## Food security and better livelihoods

## for rural dryland communities

## Farming household types and their characterization in complex crop-livestock smallholder agricultural systems for contextual analysis and extension intervention: case of Riviridzi Catchment in Ntcheu

### December 2015

Powell Mponelaa, Lulseged Tamenea\*, Gift Ndengua, Bao Leb

a International Centre for tropical Agriculture, Lilongwe, Malawi

b CGIAR Research Program on Dryland Systems, Amman, Jordan

[\*lt.desta@cgiar.org](mailto:*lt.desta@cgiar.org)

**Abstract**

Typology of farming units (households and farms) is deemed essential for targeted research and development programs. We employ the sustainable livelihood framework to collect integrated dataset for household, landscape and infrastructural attributes to quantitatively group farming units into plausible types. Interestingly it is noted that small scale farmers in the maize mixed farming system are heterogeneous and could be potentially grouped using principal component analysis (PCA) and K-mean cluster analysis (CA) into 3 classes. Income was the variable with the most discriminating power that significantly distinguished the classes into plausible types. Variables with high discriminating power between types I and II include family labour, transport facilities, household and farm equipment and tropical livestock units per person. Household types I and III differ significantly in terms of age and level of education of the household head. The types II and III are significantly distinguished only by income levels. The types identified are homogenous within a range of attribute values which can be used for technology targeting, extrapolation domain for supporting out-scaling of impacts and used for system modelling that copes with socio-ecological diversity.

Keywords: *Typology, socio-ecological, cluster analysis, system modelling, extrapolation domain.*

# Acknowledgements and disclaimer

This work was undertaken by the International Center for Tropical Agriculture (CIAT), which forms part of, and funded by, the CGIAR Research Program on Dryland Systems. This paper has not gone through CIATs standard peer review procedure. The views expressed in this document cannot be taken to reflect the official opinions of CGIAR.

# Introduction

Characterizing farm types is increasingly becoming fundamental in designing type specific interventions in order to address the diverse challenges faced by small-scale farmers (McConnell and Dillon, 1997; Bellec *et al.*, 2011; Garrity et al 2012; Le *et al*., submitted). However, efforts are hampered by lack of differentiated data about the real situation of households and their farms, complicating development of decision support system for agricultural production planning (Riveiro-Valiño et al. 2009). Development of farming system decision tools requires an understanding of both proximate and underlying determinants of farmers’ motivation to land use choice, farm transition alternatives, and farming practices. As highlighted by proponents of precision agriculture, the multiple causal relationships of agricultural production inputs and outputs are farm/household context specific (Cassman, 1999, McBratney *et al*., 2005). More challenging is that small-scale farming conditions are varied and complex to apply the whole farm approach. Farmers generally react according to production rules (natural resource extraction) but also optimises (e.g. land for cultivation). With different satisfaction and aspirations, the behaviour patterns and feedback loops differ (Le *et al*., 2012). There is thus a possibility of clustering households with similar attributes. This paper develops farm household types based on observable/proxy characteristics of the households and their environment. Ultimately, these typical farm household types would be used later as basis to develop representative decision and production functions.

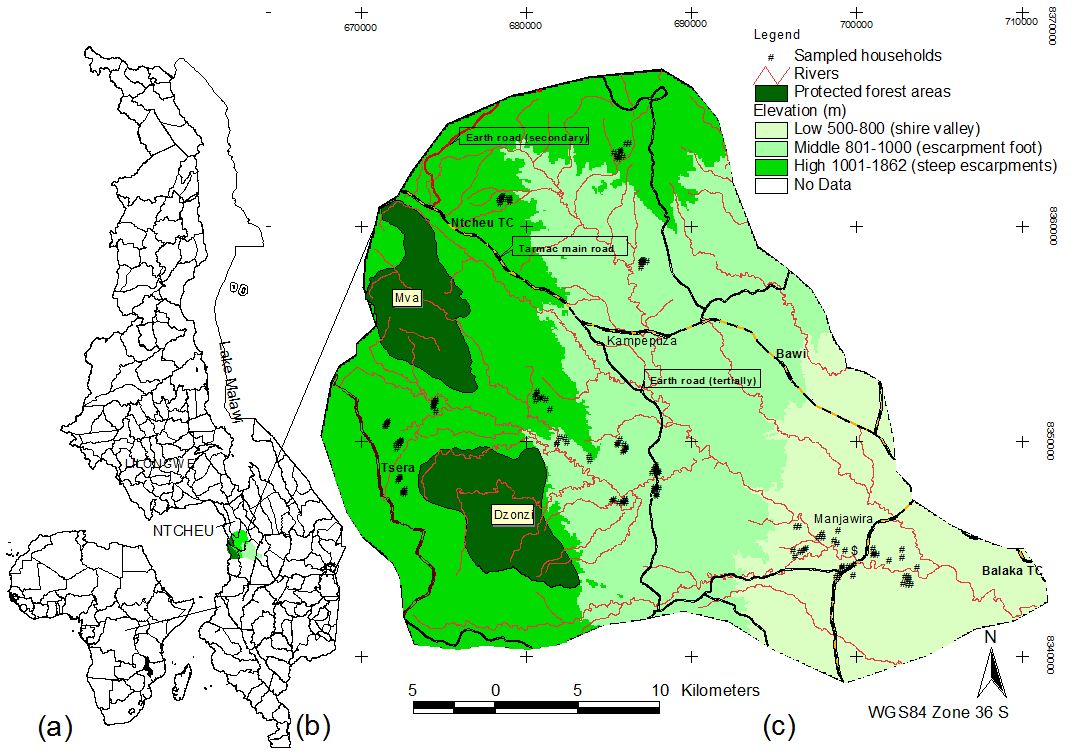
To enable better targeting of technologies, global and national stratifications of agricultural systems into agroecological zones (Fischer *et al*., 2012), farming systems (Dixon *et al*., 2001) have been conducted. At global and continental scale, farming systems classifications are based on main crops (e.g., Dixon *et al*. 2001). At country level, farming systems have been categorised based on crop suitability into agroecological zones (FAO, 1996) and rural masses grouped according to living conditions combining biophysical and socio-economic determinants into livelihood zones (Grillo and Holt, 2009). In addition, farmers of different resource endowment, education level, age, livelihood sources, access to information and market have different decision making abilities and their level of technology adoption also vary significantly (Ronner and Giller, 2012). The mechanisms by which they cope to risks and their ability to bounce back from stresses also differ (Darnhofer, 2014).

