ANALYSIS AND REVISION OF THE N2AFRICA FOCAL ADAPTATION TRIAL SURVEY, A TOOL FOR MONITORING TECHNOLOGY PERFORMANCE AND UNTANGLING YIELD VARIABILITY

Susana Prieto Bravo
MSc Internship Report (PPS- 70424)

WAGENINGEN
UNIVERSITY & RESEARCH

N2Africa Project

Analysis and revision of the N2Africa focal adaptation trial survey, a tool for monitoring technology performance and untangling yield variability

Susana Prieto Bravo

July 2017

MSc Internship Report (PPS – 70424)

Supervisor: Joost van Heerwaarden

Examiner: Maja Slingerland

Summary

N2AFRICA is a large-scale project with the aim of putting nitrogen fixation to work for smallholder farmers growing legume crops in Africa. It is a science-based "research-in-development" project, with constant learning loops to find the best technology for every farmer. The main goal of N2africa's agronomic research is to understand the major constrains on legume productivity, with a special focus on the causes of yield variability and how to reduce it. One of the four activities in this cluster is the adaptation trials, that evaluate the performance and adaptation of the proposed technologies under farmer's management. Focal farmers are chosen to collect detailed information about changes and management of the proposed technologies, as well as agronomic and household characteristics that might affect yield. The surveys are done using electronic tablets with the ODK software. Since a fundamental step in the N2africa learning loops is having enough and reliable data, the first objective of this internship was to revise the quality of the data collected in the Focal adaptation trials from 2015 and 2016. The second objective was to evaluate if the data collected gives information about a) the changes in the technologies, b) their performance and c) information that will help targeting the technology. Several household and farm characteristics were evaluated for their relationship with yield with a linear mixed model. Some of the variables were chosen for principal component analysis (PCA) and subsequent hierarchical clustering per each country to separate farms into separate groups. Overall, the data was complete but some inconsistences were found, ambiguous or incomplete questions were further analysed and changes were suggested. Frequency of hired labour, education level, farm size and inputs (among others) were significantly related with yield, and an interaction with treatment shows they might be relevant for the performance of the N2Africa technology. The clustering for each country was rather arbitrary due to the big amount of NAs values, however, farm size, the ownership of livestock, the labour dynamics and the market orientation of the farm showed differences among clusters which suggest these variables could be the base of a farm characterization. Labour requirements and marketability play fundamental rolls for adoption, so understanding the several types of farms, will improve the targeting of the technologies. Based on the results and feedback from N2Africa stuff, the survey for 2017 was adapted. This included changing the general structure of the survey, questions that recorded the changes of the farmers in the technologies were systematized for easier analysis and others were removed or changed. The guidelines for the focal adaptation trials were also updated. Finally, recommendations were made on how to adapt the survey to have more information about the potential of farmers to adopt certain technology, which will help in targeting the best-fit legumes for specific situations.

Table of Contents

1. Introduction

1.1 General Context

Nitrogen is an essential nutrient for plant growth but is often a limiting factor in the field, which reduces the yield of important crops around the world. In West Africa, 76% of soils are characterized as naturally poor (Bationo, et al., 2011). Additionally, agricultural soils have now a nitrogen deficiency due to continuous cultivation with insufficient input of mineral fertilizer or organic manures (Nezomba, et al., 2008). The lack of agricultural inputs is usually related with a low investment potential or poor resource endowment, especially in small holder farmers, who account for 75% of agriculture production in Easter African countries (Kenya, Ethiopia, Uganda and Tanzania) (Salami, et al., 2010).

Unlike other species, plants from the leguminous family form symbiotic relationship with nitrogen fixing bacteria from the genus *Rhizobium*. The bacteria take up atmospheric nitrogen (not available for the plant) and make it into organic form available for the legumes. Thanks to this association, legumes can easily grow in poor soils, furthermore, the up-taken nitrogen will make them relatively rich in protein compare to other food crops (Herridge, 2008). Moreover, If the vegetative biomass is added to the soil in the form of manure, the nitrogen will be released contributing to the needs of the subsequent crop. According to Giller (2001), leaving the legume's residues in the field can increase the supply of nitrogen in the soil up to 140 kg N ha^{-1}.

In the poorest African countries, grain legumes constitute the major source of protein in the diet, however, the productivity of the major legume crops is low compared to the attainable yield (Table 1) (Bationo, et al., 2011). In addition to low soil fertility, some of the challenges for legume productivity are low soil pH, high salinity, drought and flooding, poor access to improved germplasm, diseases, pest and weeds (Ayuk , 2011). Low legume productivity, which means less useful residues, is also related with low nitrogen input to the soil, creating circles of poverty from which it is difficult to get out.

Table 1. Average national yields (ton/ha) and higher reported potential yields for some economically important legumes. Source: **(FAO, 2014)** and **(PROTA4U, n.d.)**

	Tanzania	Nigeria	Ghana	Ethiopia	Uganda	**Attainable**
Soya Bean	1.0	0.9	NA	2.0	0.5	**3.0**
Climbing Bean	0.9	NA	1.2	1.6	1.3	**3.4**
Groundnut	1.0	1.2	1.3	1.6	0.7	**5.0**
Cowpea	0.9	0.6	NA	NA	0.4	**3.0**

N2 Africa is a project in sub Saharan Africa with the main goal of "putting nitrogen fixation to work" for small holder farmers (N2Africa.org). This means increasing grain legume productivity which will consequently increase the nitrogen input to the system. Technologies based on the use of nitrogen-fixing plants are more likely to be accessible and used by farmers in the tropics (Giller, 2001). Legume technologies are also part of several agro-ecological technologies for restoration of exhausted and degraded soils, and are proposed under the framework of ecological intensification of the farming systems (Tittonell, et al., 2010).

1.2 About N2 Africa

N2Africa is a development to research project based on the collection of data from different on farm trials that are monitored and evaluated (M&E); the research outcomes are then send back to delivery and dissemination (D&D) creating learning loops (figure 1). The project is organized according to several master plans that compiled the guidelines for seven areas[1] in order to have a common approach in all the countries the project is working. The main goal of N2africa's agronomic research is to understand the major constrains for legume productivity, with a special focus on the causes of yield variability and how to reduce it. The agronomic master plan is organized around four different activities: The Diagnosis, Researcher-managed agronomy, Demonstration and Adaptation (Vanlauwe, et al., 2014). This intership will focus on the Adaptation trials. Detailed information about the other clusters is available in the Agronomy Master Plan in the N2 Africa's intranet [www.n2africa.org].

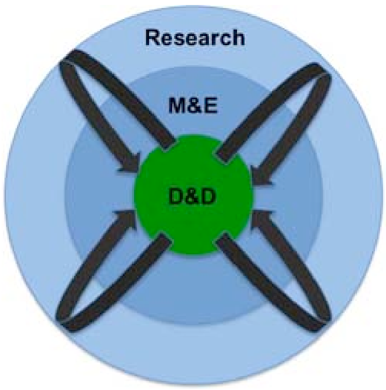

Figure 1 Learning loops are the vision of success of N2Africa. The results of research, based on the monitoring and evaluation (M&E) of farm trials feed constantly the delivery and dissemination (D&D) cluster which brings new (or improve) legume technologies. Source: www.N2Africa.org

In the research-managed activity several best-bet interventions (fertilization regimes, improved varieties and management practices) are proposed to solve some of the constrains for legume productivity found during the diagnosis phase. After presenting these best-bet strategies to the farmers during the demonstration activities, the adaptation trials evaluate the performance and **adaptation** of the proposed technologies under farmer's management (figure 2). **The goal is to**

[1] Agronomy, Dissemination, Monitoring & Evaluation and Data Management, Rhizobiology, Communication, Gender and Innovation platforms

observe how household characteristics and management practices affect the performance of the best bet-options. With the collection of yield data and the associated agronomy and household observations, the best-bet options can be translated into best-fit technologies, adapted to specific biophysical and socioeconomic conditions (Vanlauwe, *et al* ., 2014).

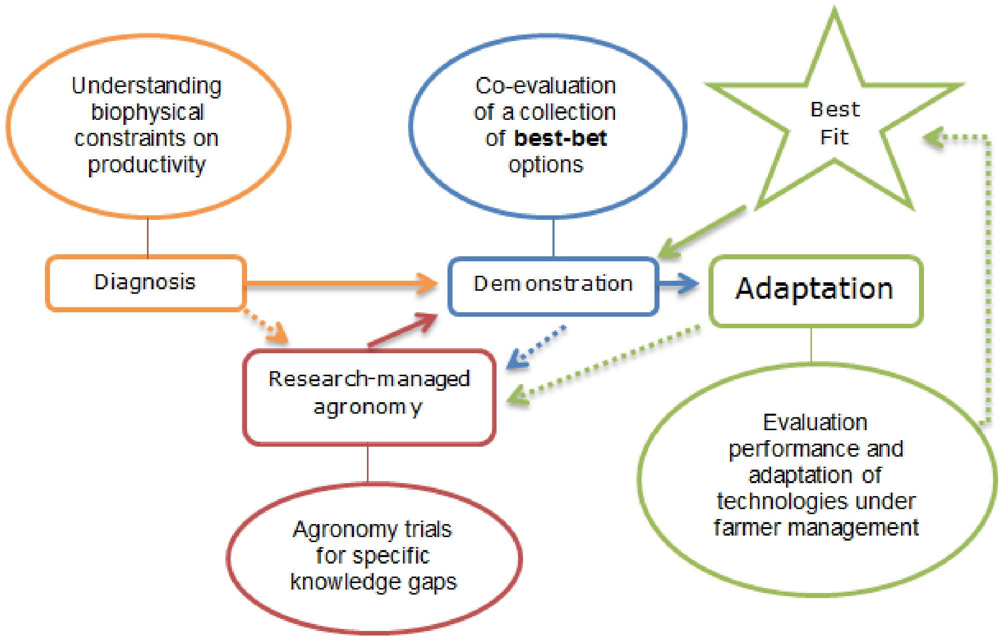

Figure 2. Cluster of the Agronomy master plan for N2 Africa. The dotted arrows represented transfer of knowledge and the full arrows transfer of technologies. Based on Vanlauwe et al, 2014.

The adaptation activity cluster includes a large number of trials for which data collection is done by short questionnaires to the farmers. The technology is distributed with inputs and management recommendation; however, farmers are free to make changes in the implementation and management. In other legume trials, it has been observed that farmers commonly modified the planting patterns of legumes (Kamanga, et al., 2014). These adaptations are recorded and they will be part of the learning loops mentioned before. **Focal farmers** are chosen for more intensive data collection done by research partners under supervision of N2Africa staff. Data is collected with the use of mobile ODK[2] forms following standardized protocols so the data collected across all experiments can be combined and compared.

As mentioned before, N2AFRICA is a large scale, science-based "research-in-development" project. This means the core of the project is the Delivery and Dissemination (D&D) of new legume technologies. Trials supply scientific data for evaluation and monitoring and the research findings will

[2] Open Data Kit (ODK) is an open-source suite of tools for mobile data collection solutions.

go back to D&D again for new (or improved) trials. The adaptation trials are part of this learning loop, thus having useful and reliable data will help to identify best-fit technologies for different contexts.

Additionally, one of the big questions in the N2 Africa project is how to target legume interventions, or in other words, how to make the best-bet technologies into best-fit, adapted to the different types of farmers participating in the project. As mentioned by Ojeim *et al* (2006) , it is not correct to assume that any introduced technology would be successful [increasing the wealth of the farmer]. Franke et al. (2014) using the model NUANCES-FARMSIM and a detailed farm characterization in Malawi created several virtual scenarios for expanding and intensifying legume production, however, they find difficulties when predicting yield variability. In a later paper, they observed that household class served as a better indicator of yield and response to fertiliser than many soil and crop management characteristics which agrees with Ronner et al., (2016).

1.3 Objectives of the internship

Several experiences with different trials in N2africa have found that what happened in the field differ greatly to what is expected from the project (Personal Communication by Thuijsman, E.). Hence, having effective data collection methods (complete surveys, appropriate software, trained technicians) is essential for the success of a big scaled project like N2Africa.

Based on the above, **the aim of this project was to evaluate the Focal Adaptation surveys form 2015 and 2016**. Two specific objectives were proposed:

I. Evaluate the **quality** of the data collected in the Focal Adaptation trials during 2016 to improve the survey for 2017. This includes evaluating the completeness, validity and consistency of the data.

II. Evaluate the informativeness of the data collected in the Focal Adaptation survey. This includes evaluate if the information is giving insights into the **farmers changes** in the implementation and management of the trials. Evaluate if the information collected about the household is related with the **performance of the N2Africa technology, and** evaluate if the information can help **targeting the N2Africa technologies.**

This is motivated by Franke *et al* (2016) who found that household class in Rwanda was a better indicator of yield than other soil and crop characteristics, additionally, differences in soil fertility are related with farmers allocation of resources (e.g. organic manures, labour for weeding) (Giller, 2011).

The results from this internship will be used to propose changes in future surveys to improve the data collection of the Adaptation trial. Furthermore, recommendations will be formulated for better technology targeting based on survey data collection.

2. Materials and Methods

2.1 Data

The data available for analysis was collected from focal farmers that participated in the adaptation trials during 2015 and 2016. There was data from Tanzania, Ethiopia, Ghana, Uganda and Nigeria from trials with soya bean, bush bean, cowpea, groundnut and climbing bean. Yield data was collected from experimental 10x10m plots. Farmers received inputs and management recommendations for the plot, however, they were free to manage the trial. Each farmer also had a control plot (referred to as own_plot), which was either the standard variety/input combination or the farmer's current practice (i.e. a 10 x 10m section in the farmer's main legume field). The data was accessed and downloaded from the N2africa intranet.

2.1.1 Changes in the data

Additional variables were calculated based on the information available in order to do better comparison and analysis: The total number of livestock per farm was converted into Tropical livestock units (TLU) assuming that 1 cattle = 0.7 TLU, 1 sheep = 0.1 TLU, 1 pig = 0.2 TLU, poultry = 0.01 TLU and horse/donkey = 0.7 TLU (Jahnke et al, 1986). The farm area, originally recorded with different units, was transformed to hectares. A new variable was added with the number of month with food sufficiency (12 – number of month with food shortage).

2.2 Analysis

2.2.1 Statistical methods

All the statistical analyses were made with the R statistical software, version 3.3.3 (R Core Team, 2017), Packages car, lme4, FactoMineR and predictmeans. Tables 2 and 3 summarize the variables evaluated in the different analysis presented below.

To analyse the relation between the selected variables and yield, a linear mixed-effect model was chosen. Variables of interest were included as fixed effects whereas other variables potentially affecting variation in the response variable, and potentially confounded with the explanatory variables, were included as random effects. These are terms in a linear predictor expression that evaluates the mean of the response variable (grain yield). Farm and district were included as random nested factors to correct for variability at field and regional level (Bates, 2010, Welham, et al., 2004).

Table 2 Farm and management factors evaluated for the performance of the technology. Variables with * where also used for the PCA (targeting)

Variable/factor	Levels (or unit)
Farm	
Total Farm Size *	hectare
TLU*	Tropical livestock units
Soil depth	meters
Management (for each plot)	
Weeding frequency	1 weeding 2 weeding 3 weeding
Intercropping	Yes No
Row planting	Yes No
Inputs (for each plot)	
Mineral fertilizer	Yes No
Herbicides	Yes No
Pesticides	Yes No
Zero inputs	Yes No

Table 3 Household characteristics evaluated. for the performance of the technology and the targeting (marked * where used in the PCA))

Variable/factor	Levels (or unit)
*Age of the head of the household **	Years
Gender of the head of the household	F M
*Family size **	Number of people living in the household
Highest education of the head of the household and highest level of education in the household	Primary Secondary Post-secondary University Other
*Frequency of hired labour **	Permanently Regularly Sometimes No, never
*Proportion of total income from farming **	All Most farm Half farm Most off-farm All off-farm
*Proportion of produce consumed at the household **	All consumed Most consumed Half consumed Most market All market
Food shortage	Every year Never Some years
*Food Shortage.nr **	Number of months
Main source of income	Crops Livestock Trade Farm labour Off-farm labour Salary

To evaluate if there was significant effect of each factor (plus interactions) on yield, an *Anova* (package *car*) was run for each model. Since the model considers the interaction a type III analysis of deviance was chosen. Additionally, the function calculates a Wald chi-square test to evaluate significant differences. The predicted means for the response variable (grain yield) in each level of the evaluated factors (only the fixed effect, random effects were not considered) were estimated with the function *predictmeans* from the same package.

Yield data much higher than the reported maximum yield for each legume was omitted from the analysis (yields higher than 6 ton per hectares). Values of zero (0) were taken as real zeros, which means the farmer did not have any yield, on the other hand, non-reported yield values were taken as NAs. In the same way, farm and household characteristics that were at least three times the national mean according to the baseline reports were not considered. Residuals of all numerical variables were evaluated for normality and data was log transformed when necessary. No other transformations were found to be required.

2.2.2 For objective 1: Data Quality

a. To evaluate the quality of the data summary statistics were calculated and exploratory graphs plotted for all the variables. A summary table for household characteristic was built for every legume/country combination, which also gave information of the distribution of the N2Africa trials in different countries. Variables with more than 50% of missing values were identified in order to analyse their importance in the survey and propose ways to improve the response ratio.

b. To evaluate the accuracy of the data possible outliers and inconsistent values where contrasted with the N2Africa baseline report for each country[3] and other reports. Questions with a high amount of non-logical values (outliers, wrong units or incorrect answers) were marked for further discussion.

2.2.3 For objective 2: Informativeness of data

a. One of the goal of the focal adaptation trials is to have insights into the farmers changes in the implementation and management of the N2 Africa technology. Most of these changes are recorded in the survey as open questions, which makes the analysis complicated and time consuming. With selected key words, most of the open questions were categorized in multiple answer questions and adapted in the new survey.

[3] The baseline is to establish the current status of livelihoods, through assessment of household characteristics (including education, occupations, sources of income), agricultural production, nutrition and market access (N2africa.org)

b. To analyse information about the performance of the N2Africa technology several linear mixed models were run, changing the type of interaction and the response variables (function *lmer from* the package *lme4*). The model incorporates fixed effects, in this case the selected factors (Tables 2 and 3) and the legume species, and random effects, which are for this analysis the farm (because yield from both plots will be considered) and the district (assuming there is a variation in yield due to location) (box 1). Since it was not in the scope of this project to analyse specific factors for each legume the interaction with the legume species was not considered. It is assumed that part of the variation is already included adding district and farm as grouping factors, since the different legumes are distributed more or less homogenous by region.

Box 1. Linear Mixed model to evaluate yield variability

lmer (estimated.grain.yield ~ **fixed effect 1** + legume + (1| district/farm), REML = FALSE)

To analyse if factors are affecting performance of N2Africa technologies the interaction between factor and treatment (N2Africa or own plot) was introduced in the model (box 2).

Box 2. Linear Mixed model to evaluate the effect of farm variables in plot performance

lmer (estimated.grain.yield ~ **fixed effect 1***treatment + legume + (1| district/farm), REML = FALSE)

c. The first step to target better the N2Africa technologies is to analyse if there are significant household differences within the focal adaptation farmers, and if it's possible to define several types of farmers. To reduce the dimension of the dataset a Principal Component Analysis (PCA) on the selected household variables was run per country, assuming that important differences in household types exist between countries. PCA can be used to describe the variation of a set of correlated household and social variables as coordinates along independent axes, with different contributions from the original variables. This method is chosen based on the work of Tittonell *et al* (2010) who used a PCA to observe the diversity of rural households and its influence in soil fertility and yield. The variables were selected based on the same paper.

