Farm typologies for West Nile Uganda, with implications for tobacco control policy

Catherine Namome[a], Beatrice Conradie[b], Corne van Walbeek[c], Tony Leiman[b]

[a] Graduate student, School of Economics, University of Cape Town. Private Bag X3, Rondebosch 7701, South Africa
[b] Associate Professor; [c] Professor. School of Economics. University of Cape Town. Private Bag X3, Rondebosch 7701, South Africa

Abstract:

Tobacco control regulation in the first world reduced the cultivation of the crop, prompting a shift in production to under-developed countries such as Uganda. There is political will in the first world to identify and support viable agricultural alternatives for tobacco farmers, but the paucity of data for these production systems makes formulating the appropriate policies difficult. This paper attempts to rectify this deficiency by describing prevailing production systems. A survey conducted in West Nile Uganda gathered data on 126 farm producers. A factor analysis technique (principal component analysis) gave seven principal components − related to landholdings, labour, farm inputs, gross output and revenue that characterised the farms. A two-step cluster analysis was performed on the factor scores, identifying three homogenous farm types. A detailed description is presented of the production system for each of these three farming types. The dominant farm type in West Nile is a traditional tobacco farming group. This farm type holds a mean of 4 acres per farm household and cultivates at least three crops. The next most prevalent type is a medium sized mixed farming type with a diversified production system. The least common farm type comprises of modern coffee and cassava farms using moderate farm inputs but maintaining higher crop yields and ultimately higher profit margins per acre. .

Keywords: Tobacco control; Farm typology; Principal Component Analysis; Cluster analysis; West Nile

1. Introduction

In the past century, most of the world's tobacco production took place in North America, China and in Zimbabwe and South Africa (FAO, 2015). The synthesis of knowledge on the impact of smoking, coupled with concerted public action over the past decades, has made the first world squeamish about tobacco production, to the point where North America currently produces only 19% of the world crop (Faostat, 2015). Production has shifted to less developed countries. In 2012, five African countries were among the top 20 producers of leaf tobacco in the world: Malawi (6^{th}), Tanzania (8^{th}), Zimbabwe (9^{th}) and Mozambique (17^{th}) Hu and Lee (2015).

Tobacco production continues in Africa because, unlike the first world, local people still believe that the economic benefits outweigh the health risks. The rest of the world no longer thinks this way (Geist et al., 2009). This led the World Health Organisation (WHO) to adopt the Framework Convention on Tobacco Control (FCTC), an international treaty to reduce demand for tobacco products and address supply concerns. On the demand side the treaty provides for taxes on tobacco products, bans on tobacco advertising of any sort, protecting individuals from tobacco smoke and mandatory warnings about the dangers of tobacco on all tobacco packaging (WHO, 2015). On the supply side the treaty has focussed on a search for alternative livelihoods for tobacco growers.

Policy interventions need to be based on a sound understanding of how those farmers currently operate, and on options for change that are based on the realities they face. Thus capturing farm heterogeneity through the analysis of farm typologies is vital to design sustainable policies. This study examines and characterises the farm types in West Nile, Uganda for that purpose. The analysis can be used as a basis for comparing the viability of tobacco farming to other cropping systems. This may make a valuable contribution to the task of accelerating tobacco control policies in Africa, since farm typologies constitute an essential step in explaining uneven performance of farming systems. As the home to many smallholder tobacco farmers, West Nile region in Uganda, occupies a position of special importance. To develop the typologies, the procedure used by Goswami et al. (2014), Milán et al. (2003) and Gaspar et al. (2008) was followed. This method combines principal component analysis with k-means clustering and single variable analyses of variance to establish significant differences between types of farms.

The paper is divided into six sections. The next section describes the study area, data collection process as well as the statistical methods used for the analysis. Section 3, is a results section which presents the findings and section 4 discusses the resultant typologies from the various statistical tests, provides a more qualitative description of each production system, and identifies the major constraints to production faced by farmers in each category. Section 5 presents the implications of these findings for policy. The paper ends with brief conclusions in section 6.

2. Data and methods

2.1 The study area

West Nile is located in the north-west part of Uganda. Two districts, Arua and Maracha were surveyed (see Figure 1). The region is a plateau that is divided into low ridges and by many small streams that thread their way across it. A more detailed description of the biophysical environment of West Nile is given in (Middleton, 1965, Middleton and Greenland, 1954). The region's average rainfall is 1340 millimetres per annum, with one rainy season from April to November (Kraybill and Kidoido, 2009). The peak levels of precipitation occur between May and July. This distribution allows cultivation of at least two crops annually and makes irrigation unnecessary for most farmers. The dry periods are during the months of December to January.

