

Farm Typology for targeting fertilizer advisory -

Digital Green and Fertilizer Ethiopia Use Cases

Version 2.0

Published: May 2024

Rohit R. Pawar¹, Andrea Diaz Ismael¹, Katrien Descheemaeker¹, Feyera Liben³, Mohammed Ebrahim³, Mern Eshete³, Abiro Tigabie⁴, Betelhem Tsedalu⁴, Gizaw Desta², Lulseged Tamene³, Celine Aubert⁵, Banchayehu Assefa⁶, Elke Vandamme⁷, Kalimuthu Senthilkumar⁸

This report was written by Plant Production Systems, group of Wageningen University and Research, with contributions from Excellence in Agronomy Ethiopia Use Cases and Add-on Segmentation team. This study was commissioned and financed by Excellence in Agronomy (Project 318 3600710).

CONTENTS

Acknowledgements	4
Project Background	5
1. INTRODUCTION	6
2. MATERIALS AND METHODS	8
2.1 Typology construction	8
2.1.1 Farm household survey	8
2.1.2 Objectives and hypothesis	8
2.2 Factor analysis on mixed data (FAMD)	9
2.2.1 Data cleaning and processing	9
2.2.2 Statistical analysis	10
3. RESULTS	13
3.1 Typology 1 – Region level	14
3.2 Typology 2 – Cropping system level	15
3.2.1 High intensified wheat farms in highlands (HIW)	16
3.2.2 Medium intensified diverse farms in mixed altitudes (MIF)	17
3.2.3 Low intensified sorghum farms in lowlands (LIS)	19
4. DISCUSSION	23
5. CONCLUSIONS	27
6. RECOMMENDATIONS	28
7. REFERENCES	29
8. APPENDIX2	31
9. SUPPLEMENTARY MATERIAL	36

ACKNOWLEDGEMENTS

We acknowledge Excellence in Agronomy (EiA), a CGIAR Initiative to improve agronomic practices with the goal of increase productivity and quality of food of smallholders of farms, for funding this work.

As part of CGIAR, the Excellence in Agronomy (EiA) is an Initiative that focusses to generate impact on poverty reduction as a primary impact area, followed by Nutrition, health and food security, Gender equality, Climate adaptation and mitigation, and Environmental health and biodiversity as related impact areas. With an estimate of 80% of smallholding in the world's farms (CGIAR, 2023), the improvement of agronomic practices represents an opportunity to increase productivity and quality of food for many farming households. Research organizations involved in core partnership with Excellence in Agronomy Initiative are AfricaRice, Alliance Bioversity and CIAT, CIMMYT, CIP, ICARDA, ICRISAT, IFPRI, IITA, IRRI, IWMI, World Agroforestry and National Agricultural Research System of Ethiopia and World Agroforestry. www.eia.cgiar.org

We acknowledge Alliance Bioversity – CIAT and ICRISAT for implementing and the enumerators on field for conducting the survey.

CGIAR is a global research partnership for a food-secure future dedicated to transforming food, land, and water systems in a climate crisis. www.cgiar.org



We acknowledge Thomas Delaune and all the colleagues from Plant Production Systems chair group for their support in the initial stage of the data analysis.

Wageningen University & Research (WUR) is the partnership between Wageningen University and the Wageningen Research Foundation with the Mission statement “To explore the potential of nature to improve the quality of life”. More than 7,600 employees, 13,100 students and more than 150,000 Lifelong Learning participants from more than 100 countries contribute to this. They do this not only in Wageningen but all over the world. They do this in the field of healthy food and living environment with the help of their own research for governments and the business community. www.wur.nl

PROJECT BACKGROUND

The EIA Initiative is organized around Use Cases located in the Global South, including Cambodia, Colombia, Democratic Republic of the Congo, Egypt, Ethiopia, Ghana, Ivory Coast, Kenya, Malawi, Mali, Mexico, Morocco, Nigeria, Peru, the Philippines, Rwanda, Senegal, Uganda, Vietnam, Zambia and Zimbabwe (CGIAR, 2023). The lessons learned on each Use Case will provide evidence to develop lessons at global application. Each Use Case has been working on the co-creation of Minimum Viable Products (MVP) that provides agronomic advice at scale.

In order to maximize the impact potential of the MVP by each Use Case, a household survey has been performed at country scale, initially in Ethiopia to gain insights into relevant areas such as gender dynamics and farmer diversity that may affect adoption potential of the MVP. The collected data is the base to construct and analyse the different farm typologies across different geographies and commodities. To do the Farm Typology analysis of the Ethiopian Use Cases, EIA approached Plant Production Systems, group of Wageningen University (WUR), with the request to clean, process and analyse the collected data, elaborate a Farm Typology and socialize the methodology of this report in a format of protocol to be use by Use Case leaders.

This report involves three steps: 1) data analysis framework; 2) construction of the farm typology; and 3) presentation of a protocol to understand and execute the R scripts developed during the data analysis.

This document represents all three steps in a systematic manner. It starts with an introduction indicating the aims of the farmer segmentation and the used data framework (chapter 1), followed with the methods we used for the typology construction (chapter 2) and the outcomes of the analysis as results and discussion (chapter 3 and 4).

1. INTRODUCTION

Complexity of agro-ecological, socio-economic and resource endowment conditions is a major challenge for the characterization of farming systems. Temporal and spatial variability in smallholder communities is normally driven by opportunities and constraints that are linked to the diversity of landscape, culture, market, resources, agricultural and off-farm activities, land access, among others (Alvarez et al., 2018).

The construction of typologies is a process that summarizes variability of the existing households and farming systems. The process involves the selection, classification, description, comparison and interpretation of variables. However, the outcomes depend on the research questions and main purpose of the typology such as: identification of diversity and its underlying causes; analyse agricultural trajectories; or support development (selection of representative farms), implementation (scale-out appropriate interventions) and monitoring (scale-up of impact assessment) of agricultural projects (Alvarez et al., 2014; Kuivanen et al., 2016). Furthermore, farm typologies can be analysed at landscape, field or household level (e.g. resource endowment), leading to different approaches and methodologies to construct the typology (Alvarez et al., 2014; Kuivanen et al., 2016).

The EiA Initiative aims to generate actionable insights into farm diversity that can be used to increase impact potential of the Use Cases through a Minimum Viable Product (MVP), by informing the (re-) design and/or the scaling strategy of the MVP (EiA Standard Operating Procedure for farmer segmentation). The MVP for the Ethiopia Use Cases is described as a digital advisory tool, currently under construction, able to deliver agronomic advice on fertilizer application for wheat, teff and sorghum to end users such as farmers, extension agents and agro-dealers. The agronomic advice is focused on fertilizer applications relevant for farmers' biophysical and socio-economic context.

Hassall et al. (2023) illustrates the applications of structural and functional typologies. The first includes understanding farm differences based on their components, such as household composition, land, resources and capital assets. The second covers more nuances, understanding differences, for example, in household dynamics, livelihood strategies, attitudes, agricultural income and diversity between regions. Identification of farm types based on household opportunities and constraints, targeting agricultural interventions and innovations, are some of the outcomes when structural and functional typologies are combined (Hassall et al., 2023).

Given the fact that the MVP is being developed, the construction of farm segmentation (combining structural and functional typologies) at landscape and farming system scale will be of a great support for the Use Cases to develop and scale the described innovation. Although all the Use Cases are aligned to the main goal of the EiA Initiative, it is important to note that each Use Case has specific MVPs, objectives and contexts. Thus, the participation of the Use Cases in defining the specific objective of the farm

typology and the selection of variables has been crucial for the farmer segmentation in this report. This process has been guided by a Standard Operating Procedure (SOP) for farmer segmentation in EIA that was developed by a team of experts in EIA and made available to the Use Cases (EIA Standard Operating Procedure for farmer segmentation).

The objective of this report is to provide a methodological approach to construct the Farm Typology analysis for Ethiopia, based on the interests and specific hypothesis of two Use Cases (Digital Green and Fertilizer Ethiopia).

In parallel with the construction of the typology, we aim to illustrate the process of statistical analysis of continuous and categorical variables as part of the typology.

2. MATERIALS AND METHODS

1.1 TYPOLOGY CONSTRUCTION

1.1.1 Farm household survey

The data used in the typology construction was collected by the Fertilizer Ethiopia and Digital Green Use Case in Ethiopia as part of the Excellence in Agronomy Initiative of CGIAR. Each Use Case collected information via a standardized and modular household survey tool based on the RHoMIS survey (EiA Standard Operating Procedure for farmer segmentation). Data has been collected using the ODK Collect app, and the ONA data management platform. Consequently, two datasets were generated, differing in the number of surveyed households and districts, but sharing the same set of questions. The two datasets were combined to get household data of 889 households from Goba, Kewet, Lemo, Siyadebirnawayu and West Belessa districts. The survey was focused on the three most important crops in these regions: wheat, teff and/or sorghum. The survey was conducted in 2023 (2015 for Ethiopian calendar) including modules that are at farm and plot level.

1.1.2 Objectives and hypothesis

The wide application of typology constructions goes along with a pallet of different methods, depending on their objectives, the type of available data and size of the sample (Alvarez et al., 2018). Therefore, the results of the typology are largely influenced by the data collection, selection of variables, and methods for dimension reduction and clustering.

EiA has developed a Standard Operating procedure (SOP) for farmer segmentation (EiA Standard Operating Procedure for farmer segmentation) where the starting point is the “Use Case MVP”, followed by “objective of farmer segmentation- selection of variables”, “data collection-survey”, “typology construction” to finally “redefine the MVP scaling strategies”. The Use Cases defined their specific objectives before implementing the EiA household survey. They provided a list of desirable variables to tailor the design of the survey based on the selection of modules defined by EiA. During the exercise of this study the selection of variables for typology construction was performed after the data collection, following the objectives of the Use Cases. Within the scope of this report, and the SOP provided by EiA, we translated the purpose of the typology to the implementation of the Use Case, i.e. to scale-out appropriate and tailored interventions.

Through interactive breakout sessions and consultations with Ethiopian Use Cases, we selected variables and developed objectives and hypothesis, providing results for two typologies:

First typology (T1): The objective was to classify farming households at the farm production system level, in order to support the development of the Use Case MVP. The hypothesis was that agricultural households can be grouped according to their crop production system such as wheat, teff and/or sorghum.

After feedback from the Use Cases, the first typology was too general for implementation of the MVP. Hence, a second objective was set to explore the data further.

Second typology (T2): The objective was to differentiate farming households at the cropping system level (integrating the results from T1), in order to support the implementation of the Use Case MVP based on their resources, production orientation and use of agricultural inputs. The hypothesis was that farming households can be classified according to their resource endowment (TLU, farm size, access to technology) and production orientation.

The above-mentioned objectives align with EIA's SOP for farmer segmentation, which states: "The objective of conducting farmer segmentation in EIA Use Cases is to generate actionable insights into farm diversity that can be used to increase the impact potential of the Use Cases' MVP's and realise inclusive impact across different types of farmers".

The surveyed households are distributed across 5 districts, which are located in either highland or lowland agro-ecologies. The characteristics of each district are linked to the attributes of the agro-ecologies but also to the socioeconomic conditions. Therefore, the district variable was considered only for farm description in both typologies (T1 and T2).

1.2 FACTOR ANALYSIS ON MIXED DATA (FAMD)

1.2.1 Data cleaning and processing

Data on household, labour, livestock, income, production, agricultural inputs and access to technology was cleaned, processed and transformed for statistical analysis (Table 1).

Variables included in multivariate analysis

The raw dataset consisted of nested variables across the plot level and farm level. Farm size was calculated as the sum of plot sizes for each household. As the questionnaire was designed to collect information on ten plots only, farm sizes of households having more than ten plots is underestimated in the final calculation. The number of farm households with ten plots was inflated in the dataset suspecting that number 10 was entered in the ODK even though farmers had more plots (supplementary material). Tropical livestock unit (TLU) was calculated using the method by Jahnke (1982): cattle was assumed to be equivalent to 0.7 TLU, horses 0.8 TU, donkey and mules 0.7 TLU, goat and sheep 0.2 TLU, and chicken 0.01 TLU. Cattle ratio and small ruminant ratio (goat and sheep)

was the ratio of their respective TLU and total household TLU. As the survey was focused on three crops, wheat, teff and sorghum, total production of only these three crops was considered. The number adults in the household (age 15 and above) was considered as a proxy for household labour availability. Household head education was a categorical variable with multiple levels (no education, adult education, primary, secondary, post-secondary). This variable was converted to binary where low education was primary and below primary level, and high education was above primary level. Other categorical variables were binary and used as they were in the dataset (Table 1).

Variables used for farm description

Crop area proportions were calculated based on the area of plots under a specific crop for a farming household. In few cases, there were multiple crops per plots, thus such households were left out during this calculation. As most of the households were not geo tagged (no coordinates), district (woreda) names were used to assess their spatial distribution in Ethiopia. As the district name was a text input in the ODK, many typing errors during the data collection resulted in over 50 unique names. Each name was assessed based on the spelling errors and the cleaning resulted in five unique districts in three different regions in Ethiopia. Land slope profile per plot was a categorical variable at plot level with three levels (flat, slope and steep slope). To transform the variable to farm level, proportion of plot area under three levels of slope (flat, slope, steep slope) was calculated, hence, we got three variables corresponding to the three levels of slope. Agricultural training was a nested variable where answers were on three levels, i) training received 1 year ago, ii) training received 5 years ago, iii) never received training. This variable was transformed to binary where 1 was training received (either 1 year ago or 5 years ago) and 0 was never received training.

In this report we mostly worked on the modules that provided information at farm level, identifying more than 90 variables. A total of 28 variables were used in the typology analysis and 34 were used for the description of the farms. The reasons for excluding the rest of the variables from both typology and farm description were: no variation, incorrect units, had many NAs due to incomplete data collection or they were not relevant to the hypothesis (supplementary material). The type of data was mostly categorical, representing 83% of the variables, while continuous data represented 17% of the variables. In the dataset, some households lacked plot level information for farm characteristics. Thus, before the PCA and FAMD, the dataset was filtered for household with plot level information which resulted in 848 households from 889. Further, the dataset was filtered for NA in at least one of the selected variables for each household and the final number of households for analysis was 758.

1.2.2 Statistical analysis

Analysis consisted of four steps, i) Data exploration, outlier handling and transformation, ii) Correlation analysis and variable selection, iii) Multivariate analysis (PCA and FAMD), and iv) Clustering and farm description (Alvarez et al., 2014, 2018; Barba-Escoto, 2019; Shukla et al., 2019). All the statistical analysis and graphical output were obtained using R software (version 4.3.1) with RStudio using

various packages (see below). Some limited amount of data handling was done using Microsoft Excel.

As distribution of the data is important for PCA, continuous variables were checked for normal distribution. Log and square root transformation was used to get the variables close to normal distribution as all the variables were skewed. Further, boxplots were created to check for outliers in the variables after transformation. Farm size, herd size and number of crops had few outliers for which the corresponding HHs were filtered out.

A correlation matrix was calculated for all the continuous variables using `cor` function from 'stats' package using Pearson's method. Variables with a correlation coefficient above 0.9 and variables without any significant correlation with any other variable were removed from the analysis (figures and tables in supplementary material). This step is necessary for dimension reduction as highly correlated or uncorrelated variables can create higher weightage or unnecessary noise (in terms of additional dimension) in the data respectively. For binary variables, percentages of the values (1,0) were calculated and variables with 90% of either 1 or 0 were removed from the analysis. This type of filtering of categorical variables improves the quality of the dataset by removing the variables with low variation. This prevents unnecessary noise in the dataset during analysis. Steps till now were repeated for all the typologies.

For the first typology, principal component analysis (PCA) was used as only continuous variables were selected. We used `dudi.pca` from 'ade4' package for PCA. For the second typology, we used a combined method with Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) called Factor Analysis for Mixed Data (FAMD) (Husson et al., 2008). This method allows using categorical variables along with continuous variables in their original format. For both the typologies, axes with eigenvalue more than one (as per the Kaiser criterion) were retained for clustering.

Hierarchical clustering was performed using Ward's method with a distance matrix generated in PCA and FAMD. Further, 'Nbclust' package was used to derive the optimal number of clusters using "ward.D2" and "kmeans" methods. This package provides a summary of 23 indices, and the cluster with the maximum number of indices was selected. If the highest number of indices indicates 3 clusters, then that specific number of clusters was chosen. This was further checked with subjective inspection of the dendrogram after which the number of clusters was finalised. After farm types were derived, Kruskal Wallis method was used to analyse the variance among farm types in continuous variables. Categorical variables were analysed for variation using percentages and heat map.

