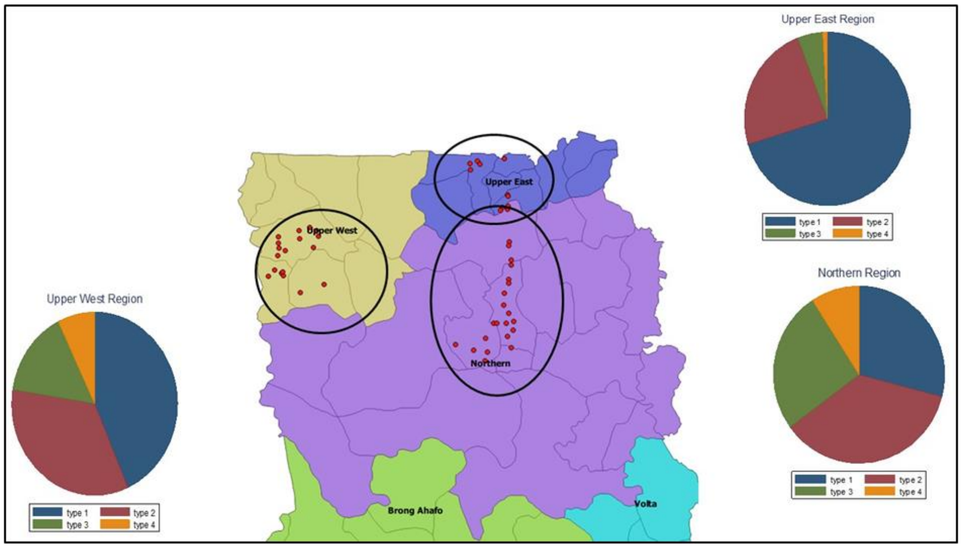
Typology characterization of farmers in Africa RISING sites in Ghana

Sara Signorelli, Carlo Azzarri and Beliyou Haile



Produced by International Food Policy Research Institute

Published by International Institute of Tropical Agriculture

March 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 Ghana Africa RISING Baseline Evaluation Survey (GARBES) (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 Ghana).

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.

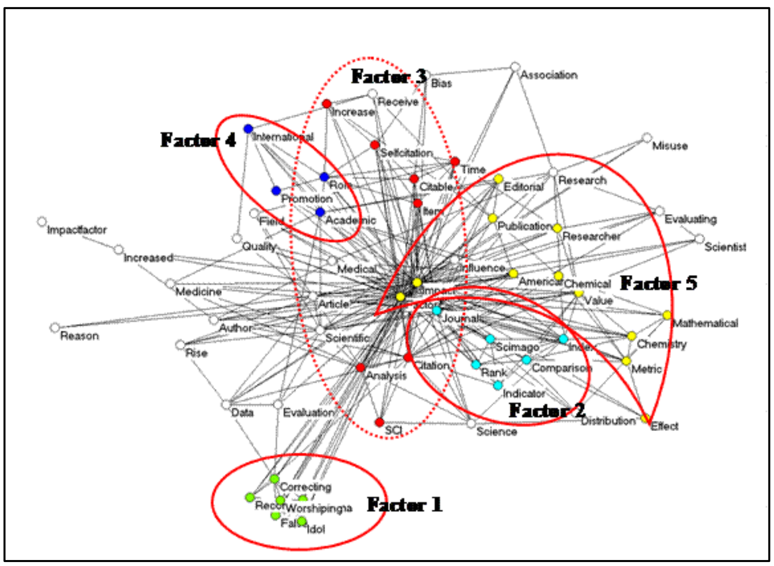


Figure 1: Example of factor analysis[1]

Our factor analysis based on GARBES 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 GARBES 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.

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

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.

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 four factors (represented by the first four dots at the top of the line graph) are highly relevant but that the 5th factor starts to be less important in explaining the variation (smaller vertical jump).

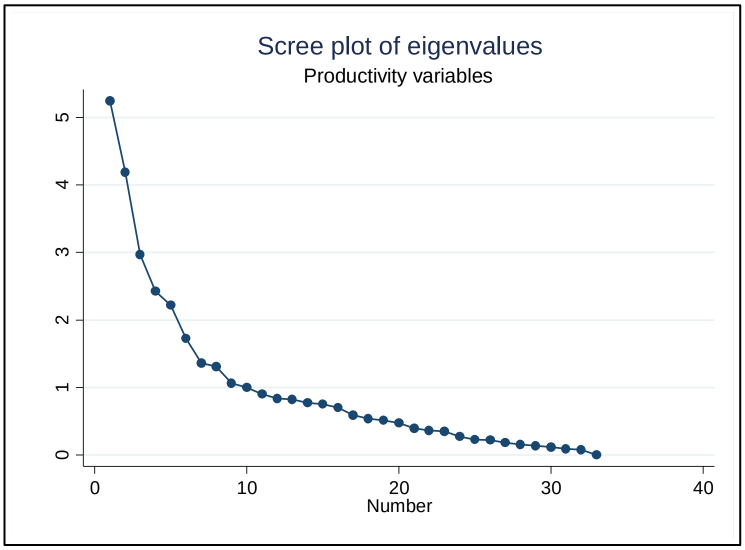


Figure 2: Scree plot of productivity variables

Table 1 shows the rotated matrix of factor loads for the four 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 size of the land, cereal production and fertilizer usage. Factor 3 captures the production of legumes. Finally, factor 4 captures livestock ownership. The final selection of variables for the cluster analysis include share of households doing intercropping and average number of intercropped plots for factor 1, area cultivated with cereals and total cereal production for factor 2, share of households cultivating legumes and total production of legumes for factor 3, and share of households owning mixed livestock and number of livestock types owned for factor 4

Table 1: Factor loads of productivity variables

Variable	Factor1	Factor2	Factor3	Factor4
Land size (Ha)	0.0334	0.7285	0.3413	0.0496
N. parcels	-0.0394	0.2328	0.086	0.1844
Min distance plot	-0.115	0.0634	0.1099	-0.1891
Max distance plot	0.0023	0.1405	0.0296	0.0418
N. trees	0.0847	0.4484	0.4054	0.0344
N. crops	0.8717	-0.0569	-0.0614	-0.0455
N. plots	0.0414	0.1837	0.4045	0.2119
HH does intercropping	0.9251	-0.0488	-0.0052	0.0401
HH does intercropping with legumes	0.8594	-0.1088	0.1931	0.0028
N. of intercropped plots	0.9057	-0.0313	-0.0044	0.0355
Size intercropped land (Ha)	0.8034	0.2789	0.0482	0.0497
Size legumes-intercropped land (Ha)	0.7674	0.0639	0.2367	0.0283
Ownership mixed livestock	0.0371	0.0019	0.0597	0.7974
N. livestock types owned	0.0674	0.068	0.0445	0.8927
Maize only crop	-0.0527	-0.0235	-0.0743	-0.0316
Mixed crops	0.0527	0.0235	0.0743	0.0316
Cultivation of cereals	0.0755	0.0795	-0.027	-0.0099
Cultivation of vegetables	-0.0301	0.007	0.002	-0.0391
Cultivation of legumes	0.1425	-0.184	0.7438	0.0646
Area cultivated with cereals (Ha)	0.0173	0.8582	-0.0184	0.0831
Area cultivated with vegetables (Ha)	0.0192	0.076	0.0099	-0.0568
Area cultivated with legumes (Ha)	0.1922	0.1766	0.7373	0.0322
Production cereals (Kg)	-0.0443	0.7973	0.0584	0.0936
Production vegetables (Kg)	-0.0428	-0.0172	-0.0336	0.0006
Production legumes (Kg)	0.0338	0.2815	0.7856	0.0537
Yield cereals (Kg/Ha)	-0.0508	0.2248	0.0722	0.0518
Yield vegetables (Kg/Ha)	-0.0498	-0.0421	-0.0318	0.0162
Yield legumes (Kg/Ha)	0.0076	0.0328	0.7139	0.0518
TLU small ruminants	-0.0223	0.2027	0.0644	0.7401
TLU big ruminants	-0.0237	0.2078	0.0919	0.4352
TLU poultry	0.0141	0.2111	-0.0077	0.4049
Fertilizer used (Kg)	-0.0762	0.6632	0.0433	0.1355
HH does irrigation	-0.0184	-0.105	-0.1007	0.0804

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). Table 2 shows that factor 1 captures elements linked to labor inputs while factor 2 captures quantities of crops harvest and their use. Dwelling characteristics and non-labor inputs do not seem to account for much of the data variation. The final list of variables considered includes total and male person days used for crop cultivation (factor 1) and total harvest and harvest used for seeds (factor 2).

