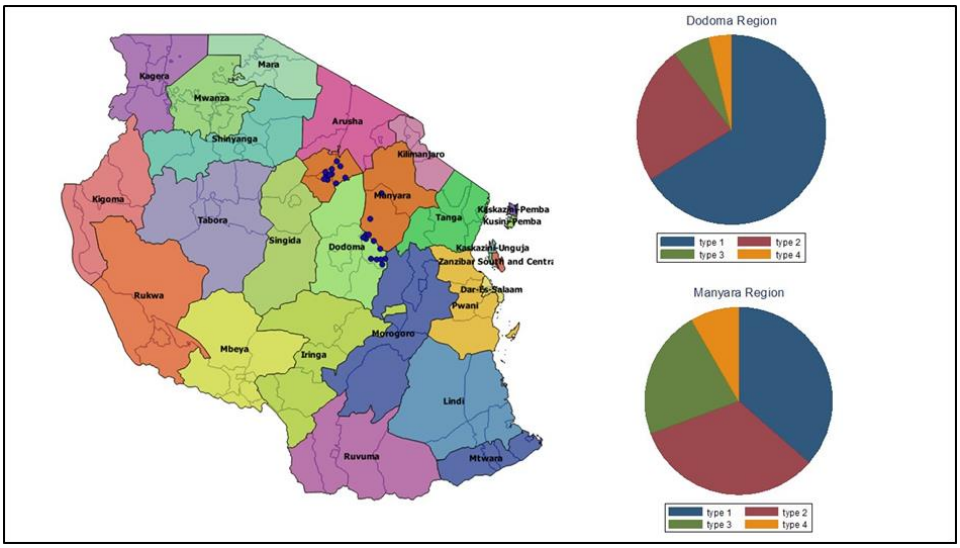
Typology characterization of farmers in Africa RISING sites in Tanzania

Sara Signorelli, Carlo Azzarri and Beliyou Haile



Produced by International Food Policy Research Institute

Published by International Institute of Tropical Agriculture

April 2016

www.africa-rising.net

The Africa Research In Sustainable Intensification for the Next Generation (Africa RISING) program comprises three research-for-development projects supported by the United States Agency for International Development as part of the U.S. government's Feed the Future initiative.

Through action research and development partnerships, Africa RISING will create opportunities for smallholder farm households to move out of hunger and poverty through sustainably intensified farming systems that improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base.

The three regional projects are led by the International Institute of Tropical Agriculture (in West Africa and East and Southern Africa) and the International Livestock Research Institute (in the Ethiopian Highlands). The International Food Policy Research Institute leads the program's monitoring, evaluation and impact assessment. http://africa-rising.net/

This document was made possible with support from the American people delivered through the United States Agency for International Development (USAID) as part of the US Government's Feed the Future Initiative. The contents are the responsibility of the producing organization and do not necessarily reflect the opinion of USAID or the U.S. Government.

Table of Contents

Introduction

Africa RISING is testing alternative technology options with heterogeneous populations of farmers that will likely respond to the technologies differently. Creating farm typologies is one approach to design targeted interventions that adequately address the needs of different types of farmers. Notably, creating typologies can help:

- **Identify suitable farms to target innovations (ex-ante)**: we assume that not all innovations are appropriate for all farms, and that structuring into groups would support the identification of technology-specific suitable farming systems.
- **Scale out innovations**: on the basis of the heterogeneity in a population we can formulate extension messages, policies and other incentive schemes to further spread the use of designed innovations.
- **Assess agro-economic effects (ex-post)** Explaining trends and farmer 'behavior' (functional characteristics, including sustainable intensification indicators) and verification of the agro-economic effects of the interventions for different farm types.

This document presents a summary of a typology study done using quantitative statistical methods (discussed below) applied to micro data from the Tanzania Africa RISING Baseline Evaluation Survey (TARBES) (conducted in 2014) and secondary data on environmental/biophysical variables from various source. The quantitative approaches have the advantage that they are reproducible and do not impose any ex-ante structure to the clustering process, while more qualitative approaches can potentially incorporate less tangible insights such as cultural patterns. Once the different farm types are identified through systematic quantitative analysis, they need to be validated with input from Africa RISING colleagues (especially working in Tanzania).

Methodological steps

We apply a combination of factor and cluster analysis to obtain the final groups, or "types" (See Cunningham & Maloney, 1999 for an empirical application). We first use factor analysis to reduce the number of socio-economic variables to characterize the farms by selecting the most relevant ones in differentiating the sample. Factor analysis is often used to discover underlying patterns in data and its aim is to explain the largest portion of the entire dataset variation with the lowest possible number of factors. Factors are unobserved variables that summarize the correlation among several observed variables and factor analysis allows us to divide the dataset into different factors, or dimensions, and categorize each variable into one of the factors. Figure 1 shows an example of how the variables in a dataset are divided into different dimensions to explain the total variation in the data. The analysis also allows us to rank the factors by their importance in explaining the variation in the data and to further rank each variable by its explanatory power within the factor.

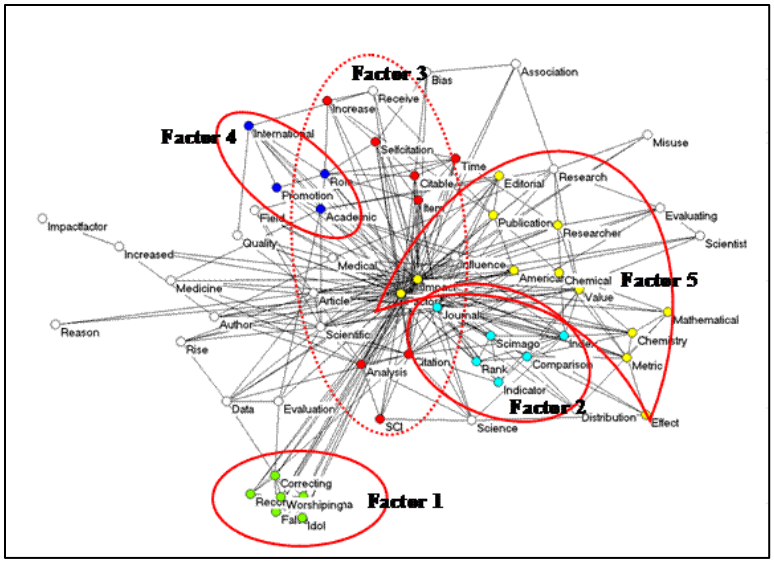


Figure 1: Example of factor analysis[1]

Our factor analysis based on TARBES data involves the following main steps (see for example McDonald; 2014. Basilevsky; 2009. Mulaik; 2009 for a discussion on the methods):

1. We divide the variables in TARBES into the five domains of sustainability that have been identified within Africa RISING to gauge progress: **productivity, economic, environment, social and human**.
2. We perform separate factor analysis on each domain to select the variables that explain the largest portion of the variation in the data.
3. We use scree plots to define the number of factors to look at and, within each of the selected factors, we consider the two variables with the highest absolute values of factor loads, conditional on them being greater than 0.5 (or smaller than -0.5).
4. Finally, we obtain a parsimonious set of socio-economic variables that explain most of the variation in the data and thus are highly relevant in defining the different farm types.

The sub-set of variables obtained using steps (1) to (4) are used to perform a cluster analysis, which divides the total sample into a chosen number of clusters (Kaufman & Rousseeuw; 2009. Romesburg; 2004. Galbraith et Al.; 2002). The numbers of clusters are chosen in order to represent groups that are different enough from each other while ensuring that each group to be included has a sufficient amount of observations. There are several different methods to perform cluster analysis, some hierarchical and some non-hierarchical. We chose the hierarchical method using medians, where the distance between two clusters is calculated as the median distance between all pairs of subjects in the two clusters. The results obtained and the characteristics of each group formed are reported in the next section.

[1] http://www.leydesdorff.net/words/

Results

Factor analysis of productivity variables (sustainability domain 1)

The scree plot of the factorization of the productivity variables (Figure 2) shows that the first three factors (represented by the first three dots at the top of the line graph) are highly relevant but that the 4th factor starts to be less important in explaining the variation (smaller vertical jump).