However, advances in development and deployment of context specific technologies have focused more on crops of economic value not taking into account the heterogeneity in landscapes and communities (Snapp *et al*. 2003; Rware *et al*. 2014). Without detailed information related to site- and context-specific, planners and decision makers when faced with farm/household heterogeneity resort to the so called blanket recommendations that have yielded skewed impacts (Snapp *et al*. 2003). These call for explicit examination of the combined role of households’ characteristics, farm and neighbourhood biophysical attributes and linked external socio-institutional factors on farmers’ behaviour. The overall aim of the study is thus to understand the heterogeneity among farming households by classifying them using household, plot and ecological variables into types. The study will further support type specific analysis of drivers and livelihood and sustainability outcomes for targeting agricultural intensification interventions.

# Methodology

## The study site and sampling

The Riviridzi River forms the major tributary of the shire basin which drains into the Zambezi River. The river flows 84 km from the mountain near the border with Mozambique to the Shire River (Fig. 1). The catchment supports agriculture and ecosystem functions to the communities and nearby towns. It is also a source of sediment, thereby posing negative externality downstream especially to hydropower generation as rivers dry and siltation/sedimentation is increasingly becoming expensive to dredge (Chimtengo *et al*., 2014).



**Figure 1** The location of sampled households with respect to elevation and main roads within Riviridzi landscape (c) in central Malawi (b) and southern Africa (a)

Topographically, the landscape can be stratified into three terrain classes: high, middle and lower (Figure 1). An expert interview with extension officers at Nsipe extension planning area (EPA) revealed that different crops are grown at high altitude verses those grown at low altitude. The higher elevation is generally characterized by altitude above 1000 m asl and steep slopes along with shallow stream valleys. Most of it has fertile hills with deep reddish soils which are used for growing maize intercropped with common beans and/or in main rotation with Irish potato while stream valleys are used for vegetable gardens especially during dry season. The middle altitude zone (Fig. 1) is a transition zone with moderate slopes. Main crops grown include sole maize in rotation with ground nuts and to a smaller extent with tobacco. The stream valleys especially those in the area west of Ntcheu are used for vegetable gardening and sweet potato production during the dry season. The low altitude is part of the flat plain adjoining the Shire River (Fig. 1). In these areas, temperatures are relatively hotter and rainfall lower than that of the higher altitude areas. Main crops grown include maize intercropped with pigeon peas and cowpeas in rotation with cotton. The livestock populations (small ruminants) increases with decreasing altitude becoming more common in semi-arid areas of shire valley (FAO, 2005). In the study area, maize production is the main livelihood strategy. Dixon’s map shows that maize mixed farming system is the most dominant for east and southern Africa and it stretches from Ethiopia to South Africa. As of 2001, the system supported 100 million people, of which 60% made up the agricultural population and 75% of the population lived in rural areas. The system had one of the highest populations of stunted children (more than 6 million) and of people living below the two dollars a day poverty line (about 70 million). Within the maize based farming systems, some farmers intercrop the cereal with legumes. Such localized differences within the main system differentiates farmers’ behaviour in usage of production enhancing technologies as well as different crop combinations.

In addition, households supplement their necessities by collecting natural resources and engaging in *ganyu,* which is paid casual labour. Fuelwood obtained from woodlands, forests and croplands is the main source of energy for cooking even for the growing urban populations. Charcoal from the area is transported to Lilongwe and the surrounding towns to satisfy the increasing demand in the peri-urban areas. Off-farm livelihood sources are adaptive strategies that most smallholder farmers resort to when food production is low.

The study employed stratified random sampling to arrive at the households to be surveyed. Determination of sample size was based on nature of the landscape and expert interview regarding cropping patterns. Based on the above mentioned three agro-ecological strata, the sampling was stratified such that data can be collected from the upper, middle and lower elevation terrain classes. From each terrain class, 4 villages were randomly sampled. This gives a hierarchical subject structure with households and their landholdings at lower level, constituting the village population and landscapes, and the entire study area is at the highest aggregate catchment level (**Table 1**). Sampling of households within villages was not screened by stratification, hence the survey was dynamic and allowed for a plausible understanding of household’s strategies (Tittonell *et al*., 2005).

**Table 1**. Sampling strategy for household data collection in the Riviridzi landscape, Ntcheu, Malawi showing number of villages….

[We vave got the list of villages with number of households in 2013/14 season, without coordinates. Being geo-referencing with existing shapefile of villages.]

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Terrain class | Upper sloppy | Upper dissected | Middle (dam + AGORA) | Transition (middle and lower) |  |
| Number of villages\* | X | X | X | X | x |
| Number of total households | X | X | X | X | x |
| Villages sampled | 3 | 4 | 5 | 4 | 16 |
| Households sampled | 30 | 40 | 50 | 40 | 153 |

\*excluding Ntcheu town

## Conceptual and analytical approaches

The livelihood strategies undertaken by farming household in the maize mixed farming system are hypothesised to be dictated by production factors including the livelihood resources and institutional processes that influence how resources can be used to realize different household objectives. Understanding the impact of the different drivers on different household categories can help design context-specific interventions. Several approaches are used to study diversity and differentiating factors of farming households and farms. Some examples include, the use of land size as a proxy for household types (Chulu and Wobst 2001; Peters 2007), gender as proxy for income level (Brown *et al.*, 1996; Peters 2007), geographic region as determinant of agricultural opportunities (Simler 1994; Dorward 2002), combination of agro-ecological and socioeconomic characteristics for determination of livelihood in terms of food access and coping strategies (MVAC, 2005; Dorward 2002; Douillet and Toulon 2014) and anthropological panel analysis of household income (Peters 2006). However, these approaches fail to address the challenge of capturing the real situation of farms and households especially the spatial and temporal dimensions. This study considers maize-mixed farming community to be a typical social-ecological system and uses two frameworks to choose variables and methodological approach (Binder *et al*. 2013). The Sustainable Livelihood Framework (SLF) by Sconnes (1998) was used to identify variables for the five production factors that define farmers’ decisions and abilities to undertake practices leading to a livelihood strategy. The Framework of Social-Ecological Systems (FSES) by Ostrom (2009) was used to draw relations among factors in the social and ecological domains that yield functional types (Binder *et al*., 2013).