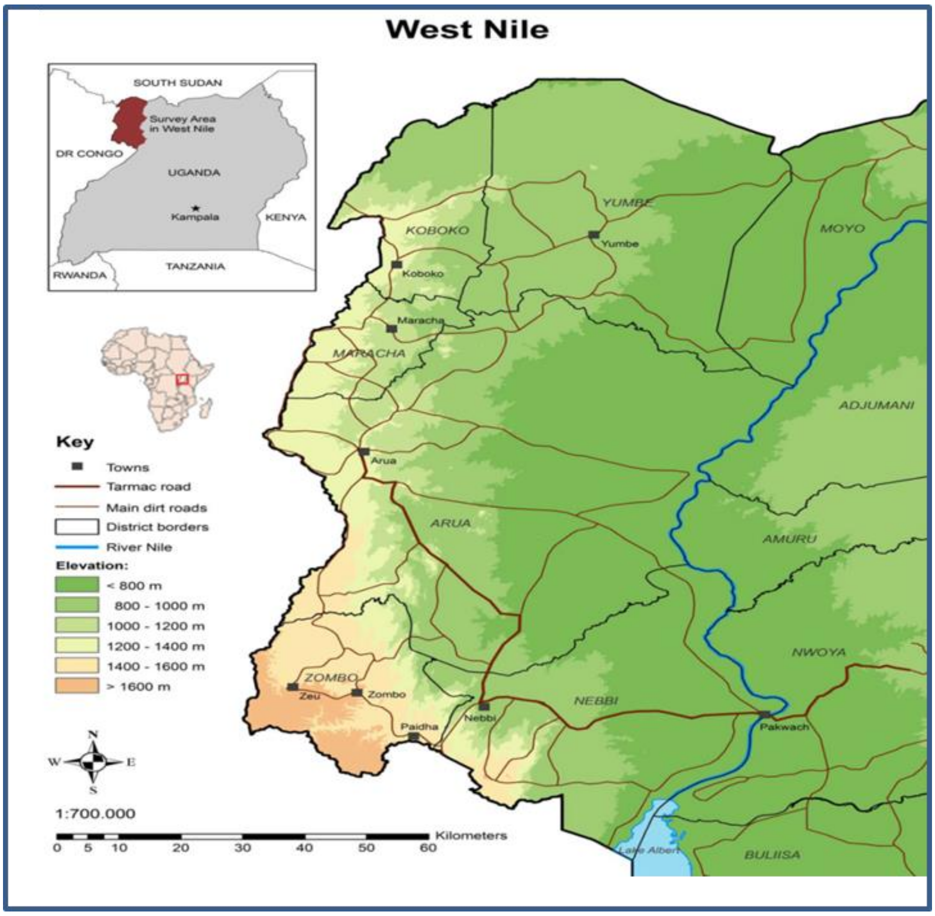

Figure 1: Location of the study area ***(Source: Ministry of finance, Uganda)***

The household farming structure in West Nile is mainly founded on three types of fields which gives equitable access to the same portfolio of soil qualities. The *amvu akua* (home gardens) are used to grow vegetables. Close proximity to livestock stabling pens means that these fields are richly manured, which counters the poor soil quality characteristic of other land. The *yimile* (riverine fields) are also relatively fertile and are well irrigated as they are riparian. They are used to grow staples such as sweet potatoes, cassava, beans and maize. These crops are sown in mixed stand and maintain vegetative cover during most of the growing season which protects the soil structure. Where necessary land is left fallow for a period of up to three years, after which the soil quality is normally restored.

The *amvu amve* (fields outside) are the fields in which most of the tobacco is cultivated. Typical crops grown here are tobacco, cassava, sesame, beans and maize. Livestock is minimal in the area, though some households do rear a few goats, sheep and cattle. The main source of labour is from members of the family and neighbours. The details of the household farm characteristics are provided in the next section in Table 1.

2.2 Data collection

A cross-sectional survey of tobacco and non-tobacco farmers was conducted in two districts of West Nile, Uganda, in August 2014. Since villages vary considerably in size, it was decided to sample 10% of each village visited. In the end 67 randomly selected villages were visited in 126 sets of household and farming data collected. By Kotrlik and Higgins (2001) guidance, 5% is considered as being adequate for cross-sectional household surveys, this rate of sampling was sufficient. In addition, the villages of West Nile exhibit a high degree of homogeneity, which allows the use of a small sample for this study. The unit of analysis was the farm household and these were selected from the lists of farmers collected from the District Agricultural Offices.

Heads of households were selected as survey participants mainly because they usually had the decision making power for farming and allocation of household resources. Interviews were conducted by the author and two field workers who are employed by the national agricultural research organisation in Uganda. The typical interview took forty five minutes and in most cases was attended by the head of the household (husband) and occasionally by other family members too. Table 1 below shows some characteristics that formed part of the interviews.

Table 1: Some farm characteristics for the farming systems (n=126)

Farm characteristic	Mean	SD*
Education (years)	9.85	3.92
Farm size (acres)	5.14	3.71
Farmers cultivating tobacco only	0.54	0.50
Farmers using hired labour only	0.83	0.38
Farmers using family labour only	0.89	0.32
Farmers using both family and hired labour	0.24	0.31
Farmers using inorganic fertilisers	0.68	0.47
Farmers accessing extension services	0.98	0.13

*Standard deviation

The majority of farmers (83%) had attained an education either at primary, secondary or degree level. The average number of years of formal education is just under ten years. Also most farmers (81%) owned a piece of land through different land tenure systems such as freehold, private, and communal. The average farm size is just over 5 acres per household. Communal and freehold are common and legally recognised forms of land tenure in the region. Roughly a third (35%) of the surveyed farms cultivated tobacco only and could be described as commercially orientated. These farmers also rent land from members of the community in order to expand their tobacco production. Nine percent of surveyed farmers used hired labour, which was common on tobacco farms. Family labour is a significant resource and 15% of farm households indicated to be utilising family members in most of their farm activities. Family members are mainly employed to work on activities such as ploughing, planting and harvesting of the crop. In some cases, tobacco farmers also indicated that they use family members in the curing process of leaf tobacco. Nearly three quarters of farmers used both hired and family labour on their farms.