Table 1. List of variables used in typologies (T1 and T2) for two Use Cases in Ethiopia. Typology 1 resulted in three different farm types (HIW, MIF, LIS) that were subsequently used in T2. Continuous variables from T1 were retained for T2.

Category	Variable	Code	Description	Unit	Data type	T1	T2		
							HIW*	MIF*	LIS*
Household	Household size	hysize	Number of members in the household	number	continuous	x	x	x	x
	Education of household head	person_education	Education level of household head	Binary	categorical		x	x	
Labour	HH labour availability	count_adults	Only adults are taken into account (from hh size)	number	continuous	x	x	x	x
	Homegarden	homegarden	Presence of homegarden in the farm	binary	categorical		x	x	x
	Oxen	oxenpair	Presence of draught power in the farm	binary	categorical			x	
Farm characteristics	Farm size	farmsize	Total area of the farm	ha	continuous	x	x	x	x
	Number of crops	num_crops	Number of crops on farm	number	continuous	x			
	Number of rent in plots	rent_in_land	Number of plots rented in	number	continuous		x	x	x
Livestock	Herd size	tlu	Total herd size	TLU	continuous	x	x	x	x
	Cattle	cattleratio	ratio cattle: total based on TLU	-	continuous		x		
	Small ruminant	smallrumratio	Ratio small ruminants: total based on TLU	-	continuous		x		
Income	Off farm income	offfarm_income_any	hh members working outside the farm	binary	categorical		x	x	x
Main crops production	Teff production	teff_prod	Teff production	kg	continuous	x	x	x	x
	Sorghum production	sorghum_prod	Sorghum production	kg	continuous	x		x	x
	Wheat production	wheat_prod	Wheat production	kg	continuous	x	x		
Agricultural inputs and practices	Compost	compost	Use of organic fertilizer	binary	categorical				x
	Manure	manure	Use of manure	binary	categorical			x	x
	Pesticides	pesticides	Use of pesticides	binary	categorical		x	x	
	Hybrid seeds	hybridseeds	Use of improved seeds	binary	categorical		x	x	x
	Fertilizer input rates	fert_input_rates	Chemical fertilizer inputs/ farm size	kg/ha	continuous	x	x	x	x
	Fertilizer on wheat	fertilizer_crops.wheat	Crops receiving fertilizer	binary	categorical				
	Fertilizer on teff	fertilizer_crops.teff	Crops receiving fertilizer	binary	categorical		x		
	Fertilizer on sorghum	fertilizer_crops.sorghum	Crops receiving fertilizer	binary	categorical				x
	Use of Urea	urea	Use of urea on farm	binary	categorical			x	x
	Use of NPS*	NPS	Use of NPS on farm	binary	categorical		x		
Use of NPSB*	NPSB	Use of NPSB	binary	categorical		x	x		
Access to technology	Agricultural training	agtraining	Sum of hh members with training	binary	categorical		x	x	x
	Household membership	hh_membership	Household having any membership	binary	categorical		x	x	x
Sample size (households)						758	286	338	134

*Note: HIW (High intensified Wheat), MIF (Medium intensified farms), LIS (Low intensified Sorghum) farm types were identified in T1. Selection of variables in T2 included all the already selected variables in T1, with exception of “number of crops”. NPS- Nitrogen, Phosphorus, Sulphur; NPSB- Nitrogen, Phosphorus, Sulphur, Boron.

3. RESULTS

1.3 TYPOLOGY 1 – REGION LEVEL

Farm types were defined using PCA results in which 66% of the variability is explained by three principal components (Figure 1). The first dimension was related to the production of main crops (wheat and sorghum) and the fertilizer input rates, the second dimension was related to the characteristics of the farm (size of the farm, TLU), and the third dimension was related to the household (size of the household, number of adults). Teff production showed a lower variability among the dimensions than the other main crops, indicating that this crop is present in all types of farms but in variable proportions.

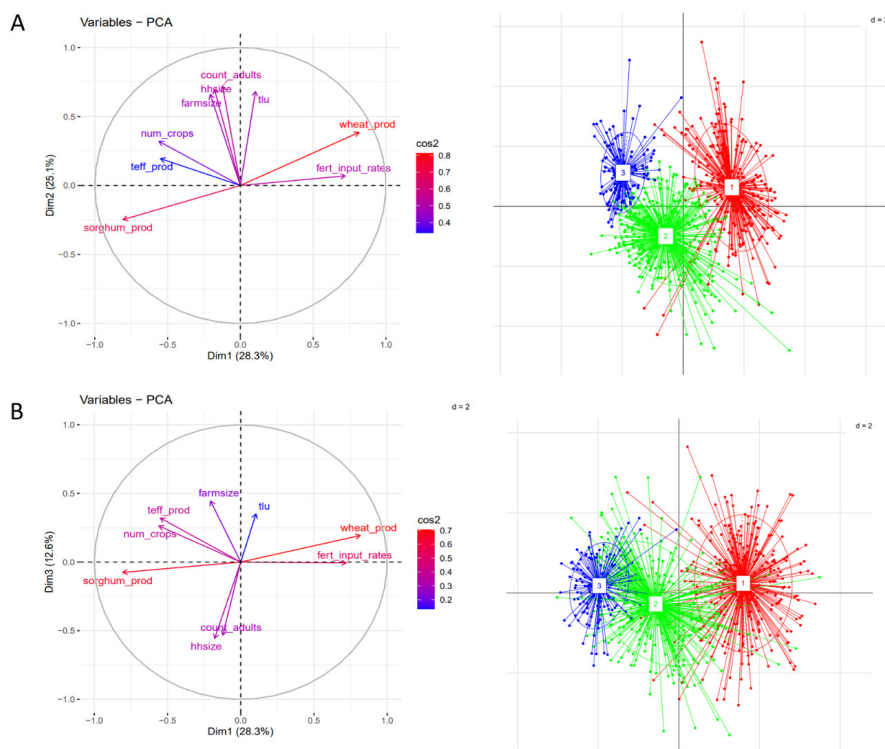


Figure 1. Typology 1- Output of PCA and clustering analysis defined by three principal components (Dim 1, Dim2, Dim 3). A) PC1 and PC2, B) PC1 and PC3. The gradient from red to blue colour indicates the weight of the variable on the axes of the PC space.

The three resulting farm types correspond to Farm type 1 (called HIW): highly intensified wheat farms in highlands, representing 38% of the 758 farms, Farm type 2 (called MIF): medium intensified diverse farms in mixed altitudes, representing 45% of the total farms, and Farm type 3 (called LIS): low intensified sorghum farms in lowlands, representing 17% (Figure 1). Most of the HIW farms belong to Goba and Siyadebirnawayu districts, which are located in the highlands. MIF farms are in both high and lowland districts like Kewet, Lemo and West-Belessa. The third farm type, LIS, is mainly located in the West-Belessa district, which is a lowland area. Highly Intensified Wheat (HIW) farms have high fertilizer input rates followed by MIF and LIS. Farms that are HIW have a higher proportion of highly educated household heads compared to MIF and LIS farms. HIW farm type has low proportion of farmers applying manure (13%) mainly due to focus on fertilizers. MIF farm type has 31% and LIS farm type has 63% farmers applying manure where higher number in LIS correlates with larger herd sizes and thus higher manure availability in this farm type.

In the entire dataset, all farmers apply fertilizers to their crops. In HIW, 97% apply to wheat followed by 61% to teff. For MIF and LIS, more proportion of farmers apply fertilizers on teff than sorghum. Compound fertilizers like NPS (19%N, 38% P₂O₅ and 7% S) and NPSB (18.9 % N, 37.7 % P₂O₅, 6.95 % S, and 0.1 % boron) are common in HIW, but in MIF and LIS only NPSB. Urea is common in all farm types. Around 30% of farmers in each farm type have off-farm income sources and the proportion of off-farm income to total income is around 13%. Thus, selling farming products is the main income source for all the farmers in all farm types. Area proportion of crops other than wheat, sorghum and teff is significant (around 40%) and important in all the farm types, especially in MIF and LIS where wheat is less predominant (Table 3). Thus, even though the survey was designed for the three main crops, the farms are in general very diverse.

Recalling the hypothesis of this first typology (which suggest that agricultural households can be grouped according to their production system such as wheat, teff and/or sorghum), we confirm the clear definition of three farm types determined by agro-ecologies (e.g. highland, lowland) and production systems. This confirmation will greatly contributes in the improvement of the Use Case MVP, specifically focusing on the cropping systems. However, no variation among the farm types was observed in variables other than mentioned above which were selected for the farm description. This implies that the current grouping is very general, prompting the need for a more detailed typology within these farm types, hence the objective 2.

1.4 TYPOLOGY 2 – CROPPING SYSTEM LEVEL

The general outcomes of the initial typology demonstrates the collaboration of two distinct Use Cases operating in overlapping locations and farming systems. To further identify farm types that effectively support the implementation of the Use Case MVP, a second typology was executed. Typology 2 was performed using farm types identified in typology 1 and by adding more continuous variables related to the hypothesis of objective 2 and categorical variables for better detailed farm description (Table 1). Typology 2 was divided into 3 sections, corresponding to the already identified farm types of typology 1: i) Farm type 1 (HIW): high intensified wheat farms in highlands, ii) Farm type 2 (MIF): medium intensified sorghum farms in mixed altitudes, iii) Farm type 3 (LIS): low intensified sorghum farms in lowlands.

Based on the objectives, the farms in typology 2 were described based on resource endowment (RE) and production orientation. Farm size, herd size and oxen-pair ownership were considered for resource endowment. The thresholds for the levels of resource endowment are calculated through an iterative process with the local teams, in our case the Use Case leaders and studying the average and dispersion of the variables (Tittonell et al., 2010). The thresholds for the levels of resource endowment (low, medium and high) were calculated using three quantiles for each of the variables. For instance, 33% and 67% quantile was calculated for farm size in the whole dataset, and farmers below 33% were Low Resource Endowment (LRE), between 33% and 67% were Medium Resource Endowment (MRE) and above 67% were High Resource Endowment (HRE). These quantiles are assumptions prior to the typology construction, validation of the boundaries is requested from the Use cases.

Crop use (sold or consumed), land tenure status (renting-in land) and fertilisation was considered in production orientation. Based on production orientation, farmers were categorised into three types: Market oriented, mixed oriented and subsistence oriented. Farm types were considered market oriented when they had relatively high proportion of crop sold in markets, a high proportion of land rented in, and high fertilisation rates, which is the case for some HIW farms. Mixed oriented farm types were farms who were mostly subsistence oriented but exhibited some market-oriented characteristics, as seen in farms like MIF and LIS situated in the West Belessa district, which cultivate sorghum and teff.

Figure 2 offers an overview of the various farm types identified in both typology 1 and 2. The labels used in the diagram remain consistent throughout this report.

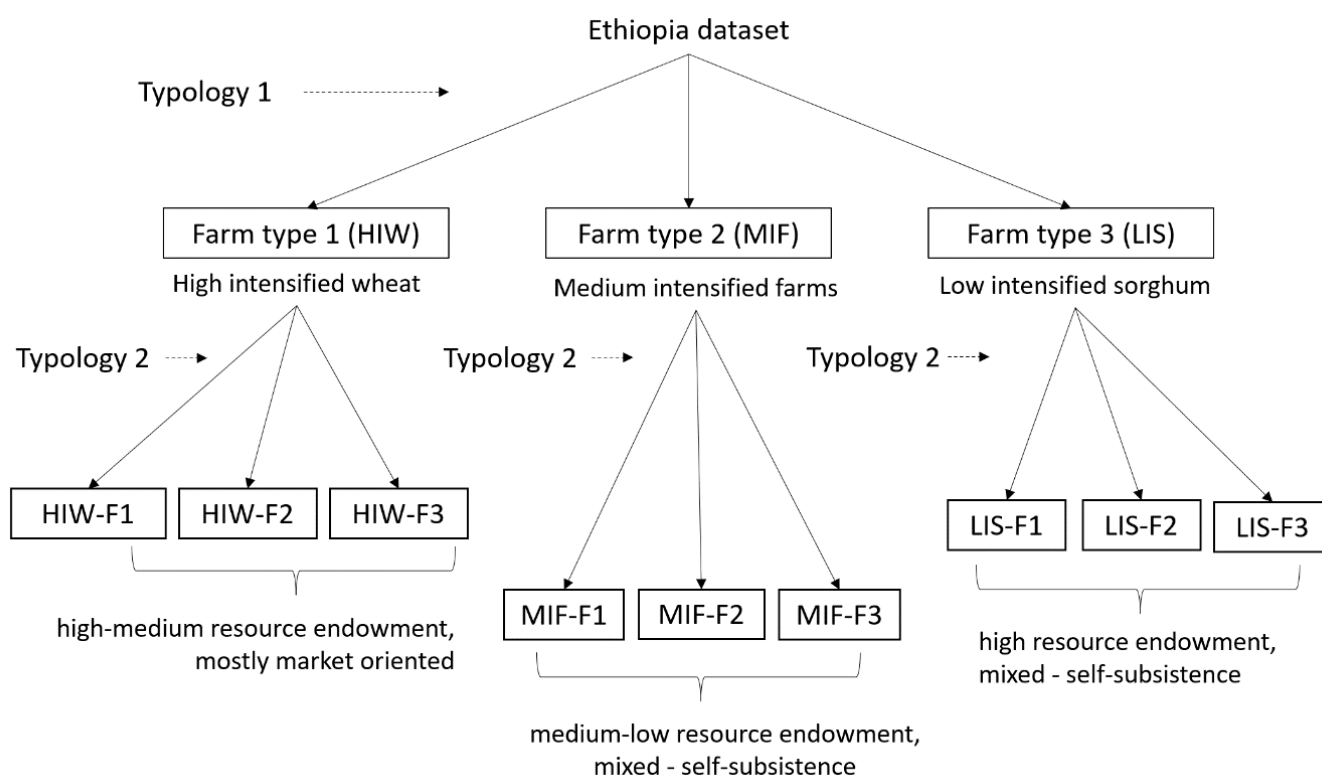


Figure 2. General overview of the identified farm types. In the first typology, intensification (high, medium and low) of the crop types (wheat and sorghum) were labelled accordingly. Farm types identified in typology 2, designed as F1, F2 and F3, were named uniquely based on the outcomes derived the typology 1.

1.4.1 High intensified wheat farms in highlands (HIW)

All continuous variables used in T1 were retained in this typology except sorghum production as the majority of the farmers were predominantly wheat and teff farmers. The selection of categorical variables was done based on the 10% variation rule mentioned above in the methods. For instance, over ~90% of household heads were male, did not use manure and compost, used urea and fertilized wheat with a type of fertilizer. Thus, categorical variables HH head gender, manure use, compost use, urea use and chemical fertilizer used on wheat were not included in the FAMD. The FAMD resulted in 7 dimensions explaining 65.9% of the data of which

first two are represented in the Figure 3. Within FAMD, there are two analysis, one PCA for continuous variables and other MCA for the categorical variables. For PCA, first dimension was related to teff-production and second dimension related to farm size and TLU. For MCA, first dimension was related to fertilizer type applied and second dimension was related hybrid seeds and pesticides use. The clustering resulted in three farm types.

Type 1 (HIW-F1): High resource endowed, market oriented, wheat dominant farms (n=101)

This farm type is mostly present in Goba district. Features: The fertilizer type use is NPS rather than NPSB, high percentage of hybrid seed users, high percentage of memberships in social groups, 15% land rented-in, high proportion of wheat sold, households with high proportion of plots on slope have high fertilizer application rates compared to other farm types, similar pattern was observed by Desta et al. (2023).

Type 2 (HIW-F2): Medium resource endowed, mixed oriented, diverse farms (n=58)

This farm type is mostly present in Lemo district. Features: High proportion of home gardens and relying on off-farm income, low percentage of memberships in social groups and low hybrid seeds use, Low percentage of rented-in land, high proportion of wheat consumed, around 48% of household face food shortage (highest compared to other types).

Type 3 (HIW-F3): High resource endowed and market oriented, high intensified, wheat-teff farms (n=127)

This farm type is mostly present in Siyadebirnawayu district. Features: highest fertilizer application rates, fertilizer type use is NPSB and Urea, high use of hybrid seeds and pesticides, high proportion of trained farmers and memberships in social groups, high proportion of wheat sold.