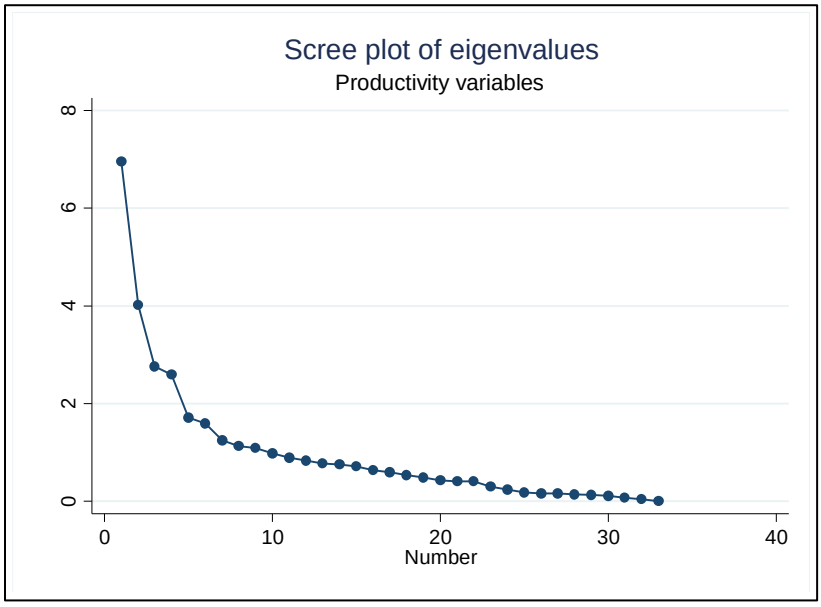


Figure 2: Scree plot of productivity variables

Table 1 shows the rotated matrix of factor loads for the three factors we have chosen, with the relevant variables highlighted (>0.5 or <-0.5). Factor 1 captures elements related to the crop diversification and intercropping practices. Factor 2 captures elements of legumes cultivation as well as the land size dedicated to intercropping. Factor 3 captures the total number of parcels and plots owned by the household as well as information on cereal cultivation. The final selection of variables for the cluster analysis include share of households doing intercropping and share of households doing intercropping with legumes for factor 1, total area intercropped and area intercropped with legumes for factor 2, and number of parcels and plots cultivated by the household for factor 3.

Table 1: Factor loads for productivity variables

Variable	Factor1	Factor2	Factor3
Land size (Ha)	-0.2532	0.3716	0.2318
N. parcels	-0.0158	0.1534	0.878
Min distance plot	-0.1461	0.0167	-0.2057
Max distance plot	-0.0079	-0.0268	0.4196
N. trees	0.071	-0.0081	0.1221
N. crops	0.7803	0.2079	-0.1504
N. plots	0.0567	0.1362	0.9091
HH does intercropping	0.796	0.1333	-0.0244
HH does intercropping with legumes	0.808	0.1524	-0.0534
N. of intercropped plots	0.5875	0.2909	0.429
Size intercropped land (Ha)	0.1304	0.891	0.1383
Size legumes-intercropped land (Ha)	0.1738	0.9074	0.1171
Ownership mixed livestock	0.1974	-0.076	0.1094
N. livestock types owned	0.1775	-0.0084	0.1083
Maize only crop	-0.2889	-0.0346	-0.1067
Mixed crops	0.2889	0.0346	0.1067
Cultivation of cereals	0.5929	-0.0771	0.2808
Cultivation of vegetables	0.0121	-0.0251	0.0407
Cultivation of legumes	0.6971	0.1344	0.044
Area cultivated with cereals (Ha)	-0.2748	0.4333	0.5274
Area cultivated with vegetables (Ha)	0.0367	0.0301	-0.0138
Area cultivated with legumes (Ha)	0.3197	0.5651	0.1995
Production cereals (Kg)	-0.0164	0.3687	0.5807
Production vegetables (Kg)	-0.0235	-0.0015	-0.036
Production legumes (Kg)	0.1568	0.6399	0.2545
Yield cereals (Kg/Ha)	0.4531	-0.054	0.1257
Yield vegetables (Kg/Ha)	-0.0398	-0.0231	0.0353
Yield legumes (Kg/Ha)	0.4144	-0.0499	0.0697
TLU small ruminants	-0.144	0.1424	0.0159
TLU big ruminants	-0.0756	0.1436	0.1568
TLU poultry	0.1087	0.0666	0.1369
Fertilizer used (Kg)	0.1771	0.2456	0.2016
HH does irrigation	-0.0669	-0.035	0.0746

Note: "N" stands for number. "HH" stands for household. "TLU" stands for Tropical Livestock Units

Factor analysis of economic variables (sustainability domain 2)

For the economic variables we considered, the relevant factors seem to be the first two (Figure 3).

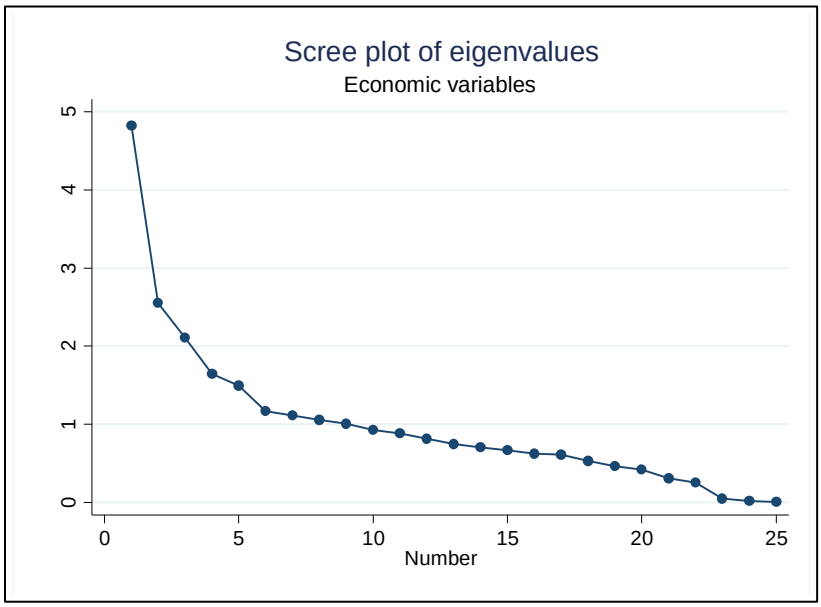


Figure 3: Scree plot of economic variables

Table 2 shows that factor 1 captures quantities of crops harvest and their use while factor 2 captures labor inputs. Dwelling characteristics and non-labor inputs do not seem to account for much of the data variation. The final list of variables considered includes Kg of harvest used for seed and Kg of harvest sold (factor 1) and use of community and hired agricultural labor by the household (factor 2).

Table 2: Factor loads for economic variables

Variable	Factor1	Factor2
Fertilizer cost	-0.0146	-0.0043
Traditional seeds cost	0.007	0.0553
Improved seeds cost	0.1772	0.119
Pesticide cost	0.1222	0.0992
Other non-labor cost	0.0315	0.1963
Animal feed cost	0.1121	0.0834
Agricultural wage	0.1831	0.7761
HH uses community labor	0.0477	0.9588
HH uses hired labor	-0.0317	0.9577
Total PD used for crops	0.262	0.22
Total harvest of grains (Kg)	0.9636	0.0676
Total harvest of stover (Kg)	0.0217	0.0661
Total harvest used for animal feed (Kg)	-0.0015	0.0751
Total harvest used for crop residual (Kg)	-0.0067	-0.0517
Total harvest used for seeds (Kg)	0.9748	-0.0053
Total harvest used for gifts (Kg)	0.0036	0.1036
Total harvest used for own consumption (Kg)	0.0907	0.1064
Total harvest used for other reasons (Kg)	-0.0082	0.0227
Total harvest sold (Kg)	0.9779	0.0483
Agri wealth index	0.2382	0.0749
Non-agri wealth index	0.231	0.165
Good floor material in dwelling	0.0945	0.1476
Good source of drinking water	0.0251	0.0451
Good toilet facility	0.0155	-0.0317
Good lighting source	0.0274	0.0842

Note: "HH" stands for household and "PD" refers to person-days.

Factor analysis of environment variables (sustainability domain 3)

For the environment domain, we identified one relevant factor concerning the problems of soil erosion and the absence of preventive measures.