The majority of farmers fertilise their fields and 68% of surveyed farm households indicated to be applying specifically the inorganic type of fertiliser. A dominant type of inorganic fertiliser is the combination of nitrogen, phosphorus and potassium (NPK). Animal manure is also widely used but specifically by farmers cultivating non-tobacco crops. Access to agri-extension services is important and nearly all (98%) households had a single visit at least every two months from an extension officer. Tobacco farmers receive extension services from tobacco companies, while those cultivating non-tobacco crops receive their extension services through local government agricultural services.

2.3 Analytical strategy

Means and standard deviations (for numerical variables) and frequencies (for categorical and discrete variables) were calculated for the most informative variables. A range of variables relating to three broad constructs, namely for farm household factors (education, tenure status, farm size, renting land, financing inputs), the costs associated with various farm inputs (planting materials, fertilizer, pesticides, labour and other specific inputs) and output and revenue (measures as gross output and sales for first crop, second crop and third crop) were used in the analysis. The first crop represents crops that are regarded as cash crops (for commercial purposes), are planted usually at the beginning of the farm calendar and often take a larger share of farm inputs. Second crop are categorised as food and partially cash crops, and the third crop represents crops produced entirely for food security but given a surplus can eventually be sold to increase household income.

In order to develop farm types, cluster analysis was used due to its ability in defining homogenous groups of farms. However, cluster analysis with cross-sectional data is usually associated with problems of multicollinearity among variables (Iraizoz et al., 2007). Ketchen and Shook (1996), suggest that multicollinearity can be reduced when a data reduction technique such as factor analysis is used. Factor analysis defines the underlying structure in a data matrix, analysing the nature of interrelationships among a vast number of variables by defining a set of common underlying factors (Iraizoz et al., 2007). Data reduction is used to achieve factor scores for each underlying dimension which substitutes them for the many original variables (Hair et al., 2006).

In our case, factor analysis was adopted with the extraction method of principal component analysis and varimax rotation, as applied by Kaiser (1970) and Abdi and Williams (2010). Principal component analysis was applied to 21 original variables to reduce the number of factors needed for cluster analysis. This method assures that the obtained factors are orthogonal and so avoids the problem of multicollinearity between variables used in the cluster analysis. Factors presenting an eigenvalue greater than one were retained for cluster analysis (Hair et al., 2006, Joseph et al., 2014, Kaiser, 1970). Principal component analysis extracted seven linear combinations (principal components) of the original variables whose weights corresponded to the eigenvectors of the correlation matrix. A varimax rotation of the component matrix was performed for the interpretation of the factors. For interpretation purposes, factor loadings greater or equal to 0.5 were used in each factor (component).

The factor scores from the retained principal components (factors) where then subjected to a cluster analysis. The cluster analysis was performed in two stages; first, a hierarchical cluster analysis was used to identify the first approximation of clusters and profile the cluster centres. The algorithm used in the hierarchical technique was the Ward's method based on squared Euclidean distances as a distance measure. Then, the observations were clustered by a k-means method (non-hierarchical clustering) with the cluster centres from the hierarchical results used as initial starting values. Using the k-means algorithm on a case by case basis made it possible to produce the number of clusters that seemed more realistic and meaningful as a final solution. This combined procedure allows one to benefit from the advantages associated with both hierarchical and k-means methods (Iraizoz et al., 2007).

Cluster analysis is important as it seeks to divide a set of principal components into a small number of groups which are relatively homogenous within themselves and heterogeneous between each other (Goswami et al., 2014). Cluster analysis accomplishes two functions: it computes similarities among the observational units, and it differentiates observational units into subgroups (Joseph et al., 2014). To profile and qualitatively describe the formed clusters, one-way analysis of variance (ANOVA) tests were performed to establish statistically significant differences in the means across farm types. Farm attributes from household characteristics, farmers' crop choices, farm endowments and yield and per-acre profitability where used. The statistical computer package used for this analysis was SPSS version 23.

3. Results

3.1 Principal Component Analysis

Two tests that indicate the validity of principal component analysis were determined. The Kaiser-Meyer-Olkin measure of sampling adequacy (Kaiser, 1970) was 0.73, indicating that the data matrix has sufficient correlation to justify the application of principal component analysis. Bartlett's test of sphericity was large and statistically significant at $p \leq 0.001$ and, so the hypothesis that the correlation matrix is the identity matrix can be rejected. These two tests indicate that the set of variables used in principal component analysis were appropriate.

A seven factor solution emerged and factors that presented an eigenvalue that was greater than one (see Table 2) were retained for further analysis. This solution explained a 77 per cent of total variance in the data set, which is more than satisfactory (Hair et al., 2006).

Table 2: Eigenvalues corresponding to each principal component (PC) and the relative total variance explained

PC	Eigenvalues	%variance explained	%variance accumulated
Comp 1	5.67	26.99	26.99
Comp 2	2.74	13.06	40.05
Comp 3	2.34	11.15	51.20
Comp 4	1.66	7.93	59.13
Comp 5	1.46	6.97	66.10
Comp 6	1.18	5.63	71.73
Comp 7	1.09	5.18	76.91

The retained principal components were extracted using the varimax rotation of the component matrix. Table 3 shows the rotated component matrix for each independent variable with corresponding factor loadings. A closer look at each column of Table 3 helps us to define each component according to the strongly associated variables (high loadings are in bold).