Overall, though the proportion of memberships in this type is high, most farmers of them are in social welfare groups, thus very few in social credit groups or cooperatives. For digital access, around 40-45% farmers in HIW-F1 and HIW-F2 have smartphones opposed to 22% in HIW-F3. Around 60-70% farmers have cell phones in all farm types.

In terms of dissemination of information preferences, most of the farm types prefer through government and relatives but HIW-F1 and HIW-F3 have other preferences like radio, TV (Supplementary material).

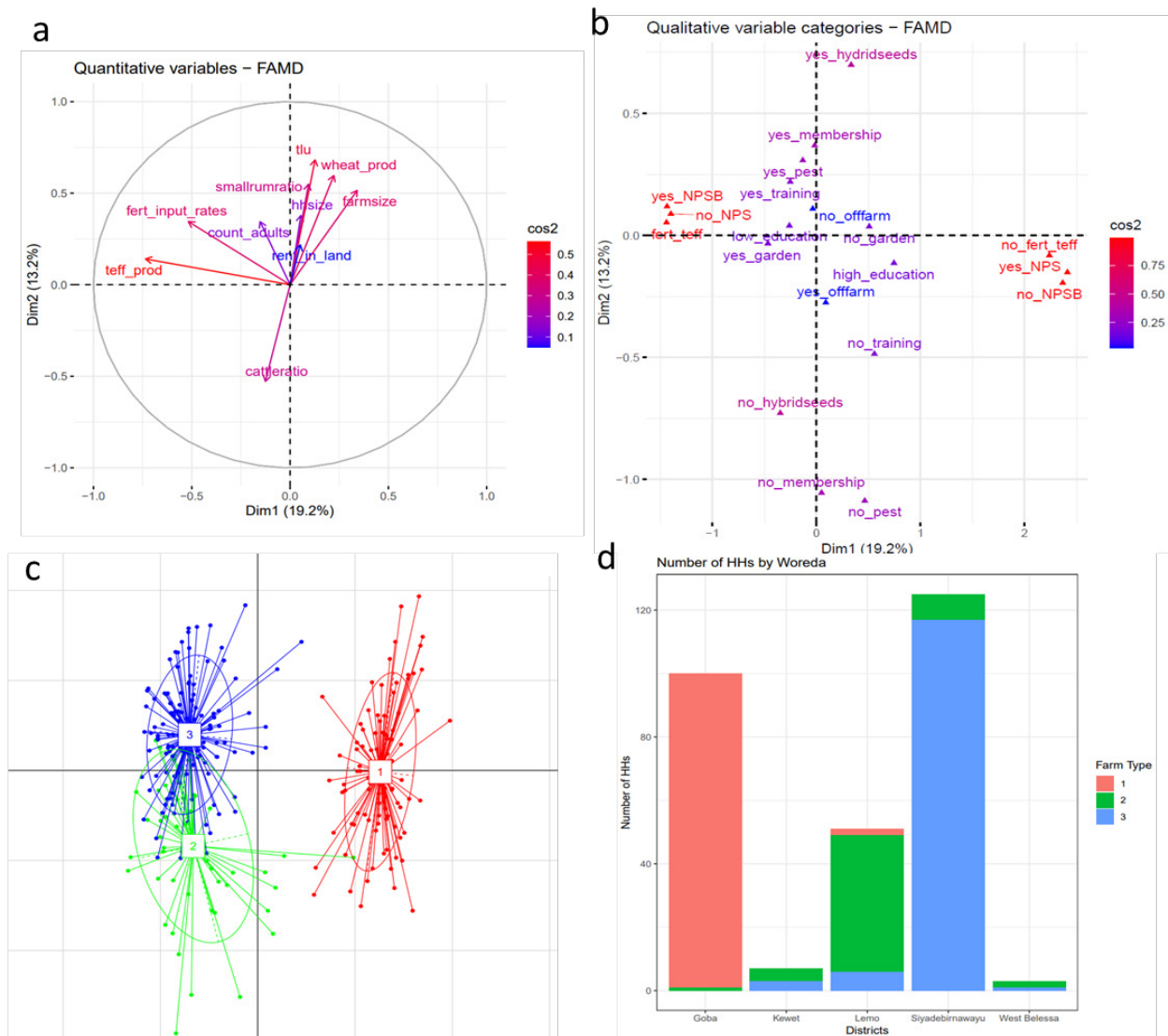


Figure 3. Typology 2 (HIW)- Output of PCA dimension 1 and 2 (a), MCA dimension 1 and 2 (b), clustering analysis (c) and the identified farm types (d) among high-intensified wheat farmers in highlands (HIW).

1.4.2 Medium intensified diverse farms in mixed altitudes (MIF)

All continuous variables used in T1 were retained in this typology except wheat production as majority of the farmers were predominantly sorghum and teff farmers. The FAMD resulted in 8 dimensions explaining 62% of the variation (Figure 4). The first dimension was related to sorghum production and membership, and the second dimension was related to household size, HH labour availability, manure use and home garden ownership. Most of the sorghum and teff (60-70%) produced by these farms is consumed at home.

Type 1 (MIF-F1): Medium resource endowed and mixed oriented sorghum-teff farms (n=207)

This type of farms is mostly located in West Belessa and Kewet district. Features: High proportion of farms rent in land, have low fertilizer input rates but most apply manure and pesticides, over 75% of the farmers have memberships in social groups.

Type 2 (MIF-F2): Medium resource endowed and mixed oriented sorghum farms (n=75)

This type of farms is mostly found in Kewet. Features: Low proportion of farms rent in land, have high fertilizer application rates, high proportion of farmers have oxen pair, low manure application, 50% farmers have memberships.

Type 3 (MIF-F3): Low resource endowed and self-subsistence teff farms (n=56)

This type of farms is mostly located in Lemo district. Features: Low fertilizer application rates, 30% area under teff and rest on other crops, higher proportion of households have homegarden, low membership, low manure use.

Even though MIF-F2 and MIF-F3 have significantly smaller farm sizes than MIF-F1, they have higher fertilisation rates. This suggests that the former types apply more fertilizer to overcome land constraints compared to the latter type. Overall, 50% of the households face food shortage where MIF-F3 which is LRE and subsistence-oriented have 64% farmers who face food shortage. Farms with high percentage of plots on slopes have low fertilizer application rates (MIF-F1), unlike what was observed in previous farm type such as HIW-F3. Most of the farmers did not use hybrid seeds in this group. Memberships in these farm types are in social welfare groups and social credit groups but very few in cooperatives and water committee. For digital access, around 25-40% farmers in all the farm types have smart phones and 60-70% have cell phones.

In terms of dissemination of information, 80% of farmers in MIF-F1 and MIF-F3 prefer through government whereas 60% in MIF-F2. 40-50% farmers prefer through relatives in MIF-F1 and MIF-F2, whereas 22% in MIF-F3 (Supplementary material). Other modes are preferred by less than 10% of the farmers and thus, the preference for other modes is low.

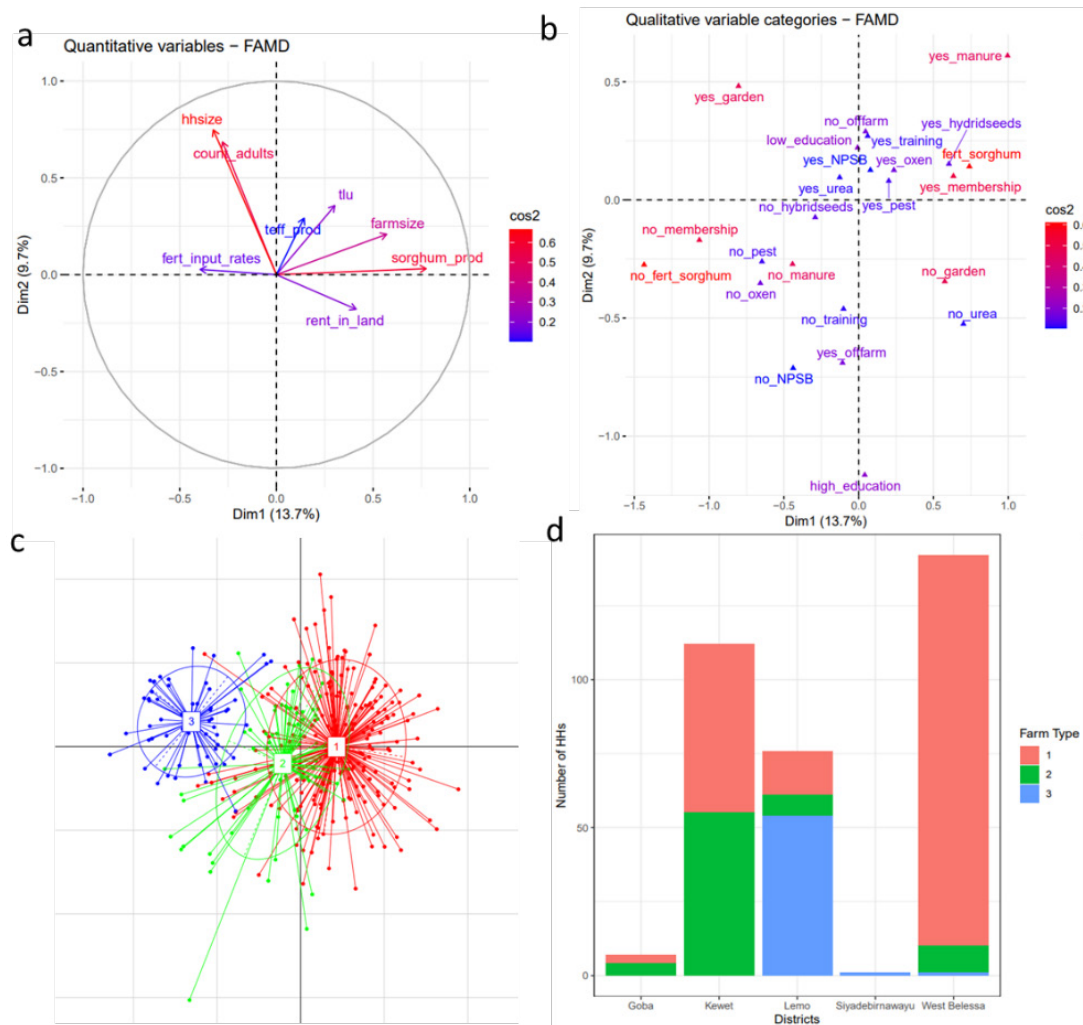


Figure 4. Typology 2 (MIF)- Output of PCA dimension 1 and 2 (a), MCA dimension 1 and 2 (b), clustering analysis (c) and the identified farm types (d) among medium-intensified sorghum farmers in mixed altitudes (MIF).

1.4.3 Low intensified sorghum farms in lowlands (LIS)

All continuous variables used in T1 were retained in this typology except wheat production as majority of the farmers were predominantly sorghum and teff farmers. The FAMD resulted in 7 dimensions explaining 62% variation in the data (Figure 5). The first dimension is related to TLU, fertilizer application rates and second dimension is related to manure use, membership and sorghum production. Almost all the farm types are in West Belessa district.

Type 1 (LIS-F1): High resource endowed farms (HRE), self-subsistence sorghum-teff farms (n=85)

Features: High manure application rates, high proportion of farmers with memberships, 38% of farmers have homegarden, 60% farmers with hybrid seeds, low percentage of farms rented-in land and low proportion of produce sold in markets. They are HRE because of large farm size and herd size but self-subsistence because of low fertilizer application rates, low land renting-in and low proportion of crop produce sold in markets.

Type 2 (LIS-F2): High resource endowed farms (HRE), mixed oriented sorghum-teff farms (n=24)

Features: Low manure application rates, low proportion of farmers with membership, high percentage of farms rented-in land but low proportion of crop produce sold in the markets, low home gardens proportions. They are HRE because of high farm size and TLU. They are mixed oriented (self-subsistence and low input market oriented) because of low inputs but high land renting in which suggests strategy to increase production.

Type 3 (LIS-F3): High resource endowed farms, mixed oriented teff farms (n=25)

Features: Farms with the lowest fertilizer application rates in the entire survey region, less farms apply fertilizers on sorghum, high proportions of land renting and high proportion of households with home gardens. Similar to LIS-F2, these farms are HRE because of high farm size and TLU. They are mixed oriented (self-subsistence and low input market oriented) because of low application of inputs but high land renting in which suggests strategy to increase production.

Overall, 40% of the households face food shortage but lower proportion of farmers in LIS-F2 compared to LIS-F1 and LIS-F3. Memberships in these farm types are in social welfare groups and social credit groups but very few in cooperatives and water committee. For digital access, 16-20% farmers in LIS-F1 and LIS-F2 have smartphones, but only 8% farmers in LIS-F3 have smartphones. Most farmers in all the types have cell phones.

In terms of dissemination of information, 85-95% of farmers in all the farm types prefer through government, 30-50% farmers prefer through relatives. 17% in LIS-F3 prefer through traders and 26% in LIS-F3 prefer through radio. Rest all modes of dissemination are preferred by less than 5% of the farmers, hence negligible (Supplementary material).

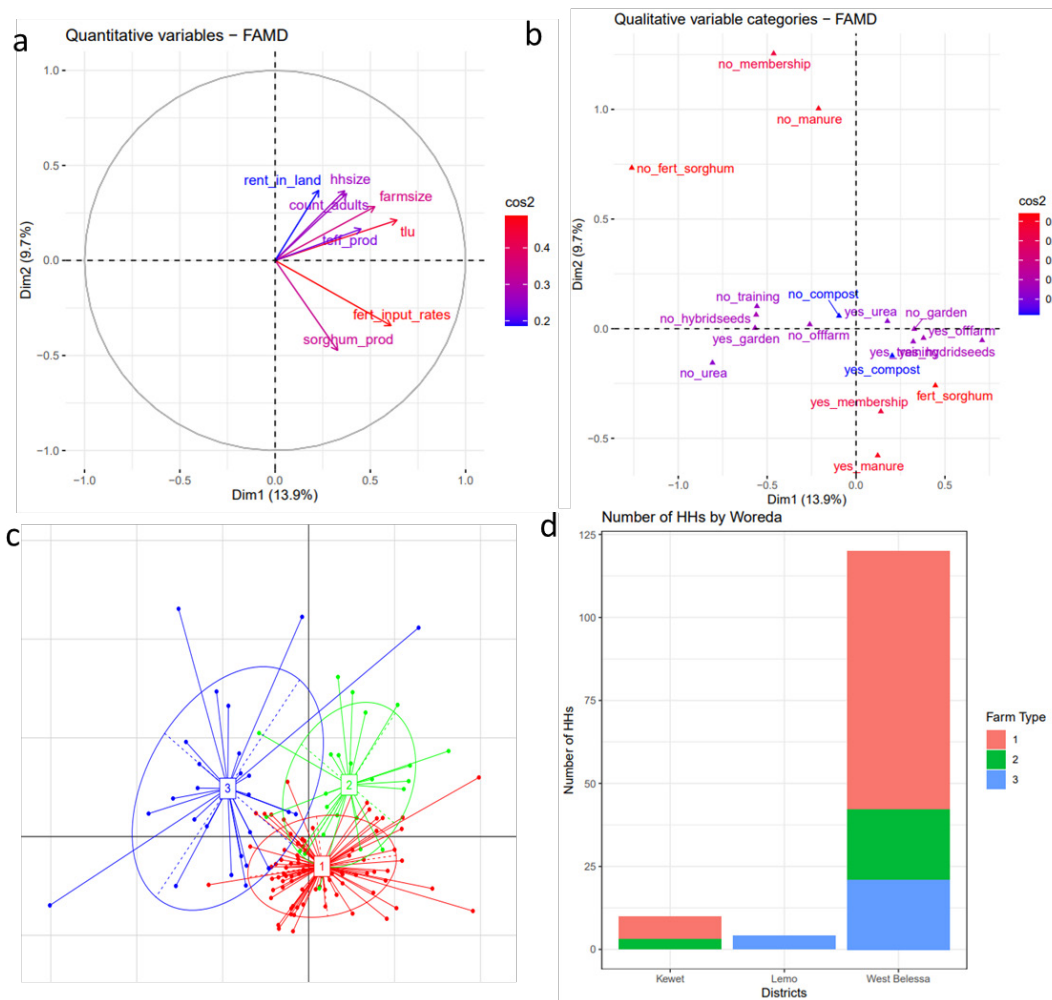


Figure 5. Typology 2 (LIS)- Output of PCA dimension 1 and 2 (a), MCA dimension 1 and 2 (b), clustering analysis (c) and the identified farm types (d) among low-intensified sorghum farmers in lowlands (LIS).