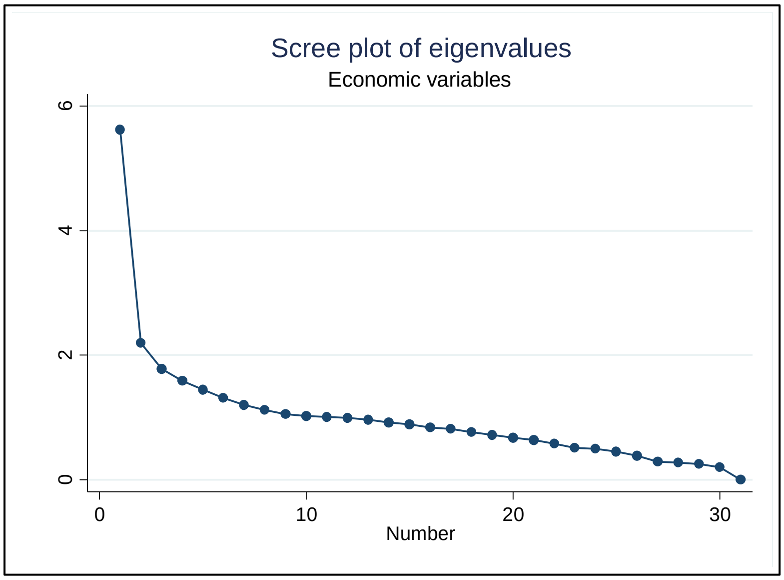


Figure 3: Scree plot of economic variables

Table 2: Factor loads of economic variables

Variable	Factor1	Factor2
Fertilizer cost	0.2774	0.1762
Traditional seeds cost	0.3377	-0.0552
Improved seeds cost	0.0257	0.2559
Pesticide cost	0.2531	0.2106
Other non-labor cost	0.256	-0.1221
Animal feed cost	0.0453	-0.2142
Agricultural wage	0.0097	-0.0341
HH uses community labor	0.2283	0.0905
HH uses hired labor	0.0649	0.0544
Total PD used for crops	0.9185	0.2251
Male PD used for crops	0.8236	0.252
Female PD used for crops	0.8477	0.1267
Family PD used for livestock	0.3445	-0.2393
Hired PD used for livestock	0.0992	0.1072
Total harvest of grains (Kg)	0.2604	0.8369
Total harvest of stover (Kg)	0.0125	0.0991
Total harvest used for animal feed (Kg)	0.0543	0.0485
Total harvest used for crop residual (Kg)	-0.0675	-0.013
Total harvest used for seeds (Kg)	0.179	0.7624
Total harvest used for gifts (Kg)	0.2097	0.6141
Total harvest used for own consumption (Kg)	0.3623	0.638
Total harvest sold (Kg)	0.0055	-0.0143
Total harvest used for other reasons (Kg)	0.1845	0.5797
Agri wealth index	0.3968	0.2294
Non-agri wealth index	0.2822	0.0745
Good floor material in dwelling	-0.0934	0.0614
Good source of drinking water	-0.1344	0.0776
Good toilet facility	0.0787	-0.0459
Good lighting source	0.0093	-0.0021

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 three relevant factors. The first concerns farmers' experience of soil erosion and the use of mitigating measures to prevent it, the second includes fallowing practices and the number of trees grown on the land and the third captures soil characteristics in terms of material (clay/loam) and color (black/brown). These six indicators are selected for the cluster analysis.

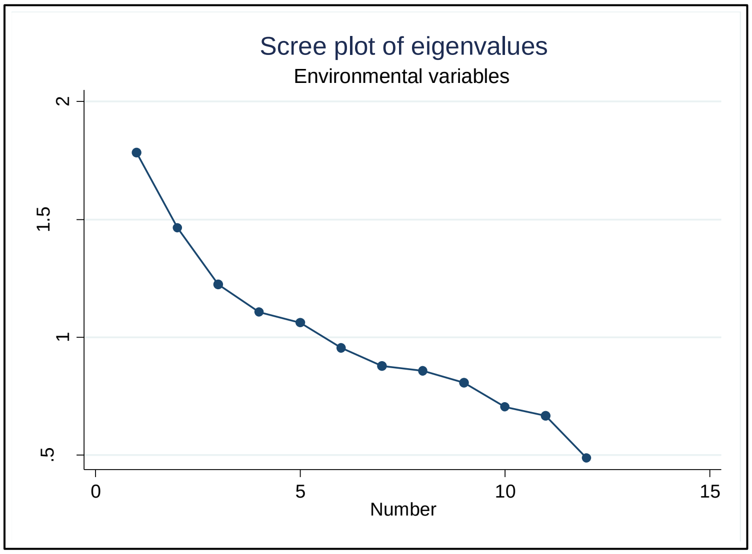


Figure 4: Scree plot of environment variables

Table 3: Factor loads of environment variables

Variable	Factor1	Factor2	Factor3
HH uses irrigation	0.0829	-0.3568	-0.2402
HH uses crop rotation	0.2095	0.294	-0.4721
HH uses fallowing	0.1724	0.6635	-0.0267
HH uses alternative tillage	-0.1276	0.1716	0.1044
HH uses manure	0.2884	-0.3357	-0.0342
HH uses urea	-0.0975	0.033	-0.0606
HH experiences soil erosion	0.8349	-0.0426	-0.017
HH experiences soil erosion and does not takes any preventive measure	0.7983	0.089	-0.0211
Share of parcels with clay or loam soil	0.0843	0.0253	0.6706
Share of parcels with black or brown soil	-0.0487	-0.03	0.7372
Share of parcels with incrusted soil	0.3938	-0.1166	0.0785
Number of trees owned	-0.0673	0.7535	-0.0559

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-only managed plots and livestock as the main variables of interest.

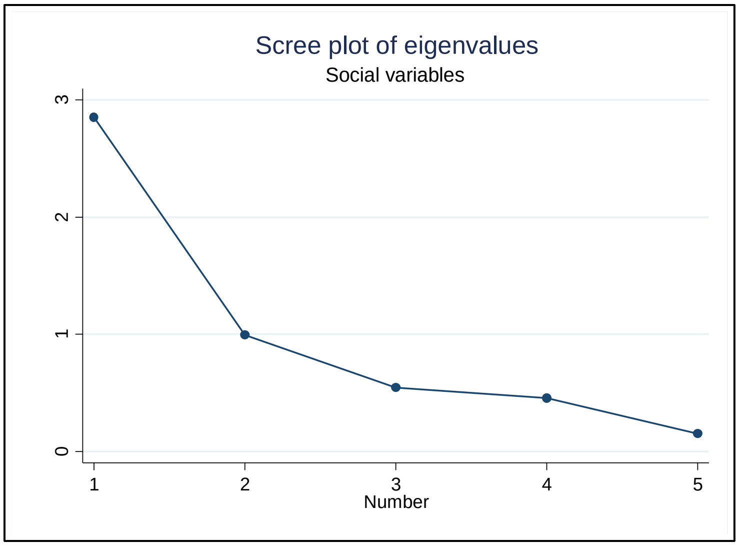


Figure 5: Scree plot of social variables

Table 4: Factor loads of social variables

Variable	Factor1
Females also responsible for plots	0.8528
Females only responsible for plots	0.8532
Females also responsible for livestock	0.8108
Females only responsible for livestock	0.8527
Gender ag. wage gap	0.116

Factor analysis of human variables (Sustainability Domain 5)

The final sustainability domain we focus on 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 size and gender composition of the household (factor 3), and education of household members (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 household size, the number of female adults, young dependency ratio, total dependency ratio, mean age among adults and among all members, and education level and literacy of the head.