The chosen function was *PCA()* from *FactoMineR* package, because of its detailed output and versatile diagnostic tools. The variables chosen for PCA are marked with an asterisk in tables 2 and 3 (section 2.2.1). Numerical variables were scaled before the analyses and ordinal categorical variables were coded as ordered integer values. So as not to reduce the sample size, missing values were imputed for this analysis. The analysis was carried out with the original and the imputed data to also analyse the effect of imputation in the ultimate results

To categorize farms into separate groups, a Euclidean distance matrix was calculated from the first 5 principal components, since in all the countries these represent at least 70 % of the total variation. This was followed by a hierarchical cluster analyses using Ward's method. With this function, the algorithm run iteratively joining similar objects until having one single cluster. At each stage, the distances are recomputed with the Ward's minimum variance method which aims to find compact clusters (Maechler, 2017). Based on the number of principal components chosen to calculate the distance (5) the dendrogram was then divided in five groups with the *cutree* function. 5 groups were selected based in other farm typologies constructed for sub-Saharan Africa (Tittonell, et al., 2010). Finally, for each variable an analyses of variance was run (box 4) to see if there were differences among clusters and among location (district).

Box 4. Analysis of Variance for the fitted clusters

aov (household characteristic ~ fitted cluster* district)

2.3 Design of 2017 survey

Feedback from the technicians doing the surveys was considered together with literature research and evaluation of other forms used in different projects. Based on the above the survey from 2016 was adapted. The adaptations included elimination of questions that were not giving relevant information, as well as addition or changes in questions to get insight into the investment potential of the households which will in turn help to better targeting of the technologies. Answers for open questions were categorized and turned into multiple option questions, based on the most common responses in the previous season. The 2016 survey was edited as an Excel XLSForm (formerly XLS2Xform). The form was then converted to a Xform which can be upload to the ODK software for testing. As part as the adaptation of the ODK survey, and based on the results from the data quality the guidelines for the focal adaptations were also updated.

3. Results

3.1 Quality of the data

In total, there were 856 adaptation trials from 2015 and 2016. Figure 3 shows the distribution of the sampled plots per country and legume for each year. The absence of information, especially of yield from the control plots (the farmers own plots) is also visible. The number of farmers participating increased considerably from 2015 to 2016.

- *About the N2Africa package and the farmers practices*

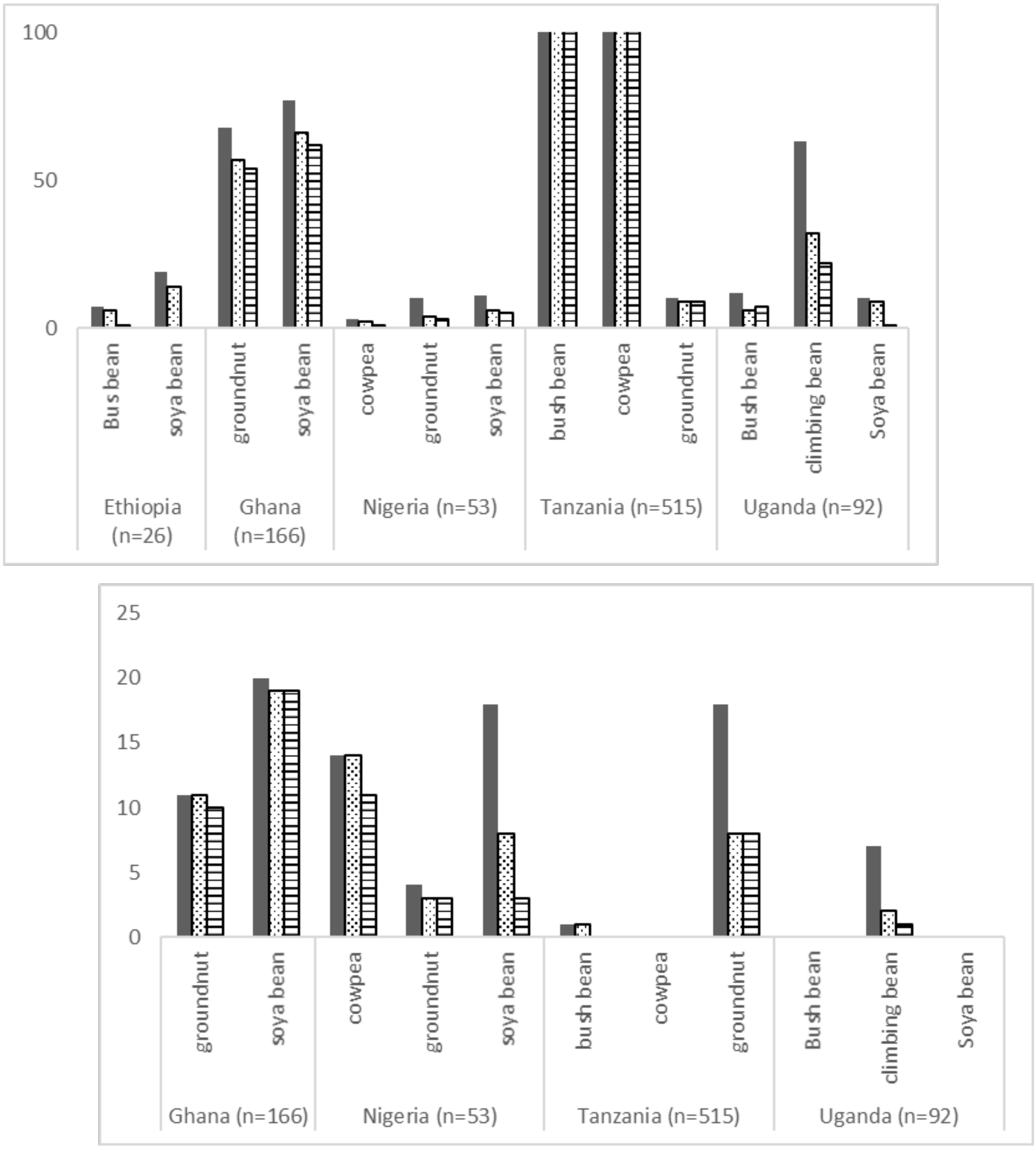

Figure 3. Number of adaptation trials for 2015(a) and 2016 (b). Solid bars are the total number of trials, doted bar are the trials with yield data for the n2 Africa plot and striped bars represent the trials with yield data for the farmers own plot. Ideally, all three bars should have the same height.

Overall, Information about the N2 Africa package, legume type, variety, and intercropping or row planting is complete, however, the amount of empty answers increases for the characteristics of the legume of the farmer and his practices (table 4) This information is important because it allows for possible the yield comparison between the current practices and the N2Africa proposed technology. Nakasaka (2016) observed that the extensions officer usually avoids asking some questions during the interview (to save time), especially the name of variety, the name of fertiliser and the farmer's own inputs/practices.

Table 4 Number of NA answer for questions about practices in the N2 Africa plot and in the farmer's plot (n=856)

Question in 2015-16 survey	N2Africa plot	Own Plot
row spacing	238	610
plant spacing	237	611
plant per hole (number)	238	603
Width of harvested plot	25	372
Depth of harvested plot	22	369
Grain weight	159	261
Legume species	0	101
Legume variety	0	103
Type_of_fertilizer	85	660
Practices	17	132

- *Incomplete data*

In general, questions that ask about past events have a low response ratio and the accuracy of the answer depends on the "farmers memory". There is also no answer when asking for the date of staking, herbicide and insecticide application, and when asking about inputs from more than one season ago the number of NA's duplicates (Table 5).

Table 5 Number of NA answer for questions about one (previous season) and two (season before previous season) seasons ago in the N2 Africa field (n=856)

Question in 2015-16 survey	NAs
crop_1_previous_season	82
other_crops_previous_season	373
type_of_mineral_fertilizer_previous_season	112
type_of_organic_fertilizer_previous_season	120
crop_1_season_before_previous_season	115
other_crops_season_before_previous_season	419
type_of_mineral_fertilizers_season_before_previous_season	148
type_of_organic_fertilizers_season_before_previous_season	161

In the original survey, information for a maximum of 4 fields was recorded, however, the only field with complete data is field 1 (the one where the N2 Africa plot is located) and the other fields have incomplete information (Table 6). Additionally, under the question about the main reasons for not having a good harvest (for legumes and no legume crops) there was no answer in any of the farms.

Table 6 NAs answers for questions about the field history for the field where the N2Africa pot is located (field 1) and the most important field for the farmer (field 2). (n=856)

Question in 2015-16 survey	NAs	
	Field 1	Field 2
area_field	0	0
walking_distance_field_	58	621
perceived_fertility_field	50	620
x1st_crop_field	73	186
variety_crop	61	176
seed_source	60	173
other_crops	423	752
type_of_mineral_fertilizer	134	651
proportion_sold_most_important_crop	87	191
crop_residues	113	211

- *Inaccurate or confusing data*

Farm area and unit show some inconsistencies, especially when the enumerator reported square meters as unit with values below 1. The absence of logical connexion between the value and the reported unit makes the information unreliable. Approximately 500 farmers (from 856) reported to suffer from some pest, disease or weeds, however, under the definition of “type of pest” or “type of disease” there was no clear boundary. Some diseases are reported under pests or the other way around (or the same is reported twice). Is important to clarify, especially with the enumerators doing the surveys, what is defined as pest and /or diseases to have more clear information on what is affecting the yield.

- *Other issues*

Some questions that had more than half of missing values or showed a lot of inconsistencies were chosen for further analysis to check if the information is relevant or informative. Some of them are listed below (table 7) with a brief discussion about the issues to choose them. These four variables were included in section 2 of this report in order to observe if they can give valuable information about the changes of the technology, the performance or the targeting (see section 3.1.2 of the results)

Table 7 Selected questions that were further analysed to add them (or not) in the 2017 survey

Question	***Issues***
Soil depth	According to the feedback from the technicians measuring soil depth is time and labour consuming during the interview. Almost half of the data of measured soil depth is missing (363 NA).
Famers perception of the field (drainage and fertility)	Since this variable is subjective it doesn't always give accurate agronomic information.
Farmer gender and age	Gender and age of the interviewed farmers is recorded correctly, however, 360 of the interviewed farmers are not the head of the household, and head of the household age is recorded incompletely and the gender is not recorded.
Education	At least 120 (out of 856) farmers mention to have "other" type of education, however it was never specified what "other" stands for.

3.2 Informativeness of data

3.2.1 About changes in the technology

According to the N2Africa learning loops it is important to know why (or why not) farmers are changing the legume technologies proposed by the project. This is a necessary step to translate the best bet solutions to best fit for the farmers. Most of the explanations regarding management changes are recorded in the survey as open questions. This reduces the amount of valid answers (table 8) and makes the analysis complicated and time consuming. Note that the form also contains information in the application of inputs, spacing and other planting and management practices, however, from the current data is not possible to know what was recommended, so not all changes can be tracked.

Table 8 Number of NA answer when asking the reasons for intercropping or row plating in each plot (total = 856)

Questions in the survey	N2 plot	Own plot
Reason for (not) intercropping	295	367
Reason for (not) row planting	147	218

With selected key words, some of the open questions were categorized in multiple answer questions and adapted in the new survey. The three questions adapted for the 2017 explain the farmer's reason for (or not) intercrop, row plant and choosing the package (the legume technology).

The chosen package could also be related with the way farms adapt and adopts the technology, since different motivations to try a new variety/legume/practice can influence the management decisions. When choosing the package (figure 4) there were four main reasons to do it: because it was instructed or simply provided by the N2Africa partners in the area, because there was a new variety/species and the farmers (most of the time) saw it in the demonstration trials and now want to try it [curiosity], because he/she thinks this option will have good yields or based on seed characteristic (is bigger, has a better shape, is more nutritious or has early maturity, among others). In the "other" reasons, according to Nakasaka (2016) some farmers report not having more options, since this was the package delivered by the extension officer.

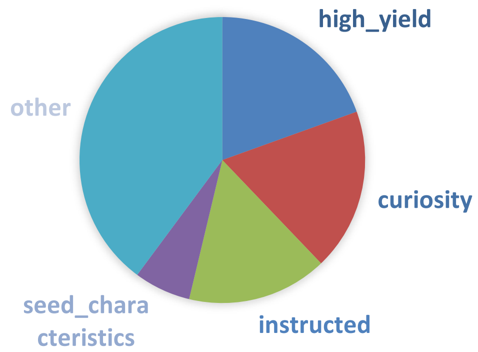

Figure 4 Proposed categories for the multiple answer questions "Why did you choose this package?" and proportion of answers from the 2015 and 2016 surveys.

Under the reasons for intercropping and row planting (or not) the answers were (Figure 5): Because it will lead to better yields, it was instructed by the N2Africa technicians to not intercrop or to plant in rows, because there was no more land available for the trial, so for default, it was intercropped, because it makes management easier. Some farmers choose to intercrop as an insurance, to have food security in case one of the crops failed [diversity]. In some cases, the farmers decided NOT to intercrop because they wanted to compare the yield of the legume/variety proposed by N2Africa with their own. Some farmers choose to plant in rows because they have always done it in this way or to avoid diseases.

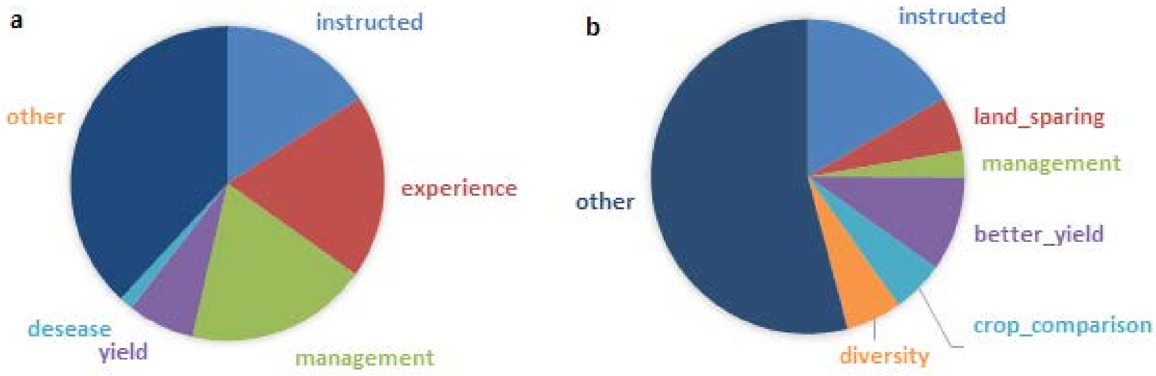

Figure 5. Farmers answer for the reasons for row planting (a) and intercropping (b)

3.2.2 About the performance of the technology

One of the biggest challenges of a big scale project like N2Africa is how to explain variation in yield and response (figure 6). Thanks to several agronomic trials it is known how biophysical and management factors can affect yield, however, there is still a high variation among farmers. Several linear mixed models were run to have an insight into which farm and household characteristics (tables 2 and 3) may be informative for understanding yield differences.

When analysing the effect of the N2Africa technology, not considering the country, treatment has a significant effect in yield and a significant interaction with the legume species ($P < 0.005$). Districts and farm ID were added as random effect, together, they account for 0.5 of the variation in yield (with 0.2 in the residual), which suggests that site plays a key role. Table 9 summarize the means for each legume per country for the N2Africa plot and the own plot. As mention before, the large variation inside a country is evident in the size of the bars in figure 6. It strikes the results of Uganda where soya bean and bush bean had higher yields in the own plot,

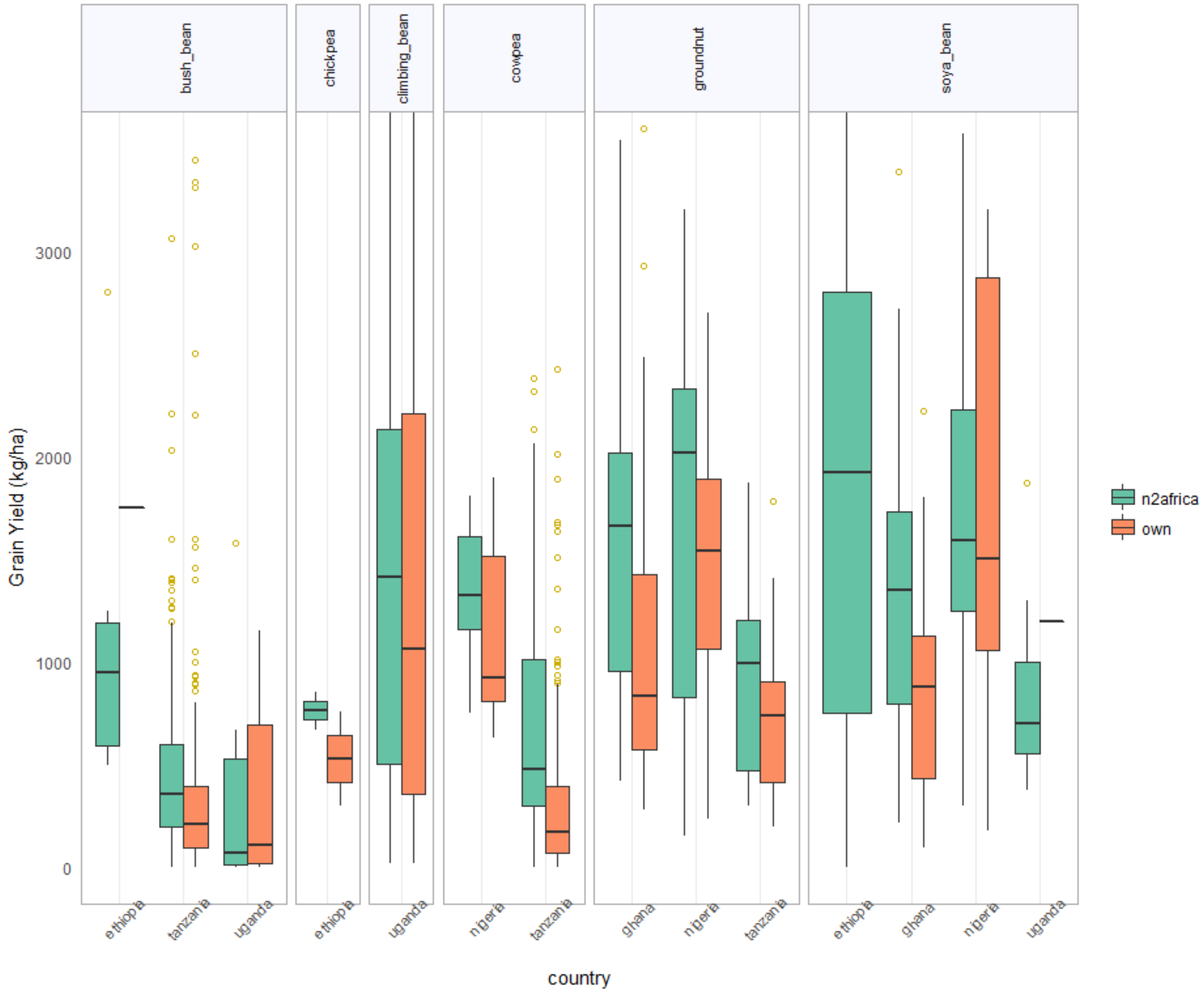

Figure 6. Variation in yield per legume and country**.** The black line represents the median.