The first component explains 26% of the total variance and is strongly correlated with farm input expenditures on access to finance, pesticides, jute twine, fuel-wood and fuel pipes. Thus the component may represent tobacco farms that pay much attention on how to obtain and maintain these types of inputs as their predictor of farm productivity. The loadings for these variables are 0.563, 0.620, 0.856, 0.849 and 0.920 respectively. The highest scoring farms on this component are more input oriented, and we would name component 1 as farms utilising *modern inputs*.

The second principal component with an explained variance of 13% has high positive correlation coefficients with farm expenditures mainly on labour. Labour requirements for planting, weeding, harvesting and curing have factor loadings of 0.763, 0.705, 0.688 and 0.822 respectively. Farms scoring high on this factor are characterized by values of their labour management, thus the component represents farms concerned with labour requirements and may be named as *labour oriented farms.*

Table 3: Rotated component matrix for each of the 21 descriptive variables according to the eight principal components

	Principal Components						
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**
Farm household factors							
Farm size	-0.008	-0.011	0.176	-0.007	**0.938**	0.021	0.056
Rented land	0.020	-0.047	0.242	0.038	**0.931**	0.007	0.031
Access to finance	**0.563**	0.378	0.279	0.064	0.139	-0.147	0.280
Farm expenditure (cost of planting material, pesticides, labour and other inputs)							
Planting material (first crop)	0.412	0.200	**0.700**	0.094	0.195	-0.050	0.030
Planting material (second crop)	-0.055	0.028	-0.089	**0.742**	-0.019	0.194	0.271
Pesticides	**0.620**	0.457	0.344	0.094	0.101	-0.117	0.143
Jute twine	**0.856**	0.175	0.058	-0.187	-0.036	-0.065	-0.005
Fuel-wood	**0.849**	0.245	0.030	-0.117	0.015	-0.033	0.046
Fuel pipes	**0.920**	0.036	0.049	-0.081	-0.048	0.028	-0.024
Labour: planting	0.301	**0.763**	0.242	0.035	-0.068	-0.013	-0.047
Labour: weeding	0.193	**0.705**	0.170	-0.061	-0.046	0.272	-0.164
Labour: harvesting	0.004	**0.688**	-0.089	-0.185	0.088	-0.142	0.314
Labour: curing	0.250	**0.822**	0.066	0.066	-0.056	-0.027	-0.108
Labour: marketing	0.002	0.221	0.011	0.132	0.016	-0.183	**-0.668**
Family labour	0.118	0.165	0.199	0.082	0.101	-0.150	**0.654**
Farm output and revenue							
Gross output (first crop, kg)	-0.057	-0.010	**0.915**	-0.077	0.146	0.007	0.058
Gross revenue (first crop)	0.167	0.236	**0.875**	-0.008	0.210	-0.073	0.097
Gross output (second crop, kg)	-0.087	0.017	-0.018	**0.938**	0.022	0.082	-0.059
Gross revenue (second crop)	-0.129	-0.111	0.087	**0.777**	0.031	-0.033	-0.256
Gross output (third crop, kg)	-0.077	0.003	-0.067	0.043	0.008	**0.898**	0.004
Gross revenue (third crop)	-0.048	0.012	-0.016	0.167	0.021	**0.917**	0.027

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization

Principal component three with an explained variance of 11% has high positive correlation coefficients with variables such as gross output for a first crop, gross sales (first crop) and the cost of planting material for the first crop. Farms scoring high on this component are characterised by the presence of producing cash crops. The representative factor loadings are 0.915, 0.875 and 0.700 respectively. This component may represent farms that are *commercially oriented*.

The fourth component (8% of variance) is highly correlated with coefficients associated with measures of gross output, sales and the cost of planting material for the second crop. Representative factor loadings are 0.938, 0.777 and 0.742 respectively. Component four could be survivalist farms oriented toward food crop production. Component 5 produces a variance of 7% and is positively correlated with own resources (farm size; 0.938) and with foreign inputs (rented land; 0.931). Farms under component five seem to be on an expansionist strategy and may be named as *farm expansionists*. The sixth component has high loadings on gross output and sales of the third crop with factor loadings of 0.898 and 0.917, and is perhaps most usefully called *third crop*. The seventh component is correlated with family labour (0.654), and negatively correlated with hired labour for marketing their farm produce (-0.668).

3.2 Cluster analysis

The factor scores from the retained principal components formed the basis of the cluster analysis. Using both hierarchical and K-means methods, a three cluster solution was settled on as it produced a reasonable trade-off between meaningful cluster differences and reasonable cluster sizes. Table 4 reports some descriptive statistics by cluster membership of respondents interviewed. The means were compared across each farm type using the F-test. The test reveals if farm differences between clusters are statistically significant. Cluster 1 contained 43 farms (35% of the total), Cluster 2 had 53 farms (43%) while Cluster 3 included 27 farms (22%). There are four main differences between the three homogenous farm types identified here. These include differences in farmers' household profiles, their crop choices, farm endowment and yield and per-acre profitability.

3.2.1 Differences in household profiles across clusters

From Table 4 below, it is apparent that these farm types do not significantly vary across their age groups. Farmers in these farm types are overwhelmingly men and in their forties, who are married, are heads of households and are Lugbara speaking[1]. Virtually all farmers in these farm types mention that farming is their main occupation and the vast majority of households rely entirely on agriculture for an income. The average household size is around nine to ten people, of which half are adults and half children below the age of twelve. This suggests a substantial source of household labour. The percentage of households with off-farm income varied between 7% and 14%. The only household characteristic which varied significantly across cluster membership was the percentage of respondents who have primary education only. It was 65% and 64% respectively for clusters 1 and 2 and just 41% for cluster 3. Although the majority of the variables comprising the household profile did not vary significantly across cluster membership, there was a tendency for the farmers in cluster 2 to be younger and cluster 3 farmers were more likely to be female, and were definitely better educated.