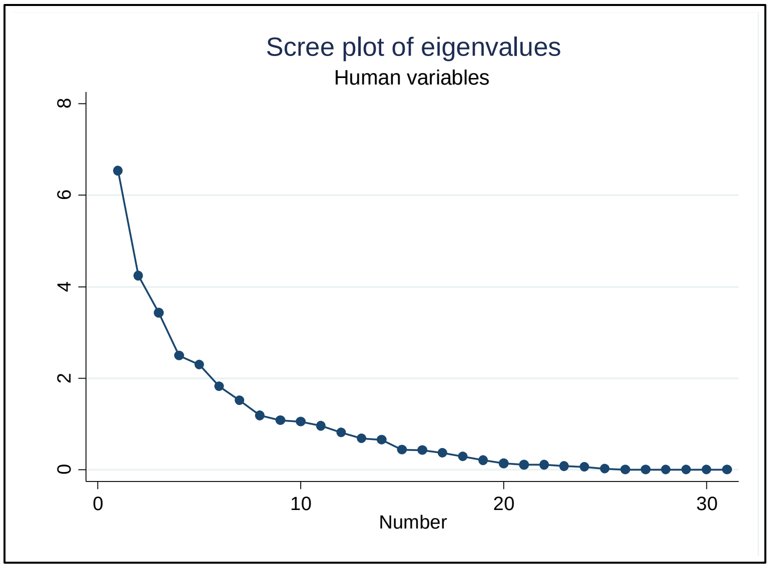


Figure 6: Scree plot of human variables

Table 5: Factor loads of human variables

Variable	Factor1	Factor2	Factor3	Factor4
HH size	0.1382	-0.1157	0.9606	-0.0315
Head is married	0.1006	-0.045	0.0625	-0.0117
Head is widow	-0.0651	0.0175	-0.0223	-0.021
Head is single	-0.0733	0.044	-0.0475	0.0269
Head is female	-0.02	0.032	-0.1622	-0.0113
Head is female and single	-0.0323	0.028	-0.0042	0.0086
Head is male and single	-0.0675	0.0385	-0.0477	0.0255
Head's age	-0.08	0.567	0.2733	-0.1428
Head's years of educ	0.0171	0.0106	-0.1027	0.8879
Head is literate	0.036	0.0217	-0.0752	0.8233
Mean years of edu.	-0.2565	-0.1202	-0.0005	0.7993
Highest years of edu.	-0.3299	-0.0707	0.2854	0.6672
Mean age	-0.4352	0.8617	-0.146	0.0187
Mean adult age	0.2775	0.907	-0.0929	-0.0839
N. of males adults	-0.363	-0.0249	0.738	-0.0009
N. of females adults	-0.0785	0.008	0.868	0.0256
Children	0.3511	-0.2106	0.7579	-0.0715
Young dep. Ratio	0.9667	-0.1035	0.0126	-0.0616
Old dep. Ratio	0.2171	0.7676	-0.0084	0.0487
Total dep. Ratio	0.9442	0.1821	0.0083	-0.0377
Share of 0-14 y.o.	0.8797	-0.3323	0.1008	-0.0989
Share of 15-29 y.o.	-0.6277	-0.2955	0.0541	0.1414
Share of 30-44 y.o.	-0.062	-0.1848	-0.0577	-0.017
Share of > 45 y.o.	-0.2379	0.8532	-0.1199	-0.0293
HH worries for food shortages	0.0569	0.0181	-0.0051	-0.0775
Months experienced food shortages	-0.0445	0.056	-0.0486	-0.0124

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 8 productivity variables, 4 economic variables, 6 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.

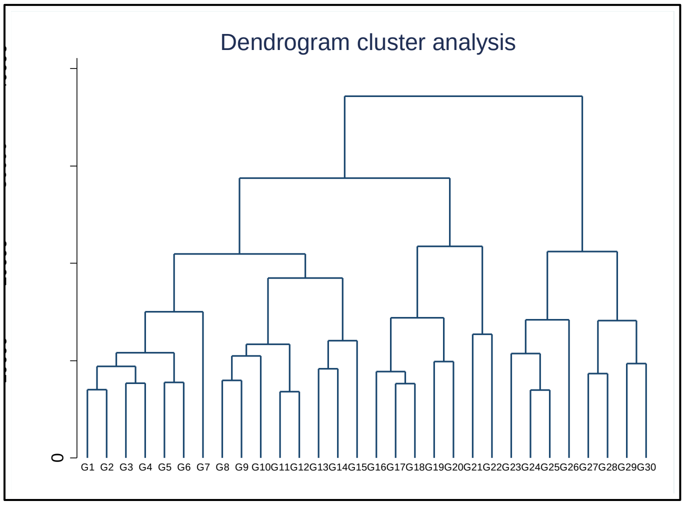


Figure 7: Dendrogram

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-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 527 of the farmers in the sample (41%). Type two defines 424 farmers (33%), type 3 accounts for 238 farmers (19%) and finally type 4 is the smallest, with 90 farmers (7%).

Table 6a: distribution of characteristics by type in the productivity domain

	Type 1	Type 2	Type 3	Type 4
Productivity Domain				
Total land size (Ha)	2.00***	3.43**	4.76***	5.31***
	[0.06]	[0.11]	[0.20]	[0.30]
Share of households doing intercropping	0.3	0.34**	0.24**	0.28
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households doing intercropping with legumes	0.2	0.25**	0.16**	0.21
	[0.02]	[0.02]	[0.02]	[0.04]
Area of intercropped plots	0.28***	0.54	0.68***	0.72**
	[0.02]	[0.05]	[0.12]	[0.15]
Area of plots intercropped with legumes	0.14***	0.27***	0.21	0.29
	[0.01]	[0.03]	[0.04]	[0.07]
Share of households owning mixed livestock	0.72***	0.81	0.89***	0.88**
	[0.02]	[0.02]	[0.02]	[0.03]
N. of different livestock types owned	2.30***	2.57	2.74***	2.83***
	[0.05]	[0.06]	[0.07]	[0.11]
Share of households cultivating maize only	0.03*	0.02	0.01	0
	[0.01]	[0.01]	[0.01]	[0.00]
Share of households growing cereals	0.94***	1.00***	1.00***	0.99
	[0.01]	[0.00]	[0.00]	[0.01]
Share of households growing vegetables	0.05	0.06	0.07	0.1
	[0.01]	[0.01]	[0.02]	[0.03]
Share of households growing legumes	0.70***	0.78**	0.71	0.81*
	[0.02]	[0.02]	[0.03]	[0.04]
Area of cereals(ha)	1.03***	1.83	2.84***	2.83***
	[0.03]	[0.06]	[0.12]	[0.24]
Area of vegetables(ha)	0.02**	0.03	0.04*	0.04
	[0.00]	[0.01]	[0.01]	[0.01]
Area of legumes(ha)	0.54***	0.95***	1.10***	1.23***
	[0.03]	[0.04]	[0.07]	[0.12]
Production of cereals(kg)	419.96***	1206.48*	2903.90***	3193.95***
	[16.62]	[40.14]	[143.48]	[329.45]
Production of vegetables(kg)	18.34	26.97	39.83	109.33***
	[7.66]	[10.61]	[18.04]	[57.91]
Production of legumes(kg)	159.76***	459.34	763.95***	1174.61***
	[9.60]	[30.14]	[57.56]	[143.40]
Yield of cereals(kg/ha)	493.86***	776.07*	1041.00***	1099.75***
	[16.53]	[23.03]	[35.88]	[85.06]
Yield of vegetables(kg/ha)	1672.98	1817.92	2332.62	3810.63*
	[500.64]	[624.97]	[811.48]	[1497.51]
Yield of legumes(kg/ha)	335.08***	516.26	709.83***	832.59***
	[14.56]	[23.98]	[45.87]	[79.05]
TLU small ruminants	0.62***	0.86	1.11***	1.30***
	[0.03]	[0.04]	[0.07]	[0.11]
TLU big ruminants	0.48***	0.91	2.43***	2.33***
	[0.07]	[0.14]	[0.52]	[0.62]
TLU poultry	0.10***	0.14***	0.15***	0.18***
	[0.00]	[0.01]	[0.01]	[0.02]
Kg fertilizer used	153.87***	303.09	492.26***	661.47***
	[12.07]	[15.77]	[31.57]	[74.50]
N. of observations	527	424	238	90