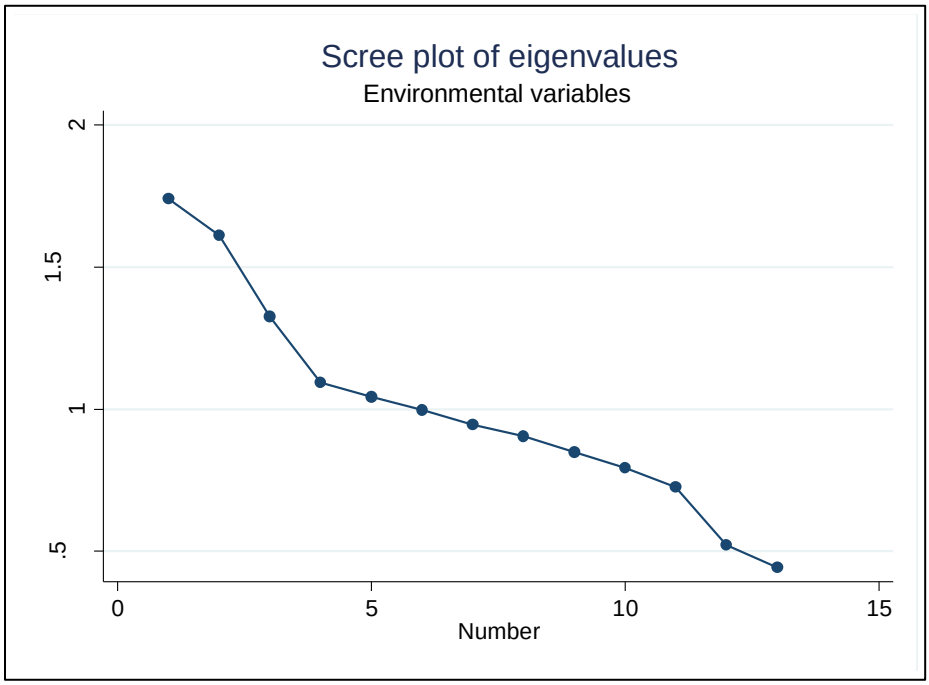


Figure 4: Scree plot of environment variables

Table 3: Factor loads for environment variables

Variable	Factor1
HH uses irrigation	0.0584
HH uses crop rotation	0.1213
HH uses fallowing	-0.0177
HH uses alternative tillage	0.309
HH uses manure	0.1049
HH uses urea	-0.0212
HH experiences soil erosion	0.8532
HH experiences soil erosion and does not takes any preventive measure	0.8426
Share of parcels with clay or loam soil	-0.0852
Share of parcels with brown or black soil	-0.0448
Share of parcels with incrusted soil	0.1813
N. of leguminous trees	-0.0267
N. of fruit trees	-0.1301

Factor analysis of social variables (sustainability domain 4)

Our dataset has a relatively small set of variables capturing social aspects, focusing on gender disparities. We thus chose only the first factor, which highlights the presence of females and females-only managed livestock as the main variables of interest.

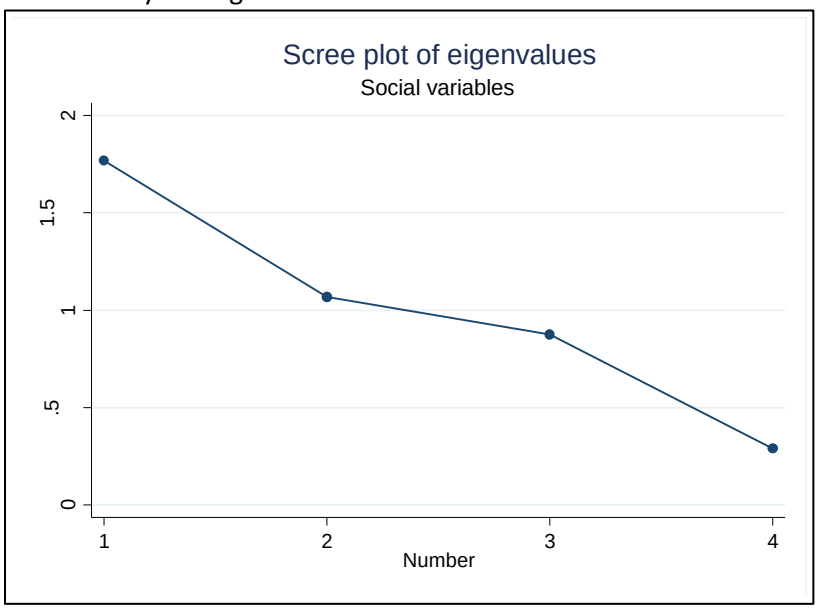


Figure 5: Scree plot of social variables

Table 4: Factor loads for social variables

Variable	Factor1
Females also responsible for plots	0.2586
Females only responsible for plots	-0.007
Females also responsible for livestock	0.922
Females only responsible for livestock	0.735

Factor analysis of human variables (sustainability domain 5)

The final sustainability domain we focus on is human capital. We select the first four factors, which capture the age composition of household members, including the prevalence of younger age groups from 0 to 29 years old (factor 1) and older age groups above 45 years old (factor 2),the level of education of household members (factor 3), and the basic characteristics of the household head (factor 4). Experiencing food shortages in the 12 months preceding interview date do not appear to play a key role in differentiating the sample. We finally select young and total dependency ratio (factor 1), mean age and mean adult age in the household (factor 2), mean level of education in the household and years of education of the household head (factor 3) and indicators of whether the household head is widow or female (factor 4).

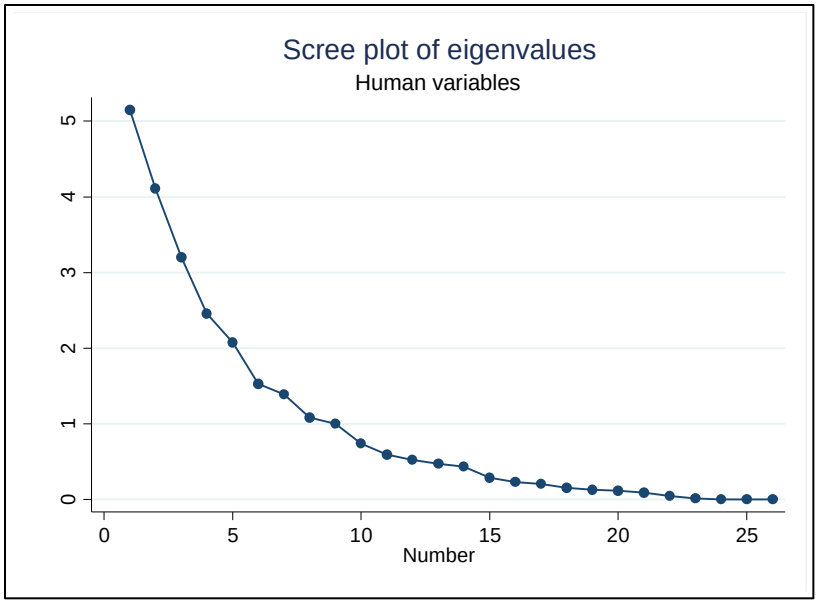


Figure 6: Scree plot of human variables

Table 5: Factor loads for human variables

Variable	Factor1	Factor2	Factor3	Factor4
HH size	0.4227	-0.1801	0.0555	-0.126
Head is married	0.1028	-0.0287	0.0446	-0.7998
Head is widow	-0.0084	0.0732	-0.0771	0.9123
Head is single	-0.0535	-0.0611	0.0219	0.0807
Head is female	0.0194	-0.0058	-0.1051	0.8408
Head is female and single	0.0571	-0.0175	-0.0086	0.1131
Head is male and single	-0.1165	-0.0638	0.0353	0.0091
Head's age	-0.1634	0.7008	-0.1236	0.1808
Head's years of educ	0.0345	-0.1141	0.8896	-0.1521
Head is literate	0.0828	-0.0959	0.7994	-0.1997
Mean years of edu.	-0.0954	-0.1668	0.9016	0.0323
Highest years of edu.	-0.2058	-0.0082	0.7497	0.0588
Mean age	-0.5321	0.8121	-0.0621	0.0365
Mean adult age	0.1472	0.9291	-0.1131	-0.0212
N. of males adults	-0.3606	-0.046	0.0673	-0.1536
N. of females adults	-0.0661	0.0551	0.0291	0.0881
children	0.666	-0.2933	-0.0733	-0.1577
Young dep. Ratio	0.9642	-0.0892	-0.0195	0.0166
Old dep. Ratio	0.1987	0.7299	-0.205	0.1209
Total dep. ratio	0.9601	0.1252	-0.0769	0.0501
Share of 0-14 y.o.	0.8875	-0.3101	0.0015	-0.0819
Share of 15-29 y.o.	-0.5595	-0.3657	0.0406	0.1289
Share of 30-44 y.o.	0.0006	-0.1414	0.0255	-0.006
Share of > 45 y.o.	-0.358	0.81	-0.0638	-0.0404
HH worries for food shortages	0.0633	0.0047	-0.0738	0.0793
Months experienced food shortages	0.0903	0.035	-0.0595	0.0428

Cluster analysis

The analysis summarized in the preceding section informed the selection of a list of factors that we used in the cluster analysis. These are 6 productivity variables, 4 economic variables, 2 environmental variables, 2 social variables and 8 human variables. Figure 7 shows the dendrogram illustrating how the farm households in our sample can be split into different groups (or types) based on these variables we have identified. The vertical distance between separations illustrates the distance of the different groups to each other.