Farming household functional types are distinguished by *state variables* that characterise households and their livelihood supporting landscapes at a given time (Polhill *et al*., 2008) (in our case during the survey period), but also are related to each other through coupled interactions defining the behaviour of the system (Railsback and Grimm, 2009). In smallholder agricultural systems, diversified behavioural portfolios (Shefrin and Statman, 2000) affect the pursuit of different livelihood strategies by households and dictates sustainability of the ecosystem. Understanding livelihood strategies, therefore, requires attention to the interactions among household and landscape *state* variables and the ways in which they may be clustered, sequenced or substituted to enable different livelihood production strategies (Scoones, 1998). As individual households pursue their goals, they are bounded by available assets and can be classified into distinct categories. One of the approaches commonly used to do so is through building typology (Le *et al.* 2012). Household types emerge as household tend to categorise itself into the most similar type, based on comparing and ranking dissimilarities in state variables of itself and its environment, with those of its neighbours.

Principal components analysis (PCA) was used to empirically identify the principal factors differentiating the household types. Using principal components (PCs) with eigenvalues over 1.0 enabled reduction of a larger number of initial variables into a smaller set without losing important defining information (Campbell *et al*. 2001; Le *et al*. 2005; Thiombiano *et al*. 2015.). The PCs are optimal linear combinations of initial variables explaining the variance in descending order. Because all of the variables have some degree of association with all the factors there is generally a problem of factor indeterminacy. To address this and achieve a simple structure, where important variables have high loading of > 0.3 on single PC and lower loadings < 0.3 on all others, we used the Varimax orthogonal rotation and Kaiser Normalisation (Williams *et al*., 2010). The factor loadings/weights signifies the importance of the variable for a particular component and the ones with highest loadings of > 0.6 were identified (Adimassu *et al*. 2012). Since PCs are independent, variables with highest loading on each PCs were used in subsequent cluster analyses thereby addressing the problem of multi-collinearity (Naes and Mevik, 2001).

Subsequent determination of household types was done using the K-mean cluster analysis (K-CA) (Lesschen *et al*., 2005). K mean is the most commonly used clustering algorithm which was developed by Mac Queen in 1967 and it is the most effective for small data sets (Alnaji and Ashour, 2011). Basically, K-Means clustering is a partitioning method that treats observations of the data as objects based on locations and distance between various input data points. Partitioning the objects into mutually exclusive clusters (K) is done in such a way that objects within each cluster remain as close as possible to each other but as far as possible from objects in other clusters. Each cluster is characterized by its centre point i.e. *centroid* and each of the records is assigned to the nearest cluster centre. A centroid is the point whose coordinates are obtained by means of computing the average of each of the co-ordinates of the points of samples assigned to the clusters (Ghosh and Dubey, 2013). The K-means converges to one of many local minima because it minimises distance measure between each data and its nearest cluster centre thereby minimising the intra-cluster variances while maximising the intra-cluster distances (Rokach and Maimon, 2005; Alnaji and Ashour, 2011). K-CA maximises the sum of the squared error (SSE) by measuring the total squared Euclidean distance of observations from the cluster centroids.

With highly heterogeneous dataset (9 PCs with eigenvalues > 1), number of centroids could be large resulting in several clusters. Optimal number of clusters *k* was determined using the *knee* method, with optimal *k* value on the inflexion point/bend on the curve of sum of distances of clusters from the centroid against the number of clusters (Salvador and Chan, 2004).

The observed and measured values for the selected variables were normalised to fit the data within unity so that they can be drawn to the same axis and used to characterise household types. The scores were computed using the following formula (Mohamad and Usman, 2013). To determine whether the clusters are conceptually and statistically distinguishable, the differences in state variables were tested using least significant differences (LSDs) after analysis of variance using unbalanced structure (Sarstedt and Mooi, 2014)). The unbalanced structure run in GenStat ensures complete fit of unbalanced sample sizes between clusters (Payne, 2012). After distinguishing the clusters, we can also try to come up with a meaningful name or label for each cluster; that is, one which adequately reflects the objects in the cluster.

## Data sources, attributes and assumptions

As indicated in section 2.2, based on SLF, household and plot variable attributes of interest were grouped into 5 categories depicting the slow changing quantitative assets that define the livelihood structure and dynamics of the study population and its surrounding environment (**Table 2**). Selection of variables for typology was on the basis that they do not undergo rapid change/shift within the short to medium term (Adams and Adams, 2008). It is assumed that these variable attributes will have a strong influence on distinguishing the types which once formulated can be slow to change. This is crucial for types to be functionally different from one another and households/land parcels may be assigned to the closest centroid based on overall nature of the attributes rather than on a few unstable variables.

Social assets represents the knowledge, comping mechanisms as well as strength of community and family relationships upon which people draw when pursuing different livelihood strategies requiring co-ordinated actions. Considering that some ISFM interventions transcends the boundary of the farm/household and take long time to show positive results, they tend to be information intensive. Education and the age of the household head as a decision maker has been used as a proxy for differences in knowledge of the technical aspects of the technologies and ability to engage in community discussions (Lin, 1990; Mponela *et al*., 2011).

Human capital represents the potential of the household to carry out an activity. Owing to the labour intensive nature of some agricultural technologies, labour availability and constraint has been proxies in terms of number of able-bodied family members and labour constraints in terms of number of dependents to be fed (Sibanda, 2014).

**Table 2** Description of livelihood assets used in the PCA and CA

|  |  |  |
| --- | --- | --- |
| Assets | Attribute | Description |
| Social | Hage | Age of the household head (years) |
|  | Heduc | Level of formal education of household head (0 =never attended school; 1 = primary 1-8; 2 = secondary 9-12; 3 = post-secondary >= 13 |
| Human | Hlabour | Number of workers available in the household (potential worker age 18-65 = 1; while the children age 9-18 and old >65 = 0.5) |
|  | Hdepend | Dependency ratio= no. of dependants (children + old)/no. of potential workers active age group |
| Physical | Hcom | Information and communication index in terms of monetary value of phones, TV set and radios owned by the household (see text) |
|  | Htransp | Transport index in terms of monetary value of bicycles and motorcycles owned by the household (see text) |
|  | Hroad | Distance from house to the nearest all-weather main road (meters) |
|  | Hcenter | Distance from house to the nearest trading centre (meters) |
|  | HEquip | Farm equipment index owned by the household (see text) |
|  | Hlu | Average tropical livestock units owned by the household |
| Natural | Hnr | Value of the natural resources collected by the household in a year (MK) |
|  | Hland | Total land used or owned by the household during 2013/14 growing season (acres) |
|  | HlandP | Total holding divided by number of persons in a household (acres/person) |
|  | HPlot | Total land holding divided by number of fragmented plots (total holding acres/no. of plots) |
|  | HlandFC% | Share of land allocated to food crops (floating between 0-1) |
|  | HlandCC% | Share of land allocated to cash crops (floating between 0-1) |
|  | HlandNCA | Size of land uncultivated during 2013-14 growing season (acres) |
| Financial | HInc | Average household annual income per person (MK/person) |
|  | HIncLS% | Share of annual Income from sale of livestock and livestock products (floating between 0-1) |
|  | HIncOF% | Share of annual income from off own farm activities (ganyu, employment, trade and remittances) (floating between 0-1) |

