statistical distribution for either the dependent or independent variables; (ii) a mixture of categorical and continuous explanatory variables may be used; (iii) the analysis is not sensitive to outliers, multi-colinearity, heteroskedasticity, or distributional error structures that affect parametric methods; (iv) it has ability to reveal variable interactions". An outline of clusters of different types of variables included in the analysis and corresponding indicators is provided in Table 1.

Table 1
Variable entries for the regression tree analysis.

Target value or cluster type	Description
Target parameter used	Yield estimation of red cherries (kg/plant)
	Leaf area (m^2/m^2)
	CLR (%)
	Dieback affected branches (%)
Plant health status	Foliar N (%)
	Foliar P (%)
	Foliar K (%)
	Foliar Mg (%)
	CLR incidence on leaves (%)
	Dieback incidence on branches (%)
	Coffee age (years)
	SPAD reading
	Coffee height (m)
Soil condition and landscape	Clay content in the soil (%)
	Soil organic carbon (%)
	pH, water suspension test
	Available P, Mehlich-3 extracting solution (ppm)
	Exchangeable K (ppm)
	Exchangeable Ca (ppm)
	Exchangeable Mg (ppm)
	Slope (%)
Management	NPK application (kg/year/plant)
	Manure application (kg/year/plant)
	Frequency of pruning (year)
	Shade density (%)

Number of stems
Plant density (ha)

The regression trees generated for this experiment were obtained in an iteratively way. First, the entire data set was analyzed, including yield as the target variable. Later, leaf area appeared as an important variable to explain overall yield variability. Therefore, in a second run of the regression tree, yield was replaced by leaf area as the target variable. After identifying CLR and dieback as important yield limiting indicators, a third and fourth regression trees were later obtained, placing CLR and dieback as target variables.

2.3.4. Linear Mixed Effect Model

The mixed model approach was chosen since it is best suited for the nature of the data (Winter 2014), where there are random missing values from a single studied object (e.g. missing pH values due to equipment failure). Linear regression models cannot deal with this type of missing values problems. Furthermore, mixed model can handle the unevenness of repeated measurement (e.g. different conditions for coffee plants found in different farms). More details on mixed effect models may be obtained elsewhere (Winter 2014)

The mixed effect model was used to predict coffee yields and leaf area for different levels of shade, number of stems and plant density. Therefore, the explanatory variables were placed as fixed effect, while farm (24 different farms were studied) was considered as random effect. Following the approach explained by Zuur (2009), the subsequent models were generated for each dependent variable: (i) the random intercept model; (ii) the random intercept and slope model; and (iii) the random effect model . To compare the different models for each dependent variable, the smallest AIC (Akaike Information Criterion) was used to identify which of the three models was more adequate. The *P* values were obtained by comparing two models in an ANOVA table, one with the effect in question (shade density) and other without it, this method is called Likelihood Ratio Test, as explained by Winter (2014). Following the selection of the model, a visual examination of the normal distribution and homoscedasticity of the residuals was performed. Homoscedasticity assumes that the variability of the data is approximately the same across the predicted values. A violation in the homoscedasticity can be translated to unequal variances, which should be corrected by transforming the data (e.g. log-transformation) (Bates, Maechler et al. 2013, Winter 2014). The AICs were obtained from the model output in R 3.0.2 GUI 1.62, using the “lme4” package (Bates, Maechler et al. 2013).

2.4. Combining Different Outcomes

To investigate, understand and compare the most distinct groups obtained from the terminal nodes of the regression trees and the outcome from the linear mixed effect model, key explanatory variables were plotted in a Spider Chart for a better visual presentation of the data. All the variables used in this analysis are listed in Table 3.

2.4.1. Principal Component Analysis

A Principal Component Analysis (PCA) was conducted using R 3.0.2 GUI 1.62 for Mac, using the "stats" package (R Core Team 2013); however, this analysis proved to only produce highly scattered points, with principal components presenting low proportion of variance. Therefore it was dropped out from the analysis, nevertheless, an outcome from the PCA can be found in the Appendix 2.

3. Results

3.1. Yield Variability

The first step of this analysis was to understand intrinsic yield variability of coffee. To do so, we first compared all the farms evaluated in this study. The yield/plant showed substantial variability for both among and within farms (Figure 3). Coffee yields ranged from 0.12 to 6.08, 0.15 to 11.48, and 0.10 to 10.30 kg of red cherry per plant for the low, medium and high shade regimes, respectively (Table 2). The average yield for plants belonging to the medium (3.47 kg of red cherries) and high (3.10 kg of red cherries) shade regime was significantly higher ($P \leq 0.05$) than the ones belonging to the low shade cluster (2.20 kg of red cherries). Moreover, the low shade group had the highest ($P \leq 0.05$) incidence of coffee leaf rust on leaves (0.44 %), followed by the medium shade group (0.32 %) and lastly by the high shade group (0.22 %). The low shade group also showed the highest ($P \leq 0.05$) incidence of dieback (0.14 %), while values for the medium (0.037 %) and high (0.0 %) shade groups were similar. There were no significant differences among the three groups for all other assessed variables (Table 2).

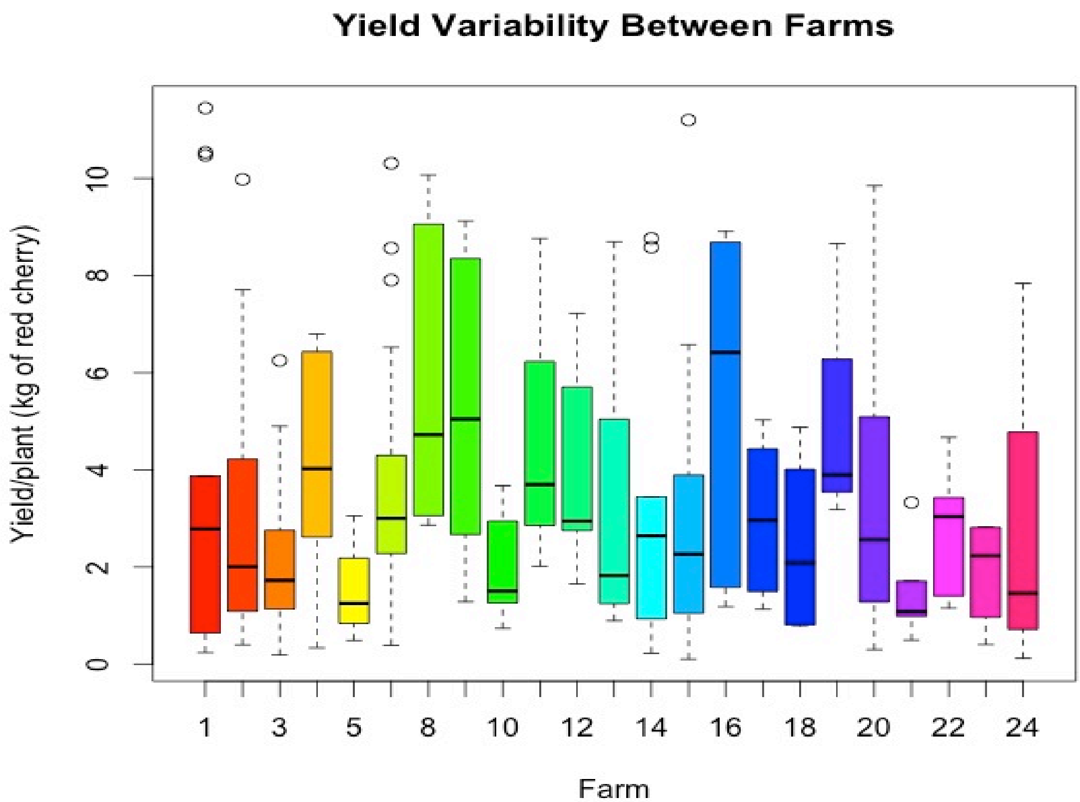

Figure 3: Yield variability (kg of red cherry) within and across all monitored farms.

Table 2
Descriptive statistics of yield, plant management, phytosanitary crop status and soil indicators of three shade regimes: low (31 – 50 %), medium (51 – 70 %) and high (71 - 90 %) shade.

Variables	Low Shade (*n*=84)			Medium Shade (*n*=164)			High Shade (*n*=97)		
	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean
Yield (kg of red cherries per plant)	0.12	6.08	2.20b	0.15	11.48	3.47a	0.10	10.30	3.10a
Slope (%)	0	11.5	4.19	0	14	4.59	0	14	4.98
Coffee Age (year)	5	60	24.57	5	60	26.87	9	53	24.66
Distance Between Coffee and Shade Tree (m)	2.05	20.20	11.04	0.7	26.55	6.79	0.8	10	3.08
Plant Density (plants/ha)	1460	5376	2845	1460	5376	2773	1460	5376	2575
pH	5.90	7.70	6.46	5.20	7.70	6.53	5.20	7.05	6.44
Soil Organic Matter (%)	2.96	6.10	4.28	2.89	6.73	4.39	2.22	6.60	4.53
Soil N (%)	0.16	0.28	0.21	0.16	0.29	0.21	0.13	0.27	0.22
Available P (ppm)	3.5	135.46	46.05	3.39	142.63	47.65	3.39	141.11	41.87
Sum of Bases (ppm)	5.08	7.41	6.43	5.02	8.04	6.52	5.16	7.55	6.54
Coffee Leaf Rust Incidence on Leaves (%)	0	1	0.44a	0	0.9	0.32b	0	0.89	0.22c
Dieback (% of affected branches)	0	0.84	0.14a	0	0.64	0.037b	0	0	0b

Pest and Disease Control (application/year)	0	3	2	0	3	1	0	3	1
Pruning (frequency/year)	1	5	2	0	6	2	0	5	2.26
Manure Application (kg/year)	0	9	3.2a	0	9	2.5b	0	9	4.1a
NPK (17-17-17) application (kg/year)	0	0.6	0.2	0	0.6	0.2	0	0.6	0.2

Different letters within a row denote significant difference between means ($P \leq 0.005$).