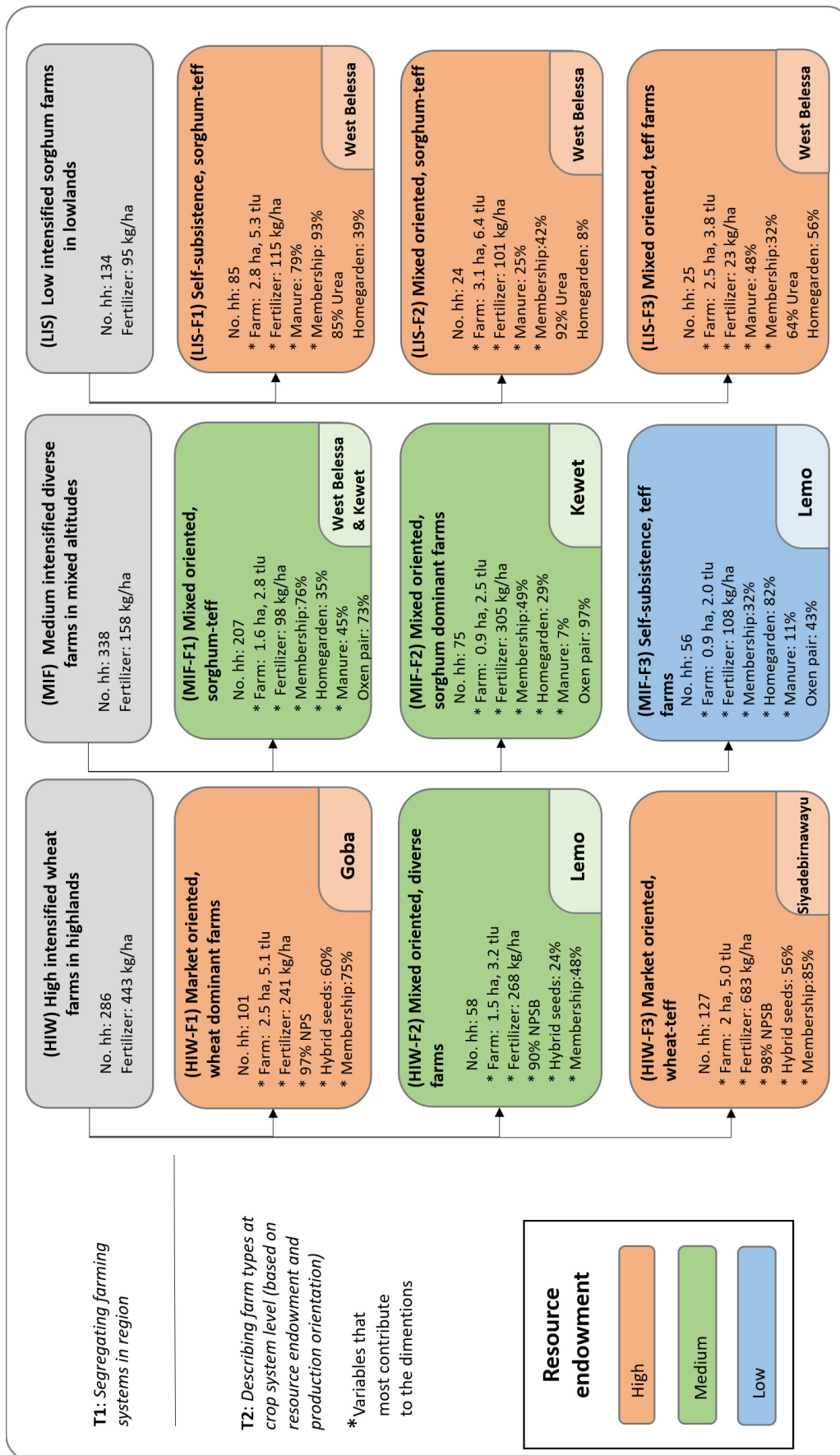


Figure 5. General overview of the outcomes of typologies 1 and 2. The grouping of the resulting farm types is represented in colours, according to their resource endowment. The production orientation and main crop(s) are indicated in the farm type name; the district where the farm type is mostly located is presented in the lower right corner; main farm type characteristics are listed.

4. DISCUSSION

Hypothesis based statistical farm typology is often used to segment farmers in a region based on their structural and functional characteristics (Alvarez et al., 2018). The variable selection depends on the scale of analysis, objective and the hypothesis of the local experts, in our case the Excellence in Agronomy (EiA), Ethiopia Use Cases. Usually, the survey questionnaire during data collection is designed based on the objective and the hypothesis of the typology. In this study, it was partially the case as the surveys were conducted in wheat, sorghum and teff growing regions, which were the Use Case locations, based on common objectives regarding farmer segmentation in the two Use Cases generalized. The data collection was generalized across various farm characteristics with a mix of structural and functional variables of continuous and categorical nature. To dive into the specific objectives of the Use Cases, exploratory analysis and segmentation of dataset was necessary to generate a meaningful typology of the system. Thus, two typologies were constructed, first on the entire dataset, at regional level, to slice the dataset into multiple sub-datasets and second typology for each of the sliced portions of datasets separately based on the farm types generated in typology 1.

Previous studies handled region level datasets by performing typologies by districts (Alvarez et al., 2018; Tittonell et al., 2010). In our case, the sample sizes per district were low to include all the relevant variables for the analysis and the agroecological difference was mostly on altitude differences (high and low), thus the district (woreda) variable was left out. The continuous structural and functional variables in the dataset were used to segment farms at regional level. Despite the exclusion of districts from the analysis, a distinct segmentation of farms based on agroecology was evident. Three distinct farm types were identified, aligning with three groups characterized by agroecological factors typical for their altitudinal location.

The categorical structural and functional variables provided a detailed description of farm characteristics. Using them initially along with the continuous variables in the first typology caused “masking” of the variation, and hence they were left out. Once the dataset was segmented based on the general variation among the continuous variables in typology 1, categorical variables proved vital to reveal the “hidden” differences among the farm types in typology 2.

Differences in farm types according to their resource endowment and production orientation

The first typology segmented the farms in two cropping systems, wheat dominant and sorghum dominant, and further segmented sorghum dominant systems in two groups: medium and low intensified based on fertilisation as wheat dominant system were high intensified due to highest fertilizer application rates. The association between resource endowment and cropping systems was not observed, as farmers engaged in both wheat and sorghum cultivation exhibited varying levels of resource endowment. Production orientation was associated to cropping systems as wheat farmers were either market oriented or mixed oriented (self-subsistence and market oriented). For wheat farmers, market and mixed orientation was mostly associated to high fertilizer application and high share of produce sold in market. For sorghum farmers, mixed orientation was mostly associated to farmers renting-in land, which can

be seen as a strategy to expand production (Berre et al., 2019). Sorghum farmers exhibited low levels of market-oriented behaviour, as indicated by the minimal share of their products sold in the market and the low application of fertilizers.

The validation of these farm characterizations involved examining a subset of continuous and categorical variables. Home gardens play a crucial role in Ethiopian farms, contributing significantly to household food security. According to Motbainor et al. (2022) households with home gardens prioritise food security more than those without. Farm types categorized as mixed-oriented or self-subsistence-oriented exhibited a higher percentage of home gardens compared to market-oriented farm types. Notably, farmers exhibiting both medium-low resource endowment (RE) and predominantly self-subsistence orientation, as well as those who own home gardens within the same group, experience food shortage. Contrastingly, a lower proportion of farmers in farm types with market orientation and HRE experience food shortage with few exceptions like LIS-F1 and LIS-F3. Even though these are HRE farms, their low intensified system might be a main driver for their food insecurity. Overall, wheat dominant farms are more food secure than sorghum dominant farms. Memberships can be an indicator of social status. Similar patterns like in food shortage is observed for membership where HRE and market-oriented farms have high participation in social groups.

Engaging in off-farm activities implies that the household relies on income beyond what is generated from their farm produce (Berre et al., 2019). In the case of high intensified wheat (HIW), a greater proportion of mixed-oriented farmers had off-farm income compared to market-oriented farmers. Similarly, for high intensified wheat (HIW), greater proportion of farmers with market orientation used hybrid seeds compared to mixed oriented farmers. This pattern was not consistent for MIF and LIS as most of the farmers had self-subsistence orientation. Ownership of devices like smartphones and cell phones are also indicators of resource endowment (van den Brand, 2011)

In this study, there was no variation in ownership of devices among any of the farm types. Households with self-subsistence and mixed orientation had higher number of crops on the farm compared to farmers with market orientation. Mutyasira (2020) observed this crop diversification in low resource endowed farmers. In this study, this phenomenon is more associated with production orientation rather than resource endowment.

Differences in farm types within different districts

Farmers in Goba and Siyadebirnawayu are similar in terms of resource endowment with farmers in West Belesa. Due to different cropping system and production orientations, the former types are highly intensified and the latter are low intensified but have high potential of intensification.

There are two farm types in Lemo district, HIW-F2 and MIF-F3. These two farm types are both diversified and the proportion of crops other than wheat, and teff was high. Even though they are in the same district, there are striking differences in these two farm types. HIW-F2 are medium resource endowed with wheat as a main crop whereas MIF-F3 are low resource endowed farms with teff as their main crops. HIW-F2 has significantly higher fertilizer application than MIF-F3 and are mixed oriented as opposed to self- MIF-F3 being subsistence. This suggests that wheat farmers are more market oriented and their farms are intensified compared to sorghum farmers. In this case, production orientation is a driver of wealth, as found earlier in Malawi (Franke et al., 2014). This suggest that market-oriented farmers are more likely to adopt new production enhancing technologies compared to those oriented towards self-subsistence.

There are two farm types in Kewet district, sorghum-teff (MIF-F1) and sorghum dominant farms (MIF-F2). Both these farm types are medium resource endowed, where most of the farmers have oxen-pair, whose ownerships is associated with capital availability (Silva et al., 2019, 2021). Both these farm types are categorised as mixed oriented yet they have different strategies for production. Sorghum-teff farms are larger (1.5 ha) compared to sorghum dominant farms (0.9 ha) and increase production by renting-in land and mostly rely on manure from the cattle as inputs on farm. On the other hand, sorghum dominant farms apply more fertilizer rather than renting-in land and use less manure even though they have similar cattle numbers as the sorghum-teff farmers. Thus, sorghum dominant have high intensified production and sorghum-teff have high extensified production strategy. Previous studies report farming households facing land pressure tend to allocate their land to income-generating crops and increase intensification (Berre et al., 2019; Boere et al., 2016). At Kewet, technologies for sustainable intensification of these two farm types could include crop-livestock integration (fertilizers and manure management) to address production and land pressure.

West Belesa is distinguished from the other districts by having the largest farm size and TLU, despite maintaining the lowest rate of fertilizer input (Table 3). Variations between farm types in West Belesa mainly revolve around crop selection and production orientation. Those exclusively producing teff show significantly lower input rates and a higher proportion of home gardens compared to sorghum-teff producers. Furthermore, the degree of production orientation (from self-subsistence to market) is linked to the need to rent additional land for cultivation, as highlighted by Mutyasira (2020). These findings suggest a high production capacity and potential for intensification, particularly encouraging synergies between crop and livestock production across all farm types in West Belesa.

Limitation of the dataset for the analysis

The details on exclusion and limitations in variable selection are in the supplementary material, few examples are discussed here. Notably, the variable of labour utilization and hired human power (measured in human-hours per day or week) essential for assessing production orientation and resource endowment was absent in the dataset. However, the data on labour use for each management practice can be found in Part III of the survey tool which is at plot level. We therefore used a proxy variable for labour, such as the presence of adults in the household. However, this proxy failed to provide a comprehensive depiction of the labour category in the typology as it highly correlated with household size.

Position of the plots in the landscape was an important variable for the Use Cases in the study. It was collected at three levels (flat, slope, steep), which was inconsistent with how it was measured in the past (valley bottom, lower slope, middle slope, upper slope, plateau) in the same study regions (Desta et al., 2023). Further, fertilizer application was measured at farm level for the Farmer Segmentation (Part I & II). The plot level measurements of fertilizer application rates are collected under Part III. This caused loss of information in estimating the differences in fertilizer application by slope type and also crop type. When the slope information was included in the current form, the variables did not correlate significantly with any other variables in the dataset and hence were not included in the typology analysis but were used for farm description.

5. CONCLUSIONS

In this study, two typologies at different scales were used to identify farm types to customize MVP advisory in three regions of Ethiopia. Our method showed that typology using only continuous variables explaining general farm characteristics in the farming system may lead to blanket recommendations due to over generalization of the typology. Combined use of both continuous variables defining broad trends in the system and categorical variables at multiple scales is necessary for detailed farm characterization. In the region level typology, farm types were segmented primarily based on cropping system and fertilization intensity. In the farming system level typology, farms were segmented based on resource endowment and production orientation. Drawing upon the second typology hypothesis, which states that farming households can be categorized based on their resource endowment (measured by TLU and farm size) and production orientation, we affirm that three distinct farm types were derived within each initially segregated farms identified in typology 1. The nine farm types identified in five districts of Ethiopia differed in cropping system, land size and tenure status, capital availability based on livestock ownership and production orientation, all of which has an influence on their farm management and type of intervention/solutions necessary for intensification. We conclude that the wheat farmers mostly have high resource endowment and are market oriented with high social standing and high food security. Wealthy sorghum farmers are also similar to wealthy wheat farmers except production orientation is self-subsistence, thus have a high potential of intensification through targeted interventions. Dissemination through government and relatives is highly preferred by all the farm types irrespective of resource endowment and production orientation. Digital access was not correlated with resource endowment.

6. RECOMMENDATIONS

- The Use Cases need to have interactive discussions with the demand partners to validate the typologies generated and tailor their MVPs for each Farm Types for effective scaling.
- There is a need to identify specific strategies to disseminate MVPs based on the requirement of each Farm Type identified.
- Further analysis can be carried out to identify specific farm typologies based on the need of Use Cases.
- Expand the network of partners for strategic dissemination and implementation of the MVP.