Physical assets include infrastructure, production equipment and technologies. Assets are relatively slow changing variables as they are usually accumulated over time and last longer (Filmer and Pretchett, 1998). However, assigning values to assets is challenging as they lack comparability and mutual substitution (Moser and Felton, 2007). In most rural areas, the value of equipment and their relative contribution to household livelihoods are largely unknown making it more complex. We employ the intuitive approach to weight the assets using the monetary value at the current market price forming an index (Moser and Felton, 2007). The total monetary value of assets (index) is estimated by multiplying the number of items by the market prices of most common types during the survey period. A communication index was computed as *Hcom* = *radios \* 10 (000) + cellphones \* 9 + television \* 50*. With advance in telecommunication, mobile and radio based communications are becoming the most common modes of information and agricultural advice for farmers. Similarly transport index was computed as *Htrans = bicycles \* 1 (40,000) + motorcycles \* 10*. Important for accessing input and output markets, bicycle has been the most common mode but recently usage of motorcycles is increasing. Likewise farm equipment index was computed as *HEquip = motor-pumps \* 100 (000) + solar panels \* 150 + oxcarts \* 250 + sewing machines \* 70.* The geographic location of household with respect to main roads and trading centres for input and output markets were estimated in ArcView GIS. We used the grid layers for roads and trading centres to estimate distances within the study area and clipped the grid values for the house way points that were collected during the survey.

Natural resource stocks and environmental services from which resource flows and services useful for livelihoods are derived considered include livestock units, natural resources and land holding production orientation. Within a largely agrarian society, access to land for agriculture and what people do with that land are the most important natural attributes. As is the case in most farming systems, household production orientations do not vary in the short to medium term. As a result, income composition and land allocation to the main crops (for cash, food and fallow) have been used as proxies for production orientation and livelihood strategies. Natural resources collected, livestock units. The household’s reliance on natural resources collected from uncultivated areas was expressed in monetary value calculated using annual collection of different products multiplied by their respective average selling price during the surveyed period. The livestock which comprise mainly of poultry and goats but fewer pigs and cattle were converted into standard livestock units (LU) using nutritional and feed requirement factors for sub-Sahara Africa (Chilonda and Otte 2006, Wikipedia 2013).

Financial assets are the capital base in terms of cash, credit/debt, savings, and other which are essential for the pursuit of a livelihood strategy. Sell of selling livestock and livestock products is one of most coping strategies among households in times of risks and shocks (Sibanda, 2014). Thus households that sell livestock could be assumed to be resilient. Non-farm sources captured included formal employment and small trade, sell of *ganyu* labour and remittances from relatives.

# Results and discussion

## Descriptive statistics, PCA results and preliminary tests

Prior to running PCA, preliminary tests were done to ensure that minimum requirements are met. Tests for adequacy of sampling shows that the variables entered have high sampling adequacy with anti-image correlation and commonalities of more than 0.5 and the combined test using Kaiser-Meyer-Olkin Measure of 0.52 (Table 3). This validates the use of the variables in cluster analysis and that their combination gives a robust result with significant Bartlett's Test of Sphericity (P<0.001).

**Table 3** Characteristics of the sampled households and preliminary tests for cluster analysis in the Rivirivi watershed, Ntcheu, Malawi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Descriptive Statistics | | anti-image | communality |
|  | Mean | Std. Deviation | correlation | extracted |
| Hage | 46 | 15.16 | .590 | .546 |
| Heduc | 1 | 0.62 | .561 | .733 |
| Hlabour | 2.36 | 1.28 | .253 | .834 |
| Hdepend | 1.38 | 0.96 | .330 | .824 |
| Hcom | 13.63 | 17.90 | .764 | .737 |
| Htransp | 0.70 | 1.47 | .683 | .814 |
| Hroad | 6627 | 4006.42 | .536 | .948 |
| Hcenter | 7625 | 3476.08 | .537 | .953 |
| HEquip | 18.19 | 66.85 | .704 | .729 |
| Hlu | 0.09 | 0.16 | .588 | .736 |
| Hnr | 17058.62 | 25044.93 | .461 | .809 |
| Hland | 1.15 | 0.83 | .548 | .873 |
| HlandP | 0.29 | 0.35 | .517 | .829 |
| HPlot | 0.60 | 0.57 | .624 | .716 |
| HlandFC% | 0.66 | 0.39 | .427 | .979 |
| HlandCC% | 0.30 | 0.38 | .406 | .979 |
| HlandNCA | 0.08 | 0.36 | .301 | .763 |
| HInc | 116315.36 | 212027.84 | .691 | .729 |
| HIncLS% | 0.08 | 0.22 | .536 | .705 |
| HIncOF% | 0.61 | 1.10 | .362 | .815 |
|  |  |  |  |  |
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .517 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1552.226 |
| df | 190 |
| Sig. | .000 |

Variable independence test using pairwise correlation revealed low levels of correlations among most variables r < 0.3, warranting the use of Varimax rotation on the assumption that variables are independent. With 149 households sample size and 20 variables for PCA with variable to sample size ratio of 7.45, we use Eigenvalues of greater than 1.0 as cut off to select 9 PCs which explained 80.2% (**Table 4**). The scree plot (Figure 2) also shows that after 9 PCS the change in variation explained is minimal and stabilises.