Yield was negatively correlated with slope and distance between shade tree and coffee ($P \leq 0.05$) for the intermediate shade group, and also with plant density and coffee leaf rust for the high shade group (Appendix 1). For the low shade group, yield was correlated positively with NPK application (Appendix 1).

Based on the boundary line analysis results, attainable yield per plant increased for all three shade regimes when soil organic matter, soil N , available P, sum of bases (Ca + Mg + K), pest and disease control and manure increased (Figures 4F,G,H, I). In contrast, the attainable yield decreased, for all shade regimes, with an increase of slope and/or incidence of coffee leaf rust and dieback (Figures, 4A, J, K). Coffee age had little impact on the attainable yield of every shade group (Figure 4 B).

As soil pH increased, so did the attainable yield of all the groups up to a point when it started to decrease when pH exceeded values around, 6.2, 6.5 and 6.7 for the low, intermediate and high shade groups, respectively (Figure 4 E).

Across all shade regimes, coffee plants did not tolerate to be overly close to the shade tree, while for the low shade regime, attainable yields increased as the distance between the shade trees and coffee plants increased to about 11 m. For the medium and high shade regimes, a distance to the shade three of 2.5 m seems to be optimal, (Figure 4 C).

The variation in attainable yield for the low shade regime appeared to be not strongly governed by plant density. However, the two other shade regimes, having, respectively, a plant density above around 2200 and 1700 plants/ha for the intermediate and high shade groups resulted in decreased attainable yields (Figure 4 D).

Pruning frequency, up to two times a year, enhanced attainable yield for the intermediate shade group, while it decreased when the pruning frequency exceeded this value (Figure 4 M). The effect of pruning with high shade was also positive with a frequency of one pruning a year, being optimal. The frequency of pruning negatively affected the attainable yield of the low shade group.

Application of NPK (17-17-17) had a positive effect on attainable yield for the low and high shade regimes, although overall yields were highest for the high shade regime. For the intermediate shade regime, application of NPK followed a counter intuitive pattern with higher NPK application being associated with lower attainable yield (Figure 4 O).

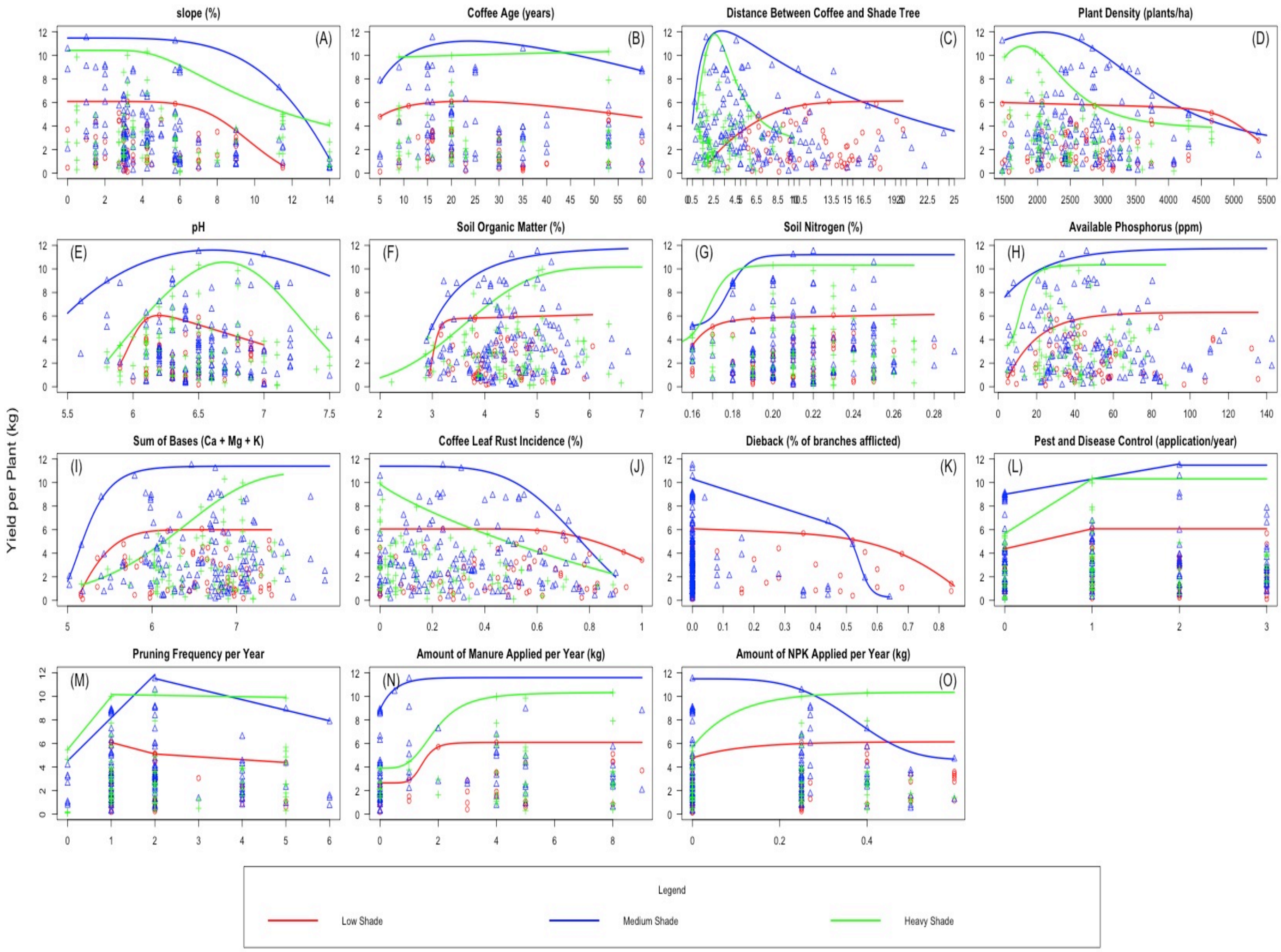

Figure 4: Relationships between coffee yield and yield predictors for low (31 – 50 %), medium (51 – 70 %) and high (71 -90 %) shade densities.

3.2. Grouping the Data

The descriptive analysis presented above showed that there were no major differences among shade regimes for included variables, except for yield, coffee leaf rust infestation and dieback incidence. The boundary line analysis thus provided a tool to better depict and understand the intrinsic complexity of yield variability as related to attainable yield responses to different independent variables evaluated in this study. Unfortunately, the two preceding analysis did not explain what are the main drivers governing yield variability. Therefore, there is a need of defining more homogenous groups of observation in order to explain the yield variability more into depth. For this reason, we generated regression trees as well.

The first regression tree was implemented using the entire data set (n = 318) and by making yield/plant the main target. In this case, coffee foliar area appears right on the top (node 1, Figure 5), as the most important variable that explained a major section of yield variability. In this context, two data clusters emerged based on a threshold of 5.56 m^2/plant. The group with low (<5.56 m^2/plant) leaf area (n = 136), which appears as terminal node (node 2, Figure 5), had an average yield of 2.12 kg of red cherry per plant. The group with high (>5.56 m^2/plant) leaf area (n = 182) had a yield of 4.92 kg. In the proceeding step of this tree, however, slope becomes the next most important variable to explain overall yield variability. This logic is applied several times, where each new node (from a parent node) is split until a terminal node is established, grouping then a set of data with more homogeneous characteristics. The group with the highest yield (13 kg/plant) is represented by node 15. In that group, the slopes were below 3.8 % while plants had a SPAD reading of less than 58 and leaf area exceeded 16 m^2/plant. It is important to notice that shade only appears on the bottom of the tree, and it is only important to explain the yield variability of terminal node 9 and 10. Nutrients also did not appear as major driving variables for explaining yield, and only soil available P appeared in the regression tree, for node 12 (Figure 5).

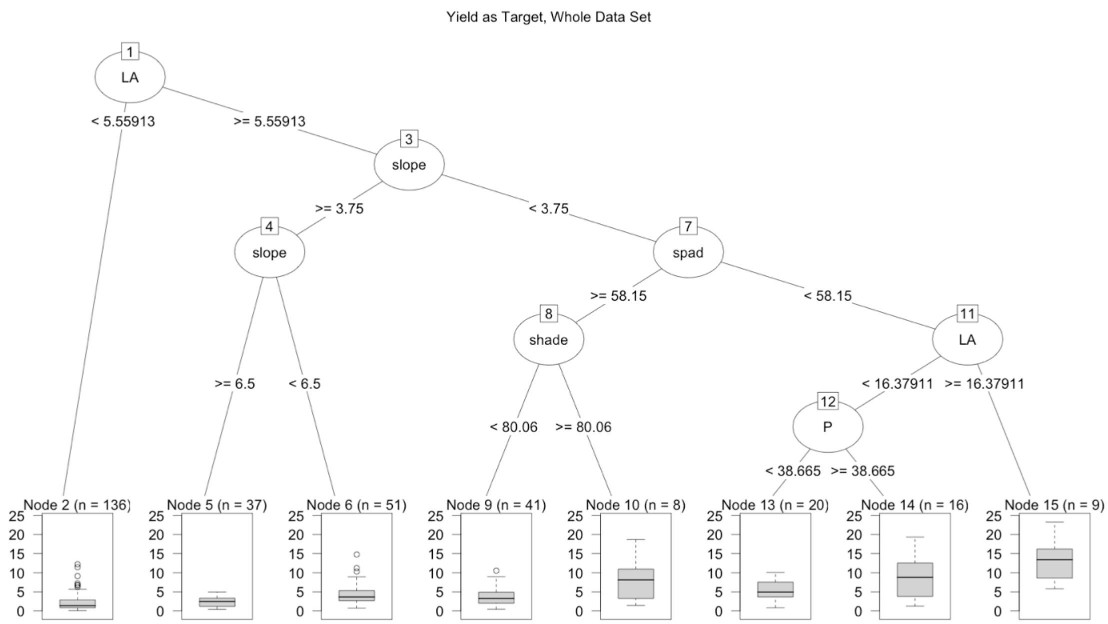

Figure 5: Regression tree describing yield variability (kg/plant) of coffee as a function of different variables including soil, plant and management characteristics. Node 1 is representing the entire population of coffee plants (n=318). The bottom groups represent the terminal nodes, including a graph showing the yield distribution within that group. Every node between the very top group to the terminal nodes is the path taken, and it shows the discerning thresholds, for corresponding variable determining in the terminal nodes.