Household profiles already predict how these homogenous types of farmers are likely to engage with the wider world. On the one hand, one would expect that households who already have one member working in an off-farm occupation to have better networks to the outside. Links to the wider world may make such households more likely to experiment with modern inputs, find it easier to source and pay for modern inputs and perhaps to have the strongest relationships with tobacco companies. On the other had one could say the same of the members of cluster 3, who should find it easier to practice modern farming (and all that comes with it) on

[1] Middleton (1965) explicitly describes the lugbara speaking people of West Nile Uganda.

account of their better education. In summary then, these household profiles suggest that cluster 2 is likely to contain the most traditional farm operators.

Table 4: Farm and farmer differences across clusters (ANOVA)

Cluster description	**Cluster 1** N=43	**Cluster 2** N=53	**Cluster 3** N =27	**ANOVA (F-test statistic)**
	Medium sized mixed farms	Small traditional tobacco farms	Small modern coffee &cassava farms	F-statistic
Household attributes				
Age of household head (years)	49	46	43	1.87^{ns}
% farmers who are head of household	95	92	93	0.18^{ns}
% farmers who are men	98	91	85	1.85^{ns}
% farmers, working as farmers	100	98	96	0.72^{ns}
% farmers with only primary education	65	64	41	2.53*
Size of household	10.2	9.8	9.0	0.80^{ns}
% households with off-farm income	14.0	7.5	7.4	0.65^{ns}
Selected farm endowments				
Farm (acres)	7.3	4.2	4	12.46***
Land under crops (acres)	3.26	3.15	3.93	1.72^{ns}
% farms on clan / family land	72	55	85	4.29**
% farms using company credit	44	64	19	8.45***
% farms using bank loans	40	13	56	9.42***
Crop choices				
% of farms growing each crop				
Tobacco	47	75	22	12.89***
Cassava	53	42	74	3.96**
Sweet potatoes	26	26	48	2.45*
Maize	35	43	48	0.66^{ns}
Coffee	9	2	19	3.48**
% of farms practicing mono cropping	42	45	22	2.13†

*** signifies p ≤0.001 ** signifies ≤≤0.05, * signifies ≤ 0.10, †signifies ≤0.15, ^{NS} signifies Not significant

3.2.2 Differences in distribution of farm endowments

In the previous section we tried to argue that certain household profiles predict a greater likelihood to engage in modern agriculture. In this section systematic differences in farm endowments are examined for the same purpose. Farm endowments are complex. The aspects focussed on in Table 4 are farm size, tenure system, type of production system and access to credit. Farm size comprises own and rented land. Renting is a common phenomenon, with most operations renting roughly half their total area cultivated. The tenure system can be communal, freehold and clan land. Since the majority of respondent indicated that they obtained their land through inheritance a dummy variable was constructed where 1 = operate on clan or family land and 0 = otherwise. The majority of farmers we interviewed indicated that they use credit to finance farm inputs. The two main sources of credit were bank loans and advances made by tobacco companies. Hence two dummy variables were constructed to capture if a farmer accessed either of these two sources of credit.

Whereas there was almost no difference in the household profiles across clusters, all aspects of the farm endowments considered here varied significantly across clusters. Cluster 1 operates on almost double the farm size of clusters 2 and 3, which were similar. About three quarters of the larger farms are on clan land. In Cluster 2 almost half of farms are not on clan land, while in cluster 3 the clear majority are on clan land. Exactly the opposite of what emerged from the respondent profiles we find some evidence here that cluster 2 farmers might be more oriented towards modern farming as they are not bound by clanship ties which speaks of tradition.

It is clear from the data in Table 4 that company credit and bank loans are complementary. Between 70% and 80% of respondents, use either of these sources as the main way of financing their next crop. For cluster 1 the proportions were similar; 44% indicated that they use company credit and 40% indicated that they use bank loans. Cluster 2 farmers depend on company credit rather than bank loans, while cluster 3 farmers are more likely to depend on bank loans than any of the other groups, perhaps because they do not have access to company credit. Company credit is only available for tobacco growing. While it is normal to think of farm endowments determining crop choices, here crop choices might determine what resources the landholder has access to. Therefore, the next section examines differences in crop choice.

3.2.3 Differences in crop choices across clusters

Coffee and tobacco are specialised cash crops in West Nile. They also seem to be mutually exclusive on most farms as only two farms grew both in 2014. Coffee is less prevalent as compared to cassava in the region. Cassava has become the upcoming cash crop, putting Uganda as the six largest producer of the crop in Africa. The sample contains 66 tobacco producers and ten coffee growers. In addition farmers grow maize, cassava, sweet potatoes, beans and bananas and several minor crops (e.g. onions), some of which is for market and the rest for home consumption.

Almost every other cluster 1 farmer grew tobacco in 2014. There was no clear correlation between tobacco growing and the choice of any other crop, except cassava. Cassava growing was negatively correlated with tobacco growing ($r = -0.3457$, $p = 0.0232$), which suggests that although these medium sized operations have a high prevalence of tobacco growing, they should not be described as tobacco growers. Cassava production was positively correlated with the production of sweet potatoes ($r=0.3330$, $p = 0.0291$), but unrelated to maize growing ($r = 0.0023$, $p = 0.9884$). It is possible that within this group farmers specialise in either tobacco or cassava to generate cash incomes.