7. REFERENCES

- Alvarez, S., Paas, W., Descheemaeker, K., & Tiftonell, P. (2014). Typology construction, a way of dealing with farm diversity General guidelines for Humidtropics. <http://humidtropics.cgiar.org>
- Alvarez, S., Timler, C. J., Michalscheck, M., Paas, W., Descheemaeker, K., Tiftonell, P., Andersson, J. A., & Groot, J. C. J. (2018). Capturing farm diversity with hypothesis-based typologies: An innovative methodological framework for farming system typology development. *PLOS ONE*, 13(5), e0194757. <https://doi.org/10.1371/journal.pone.0194757>
- Barba-Escoto, L. P. A. K. L.-R. S. R. J. M. L. , T. J. P. A. S. . (2019). Researchers' Manual for Quantitative Farming Systems Typologies Applications using the R Statistical Tool .
- Berre, D., Baudron, F., Kassie, M., Craudurd, P., & Lopez-Ridaura, S. (2019). DIFFERENT WAYS TO CUT A CAKE: COMPARING EXPERT-BASED AND STATISTICAL TYPOLOGIES TO TARGET SUSTAINABLE INTENSIFICATION TECHNOLOGIES, A CASE-STUDY IN SOUTHERN ETHIOPIA. *Experimental Agriculture*, 55(S1), 191–207. <https://doi.org/10.1017/S0014479716000727>
- Boere, E., Mosnier, A., Bocqueho, G., Krisztin, T., & Havlik, P. (2016, September 23). Developing country-wide farm typologies: An analysis of Ethiopian smallholders' income and food security. 5th International Conference of the African Association of Agricultural EconoMIFts. <http://ageconsearch.umn.edu/handle/246924>
- CGIAR. (2023, November 3). Initiative Excellence in Agronomy. <https://www.cgiar.org/initiative/excellence-in-agronomy/>.
- Desta, G., Legesse, G., Agegnehu, G., Tigabie, A., Nagaraji, S., Gashaw, T., Degefu, T., Ayalew, B., Addis, A., Getachew, T., Managido, D., Bazie, Z., Abathun, T., Abera, A., Dache, A., Adissie, S., Sebnie, W., Feyisa, T., Yakob, G., ... Harawa, R. (2023). Landscape-based nutrient application in wheat and teff mixed farming systems of Ethiopia: farmer and extension agent demand driven approach. *Frontiers in Sustainable Food Systems*, 7. <https://doi.org/10.3389/fsufs.2023.1241850>
- 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. <https://doi.org/10.1016/j.eja.2014.04.002>
- Hassall, K. L., Baudron, F., MacLaren, C., Cairns, J. E., Ndhlela, T., McGrath, S. P., Nyagumbo, I., & Haefele, S. M. (2023). Construction of a generalised farm typology to aid selection, targeting and scaling of onfarm research. *Computers and Electronics in Agriculture*, 212, 108074. <https://doi.org/10.1016/j.compag.2023.108074>
- Husson, F., Josse, J., & Lê, S. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, 25(1). <https://doi.org/10.18637/jss.v025.i01>
- Jahnke, H. E. (1982). *Livestock Production Systems and Livestock Development in Tropical Africa* (ISBN 3.92255312-5). KIELER WISSENSCHAFTSVERLAG VAUK .
- Kuivanen, K. S., Michalscheck, M., Descheemaeker, K., Adjei-Nsiah, S., Mellon-Bedi, S., Groot, J. C. J., & Alvarez, S. (2016). A comparison of statistical and participatory clustering of smallholder farming systems – A case study in Northern Ghana. *Journal of Rural Studies*, 45, 184–198. <https://doi.org/10.1016/j.jrurstud.2016.03.015>

- Motbainor, A., Arega, Z., & Tirfie, M. (2022). Comparing level of food insecurity between households with and without home gardening practices in Zege, Amhara region, North West Ethiopia: Community based study. *PLOS ONE*, 17(12), e0279392. <https://doi.org/10.1371/journal.pone.0279392>
- Mutyasira, V. (2020). Prospects of sustainable intensification of smallholder farming systems: A farmer typology approach. *African Journal of Science, Technology, Innovation and Development*, 12(6), 727–734. <https://doi.org/10.1080/20421338.2019.1711319>
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Posit team (2023). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.
- Shukla, R., Agarwal, A., Gornott, C., Sachdeva, K., & Joshi, P. K. (2019). Farmer typology to understand differentiated climate change adaptation in Himalaya. *Scientific Reports*, 9(1), 20375. <https://doi.org/10.1038/s41598-019-56931-9>
- Silva, J. V., Baudron, F., Reidsma, P., & Giller, K. E. (2019). Is labour a major determinant of yield gaps in sub-Saharan Africa? \ A study of cereal-based production systems in Southern Ethiopia. *Agricultural Systems*, 174, 39–51. <https://doi.org/10.1016/j.agsy.2019.04.009>
- Silva, J. V., Reidsma, P., Baudron, F., Jaleta, M., Tesfaye, K., & van Ittersum, M. K. (2021). Wheat yield gaps across smallholder farming systems in Ethiopia. *Agronomy for Sustainable Development*, 41(1), 12. <https://doi.org/10.1007/s13593-020-00654-z>
- Tittonell, P., Muriuki, A., Shepherd, K. D., Mugendi, D., Kaizzi, K. C., Okeyo, J., Verchot, L., Coe, R., & Vanlauwe, B. (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. <https://doi.org/10.1016/j.agsy.2009.10.001>
- van den Brand, G. (2011). *Towards Increased Adoption of Grain Legumes Among Malawian Farmers – Exploring Opportunities and Constraints Through Detailed Farm Characterization*. Wageningen University.

8. APPENDIX

Table 2. Variables used for farm description

Category	Variables	Code	Description	Unit	Data type			
Household	District	woreda	Location base on district names		categorical			
	Age of household head	person_age	Age of the head of the household	years	continuous			
	Gender	person_gender		binary	categorical			
	Food shortage	food_security	Shortage of food at some point of a year	Binary	categorical			
Farm charc.	Wheat area proportion	wheat_area_proportion	Percentage of the total operated area allocated for Wheat	%	categorical			
	Teff area proportion	teff_area_proportion	Percentage of the total operated area allocated for teff	%	categorical			
	Sorghum area proportion	sorghum_area_proportion	Percentage of the total operated area allocated for Sorghum	%	categorical			
Livestock	Number of plots	num_plot	Number of plots on farm cultivated	number	continuous			
	poultry	poultryratio	ratio from TLU	%	continuous			
Technology	Smartphone access	smartphone	Score on the levels of access to smartphone	binary	categorical			
	Cellphone access	cell_phone	Score on the levels of access to cellphone	binary	categorical			
	Dissemination	Government	government.1	Mode of dissemination of information to farmers	binary	categorical		
		Private	private.1		binary	categorical		
		Trader	trader.1		binary	categorical		
		Association	association.1		binary	categorical		
		Relatives	relatives.1		binary	categorical		
		Other farmer	other_farmer.1		binary	categorical		
		Radio	radio.1		binary	categorical		
		TV	tv.1		binary	categorical		
		Social media	social_media.1		binary	categorical		
		App	app.1		binary	categorical		
		Water conservation practices	Soil stone bunds		soil_stone_bunds	Water conservation practices adopted by the households	binary	categorical
			Check dams		check_dams		binary	categorical
Cut off drain	cut_off_drain		binary	categorical				
Ridge furrow	ridge_furrow		binary	categorical				
Contour ploughing	contour_ploughing		binary	categorical				
Terraces	terraces		binary	categorical				
Membership	Strip planting	strip_planting		binary	categorical			
	Social welfare	social_welfare		binary	categorical			
	Saving credits	saving_credits		binary	categorical			
Slope	Water committee	water_committee	Memberships in various organizations/groups	binary	categorical			
	Multi cooperative	multi_cooperative		binary	categorical			
	Flat land	flat_land_area	proportion of flat land area on farm	%	continuous			
	Slope land	slope_land_area	proportion of slope land area on farm	%	continuous			
	Steep slope land	steep_land_area	proportion of land area with steep slope on farm	%	continuous			

Table 3. Farm type description for farm typology 1. The values are means with standard deviation in the brackets. Letters in superscripts on values show significant difference.

Category	Variable	Farm type (production system)*			Total	
		HIW (Wheat)	MIF (Sorghum-teff)	LIS (Sorghum-teff)		
Household	Household size	5.6 (2.4) ^b	5.1 (1.8) ^b	6.8 (1.5) ^a	5.6 (2.1)	
	Adult labour	3.5 (1.6) ^b	2.9 (1.2) ^c	4.1 (1.2) ^a	3.3 (1.4)	
	HH head age (yr)	45.2 (13.2)	43.3 (13.0)	50.5 (10.6)	45.3 (12.9)	
Farm characterization	Farm size (ha)	2.1 (1.1) ^b	1.3 (0.7) ^c	2.8 (1.1) ^a	1.9 (1.1)	
	Fertilizer input rate (kg/ha)	443.1 (417.7) ^a	157.7 (258.5) ^b	95.4 (330.2) ^c	254.4 (369.7)	
	Number of crops on farm	3.5 (1.4) ^c	4.0 (1.6) ^b	5.7 (1.4) ^a	4.1 (1.7)	
	Number of plots on farm	4.3 (2.4) ^b	3.8 (2.0) ^c	6.7 (2.2) ^a	4.5 (2.4)	
	Wheat production (kg)	4345.2 (4329.7) ^a	103.8 (351.3) ^b	71.3 (256.2) ^b	1698.4 (3372.7)	
	Teff production (kg)	373.7 (572.9) ^b	296.6 (293.8) ^b	689.4 (539.3) ^a	395.1 (482.6)	
	Sorghum production (kg)	26.2 (252.0) ^c	753.4 (683.3) ^b	1598.8 (3013.4) ^a	628.5 (1462.1)	
	Wheat area proportion (%)	44.9 (27.8) ^a	3.3 (13.1) ^b	2.5 (6.3) ^b	18.9 (28.0)	
	Teff area proportion (%)	17.0 (19.8) ^c	23.6 (24.0) ^b	26.6 (15.1) ^a	21.7 (21.4)	
	Sorghum area proportion (%)	1.1 (6.4) ^b	25.6 (24.3) ^a	22.6 (13.3) ^a	15.8 (21.0)	
	Other crops area proportion (%)	37.1 (30.1) ^b	47.5 (29.4) ^a	48.2 (18.2) ^a	43.7 (28.5)	
	TLU	4.7 (2.3) ^a	2.6 (1.6) ^b	5.2 (2.6) ^a	3.8 (2.4)	
	Livestock	Cattle ratio	0.7 (0.2) ^b	0.8 (0.3) ^a	0.7 (0.2) ^b	0.8 (0.2)
		Small ruminant ratio	0.1 (0.1) ^b	0.2 (0.2) ^b	0.3 (0.2) ^a	0.2 (0.2)
		poultry ratio	0.0 (0.1) ^b	0.0 (0.1) ^a	0.0 (0.0) ^a	0.0 (0.1)
Land features	Rented in land	0.2 (0.3) ^b	0.1 (0.2) ^b	0.2 (0.2) ^a	0.1 (0.2)	
	Area with flat land (%)	77.1 (31.1) ^a	75.5 (30.8) ^a	70.3 (26.0) ^b	75.2 (30.2)	
	Area with slope land (%)	17.7 (27.5) ^c	22.9 (29.4) ^b	25.1 (22.8) ^a	21.3 (27.7)	
	Area with steep slope land (%)	3.9 (13.7) ^a	0.8 (5.1) ^b	2.8 (8.5) ^a	2.3 (9.8)	

*Most representative production system

Table 4. Farm type description of farm typology 2-HIW: F1, F2, F3. The values are means with standard deviation in the brackets.

Letters in superscripts on values show significant difference.

Category	Variable	Farm type (district)*			Total
		HIW-F1 (Goba)	HIW-F2 (Lemo)	HIW-F3 (Siya-debirnawayu)	
Household	Household size	5.8 (2.6) ^{ab}	6.1 (3.0) ^a	5.2 (2.0)	5.6 (2.4)
	Adult labour	3.2 (1.5) ^b	4.4 (2.0) ^a	3.4 (1.4)	3.5 (1.6)
	HH head age (yr)	42.6 (12.0) ^b	50.4 (14.9) ^a	44.9 (12.7) ^{ab}	45.2 (13.2)
Farm characterization	Farm size (ha)	2.5 (1.3) ^a	1.5 (0.8) ^c	2.0 (0.9) ^b	2.1 (1.1)
	Fertilizer input rate (kg/ha)	241.3 (294.9) ^b	268.8 (207.7) ^b	683.3 (448.6) ^a	443.1 (417.7)
	Number of crops on farm	3.0 (1.3) ^b	4.7 (1.7) ^a	3.3 (1.1) ^b	3.5 (1.4)
	Number of plots on farm	3.5 (1.7) ^b	2.9 (1.6) ^c	5.6 (2.5) ^a	4.3 (2.4)
	Wheat production (kg)	5347.2 (4415.6) ^a	1418.1 (2561.5) ^b	4885.0 (4349.4) ^a	4345.2 (4329.7)
	Teff production (kg)	0 (0) ^c	369.0 (413.6) ^b	673.0 (680.3) ^a	373.7 (572.9)
	Sorghum production (kg)	0 (0) ^b	119.0 (547.9) ^a	4.7 (53.2) ^b	26.2 (252.0)
	Wheat area proportion (%)	55.6 (28.3) ^a	16.8 (22.5) ^b	49.2 (20.3) ^a	44.9 (27.8)
	Teff area proportion (%)	0.8 (5.7) ^c	24.9 (25.7) ^b	26.3 (15.4) ^a	17.0 (19.8)
	Sorghum area proportion (%)	0 (0) ^b	4.0 (12.1) ^a	0.5 (4.5) ^b	1.1 (6.4)
	Other crops area proportion (%)	43.7 (28.1) ^a	54.3 (34.9) ^a	24.0 (23.1) ^b	37.1 (30.1)
	TLU	5.1 (2.7) ^a	3.2 (1.7) ^b	5.0 (2.0) ^a	4.7 (2.3)
	Livestock	Cattle ratio	0.7 (0.2) ^b	0.9 (0.1) ^a	0.7 (0.2) ^b
Small ruminant ratio		0.2 (0.2) ^a	0.0 (0.1) ^b	0.2 (0.1) ^a	0.1 (0.1)
poultry ratio		0.0 (0.1)	0.0 (0.0) ^a	0.0 (0.0) ^a	0.0 (0.1)
Rented in land proportion		0.2 (0.3) ^{ab}	0.1 (0.2) ^b	0.2 (0.3) ^a	0.2 (0.3)
Land features	Area with flat land (%)	82.6 (26.1) ^a	86.7 (26.6) ^a	68.5 (34.3) ^b	77.1 (31.1)
	Area with slope land (%)	11.5 (23.0) ^b	11.4 (23.0) ^b	25.4 (30.6) ^a	17.7 (27.4)
	Area with steep slope land (%)	4.2 (13.1) ^a	1.8 (8.6) ^a	4.6 (15.8) ^a	3.9 (13.7)
Production use	Wheat consumed (%)	37	55	39	41
	Teff consumed (%)		71	61	63
	Wheat sold (%)	47	37	46	45
	Teff sold (%)		16	21	
Categorical variables	HH head high education (%)	38	29	15	27
	Pesticides use (%)	70	74	86	77
	Hybrid seeds use (%)	60	24	56	47
	Fertilizer use on teff (%)	1	90	95	62
	NPS fertilizer (%)	97	10	1	36
	NPSB fertilizer (%)	1	90	98	63
	Offfarm income source (%)	27	40	24	30
	Homegarden (%)	40	74	52	55
	Membership (%)	75	48	85	70
	Food shortage (%)	33	48	13	31
	Agricultural training (%)	58	67	78	68

*Most representative district

Table 5: Farm type description of farm typology 2 –MIF:F1, F2, F3. The values are means with standard deviation in the brackets.

Letters in superscripts on values show significant difference.

Category	Variable	Farm type (district)*			Total
		MIF-F1 (West-Beles- sa)	MIF-F2 (Kewet)	MIF-F3 (Lemo)	
Household	Household size	4.9 (1.7) ^b	4.6 (1.6) ^b	6.6 (1.7) ^a	5.1 (1.8)
	Adult labour	2.7 (1.0) ^b	2.8 (1.3) ^b	3.6 (1.2) ^a	2.9 (1.2)
	HH head age (yr)	41.7 (12.5) ^b	43.2 (14.1) ^b	49.7 (11.6) ^a	43.3 (13.0)
	Farm size (ha)	1.6 (0.7) ^a	0.9 (0.4) ^b	0.9 (0.4)	1.3 (0.7)
Farm characterization	Fertilizer input rate (kg/ha)	98.1 (108.5) ^b	305.3 (476.8) ^a	180.1 (126.3) ^a	157.7 (258.5)
	Number of crops on farm	4.3 (1.5) ^b	2.8 (1.2) ^c	4.9 (1.7) ^a	4.0 (1.6)
	Number of plots on farm	4.7 (2.0) ^a	2.6 (1.2)	2.2 (0.9) ^b	3.8 (2.0)
	Wheat production (kg)	64.5 (338.7) ^b	128.0 (371.8)	217.0 (347.4) ^a	103.8 (351.3)
	Teff production (kg)	355.3 (285.0) ^a	136.0 (254.4)	294.6 (299.3) ^a	296.6 (293.8)
	Sorghum production (kg)	903.5 (662.0) ^a	901.7 (625.3) ^a	0.0 (0.0) ^b	753.4 (683.3)
	Wheat area proportion (%)	2.1 (9.6) ^a	5.7 (20.1) ^a	4.5 (12.0) ^a	3.3 (13.1)
	Teff area proportion (%)	24.0 (18.9) ^a	15.9 (25.8) ^b	32.7 (33.4) ^a	23.6 (24.0)
	Sorghum area proportion (%)	28.3 (20.8) ^a	35.7 (29.7) ^a	1.6 (7.3) ^b	25.6 (24.3)
	Other crops area proportion (%)	45.5 (23.5) ^b	42.7 (34.4) ^b	61.2 (37.6) ^a	47.5 (29.4)
Livestock	TLU	2.8 (1.6) ^a	2.5 (1.8) ^{ab}	2.0 (0.9) ^b	2.6 (1.6)
	Cattle ratio	0.7 (0.3) ^b	0.9 (0.2) ^a	0.9 (0.2) ^a	0.8 (0.3)
	Small ruminant ratio	0.2 (0.3) ^a	0.1 (0.2) ^b	0.1 (0.2) ^b	0.2 (0.2)
	poultry ratio	0.0 (0.1) ^a	0.0 (0.1) ^b	0.0 (0.1) ^{ab}	0.0 (0.1)
Land features	Rented in land proportion	0.2 (0.2) ^a	0.0 (0.1) ^b	0.0 (0.0) ^b	0.1 (0.2)
	Area with flat land (%)	71.5 (29.6) ^b	79.1 (32.2) ^a	85.3 (30.7) ^a	75.5 (30.8)
	Area with slope land (%)	26.3 (27.8) ^a	20.3 (32.3) ^b	13.4 (29.4) ^b	22.8 (29.4)
	Area with steep slope land (%)	1.1 (6.1) ^a	0.0 (0.0) ^a	0.6 (4.5) ^a	0.8 (5.1)
Production use	Sorghum consumed (%)	70	78		73
	Sorghum sold (%)	14	12		13
	Teff consumed (%)	61	65	70	65
	Teff sold (%)	21	20	18	20
Categorical variables	HH head high education (%)	19	11	11	14
	Manure use (%)	45	7	11	21
	Pesticides use (%)	80	68	73	74
	Hybrid seeds use (%)	38	20	29	29
	Fertilizer use on sorghum (%)	75	84	9	56
	Urea use (%)	77	95	98	90
	NPSB fertilizer (%)	86	80	88	84
	Offfarm income source (%)	33	19	30	27
	Oxen pair (%)	73	97	43	71
	Homegarden (%)	35	29	82	49
	Membership (%)	76	49	32	52
	Agricultural training (%)	60	72	61	64
	Food shortage (%)	42	43	64	50

*Most representative district

Table 6: Farm type description of farm typology 2 – LIS: F1, F2, F3. The values are means with standard deviation in the brackets.