Table 9 Means of the estimated grain yield per legume per country. The shadow rows represent the N2Africa trial, the white bar is the own plot. The stars indicate significant differences (p < 0.05) between the N2Africa plot and the farmers plot. The number on the left column for each country is the total number of trials(the same for both treatments), the number between brackets is the number of trials that reported yield data. Compiled data from 2015 and 2016

	Tanzania		Nigeria		Ghana		Ethiopia		Uganda		*Trials per legume*
Soya Bean	0		29	1763 (14)	97	1307 (85)*	19	1896 (14)	10	870 (9)	
				1773 (8)		860 (81)		NA		1200 (1)	155
Climbing Bean	0		0		0		0		70	1566 (34)*	
										1417 (23)	70
Groundnut	28	946 (17)	14	1671 (7)	79	1630 (68)*	0		0		
		756 (17)		1488 (6)		1028 (64)					121
Cowpea	172	687 (172)*	17	1302 (16)	0		0		0		
		330 (172)		1156 (12)							189
Bush Bean	315	498* (287)	0		0		7	1158 (6)	12	400(6)	
		552 (245)						1750 (1)		377 (7)	334
Trials per country	515		60		176		26		92		**869**

a. Farm characteristics

There is a significant relation (P = 0,009) between farm size and yield (figure 7), which has also been reported by several authors (Walker et al 2006, Fan et al 2013), however the R^2 is extremely low (0.02). The total number of livestock didn't show significant relation with yield. On the same way, the soil depth in the field (from 3 measured points) and the yield didn't show any strong correlation (p = 0.3897, R^2 = 0.004) The relationship between the perceived fertility and yield was only slightly significant for some cases but for most of them is even opposite to the description (figure 8). The situation was similar with the perception of drainage.

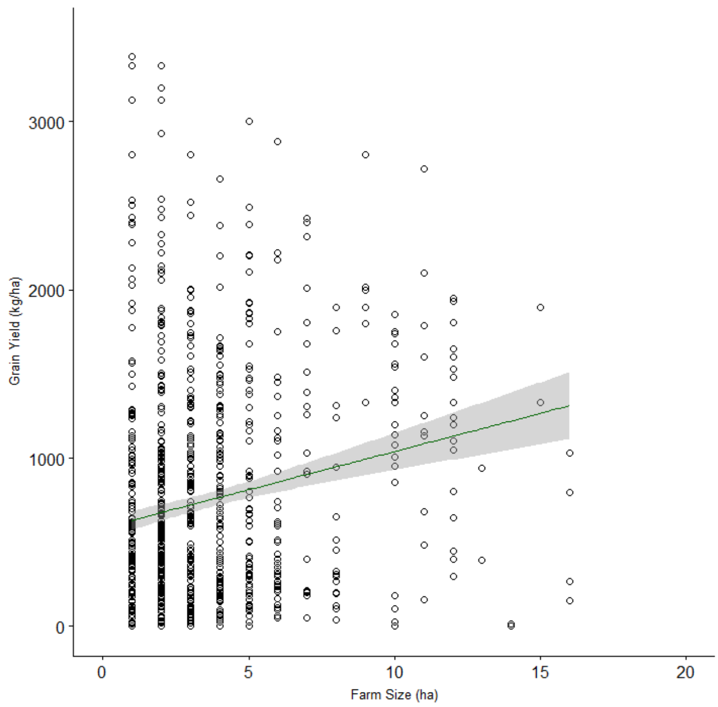

Figure 7. Relation between Grains Yield and Farm size. The line represents de linear regresion line (R^2 = 0.018) and the 95% confidence interval

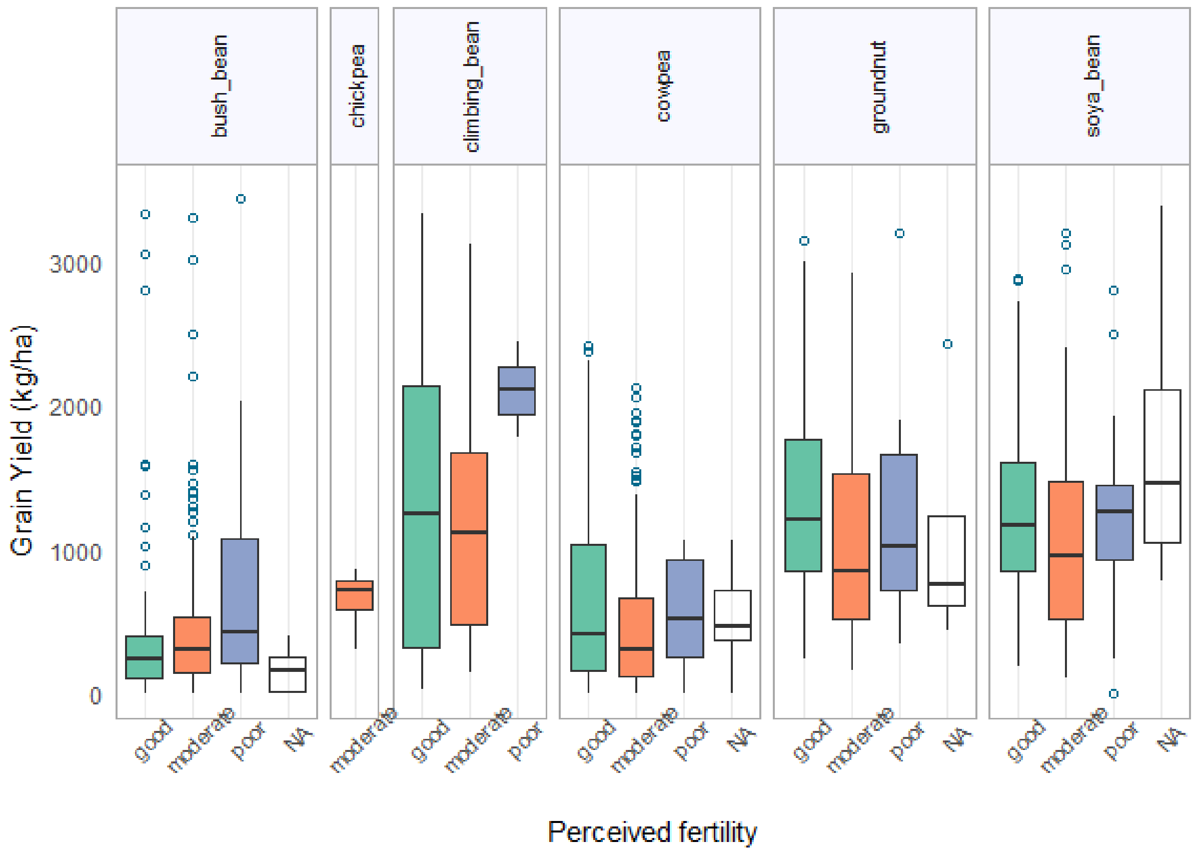

Figure 8 Farmers perception of fertility of the field where the n2africa plot is located and its relation with the estimated yield.

b. Management and inputs

There is significant difference ($p < 0.005$) in yield were the farmer weeds 3 times compared with only one or two weedings. However, there is no significant difference between weeding one or two times. Additionally, when comparing the effect of weeding in both plots (the N2Africa and the farmers own plot), weeding frequency has a significant effect in the yield of both plots, however, the effect is higher on the own plot (0.00007 vs 0.03). Figure 9 shows the predicted means for the interaction of the weeding frequency and treatment. It is possible to see how management can affect the end results of the trials.

Row planting presented significant differences ($p = 0.01$), with higher yields when the farmer planted in rows. Intercropping had a significant effect in yield when the interaction with treatment was considered ($P < 0.05$). For the own plot, the yield didn't differ between intercropped or not legumes, however, for the N2Africa plot, intercropping lead to significant lower yields (Figure 10). Either way, the yield in the N2Africa plot was always higher that the own plot. Intercropped is never part of the N2 Africa recommendations, but for different reasons (see section 3.2.1) farmers choose to do it (200 farmers). This result shows how farmers changes in the technology can affect the performance.

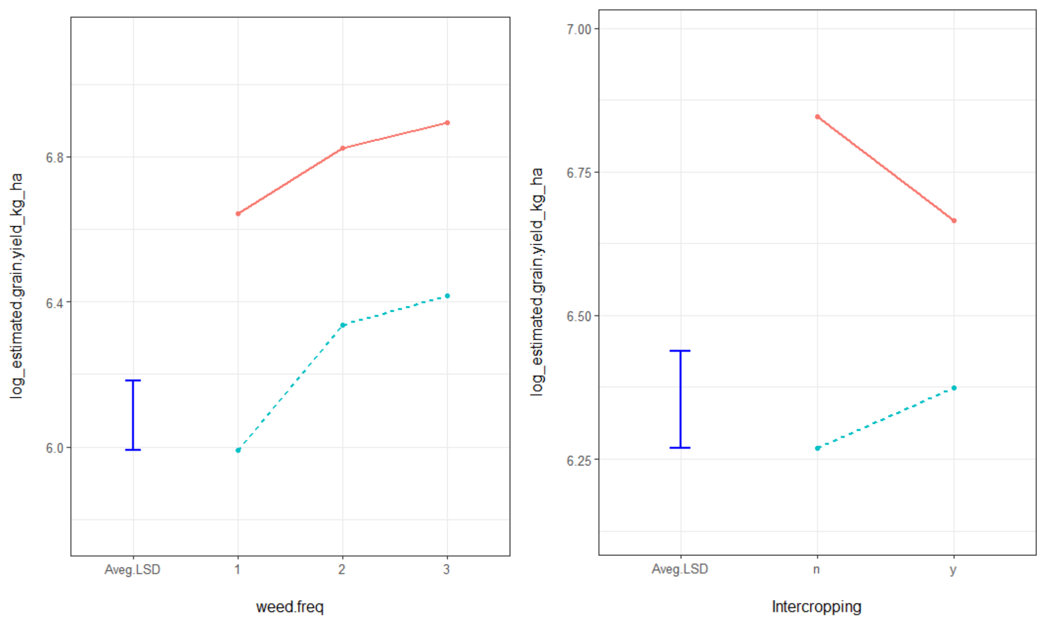

Figure 9. Relation between the weeding frequency (left) and treatment and the effect of intercropping (right) in grain yield. The solid shows the yield in the N2Africa plot, the dotted line the yield in the own plot

All inputs had significant effect in yield regardless the treatment (Figure 10). Considering the interaction between the treatment and adding or not inputs in each plot, all of them had a significant interaction, suggesting that the different plots responded different to the addition of inputs. The own plot is much more responsive (visible in the slopes in figure 10) than the N2Africa plot. Some farmers answer they did not apply any inputs (either herbicide, insecticide or fertilizer) (bottom right corner of the graph). strikes the increase in N2Africa yield even when no inputs were applied.

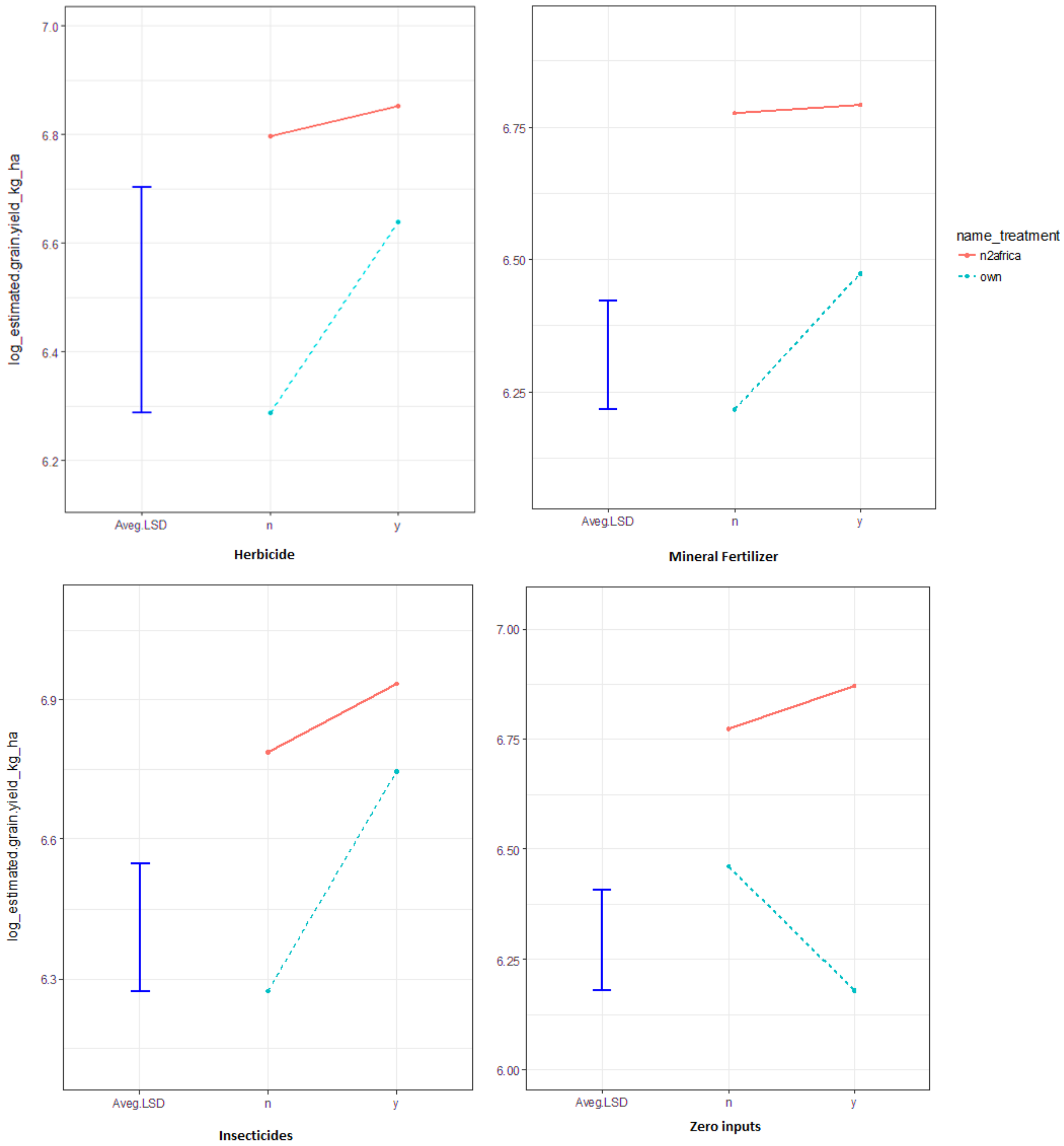

Figure 10 interaction of inputs and treatment and their effect on yield. The farmers were asked about all the inputs they applied in both plots, the data showed is a yes/no variable. Note that in "zero inputs" 'y', means that there were NO inputs, and "n" means, either fertilizer, herbicide or insecticide was applied.

c. Household characteristics

The gender of the farmer in charge of the N2 plot show a slightly significant ($p = 0.05$) influence in yield, with male farmers having higher yields. This has been reported before by Tittonell et al (2010) Additionally, there was a significant interaction ($P < 0.0005$) between farmer's gender and treatment. For the yield in the own plot, male farmers had higher yields compared to female farmers. Some authors have proposed that female farmers are more flexible when adapting new technologies (Mugi-

Ngengaa, et al., 2016) and even if the results don't show that, the N2Africa yield differs less between females and male farmers (compared to the own plot) (figure 11)

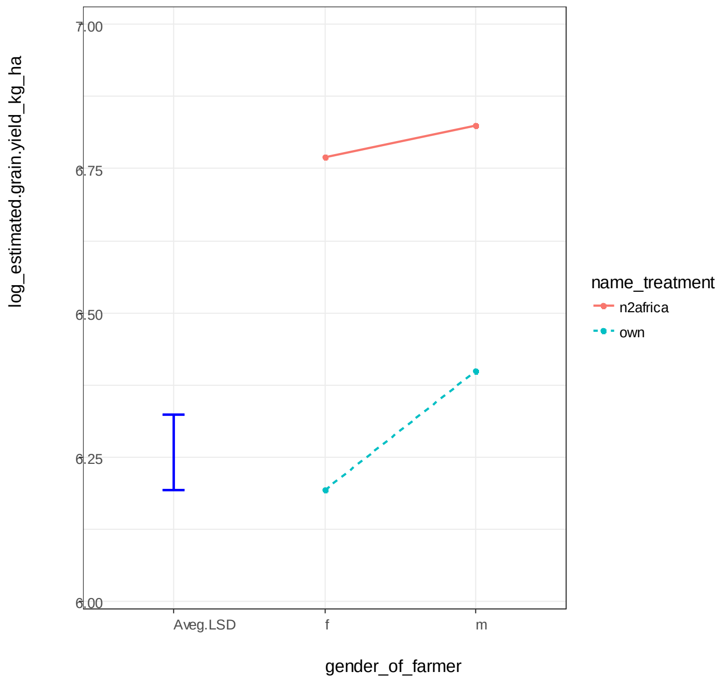

Figure 11 Relation between gender of the farmer and treatment and their effect in yield. Note that this is the gender of the interviewed farmer (the one in charge of the N2Africa plot), the gender of the head of the household was not recorded.

Neither the age of the farmer or the age of the household head showed a clear relation with yield (P > 0.1) (figure 5), however, some authors report that age of the head of the household is related with the grade of adoption of new technologies, which might not be related with the performance but with the changes and targeting of the proposed technologies (Mutuma, et al. , 2014; Mugi-Ngenga et al 2015). Family size didnt affect yield either.

There was significant difference in yield (P < 0.05) from farms where the highest level of education is post-secondary or university compared with households with primary or secondary education only (figure 12). As mentioned before, strikes the higher yields when the response is "other" type of education, and it will be interesting to know more about what formation are the farmers having. As a clarification, this is the highest level of education in the households which does not always match the education of the head of the household. The education of the head of the households didn't show significant relation with yield (P = 0.3). Interesting, there was also a significant relation (P = 0.003) between education and the yield when the interaction with treatment was considered. As show in figure 12, if a farmer has a higher education this will mean that the performance of the N2Africa technology will be even better (in terms of yield).