In cluster 2 three in four farmers grew tobacco in 2014. No other crop had this uptake and tobacco production was uncorrelated with the production of any other crop. Within cluster 2 the production of stables is associated with the production of other stables. Maize production was significantly correlated with cassava production (r = 0.3340, p = 0.0117) and sweet potato growing (r = 0.2525, p = 0.0681), whilst cassava production was highly correlated with sweet potato growing (r = 0.5375, r = 0.0000). Here the impression is that tobacco is the only cash crop of note whilst cassava and sweet potatoes are cultivated for subsistence purposes only. These farmers probably reserve their outlying fields for tobacco (and not maize) while their stables are cultivated on riparian land.

In cluster 3 there are three tobacco growers, eight coffee growers and 15 farmers who grow cassava. In this group the production of stables is positively correlated with the production of other staples; maize production is correlated with the production of cassava (r =0.4009, p = 0.0382) and bananas (r = 0.3669, r = 0.0598), whilst cassava growing was correlated with sweet potato growing (r =0.4009, p = 0.0382). Tobacco production was negatively correlated with the production of all other crops, and significantly so with the production of maize (r = -0.3368, p = 0.0858), which again suggests that outlying field are reserved for tobacco over maize growing. Coffee production was marginally positively correlated with sweet potato (r = 0.3039, p = 0.1233) and cassava (r = 0.2820, p = 0.1541) production, which suggest that these farms might have relatively better access to reliable water sources. In the next section we examine differences in Input uptake, yield and per acre profitability to extend the argument on crop choices.

3.2.4 Differences in uptake of modern inputs, yield and per-acre profitability

In this section, we look at Table 5, and try to give cluster differences on uptake of modern inputs such as fertilisers and pesticides. Fertiliser and pesticide uptake showed significant differences across clusters. In Cluster 1, 49% of farmers applied inorganic fertilisers, 16% applied animal and poultry manure whilst the rest (28%) did not fertilise their fields. Fertiliser expenditure varied significantly across clusters, with cluster 2 who mainly produce tobacco, spending more on their application of fertilisers but interesting, posting low crop yields. Pesticide uptake also varied significantly across clusters with 75% of farmers in cluster 2 having applied pesticides, 47% in cluster 1 and only 22% used pesticides in cluster 3. In terms of crop yield, cluster 1 performs well, when compared to the other two clusters. The average crop yield that farmers harvest in 2014 from cluster 1 was 2101 kilograms, which significantly differed from the 1865 kilograms of cluster 3 and 1386 kilograms of cluster2.

Table 5: Differences in Input uptake, yield and per acre profitability across clusters

Cluster description	***Cluster 1*** *N=43*	***Cluster 2*** *N=53*	***Cluster 3*** *N=27*	***ANOVA F statistic***
	Medium-sized mixed farms	Small traditional tobacco farms	Small modern coffee & cassava farms	F-statistic
Modern Input uptake				
Fertilisers	65	83	44	6.84***
% NPK, CAN	49	53	22	AC
% animal manure	12	30	18	AC
% poultry manure	4	0	4	AC
Pesticides	47	75	22	12.89***
Land holdings				
Farm size (acres)	7	4	4	12.24***
Renting land (acres)	5	3	3	6.27***
Total (acre)	12	7	7	AC
Physical yield (kg)				
First crop	1672	1103	1272	2.60*
Second crop	95	209	478	30.84***
Third crop	334	74	115	2.01NS
Total	2101	1386	1865	AC
Yield (kg/ac)	175	198	266	AC
Gross income (US$)				
First crop	1184	981	892	0.69NS
Second crop	49	96	282	45.36***
Third crop	124	48	72	1.98NS
Total income	1357	1125	1246	AC
Unit gross income (US$ / ac)	113	160	178	AC
Farm input expenditure (cost of production (US$))				
Planting material (first crop)	197	268	171	1.43NS
Planting material (second crop)	4	10	13	9.50***
Planting material (third crop)	2	3	5	1.58NS
Fertilizer	53	85	24	3.34**
Pesticides	6	11	3	10.37***
Jute twine	2	5	1	10.02***
Fuel pipes	6	21	3	14.34***
Fuel-wood	10	22	5	8.01***
Hired labour	29.39	32.70	13.26	AC
Family labour	13	28	9	7.43***
Total cost of production	322.39	485.70	247.30	AC
Unit cost of production (US$ / ac)	27	40.48	20.61	AC
Unit profit (US$ / ac)	86	119.52	157.39	AC

NS Not significant; * $p<0.1$; ** $p<0.05$; *** $p<0.01$; AC signifies Author calculations: Notes[2] Exchange rate: 1 US$ = 2760 UGX, 1 ZAR = 230 UGX in August 2014

As shown in the previous paragraph, cluster 2 farmers use more fertilisers and pesticides when compared to cluster 1 and 3. Farmers in cluster 1 are mainly tobacco growers who may be currently experiencing a decline in their crop yields due to the lack of crop rotations and, as mentioned earlier, a build-up of diseases in their fields. Cluster 3 with majorly cassava and coffee growers posts a higher unit profit per acre of US$157, compared to cluster 2 with US$ 119 and cluster 1, US$ 89. Less use of fertiliser, labour and pesticides, which translates into a lower cost of production, shows that the two crops (coffee and cassava) are feasible alternatives to tobacco. Tobacco, which often requires intensive labour and continuous fertilisation of the soils, has an incredibly high cost of production for cluster 2.