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6b: distribution of characteristics by type in the economics domain

	Type 1	Type 2	Type 3	Type 4
Economic Domain				
Value of fertilizer used (GHC)	137.18***	307.85	438.72***	709.80***
	[8.29]	[24.15]	[29.77]	[91.65]
Value of traditional seeds purchased (GHC)	17.63*	19.58	26.43*	32.37**
	[1.89]	[2.49]	[4.98]	[8.32]
Value of improved seed purchased (GHC)	1.70***	3.67	9.11***	17.61***
	[0.47]	[0.92]	[2.40]	[7.59]
Value of pesticides used (GHC)	24.99***	51.06	96.93***	152.64***
	[2.02]	[3.88]	[9.64]	[23.66]
Share of households using communal labor	0.55	0.58	0.58	0.74***
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households using hired labor	0.48***	0.53	0.58*	0.64**
	[0.02]	[0.02]	[0.03]	[0.05]
Total person-days used, male & female	158.70***	249.60**	310.62***	430.56***
	[4.19]	[6.53]	[10.55]	[21.59]
Total Kg of grains harvested	473.33***	1778.01***	5028.31***	16036.51***
	[13.83]	[32.26]	[122.95]	[758.52]
Total Kg harvest used for own consumption	286.83***	843.49	1597.91***	2295.34***
	[11.96]	[28.92]	[65.92]	[150.82]
Total Kg harvest sold	99.93***	525.98***	1560.51***	2445.04***
	[14.81]	[33.01]	[86.36]	[264.22]
Agricultural wealth index	-0.45***	0.07	0.51***	1.05***
	[0.03]	[0.04]	[0.07]	[0.17]
Non-agricultural wealth index	-0.25***	0.04	0.29***	0.55***
	[0.03]	[0.05]	[0.08]	[0.16]
Share of households with good floor in dwelling	0.87	0.89	0.88	0.93
	[0.01]	[0.02]	[0.02]	[0.03]
Share of households with good source of drinking water	0.11***	0.18	0.21**	0.23*
	[0.01]	[0.02]	[0.03]	[0.04]
Share of households with good toilet facility	0.06***	0.07	0.14***	0.11
	[0.01]	[0.01]	[0.02]	[0.03]
Share of households with good source of lighting	0.37	0.38	0.34	0.37
	[0.02]	[0.02]	[0.03]	[0.05]
N. of observations	527	424	238	90

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6c: distribution of characteristics by type in the social domain

	Type 1	Type 2	Type 3	Type 4
Social Domain				
Share of HH with female having shared plot responsibility	0.36***	0.20***	0.16***	0.25
	[0.02]	[0.02]	[0.02]	[0.04]
Share of HH with female having exclusive plot responsibility	0.25***	0.08***	0.05***	0.04***
	[0.02]	[0.01]	[0.01]	[0.02]
Share of HH with female having shared livestock responsibility	0.31***	0.19***	0.13***	0.23
	[0.02]	[0.02]	[0.02]	[0.04]
Share of HH with female having exclusive livestock responsibility	0.19***	0.06***	0.03***	0.03***
	[0.02]	[0.01]	[0.01]	[0.02]
*Gender wage gap (wage women/wage men*100)*	87.36	85.93	85.03	78.21*
	[2.03]	[1.80]	[3.12]	[5.70]
N. of observations	527	424	238	90

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6d: distribution of characteristics by type in the human domain

	Type 1	Type 2	Type 3	Type 4
Human Domain				
Household size	6.82***	8.87*	10.52***	11.76***
	[0.15]	[0.25]	[0.36]	[0.72]
Share of married heads	0.82***	0.93***	0.97***	0.94
	[0.02]	[0.01]	[0.01]	[0.02]
Share of female heads	0.27***	0.10***	0.07***	0.01***
	[0.02]	[0.01]	[0.02]	[0.01]
Age of the head	47.89	46.67*	48.89	48.63
	[0.65]	[0.66]	[0.98]	[1.38]
Years of education of the heads	2.42**	2.17	1.50**	1.30*
	[0.20]	[0.20]	[0.23]	[0.39]
Share of literate heads	0.21**	0.18	0.15	0.13
	[0.02]	[0.02]	[0.02]	[0.04]
Mean years of education in the household	3.28***	2.96	2.74*	2.44**
	[0.12]	[0.13]	[0.16]	[0.22]
Max years of education in the household	7.55	7.36	7.74	7.78
	[0.22]	[0.25]	[0.34]	[0.55]
Average age of adults in the household	38.47***	36.59*	35.71***	36.31
	[0.40]	[0.36]	[0.44]	[0.76]
Number of children in the household	1.09***	1.64*	2.05***	2.29***
	[0.05]	[0.08]	[0.11]	[0.23]
Young dependency ratio	0.90***	1.02**	0.99	0.97
	[0.03]	[0.03]	[0.04]	[0.06]
Old dependency ratio	0.18***	0.14	0.12**	0.11
	[0.01]	[0.01]	[0.01]	[0.02]
Share of HH worrying about food shortages	0.40***	0.29**	0.27**	0.27
	[0.02]	[0.02]	[0.03]	[0.05]
Months experiencing food shortages?	2.02***	1.48	0.96***	0.94***
	[0.11]	[0.10]	[0.11]	[0.14]
N. of observations	527	424	238	90

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6e: distribution of characteristics by type in the environment domain

	Type 1	Type 2	Type 3	Type 4
Environmental Domain				
Share of households practicing irrigation	0.03	0.03	0.03	0.04
	[0.01]	[0.01]	[0.01]	[0.02]
Share of households practicing rotation	0.59***	0.66	0.70*	0.82***
	[0.02]	[0.02]	[0.03]	[0.04]
Share of households practicing fallowing	0.19	0.17	0.19	0.36***
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households practicing alternative tillage	0.04	0.03	0.03	0.03
	[0.01]	[0.01]	[0.01]	[0.02]
Share of households using manure on (any) plot in either season	0.29***	0.20**	0.21	0.26
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households using urea on (any) plot in either season	0.02***	0.05	0.07**	0.09**
	[0.01]	[0.01]	[0.02]	[0.03]
Share of households affected by soil erosion	0.24	0.2	0.24	0.24
	[0.02]	[0.02]	[0.03]	[0.05]
Share of households with soil erosion but no erosion control measu	0.08	0.05*	0.08	0.04
	[0.01]	[0.01]	[0.02]	[0.02]
Average share of parcels with clay or loam soil	0.44***	0.49	0.49	0.53
	[0.02]	[0.02]	[0.02]	[0.04]
Average share of parcels with black or brown soil	0.59***	0.55	0.48***	0.55
	[0.02]	[0.02]	[0.03]	[0.04]
Average share of parcels with incrusted soil	0.13	0.12	0.18***	0.08**
	[0.01]	[0.01]	[0.02]	[0.02]
Number of trees owned on the land	24.81***	38.17	50.35***	69.64***
	[1.09]	[1.58]	[2.78]	[5.70]
N. of observations	527	424	238	90