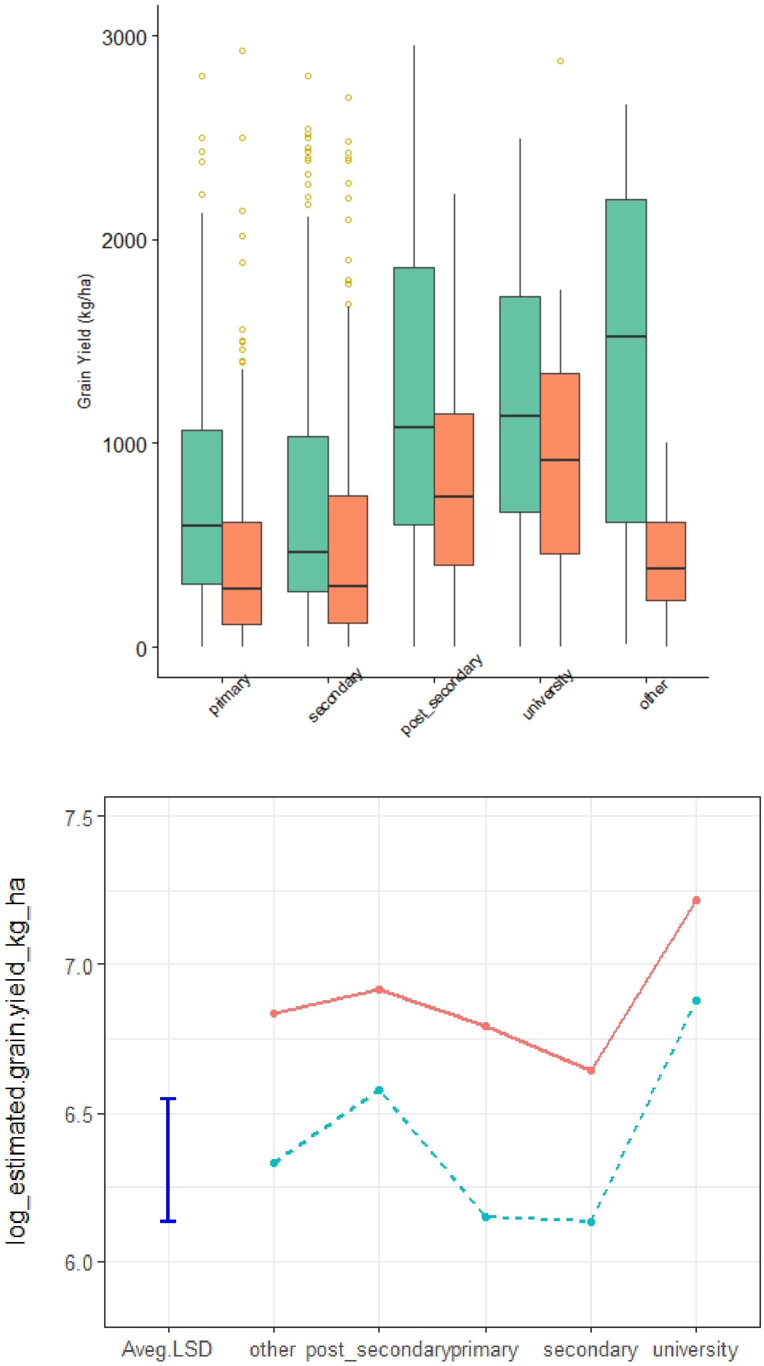

Figure 12 Relation between the highest level of education in the household and the yield. The green bars and red solid line represent the yield from the N2Africa plots; orange bars and blue dotted line the yield from the farmer's plot

The frequency on which farmers hired external labour influenced yield (figure 13). There is a significant difference (P < 0.005) in yield between farmers that never hired labour and the ones that do it sometimes, regularly or permanently. How often a farmer hired labour can be related with the family labour available or with having or not money to pay someone else to work on their farm.

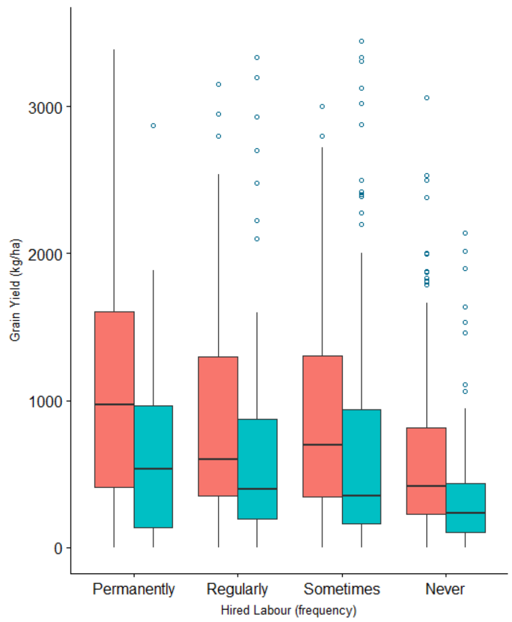

Figure 13 Frequency on which farms hired labour and its relationship with yield. Permanently (i.e. every year, throughout the cropping season); Regularly (e.g. at peak periods during the cropping season); Sometimes (e.g. not every season or peak period, only if money allows). Light bars represent the yield from the N2Africa plot.

Based on the results, it seems that the product orientation also influences the yield. In figure 14 it is visible that farmers who get most of their income from off-farm sources have significant (p = 0.013) lower yields than farmers that get most of it from the farm. Additionally, when comparing different sources of income, there are slightly significant differences (P = 0.059) among farmers that are focused on crops (or get more income form them) and the ones focus on livestock. How important agriculture is in the household (in terms of share of income) might be related to how much labour, time and money is invested in it, which will be reflected then in higher yields. The frequency of food shortage or the number of months didn’t show significant relation with yield, either if other reports have found an important link between nutrition and health and agricultural productivity (Fan, et al., 2013)

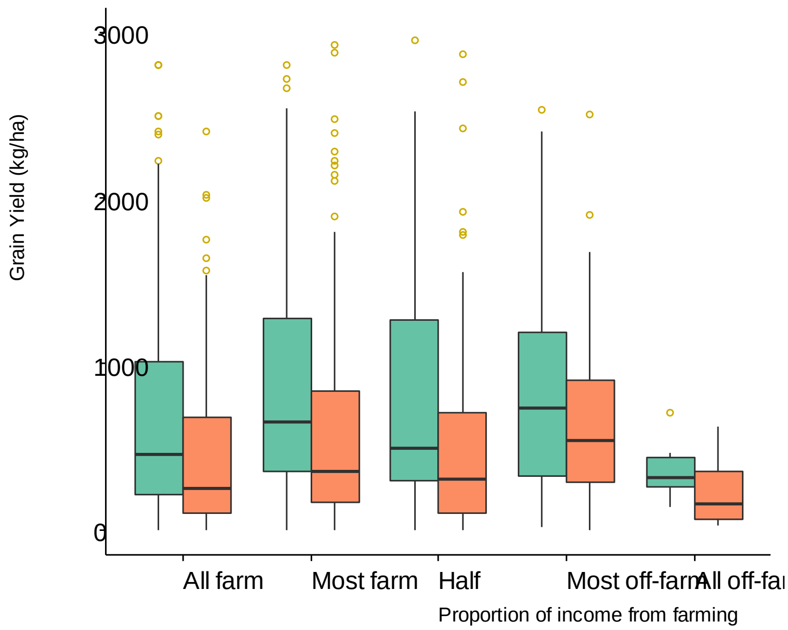

Figure 14. Relation of the proportion of income from farming and yield. The green bar represents the n2africa yield, the blue dots represent the mean.

3.2.3 Targeting the technologies

For Tanzania, the two principal components explained 23 and 14 percent of the variation. The variables with high score in the first principal component were TLU and the proportion of income from farming. Followed by family size in the second component (14 %). According to the results, farm size is closely related with the total amount of livestock, and these two are related with age of the head of the household (figure 15d). When comparing the different clusters obtained from the scores of the PCA, farm size, TLU, proportion of home consumption and income from farming show high significant differences. Additionally, farm size, TLU and home consumption had a significant interaction with district, suggesting that these two factors (the size of the farm and the production orientation) are defined by the location. Additionally, home consumption is closely related with hired labour, suggesting possible differences on resource endowment (figure 15d).

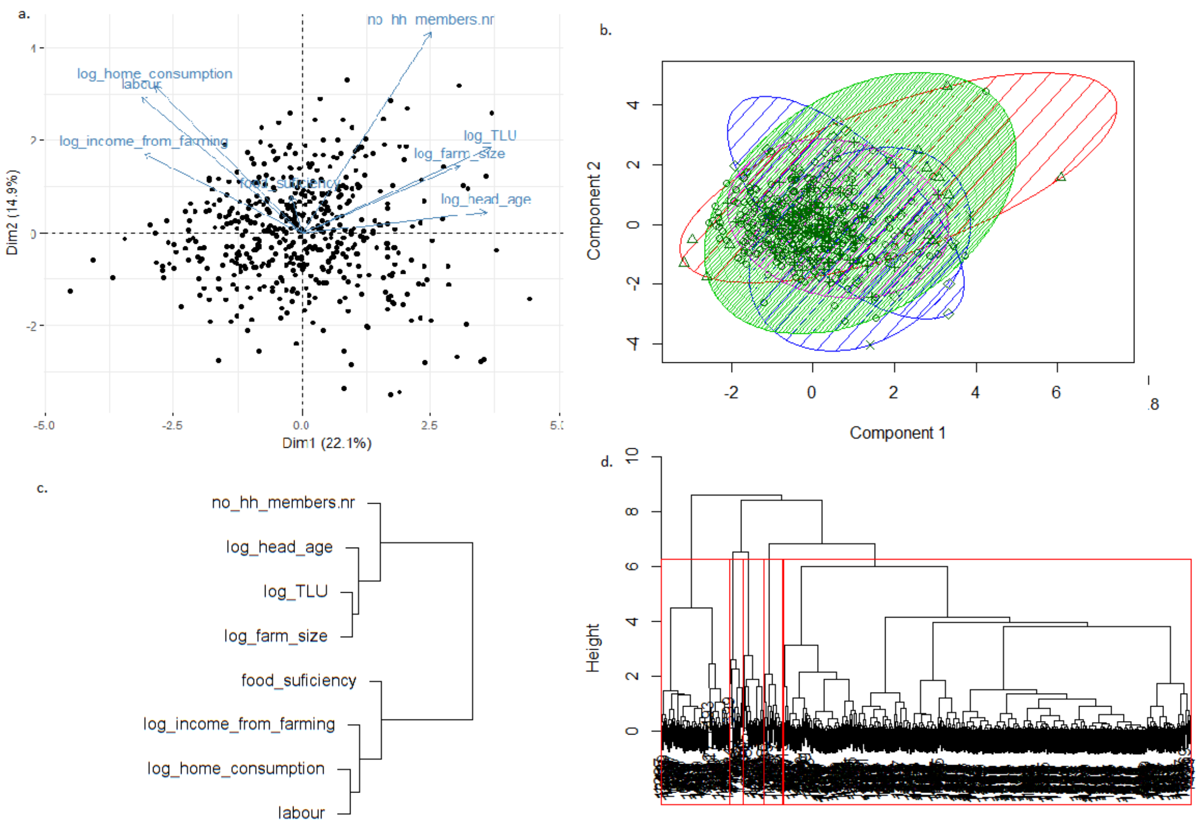

Figure 15 Results from the PCA and the cluster analysis for Tanzania. a. Biplot showing the distribution of the farmers based on the variable analysed and their relationship with the first two PC. b) Resulted clusters in relation with PC1 and PC2. c) Dendrogram for the clustering of variables following the same method d). Dendrogram for the clustering of farms based on the Euclidian distance

For Tanzania, there were significant differences between clusters for family size, number of month with food sufficient, farm size, hired labour and the proportion of income from farming. Yield didn't show any significant differences. In figure 16 is possible to see the differences among the define clusters, there is a lot of variability within cluster for most of the variables, which makes hard to delimited clear farm types. In agreement with the PCA results, Family size differ between groups and its slightly proportional with the farm size and the total number of livestock. In the N2Africa baseline report it was conclude that in Tanzania, even if most of the farms have income off-farm crops are still the most important source of income. The mean TLU for the country was reported as 2.83, higher that the ones found in the adaptation surveys, additionally, 30% of the farmers indicated they could not hire enough people to weed their own fields which is visible in the 2 group, where most households declare to hired labour only if money allows.

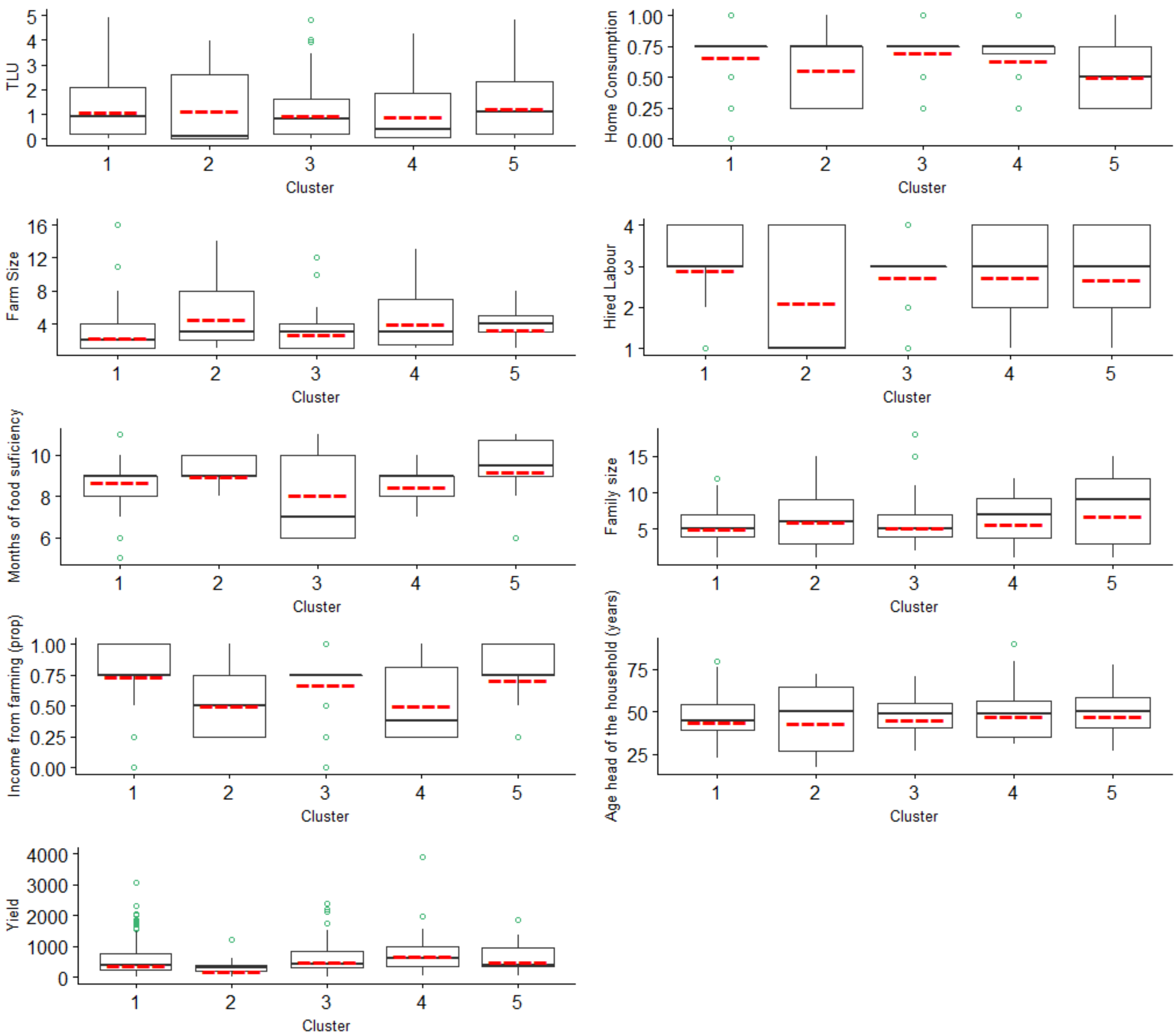

Figure 16 Differences for several farm characteristics between the defined clusters for Tanzania. The dark line represents the media and the red dotted line the mean. Number of farmers per cluster is: 1:397, 2:13, 3:67, 4:20 and 5:19. Note that the data used for the plot is the original, non-imputed data)

Table 10 Number of farmers per cluster per variable for Tanzania. On the left, the number of farmers per cluster if the data is imputed.

		Variables								
Cluster	**Imputed**	Farm size	Hired labour	TLU	Food sufficiency	Family size	Income from farming	Farmers age	Home consumption	Yield
I	**397**	226	395	397	263	397	384	310	391	364
II	**13**	11	13	13	9	13	13	11	13	13
III	**67**	41	67	67	67	66	67	54	67	61
IV	**20**	11	20	20	16	20	20	19	20	19
V	**19**	13	19	19	10	19	18	14	19	19

The sample for Ethiopia was quite small (26 farmers) an as it is clear in figure 17 and table 11. Tthe distribution of the farmers along the first principal components doesn't show any pattern, which will make the clustering rather arbitrary. However, the variables with higher loading were home consumption and TLU for the first component and family size for the second one.

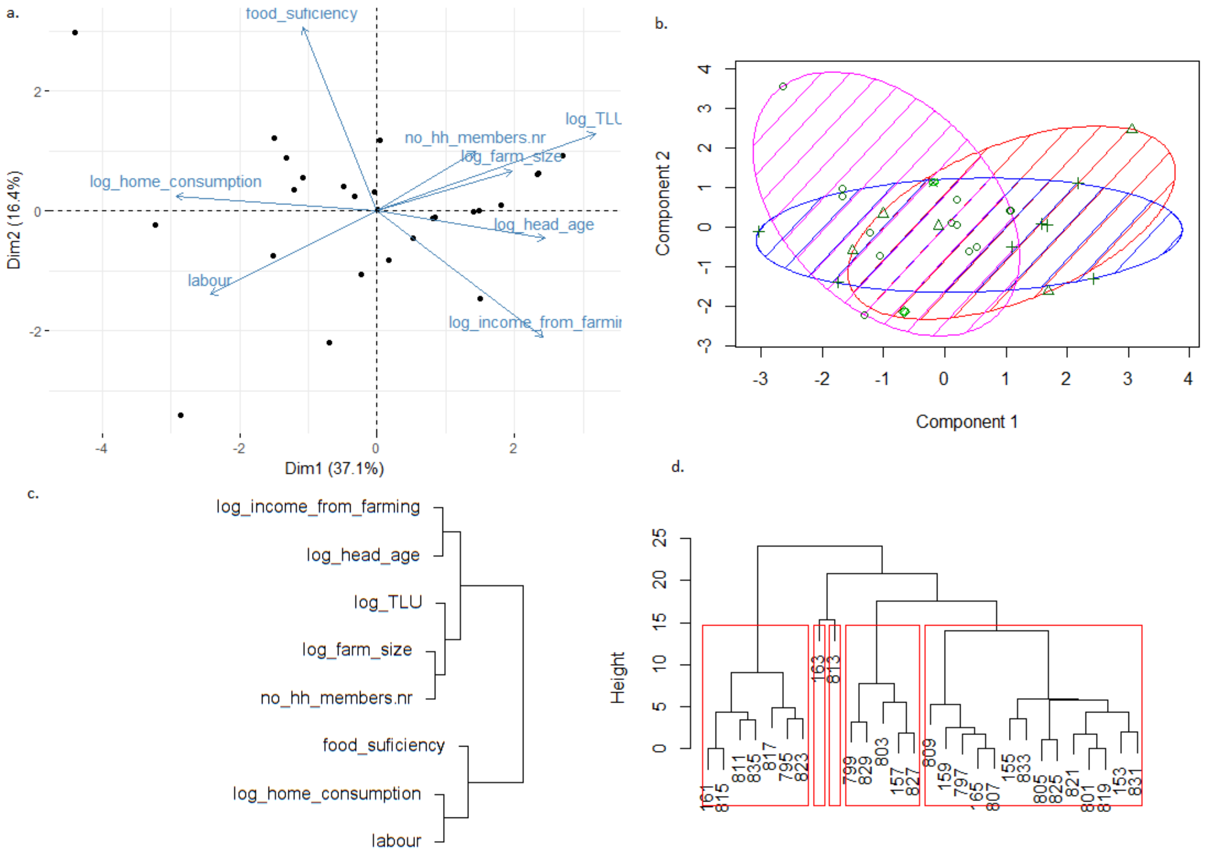

Figure 17 Results from the PCA and the cluster analysis for Ethiopia. a. Biplot showing the distribution of the farmers based on the variable analysed and their relationship with the first two PC. b) Resulted clusters in relation with PC1 and PC2. c) Dendrogram for the clustering of variables following the same method d). Dendrogram for the clustering of farms based on the Euclidian distance

Table 11 Number of farmers per cluster per variable for Ethiopia. On the left, the number of farmers per cluster if the data is imputed.