**Table 4** Principal components with Eigen values and total variance explained

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
| Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 3.038 | 15.191 | 15.191 | 3.038 | 15.191 | 15.191 | 2.661 | 13.307 | 13.307 |
| 2 | 2.864 | 14.318 | 29.509 | 2.864 | 14.318 | 29.509 | 2.419 | 12.094 | 25.402 |
|  |  |  |  |  |  |  |  |  |  |
| 3 | 2.386 | 11.932 | 41.440 | 2.386 | 11.932 | 41.440 | 1.975 | 9.875 | 35.277 |
| 4 | 1.592 | 7.960 | 49.401 | 1.592 | 7.960 | 49.401 | 1.960 | 9.799 | 45.075 |
| 5 | 1.549 | 7.745 | 57.146 | 1.549 | 7.745 | 57.146 | 1.518 | 7.590 | 52.665 |
| 6 | 1.224 | 6.122 | 63.267 | 1.224 | 6.122 | 63.267 | 1.508 | 7.542 | 60.206 |
| 7 | 1.192 | 5.962 | 69.229 | 1.192 | 5.962 | 69.229 | 1.441 | 7.207 | 67.413 |
| 8 | 1.131 | 5.654 | 74.883 | 1.131 | 5.654 | 74.883 | 1.356 | 6.778 | 74.192 |
| 9 | 1.073 | 5.363 | 80.246 | 1.073 | 5.363 | 80.246 | 1.211 | 6.055 | 80.246 |
| 10 | .770 | 3.849 | 84.096 |  |  |  |  |  |  |
| 11 | .716 | 3.582 | 87.678 |  |  |  |  |  |  |
| 12 | .554 | 2.769 | 90.447 |  |  |  |  |  |  |
| 13 | .446 | 2.230 | 92.677 |  |  |  |  |  |  |
| 14 | .432 | 2.159 | 94.836 |  |  |  |  |  |  |
| 15 | .322 | 1.609 | 96.444 |  |  |  |  |  |  |
| 16 | .273 | 1.364 | 97.808 |  |  |  |  |  |  |
| 17 | .230 | 1.152 | 98.961 |  |  |  |  |  |  |
| 18 | .111 | .554 | 99.514 |  |  |  |  |  |  |
| 19 | .081 | .403 | 99.917 |  |  |  |  |  |  |
| 20 | .017 | .083 | 100.000 |  |  |  |  |  |  |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |

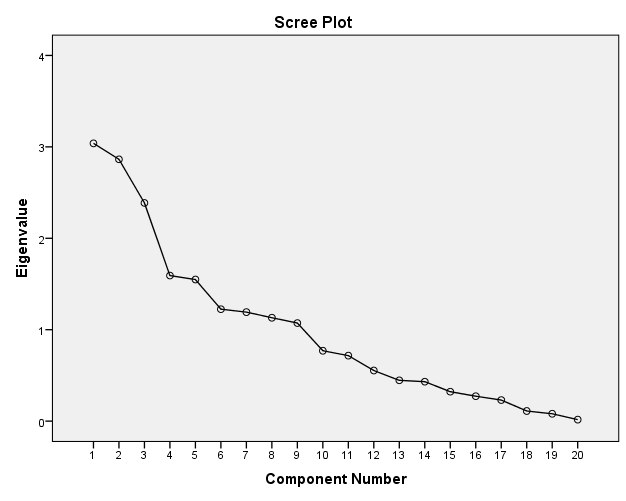


Figure 2 Scree plot of Eigen values and number of principal components

We used the 9 PCs that account for 80% of the variation (**Table 5**) to characterise the households. In some cases such as the first PC, more than one variable has higher loading that jointly carry the captured variation.

**Table 5**: Rotated Component Matrixa

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Component | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Hage | -.033 | .143 | .005 | -.034 | .159 | **-.687** | -.015 | .151 | -.048 |
| Heduc | .049 | .016 | .145 | -.038 | .044 | **.834** | .031 | .093 | -.034 |
| Hlabour | .045 | -.082 | -.029 | -.035 | .043 | .019 | **.845** | -.032 | -.325 |
| **Hdepend** | .087 | -.159 | .091 | -.043 | .009 | -.026 | **-.810** | -.086 | -.343 |
| **Hcom** | **.804** | .026 | .121 | -.139 | .070 | .176 | .134 | .045 | -.034 |
| Htransp | **.883** | -.037 | .072 | .086 | -.031 | -.125 | .007 | .064 | .021 |
| Hroad | .051 | -.039 | -.106 | **.964** | -.053 | -.026 | -.007 | .018 | -.011 |
| **Hcenter** | .014 | .007 | -.152 | **.960** | -.085 | .021 | .010 | -.020 | .008 |
| **HEquip** | **.820** | -.006 | -.086 | .118 | -.083 | -.043 | -.137 | .047 | .067 |
| Hlu | .276 | .033 | -.003 | .081 | .294 | -.131 | -.034 | **.730** | .119 |
| Hnr | -.006 | -.018 | -.067 | -.008 | .022 | -.009 | -.018 | .013 | **.896** |
| **Hland** | .085 | **.788** | .131 | -.027 | .451 | .040 | .013 | .049 | -.138 |
| HlandP | -.038 | **.860** | .001 | -.033 | .074 | -.082 | -.041 | .072 | .262 |
| **HPlot** | -.033 | .563 | .067 | -.047 | **.600** | -.058 | -.043 | .080 | -.138 |
| **HlandFC%** | -.019 | -.175 | **-.956** | .146 | .014 | -.053 | .041 | .045 | .073 |
| HlandCC% | .041 | -.055 | **.967** | -.127 | .100 | .090 | -.065 | -.009 | -.017 |
| **HlandNCA** | -.020 | **.806** | .014 | .029 | -.258 | -.081 | .116 | -.114 | -.112 |
| **HInc** | **.676** | .011 | -.054 | -.014 | .041 | .487 | -.050 | .024 | -.162 |
| HIncLS% | -.029 | -.009 | .035 | -.105 | **.821** | -.074 | .048 | -.068 | .077 |
| HIncOF% | -.037 | -.008 | -.049 | -.062 | -.242 | .044 | .067 | **.860** | -.063 |
| Extraction Method: Principal Component Analysis.  Rotation Method: Varimax with Kaiser Normalization. | | | | | | | | | |
| 1. Rotation converged in 6 iterations. | | | | | | | | | |

## Typology of farming households in maize mixed farming system

As noted from **Table 6** the cluster analysis resulted in several class prototypes for the farming households at different k-values and variable sets. From the K-mean cluster analysis, we see (in **Figure 3**) that with 20 variables, the optimal number of clusters is obtained at the point of inflexion where k=5. We also run the CA using the variables with highest loading and profoundly, using 20 variables gave fewer outlier clusters with small number of households.