Note: The stars represent significance levels of mean difference tests between the type under consideration and the other three types combined.* significant at 10%; ** significant at 5%; *** significant at 1%

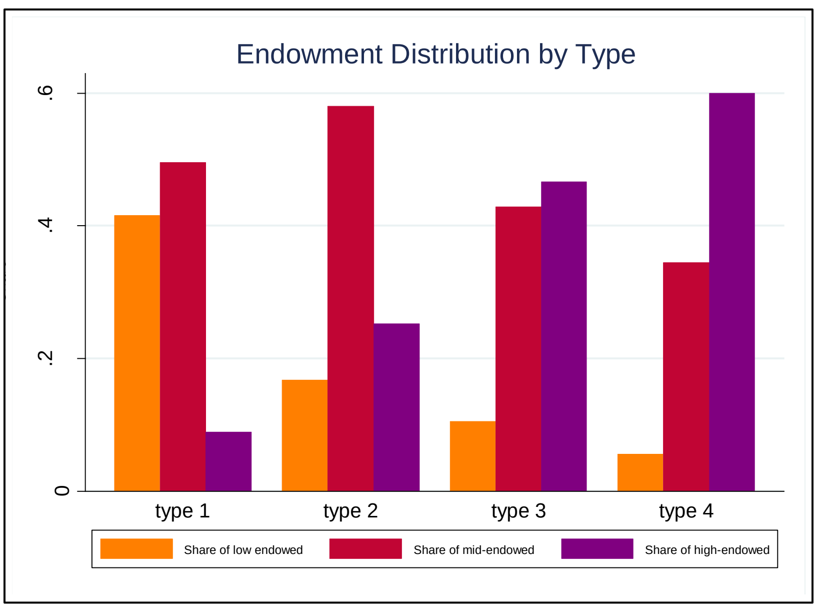


Figure 8: Level of Endowments by Type

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 households with low to medium levels of endowments

- High number of female headed households, with heads less likely to be married and with high education attainments and literacy rates. This also comports higher level of gender equality with respect to the other types.
- High food insecurity.
- Relatively smaller household size and high elderly concentration.
- Little asset ownership (small land, little livestock, low agricultural and non-agricultural wealth).
- 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).
- Generally bad soil quality and little conservation practices in place.

Type 2: Young medium-endowed households

- Relatively small and young households with high share of children and relatively high levels of education.
- Good performance in gender equality measures.
- Little assets and low productivity, even though better than type 1.
- More likely to grow vegetables than other groups.
- Fairly good soil quality but little conservation practices in place.

Type 3: Medium to highly endowed households breeding cattle

- Large households with young, married heads and many children.
- Bad performance in gender equality measures.
- More likely to do intercropping than other groups.
- Medium to high levels of crop production and productivity coupled with high input use.
- High ownership of cattle.
- Serious problems of incrusted soil as well as soil erosion against which no measures are taken. Some conservation practices in place.

Type 4: High yield households with high endowments

- Very large male headed households composed mostly by active population (15-65 y.o.).
- Low levels of education and head literacy rates, low levels of gender equality except for the wage gap measure.
- Higher levels of food security with respect to other groups.
- Extremely high asset ownership (large land, number of livestock units, agriculture and non-agriculture index).
- High production and productivity of crops with high input use (fertilizer, pesticides, and improved seeds).

- Good soil conditions and conservation practices.

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 households with low to medium levels of endowments	Low crop production and productivity. Little livestock owned.	Low wealth (agri and non-agri), low input expenditure, small quantities of harvest going to sales and own consumption	Low levels of soil-preservation practices and bad soil quality.	High gender equality.	Female heads, high share of old members. High level of education and literacy. Low food security.
Type 2: Young medium-endowed households	Low crop production and productivity. Little livestock owned. Vegetable growers.	Low wealth (agri and non-agri), low input expenditure, small quantities of harvest going to sales and own consumption	Fairly good soil conditions but little conservation practices in place.	High gender equality.	Small households with many children. High level of education and literacy. Low food security.
Type 3: Medium to highly endowed households breeding cattle	High crop production and productivity. Frequent intercropping. High ownership of cattle.	High wealth (agri and non-agri), high input expenditure, large quantities of harvest going to sales and own consumption.	High percentage of incrusted soils and generally bad soil conditions. Some conservation practices in place.	Low gender equality.	Large households with married heads and many children. High food security.
Type 4: High yield households with high endowments	Very high crop production and productivity. Overall high livestock ownership of all kinds.	Very high wealth (agri and non-agri), high input expenditure, large quantities of harvest going to sales and own consumption	Fairly good soil conditions and conservation practices in place.	Low gender equality.	Very large households with male heads and large share of active members. Very low education and literacy rates. High food security.