Letters in superscripts on values show significant difference.

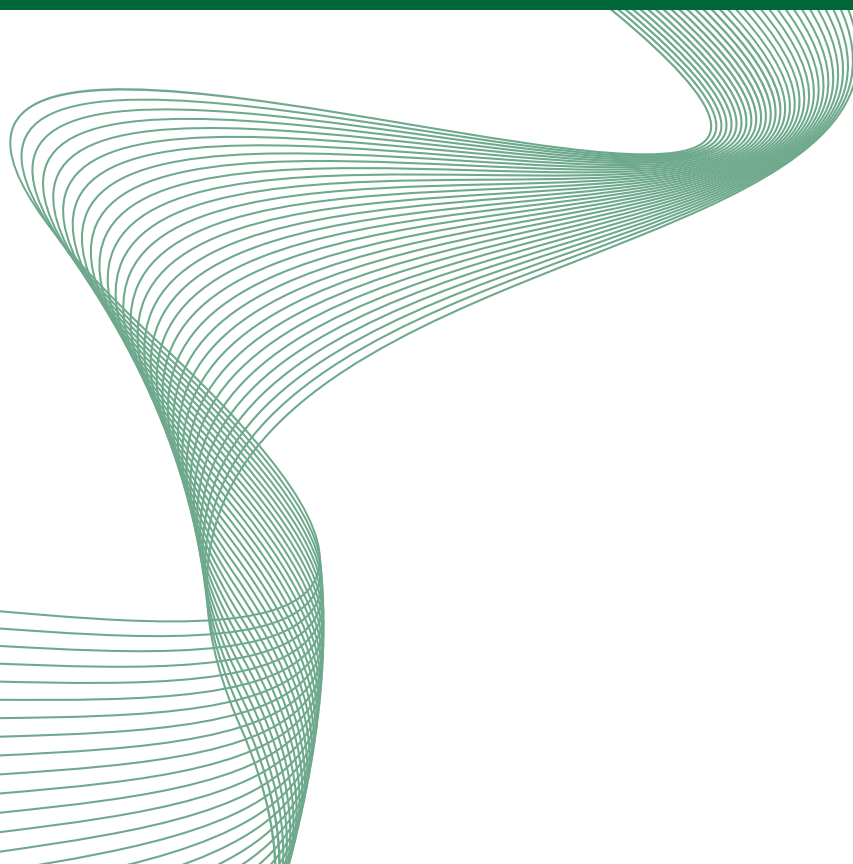
Category	Variable	Farm type (district)*			Total	
		LIS-F1	LIS-F2	LIS-F3		
		(West-Belessa)	(West-Belessa)	(West-Belessa)		
Household	Household size	6.8 (1.5) ^a	7.3 (1.4) ^a	6.6 (1.4) ^a	6.8 (1.5)	
	Adult labour	4.2 (1.2) ^a	4.1 (1.2) ^a	3.8 (1.1) ^a	4.1 (1.2)	
	HH head age (yr)	50.9 (10.6) ^a	48.9 (9.2) ^a	50.8 (12.1) ^a	50.5 (10.6)	
	Farm size (ha)	2.8 (1.1) ^{ab}	3.1 (1.0) ^a	2.5 (1.4) ^b	2.8 (1.1)	
	Fertilizer input rate (kg/ha)	114.9 (402.3) ^a	101.5 (178.3) ^a	22.9 (23.5) ^b	95.4 (330.2)	
Farm characterization	Number of crops on farm	5.8 (1.4) ^a	5.3 (1.3) ^a	5.4 (1.6) ^a	5.7 (1.4)	
	Number of plots on farm	7.0 (1.9) ^a	7.2 (2.4) ^a	5.3 (2.2) ^b	6.7 (2.2)	
	Wheat production (kg)	47.6 (144.9) ^a	8.3 (40.8)	212.0 (511.8) ^a	71.3 (256.2)	
	Teff production (kg)	639.2 (532.6) ^b	925.0 (454.4) ^a	634.0 (594.4) ^b	689.4 (539.3)	
	Sorghum production (kg)	1754.7 (3721.0) ^b	1924.2 (781.3) ^a	756.6 (585.9) ^c	1598.8 (3013.4)	
	Wheat area proportion (%)	2.3 (5.4) ^a	3.2 (6.7) ^a	2.5 (8.5) ^a	2.5 (6.3)	
	Teff area proportion (%)	24.8 (12.8) ^a	28.6 (11.5) ^a	30.9 (23.1) ^a	26.6 (15.1)	
	Sorghum area proportion (%)	23.2 (12.5) ^a	24.0 (12.9) ^a	19.3 (15.9) ^a	22.6 (13.3)	
	Other crops area proportion (%)	49.6 (17.8) ^a	44.2 (12.6) ^a	47.3 (23.5) ^a	48.2 (18.2)	
	TLU	5.3 (2.4) ^a	6.4 (3.2) ^a	3.8 (2.1) ^b	5.2 (2.6)	
	Livestock	Cattle ratio	0.7 (0.2) ^b	0.6 (0.2) ^b	0.8 (0.2) ^a	0.7 (0.2)
		Small ruminant ratio	0.3 (0.2) ^a	0.3 (0.2) ^a	0.2 (0.2) ^b	0.3 (0.2)
poultry ratio		0.0 (0.0) ^a	0.0 (0.0) ^a	0.0 (0.0) ^a	0.0 (0.0)	
Rented in land proportion		0.1 (0.1) ^c	0.4 (0.2) ^a	0.2 (0.3) ^b	0.2 (0.2)	
Land features	Plots with flat land (%)	68.8 (25.0) ^a	80.1 (21.9) ^a	65.8 (31.5) ^a	70.3 (26.0)	
	Plots with slope land (%)	27.3 (22.3) ^a	19.2 (20.9) ^a	23.4 (25.6) ^a	25.1 (22.7)	
	Plots with steep slope land (%)	3.2 (9.2) ^a	0.5 (2.6) ^a	3.7 (7.2) ^a	2.8 (8.4)	
Production use	Sorghum consumed (%)	65	69	68	67	
	Sorghum sold (%)	13	16	15	15	
	Teff consumed (%)	60	66	63	63	
Categorical variables	Teff sold (%)	16	12	27	19	
	Manure use (%)	79	25	48	51	
	Compost use (%)	32	46	20	33	
	Hybrid seeds use (%)	60	71	48	60	
	Fertilizer use on sorghum (%)	88	79	20	62	
	Urea use (%)	85	92	64	80	
	Offfarm income source (%)	28	17	32	26	
	Homegarden (%)	39	8	56	34	
	Membership (%)	93	42	56	64	
	Agricultural training (%)	65	58	64	62	
	Food shortage (%)	47	17	56	40	

*Most representative district



Supplementary Material

Farm Typology for Ethiopia Use Cases
(analysis at the landscape, crop and use
case level)



The following table 1 illustrates the list of excluded variables in the analysis. The reasons why they were excluded are indicated in the last column.

Table 1. List of variables not included in the typology, nor in the farm description

No used variables due to no variation or problems in the dataset					Reasons	
Household	Gender	person_gender		binary	categorical	No variation
Labour	Work on farm activities		Total labour use in the farm per year	%	categorical	No units were provided
	Hired labour		Sum of plots with hired labour > area proportion with hir	%	categorical	No units were provided
	Homegarden labour participation		Ratio of hh members that contribute in the homegarden	%	categorical	No units were provided
Land use	Plot tenure status				binary	many NA
	Intercropping ratio		Percentage of the total operated area doing intercroppin	%	binary	No relevan for the hypothesis
Livestock	pigs		Percentage of the total herd size	%	continuous	No relevan for the hypothesis
	poultry	poultryratio	ratio from TLU	%	continuous	No significant correlation
	other animals		Percentage of the total herd size	%	continuous	No relevan for the hypothesis
Income	Credit				categorical	many NA
	On farm income	on_farm_income	Total income per year	ETB?	continuous	Only referes to one main crop
	Off farm income proportion	off_farm_income_propor	Proportions of off-farm income	5 levels	categorical	
Main crops: production and allocation	Total production		Sum of main crops production	kg	continuous	Includes Maize
	Crop sold		Proportion of the total crop production	RHoMIS	categorical	Many NAs
	Crop feed for livestock		Proportion of the total crop production	RHoMIS	categorical	Many NAs
	Crop saved for seed		Proportion of the total crop production	RHoMIS	categorical	Many NAs
Agricultural inputs and practices	Agricultural inputs		Proportions of the use of agricultural inputs	%	categorical	No variation
	Use of MOP		Ratio from the total used fertilizers	%	categorical	No variation
	Use of Ammonium suphate		Ratio from the total used fertilizers	%	categorical	No variation
	Use of NPSZn	NPSZn	Ratio from the total used fertilizers	%	categorical	No variation
	Use of DAP	DAP	Ratio from the total used fertilizers	%	categorical	No variation
	Use of TSP	TSP	Ratio from the total used fertilizers	%	categorical	No variation
	Use of Potassium sulphate		Ratio from the total used fertilizers	%	categorical	No variation
	Use of SOP		Ratio from the total used fertilizers	%	categorical	No variation
	Irrigation		proportion of households using irrigation	%	categorical	No relevan for the hypothesis
Water conservation practices		Proportions of most relevant water conservation practica	%	binary	No relevan for the hypothesis	
Access, use and constraints of technologies	Irrigation use constraints		Proportions of most relevant irrigation constraints	%	categorical	No relevan for the hypothesis
	Environmental constraints		Proportions of most relevant constraints	%	categorical	No relevan for the hypothesis
	Fertilizer use constraints		Proportions of most relevant fertilizer use constraints	%	categorical	No relevan for the hypothesis
	Organic fertilizer use constraints		Proportions of most relevant org. fertilizer use constrain	%	categorical	No relevan for the hypothesis
	Land preparation tools		Proportions of most relevant used tools	%	categorical	Mechanization?
	Machinery access		Proportions of types of access to machinery	%	categorical	Ownership is NA
	Machinery use constraints		Proportions of most relevant used machinery constraint	%	categorical	No relevan for the hypothesis
	Distance to market	market_distance	Distance to market using the most common way of trans	minutes	continuous	Unit not valid
Digital access-laptop, tablets	laptop.tablets	Score on the levels of access to laptop, tablet	laptop, t	categorical	No variation	

Output of dendrograms (figures 1-4) from Typology 1 and Typology 2.

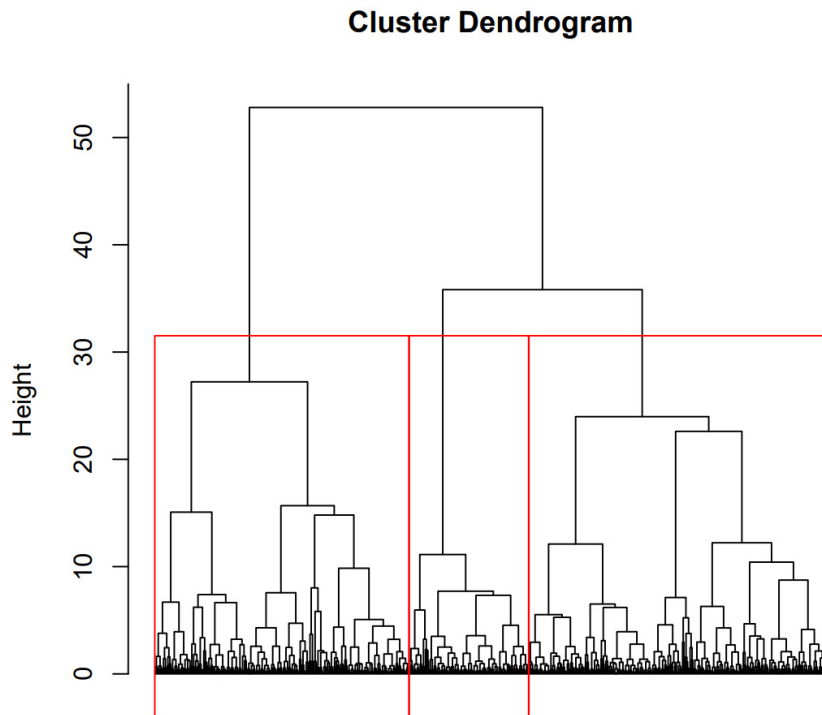


Figure 1. Dendrogram from Typology 1. Segregation of farming systems in the region (HW, MS and LS)

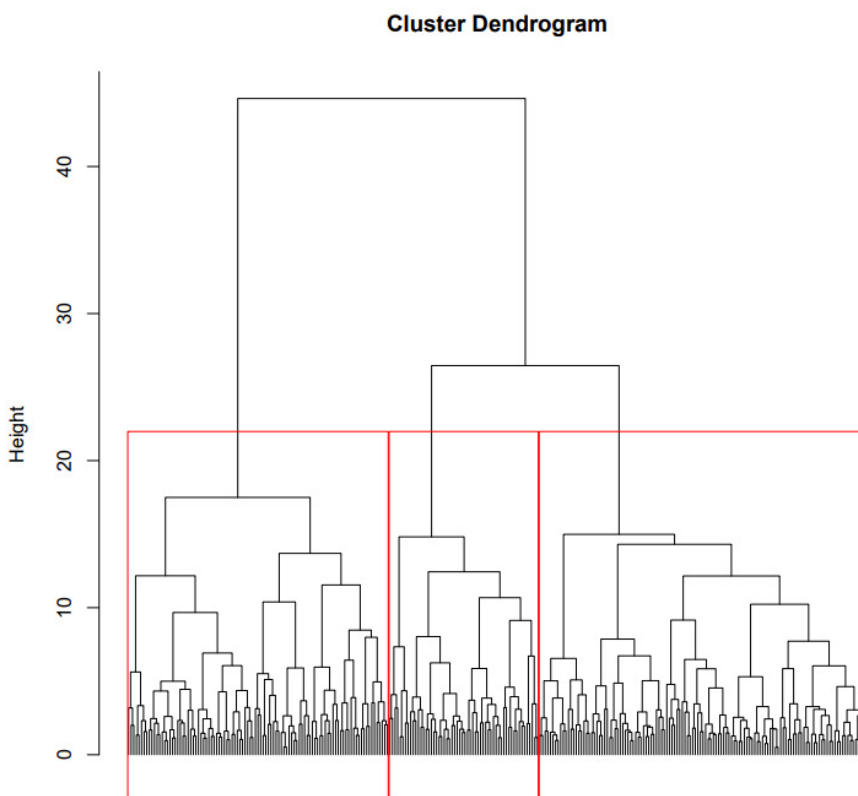


Figure 2. Dendrogram from Typology 2. High intensified wheat in highlands.

