Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar
Spanish Journal of Agricultural Research 2010 8(2), 460-471
ISSN: 1695-971-X

Creating a typology of tobacco farms according to determinants of diversification in Valle de Lerma (Salta-Argentina)

M. D. Chavez[1,2]*, P. B. M. Berentsen[2] and A. G. J. M. Oude Lansink[2]

[1] *Instituto Nacional de Tecnología Agropecuaria (INTA). Estación Experimental Salta. Grupo de Estudios Económicos y Sociología Rural. Ruta Nacional 68, km 172. 4403 Cerrillos (Salta). Argentina*
[2] *Business Economics Group. Department of Social Sciences. Wageningen University. Hollandseweg, 1. 6706 KN Wageningen. The Netherlands*

Abstract

The objective of this article is to identify typical tobacco farms according to determinants of diversification that can be used to explore possibilities of diversification in the province of Salta (Northwest of Argentina). National Agriculture Census data of 278 farms in the main tobacco production area of Salta were used for the analysis. The variables selected concerning determinants of diversification were: land area, irrigation, general capital goods and specific capital goods, ownership of land, education, off-farm work, and labour availability. The analysis of the principal components applied to 16 selected variables allowed to reduce the dimensionality of the data to four components. Those components were used to apply K-means cluster approach to classify the farms. Four clusters were determined. Cluster 1 and Cluster 2 are the largest clusters. These concern highly specialized tobacco farms. They differ regarding determinants for diversification due to different levels of education of the farmer and different levels of off-farm work. Both clusters are interesting for further analysis regarding diversification alternatives to maintain or improve income and to reduce soil degradation. Cluster 3 concerns large tobacco farms being somewhat less specialized than the farms in Clusters 1 and 2. Intensive tobacco production makes this group interesting for exploring diversification alternatives to improve soil conditions. Farms in Cluster 4 already have a high level of diversification with substantial livestock production. The presence of perennial pastures suggests a better soil management than the other clusters. This cluster looks appealing to investigate what can be done regarding diversification.

Additional key words: clusters, determinants of diversification, income, *Nicotiana tabacum*, principal components, soil degradation.

Resumen

Tipología de explotaciones tabacaleras de acuerdo a determinantes de diversificación en el Valle de Lerma (Salta-Argentina)

El objetivo de este artículo es identificar explotaciones de tabaco de acuerdo a determinantes de diversificación para ser utilizadas en la exploración de posibilidades de diversificación en la provincia de Salta (Noroeste de Argentina). Se utilizaron para el análisis datos del Censo Nacional Agropecuario correspondientes a 278 explotaciones de la principal área de producción de tabaco de Salta. Las variables seleccionadas como determinantes de la diversificación fueron: superficie de tierra, capital general, capital específico, propiedad de la tierra, educación, trabajo extra predial y mano de obra. Se aplicó análisis de componentes principales a 16 variables seleccionadas y así se redujo la dimensión de los datos a cuatro componentes, que fueron utilizados en el análisis de conglomerados (K-means) para clasificar las explotaciones. Se determinaron cuatro conglomerados. Los conglomerados 1 y 2 son los más numerosos; están compuestos por explotaciones altamente especializadas en tabaco y difieren entre sí por el nivel de educación del productor y por el grado de trabajo extra predial que el mismo posee. Ambos grupos son interesantes para posteriores análisis relacionados a alternativas de diversificación para mantener o mejorar el ingreso y reducir la degradación del suelo. El conglomerado 3 consiste en grandes explotaciones tabacaleras menos especializadas que las anteriores. La intensa producción de tabaco hace a este grupo interesante para la exploración de alternativas de diversificación tendientes a mejorar las condiciones de suelo. Las explotaciones del conglomerado 4 tienen un nivel alto de diversificación con importante producción de ganado. Este grupo es atractivo para investigar lo que puede hacerse en relación a la diversificación.

Palabras clave adicionales: componentes principales, conglomerados, degradación de suelos, determinantes de diversificación, ingresos, *Nicotiana tabacum.*

* Corresponding author: dchavez@correo.inta.gov.ar
Received: 12-06-09; Accepted: 25-03-10.