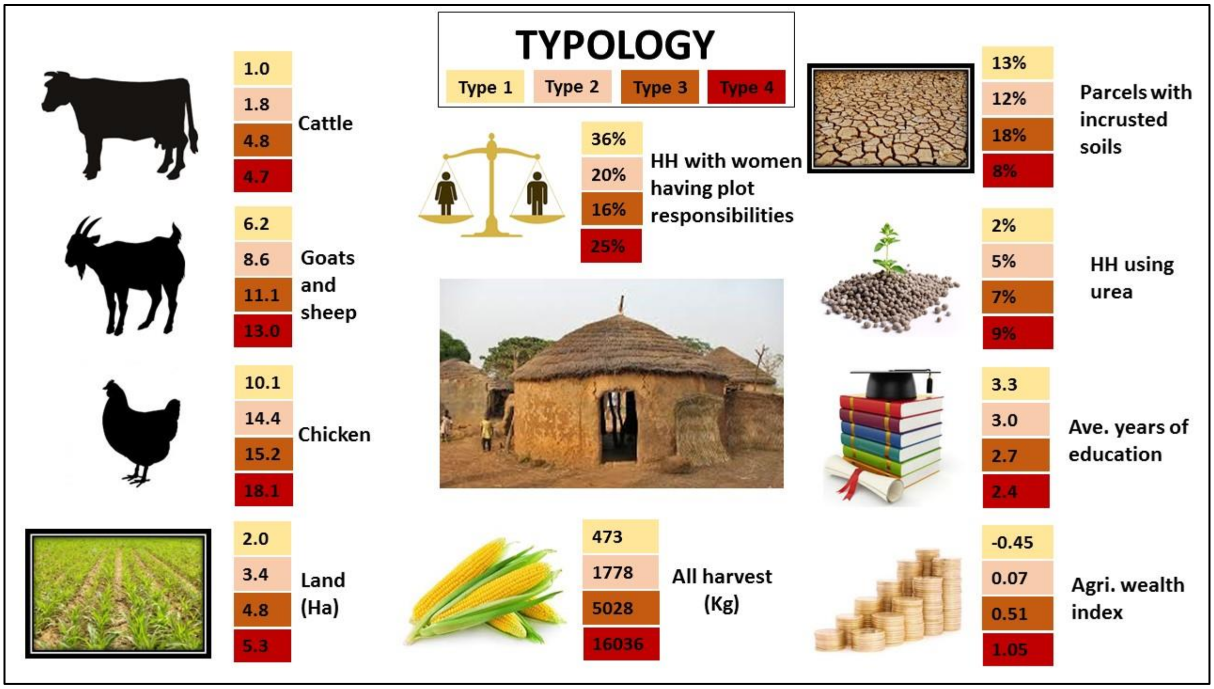


Figure 9: Graphic representation of types

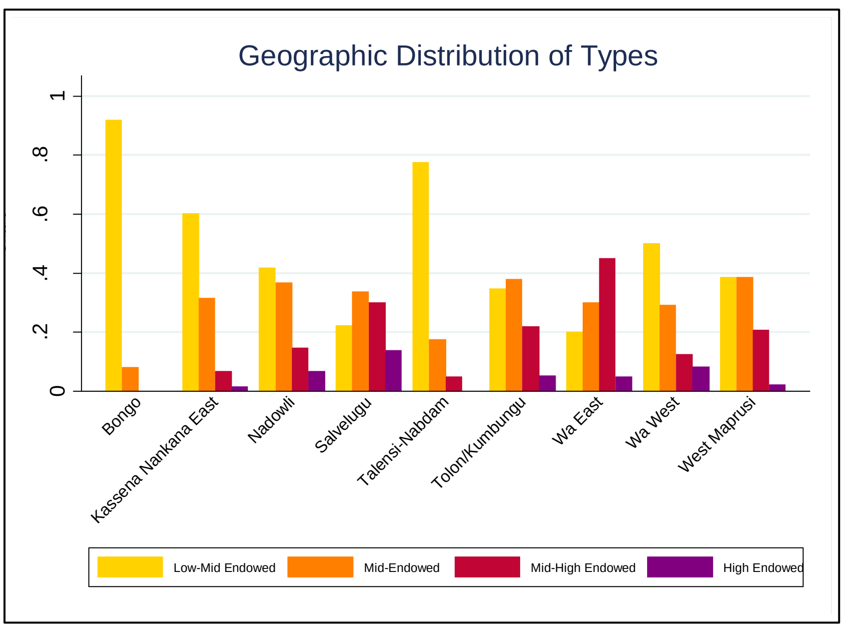


Figure 10: Distribution of Typologies by Districts

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 Bongo and Talensi-Nabdam there is an over representation of female-headed households with low to middle levels of endowments (type 1), Wa East and Salvelugu concentrate the highest shares of Mid-High endowed and high endowed households (type 3 and 4). Similar differences appear when we look at the regional typology distribution (Figure 11), with the Northern region hosting a large portion of richer types while the Upper East 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.

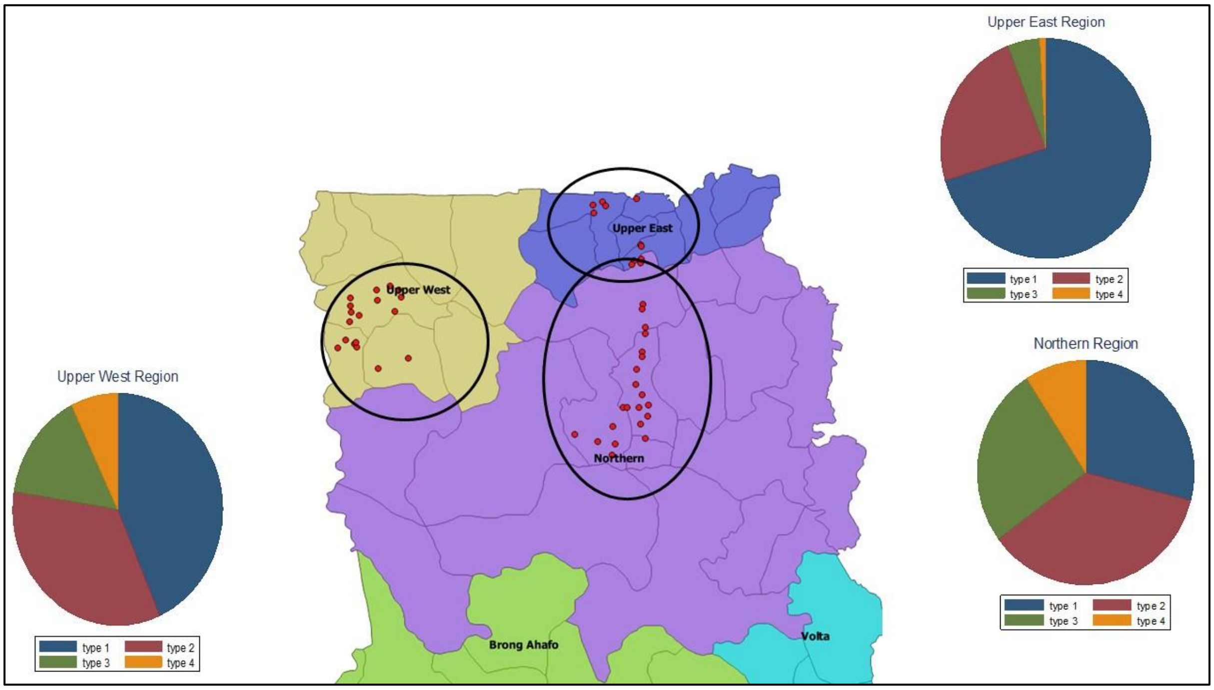


Figure 11: Distribution of Typologies by Regions

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:

- Types 3 and 4 – the medium to high endowed farmers – perform the highest in the productivity and economic domains but the opposite is true for the social and human domains, where especially type 4 is lagging behind (except for the food security measures, which are not included in this representation).
- Despite being much less wealthy, types 1 and 2 – the low to medium endowed farmers – have higher average levels of education and present higher gender equality. This observation may indicate the existence of low demand for human capital in the current Ghanaian agricultural system. This structural issue should be addressed in order to achieve long-term development.
- Finally, only the very wealthy farmers seem to engage in preserving the natural resource base. Type 3, despite the relatively high endowments levels and the favorable climatic conditions, present serious problems of soil erosions coupled with little action to improve the situation. This poses a big threat in terms of sustainability and thus calls for a large effort to spread awareness on the importance of maintaining a fertile soil.

Recommendations:

- Types 1 and 2 constitute the majority of AR farmers, they possess higher levels of education and women have access to some resources and responsibilities. AR should support the diffusion of affordable productive agricultural technologies in order to help these households increase their productivity and economic status. In addition, trainings can be delivered to sensitize the farmers about the importance of preserving the natural resource base.
- Type 3 presents' rather high levels of productivity and economic endowments but has the worst environmental conditions. The AR project can focus on training these type of farmers on how to better restore and preserve good soil quality.
- Type 4 includes a minority of well-off farmers that also benefit from favorable soil conditions. They severely lack human capital endowments, especially in terms of education, but AR has limited scope to directly act in this area. These farmers can be involved in the demonstration and diffusion of technologies with their neighbors in the village, since their productive capacity can serve as a good example.
- Women's inclusion is generally quite poor all across the groups, with the exception of type 1. Therefore, the AR gender specialists should be closely involved to design interventions able to directly improve gender equality.

The appendix includes additional graphs characterizing the obtained typologies.