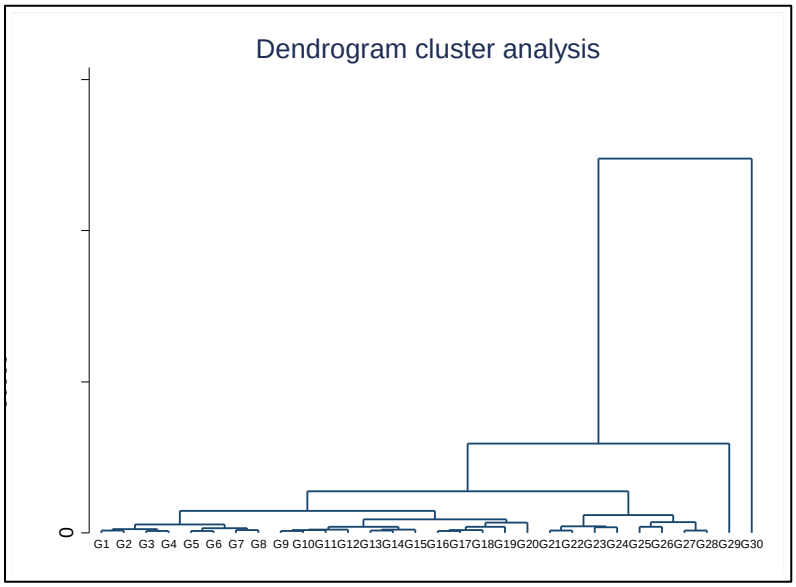


Figure 7: Dendrogram

Considering the number of observations within each group and differentiation of characteristics between groups, we decided to create four final groups, or "types" of farmers. Tables 6a to 6e illustrate the distribution of characteristics across these types and sustainability domains discussed before. Because the clusters were defined using the variables accounting for most of the data variation, as captured by the factor analysis, most of the characteristics differ significantly across every type. Type 1 is the biggest one and includes 331 of the farmers in the sample. Type two defines 242 farmers, type 3 accounts for 149 farmers and finally type 4 is the smallest, with 58 farmers.

Table 6a: Distribution of characteristics in the economic domain

	Type 1	Type 2	Type 3	Type 4
Economic Domain				
Value of fertilizer used (GHC)	2521.15	2818.87	4651.01	2722.76
	[836.09]	[1038.45]	[1955.15]	[1487.98]
Value of traditional seeds purchased (GHC)	3160.42***	6607.02	12398.47***	15344.83**
	[623.05]	[1462.87]	[3264.18]	[6810.44]
Value of improved seed purchased (GHC)	23135.83***	44942.63	87839.26***	145139.02***
	[2122.67]	[3169.06]	[6626.91]	[21441.80]
Value of pesticides used (GHC)	1912.39***	6012.19	11130.20***	19548.85***
	[442.01]	[1256.08]	[2415.51]	[7752.47]
Share of households using communal labor	0.47***	0.69*	0.85***	0.86***
	[0.03]	[0.03]	[0.03]	[0.05]
Share of households using hired labor	0.47***	0.68*	0.85***	0.81***
	[0.03]	[0.03]	[0.03]	[0.05]
Total person-days used, male & female	66.41***	105.27	159.09***	216.02***
	[2.93]	[5.37]	[9.97]	[19.07]
Total Kg of grains harvested	876.53***	2117.26*	4234.87***	11362.09***
	[39.08]	[58.42]	[131.23]	[2395.40]
Total Kg harvest used for own consumption	628.23***	973.96	1296.79***	1741.22***
	[28.64]	[41.32]	[62.10]	[132.28]
Total Kg harvest sold	107.19***	873.96**	2387.48***	8533.34***
	[7.09]	[20.40]	[50.47]	[2058.45]
Agricultural wealth index	-0.29***	-0.02	0.30***	1.08***
	[0.04]	[0.06]	[0.07]	[0.15]
Non-agricultural wealth index	-0.24***	-0.04	0.27***	0.95***
	[0.04]	[0.05]	[0.08]	[0.21]
Share of households with good floor in dwelling	0.10***	0.14	0.26***	0.34***
	[0.02]	[0.02]	[0.04]	[0.06]
Share of households with good source of drinking water	0.60***	0.67	0.72*	0.78**
	[0.03]	[0.03]	[0.04]	[0.06]
Share of households with good toilet facility	0.01	0.01	0.03	0.05**
	[0.01]	[0.01]	[0.01]	[0.03]
Share of households with good source of lighting	0.25***	0.39	0.49***	0.62***
	[0.02]	[0.03]	[0.04]	[0.06]
N. of observations	331	242	149	58

Standard errors of means in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6b: Distribution of characteristics in the social domain

	Type 1	Type 2	Type 3	Type 4
Social Domain				
Share of HH with female having shared plot responsibility	0.66	0.64	0.6	0.58
	[0.02]	[0.03]	[0.04]	[0.06]
Share of HH with female having exclusive plot responsibility	0.25***	0.16*	0.15	0.12
	[0.02]	[0.02]	[0.03]	[0.04]
Share of HH with female having shared livestock responsibility	0.09***	0.12	0.13**	0.16***
	[0.01]	[0.01]	[0.01]	[0.02]
Share of HH with female having exclusive livestock responsibility	0.04	0.04	0.04	0.06
	[0.00]	[0.00]	[0.01]	[0.01]
N. of observations	331	242	149	58

Standard errors of means in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6c: Distribution of characteristics in the productivity domain

	Type 1	Type 2	Type 3	Type 4
Productivity Domain				
Total land size (Ha)	1.65***	1.85**	2.69***	3.92***
	[0.11]	[0.11]	[0.20]	[0.39]
Share of households doing intercropping	0.69***	0.86***	0.92***	0.90*
	[0.03]	[0.02]	[0.02]	[0.04]
Share of households doing intercropping with legumes	0.60***	0.77**	0.87***	0.83*
	[0.03]	[0.03]	[0.03]	[0.05]
Area of intercropped plots	0.68***	1.2	2.54***	3.25***
	[0.06]	[0.07]	[0.45]	[0.33]
Area of plots intercropped with legumes	0.22***	0.4	0.89***	1.21***
	[0.02]	[0.02]	[0.16]	[0.15]
Share of households owning mixed livestock	0.59***	0.77**	0.83***	0.83**
	[0.03]	[0.03]	[0.03]	[0.05]
N. of different livestock types owned	2.11***	2.66**	2.72**	3.07***
	[0.08]	[0.09]	[0.11]	[0.17]
Share of households cultivating maize only	0.12***	0.05*	0.03**	0.03
	[0.02]	[0.01]	[0.01]	[0.02]
Share of households growing cereals	0.96**	0.99	0.99	1
	[0.01]	[0.01]	[0.01]	[0.00]
Share of households growing vegetables	0.03	0.02	0.05	0.02
	[0.01]	[0.01]	[0.02]	[0.02]
Share of households growing legumes	0.64***	0.81**	0.89***	0.93***
	[0.03]	[0.03]	[0.03]	[0.03]
Area of cereals(ha)	0.95***	1.03*	1.32	2.60***
	[0.06]	[0.08]	[0.12]	[0.39]
Area of vegetables(ha)	0.01	0	0.01	0
	[0.00]	[0.00]	[0.00]	[0.00]
Area of legumes(ha)	0.25***	0.44	0.75***	1.26***
	[0.02]	[0.02]	[0.05]	[0.13]
Production of cereals(kg)	734.91***	1621.09*	3080.93***	6804.91***
	[36.23]	[62.80]	[170.13]	[992.02]
Production of vegetables(kg)	5.71	11.61	36.55**	6.9
	[2.14]	[7.75]	[19.24]	[6.90]
Production of legumes(kg)	99.83***	291.78	601.46***	1107.91***
	[7.07]	[19.78]	[51.10]	[123.68]
Yield of cereals(kg/ha)	1447.62***	2198.97**	2710.67***	2906.20***
	[71.30]	[87.61]	[106.79]	[189.78]
Yield of vegetables(kg/ha)	2246.08	2652.21	5090.14*	2635.73
	[911.10]	[628.58]	[1765.15]	[.]
Yield of legumes(kg/ha)	537.05***	722.26	907.02***	936.84***
	[29.27]	[36.32]	[45.66]	[92.75]
TLU small ruminants	0.45***	0.59	0.61	1.04***
	[0.05]	[0.06]	[0.07]	[0.14]
TLU big ruminants	1.92***	2.45	3.17**	6.03***
	[0.18]	[0.18]	[0.24]	[0.71]
TLU poultry	0.03***	0.05	0.07***	0.09***
	[0.00]	[0.00]	[0.01]	[0.01]
Kg fertilizer used	785.62***	1478.83	2623.97***	3267.33***
	[87.81]	[142.31]	[246.03]	[533.49]
N. of observations	331	242	149	58