Since the leaf area was clearly the most important variable to explain intrinsic yield variability, the following step of this analysis was to place that variable as a target, instead of yield (Figure 5). This analysis was performed again for the whole dataset, for which the value for the leaf area of coffee plants was determined (n = 344) (Figure 6). The most important variable to explain the variability of leaf area was the number of stems (Node 1), with a threshold of 5.5 stems/plant. Node 4 had the lowest average leaf area (2.8 m^2/plant) and featured a relatively high presence of dieback (DB, with more than 32 % of the twigs being affected by). Node 7 represented the second lowest leaf area group (on average 4 m^2/plant) with coffee plants having a high incidence of CLR, with 56 % or more leaves affected by that disease. Both nodes 4 and 7 coincided with a shade regime below than 78 % of shade. Node 14 had the highest average leaf area, however, it included only eight observation.

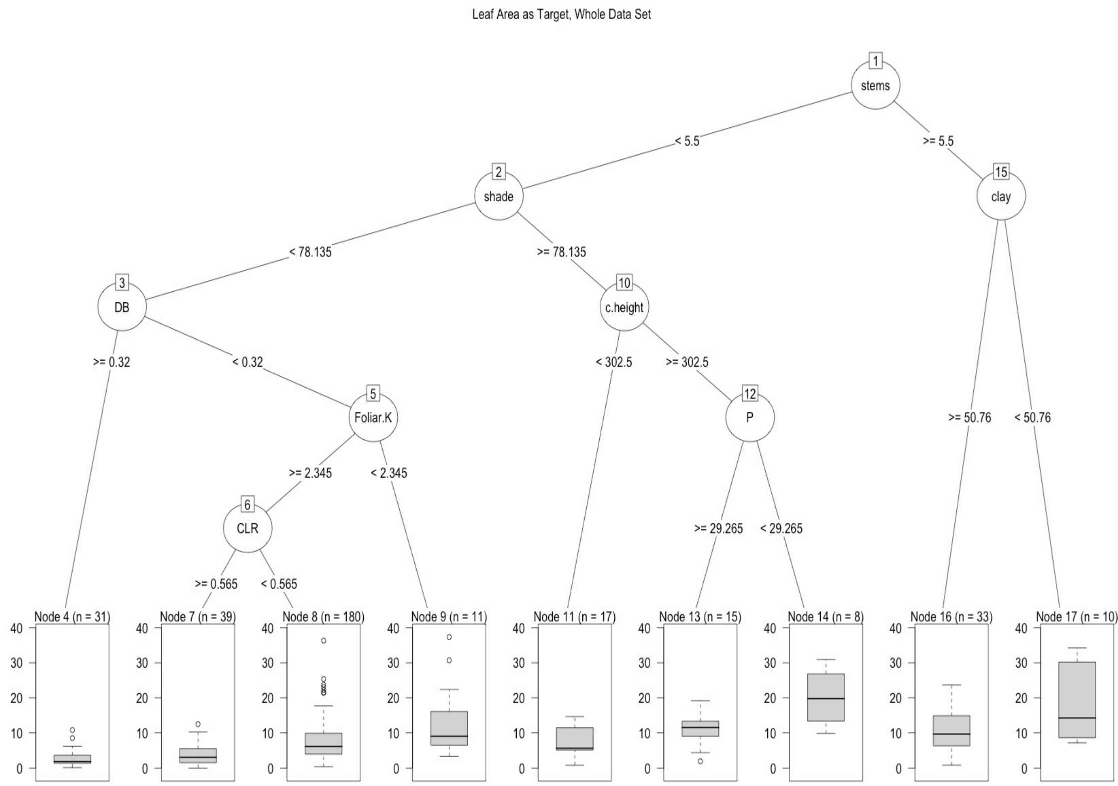

Figure 6: Regression tree describing leaf area (m²/plant) variability of coffee as function of different variables comprising soil, plant and management characteristics. Node 1 is representing the entire population of coffee plants (n=318). The bottom groups represent the terminal nodes, including a graph showing the yield distribution within that group. Every node between the very top group to the terminal nodes is the path taken, and it shows the discerning thresholds, for corresponding variable determining in the terminal nodes.

Because dieback and CLR were probably among the major causes reducing leaf area, these two variables were included as target for the following analyses. The resulting tree for dieback being the target was very small and only had two bifurcations (Figure 7). Node 5 had the highest presence of dieback, with an average of 0.37 % of twigs being affected. In this group, the shade density was relatively low (less than 49 %), while the incidence of CLR was high (more than 0.79).

The following step was to generate the regression tree to outline CLR-based variability (Figure 8). Node 16 had the lowest plant density (less than 2194 plants/ha), while many twigs were affected by dieback (more than 18 %). The average of CLR of node 17 was 66 %. The plants in node 3 had a high shade density (61.5 % or more) and lower incidence of CLR (23 %) when compared to node 6, which had a less than 61.5 % of shade density and an average of 35 % of the leaves affected by CLR (Figure 8, averages not shown).

The results from these regression trees did not provide clear evidence that shade was the most important variable to define either yield or leaf area. However, there is some indication that more shade may reduce dieback and CLR incidence, and thereby may be contribute to higher leaf area values. It is for this reason that separate regression trees for the different shade groups used in the boundary line analysis were not performed.

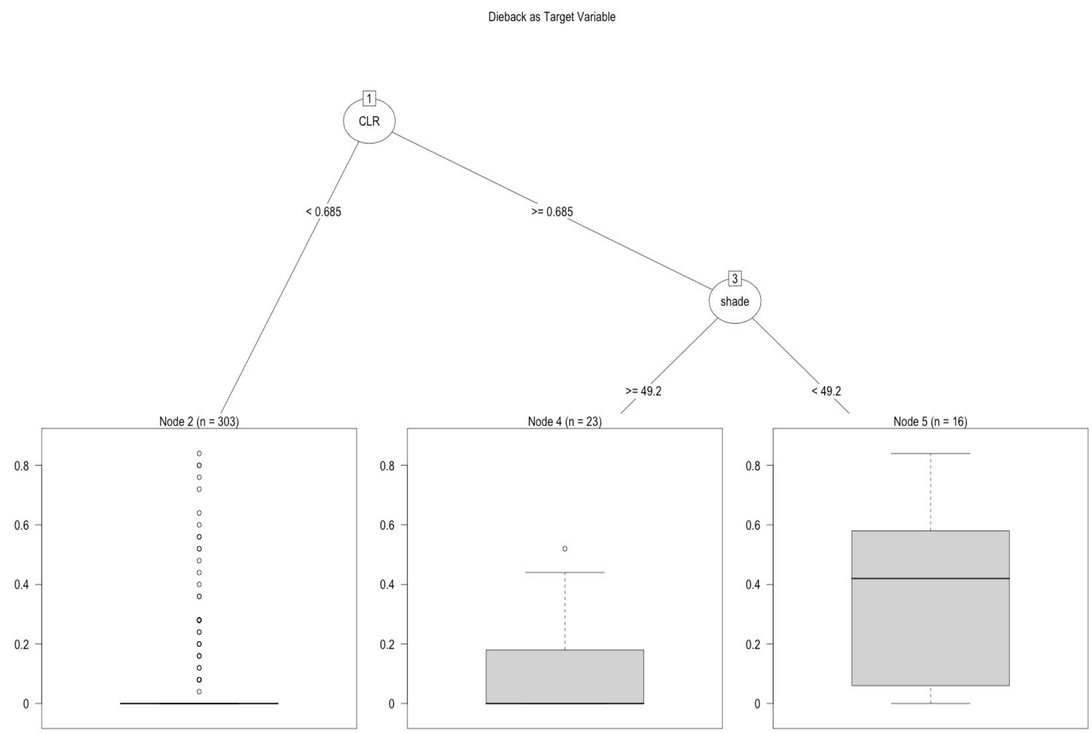

Figure 7: Regression tree to describe the branches affected by dieback (%) variability of coffee as function of different variables comprising soil, plant and management characteristics. Node 1 is representing the entire population of coffee plants (n=318). The bottom groups represent the terminal nodes, including a graph showing the yield distribution within that group. Every node between the very top group to the terminal nodes is the path taken, and it shows the discerning thresholds, for corresponding variable determining in the terminal nodes.