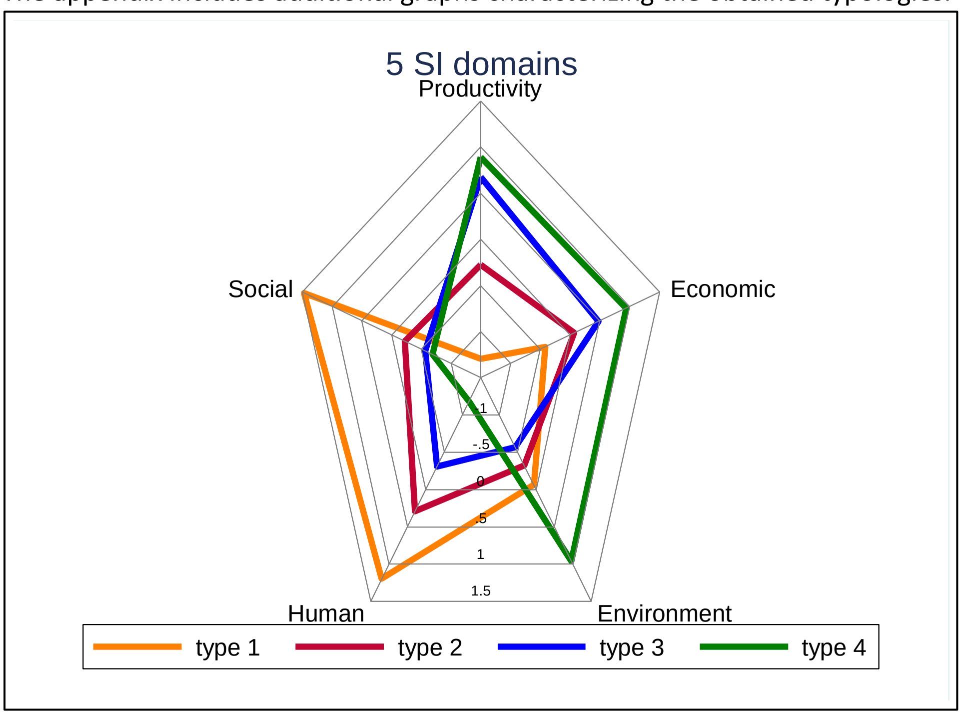


Figure 12: Typologies performance by sustainability domain

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); exclusive female responsibility in managing certain plots (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

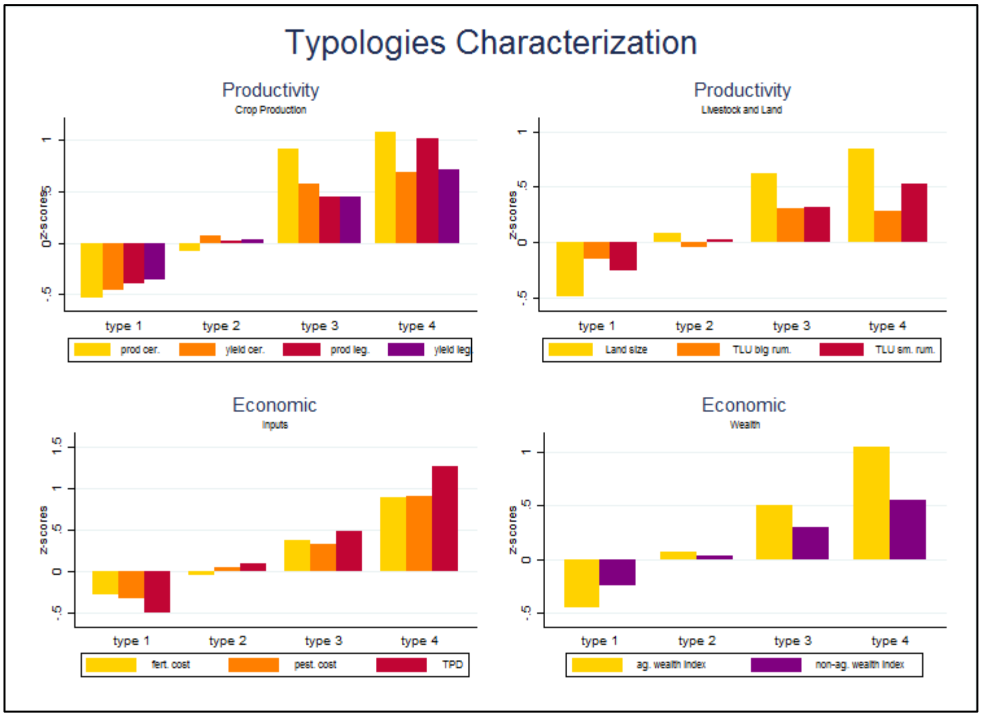


Figure A1: Typologies by domain (productivity and economic)

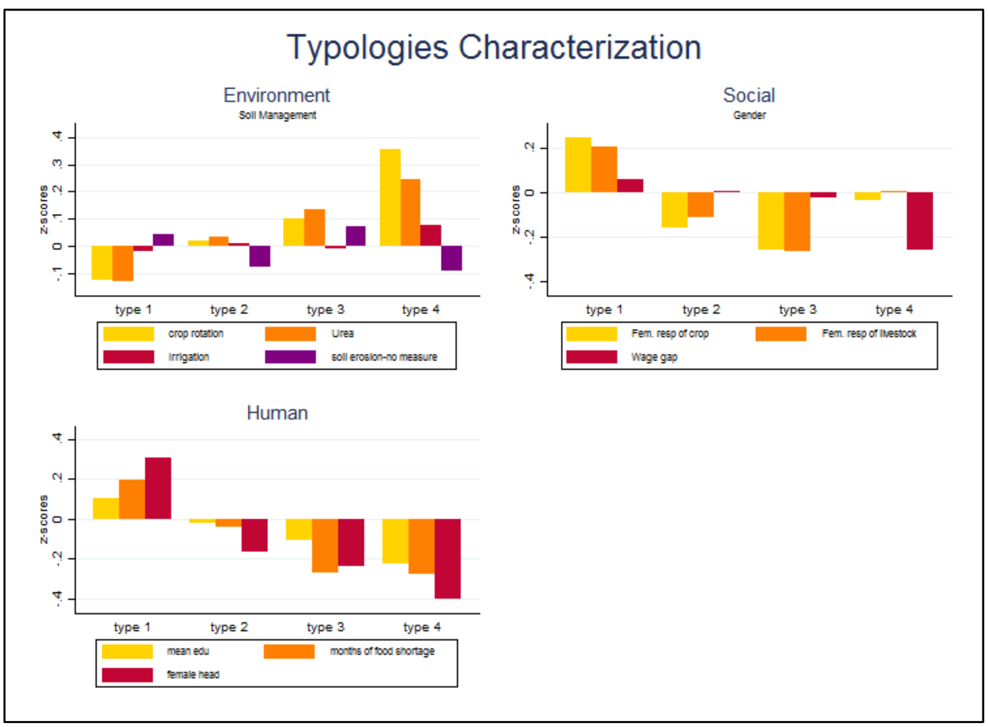


Figure A2: Typologies by domain (environment, social and human)

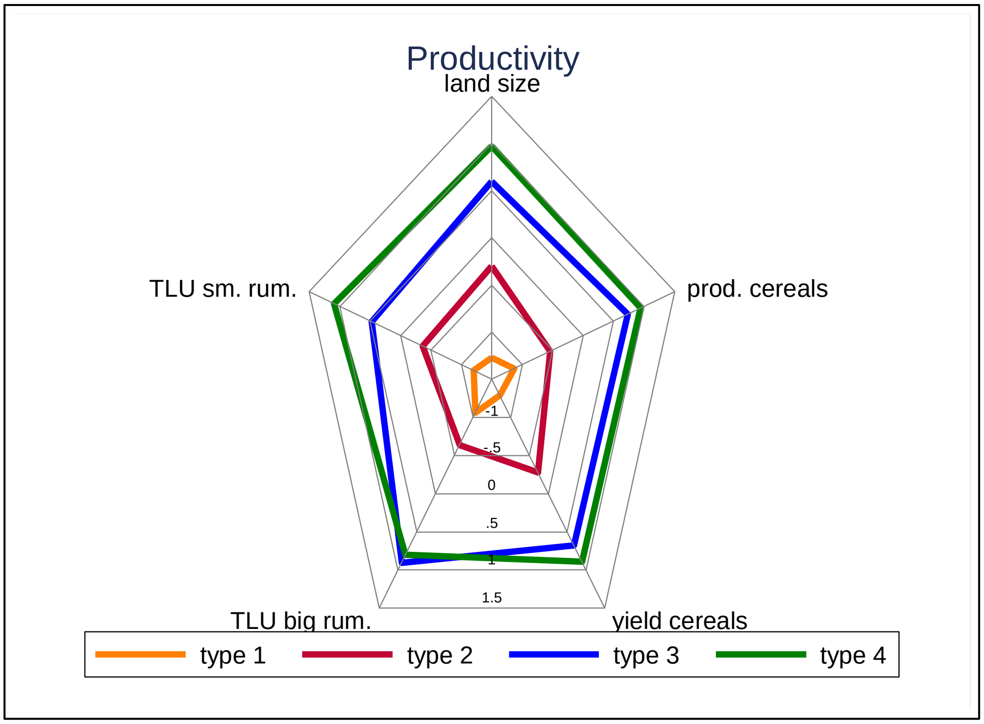


Figure A3: Radar graph – productivity (z-scores)