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6d: Distribution of characteristics in the environmental domain

	Type 1	Type 2	Type 3	Type 4
Environmental Domain				
Share of households practicing irrigation	0.01***	0.03	0.03	0.07**
	[0.00]	[0.01]	[0.01]	[0.03]
Share of households practicing rotation	0.11***	0.21	0.30***	0.47***
	[0.02]	[0.03]	[0.04]	[0.07]
Share of households practicing fallowing	0.03	0.02	0.05	0.02
	[0.01]	[0.01]	[0.02]	[0.02]
Share of households practicing alternative tillage	0.02***	0.00*	0	0.02
	[0.01]	[0.00]	[0.00]	[0.02]
Share of households using manure on (any) plot in either season	0.47***	0.59	0.74***	0.6
	[0.03]	[0.03]	[0.04]	[0.06]
Share of households using urea on (any) plot in either season	0.01	0.02	0.02	0.03
	[0.00]	[0.01]	[0.01]	[0.02]
Share of households affected by soil erosion	0.12	0.14	0.11	0.16
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households with soil erosion but no erosion control measure	0.03	0.04	0.04	0.07
	[0.01]	[0.01]	[0.02]	[0.03]
Average share of parcels with clay or loam soil	0.53***	0.62	0.72***	0.85***
	[0.03]	[0.03]	[0.03]	[0.04]
Average share of parcels with black or brown soil	0.26	0.25	0.28	0.23
	[0.02]	[0.02]	[0.03]	[0.04]
Average share of parcels with incrusted soil	0.35	0.35	0.38	0.29
	[0.03]	[0.03]	[0.04]	[0.06]
N. of leguminous trees owned	0.86*	1.32	1.54	0.95
	[0.13]	[0.29]	[0.31]	[0.46]
N. of fruit trees owned	3.45**	6.23	4.93	12.36**
	[0.81]	[1.75]	[1.49]	[4.72]
N. of observations	331	242	149	58

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6e: Distribution of characteristics in the human domain

	Type 1	Type 2	Type 3	Type 4
Human Domain				
Household size	5.82***	6.11	6.88***	6.81*
	[0.14]	[0.17]	[0.24]	[0.39]
Share of married heads	0.76***	0.83	0.84	0.97***
	[0.02]	[0.02]	[0.03]	[0.02]
Share of female heads	0.19***	0.11	0.09	0.02***
	[0.02]	[0.02]	[0.02]	[0.02]
Age of the head	46.5	45.91	47.29	46.45
	[0.81]	[0.90]	[1.08]	[1.70]
Years of education of the heads	5.01***	5.77	5.65	6.28*
	[0.22]	[0.22]	[0.24]	[0.36]
Share of literate heads	0.70***	0.8	0.82	0.90**
	[0.03]	[0.03]	[0.03]	[0.04]
Mean years of education in the household	5.32***	6.48***	6.59***	6.64*
	[0.16]	[0.16]	[0.18]	[0.30]
Max years of education in the household	7.42***	8.47	9.12***	8.91*
	[0.20]	[0.20]	[0.24]	[0.38]
Average age of adults in the household	24.38	23.85	24.47	24.35
	[0.60]	[0.66]	[0.82]	[1.22]
Number of children in the household	1.08	1.1	1.05	1.19
	[0.06]	[0.07]	[0.09]	[0.15]
Young dependency ratio	1.11	1.06	1.06	1.1
	[0.05]	[0.06]	[0.07]	[0.11]
Old dependency ratio	0.12	0.1	0.11	0.11
	[0.02]	[0.02]	[0.02]	[0.03]
Share of HH worrying about food shortages	0.26***	0.11***	0.13	0.03***
	[0.02]	[0.02]	[0.03]	[0.02]
Months experiencing food shortages?	0.71***	0.24**	0.13***	0.00***
	[0.09]	[0.06]	[0.05]	[0.00]
N. of observations	331	242	149	58

Standard errors of means in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

The four types differ from each other across all of the five domains, as shown in table 6. One of the striking characteristic that stands out in differentiating them is the level of endowments, as measured by a wealth index including dwelling characteristics, size of the cultivated land and ownership of agricultural and non-agricultural assets (figure 8). We defined low-endowed households as the ones in the bottom quartile of the wealth distribution, mid-endowed households as the ones in the 2nd and 3rd quartile and highly endowed households as the ones in the top quartile of the asset distribution. Figure 8 shows in which of the endowments category fall most of the households in our typologies. More broadly, the types can be characterized as following:

Type 1: Female-headed, low educated households with low levels of endowments

- High number of female headed households, with heads less likely to be married and with low education attainments and literacy rates.
- High proportion of women with plot responsibilities but low proportion of women with livestock responsibilities.
- Very high food insecurity.
- Little asset ownership (land below 2 Ha, little livestock, low agricultural and non-agricultural wealth), and bad dwelling conditions.

- Low production and productivity of all major crops, also due to low input use (both in terms of labor inputs, which are mainly composed by family labor, and non-labor inputs).
- Most crop production devoted to own consumption.
- Little use of soil conservation practices.

Type 2: Young medium-endowed households

- Relatively small and young households with mid-levels of education.
- Low productivity and input use, even though better than type 1, and mid-levels of endowments. Low land size (below 2 Ha).
- More likely to grow vegetables than other groups.
- Little use of soil conservation practices.

Type 3: Medium-endowed households growing vegetables and practicing intercropping

- Large households with high levels of educational attainment
- More likely to do intercropping and grow vegetables than other groups. Second group most likely to grow legumes after group 4.
- Medium levels of crop production and high productivity, coupled with very large use of fertilizer and hired labor.
- Medium levels of endowments, with average land size between 2 and 3 Ha.
- Frequent use of manure but also problems of soil incrustation.

Type 4: Highly endowed households breeding livestock and growing legumes

- Male headed households with high levels of literacy rates and educational attainments.
- High percentage of women with some livestock responsibilities.
- Very high levels of food security (Average months of food shortages close to zero).
- Extremely high asset ownership (large land above 3 Ha, high number of livestock types and units, high agriculture and non-agriculture index) and very good dwelling conditions.
- High production and productivity of crops with high input use. Very high share of households growing legumes (93%).
- Frequent use of crop rotation and irrigation and high share of clay or loam soils.

Table 7 summarizes the main characteristics of every type relative to each sustainability domain, providing a simplified framework for classifying farm households into a particular type. Figure 9 shows a graphic representation of the main characteristics of each type.

Table 7: Matrix of performance for each SI domain

	Productivity	**Economic**	**Environment**	**Social (gender)**	**Human**
Type 1: Female-headed, low educated households with low levels of endowments	Low crop production and productivity. Little livestock owned.	Low wealth (agri and non-agri), land size below 2 Ha, low input expenditure, most harvest going to own consumption rather than sales.	Little use of soil conservation practices.	High frequency of female responsibility for crops but opposite for livestock.	Female heads with low levels of literacy and education. Very low food security.
Type 2: Young medium-endowed households	Low crop production and productivity. Little livestock owned.	Low-medium wealth (agri and non-agri), land size below 2 Ha, low input expenditure, same proportion of harvest going to own consumption and sales.	Little use of soil conservation practices.	Average gender equality.	Small households with low dependency ratio. Relatively low food security.
Type 3: **Medium-endowed households growing vegetables and practicing intercropping**	High crop production and productivity. Frequent intercropping. Vegetable growers and second largest legume growers.	Medium-high wealth (agri and non-agri), high input use (especially fertilizer and hired labor), harvest going to sales twice the amount going to own consumption.	Frequent use of manure but also problems of soil incrustation.	Average gender equality.	Large households with married male heads and high levels of education and literacy.
Type 4: **Highly endowed households breeding livestock and growing legumes**	Very high crop production and productivity. High livestock ownership. Legume growers.	Very high wealth (agri and non-agri), high input use, harvest going to sales five times the amount going to own consumption. Good dwelling conditions.	Frequent use of soil conservation practices but problems with soil erosion.	High frequency of female responsibility for livestock but opposite for crops.	Very large households with married male heads and high levels of education and literacy. Very high food security.