**Table 6** Number of households per cluster (K) for all 20 variables by PCA

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| K = | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 |
|  | 63 | 63 | 74 | 75 | 79 | 89 | 89 | 89 | 122 | 125 | 144 |
|  | 26 | 26 | 30 | 34 | 34 | 34 | 38 | 38 | 22 | 22 | 5 |
|  | 19 | 20 | 15 | 15 | 17 | 14 | 17 | 17 | 3 | 2 |  |
|  | 14 | 14 | 14 | 14 | 14 | 7 | 3 | 3 | 2 |  |  |
|  | 9 | 10 | 6 | 6 | 2 | 3 | 1 | 2 |  |  |  |
|  | 7 | 8 | 5 | 2 | 1 | 1 | 1 |  |  |  |  |
|  | 3 | 3 | 2 | 1 | 1 | 1 |  |  |  |  |  |
|  | 3 | 2 | 1 | 1 | 1 |  |  |  |  |  |  |
|  | 2 | 1 | 1 | 1 |  |  |  |  |  |  |  |
|  | 1 | 1 | 1 |  |  |  |  |  |  |  |  |
|  | 1 | 1 |  |  |  |  |  |  |  |  |  |
|  | 1 |  |  |  |  |  |  |  |  |  |  |

**Figure 3** Knee bend plot of average distance to cluster centres verses number of cluster

## Household types

As seem in **Table 7,** income is the variable with the most discriminating power that significantly distinquish the classes into plausible types. Type I is distiquishable also by having significantly fewer communication facilities. Other variables with high discriminating power between types I and II include family labour, transport facilities, household and farm equipment and tropical livestock units per person. Household types I and III differ significantly in terms of age and level of education of the household head. The types II and III are signifiacntly distinqushed only by income levels.

Standardised Z-score to characterise household types for classes with more than 5 households are presented in **Table 7** and **Figure 4**. The standard Z-score shows that clusters differ in terms of having attributes with 0.02 to 0.2 difference. The summary statistics shows that the household type I is characterised by low income (averaging MK18,652 and range of 0.00 to 67,500), have fewer household equipment (mean monetary value MK3,260 and range of 0.0 to 150,000.00) and fewer livestock per person (average 0.06 and range 0.00 to 0.9). Though they hold more land (1.08 ha), per capita area is low (0.33 ha/person). It can also be seen that this class does not have transportation and communication facilities. Moreover, they stay very farm from the road and main trading centres. Thus this group would not have good access to input and output markets and have difficulty in transporting bulky goods. They are headed by older persons that attained less education.

Type II has a similar advantage as type I in that most households have better access and extracts more natural resources and own similar areas of non-cropped land. This group differs with others in that it generates some income from selling livestock and livestock products (10% of total income) and have better transport facilities (0.09). They grow more crops for food than for cash. They do not get considerable income from off-farm activities. The households are comprised mainly of potentially working age group with fewer dependants.

Type III have high income (MK328,542/annum) and larger land holding (1.33 ha), headed by younger persons (39.59 years old on average) almost every one attaining primary and a good number reached secondary school, and they have radios and mobile phones for communication. This group typically does not rely on NRM nor selling livestock for livelihood and has little area under fallow/not cultivated. It has more dependants. They grow cash crops on larger areas compared to food crops and have more livestock that are not generally sold.

The other 5 households in clusters 1 and 4 are potentially outliers and different from the rest interms of having higher education of household head, more communication, transport and equipment and larger income.

**Table 7** Differentiating the 3 household types using standardised scores and ANOVA

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Standardised scores | | | | | Means for 3 household types | | | | | | | | | |  |
| Type I (n=89) | Type II (n=38) | Type III (n=17) | n=3 | n=2 | Type I (n=89) | | Type II (n=38) | | | | Type III (n=17) | | | |  |
| Hage\* | 0.43 | 0.4 | 0.28 | 0.23 | 0.18 | 48.27 | b | 46.53 | ab | | 39.59 | | a | |  | |
| Hlabour\*\* | 0.32 | 0.39 | 0.49 | 0.67 | 0.67 | 2.09 | a | 2.87 | b | | 2.71 | | ab | |  | |
| Hdepend | 0.26 | 0.36 | 0.34 | 0.25 | 0.25 | 1.39 |  | 1.24 |  | | 1.50 | |  | |  | |
| Heduc\*\* | 0.28 | 0.25 | 0.3 | 0.37 | 0.4 | 0.97 | a | 1.16 | ab | | 1.47 | | b | |  | |
| Htransp\*\* | 0.07 | 0.13 | 0.16 | 0.17 | 0.81 | 0.37 | a | 1.00 | b | | 0.82 | | ab | |  | |
| Hcom \*\*\* | 0.03 | 0.09 | 0.07 | 0.18 | 0.64 | 8.92 | a | 16.26 | b | | 20.29 | | b | |  | |
| HEquip\*\* | 0.46 | 0.44 | 0.42 | 0.55 | 0.55 | 3.26 | a | 33.16 | b | | 30.00 | | ab | |  | |
| Hlu\*\* | 0.48 | 0.46 | 0.44 | 0.52 | 0.54 | 0.06 | a | 0.14 | b | | 0.14 | | ab | |  | |
| Hnr | 0.01 | 0.07 | 0.07 | 0.11 | 0.56 | 18343 |  | 18077 |  | | 11400 | |  | |  | |
| Hland | 0.06 | 0.14 | 0.15 | 0.16 | 0.13 | 1.08 |  | 1.19 |  | | 1.33 | |  | |  | |
| HlandP | 0.14 | 0.14 | 0.09 | 0.05 | 0.04 | 0.33 |  | 0.24 |  | | 0.24 | |  | |  | |
| HPlot | 0.17 | 0.19 | 0.21 | 0.15 | 0.29 | 0.55 |  | 0.70 |  | | 0.69 | |  | |  | |
| HlandFC% | 0.11 | 0.07 | 0.07 | 0.05 | 0.09 | 0.65 |  | 0.70 |  | | 0.62 | |  | |  | |
| HlandCC% | 0.09 | 0.12 | 0.12 | 0.05 | 0.06 | 0.30 |  | 0.25 |  | | 0.35 | |  | |  | |
| HlandNCA | 0.65 | 0.7 | 0.62 | 0.67 | 0.65 | 0.09 |  | 0.11 |  | | 0.02 | |  | |  | |
| HInc\*\*\* | 0.3 | 0.25 | 0.35 | 0.33 | 0.35 | 18653 | a | 128753 | b | | 328542 | | c | |  | |
| HIncLS% | 0.03 | 0.03 | 0.01 | 0 | 0 | 0.08 |  | 0.10 |  | | 0.05 | |  | |  | |
| HIncOF% | 0.01 | 0.08 | 0.21 | 0.49 | 0.93 | 0.67 |  | 0.45 |  | | 0.63 | |  | |  | |
| Hroad | 0.08 | 0.1 | 0.05 | 0.04 | 0 | 6739 |  | 6396 |  | | 6148 | |  | |  | |
| Hcenter | 0.05 | 0.03 | 0.05 | 0.07 | 0.05 | 7755 |  | 7405 |  | | 7187 | |  | |  | |
| \*\*\*, \*\*, \* significant at 99%, 95%, 90% respectively (unbalanced ANOVA) | | | | | | | | | |  | |  | |  | | | |  |
| Different letters indicate significant difference (at 5%) for predicted means between clusters using LSD | | | | | | | | | | | | | | | |  |