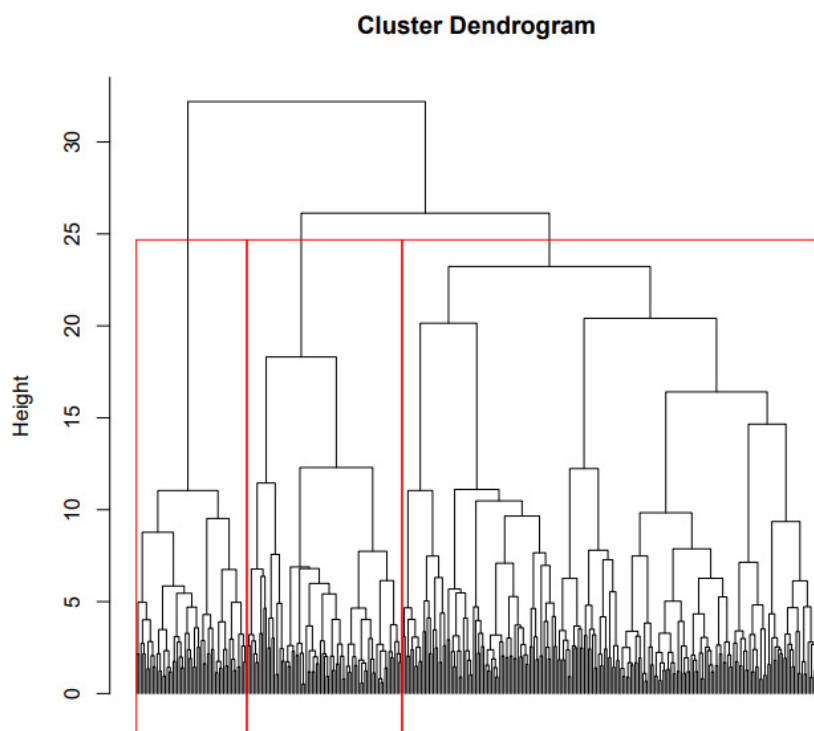


Figure 3. Dendrogram from Typology 2. Medium intensified sorghum in mixed attitudes.

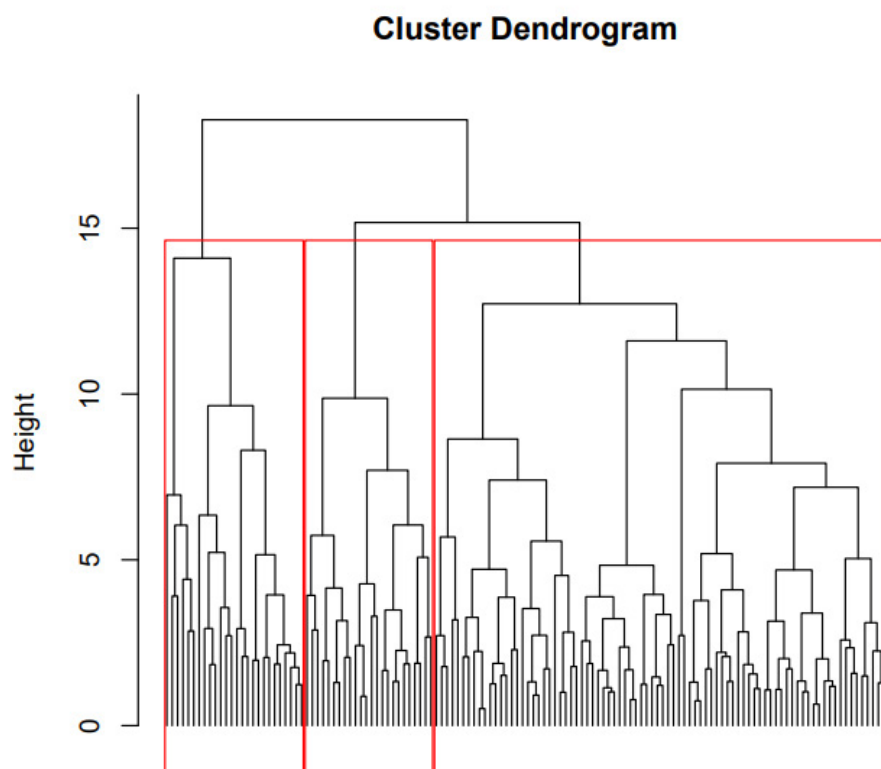


Figure 4. Dendrogram from Typology 2. Low intensified sorghum in lowlands.

Contribution of dimensions (figures 5 – 7) from Typology 2.

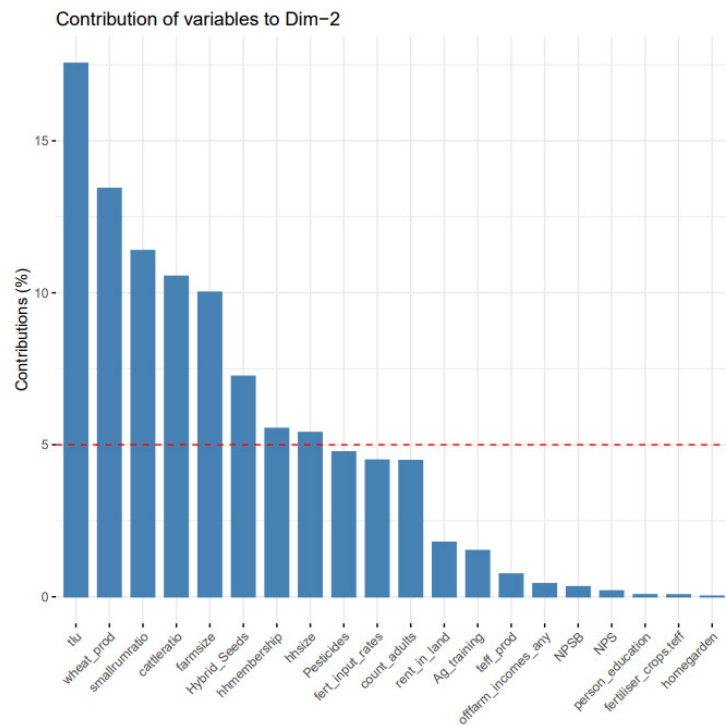
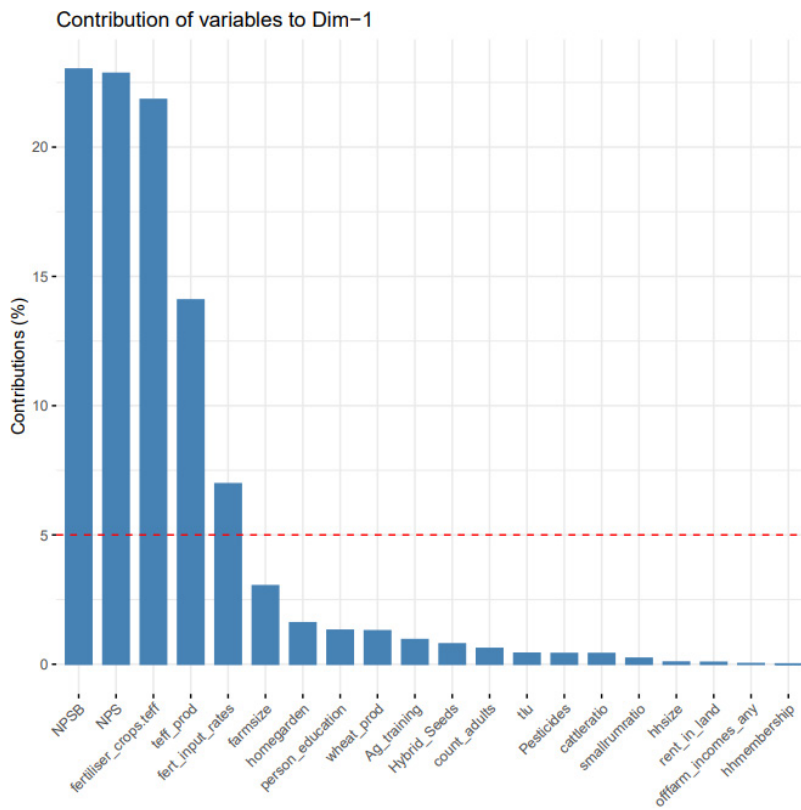


Figure 5. Dimension contribution from High intensified wheat in highlands (Dim 1 and 2).

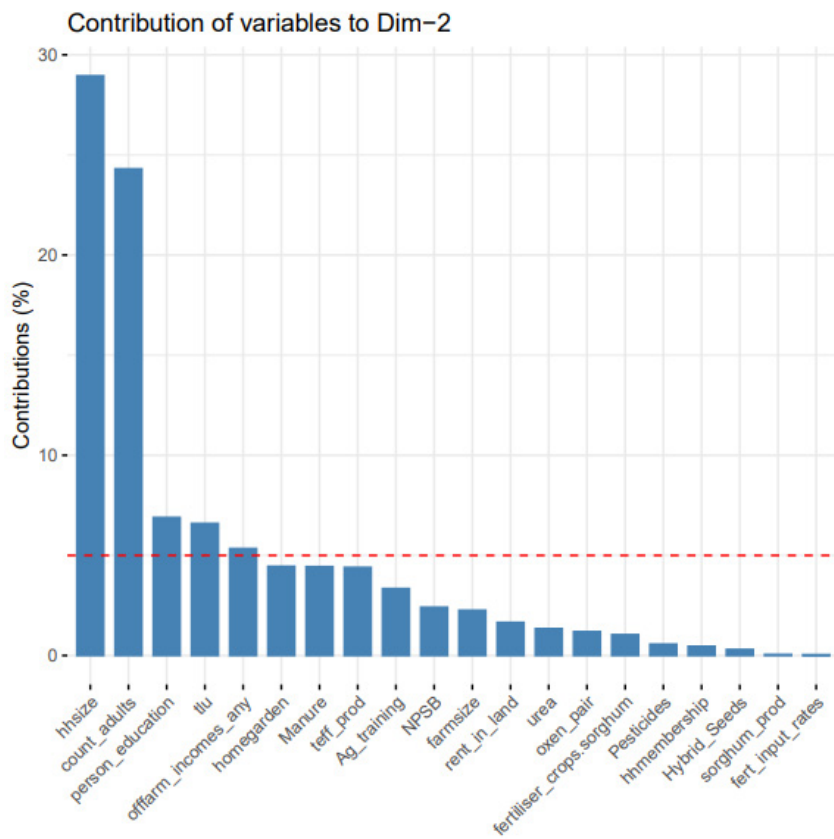
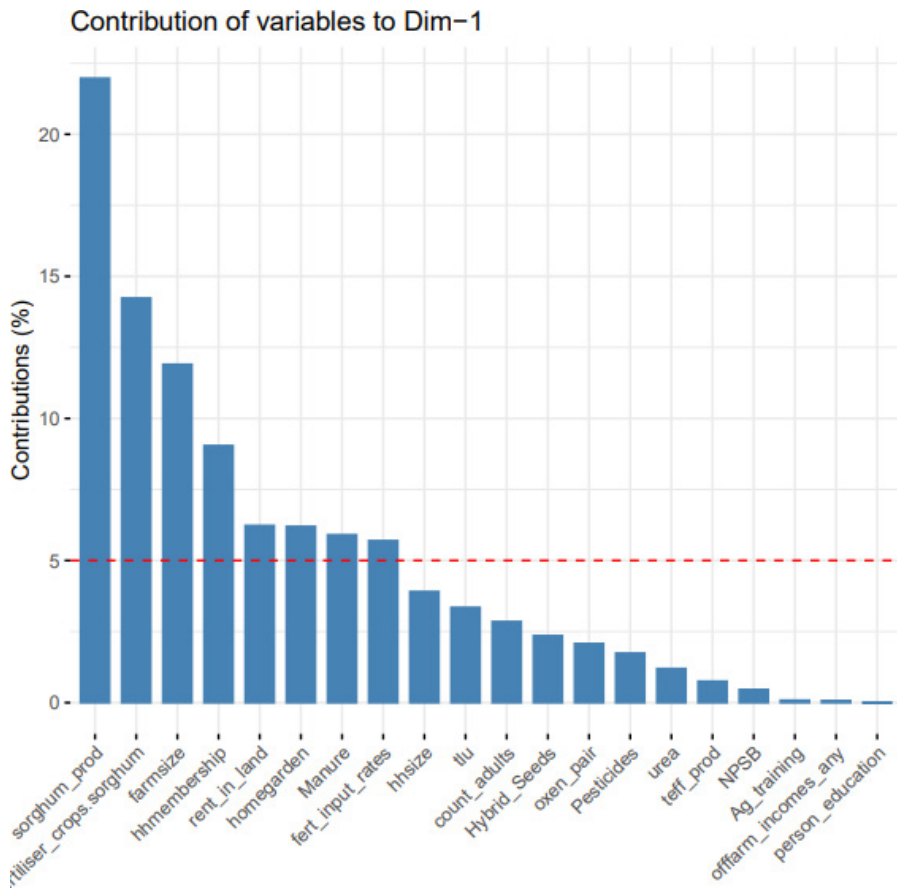


Figure 6. Dimension contribution from Medium intensified sorghum in mixed attitudes (Dim 1 and 2).

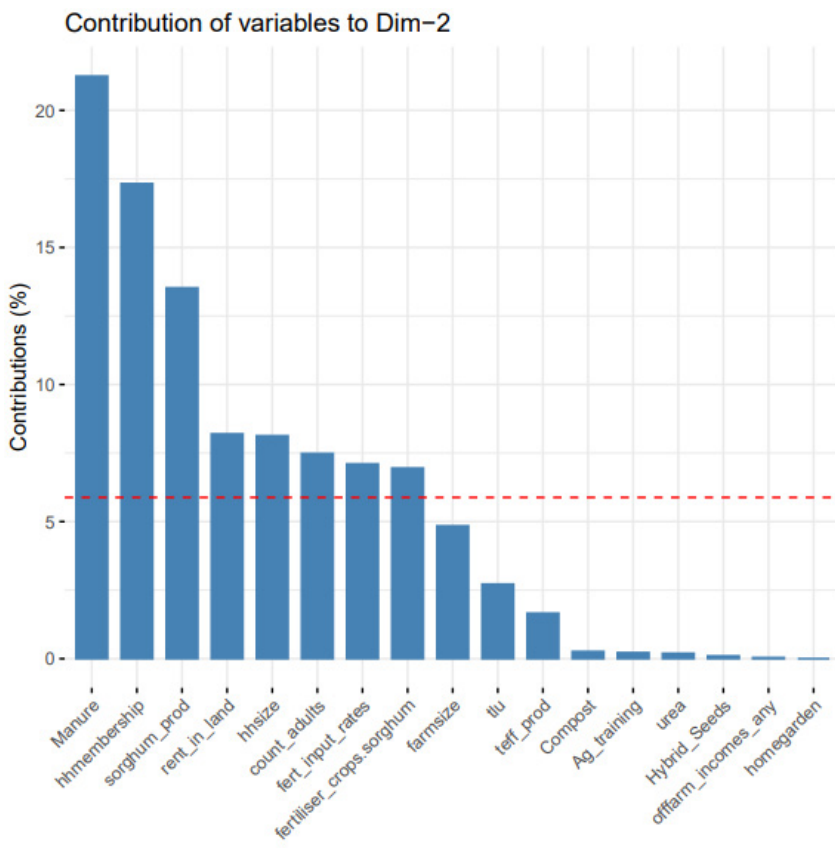
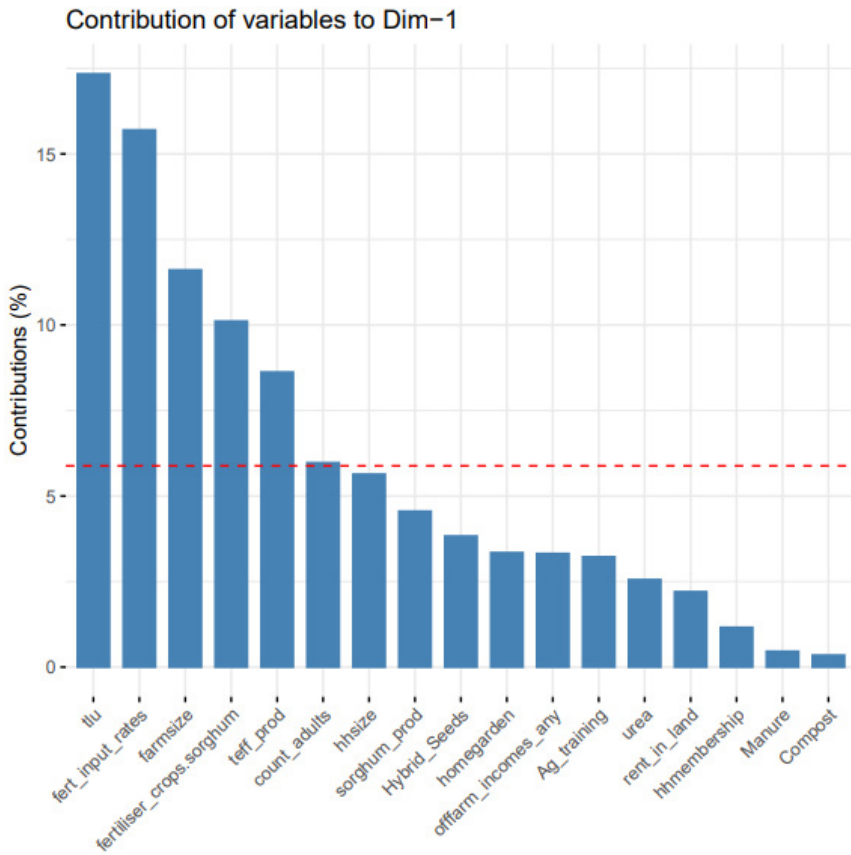


Figure 7. Dimension contribution from Low intensified sorghum in lowlands (Dim 1 and 2).