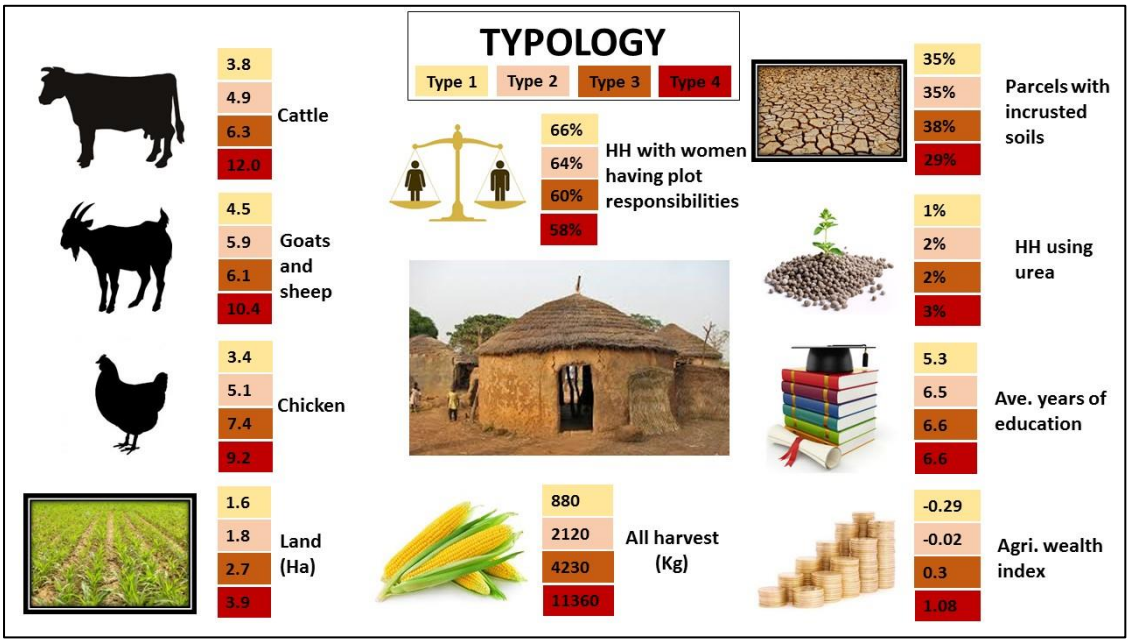


Figure 9: Graphic representation of types

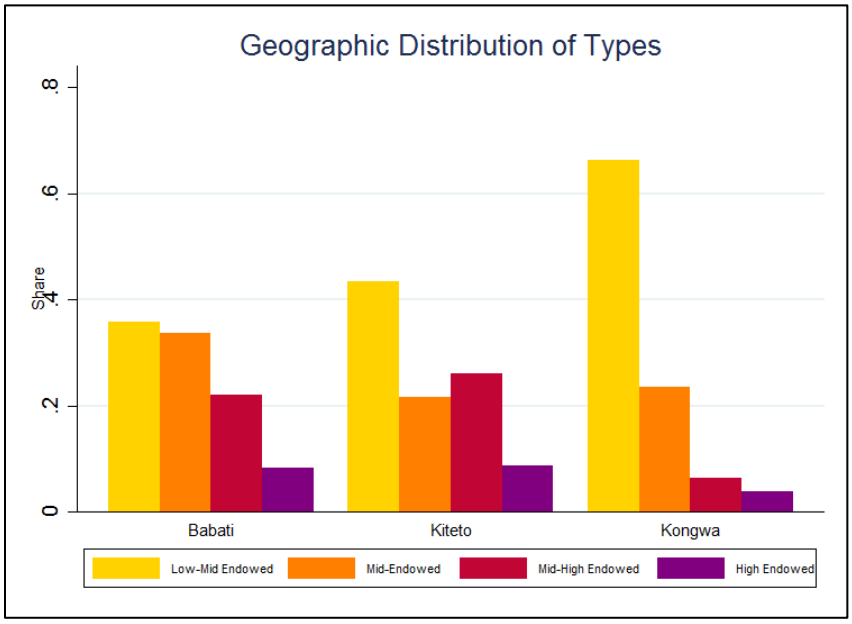


Figure 10: Distribution of typologies by Districts

The differences in climatic conditions between groups are an indication of heterogeneity of typology distribution across space. Figure 10 shows the typology composition of each district in the sample. While in Kongwa there is a very high concentration of female-headed, low educated households with low levels of endowments (type 1), Kiteto and espacially Babati concentrate high shares of Mid-endowed and high endowed households (type 3 and 4). Similar differences appear when we look at the regional typology distribution (Figure 11), with the Manyara region hosting a large portion of richer types while the Dodoma region mostly concentrates the

poorest households (type 1). The spatial distinctions are important because they can support interventions based on the most prevalent households' typologies in the area.

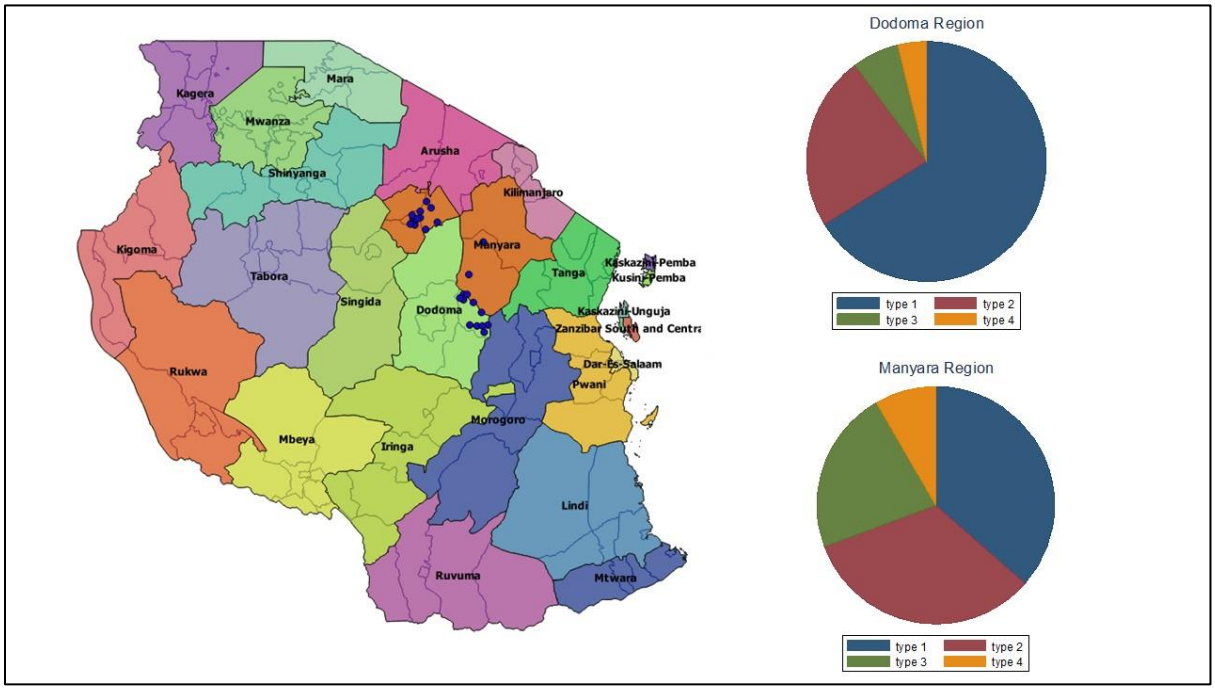


Figure 11: Distribution of Typologies by regions

The characteristics of each household type described above can be displayed clearly with a spider plot. Figure 12 summarizes the performance of each type relative by each domain as follows:

- Type 3 and 4 are the most productive groups, while type 2 shows mid-levels of productivity and type 1 is lagging far behind.
- In terms of economic endowments, type 4 differentiates itself with a very strong performance, while the other groups are fairly close to each other at a lower level.
- In terms of human endowments, here measured by educational attainments, types 2,3 and 4 are very similar, and type 1 differentiate itself with very low levels of performance.
- Finally, type 1 and type 4, despite their wide differences in productivity, economic and human endowments, perform similarly in terms of soil conservation practices and gender equality. While type 4 might have a high respect for women and the environment because of choice, in the case of type 1 this might be a necessity driven by the scarcity of resources.

Recommendations

- AR can focus on increasing the productive capacity and economic endowments for group 1 and 2 through the introduction of superior agricultural technologies. In addition, through nutrition trainings the project can improve the food insecurity of type 1 and, to a minor extent, of type 2.
- Secondly, AR can focus on sharing information about the importance of preserving soil fertility and improving gender equality. This will improve the performance of type 2 and 3 in these two domains and prevent the degradation of the scores of type 1, which may result from the improvement of his economic and productivity conditions.

The appendix includes additional graphs characterizing the obtained typologies.