**Figure 4** Standandised scores characterising the 3 optimal household types considering the most significant variables that differentaited the three classes

# Discussion

Farm household typologies serve the purpose of identifying homogeneous groups of households and their farms for targeting intervention. As Timler *et al*. (2014) puts it, type specific analysis sets a meaningful compromise between analysing every single farm and assuming broad categories such as smallholders in general. Farming systems among smallholder farmers in Africa are highly diverse. Apart from inherent variations in soils, vegetation, topography and climate, sub-optimal performance within similar locations stem from differences in farmers’ livelihood aspirations and resource endowment (factors for productivity). Using the SLF, the study has revealed that heterogeneous smallholder households can be grouped using production factors into 3 subsets (types), homogenous within a certain range of attribute values which can be used for technology targeting (Andersen et al., 2007; van de Brand, 2011). Hence, instead of providing ‘blanket’ or precise single farm recommendations for smallholder farmers in certain areas, recognizing and responding to the variability in local farm characteristics promises more appropriate, targeted and efficient design recommendations to achieve improvements in agricultural production (Ojiem *et al*., 2006; Tittonell *et al*., 2010).

Although a few typologies have been conducted in the region, the current study has pointed out the important household, plot and spatial *state* variables that significantly distinguish the types of farming households. Some studies combined *state* variables with variables for behavioural outcomes including land use choice leading to *indeterminate* types. For a study of sustainable intensification, current typologies produced for various purposes ranging from cropping system, poverty and vulnerability could not be ideal. A most recent typology by Timler characterised the farms using income sources from farming and non-farm sources but could not show significant differences in underlying attributes.

References

Adams W.Y. and Adams E.W. (2008). Archaeological typology and practical reality: a dialectical approach to artefact classification and sorting. Cambridge University Press, New York.

Adimassu Z., Kessler A. and Hengsdijk H. (2012). Exploring determinants of farmers’ investments in land management in the Central Rift Valley of Ethiopia. *Applied Geography* 35:191-198.

Alnaji K.W. and Ashour W.M. (2011). A Novel Clustering Algorithm using K-means (CUK). *International Journal of Computer Applications* *(0975 – 8887)*, 25, 1, 25-30.

Andersen, E., B. Elbersen, F. Godeschalk, D. Verhoog (2007) Farm management indicators and farm typologies as a basis for assessments in a changing policy environment, Journal of Environmental Management v82(3) 353-362.

Le Bellec F., Cattan P., Bonin M. and Rajaud A. (2011). Building a typology of cropping practices from comparison with a technical reference: first step for a relevant cropping system redesigning process – results for tropical citrus production. *Fruits*, 66, 143-159

Binder C.R., Hinkel J., Bots P.W.G. and Pahl-Wostl C. (2013). Comparison of Frameworks for Analyzing Social-ecological Systems. *Ecology and Society*, 18, 4, 26.

Brown D.G., Reutlinger S. and Thomson A.M. (1996). Malawi: Food Security in a Market Oriented Economy. USAID Research Report 1011. <http://ufdc.ufl.edu/UF00072240/00001/1j>

Campbell B., Sayer J.A., Frost P., Vermeulen S., Perez M.R. Cunningham A., Prabhu R. (2001). Assessing the performance of natural resource systems. *Conservation Ecology*, 5, 2, 22

Cassman K.G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proc. Natl. Acad. Sci. USA*, 96, 5952–5959.

Chilonda, P. and Otte, J., 2006. Indicators to monitor trends in livestock production at national, regional and international levels. *Livestock Research for Rural Development*, 18, 6.

Chimtengo M., Ngongondo C., Tumbare M., and Monjerezi M. (2014) Analysing changes in water availability to assess environmental water requirements in the Rivirivi River basin, Southern Malawi. *Physics and Chemistry of the Earth, Parts A/B/C* **67-69**, 202-213.

Chulu O. and P. Wobst (2001). A 1998 Social Accounting Matrix for Malawi. Washington D.C., IFPRI.

Darnhofer I. (2014). Resilience and why it matters for farm management. *European Review of Agricultural Economics*, 41, 3, 461-484.

Dixon J., Gulliver A. and Gibbon D. (2001). Farming systems and poverty: improving farmers livelihoods in a changing world. FAO, Rome and World Bank, Washington DC.

Douillet M. and Toulon A. (2014). Developing a typology of agricultural holdings for improved policy design: a preliminary case study of Malawi. Foundation for World Agriculture and Rurality/ world Agricultures Watch, Working paper number 6.

Dorward A. (2002). A Typology of Malawian Rural Households. Working Paper: Institutions and Economic Policies for Pro-Poor Agricultural Growth, ESCOR Project R7989. DFID.

FAO (1996). Agroecological zoning: guidelines. FAO Soils Bulletin 73, Rome.

FAO ([2005](http://www.fao.org/ag/againfo/resources/en/publications/sector_briefs/lsb_MWI.pdf)). Livestock Sector brief Malawi. FAO, Rome.

Grillo J. and Holt J. (2009). Application of the Livelihood Zone Maps and Profiles for Food Security Analysis and Early Warning: Application of the Livelihood Zone Maps and Profiles for Food Security Analysis and Early Warning. USAID.

Filmer D. and L. Pritchett (2001). Estimating Wealth Effects without Expenditure Data- or Tears: An Application of Educational Enrolment in States of India. *Demography*, 38, 1, 115-132.

Fischer G. Nachtergaele F.O, Prieler S. *et al*. (2012). Global Agro-ecological Zones: Model Documentation. GAEZ ver 3.0. IIASA/FAO.

Garrity D., Dixon J. and Boffa J.M. (2012). Understanding African Farming Systems Science and Policy Implications. Paper presented at food security in Africa: bridging research and practice, Sydney.

Ghosh S. and Dubey S.K. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. *International Journal of Advanced Computer Science and Applications*, 4, 4, 35-39

Le Q.B., Park S.J. and Vlek P.L.G. (submitted). Household typologies and socio-ecological determinants of their land-use choices in Vietnam forest margins: Implications for empirical multi-agent modelling. *Land Use Policy*.