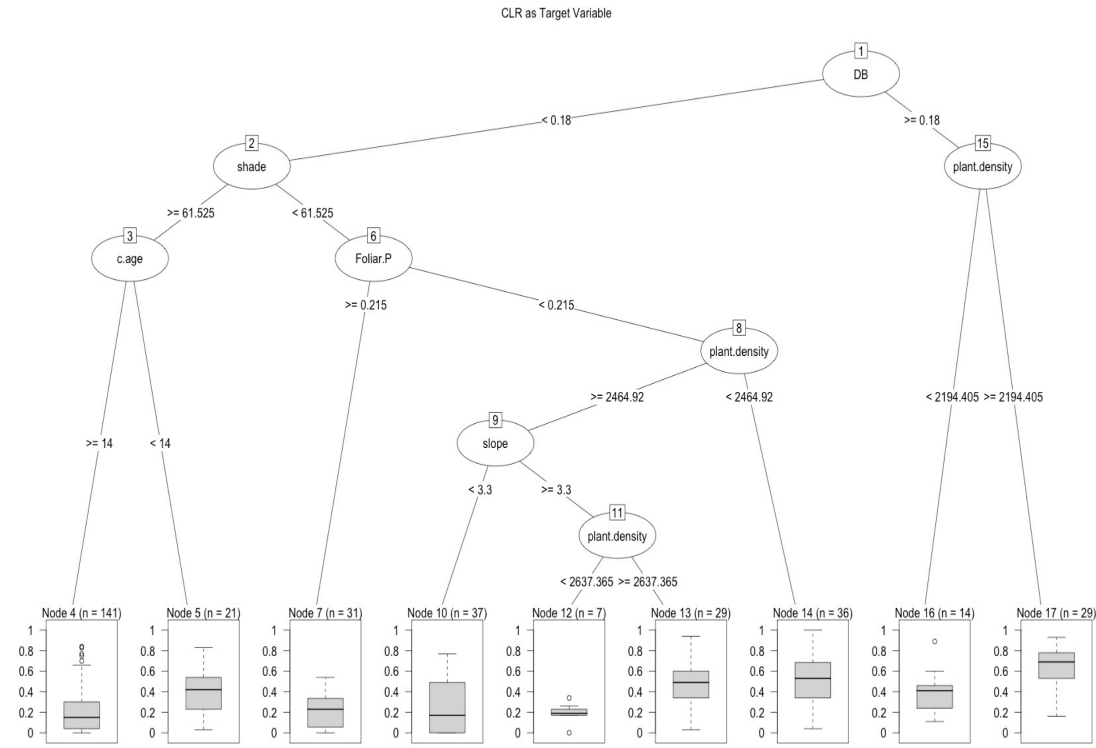

Figure 8: Regression tree to describe Coffee Leaf Rust (CLR) variability of coffee as function of different variables comprising soil, plant and management characteristics. Node 1 is representing the entire population of coffee plants (n=318). The bottom groups represent the terminal nodes, including a graph showing the yield distribution within that group. Every node between the very top group to the terminal nodes is the path taken, and it shows the discerning thresholds, for corresponding variable determining in the terminal nodes.

3.3. Shade Effect on Coffee

The boundary line analysis provided distribution patterns for attainable yield as affected by shade regime (Figure 4). Nevertheless, this analysis did not show how shade affects overall crop performance. Moreover, it was observed that the variability within and among different farms was high, indicating that there are probably high intrinsic levels of variation in every farm.

Due to the situation described above, we produced two linear mixed effect models, placing the shade density as fixed effect to predict both yield and foliar area. To assess the effect of shade on yield, we compared a range of models by the Alkaike Information Criterion (AIC) with smallest values being better. This was done to assess which model fitted best the sampled population of coffee plants.

The use of a random intercept and slope model proved to be more suitable for the data (AIC = 1686.8, Appendix 7). Based on visual inspection of residuals, it was concluded that the assumptions of a normal distribution and the homoscedasticity of the model were not valid. After a log transformation of the dependent variable, these problems were effectively addressed. It was shown that shade had no effect on overall coffee yield (chisq = 0.97, *P* = 0.32) across farms. However, shade had a positive and negative impact on coffee per farm as shown in Figure 9. Based on this figure it appears that response slopes were highly variable across the different farms. Farms number 3, 8, 11, 12 and 19 had the highest intercept and also the steepest slopes (both positive and negative) among the 23 farms included in this analysis. The other farms did not deviate too much from the overall population's intercept and the response of yield to shade was not very pronounced.

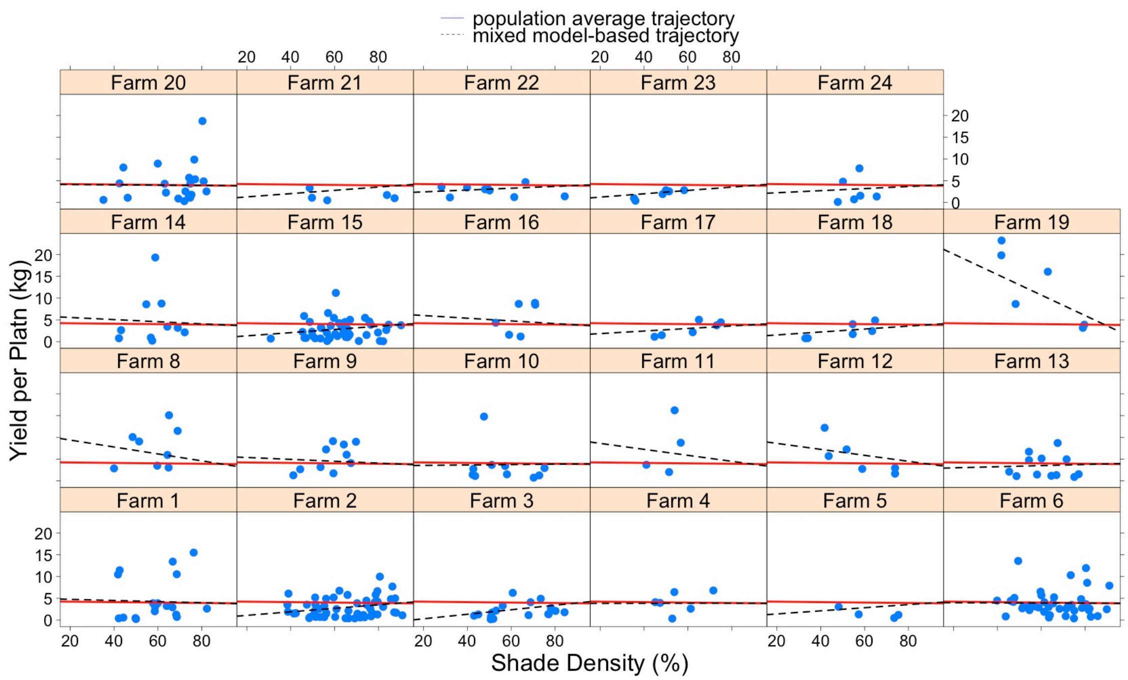

Figure 9: Relationship between shade density (%) and coffee yield (kg/plant) for different farms. Each plot represents a farm. Red lines show the result of the linear mixed model for population, and it is specified by 4.326 - 0.005 x Shade. The doted lines are the response of yield to shade for a specific farm. The blue lines are the outcome of a simple linear model for a specific farm.

Shade had a significant effect on leaf area of coffee plants, increasing it by LA=0.08 ± 0.025 m^2/plant (chisq = 49.144, *P* <0.001, Appendix 8). The chosen model had a by-farm random intercept, but no by-farm random slope. Farm 19 presented the highest deviation from the population's intercept (Figure 10).

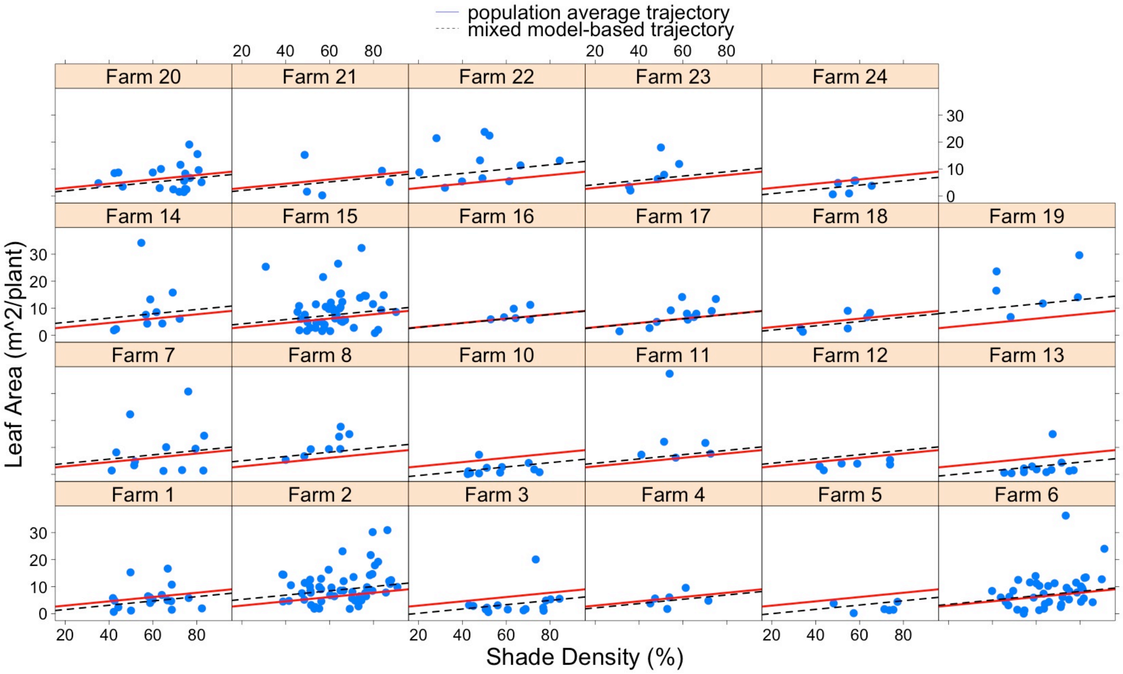

Figure 10: Relationship between shade density (%) and coffee leaf area (m²/plant) for different farms. Each plot represents a farm. Red lines are the result of the linear mixed model for the population, and it is specified by 1.373 + 0.080 x Shade. The doted lines are the response of yield to shade for a specific farm. The blue lines are the outcome of a simple linear model for a specific farm.

3.4. Yield Gap Analysis

The regression trees provided a good understanding on some of the driving factors that govern the clustering of coffee into different yield performance groups. From this point, more specifically, from the first presented regression tree (Figure 5), it was decided to form and examine three groups in order to determine what factors may explain the yield differences. To do so, node 14 (yield=8.5 kg/plant, n = 16) was compared with nodes 2 (2.1 kg/plant, n = 136) and 6 (4.4 kg/plant, n = 51) in a spider chart, for better visual comparison, similar as was done by Delmotte, Tittonell et al. (2011). Although node 15 had the highest yield (13 kg/plant), we did not include this group because of its low number of observations (n = 9).