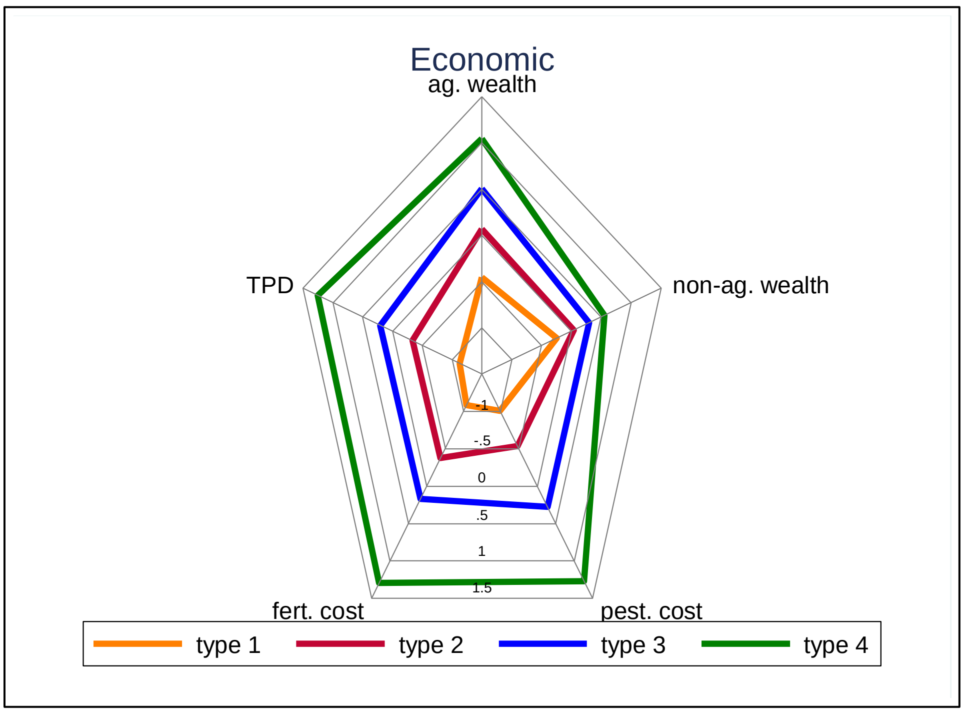


Figure A4: Radar graph – economic (z-scores)

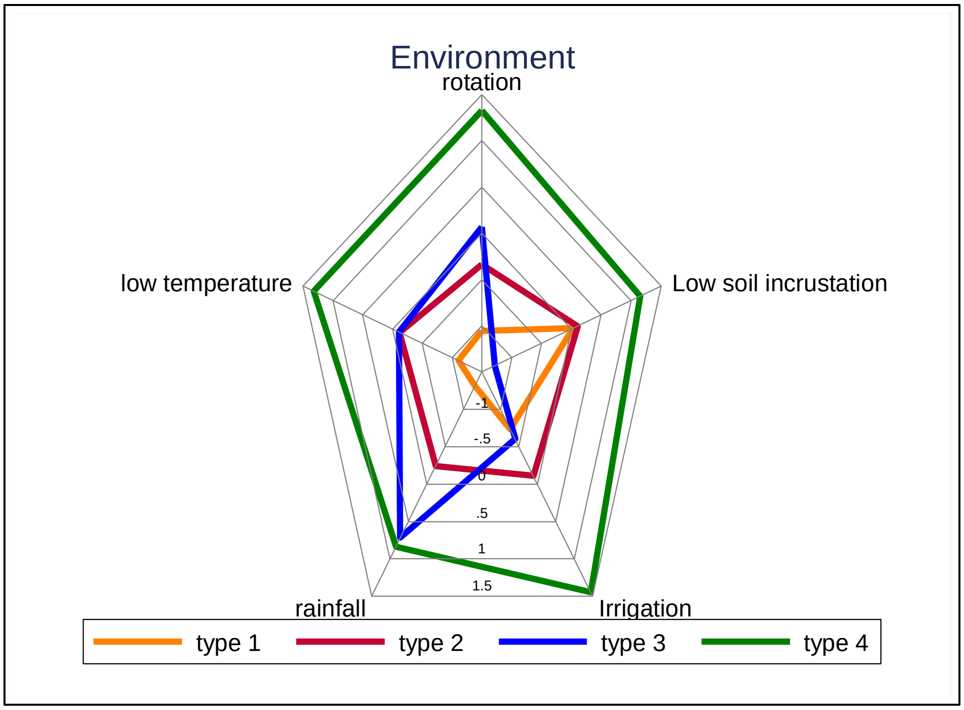


Figure A5: Radar graph – environment (z-scores)

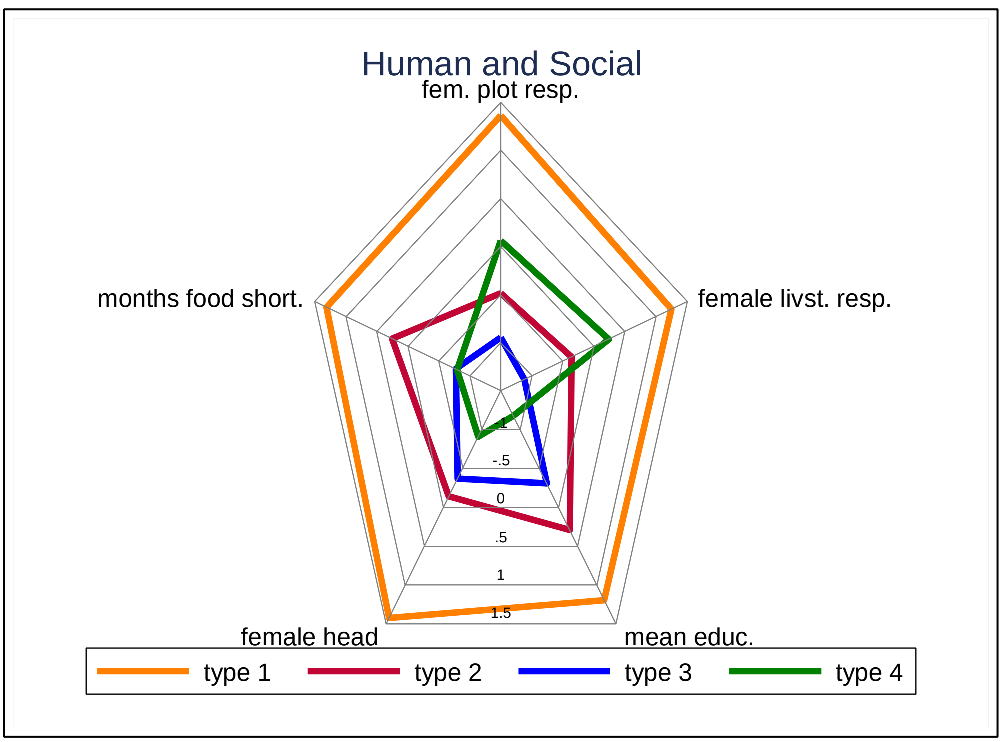


Figure A6: Radar graph – social and human (z-scores)