4. Discussion

Classification of farm households into various farm types based on their farm endowments, household attributes and farm output and revenue gained, has brought out differences that would otherwise be masked if farm households were to be discussed as a whole. Finding differences within farming systems is more delicate work and many months of field work would be necessary before objectively choosing predominant production systems. Available work using quantitative methods for creating typologies for farming systems with tobacco is limited. Several studies (Gaspar et al., 2008, Gelasakis et al., 2012, Milán et al., 2003, Usai et al., 2006, Riveiro et al., 2013) have classified livestock typologies i.e. sheep and goats but minimal studies are available for food grains and cash crops. Köbrich et al. (2003) identified farm typologies in Chile and in Pakistan with a combination of wheat and livestock. More recently, a cluster analysis was used by Chavez et al. (2010) to characterise tobacco farms based on their diversification factors in Argentina, and Goswami et al. (2014) used both hierarchical and non-hierarchical approaches to describe typologies based on food grains production systems in India.

In the present study, principal component analysis and cluster analysis successfully identified three clusters that together accounted for 100% of the farms with distinct farming systems. All factors explained differences among farming systems. Variables that were not strongly correlated with other variables were eliminated. Although the level of education does, in some cases, differentiate farm types, it was not used in the description of the clusters per se. West Nile farms (particularly those cultivating non-tobacco crops) were considerably smaller in farm size than their tobacco counterparts. Yet the same basic principles are evident; namely, the need for improved planting material, fertilisers and pesticides. Despite this fact, the different farm expenditures that exist within traditional farms, medium sized farms and small modern cassava farms were apparent in the present study.

There were significant differences especially along farm endowments, crop choices, yield and per-acre profitability. Traditional tobacco farms (Cluster 2) farms are well established and undoubtedly more complete economic units, but with improper use of farm inputs when compared to cluster 1 and 3. Cluster 2 represents farm types with small traditional tobacco and subsistence type of farms with considerably higher input use but lower crop yields. In terms of performance, this cluster is characterised by low gross returns but with a high cost of cultivation. Farms in cluster 1 have high farm input expenditures which do not translate into higher farm yields thus having a negative impact on the unit profit for the cluster. Virtually all farms in cluster 1 produce tobacco, thus this explains the low crop yields that are posted. Possible reasons why farmers in this cluster are experiencing lower crop yields could be that their soils have decreased in fertility due to a build-up of nematodes, which is also attributable to a lack of crop rotation and not spending enough money on nematocides. Farms in cluster 1 may opt to switch to coffee or cassava since crops do not require high input expenditures. Any tobacco control program that offers a reliable marketing strategy that competes fairly with tobacco while at the same time rehabilitating their soil fertility is likely to be attractive to these small sized, tobacco farms.

Medium sized mixed farms (Cluster 1) comprise 47% tobacco farms who diversify their resources into cultivating non-tobacco crops, and the rest (53%) are non-tobacco farms, and of these, 43% are previously tobacco growers. This farm type performs fairly well in their crop yields but poorly in their unit profit per acre. A significant difference between this farm type and the other two clusters is its unit profit, yet it projects high crop yields when compared to traditional tobacco (cluster 2) and modern cassava farms (cluster 3). This challenge roots from the low price per kilogram of the cluster's crops which translates to a lower gross income and ultimately low profits. Formulating strong price mechanisms that can improve their profit margins especially for non-tobacco crop would be encouraging to such farms.

In cluster 3, farms are well established and specialised coffee and cassava farms, capitalising on minimal fertiliser and pesticide use, and based on modern farm management practices such as the use of improved planting material. Education, access to finance and extension visits differentiate cluster 3 from cluster 1 and 2. The former representing a group of farms that have spent on average 15 years in an education institute and have access to farm cooperative finance. There is also a difference in input usage, where modern farms put more efforts in reducing farm inputs, and perhaps, farm operational costs. The difference in farm output and revenue for cluster 3 can be attributed to the saving option whilst maintaining higher crop yields. These farms clearly focus on stable cassava and coffee production.

The data presented here suggests that tobacco is by no means the only cash crop in West Nile. This is good news for tobacco control as it points to economically viable alternatives to tobacco. It is curious that coffee production has such a low prevalence and perhaps one of the recommendations from this study could be that the barriers to the expansion of that industry should be examined. It would be equally interesting to know what proportion of the stables grown by farmers in each of the clusters make it to the market. Finally, the next step would be to conduct a profitability analysis of maize and cassava compared to coffee or tobacco among different clusters, exploring and discovering the most beneficial levels of intensified management and on-farm input investment.

5. Implications for policy formulation

The application of farm typologies offers the opportunity to improve the efficiency of programmes that reduce the dependency on tobacco growing, through a greater understanding of the circumstances under which farmers, both tobacco and non-tobacco, are operating. This then provides the potential to tailor policies to specific needs and to target farmers' farming strategies. There are several ways in which these findings would support tobacco control policy formulation. In particular, the study findings establish that there are possibilities of alternative cropping systems for tobacco growers and emphasises the continued need for farm input investment into these cropping systems to encourage crop substitution.