The yield gaps across those groups were: 6.4 kg/plant between nodes 14 and 2; 4.1 kg/plant between nodes 14 and 6; and 2.3 between nodes 6 and 2. When comparing the average yield of node 14 with the maximum attainable yield observed in the current study (Table 2), we see a yield gap of around 3 kg/plant. This gap is even bigger when Node 2 is compared with the attainable yield, which would translate to a yield gap of around 9.4 kg/plant.

The slope of the land in this analysis seems to play an important role, which is to be expected based on the first regression tree (Figure 5). The average slope of the high yielding group was considerably lower than the other groups, being 2.2, 5.1 and 4.7 % for the high, intermediate and low yielding groups, respectively.

When comparing the variables across the three different groups, there is no strong indication that any of the soil quality and plant nutrition parameters were responsible for the observed yield gaps (Table 3). From all the soil indicators, only Ca and pH had a significant difference among the groups ($P \leq 0.05$), with the high yielding group having a slightly higher pH (6.8) than the intermediate (6.4) and low (6.4) yielding groups. Regarding soil Ca levels, the high (923 ppm) and low (856 ppm) groups presented significantly higher contents compared to the intermediate yielding group (800 ppm). Foliar K was higher in node 2 (4.4 %) than in node 6 (3.9 %). The foliar K content for node 14 (4.3 %) was not significantly different from the other two nodes. The application of NPK fertilizer did not significantly differ ($P \leq 0.05$) across the groups (data not shown).

Key indicators related to crop management, including coffee age, number of stems, plant density and foliar area were higher for node 14 when compared to node 2 ($P \leq 0.05$) (Figure 11). Moreover, CLR also was higher for node 2 (40 %) than in nodes 14 (24 %) and 6 (29 %). Even though the CLR was higher in node 2, the other two nodes still had a relatively high incidence of CLR as well.

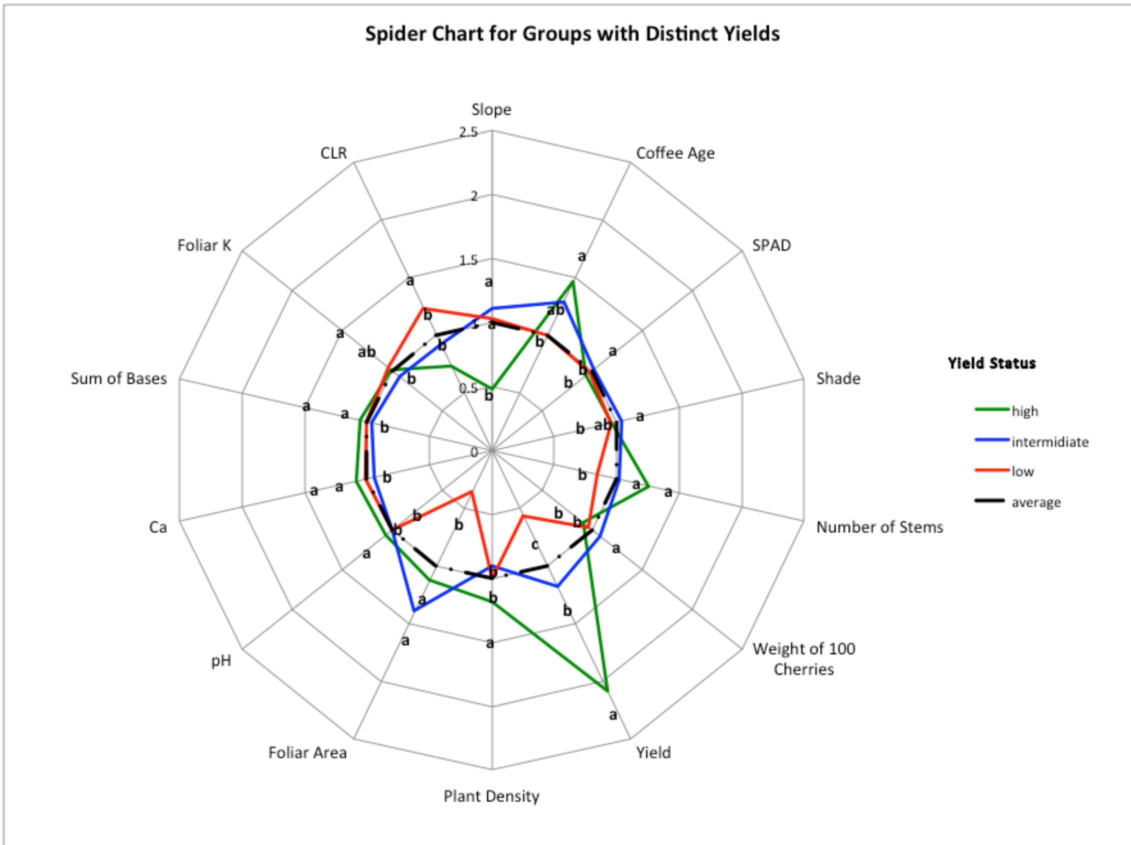

Figure 11: Spider chart showing the key variables that describe differences among the high, intermediate and low yielding groups. Expressed values are relative to the population's average, being above one higher than average. The groups were obtained from nodes 2, 6 and 14 which correspond to low, intermediate and high yields (Figure 5), respectively. Different letters within the same line express significant difference ($P \leq 0.05$)

From the results obtained from the mixed effect model, we selected nine farms, three of which that had the steepest positive slope to shade density (farms 2, 3 and 5, Figure 9), three with the steepest negative slope (farms 8, 11 and 19, Figure 9), and three for which the intercept closely matched the one for the overall population. We plotted those three groups in the same spider diagram using the same methodology explained earlier in this section. This was done in order to gain a better understanding on how the positive, negative or zero response curves to shade density link up with other management factors as well.

It can be clearly seen that the different groups had distinct differences in yields (Figure 12). The group with negative response to shade density was the one that had the highest average yield (8.02 kg/plant), followed by the group with zero response (4.23 kg/plant)to shade and finally the group with positive response (2.58 kg/plant)(Figure 9 and Table 3).

It is interesting to note that shade can probably benefit low yielding plants, while it may have the opposite effect on high yielding plants. Another aspect related to the group with positive response to shade density is that dieback affected 11 % of the branches, which is relatively high. As showed before in the regression tree (Figure 7) and in the boundary line analysis (Figure 4 K), lower incidence of dieback is associated with higher shade density. Furthermore, the group with positive response to shade had, on average, the highest slope.

High number of stems appears to be linked to the high average yield observed for the group with negative response to shade, since it relates closely to the yield. Moreover, the same group had the highest plant density of 3216 plants/ha, which is still relatively low. Other important aspect regarding this group is the high leaf area, which is also associated with higher yields. The application of manure and NPK for this group was very close to the average of the whole sampled population (Figure 11). The application of NPK and manure was the highest in the group with no response to shade density, but it was not related to the highest soil and tissue nutrient levels (Figure 11 and Table 3). Interesting enough, the group with a positive response to shade tended to have higher soil and tissue nutrient content. However, application of manure and NPK was the lowest of all three groups, which appears to be a bit surprising.

Plants in the group that did not show a response to shade density typically were the oldest stands, with the average coffee shrub age being 36 years old and also received the highest manure application.

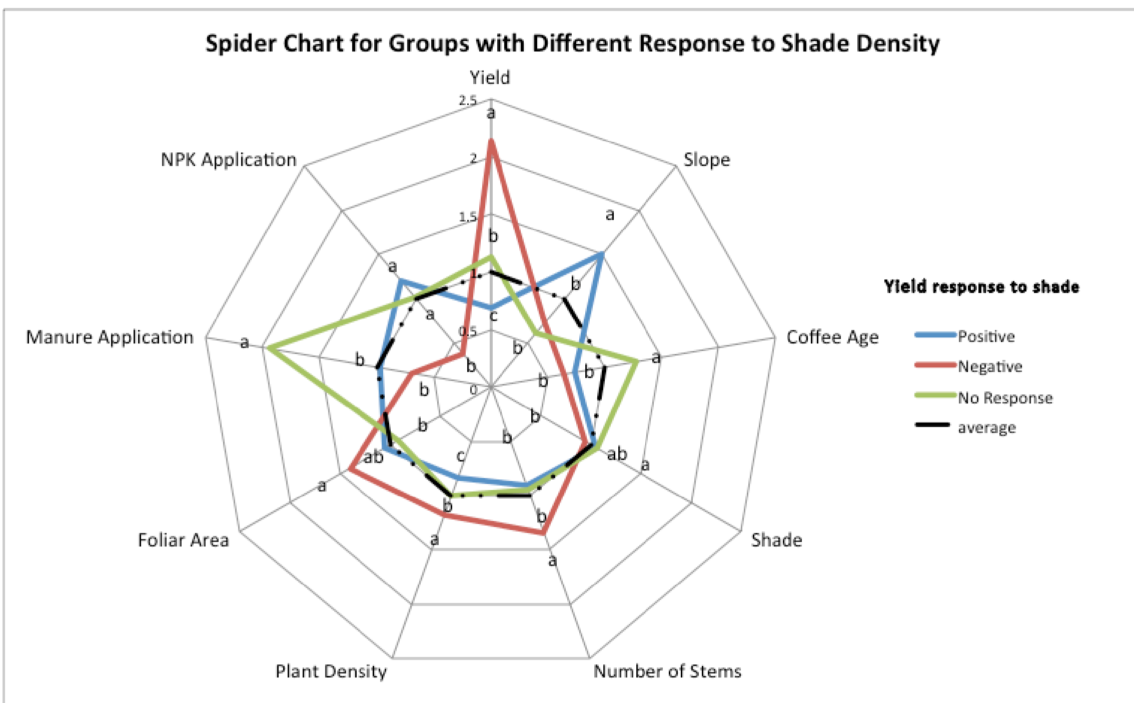

Figure 12: Spider chart containing the key variables that describe the differences between the groups with increasing, decreasing and no effect of shade density on yield. Expressed values are relative to the population's average, being above one higher than average. The groups were obtained from Figure 9. Different letters within the same line express significant difference (*P* ≤ 0.05)

Table 3:

Average values of variables related to plant status, management, and soil and landscape for the groups with increasing, decreasing and no effect of shade on yield, and for the groups with high, intermediate and low yields.