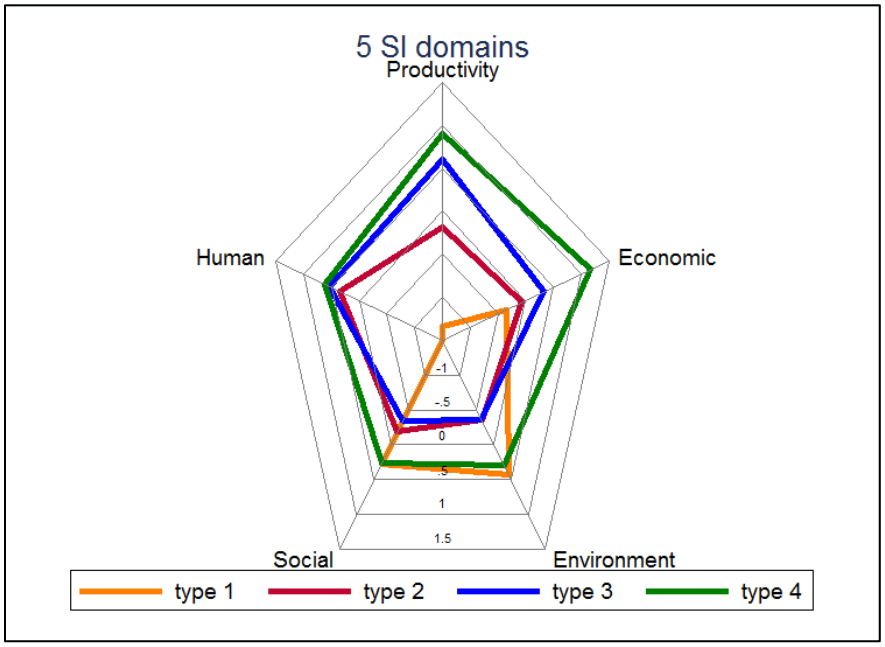


Figure 12: Typologies performance by sustainability domain

NOTE: The following variables are used to measure each domain: cereals yield (Productivity), asset-based wealth index (Economic), soil conservation index composed of crop rotation, alternative or minimum/zero tillage, experience of soil erosion without measures for mitigating it and share of parcels with incrusted soils (Environment); gender equality index composed by female responsibility in managing certain plots and livestock (Social), and average education in the household (Human).

References

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications* (Vol. 418). John Wiley & Sons.

Cunningham, W., & Maloney, W. F. (1999). Heterogeneity among Mexico's micro-enterprises: an application of factor and cluster analysis. *World Bank Policy Research Working Paper*, (1999).

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863-14868.

Galbraith, J. I., Moustaki, I., Bartholomew, D. J., & Steele, F. (2002). *The analysis and interpretation of multivariate data for social scientists*. CRC Press.

Harman, H. H. (1976). *Modern factor analysis*. University of Chicago Press.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

Kim, J. O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues* (Vol. 14). Sage.

McDonald, R. P. (2014). *Factor analysis and related methods*. Psychology Press.

Mulaik, S. A. (2009). *Foundations of factor analysis*. CRC press.

Romesburg, C. (2004). *Cluster analysis for researchers*. Lulu.com.

Sethi, S. P. (1971). Comparative cluster analysis for world markets. *Journal of Marketing Research*, 348-354.

Appendix Figures

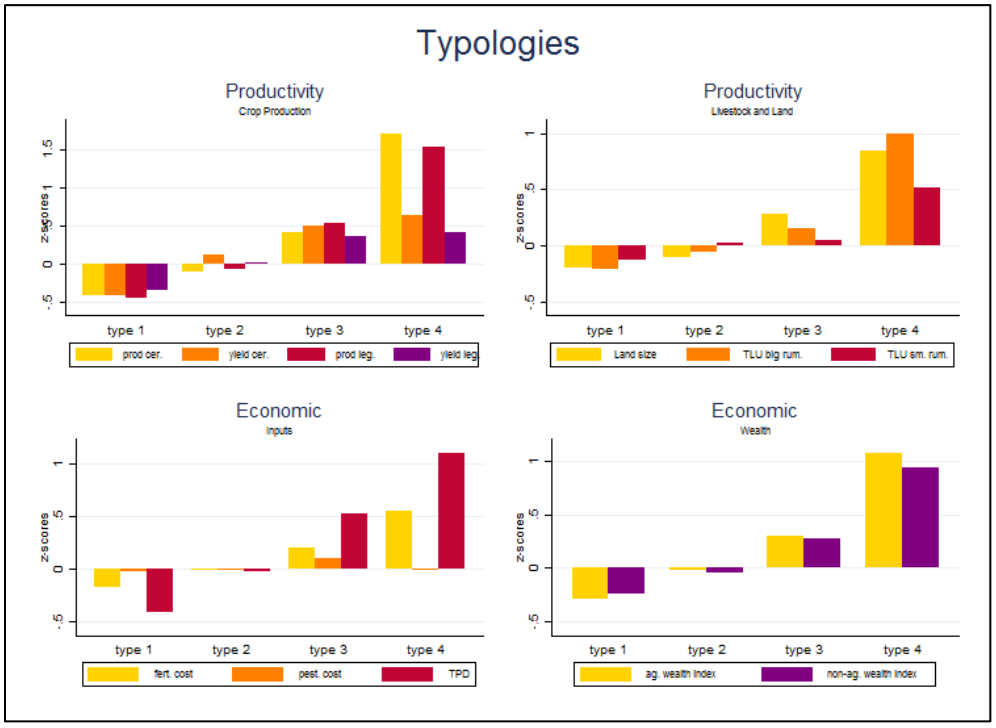


Figure A1: Typologies by domain (productivity and economic)

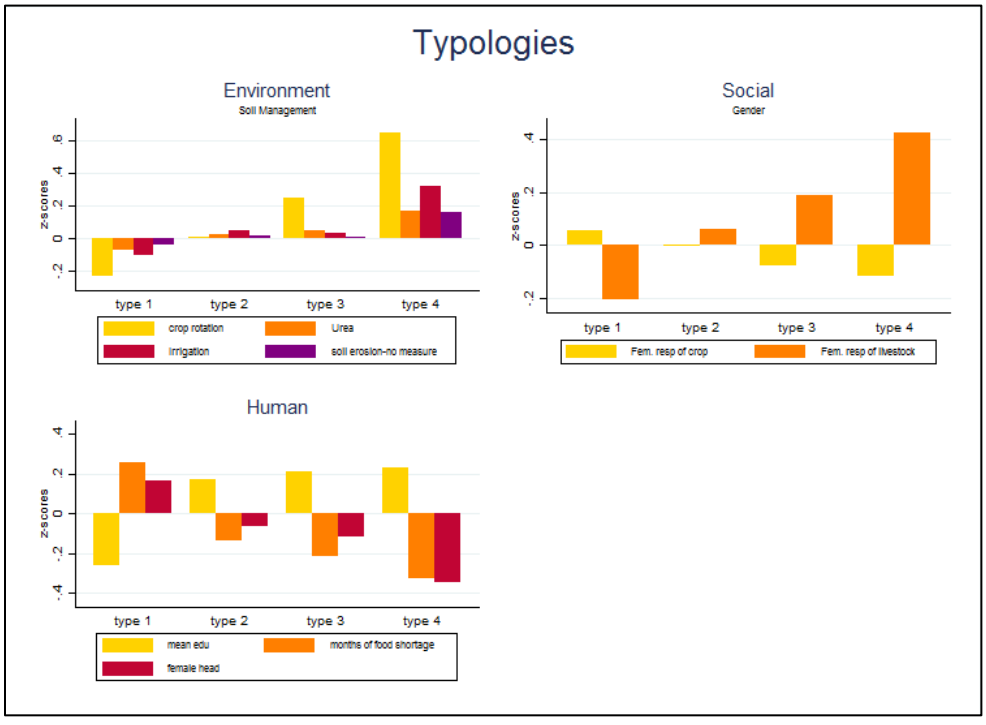


Figure A2: Typologies by domain (environment, social and human)