6. Conclusions

Farm typologies are critical for representing the diversity in farm households and enabling the creation of policies that can effectively control supply of tobacco. Different policy interventions for promoting non-tobacco crops and reducing tobacco production may appeal to different production systems. The primary objective of this paper was to investigate the predominant farming systems using farm household economic data to determine how farm households should be segmented while implementing tobacco control policies. We used principal component analysis and cluster analysis to identify three groups of farming systems in West Nile. The results indicated considerable variation among clusters, confirming that much of the variation in input use and expenditure at farm level could be explained by the variation in cropping systems.

This farm typology provides multiple avenues for future research beyond the scope of the present research. One particular extension of this typology is to incorporate it into an economic profitability model. From the results presented here, one hypothesis is that each cluster in this typology will allocate their resources (in particular family labour) differently if current farming systems are changed. Another extension is to assess the economic risks, opportunities and possibilities of farmers to shift from a tobacco-biased system to non-tobacco farming system.

Finally, this research contributes new typologies of West Nile farm households to the literature. This typology identifies intuitive clusters of African farm households, which are based on household economic data. Policy makers can assess the impacts of tobacco control measures on how these African farm households allocate their resources and by their farm sales. Combining this new typology with current assessments of alternative livelihoods for tobacco farmers can provide a new richness in analysing the implications of tobacco control policy changes for various tobacco growers.

Acknowledgements

This work is financially supported by the Bill & Melinda Gates Foundation at the University of Cape Town under the *Economics of Tobacco Control Project*. The authors are grateful for the assistance of Dr Liberty Eaton and Dr Sebnem Er for their valuable contribution. The tobacco and non-tobacco farmers are also thanked for their willingness to take part in the interviews. Arua and Maracha district agricultural officers are thanked for their insightful discussions about the study area.

References

ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics,* 2, 433-459.

CHAVEZ, M., BERENTSEN, P. & LANSINK, A. O. 2010. Creating a typology of tobacco farms according to determinants of diversification in Valle de Lerma (Salta-Argentina). *Spanish journal of agricultural research*, 460-471.

FAOSTAT, F. 2015. Statistics Division. *Food and Agriculture Organization of the United Nations*.

GASPAR, P., ESCRIBANO, M., MESÍAS, F., DE LEDESMA, A. R. & PULIDO, F. 2008. Sheep farms in the Spanish rangelands (dehesas): Typologies according to livestock management and economic indicators. *Small Ruminant Research,* 74, 52-63.

GEIST, H. J., CHANG, K.-T., ETGES, V. & ABDALLAH, J. M. 2009. Tobacco growers at the crossroads: towards a comparison of diversification and ecosystem impacts. *Land Use Policy,* 26, 1066-1079.

GELASAKIS, A., VALERGAKIS, G., ARSENOS, G. & BANOS, G. 2012. Description and typology of intensive Chios dairy sheep farms in Greece. *Journal of dairy science,* 95, 3070-3079.

GOSWAMI, R., CHATTERJEE, S. & PRASAD, B. 2014. Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: study from coastal West Bengal, India. *Agricultural and Food Economics,* 2, 1-24.

HAIR, J. F., BLACK, W. C., BABIN, B. J., ANDERSON, R. E. & TATHAM, R. L. 2006. *Multivariate data analysis*, Pearson Prentice Hall Upper Saddle River, NJ.

HU, T.-W. & LEE, A. H. 2015. Commentary: Tobacco control and tobacco farming in African countries. *Journal of public health policy,* 36, 41-51.

IRAIZOZ, B., GORTON, M. & DAVIDOVA, S. 2007. Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain. *Agricultural Systems,* 93, 143-169.

JOSEPH, F., BLACK, W. C., BABIN, B. J. & ANDERSON, R. E. 2014. *Multivariate data analysis*.

KAISER, H. F. 1970. A second generation little jiffy. *Psychometrika,* 35, 401-415.

KETCHEN, D. J. & SHOOK, C. L. 1996. The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal,* 17, 441-458.

KÖBRICH, C., REHMAN, T. & KHAN, M. 2003. Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in Chile and Pakistan. *Agricultural systems,* 76, 141-157.

KOTRLIK, J. & HIGGINS, C. 2001. Organizational research: Determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal,* 19, 43.

KRAYBILL, D. & KIDOIDO, M. 2009. *Analysis of relative profitability of key Ugandan agricultural enterprises by agricultural production zone*, International Food Policy Research Institute (IFPRI).

MIDDLETON, J. 1965. The Lugbara of Uganda. *American Anthropologist,* 67, 1317-1318.

MIDDLETON, J. & GREENLAND, D. 1954. Land and population in West Nile district, Uganda. *Geographical Journal*, 446-455.

MILÁN, M., ARNALTE, E. & CAJA, G. 2003. Economic profitability and typology of Ripollesa breed sheep farms in Spain. *Small Ruminant Research,* 49, 97-105.

RIVEIRO, J., MANTECÓN, A., ÁLVAREZ, C. & LAVÍN, P. 2013. A typological characterization of dairy Assaf breed sheep farms at NW of Spain based on structural factor. *Agricultural systems,* 120, 27-37.

USAI, M., CASU, S., MOLLE, G., DECANDIA, M., LIGIOS, S. & CARTA, A. 2006. Using cluster analysis to characterize the goat farming system in Sardinia. *Livestock Science*, 104, 63-76.

WHO 2015. WHO global report on trends in prevalence of tobacco smoking. Geneva, Switzerland.