		Variables								
Cluster	**Imputed**	Farm size	Hired labour	TLU	Food sufficiency	Family size	Income from farming	Farmers age	Home consumption	Yield
I	**14**	13	14	14	2	14	13	2	14	10
II	**5**	5	5	5	2	5	5	1	5	4
III	**7**	6	7	7	2	7	7	1	7	5
IV	**1**	1	1	1	0	1	1	0	1	1
V	**1**	0	1	1	1	1	1	0	1	1

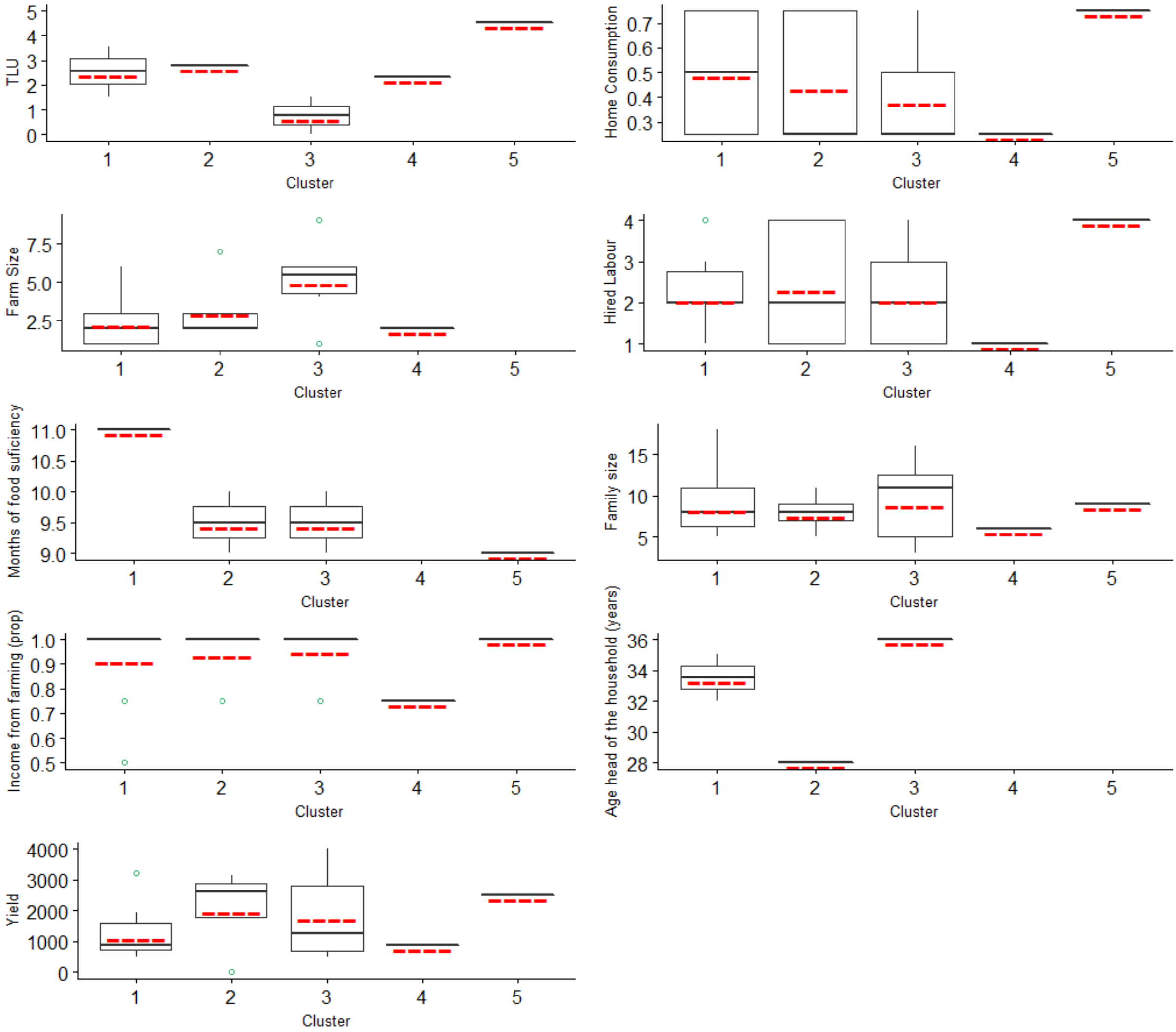

Figure 18 Clusters for Ethiopia, 1:14, 2:5, 3:7, 4:1 and 5:1

For Nigeria, the first component compresses 21% of the variability and 20% de second one. Farms size and the age of the head of the household had the higher scores in the first PC (figure 19). These two variables, according to the results, are close related. For the second component, labour had a significant higher score that the other variables. Similar as Tanzania, labour is related with the proportion of home consumption. The results from the clustering showed significant differences for family size, farm size and proportion of home consumption. Family size and months of food sufficiency showed significant interaction with district.

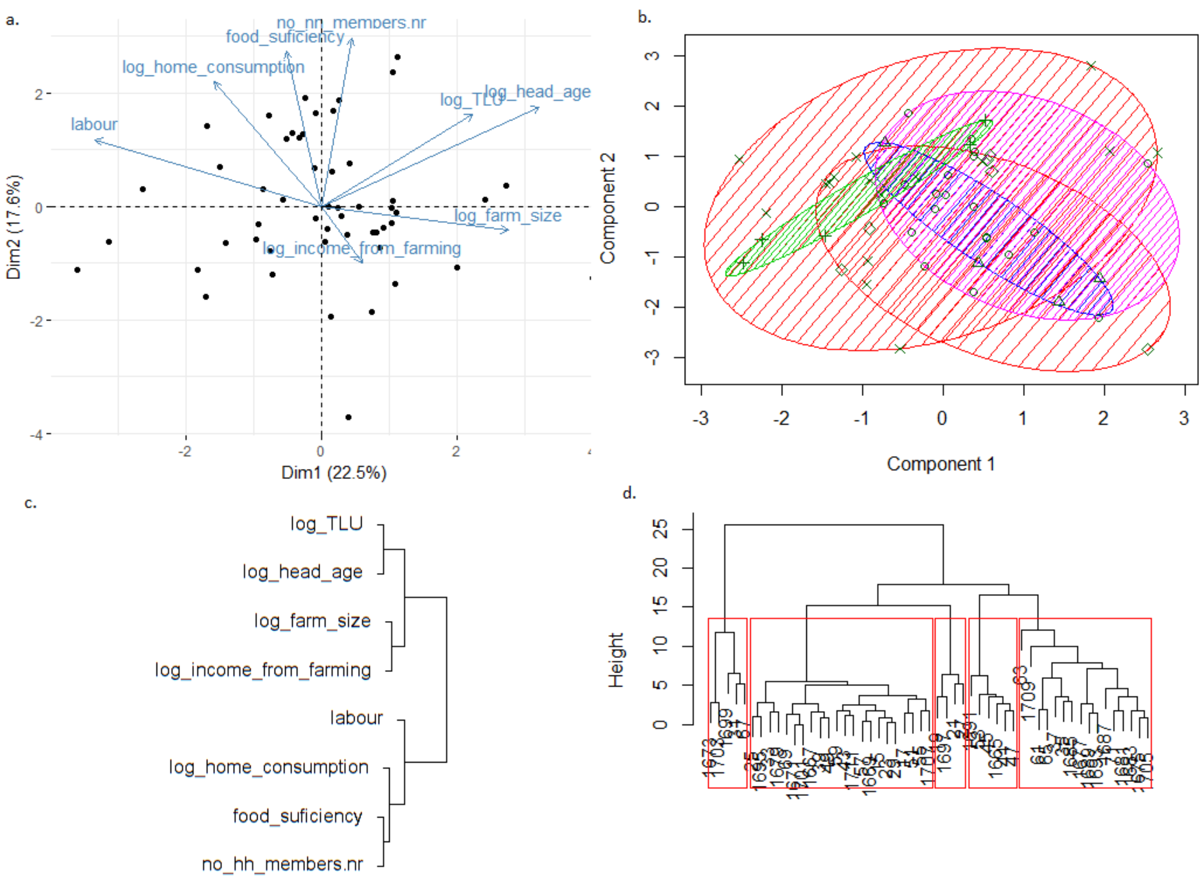

Figure 19 Results from the PCA and the cluster analysis for Nigeria. a. Biplot showing the distribution of the farmers based on the variable analysed and their relationship with the first two PC. b) Resulted clusters in relation with PC1 and PC2. c) Dendrogram for the clustering of variables following the same method d). Dendrogram for the clustering of farms based on the Euclidian distance.

Table 12 Number of farmers per cluster per variable for Nigeria. On the left, the number of farmers if the data is imputed.

Cluster	**Imputed**	Variables Farm size	Hired labour	TLU	Food sufficiency	Family size	Income from farming	Farmers age	Home consumption	Yield
I	**22**	7	19	22	6	17	19	5	19	12
II	**4**	1	4	4	3	4	4	3	4	3
III	**5**	0	5	5	4	5	5	3	5	4
IV	**16**	8	15	16	10	14	15	7	15	8
V	**6**	4	6	6	3	5	5	5	6	3

Only farm size and proportion of home consumption show significant difference between clusters. From the baseline surveys, Nigeria was the country with higher number of farmers (6%) reporting livestock as their main source of income and group two presents relative high TLU compare with the other countries (figure 20) . This group is also characterized with smaller families probably related with younger farmers also involved in other source of income (similar as group 1). Groups 3 and 4 seem to be more farmed oriented but with big families. In Kano state, farmers are reported to be larger than in other regions of the country, in the same district, a big percent of the farmers rely in off farm source for their main income.

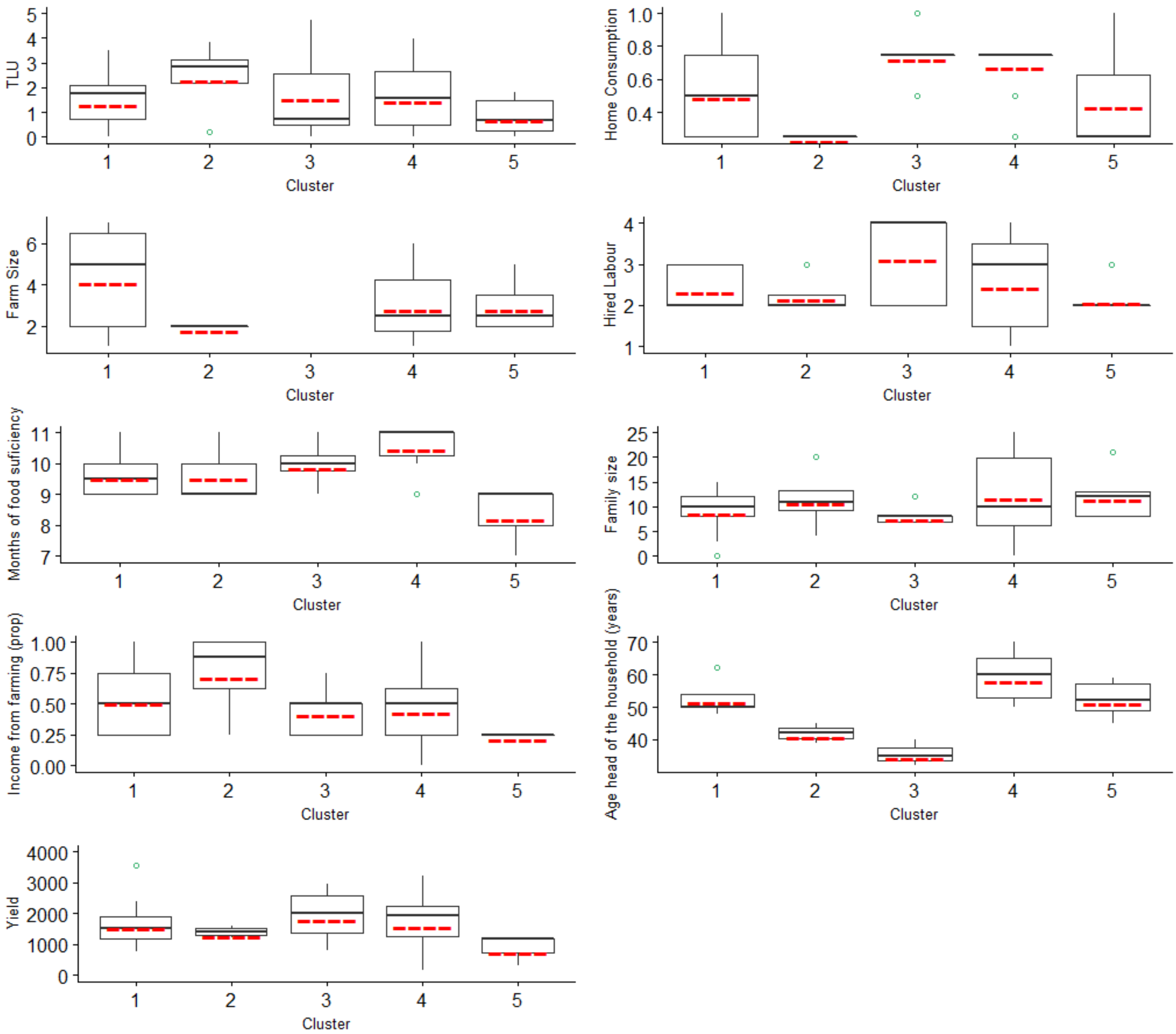

Figure 20 Differences for several farm characteristics between the defined clusters for Nigeria. The dark line represents the media and the red dotted line the mean. Number of farmers per cluster is: 1:22, 2:4, 3:5, 4:16 and 5:6. Note that the data used for the plot is the original data.

In Uganda, home consumption had high score for the PC1 (21% of the variability) followed by the age of the household in PC2 (16%). In this case, home consumption is related with the frequency of hired labour, month of food sufficiency and farm size. The results from the cluster showed significant differences only for proportion of the income from farming which is related with the number of members of the household. Proportion of home consumption has significant differences for the district suggesting that the main differences among farmers in Uganda is the market orientation and that is also related with site.

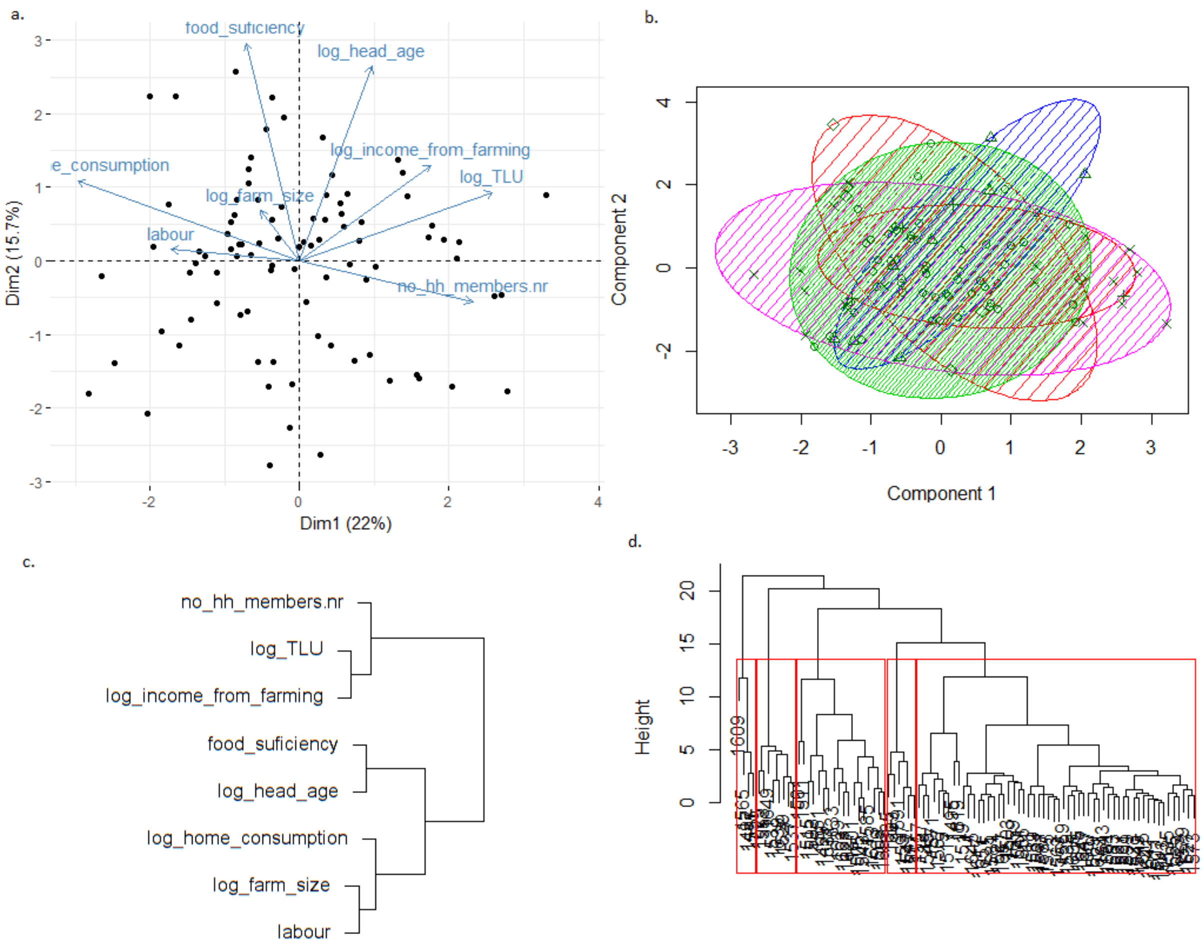

Figure 21 Results from the PCA and the cluster analysis for Uganda. a. Biplot showing the distribution of the farmers based on the variable analysed and their relationship with the first two PC. b) Resulted clusters in relation with PC1 and PC2. c) Dendrogram for the clustering of variables following the same method d). Dendrogram for the clustering of farms based on the Euclidian distance.

In the initial clustering, TLU and proportion of income from farming showed significant differences. Group 3 encloses the farms with more livestock but at the same time with smaller farms, according to the home consumption it seems this group is more market oriented, also because most of their income comes from farming. In the baseline survey, the results were that half of the crops is used for sale and only one third for home consumption. Uganda also shows smaller farm areas when compare with the other countries, as seen in figure 22. Groups 1 and 2 are similar, but they differ in family size, wich probably influences the amount if labour they hired.