Plant Status													
Yield behavior to increasing shade density	Yield (kg/plant)	Weight of 1 Cherry (g)	Coffee Age (years)	SPAD Reading	Foliar Area (m^2/plant)	Incidence of Coffee Leaf Rust on Leaves (%)	Branches Affected by Dieback (%)	Foliar Area (m^2/plant)	Foliar N (%)	Foliar P (%)	Foliar K (%)	Foliar Ca (%)	Foliar Mg (%)
Slope pattern													
Increase	2.58 c	1.6 a	19.25 b	54.48 b	8.38 ab	28 a	11.0 a	8.38 ab	2.78 a	0.26 a	4.69 a	1.05 a	0.80 a
Decrease	8.02 a	1.60 a	16.66 b	54.79 b	11.12 a	34 a	4.0 ab	11.12 a	2.73 a	0.15 b	3.80 b	0.89 b	0.77 ab
No Effect	4.23 b	1.63 a	33.26 a	57.80 a	7.30 b	27 a	2.5 b	7.30 b	3.02 a	0.18 b	4.55 a	1.11 a	0.71b
Yield													
High	7.82 a	1.49 b	38.21 b	52.58 b	8.91 b	0.23 a	3.5 a	8.91 b	2.55 a	0.16 a	4.27 a	1.22 a	0.71 a
Intermediate	4.43 b	1.77 a	33.55 ab	56.94 a	11.02 a	0.29 b	4.7 a	11.02 a	2.81 a	0.17 a	3.89 b	1.05 a	0.76 a
Low	2.11 b	1.58 b	26.10 a	54.19 b	2.80 a	0.39 b	12.4 b	2.80 a	2.77 a	0.18 a	4.36 ab	1.0 a	0.73 a
Whole Population	3.75	1.65	26.02	55.81	7.91	32.0	6.2	7.91	2.85	0.18	4.22	1.06	0.75

Management								
Slope Pattern	Distance From the Shading Tree (m)	Shade Density (%)	Number of Stems	Plant Density (plants/ha)	Desuckering Frequency (times/year)	Pest Control frequency (times/year)	Manure Application (kg/year)	NPK Application (kg/year)
Increase	5.76 a	64.64 ab	3.44 b	2279 c	1 b	1 b	4.57 a	0.25 a
Decrease	8.11 a	57.93 b	5.11 a	3216 a	1 b	2 a	3.25 a	0.08 b
No Effect	7.68 a	65.13 a	3.56 b	2746 b	3 a	1 b	9.09 b	0.21 a
Yield								
High	5.66 b	58.96 b	4.76 b	3246.28 b	2 a	2 a	4.56 a	0.10 a
Intermediate	6.76 b	64.13 a	3.88 a	2462.74 b	2 a	1 a	3.48 a	0.24 a
Low	7.92 a	58.64 ab	3.18 a	2736.46 a	3 a	2 a	5.71 a	0.23 a
Whole Population	6.88	61.44	3.79	2734.03	2	2	4.70	0.21

Soil and Landscape									
Slope Pattern	Slope (%)	pH	Organic Matter (%)	N (%)	P (ppm)	Ca (ppm)	Mg (ppm)	K (ppm)	Sum of Bases (ppm)
Increase	6.83 a	6.35 b	4.13 b	0.20 a	56.59 a	885.55 a	201.28 a	265.63 a	6.76 a
Decrease	3.36 b	6.20 b	4.52 a	0.21 a	35.56 b	751.51 c	197.44 a	250.87 b	6.02 c
No Effect	2.76 b	6.72 a	4.51 a	0.21 a	45.11 b	814.30 b	192.51 b	267.24 a	6.33 b
Yield									
High	2.16 a	6.84 b	4.42 a	0.21 a	57.21 a	922.75 a	194.27 a	267.17 a	6.89 a
Intermediate	5.08 a	6.43 b	4.31 a	0.21 a	42.79 a	800.25 b	195.96 a	255.16 a	6.26 b
Low	4.71b	6.43 a	4.31 a	0.30 a	49.52 a	856.42 a	196.99 a	259.72 a	6.56 a
Whole Population	4.56	6.46	4.38	0.26	47.88	847.94	196.99	259.05	6.52

Different letters within a same column represent significant differences between the variables ($P \leq 0.05$)

4. Discussion

This experiment aimed to elucidate what factors are governing intrinsic yield variability, yield gaps, and to also define the role of shade in this context. In this manner, the following research question was addressed: how shade density and other factors contribute to yield variability and yield gap of coffee?

Shade is an important characteristic and currently topic that generates a lot of discussion on coffee production around the world, especially as related to climate change an coffee production in relatively hot production regions. This research also defined key management points that should be addressed in order to improve coffee yields in the region of Mount Elgon, Uganda. To do so, we looked into three different scopes related to coffee production being 1) field management; 2) prevailing soil and landscape conditions; and 3) plant nutrition and phytosanitary aspects.

The inherent complexity of the system and the manifold interactions among key variables and parameters required an integrated research approach. The multiple analysis used in this thesis thus are being integrated in order to address the complexity associated with yield variability. Each taken method as such, was not able to provide a complete understanding of yield variability. However, the combination of different approaches delivered a better insight on the complexity that holds the yield variability, since different angles and aspects were being assessed . The used methods consisted in: (1) determining key distinctions among different groups under different shade regimes, through descriptive analysis; (2) assessing the response of attainable yield to multiple variables as related to different shade regimes (Boundary Line Analysis); (3) clustering the data into more homogenous and identifying the variables that could account for most of the yield variability, via the use of , through regression trees; (4) assessing the effect of shade on coffee yield, by the use of linear mixed-effect model. Ultimately, key variables governing the yield variability were identified and the role of shade in such context was examined. Unfortunately, up to date there are no other multi-analytical studies focusing on yield gap and yield variability of coffee that could be used to compare our findings. Similar study was performed on rice (Delmotte, Tittonell et al. 2011), however no similar approach was done on agroforestry systems. Nevertheless, single key aspects found in our study will be compared with the available literature.

It was observed that there was a great amount of variability within and across the different farms that were being assessed. From this point, the whole population of sampled plants was sorted into three groups, based on a specific range of shade density. Out of those groups, the low shaded group of coffee plants had the lowest attainable yield, followed by the high and then the intermediate shaded plants. There is a lot of discussion about what shade regime is optimal for coffee. (Lambot and Bouharmont 2004), for instance, stated that shade should only filter out between 25 to 40 % of incoming light, and it should never exclude more than that, since production otherwise will drop substantially. However, in a research study conducted in Costa Rica, Muschler (1998) concluded that 50 % of light reduction did not dramatically decrease the coffee production in that environment. Furthermore, light reduction decreased the amount of fruits dropped on the ground, which can be over 20 % of the total amount fruits produced.

To what extent shade is being desirable and beneficial may depend on many system properties and management aspects of coffee production systems. For example, coffee is grown all around the world, and therefore a large range of climatic conditions will have different impact on coffee. Moreover, the availability and use of different coffee varieties and diverse management practices would intensify dissimilarities around the effect of shade on coffee (Muschler 2003).

The current study showed that a shade regime of around 50 to 70 % appears to be the most suitable. Higher shade (71-90 %) was associated with a loss in attainable yield, but it was still more preferable compared to shade levels dropping below 50 %. This since increased shade density may also reduce the incidence of dieback and/or coffee leaf rust. A possible reason for higher shade density being more desirable is the relatively low altitude of the studied area, which is between 1100 to 1400 m high. In Uganda, arabica coffee is predominantly being grown above an altitude of 1400 m, since this cultivar is being better adapted to cooler tropical environments (Jassogne, Läderach et al. 2013). Arabica coffee can be found in those lower lands due to the fact that Mount Elgon is a traditional region regarding the production of *C. Arabica*, thus it is a preference of choice among farmers there rather then it being grown in its optimal cultivation niche.

Although the attainable yield of coffee proved to be higher when the shade density was between 50 to 70 %, there was no indication that shade has a direct impact on coffee yield. However, when looking at individual farms, it was observed that the yield response to shade was highly variable, with slopes being either positive, negative or close to zero. Soto-Pinto et al. (2000) observed that coffee yields, in Chiapas, Mexico, benefited by an increase in shade from 23 to 38 %. According to these authors, yield showed almost no response when the shade was raised to 48%. From that point on, the yields started to decrease with an increase of shade density. Unfortunately, these results were done for entire plots, and our study focused on rather heterogeneous stands and individual plants, making it hard to develop a more detailed comparison.

Interestingly, shade had no effect on yield of coffee plants when analyzing the whole dataset, but it had positive, neutral and negative response depending on the farm. The most striking differences among those groups were yield, land slope, plant density, number of stems, foliar area and incidence of dieback. In the case when shade had a negative effect on yield, the plants had on average, relatively high yield, number of stems, plant density, foliar area, and were planted on flatter slope and were also older. In a long term study, Jaramillo-Botero, Santos et al. (2010) observed that by increasing shade density, the yield decreased while leaf area increased. They also observed that the effect of shade was intensified during the highly productive periods of coffee.

When analyzing the group of farms that had a positive response to shade, it was observed that for this group, a relatively high number of branches were affected by dieback. Shade thus might have had a positive effect in decreasing this problem. The yield response slope of that group was higher compared to the other groups. In this case, it may be argued that shade may result in better soil protection, therefore reducing soil erosion losses. However, it was shown that this group had the same levels of nutrients (if not sometimes higher) compared to the other two groups. There was not sufficient evidence to establish a clear relation between shade and slope.