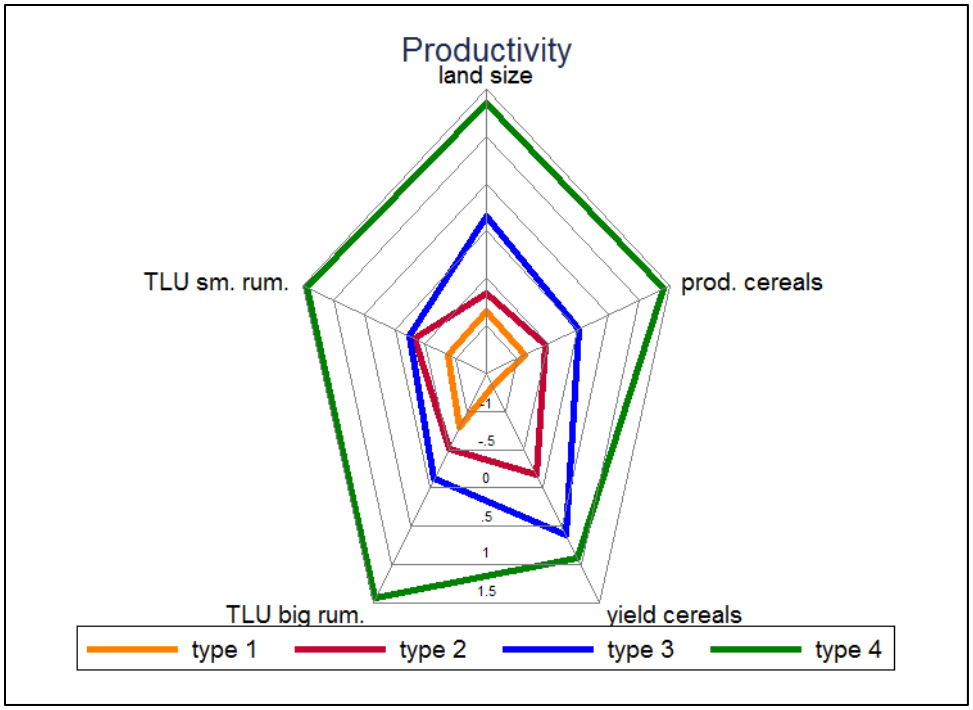


Figure A3: Radar graph – productivity (z-scores)

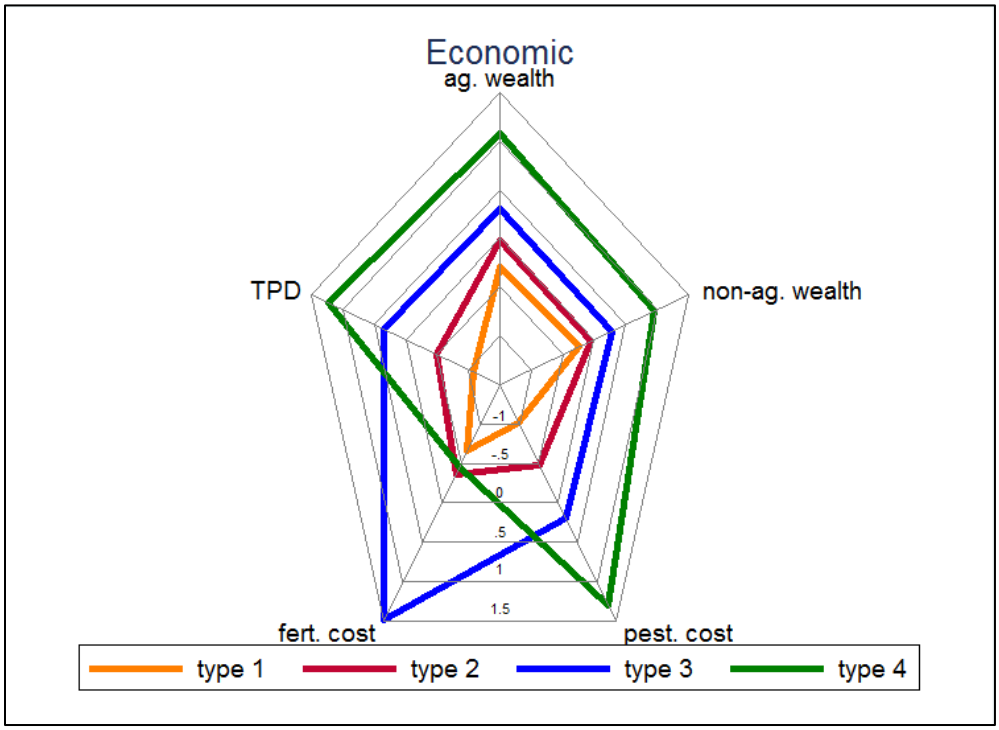


Figure A4: Radar graph – economic (z-scores)

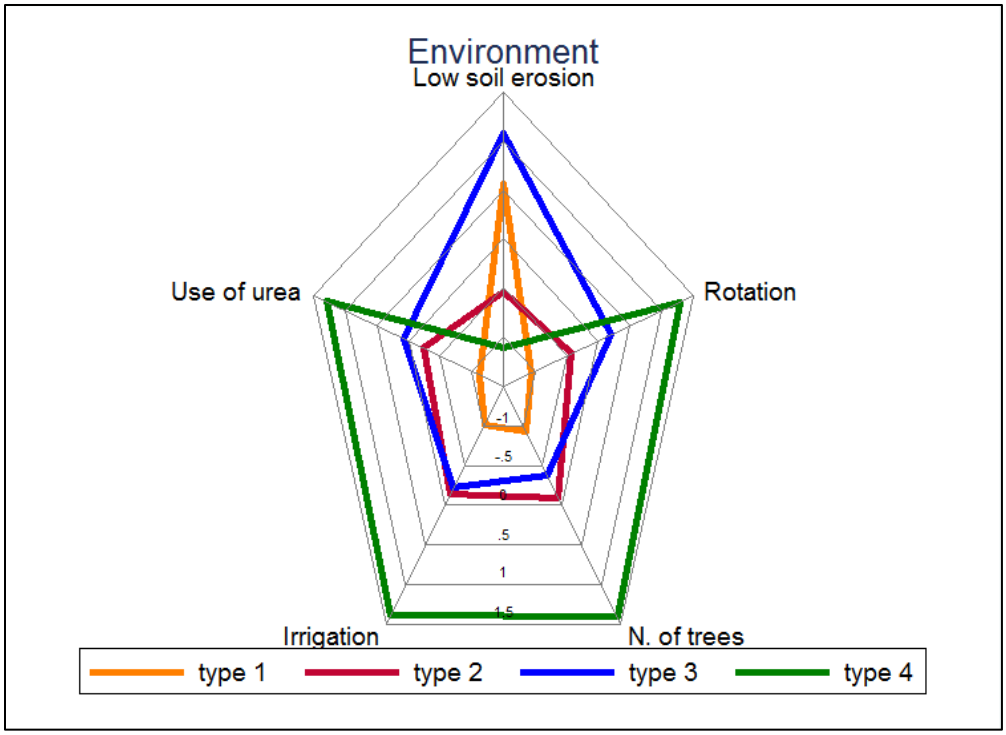


Figure A5: Radar graph – environnent (z-scores)

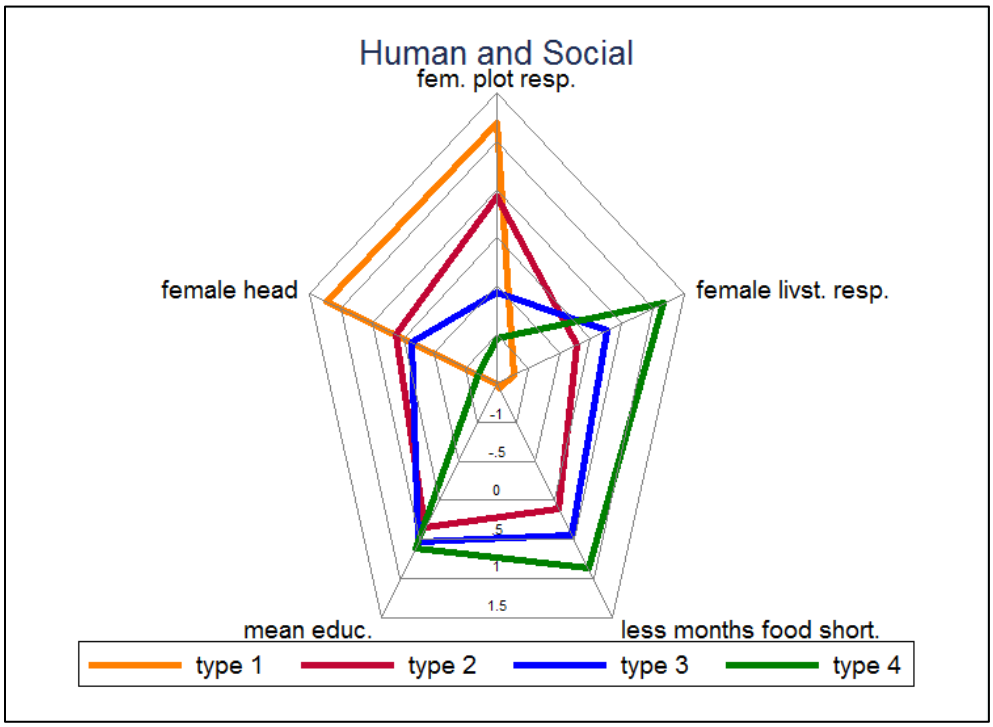


Figure A6: Radar graph – social and human (z-scores)