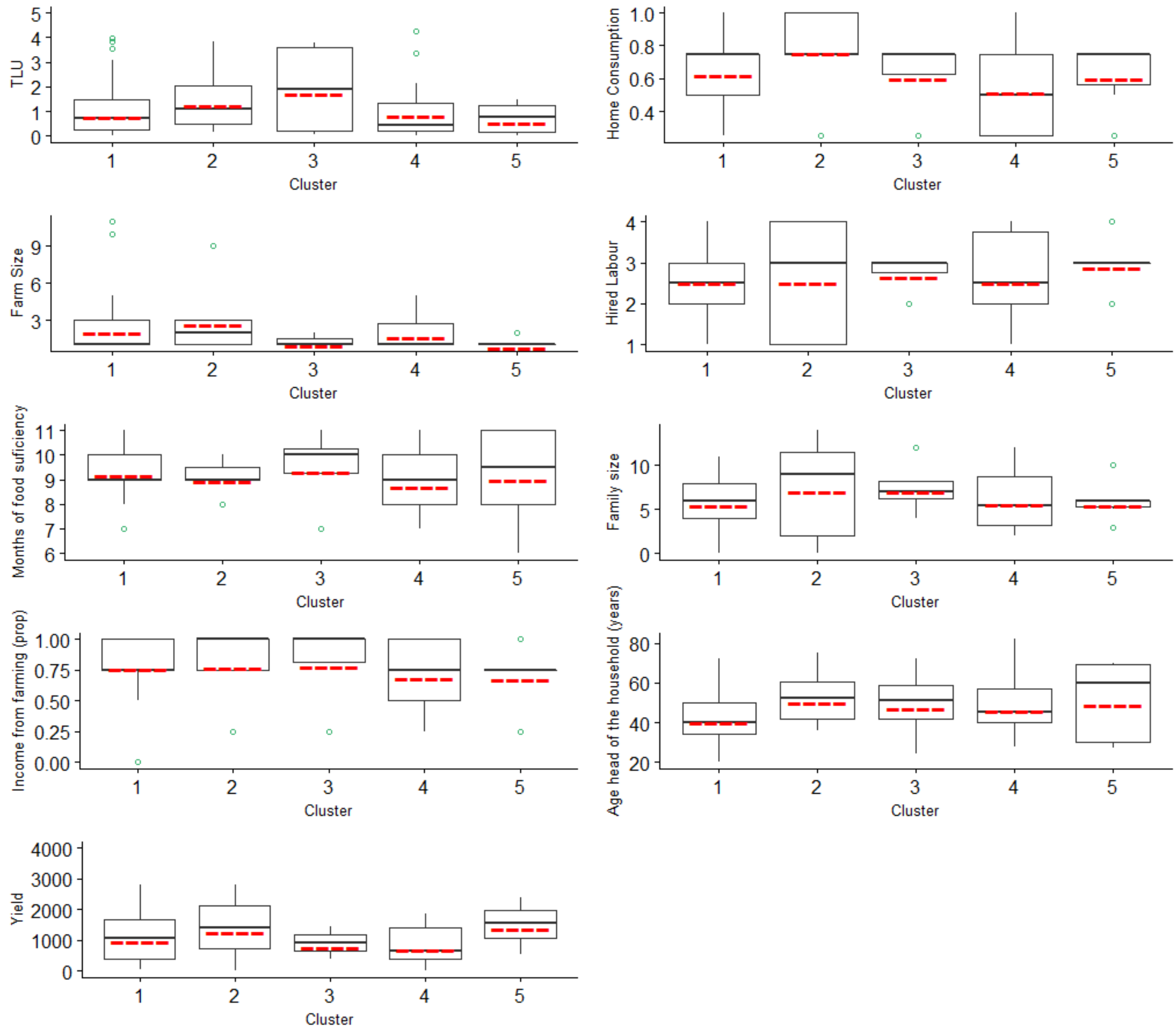

Figure 22 Differences for several farm characteristics between the defined clusters for Uganda. The dark line represents the media and the red dotted line the mean. Number of farmers per clusters is: 1:56, 2:8, 3:4, 4:18 and 5:6. Note that the data used for the plot is the original data.

Table 13 Number of farmers per cluster per variable for Uganda. On the left, the number of farmers per cluster if the data is imputed.

		Variables								
Cluster	**Imputed**	Farm size	Hired labour	TLU	Food sufficiency	Family size	Income from farming	Farmers age	Home consumption	Yield
I	**56**	49	50	56	43	53	48	47	50	27
II	**8**	6	8	8	7	8	5	7	8	4
III	**4**	3	4	4	4	4	4	4	4	2
IV	**18**	18	18	18	11	18	18	16	18	12
V	**6**	6	6	6	6	6	6	5	6	6

Finally, for Ghana, TLU and the proportion of income from farming show high loading in the PC1 (22%) while farm size for PC2 (18%). As in the other cases food sufficiency was related with home consumption and income form farming. TLU, as also found in other countries, is close related with farm size. The result from the cluster were significant for food sufficiency, farm size, TLU and home consumption and proportion of income from farming, all of them showing also a strong interaction with district.

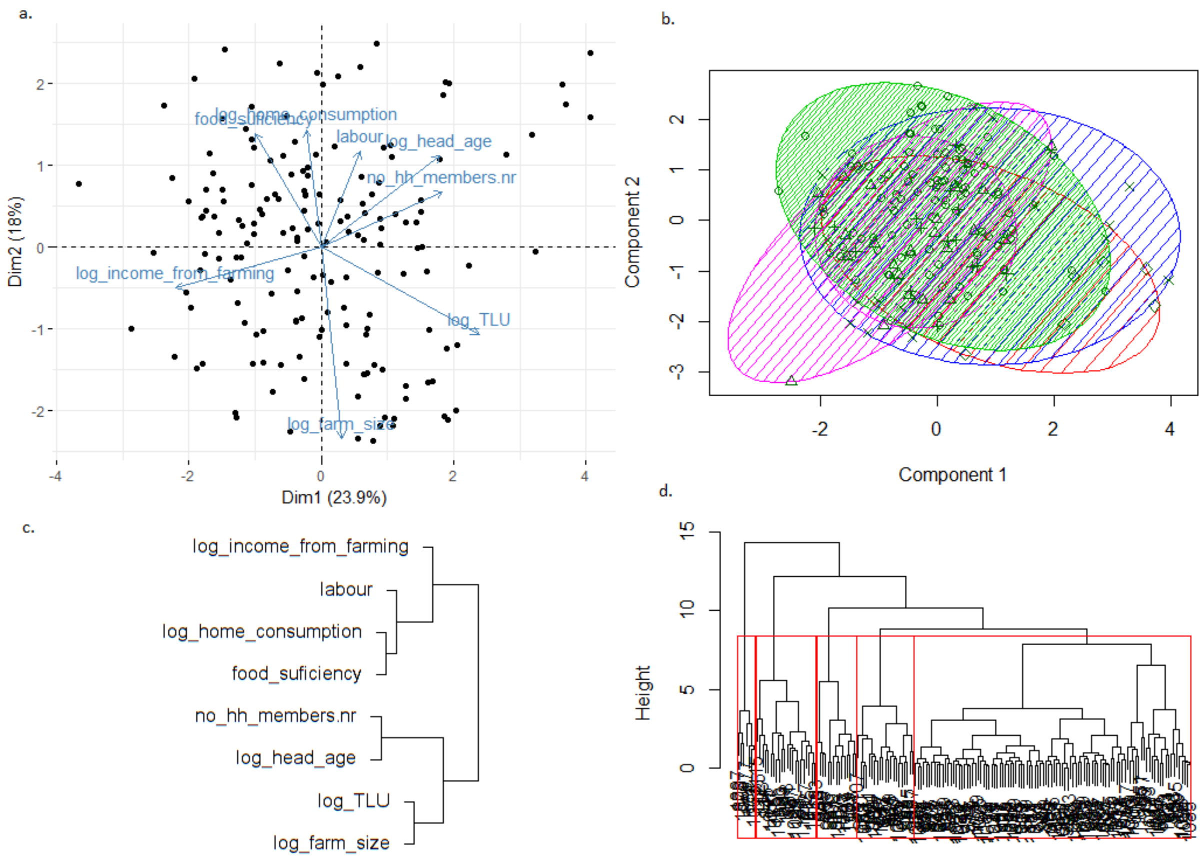

Figure 23 Results from the PCA and the cluster analysis for Ghana. a. Biplot showing the distribution of the farmers based on the variable analysed and their relationship with the first two PC. b) Resulted clusters in relation with PC1 and PC2. c) Dendrogram for the clustering of variables following the same method d). Dendrogram for the clustering of farms based on the Euclidian distance.

For Ghana Family size, Farm size, TLU, labour and proportion of income from farming show significant differences. Consistently, farm size shows significant differences between clusters, visible in the smaller farms of cluster 5 compared with the others. According to the baseline survey at least 90% of the household in Ghana rely in cropping as their main source of income but around 30% of women and men older than 17 are involved in off-farm income generation. This can be also related with the fact that at least 70% of the farmer hired labour for agricultural activities.

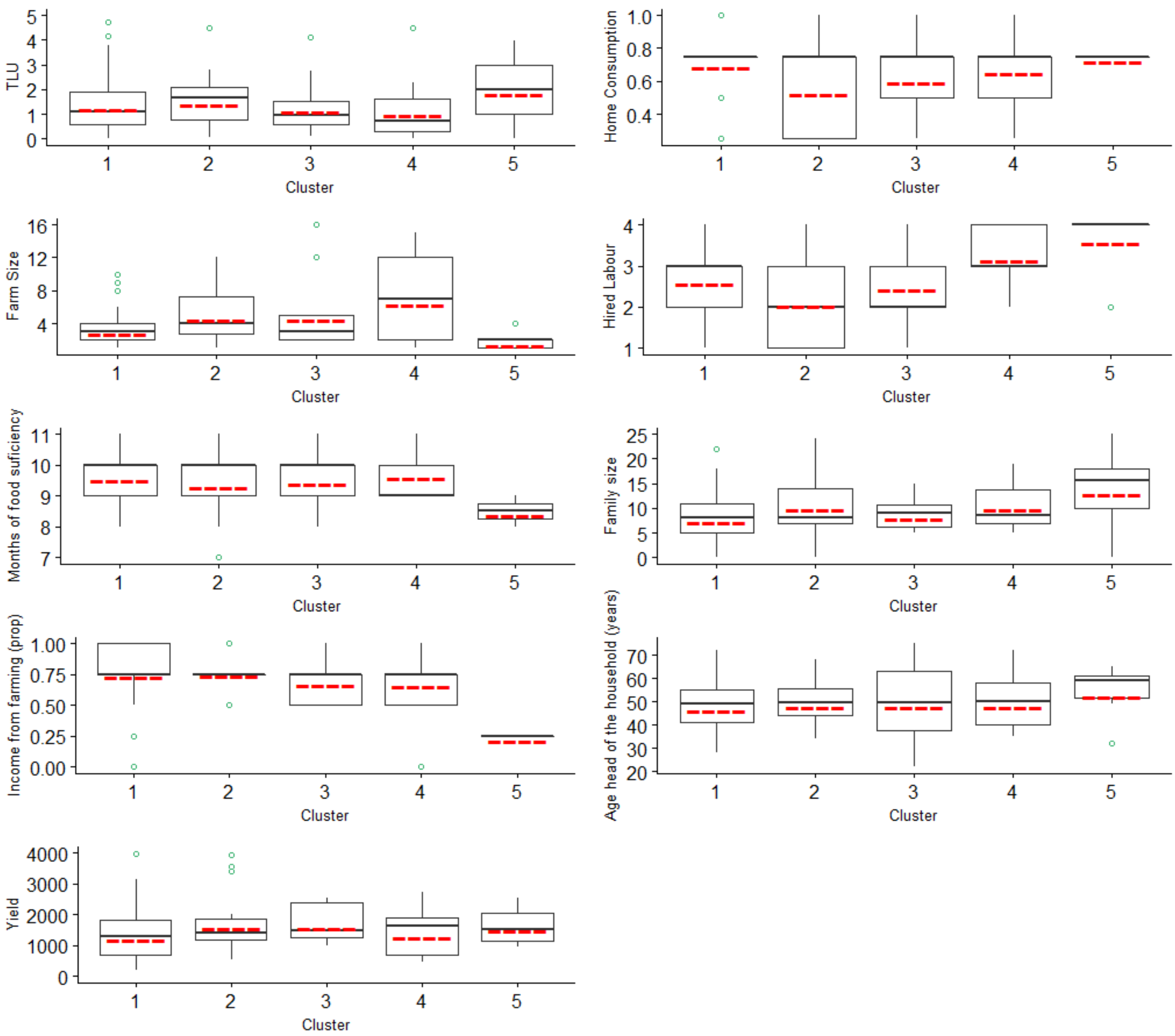

Figure 24 Differences for several farm characteristics between the defined clusters for Ghana. The dark line represents the media and the red dotted line the mean. Number of farmers per clusters is: 1:102, 2:22, 3:15, 4:21 and 5:7. Note that the data use is the original data.

Table 14 Number of farmers per cluster per variable for Ghana. On the left, the number of farmers per cluster if the data is imputed

Cluster	**Imputed**	Variables: Farm size	Hired labour	TLU	Food sufficiency	Family size	Income from farming	Farmers age	Home consumption	Yield
I	**102**	87	95	102	82	99	91	71	95	85
II	**22**	20	22	22	17	22	18	14	21	22
III	**15**	15	15	15	11	14	15	12	15	10
IV	**21**	20	21	21	17	20	21	17	21	20
V	**7**	7	6	7	2	6	6	6	6	6

After the clustering, significant differences for the household characteristics were analysed between clusters and district. Table 15 indicates in which cases the results were significant. The results from the cluster analysis are discussing further in section 4.2.

Table 15 Summary of the results for each variable and the clustering*district. the shadow row represents the re lation with cluster, the light rows the relation with district. A thick square symbolises a significant interaction be tween cluster and district. (Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1)

	Tanzania	*Ethiopia*	*Ghana*	*Uganda*	*Nigeria*
Family size	**		**		
	***			*	*
Months of food sufficiency	*				
	.		***	*	**
Farm size	**	.	***		*
	***		***		*
TLU			***	.	
	***		***	*	*
Proportion of home consumption	.			**	.
	***	**	***	***	**
Age of the head of the household	.			.	
	*		*		
Hired labour	*		**		
	.	*	***	*	*
Proportion of income form farming	***	.	***		*
	*	*	***	***	*
Yield	.				
	***		***		*

3.3 Changes in the survey

Table 16 summarizes the most significant changes in the focal adaptation survey. The main goal of the updated version was to make it short and clear. A lot of conditional sections were added, which means that some questions appear in the table based on the answer to previous one. I this way only relevant information for each farmer will be recorded and the time using the app will be relative to the amount of information wanted.

Table 16 Summary if the questions changed or removed from the Focal Adaptation survey

Question in 2016 survey	**Changes for 2017 survey**	**Reasons**
Package code	Removed	The code was not uniform among partners, now, a detail characterization of the package content is record
legume in the package, variety and inputs	Choices are provided conditioned to the region and the legume choose	To ensure the correct selection, only legumes distributed in each country can be selected for this answer.

Plot information	Conditional, the farmers report which plots he/she planted (N2Africa and control) and respective questions are available for each.	If the farmer doesn't have a control" plot, the survey will be only used for farm characteristics, bus not for performance.
Yield and information of non-legume crops	Optional to the technician. It is possible to record information of maximum two crops when the legume is not as sole crop.	In case information of other crops is wanted. Making it conditional, the technician can select to "save time" only recording the legume information. Easier[faster] surveys might ensure more complete answers in important topics
Soil depth	Removed	Time and consuming and not giving valuable information
Perceived fertility, drainage and slope	Removed	Not giving valuable information
Age and gender of the head of the households	A conditional was created, so if the interviewed farmers is NO the head of the household new questions will be available	To ensure that the information of the head of the household is always recorded
Information about other fields (largest field and field with the most important legume)	Only information about the field were the N2Africa plot is located will be recorded in the guidelines, is explain that the control plot should be in the same field	For the adaptation trials purpose, information of other fields is not relevant and it was time consuming
Inoculant questions	Removed	Not trustful information was recorded
Reasons for choosing the package	Multiple options answer (see table 10)	For easier analysis and homogenization of the answers
Reasons for intercropping and row planting		
Sources of income (Until nine different sources and corresponding ranking)	Selection of the three main sources of income	For easier analysis and answer.
Ownership of the field	Removed	Not relevant information (for the Adaptation trials)
NEW	Total amount of fertiliser used across all fields	For estimation of expenses

Table 17 summarizes the options added in the survey for the question bout reasons for choosing the package and intercropping and row planting. Approximately half of the responses from 2015 and 2016 fit in one of the new categories proposed (Figures 5 and 6).

Table 17 Introduced options for three questions in the 2017 survey.

ODK code	Farmer's answer
Reasons for choosing the package	
High Yield	High yield
Curiosity	Curiosity
Instructed	Because of advice or instructions
Seed Characteristics	Because of seed characteristics (colour, shape, size, better quality, taste)
Other	Other reasons
Reasons For Intercrop (Or Not)	
Instructed	I follow advice/instructions
Land Sparing	I have (or don't have) land available
Management	Is easier to manage
Better Yield	To have more yield (or better performance)
Crop Comparison	To see performance of the variety/species (to learn)
Diversity	To have different crops (e.g. In case one fails)
Other	Other
Reasons For Row Planting (Or Not)	
Instructed	I follow advice/instructions
Experience	I usually do it like this, for personal choice
Management	For easier management (planting, weeding, fertilizing etc.)
Yield	For higher yields (or to compare)
Disease	To prevent diseases
Other	Other

In addition to updating the form and type of questions in the ODK form, the guidelines for the adaptation trials were also updated in order to give clear instruction to the N2Africa technicians on how and what information record. The new guideline (Annex 2) highlights important parts of the process, has details step by step explanation on how to set up the N2 Africa plot and the own pilot and includes dynamic tree decisions to guide de technicians to the right type of measurements defined by certain conditions.

4. Discussion

Surveys are useful tools for collecting reliable information in agronomic projects that involve farmer participation. They are especially important in the N2Africa agronomy plan since there is an emphasis on monitored on-farm trials with the aim of documenting changes in technology implementation, on-farm performance and adoption of technologies. For these aims to be met it is important to establish that collected data is reliable and informative. Since N2Africa's focal adaptation surveys are often executed by enumerators who have received only limited training and are not directly employed by the project, data quality and completeness may be compromised, particularly for questions considered cumbersome or complicated by the implementing staff. For this reason, the first objective of this internship was to evaluate the completeness, validity and consistency of the data.

The second objective was to evaluate the informativeness of the data. Given the aims of the focal adaptation trials, informative questions are those that record changes that farmers make to legume technologies and management recommendations, those that relate to differences in measured yields and those that can be used to group farmers into groups based on production orientation, wealth and yield levels. Reducing the length and complexity of surveys is one way to improve data quality. This requires identifying a core set of questions that can be shown to be informative and removing questions that are not. This was the final aim after this internship, propose a new survey with informative questions about the topics mention before.

4.1 **Quality** of the data

Overall, the data collected for 2016 and 2015 adaptation trials was enough to draw conclusion about the performance of the N2Africa technology in terms of agronomic and biophysical factors. In the annual report for 2016 (Ampadu-Boakye et al, 2017) is presented that the ODK database provided more uniform data for reporting across the countries, as compared to 2015. However, and as it was mention before, there is an enormous yield variability among places and technologies that is no truly explain by biological factors

Factors that could explain (part of) this enormous variability, as seen in the results of this internship, are household and farm characteristic and management practices. In the current survey, some of the questions related with this topics are answer completely and can be used to analyse yield variability. However, there is still a lack on relevant information about income and expenses of the household and labour dynamics. Some valuable data is being collected completely in the survey (family size and age distribution among the members, frequency of hired labour and education), with some assumptions about labour dynamics according to age is possible to estimate the family labour and the share labour, however, this is could be risky and led to erroneous information.

From the results was also possible to observe a lack of clear information about the farm. Even if farm size is recorded most of the times, in the baseline surveys from the N2Africa project it was observed that farmers usually overestimate or underestimate the size of their farms, so these figures should be treated with care. On the other hand Salami et al (2010) conclude that the field size is more important that the total land own, and as it was clear from table 6 information about other field different than the N2africa field was not recorded.

When thinking about targeting, is fundamental to assess what are the probabilities for a farmer to adopt certain technology. These could be defined by market orientation, labour availability and overall socioeconomic status. Recording exact figures of income and expenses thru surveys is not reliable Additionally, one of the lessons learned from the N2Africa baseline surveys, is that farmers were also unable to estimate amounts of inputs used in crops. Therefore, it was not useful to calculate input use per hectare. However, asking about the total amount of fertilizer (and type) used in the farm, it might not give an estimation of Kg fertilizer per hectare but it can give an estimation of “how much” the farmers expends in fertilizer per season.