By assessing the group with no response to shade, it was observed that all tested variables were very close to the population average, except for the lower slope and the relatively high manure application. Jaramillo-Botero et al. (2010) used different fertilizer application rates in a long-term study and looked for correlations between fertilizer application, shade density and yield. They concluded that for their study, fertilization rate had no direct effect on yield across the different shade treatments (up to 48 % reduction of Photosynthetically Active Radiation, PAR). In our research, nutrients tended to be always on the bottom of the regression trees, showing thus their importance in terms of governing coffee yield is off less pronounced compared to other factors. The importance of indicators other than soil nutrients was also evident from our yield gap analysis (Table 3). In fact, based on the Snoeck and Lambot's (2004) coffee nutrient requirements guidelines, observed values for diagnostic leaf tissue values typically appeared to be in optimum range, except for K and Mg, where values actually exceeded target values for coffee.

In terms of the yield gap analysis, three groups were developed, corresponding to low, intermediate and high coffee yield/plant. The group with the lowest yield was associated with higher number of branches affected by dieback and high incidence of CLR on the leaves. Consequently the leaf area for this group was low, which ultimately also appears to negatively affect yield. On the other hand, the highest yielding group presented the opposite characteristics of the later group, while plants were also grown on relative flat slopes (2.16 %).

The inconsistent yield responses to shade regimes observed in the current study demonstrates that indeed shade is subject of variability across different production conditions (as quoted by Ricci, Costa et al. 2006). However, it was also shown that the behavior of yield response to shade is highly irregular even within farms with similar climatic conditions, soil characteristics and management practices. Based on these results it may be concluded that there are no simple and universal guidelines as to what is the best shade management for coffee, even when the conditions appear to be similar. Therefore, it is necessary to develop follow-up studies that explore how shade is interacting with multiple production factors associated in coffee-based systems, rather than exploring the interaction of yield and shade alone.

The methods used here could be further explored in future research studies and system analysis. Such effort should include different regions, and in this manner it might enhance our understanding of factors and processes governing yield variability in complex coffee-based systems. Moreover, it may then be possible to compare yield gap across different study areas. However, since coffee is a perennial crop, long-term research studies that include different growth stages and at least 2 to 4 production cycles are also needed. Our research was only conducted during the production phase of coffee. In addition, coffee has a biennial effect on its production, meaning that years of high production are being alternated with lower production. This type of alternate bearing patterns in coffee could not be captured in the current study but it may well have strong impacts on yield variation. Finally, pests and diseases are highly unpredictable. In other words, what has been observed during the current study, may not be observed during other parts of growth and/or production cycle of coffee. Therefore it is concluded that follow-up research studies should be conducted for at least a two-year period. Using constant monitoring methods throughout the year would also facilitate a better understanding of how in-season growth patterns may affect the yield gap and intrinsic yield variability. Furthermore, long-term studies may allow improved assessment if there are annual variations in terms of the effect of shading on coffee yield.

5. Production Recommendations

Summing up the information gathered during this research study and after comparing it with what is considered good agricultural practices for coffee in Uganda, a number of recommendations may be structured to improve coffee productivity and thereby decrease the yield gap. In this context, it was observed, that nutrients are not the major driver to yield gaps.

The Uganda Coffee Development Authority (2014) describes that old coffee plantations are one of the major constraints the coffee industry is facing. It was observed that indeed coffee plants in the studied area were on average old (about 26 years old). Low-input systems are also seen to be a constraint, according to the Uganda Coffee Development Authority (2014). However, there were no clear evidences that this is the case in the study region. It is worth to mention that farmer's perception regarding to low yields is associated with high incidence of pest and diseases, and also with low application of chemical fertilizers.

Following the recommendations provided by Lambot and Bouharmont (2004), coffee under shade should be stumped more frequently, or the issue can be lessened by the use of more suitable dwarf varieties, such as Caturra or Catimor. They also stated that the stumping issue could be totally avoided by topping the coffee trees. None of those recommendations is actually being put into practice in any of the farms included in the current study. In reality, farmers that do not completely stump their plants, tend to selectively cut the oldest branch and allow a new one to form. Such practice is used since they cannot afford to have an unproductive plant for 3 years (roughly the period that the coffee plant starts to produce). Moreover, observed coffee plants tended to be rather tall, which is in contrast by standard recommendations provided by Lambot and Bouharmont (2004). Therefore, topping the plants seems to be a more preferable solution in this case.

Based on the our research findings, it appears that yields may be improved through better shade and stem management, and using optimal plant densities. Shade management could be optimized by adjusting shade trees density and spacing as related to tree species being used (Lambot and Bouharmont 2004). Moreover, according to Lambot and Bouharmont (2004), in Latin America, shade trees should be pruned twice a year, one time after the main flowering and also during the dry season. In Mount Elgon, tree pruning only happens when farmers need the wood for their personal use, or when they are planning to sell it. Unfortunately they do not frequently prune the trees for the purpose of maintaining optimal shade levels. Of course the situation in Mount Elgon might be different, but it is worth to consider this during future research studies. Regarding optimal plant density and stem management, it appears that both management practices may be better adjusted. This since it was observed that there was a clear pattern-relating yield with higher plant density and number of stems. This may imply that local resources, including stand densities, could be better managed. Nevertheless, one should consider that farmers grow coffee in association with other subsistence crops. As result, intensifying coffee production may lead to a reduction in local food production and/or sovereignty.

6. Acknowledgements

I want to thank my supervisors Johannes Scholberg, Pablo Tittonell, Piet van Asten and Laurence Jassogne for their support during the implementation of this study and the subsequent data analysis and thesis writing process. I also thank David Mukasa for his technical support and data collection. Finally I thank the International Institute of Tropical Agriculture (IITA) for its financial aid and the Wageningen University for providing me the tools for the data analysis.

7. References

Aerts, R., K. Hundera, G. Berecha, P. Gijbels, M. Baeten, M. Van Mechelen, M. Hermy, B. Muys and O. Honnay (2011). "Semi-forest coffee cultivation and the conservation of Ethiopian Afromontane rainforest fragments." Forest Ecology and Management **261**(6): 1034-1041.

Bamutaze, Y., M. M. Tenywa, M. J. G. Majaliwa, V. Vanacker, F. Bagoora, M. Magunda, J. Obando and J. E. Wasige (2010). "Infiltration characteristics of volcanic sloping soils on Mt. Elgon, Eastern Uganda." Catena **80**(2): 122-130.

Bates, D., M. Maechler, B. Bolker and S. Walker (2013). lme4: Linear mixed-effects models using Eigen and S4.

Beer, J., R. Muschler, D. Kass and E. Somarriba (1998). Shade management in coffee and cacao plantations. Directions in Tropical Agroforestry Research. P. K. R. Nair and C. R. Latt, Springer Netherlands. **53:** 139-164.

Chok, N. S. (2010). Pearson's Versus Spearman's and Kendall's Correlation Coefficients for Continuous Data PeerReviewed, University of Pittsburgh.

Cilas, C. and F. Descroix (2008). Yield Estimation and Harvest Period. Coffee: Growing, Processing, Sustainable Production, Wiley-VCH Verlag GmbH**:** 593-603.

DaMatta, F. M. (2004). "Ecophysiological constraints on the production of shaded and unshaded coffee: a review." Field Crops Research **86**(2–3): 99-114.

Delmotte, S., P. Tittonell, J. C. Mouret, R. Hammond and S. Lopez-Ridaura (2011). "On farm assessment of rice yield variability and productivity gaps between organic and conventional cropping systems under Mediterranean climate." European Journal of Agronomy **35**(4): 223-236.

Fermont, A. M., P. J. A. van Asten, P. Tittonell, M. T. van Wijk and K. E. Giller (2009). "Closing the cassava yield gap: An analysis from smallholder farms in East Africa." Field Crops Research **112**(1): 24-36.

Food and Agriculture Organization of the United Nations (2014). "Uganda." **2014**.

Hamilton, A. C. and R. A. Perrott (1981). "A study of altitudinal zonation in the montane forest belt of Mt. Elgon, Kenya/Uganda." Vegetatio **45**(2): 107-125.

International Coffee Organization. (n.d.). "World Coffee Trade." Retrieved April 4, 2014, from http://www.ico.org/trade_e.asp.

Jaramillo-Botero, C., R. H. S. Santos, H. E. P. Martinez, P. R. Cecon and M. P. Fardin (2010). "Production and vegetative growth of coffee trees under fertilization and shade levels." Scientia Agricola **67**: 639-645.

Jassogne, L., P. Läderach and P. van Asten (2013). "The Impact of Climate Change on Coffee in Uganda: Lessons from a case study in the Rwenzori Mountains - rr-impact-climate-change-coffee-uganda-030413-en.pdf." **2014**.

Lambot, C. and P. Bouharmont (2004). Soil protection. Coffee: growing, processing, sustainable production : a guidebook for growers, processors, traders, and researchers. J. N. Wintgens. Weinheim, Wiley-VCH**:** LV, 975 p.

Lewis, R. J. (2000). An Introduction to Classification and Regression Tree (CART) Analysis. Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. San Francisco, California**:** 14 p.