Result from categorical variables (figures 8 – 10) from Typology 2.

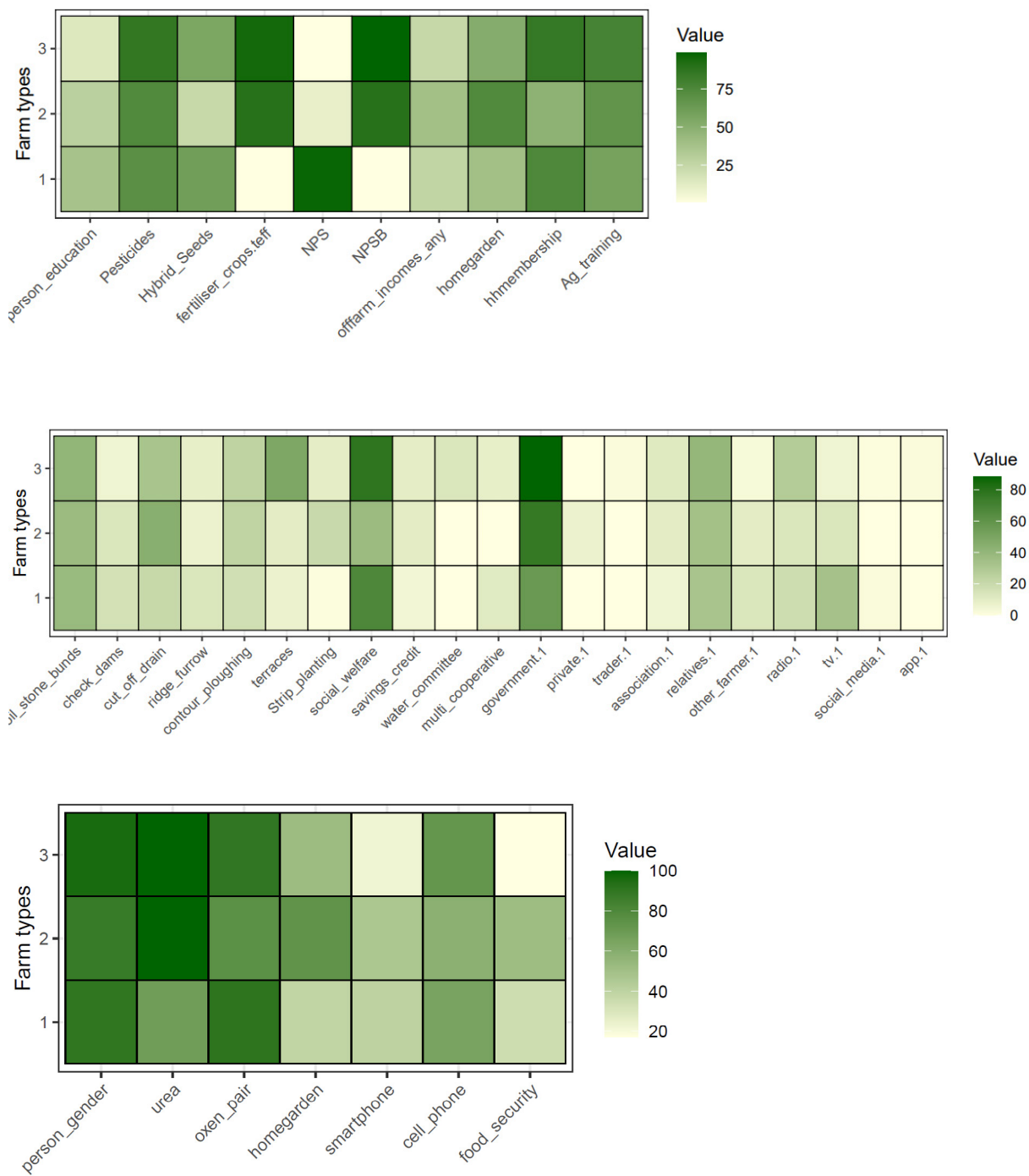


Figure 8. HW typology outputs from categorical variables: FAMD (upper), water conservation practices, memberships and dissemination (middle) and additional variables (lower position)

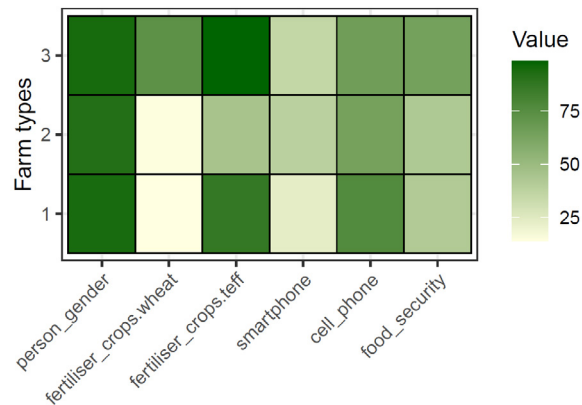
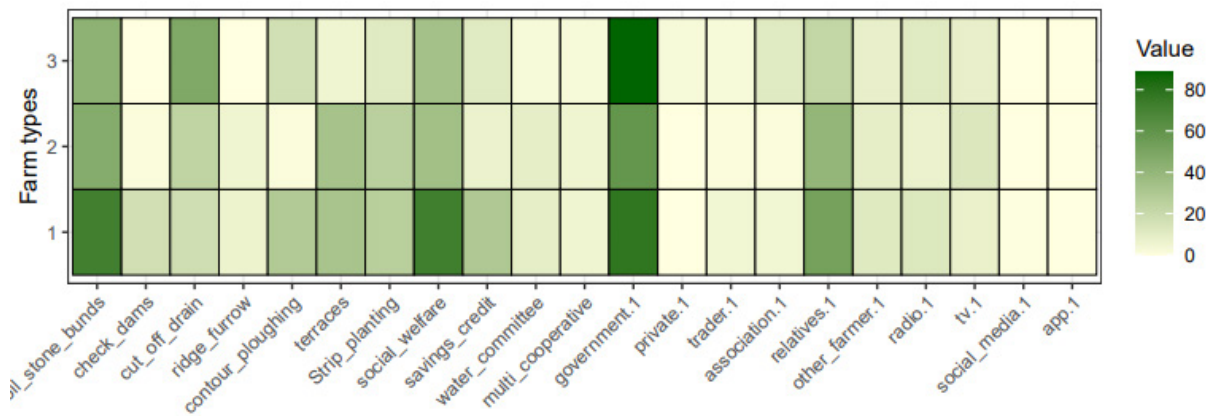
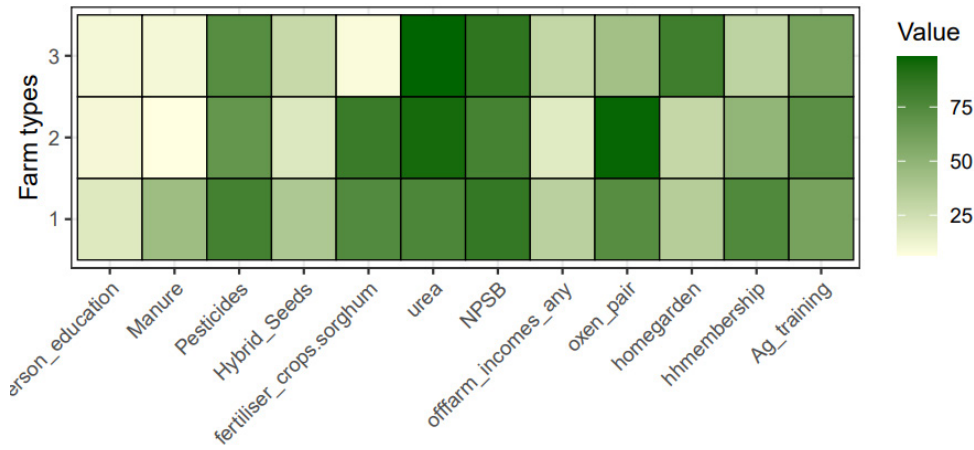


Figure 9. MS typology outputs from categorical variables: FAMD (upper), water conservation practices, memberships and dissemination (middle) and additional variables (lower position)

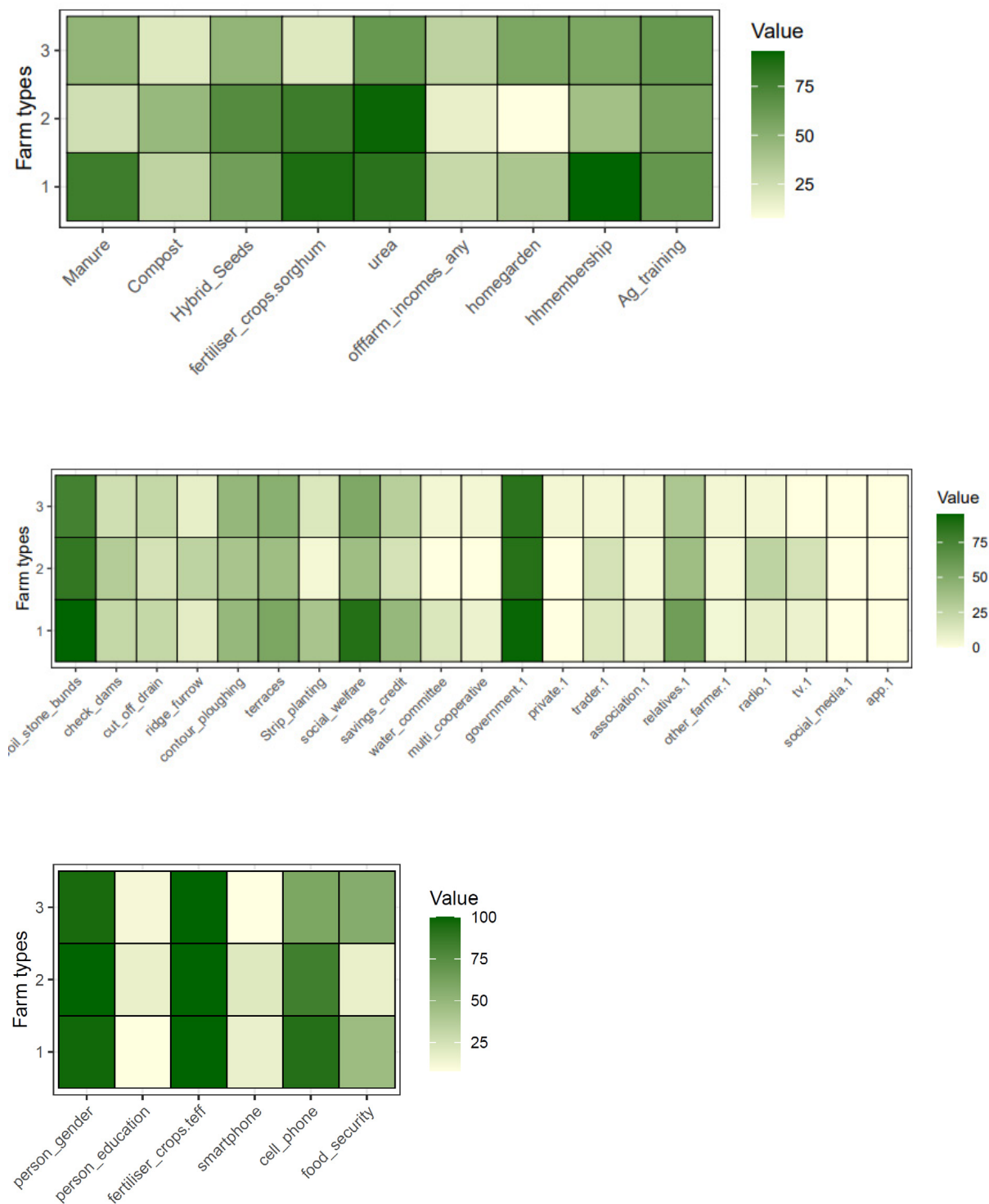


Figure 10. LS typology outputs from categorical variables: FAMD (upper), water conservation practices, memberships and dissemination (middle) and additional variables (lower position)

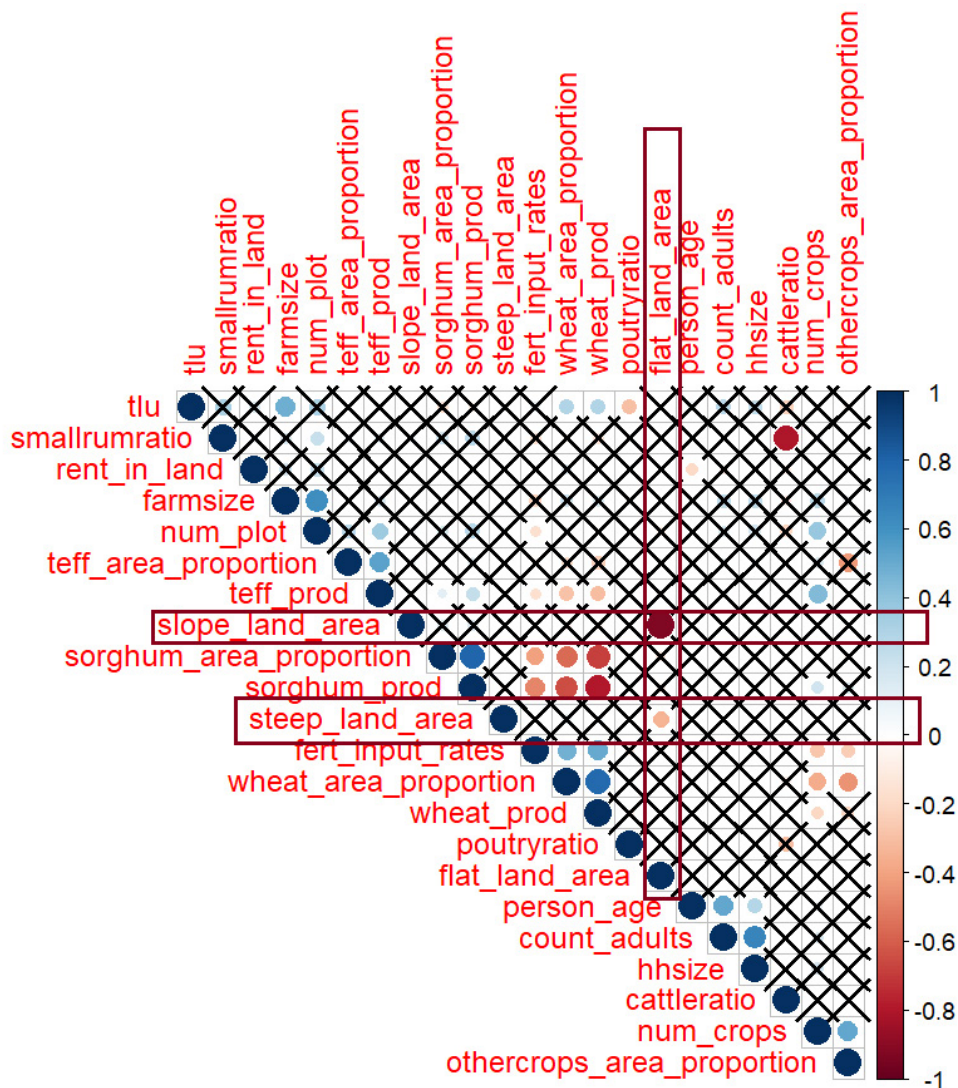


Figure 11: Correlation matrix of continuous variable. The cross marks denote non-significant correlation between the two variables. Here, inset, slope variables either highly correlating among each other or not correlating significantly with any other variation. Hence, slope variables were not included in the PCA and FAMD.