Introduction

Tobacco (*Nicotiana tabacum* L.) is the most broadly produced non-food crop in the world and it is cultivated in more than 120 countries, as it can be grown under a wide range of climatic and soil conditions. The share of tobacco produced by developing countries increased from 57% in 1961 to 86% in 2006 (ITGA, 2008; Geist *et al*., 2009). Tobacco is a controversial crop not only because of the negative impact of smoking on health, but also because of environmental issues. In fact, soil degradation, deforestation and water pollution are part of the costs of tobacco production (Geist *et al*., 2009). The World Health Organization (WHO) recommended measures to control tobacco production and consumption within the Framework Convention on Tobacco Control (FCTC). Recommended measures aimed at reducing the demand for tobacco include, among others, price and tax measures, measures to protect non-smokers from exposure to tobacco smoke and a ban on advertising, promotion and sponsorship. Regarding the reduction of supply of tobacco, the WHO suggests, among others, the promotion of economically feasible alternatives for workers, growers and sellers. Besides, protection of the environment regarding tobacco cultivation is encouraged (WHO, 2003).

Tobacco production represents around a quarter of the total gross value of the agricultural production of the Salta province, in the Northwest of Argentina, and about 175,000 people depend on tobacco production for a living (Fittipaldi, 2004; Cámara de Tabaco de Salta, 2008). Salta produced 30% of the total tobacco production in the country in 2008. From 1989 to 2008, local tobacco production increased by 93%. Virginia tobacco represents 97% of the total production in Salta (MinAgri, 2009). Virginia is a flue-cured tobacco type that is dried in closed buildings with a heating system (ITGA, 2008). The tobacco cultivated area is mainly concentrated in the Valle de Lerma, in the centre of Salta. Cerrillos, Chicoana and Rosario de Lerma departments are the main producers in the Valley; they contribute 73% of the total production in tons in Salta (MinAgri, 2008).

Tobacco production has a relevant economic and social impact in Salta. However, also in this region, the negative environmental effects of intensive tobacco production, like soil degradation are recognized (Corvalán, 1997). In addition, tobacco farming involves production and price risk. Tobacco production is sensitive to temperature, precipitation and irrigation variations. The price of tobacco paid by the industry is completed so far by the national government. Future governmental support is uncertain because of international pressure to reduce tobacco production and consumption and because of the fact that the governmental price complement is the result of a political bargain process at national level (Fittipaldi, 2004).

The need for a diversification strategy for tobacco production in Valle de Lerma is widely recognised by national and provincial authorities and farming cooperatives (Fittipaldi, 2004). A first step required to be able to explore options for diversification is an inventory of existing tobacco growing systems.

Senthilkumar *et al*. (2009) suggest that a classification of farms to investigate future alternatives is needed, as it is not possible to conduct an exploration of every farm. The variables used in a typology depend on the aim of the research. In general, variables related to farm size, capital, labor, production model, soil quality and managerial skills are included to identify types of farming systems (Köbrich *et al*., 2003). Titonell *et al*. (2005) categorized farms according to resource endowment, production orientation, main constraints faced by farmers, position in farm cycle and main source of income. Andersen *et al*. (2007) classified farms with different environmental performance. Quantitative techniques have been applied to build typologies to understand the variety of farming systems (Köbrich *et al*., 2003; Milán *et al*., 2006; Nahed *et al*., 2006; Usai *et al*., 2006; Pardos *et al*., 2008; Senthilkumar *et al*., 2009).

This article aims at building a typology to identify typical tobacco farms according to determinants of diversification in the main departments for tobacco production in the Valle de Lerma. The results will provide representative farms which will be used in subsequent research to develop prospective models and evaluate potential diversification alternatives.

For the purpose of this article, the concept of diversification entails not only the number of farm activities but also the balance or share of them (Minot *et al*., 2006). Off-farm activities are excluded from the definition of diversification.

Abbreviations used: FCTC (Framework Convention on Tobacco Control), INDEC (National Institute of Statistics and Census), KMO (Kaiser-Meyer-Olkin), SID (Simpson index of diversity), WHO (World Health Organization).

Reasons for and determinants of diversification

Reasons for diversification

Literature shows a wide variety of reasons for diversification, but all of them can be summarized in two main reasons, namely risk reduction and improvement of income.