4.2 Informativeness of the data

4.2.1 Identifying survey variables that are informative of changes to technology/practices

Recording the changes in the technology, or the changes in practices is challenging, especially because is not possible to fully know the recommendations of the technician to each farmer. Farmers are provided with a package with a known amount of seed and inputs. If the package code is recorded corrected, the species, variety and inputs are known. However, the package doesn’t include recommended practices, this are defined by the N2Afric partners in each country and are communicated by the technicians handing out the packages. In some cases, some practices are also shown in the demonstration trials.

Since the recommendations are not recorded, the changes in the technology can’t be fully known, however, knowing the reasons of the farmers for choosing certain practices can give insights in the perception of the farmers in the N2Africa technology and the possible adaptations they are doing. As seen in figure 5 some farmers decided to intercrop or row plant because it was instructed this way. The other possible reasons, however, don’t give information about farmers that were instructed do to it but decided not to.

Each best-bet technology proposed for N2Africa is linked with certain inputs and practices that ensure its maximum performance. There are two crucial steps that need to be recorded in order to transform the best-bet in best-fits: a. The instruction given to the farmers b. the interpretation and adoption of this recommendation. Changes should be made in the current methodology, either standardize the package codes and make sure the technician gives the instruction that corresponds to each package, or add some questions in the survey that record the recommendations of the technicians and then, farmers changes.

4.2.2 Identifying survey variables that are informative of performance

- Farm characteristics

According to the data (figure 7) bigger farms are related with higher yield. The relationship between farm size and data is hard to untangle as it might be related with several factors, however, bigger farms which produce more also mean more inputs (either as crop residues or manure, or as cash flow to get fertilizers) which in turn will give higher yields. From the available data, there was no significant interaction between the production orientation and the farms size, however, subsistence farmers have smaller farms and following the same pattern, farmers who gets most of their income from the farm, have bigger farms (data not show).

Walker *et al* (2006) mention that Increasing farm size and number of fields are highly correlated with reductions in the severity of poverty in sub-Saharan Africa, considering that the farm size is accounted as the land own not the land cultivated. This information must be treated carefully and is risky to draw conclusions out of it. Small farmers tend to systematically under-report the size of their plots, and it is only among the top three landholding deciles that farmers tend to over-report farm size (Carletto, Savastano, & Zezza, 2013) Additionally, Ayalew Ali & Deininge (2014) suggests that more than the total farm size the size of the plot will define agriculture productivity. In agreement, subdivide land into smaller units (which is the pattern in sub Saharan Africa) can fragment the production system and lower productivity (Salami, et al., 2010).

- Labour and management

The frequency in which the farmer hires labour presented significant differences. The yield from farms that never hired labour is significant lower that for farmers that do it permanently, or even during peak periods or when money allows. Labours constrains are directly associated with the farm productivity since important management practices like weeding or soil preparation can be delay when there is not labour available (Ronner, et al., 2016). The number of weeding also affects the yield significantly when only one weeding was carried out during the season when compared with doing it two or three times. Farmers in several surveys in sub-Saharan Africa had reported that weeding is the activity which more labour requires (Leonardo, et al., 2015) which agrees with the results found for hired labour presented before, however, there was not significant effect in the interaction between the weeding frequency and labour.

Giller et al (2001) concluded that allocation of resources (which includes allocation of labour for weeding and other practices) result in significant differences in soils fertility in the farm and consequently differences in yield. Labour availability among the farmers participating in the N2Africa trials is a key factor to consider when designing the best fit technologies, since Farrow et al (2016), included as one of the factors influencing legume adoption.

- Household characteristics

The number of member in the household didn't show significant influence in yield, neither a significant interaction with the frequency of hired labour, however, according to Mugi Nenega (2016) household size can influence the amount of labour availability in two different ways : Either the part of the family is force to off farm activities (when the production is not enough to supply the needs and extra income is necessary), or there is more labour endowment in the farm which will in turn increase the production. They key point, and what should be the focus of introducing new technologies, is help the families increase production with less labour to take them to the second group.

Another factor that also show significand effect in yields was the highest level of education in the household (not always the same as the head of the household). When the highest level of education in the household was post-secondary or university yields were significantly higher than for primary or secondary schooling. Additionally, even if not significant, when "other" education was reported, the yields were also higher than primary and secondary. This might be related with informal learning in agricultural practises. As an example, education about inoculant was a major driving force for the success of a soybean promotion program (Chianu et al., 2011). Education was also listed by Farrow et al (2016) as one of the factors affecting legume adoption.

Interestingly, even if the education of the household head was not related with yield it did show an interacting with treatment. For primary and secondary education, the yields from both plots were significant different, however, for postsecondary and university, there was no differences between both plots (figure25). Mugi Nenega et al (2016) found a positive relation between the education level and the adoption of new ideas or concepts in agriculture which could be what is observed here. More educated farmers can understand and follow the N2Africa recommendations better improving the performance of the technology

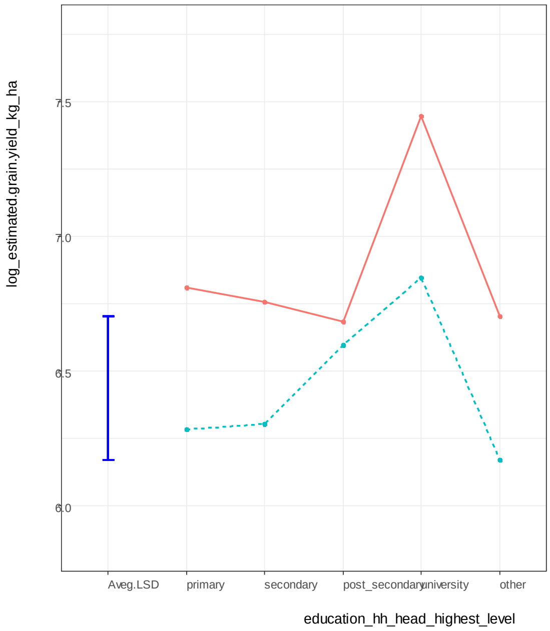

Figure 25 Interaction between the higher level of education of the head of the household and treatment and their relation with yield. The solid line represented the N2Arica plot and the dotted line the own plot.

- Market orientation

The proportion of income from farming had also a significant effect in yield. Yields were significantly higher when the household gets most of the income form the farm compared with when they get half or all from the farm. Tittonell et al (2015) included the percentage of income from farming and the number of years perceiving non-farm income in an analyses related with the diversity of rural household and they influence in soil fertility; the first one was the variable with higher loadings. Diversifying income sources by generating income from activities off the farm may increase the productivity of the farm and help reduce farmers' vulnerability to exogenous weather or price shocks (Proctorm 2014).

4.2.3 Identifying survey variables that can be used to group farmers

From the relatively small number of farmers analysed per country, and the few farm and household factor that were considered is visible the high variability among farms (even in the same region). This can somehow explain the variability found in the N2Africa trials performance. Among the factors that significantly affected yield presented in section 1, most of them presented also significant differences among clusters. The differences among cluster differ for the 5 countries evaluated. However, some variables were clearly important when defining separate groups: Farm and family size, the frequency of hired labour and the proportion of income from farming were some of them.

Farm size differed a lot in the five countries, and it was usually related with a bigger family. Contrary to what was expected based on previous results (section 3.2.2) bigger farms were not always related with better yields (Eg. Tanzania and Nigeria). It was also common to have cluster whit relatively big farms but with lower frequency of hired labour. Here, it becomes important to make a distinction between the area own (the data in this case) and the area farmed which can give a better insight into the production and the labour needs.

Is worth to mention that in the results, larger farms were not related with more TLU. In Titonell's (2010) farm typology, larger (and wealthier) farms were focus in the production of cash crops, not livestock products, so area is not always proportional to herd size. Overall, livestock is not the main source of income in sub-Saharan Africa. The baseline surveys found that at least 90% of the household in Ghana rely in cropping as their main source of income and the situation doesn't differ a lot among countries. In the base line it was also observed that differences in livestock keeping between regions appear to be more associated with cultural practices and agricultural development pathways. In Nigeria for example, at least 6% of the interviewed farmers (the higher off all countries) declare to rely mostly of livestock as main income. In the clusters, farms with higher TLU had also a considerable bigger proportion of their income from farming and less home consumption which can indicate more specialized farms.

Anyhow, having livestock not only will improve the productivity or income of the farm, but is related with the overall resource endowment of the household. In Franke *et al* (2014) farm characterization, the poorest farms are characterized with small areas and no livestock, they also get most of their income from casual labour. Consistently, in the clustering, smaller farms, with less TLU were related with less months of food sufficiency, surprisingly, these was not always corresponding with a relatively lower yield, but with a lower proportion of income from farming. From the clustering, It was also clear that are more wealthy farmers and even if they have a high home consumption (so they're not marked oriented), they hired labour frequently, have more than 11 months of food sufficiency and higher yields.

4.2.3.1 Implications for targeting the N2Africa technology's

One of the aims of N2Africa is to translate best-bet technologies into best-fit adapted for every type of farmer, but adoption, as Farrow (2014) presented, depend on multiple socioeconomic factors that include: Household access to capital and assets, output market for legume products, availability of inputs, gender, education of the household members, experience of the farm household members and land availability. As seen in the clustering, farmers can vary widely among this characteristic, and is impossible to categorize each farmer.

A first remark from the clustering results, is that yield differences among cluster were never significant. This is relevant when targeting agricultural innovations since yield is an indicator of the performance of the propose technologies. If the targeting takes yield as the main factor to separate farmers in groups, this clustering wouldn't be relevant. However, in the first part of the analyses, several household and farm characteristics were significantly related with yield, some if which, presented significance differences among clusters. This ambiguous result can be related with the fact that in the mixed linear model, country was not added as random factor while the clustering was done per country. Conclusion drawn from the clustering need to be handle with care, but anyway, discuss and comment the differences found between farmers can lead to a better targeting of the technologies when considering other factors (different than yield).

The proportion of home consumption, was related with yield (see section 3.2.2c) , and differed among clusters. This defines if a farmers is producing for the market, for self-subsistence or if the household is diversified and non-farm income plays an important role. In addition to give insight in the type of farmer, this information can also give information about the seed availability for next season which will also influence the adoption and performance of the technology in following years (Freeman, et al.2002). Mhango *et al* (2012) reported that seed and market access are two of the most important drivers for legume adoption. A seed shortage can be the consequences of poor availability of the seed in the market, lack of money or the inability to store seed from previous seasons, which in turn can be related with a food shortage or a need for cash (David & Sperling, 1999). Gather this information implies adding additional questions to the survey, to understand what percentage of the produce goes where (consumed, sold, trade or used to feed the animals) (Anderson, et al., 2016).

This differences in proportion of home consumption and income from farming are important when introducing new legume technologies, because different farms will have different needs (more food for the children vs better grain for the market etc.) Marked oriented farmers usually have higher resource endowment and a salaried job (Franke et al 2014) additionally, Rooner *et al* (2015) found that soya bean technologies showed better responded under wealthier farmers. On the other hand, farmers that's are not fully engaged in agriculture (is not their main source of income) have lower potential to adopt new technologies, especially these related with soil and water conservation practices (Mugi-Ngengaa, et al., 2016). From below, it is possible to observe the difficulties that can be found when proposing new technologies to farmers.

According to Davis *et al* (2017) the diversification of household income in sub-Saharan Africa is the norm rather than the exception, however, the drivers to look for other sources of income different that farming can be diverse for types of farmers. Studies have concluded that there is a relation among capital accumulations of adult labour and the access to non-farm activities (Proctor, 2014). Additionally, non-farm incomes are useful to minimize risk (in case of crop failure) (Proctor, 2014) and risk management might be related with the willingness of the farmer to try or invest in a new technology.

The frequency of hired labour also showed differences between clusters, however, there is not enough information in the adaptations survey to untangle the reasons to hired or not labour. This could be related with having enough family labour, or with a bigger farm or the lack of money to do it. Snapp and Silim (2002) conclude that, after return, the labour requirement is the most important criteria to choose a new legume technology. How strong these preferences are depended also from the region and the gender of the farmer, since in some places women tend to work more in the farm than men.

Available labour plays also a key role in the yield because is related with the timing of several agronomic activities. In the baseline, a big percentage of farmers report not being able to hired enough people to carry the farm activities on time, which will delay the planting and the weeding date. In Ethiopia, 25%-48% of the households indicated that farming activities were delayed, because farmers were not able to hire labour (Franke and Wolf, 2011) which will consequently affect yield

When analysing de adoption of new legume technologies, labour was often the most important constrain. Additionally, the land/labour ratio was highly related with food sufficiency (Titonell 2015) and food sufficiency is an important drive when choosing or adopting a new technology. Important questions to add to the survey, is ordered to understand better the labour dynamics can be: family labour (# members working on the farm) and family members working temporarily/permanently off-farm. Assumptions can be made based on the number of family members in different age groups, but this can be risky whit so much variability between farmers.

It is now clear how several household and farm factors can affect the adoption of certain technology and how hard is to untangle the complex relations among them. Multiple factors need to be considered in order to target better the legume technologies proposed by N2Africa. Wealthier farmers are probably more interested in grain legumes and not so much in the extra benefits in soil improvement. For larger farms than rely mostly in cash crop, legumes that can be grown in intercrop with their current crops might be a good option. More research in the performance of the technology under different intercrop systems can bring insights into this. On the other hand, poorer farmers, where food sufficiency and labour availability are the main constrains, might be interested in labour saving varieties or practices. In both cases a strong chain (or the possibility to build one) that ensures the availability of seed and other inputs will increase the probably of adoptions.

4.3 Recommendations for future surveys

With small assumptions, some of the questions in the current focal adaptation survey can be used to estimate the potential of farmers to adopt certain technology or to invest on it, however, more details in some topics is needed to have more reliable analysis. Several authors have observed that ask for expenses and incomes directly is not efficient, especially because in the social and political context of most sub-Saharan Africa the answers can be bias or the farmer simply denies to give this type of information (Vyas & Kumaranayake , 2006). Vyas (2006) suggest that is better to collected for variables that capture living standards, like household assets (e.g. TV, bicycle, cell phone) and infrastructure and housing characteristics.

Tittonell et al (2010), when analysing the diversity of rural livelihood, included also the Number of graded (improved breed) cattle and Number of oxen and ox-ploughs. The last one can also give information about labour saving tools. These were also included by Falconier et al (2015) as part of the variables that describe basic farm resources and define the potential of land and labour productivity.

Another way to collect information about income generation, without directly asking the farmers, is getting more detailed information of the income sources (Anderson et al, 2016). This includes asking what type of salaried job he/she does, how does the farmers sell the product (in the city, market, retails, trade), and to record the in-kind income from livestock (what are the animals used for). If there are not time constrains for the survey, more questions about the livestock can be made, considering that animals are most of the times live assets. Informational bout the herd size and composition (age and sex of the animals) is a basic indication of the potential of the livestock to generate income (Zezza,, et al., 2016)

The variety of economic and farming situation in sub Saharan Africa makes very difficult to characterize farmers to the detailed needed for designing bet fit technologies for each situation. The suggestion is Instead of characterizing the farms and the farmers in groups, and design legume technology for them, try to design flexible technologies that can be adaptable for specific constrains , especially labour. Additionally, understand the labour dynamics can else help targeting the technologies. Add questions that give insight into the family labour or the share labour could give valuable information.

Is important to take into account the any conclusions or assumption about economic situation will change depending on the location. In recent years the importance of on farm sources if income is decreasing and they are being replace for non-agricultural activates. Additionally, Davis et al (2017) observed that the share of non-agricultural income increases with increasing levels GDP). This could mean that farmers with (theoretically) more investment potential are not going to invest in agriculture.

Following this logins, it becomes important to understand the motives of the farmer to try or adopt certain technology in order to understand what are is the probability that this farmers invest on in tin the future. Emphasis must be made in asking (and making sure that farmer answer) the reasons for choosing certain N2Africa package.

4.4 Limitations of this study

The main goal of this internship was to revise the Focal adaptation survey used in the N2Africa project, however, the total sample was quite unbalanced (since not all legumes are grown in all countries) which made some of the statistical analyses inaccurate. Besides, the number of trials per country differ considerably, so for some countries like Tanzania more conclusions can be drawn when compare with Ethiopia for example.

On the other hand the amount of unknown or invalid data collected in the surveys did the farm characterization rather arbitrarily and is hard to conclude that this accurately represent the farms participating in the trials. Additionally, lack of social economical information can affect the grouping.

This internship didn't analyse deep the treatment effect in yield (the differences between the N2Africa plot and the own plot), however, the lack of data about the fames yield makes it risky to draw conclusions about the performance of the proposed technology compared with the current practices.

The big variability in the N2Arica trials performance is related with a big variability among the farmers. In the PCA, no variable shows considerable high loading, making the cluster arbitrary. This could be because the complex interactions among the evaluated variables. Due to time limitations during this internship it was not possible to analyse deeper the relationship between several household characteristics that can define types of farmers and/or influences yield (e.g. family members in group ages and labour or home consumption).

5. References

Ampadu-Boakye, T., M. Stadler & F, Kanampiu (2017) N2Africa Annual Report 2016, available in http://www.n2africa.org/

Anderson, J., Learch, C. E., & Gardner , S. T. (2016, April). National Survey and Segmentation of Smallholder Households in Uganda Understanding Their Demand for Financial, Agricultural, and Digital Solutions. *Working Paper on National Survey and Segmentation of Smallholder Households*, p. 90.

Ayalew Ali, D., & Deininger, K. (2014, February). Is There a Farm-Size Productivity Relationship in African Agriculture? Evidence from Rwanda. *Policy Research Working Paper Series*(6770).

Ayuk , E. T. (2011). Preface. In *Fighting Poverty in Sub-Saharan Africa: The Multiple Roles of Legumes in Integrate Soil Fertilty Management.* London: Spirnger.

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R.* Springer.

Bationo, A., Waswa, B., Okeyo, J., Maina, F., Kihara, J., & Mokwunye, U. (2011). *Fighting Poverty in Sub-Saharan Africa: The Multiple Roles of Legumes in Integrate Soil Fertilty Management.* London: Springer.

Carletto, C., Savastano, S., & Zezza, A. (2013). *Measurement, Farm Size and Productivuty.* Retrieved from Living Standards Measurement Study Brief Series - Integrated Surveys in Agriculture: http://siteresources.worldbank.org/INTSURAGRI/Resources/7420178-1294259038276/Fact_Artifact_Brief.pdf

Chianu, J. N., Nkonya, E. M., Mairura, F. S., Chianu, J. N., & Akinnifesi, F. K. (2010). Biological nitrogen fixation and socioeconomic factors for legume production in sub-Saharan Africa: a review. *Agronomy for Sustainable Development*.

David, S., & Sperling, L. (1999). Improving technology delivery mechanisms: Lessons from bean seed systems research in eastern and central Africa. *Agriculture and Human Values, 16*(4), 381-388.

Davis, B., Di Giuseppe, S., & Zezza, A. (2017). Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy, 67*, 153-174.

Falconniera, G. N., Descheemaeker, K., Van Mourik, T. A., Ousmane, O. M., & Giller, K. E. (2015). Understanding farm trajectories and development pathways: Two decades of change in southern Mali. *Agricultural Systems, 139*, 210-222.