Martinez, H. E. P., S. M. Bragança and V. Universidade Federal de (2005). "[Growth and accumulation of nutrients for the coffee plant Conilon (Coffea canephora Pierre)]." **2014**.
Mehlich, A. (1984). "Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant." Communications in Soil Science and Plant Analysis **15**(12): 1409-1416.
Muschler, R. G. (1998). Tree-crop compatibility in agroforestry: production and quality of coffee grown under managed tree shade in Costa Rica, Gainesville, FL, University of Florida, US.
Muschler, R. G. (2003). Shade management and its ffect on Coffee growth and quality. Coffee: growing processing, sustainable production : a guidebook for growers, processors, traders, and researchers. J. N. Wintgens. Weinheim, Wiley-VCH**:** LV, 975 p.
Nelson, D. W. and L. E. Sommers (1996). Total carbon, organic carbon, and organic matter. Methods of soil analysis. Part 3 - chemical methods. S. S. S. o. America. Madison, Soil Science Society of America Inc.**:** 961-1010.
Norgrove, L. and D. Hulme (2006). "Confronting Conservation at Mount Elgon, Uganda." Development and Change **37**(5): 1093-1116.
Okalebo, J. R., K. W. Gathua and P. L. Woomer (1993). Laboratory methods of soil and plant analysis: a working manual. Nairobi, UNESCO-ROSTA.
R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.
Ricci, M. d. S. F., J. R. Costa, A. N. Pinto and V. L. d. S. Santos (2006). "Cultivo orgânico de cultivares de café a pleno sol e sombreado." Pesquisa Agropecuária Brasileira **41**: 569-575.
Ritz, C. and J. C. Streibig (2005). "Bioassay Analysis using R." Journal of Statistical Software **12**(5).
Rodrigues Junior, F. A., L. B. Vieira, D. M. Queiroz and N. T. Santos (2011). "Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar." Revista Brasileira de Engenharia Agrícola e Ambiental **15**: 778-787.
Sarkar, D. (2008). Lattice: Multivariate Data Visualization with R. New York, Springer.
Shatar, T. M. and A. B. McBratney (2004). "Boundary-line analysis of field-scale yield response to soil properties." The Journal of Agricultural Science **142**(05): 553-560.
Snoeck, J. and C. Lambot (2004). Crop Maintenance. Coffee: Growing processing, sustainable production : a guidebook for growers, processors, traders, and researchers. J. N. Wintgens. Weinheim, Wiley-VCH**:** LV, 975 p.
Tittonell, P. and K. E. Giller (2013). "When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture." Field Crops Research **143**(0): 76-90.
Tittonell, P., K. D. Shepherd, B. Vanlauwe and K. E. Giller (2008). "Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—An application of classification and regression tree analysis." Agriculture, Ecosystems & Environment **123**(1–3): 137-150.

Tittonell, P., B. Vanlauwe, M. Corbeels and K. E. Giller (2008). "Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya." Plant and Soil **313**(1-2): 19-37.
Uganda Coffee Development Authority (2014). "General Coffee Production." **2014**.
Uganda Wildlife Authority (2014). "Mount Elgon National Park." **2014**.
van Asten, P. J. A., L. W. I. Wairegi, D. Mukasa and N. O. Uringi (2011). "Agronomic and economic benefits of coffee–banana intercropping in Uganda's smallholder farming systems." Agricultural Systems **104**(4): 326-334.
van Asten, P. J. A., M. C. S. Wopereis, S. Haefele, M. O. Isselmou and M. J. Kropff (2003). "Explaining yield gaps on farmer-identified degraded and non-degraded soils in a Sahelian irrigated rice scheme." NJAS - Wageningen Journal of Life Sciences **50**(3–4): 277-296.
Wairegi, L. W. I., P. J. A. van Asten, M. M. Tenywa and M. A. Bekunda (2010). "Abiotic constraints override biotic constraints in East African highland banana systems." Field Crops Research **117**(1): 146-153.
Winter, B. (2014). "A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)." Retrieved April 4, 2014, from http://arxiv.org/pdf/1308.5499.pdf.
Wintgens, J. N. (2003). Coffee: growing, processing, sustainable production. A guidebook for growers, processors, traders and researchers. Weinheim, Wiley-VCH.
Zuur, A. F. (2009). Mixed effects models and extensions in ecology with R. New York etc., Springer.

8. Table of Figures

9. Appendices

9.1. Appendix 1: Correlation Tests

	p-value	z	tau
Yield and Distance	0.012	-2.51	-0.14
Yield and Plant Density	0.049	-1.96	1453309
Yield and Phosphorous	0.033	-2.1339	-0.12
Yield and CLR	0.009	-2.63	-0.20
Yield and NPK	0.027	2.22	0.23

Appendix 1: correlation parameters and P-value

9.2. Appendix 2: Principal Component Analysis

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.97	1.77	1.44	1.34	1.30	1.15	1.08
Proportion of Variance	0.16	0.13	0.08	0.08	0.07	0.06	0.05
Cumulative Proportion	0.16	0.29	0.38	0.45	0.52	0.58	0.63

	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Standard deviation	1.04	0.99	0.98	0.93	0.85	0.83	0.76
Proportion of Variance	0.04	0.04	0.04	0.04	0.03	0.03	0.02
Cumulative Proportion	0.67	0.71	0.75	0.79	0.82	0.85	0.87

Appendix 2: Principal Component Analysis. PC = Principal Component.

9.3. Appendix 3: Regressions Tree

	CP	nsplit	rel error	xerror	xstd
1	0.147489364	0	1	1.005638853	0.160056496
2	0.076274987	1	0.852510636	0.944929355	0.14636275
3	0.070501323	2	0.776235649	0.951059577	0.14256429
4	0.069232364	3	0.705734326	0.960042681	0.143599363
5	0.029775252	4	0.636501962	0.898154536	0.117518992
6	0.022488498	5	0.60672671	0.884378047	0.114459412
7	0.021881099	6	0.584238213	0.849457146	0.110421334
8	0.01901287	7	0.562357114	0.852270191	0.110486417
9	0.016471202	8	0.543344244	0.842655419	0.109588414
10	0.011393453	9	0.526873042	0.842335794	0.108923093

11	0.010468806	10	0.515479589	0.837399954	0.108081221
12	0.01	11	0.505010782	0.838376694	0.108060911

Appendix 3: complexity parameter, number of splits, relative error, estimation of cross-validated prediction error and standard error of the regression tree presented in Figure 5.

	CP	**nsplit**	**rel error**	**xerror**	**xstd**
1	0.071832738	0	1	1.005062911	0.131866464
2	0.054374738	1	0.928167262	0.986984699	0.133900637
3	0.035951986	2	0.873792524	1.078525872	0.139307083
4	0.034654188	3	0.837840539	1.050784086	0.13548771
5	0.029825283	4	0.803186351	1.053986816	0.13434498
6	0.029580192	5	0.773361067	1.042892714	0.133156272
7	0.026693343	6	0.743780875	1.037826276	0.132511991
8	0.020876528	7	0.717087533	1.073180542	0.13223142
9	0.018601749	8	0.696211005	1.092602342	0.127618485
10	0.016331926	9	0.677609256	1.094219596	0.128349863
11	0.014038131	10	0.661277329	1.115707334	0.132744536
12	0.013974959	11	0.647239199	1.091363774	0.129669804
13	0.010273547	12	0.63326424	1.084038736	0.129394339
14	0.01	15	0.6024436	1.113635268	0.127219829

Appendix 4: complexity parameter, number of splits, relative error, estimation of cross-validated prediction error and standard error of the regression tree presented in Figure 6.

	CP	**nsplit**	**rel error**	**xerror**	**xstd**
1	0.105993933	0	1	1.008133465	0.166458251
2	0.060128592	1	0.894006067	0.991394867	0.167293794
3	0.04969223	2	0.833877475	0.99044855	0.161669867
4	0.02443267	5	0.684800785	1.068715964	0.168736718
5	0.023724633	8	0.611502776	1.038487605	0.160435389
6	0.01	9	0.587778143	0.999420849	0.150613523

Appendix 5: complexity parameter, number of splits, relative error, estimation of cross-validated prediction error and standard error of the regression tree presented in Figure 7.

	CP	**nsplit**	**rel error**	**xerror**	**xstd**
1	0.128204534	0	1	1.00626053	0.060252983
2	0.053640567	1	0.871795466	0.905634569	0.059655705
3	0.042107717	2	0.818154899	0.975328633	0.068939835
4	0.038036858	3	0.776047183	0.98233651	0.069585852
5	0.028857969	4	0.738010325	0.971292713	0.069206627
6	0.028485695	5	0.709152355	1.001166909	0.07329913
7	0.022650923	6	0.680666661	0.993320384	0.072652854
8	0.020863766	7	0.658015737	0.966626603	0.071603275

9	0.019878773	8	0.637151971	0.947355278	0.068702141
10	0.017855277	9	0.617273199	0.921384497	0.067961114
11	0.015563452	10	0.599417921	0.926912395	0.070089647
12	0.015104544	11	0.583854469	0.934224735	0.071788492
13	0.014773595	13	0.553645382	0.931435404	0.071402045
14	0.01286622	14	0.538871787	0.908953385	0.070049599
15	0.011286667	16	0.513139346	0.90963447	0.068828654
16	0.010485301	18	0.490566011	0.91699302	0.069020491
17	0.01	19	0.48008071	0.925462377	0.069793851

Appendix 6: complexity parameter, number of splits, relative error, estimation of cross-validated prediction error and standard error of the regression tree presented in Figure 8.

9.4. Appendix 4: Mixed Effect Model

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	P
RI	6	1695.99	1718.52	-841.99	1683.99	NA	NA	NA
RI + RS	8	1686.81	1716.85	-835.40	1670.81	13.18	2	0.0014**

Appendix 7: Likelihood Ratio Test comparing two mixed effect models. "RI" stands for the random intercept model, and "RI + RS" for random intercept and slope model. Both models include shade as fixed effect, and farm, CLR and dieback as random effects. "RI + RS" was the model used in fig. Figure 9. Df = degrees of freedom; AIC = Alkaike Information Criteria; Bayesian Information Criteria; logLik = Log Likelihood; Chisq = Chi Square; Chi Df = degrees of freedom associated with chisq.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	P
No Shade	5	2280.12	2299.38	-1135.06	2270.12	NA	NA	NA
Shade	6	2232.97	2256.00	-1110.49	2220.97	49.14	1	2.38E-12***

Appendix 8: Likelihood Ratio Test comparing two random intercept mixed effect models. "No shade" is a model without including the effect of shade. "Shade" is a model where the effect in question is included. Both models include farm, CLR and dieback as random effects. "Shade" was the model used in Figure 9. Df = degrees of freedom; AIC = Alkaike Information Criteria; Bayesian Information Criteria; logLik = Log Likelihood; Chisq = Chi Square; Chi Df = degrees of freedom associated with chisq.