Le Q.B., Vlek P.L.G., Denich M., Martius C., Rodgers C. and van de Giesen N. (2005). Multi-agent system for simulation of land-use and land cover change: A theoretical framework and its first implementation for an upland watershed in the Central Coast of Vietnam. Ecology and Development Series No. 29, University of Bonn.

Le Q.B., Seidl R. and Scholz R.W. (2012). Feedback loops and types of adaptation in the modelling of land-use decisions in an agent-based simulation. *Environmental Modelling and Software*, 27, 28, 83-96.

Lesschen J. P., Verburg P. H. and Staal S. J. (2005). Statistical methods for analyzing the spatial dimension of changes in land use and farming systems. LUCC Report Series No. 7. The International Livestock Research Institute, Nairobi, Kenya and LUCC Focus 3 Office, Wageningen University, the Netherlands, 80 p.

Lin J.Y. (1990). Education and innovation adoption in agriculture: evidence from hybrid rice in China. Working paper number 603. Department of Economics, University of California, Los Angels.

 McBratney A., Whelan B. and Ancev T. (2005). Future directions of precision agriculture. Invited paper, 7th International Conference on Precision Agriculture, Minneapolis.

McConnel D.J. and Dillon J.J. (1997). Farm Management for Asia: a Systems Approach. FAO Farm Systems Management Series – 13, FAO, Rome.

Mohamad I.B. and Usman D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology* 6(17): 3299-3303.

Moser C. and Felton A. (2007). The construction of an asset index measuring asset accumulation in Ecuador. Chronic Poverty Research Centre, Working Paper 87, Global Economy and Development, Washington DC.

Mponela P., Jumbe B.L. and Mwase W.F. (2011). Determinants and extent of land allocation for *Jatropha curcas* L. cultivation among smallholder farmers in Malawi. *Biomass and bioenergy*, 35, 7, 2449-2505.

Naes T. and Mevik B.H. (2001). Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics*, 15, 413–426.

Ojiem J.O., de Ridder N., Vanlauwe B. and Giller K.E. (2006). Socio-ecological niche: a conceptual framework for integration of legumes in smallholder farming systems. *International journal of agricultural sustainability*, 4, 1, 79-93.

Ostrom E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, 325, 5939, 419-422.

Payne R.W. (1993) Genstat 5 release 3 reference manual. Oxford University Press Inc, New York.

Peters P.E. (2006). Rural income and poverty in a time of radical change in Malawi. *The Journal of Development Studies*, 42, 2, 322–345.

Peters P.E. and Kambewa D. (2007). Whose security? Deepening social conflict over ‘customary’ land in the shadow of land tenure reform in Malawi.” Center for International Development, Working Paper No. 142, Harvard University.

Polhill J.G., Parker D., Brown D. and Grimm V. (2008). Using the ODD Protocol for Describing Three Agent-Based Social Simulation Models of Land-Use Change. *Journal of Artificial Societies and Social Simulation*, 11, 2, 3

Railsback S.F and Grimm V. (2011). Agent-Based and Individual-Based Modeling: A Practical Introduction. Princeton University press. pp 352.

Riveiro-Valiño J.A., Álvarez-López C.J., Marey-Pérez M. F. (2009). The use of discriminant analysis to validate a methodology for classifying farms based on a combinatorial algorithm. *Computer and electronics in agriculture*, 66(2)113-120.

Rokach L. and Maimon O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook*, pp. 321-352. Springer US.

Ronner, E. and Giller, K.E. 2012. Background information on agronomy, farming systems and ongoing projects on grain legumes in Ethiopia, www.N2Africa.org, 33 pp.

Rware H., Wairegi L., Odour G. et al. (2014). Assessing the Potential to Change Stakeholders Knowledge and Practices on Fertilizer Recommendations in Africa. Agricultural Sciences, 5, 1384-1391

Salvador S. and Chan P. (2004) Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. Proceedings of the Sixteenth IEEE International Conference on Tools with Artificial Intelligence, pp. 576–584. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Sarstedt M. and Mooi E. (2014). A Concise Guide to Market Research (2nd ed.). pp273-324. Springer.

Sconnes I. (1998). Sustainable livelihoods framework for analysis. IDS working paper 72.

Shefrin H. and Statman M. (2000). Behavioral Portfolio Theory. *Journal of Financial and Quantitative Analysis*, 35, 2, 127-151.

Sibanda L.M. (2014). African smallholder farm families have lost the elasticity to bounce back!

Simler, K. (1994). Agricultural Policy and Technology Options in Malawi: Modelling Responses and Outcomes in the Smallholder Subsector. Cornell Food and Nutrition Policy Program Working Paper 49.

Snapp S.S., Blackie M.J. and Donovan C. (2003). Realigning research and extension to focus on farmers’ constraints and opportunities. *Food Policy*, 28, 349-363.

Thiombiano B.A., Le Q.B., Odai S.N., and Ouedraogo D. (2015). Exploring soil nutrient management and production performances to support building smallholder farms’ resilience to climate change: case of South-Western Burkina Faso. PhD thesis, Kwame Nkrumah University of Science and Technology (KNUST).

Timler C., Michalscheck M., Klapwijk C., Mashingaidze N., Ollenburger M., Falconnier G., Kuivanen K., Descheemaeker K., Groot J. (2014). Characterization of farming systems in Africa RISING intervention sites in Malawi, Tanzania, Ghana and Mali. Department of Plant Sciences, Wageningen University and Research Centre / International Institute of Tropical Agriculture.

Tittonell P., Vanlauwe B., Leffelar P.A., Shepherd K.D. and Giller K.E. (2005). Exploring diversity in soil fertility management of smallholder farms in western Kenya II. Within-farm variability in resource allocation, nutrient flows and soil fertility status. *Agriculture, Ecosystems and Environment*, 110, 166–184

Tittonell P., Muriuki A.W., 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, 83–97.

van de Brand G. (2011). Towards increased adoption of grain legumes among Malawian farmers - Exploring opportunities and constraints through detailed farm characterization. MSc Thesis. Universiteit Utrecht/ Wageningen University.

Wikipedia, 2013. Livestock grazing comparison [online]. Available on: http://en.wikipedia.org/wiki/Livestock\_grazing\_comparison [Accessed 12 February 2013].

Williams B., Brown T. and Onsman A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8, 3, 1-13.