Fan, S., Brzeska, J., Keyzer, M., & Halsema, A. (2013). *From Subsistence to Profit Transforming Smallholder Farms.* Washington, DC: International Food Policy Research Institute.

FAO. (2014). *FAOSTAT*. Retrieved from Food and Agriculture Organization of the United Nations: http://www.fao.org/faostat/

Farrow, A., Ronner, E., Brand, G. J., Boahen, S. K., Leonardo, W., Wolde-Meskel, E., . . . Kantengwa, S. (2016). From best fit technologies to best fit scaling: incorporating and evaluating factors affecting the adoption of grain legumes in Sub-Saharan Africa. *Experimental Agriculture*, 1-26.

Franke, A. C., & de Wolf , J. J. (2011). *N2Africa Baseline Report.* N2Africa .

Franke, A. C., van den Brand, G. J., & Giller, K. E. (2014). Which farmers benefit most from sustainable intensification? An ex-ante impact assessment of expanding grain legume production in Malawi. *European Journal of Agronomy, 58*, 28-38.

Franke, A., Baijukya, F., Kantengwa, S., Reckling, M., Vanlauwe, B., & Giller, K. (2016). Poor Farmers – Poor Yields: Socio-Economic, Soil Fertility And Crop Management Indicators Affecting Climbing Bean Productivity In Northern Rwanda. *Experimental Agriculture*, 1-21.

Freeman, H. A., van der Merwe, P. J., Subrahmanyam,, P., Chiyembekeza, A. J., & Kaguongo, W. (2002). Assessing Adoption Potential of New Groundnut Varieties in Malawi. *Working Paper Series*(11), p. 12.

Giller, K. E. (2001). *Nitrogen Fixation In Tropical Cropping Systems* (2nd ed.). CABI Publishing.

Herridge, K. E. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil, 311*(1), 1-18.

Jahnke, H. E., G. Tacher, P. Keil And D. Rojat (1986) Livestock production in tropical Africa with special reference to the tsetse-afferent zone. Livestock production in tropical Africa with special reference to the tsetse-afferent zone. Available in http://www.fao.org/wairdocs/ilri/x5443e/x5443e04.htm

Kamanga, B. C., Kanyama-Phiri, G. Y., Waddington, S. R., Almekinders, C. J., & Giller, K. E. (2014). The evaluation and adoption of annual legumes by smallholdermaize farmers for soil fertility maintenance and food diversityin central Malawi. *Food Security, 6*, 45-59.

Leonardo, W. J., van de Ven, G. W., Udo, H., Kanellopoulos, A., Sitoe, A., & Giller, K. E. (2015). Labour not land constrains agricultural production and food self-sufficiency in maize-based smallholder farming systems in Mozambique. *Food Security, 7*(4), 857–874.

Maechler, M. (2017). *R Documentation: Hierarchical Clustering.* Retrieved from Seminar for Statistik , ETH Zurich, Switzerland.

Mhango, W. G., Snapp, S. S., & Phiri, G. Y. (2043). Opportunities and constraints to legume diversification for sustainable maize production on smallholder farms in Malawi. *Renewable Agriculture and Food Systems, 28*(3), 234-244.

Mugi-Ngengaa, E. W., Mucheru-Munaa, M. W., Mugwe, J. N., Ngetichb, F. K., Mairura, F. S., & Mugendi, D. N. (2016). Household's socio-economic factors influencing the level of adaptation to climate variability in the dry zones of Eastern Kenya. *Journal of Rural Studies, 43*, 49-60.

Mutuma, S. P., Okello, J. J., Karanja, N. K., & Woomer , P. L. (2014). Smallholder Farmers' Use And Profitability Of Legume Inoculants in Western Kenia. *African Crop Science Journal, 22*(3), 205-213.

Nakasaka, K. (2016). *Evaluating farmers' decision making on choosing technologies and practices in adaptation trials.* N2Africa project in northern Ghana. Wageningen University.

Nezomba, H., Tauro , T. P., Mtambanengwe , F., & Mapfumo, P. (2008). Nitrogen fixation and biomass productivity of indigenous legumes for fertility restoration of abandoned soils in smallholder farming systems. *South African Journal of Plant and Soil, 25*(3), 161-171.

Ojiem, J. O., de Ridder , N., Vanlauwe , B., & Giller, K. E. (2006). Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems. *International Journal of Agricultural Sustainability, 4*(1), 79-3.

Proctor, F. J. (2014, September). Rural economic diversification in sub-Saharan Africa. *IIED Working Paper*.

PROTA4U. (n.d.). Retrieved from PROTA4U Web data base : PROTA4U Web data base www.prota4u.info/database/

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria https://www.R-project.org

Ronner, E., Franke, A. C., Vanlauwe, B., Dianda, M., Edeh, E., Ukem, B., . . . van Heerwaarden, J. (2016). Understanding variability in soybean yield and response to P-fertilizer and rhizobium inoculants on farmers' fields in northern Nigeria. *Field Crops Research, 186*, 133-145.

Salami, A., Kamara, A. B., & Brixiova, Z. (2010, April). Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. *Working Paper Series*(105).

Snapp, S. S., & Silim, S. N. (2002). Farmer preferences and legume intensification for low nutrient environments. In J. A.-G. (Ed.), *Food Security in Nutrient-Stressed Environments: Exploiting Plants' Genetic Capabilities* (Vols. DPSS, volume 95, pp. 289-300). Kluwer Academic Publishers.

Tittonell, P. A., Muriuki, K. D., Shepherd, D., Mugendi, K. C., Kaizzi, J., Okeyo, L., . . . Vanlauwe. (2010). The diversity of rural livelihoods and their influence on soil fertility in agricultural systems of East Africa – A typology of smallholder farms. *Agricultural Systems, 103*(2), 83-97.

Tittonell, P., Vanlauwe, B., Leffelaar, P. A., Rowe, E. C., & Giller, K. E. (2005). Exploring diversity in soil fertility management of smallholder farms in western Kenya: I. Heterogeneity at region and farm scale. *Agriculture, ecosystems & environment*, *110*(3), 149-165.

Vanlauwe, G., Giller, K., & van Heerwaarden, J. (2014). Agronomy Master Plan. *N2 Africa Master Plan, 1.1*. Retrieved from available in http://www.n2africa.org/

Vyas , S., & Kumaranayake , L. (2006). Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan. 2006 Nov;21(6):459-68. Epub 2006 Oct 9., 21*(6), 459-68.

Walker, T. S., Boughton, D., Tschirley, D. L., Pitoro, R., & Alda, T. (2006). Using Rural Household Income Survey Data to Inform Poverty Analysis: An Example from Mozambique. Queensland, Australia: International Association of Agricultural Economists (IAAE) > Annual Meeting.

Welham, S., Cullis, B., Gogel, B., Gilmour, A., & Thompson, R. (2004). Prediction in Linear Mixed Models. *Australian & New Zealand Journal of Statistics, 46*(3), 325–347.

Zezza, A., Pica-Ciamarra, U., Muger, H. K., Mwisomba, T., & Okello, P. (2016, November). Measuring the Role of Livestock in the Household Economy A Guidebook for Designing Household Survey Questionnaires. *LSMS GUIDEBOOKS*.

Annex A

General Guidelines for 2017 (Focal) adaptations

Introduction

Adaptation trials (Task **2.2.2** in the Agronomy master plan) are small, farmer-managed try-outs that evaluate the implementation and effects of specific N2Africa best-bet technologies by a large number of farmers. There will be in the order of **200-1000 per target region (around 1000-5000 per country, per year)**. They form an important part of the **N2Africa led dissemination** and link up with the larger **demonstration** trials by offering farmers technology packages that form a subset of those included in the demonstration trials. The basic goal of these trials is to monitor the **performance of best-bet technologies** under farmer management and to register **adaptations** of these technologies by farmers.

> **Important**
>
> Ask the farmers to plant the N2Africa package in a regular (10x10 m) plot next to their own main legume field (figure 3)

Execution of adaptation trials

All farmers receiving an N2Africa technology package will be **registered** on a specially provided **"Input Distribution and Feedback Form"**. Adaptation trials are linked to a demonstration trial, so there will be usually by one form per demonstration trial. The "Input Distribution and Feedback Form" also provides space for the answers to a short set of **feedback questions** to be asked after harvest. These questions can be answered quickly either by telephone or short interviews during after-harvest field days or home visits.

The distribution of inputs provides the opportunity to **instruct** farmers on the use of inputs and best agronomic practices. It is advised to prepare a simple **instruction leaflet** for farmers explaining how to plant the N2Africa package. This leaflet can also be used to collect basic **farmer feedback** that can be recorded on the "Input Distribution and Feedback Form" at the end of the season. It is important to ask farmers to plant the N2Africa package in a regular (typically 10 meter x 10 meter) plot next to their own main legume field. This makes it easier to compare the performance of the N2Africa technology to farmer's practice.

In addition, it is important to **randomly sample** a subset of **50-100** farmers as so-called **focal adaptation** farmers, which will be subject to more intensive data collection, particularly yield measurements. The role of focal adaptation trials is to provide more detailed information on the on-farm performance of N2Africa packages and to have proper documentation of **adaptation practices**. The aim is to take **yield data** at 50-100 focal farmers so make sure to sample additional farmers if not everyone is found to have planted their adaptation package.

Preparations an execution of focal adaptation trials

Selecting focal farmers

Focal farmers should be selected **at random** from the list of farmers registered on the "Input Distribution and Feedback Form". A more or less random selection can be made by selecting the 5th farmer on the list and selecting each subsequent farmer to be exactly 10 names below on the list. Aim for selecting **at least 20** focal farmers (that have actually planted the package) per target region **(50-100 farmers in total)**. In practice this means selecting around 2 farmers from each "Input Distribution and Feedback Form". It is possible that for logistical reasons you can only select focal farmers from a **subset of localities**, in which case you will of course need to sample a larger number of farmers per form.

We are interested in comparing performance of new technologies compared to farmer's own practice. So a farmer can be selected as focal farmer <u>only if</u>:

1. He or she is planting **the same legume crop** as the one on the chosen N2Africa package
2. He or she has planted the **N2Africa package next or close to their own legume**.

Since not all farmers that have received an adaptation package will have planted the package, or may not be willing to act as focal farmer, be prepared to sample **more focal farmers** if the target number of 50-100 focal farmers that have planted is not reached.

Focal farmers will receive the same instructions as other farmers, but in addition will be asked to:

- Be willing to receive N2Africa field staff twice during the season for **monitoring** of the trial and its management.
- Participate in the **documentation** of details on agronomic practice and adaptation of the technology by means of a **field book**.
- **Save** their harvested and grain for yield measurement by N2Africa field staff.
- Have **photographs** taken of their trial field.

After planting, a minimum of **2 visits** should be planned for each focal adaptation farmer: a **mid-season** visit and a **harvest visit**. <u>All data should be recorded electronically using a tablet or phone.</u>

Mid-season visit

The mid-season visit has two main aims: Establish a **control section** within the farmer's own field and to **check the state and characteristics** of the N2Africa package and the farmer's own legume plot. Before starting the visit don't forget to **switch on the GPS on your tablet or phone,** GPS coordinates of the N2Africa plot are very important. Additionally, a number of **household characteristics will** be recorded during this visit.

IS RECOMMENDED TO **RECORD DATA ONLY FOR THE LEGUME CROPS**,

If the legume is **intercropped,** you can record information for **at least 2 more crops**. Follow the same observations list below and the same methodology for plant density in figure 2. Note that you will have to measure yield for the chosen crops in the next visit

The activities during the mid-season visit are:

1. Establishing a **control section** within the farmer's main legume field to allow yield comparison with the N2Africa package.

To stablish a control section in the farmer's legume field follow the steps figure 1.

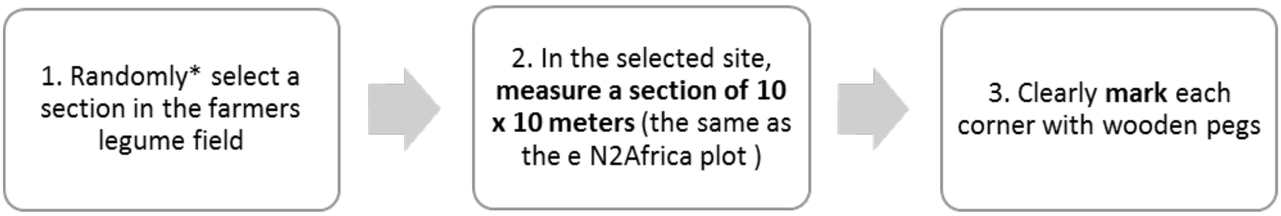

Figure 3. How to stablish a control section. *You can choose a section randomly throwing a light object (e.g. a hat or a rubber ball) into the field and placing the section wherever that object lands.

2. Checking the state and characteristics of the N2Africa package and the farmer's own legume plot.

After the control section has been established observe and record (from the N2Africa plot and the control section in the farmer's own legume field):

- Take a **photograph** of each plot
- Note the GPS coordinates
- Record exact **plot size**, **plant survival** and **planting density** (follow the steps in figure 2)
- Observe some field characteristics

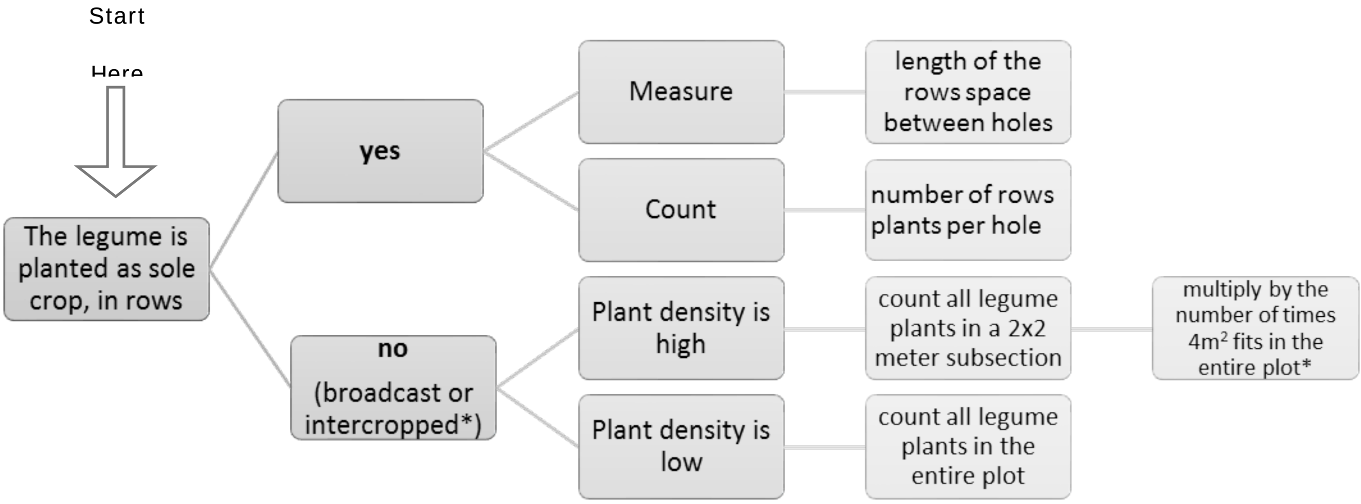

Figure 4. How to measure plant density. (* e.g. if the entire plot is 10x10 m, multiply the number of plants you count in a 2x2 section by 24, which is the number of times $4m^2$ fits in $100m^2$)

3. Record some history of the field where the N2 Africa plot ins planted, as well as a number of land, cropping and livestock characteristics

IMPORTANT! Before you finish the visit **instructed the farmer to a**. Leave the control section in his field and the n2africa plot to be **harvested by technicians** in the next visit or **b. Store the harvested grain** of the N2Africa plot and of the control section in his own legume plot in **two separate harvest bags**. If this is the case, provide the farmer with two bags of different colour (e.g. blue for the N2Africa plot, red for the control section).

Harvest visit

At the end of the season a second visit should take place to take **yield data**. Farmer's will be asked to harvest, shell and store the grain separately for the N2Africa plot and the control section in the farmer's own legume plot.

1. **Before weighting the harvest**: Ask the farmer to provide their own estimate of the yields for the N2africa plot and for their entire own legume plot. These estimated yields are then on the instruction leaflet.

2. **Taking yield data**: Grain weights will be measured exactly by weighting the two separately stored harvest bags (i.e. from the N2Africa plot and the control section) with a digital scale to a precision of two decimals (e.g. 15.45 kg). See figure 3.

3. **After measuring yield**: Ask the farmer to provide information on problems faced during the season, his agronomic management. All data will be recorded electronically using a tablet or on phone.

Note: **If you recorded information for other crops during the mid-season visit, measure the yield for those crops as well.**

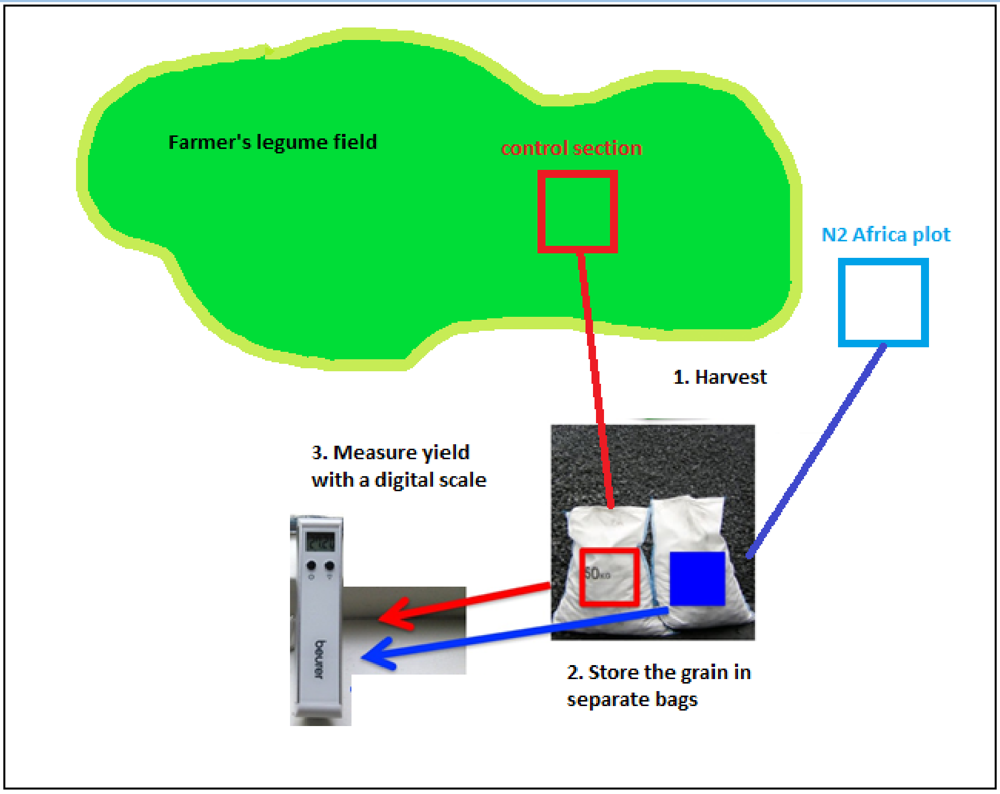

Figure 3. The principles of good harvest estimates. Mark a 10 x 10 section of the farmer's main legume field. Ask the farmer to harvest this section separately from the rest of the field and keep the harvest in a separate yield bag. Compare the weight of this bag to the weight of the 10 x 10 N2Africa plot, using a digital scale.