Risk reduction can be achieved when different sources of income have low or negative correlations. Thus, the diversification of farming activities may be a way to handle risk (Hardaker *et al*., 1997; Upton, 2004; Minot *et al.*, 2006).

An improvement in income may arise from scope economies. The concept of scope economies refers to cost savings due to joint production of products compared to costs of separate production. Cost savings were identified for different outputs in German dairy farms (Fernández-Cornejo *et al*., 1992). The shared use of inputs like labour, machinery and equipment led to cost savings in Dutch vegetable firms (Oude Lansink, 2001). Apart from scope economies, current literature reveals empirical evidence that suggests that diversification influences farmers' income positively (Bravo-Ureta *et al*., 2006). By building scenarios, Hengsdijk *et al.* (2007) found that diversification emerged as the most encouraging option to improve per capita income in traditional rice farms, compared to intensification, land expansion and exit from agriculture. Manos *et al*. (2009) observed that the implementation of alternative crops to a plan including tobacco can increase the income of farmers. Long distances to roads and markets can lead households to diversify into many activities to fulfill consumption needs. This way, transaction costs are saved (Barrett *et al*., 2001; Minot *et al.*, 2006). Another example is given by Sharma and Sharma (2005). Cost savings can be realized through a rice-wheat crop continuous growing system and replacing the use of fertilizer with the inclusion of a short duration pulse or replacing wheat or rice by other crops, which can be considered as diversification. A shift from food production for own consumption to a cash crop production contributes to improvement of income for smallholders (Minot *et al*., 2006).

Determinants of diversification

Determinants define the diversification possibilities of a farm. Land area, irrigation, capital goods, land ownership, age, education level, off-farm work and labour availability are considered determinants of diversification in current literature.

Total area of land is important in the case of arable farms. There is empirical evidence in current literature that the area of land has a positive effect on diversification (Benin *et al.*, 2004; Bravo-Ureta *et al*., 2006). The larger the area of land, the more motivated a farmer will be to devote part of it to introduce diversification.

Irrigation may have a negative influence on the decision to diversify the farm. An empirical analysis showed a positive relation between irrigation and the share of tobacco growing area at household level in India. These results suggest that irrigation does not encourage farmers to diversify (Panchamukhi, 2000).

The type of capital goods may have opposite effects on diversification. Specific capital goods may contribute to output specialization whereas general capital goods may facilitate diversification. For example, general machinery can be used more efficiently if used for different activities at different times of the year (Fernández-Cornejo *et al*., 1992; Hardaker *et al*., 1997). It can be expected that the availability of specific capital goods like tobacco curing barns, backpacks, grain machinery, and pasture machinery will prevent farmers from shifting to diversification. Conversely, general capital goods like tilling tools, tractors, sprayers and fertilizer drill, trucks and barns can motivate farmers to diversify.

Empirical data reveal very positive effects of land tenure on output diversification in Central America, suggesting that owners grow a wider variety of production items (Bravo-Ureta *et al*., 2006). A person who relies on rented land to produce will be limited in the decisions regarding land management (Caballero, 2001). The owner of the land may be more willing to experiment new activities to improve income in a medium or long term. Conversely, a farmer that rents the land may focus on making a profit in the short run.

The age of the farmer may affect diversification decisions. Empirical research found that the number of crops increase with the age of farmers, suggesting that they try new crops as they earn experience along their lives (Minot *et al*., 2006). The same was found within more diversified farms in West Midlands (UK). Farmers involved in more diversified farms have significant farming experience; a survey showed that 70% of them were over 45 years of age (Ilbery, 1991). The results of a survey carried out on growers in tobacco growing states in the southeast of the USA showed a negative relation between age and being interested in trying different activities from tobacco (Altman *et al*., 1996). The findings of another survey on tobacco farmers of North Carolina (USA) suggest that younger farmers are more interested in diversification while older growers are more likely to continue cultivating tobacco until they retire (Altman *et al*., 1998).

Education level has a strong and positive influence on the number of grown crops, stressing the importance of education and ability to understand information coming from extension services or other sources (Minot *et al.*, 2006). Bravo-Ureta (2006) found a positive effect of the average level of education for household members on diversification in Central America.

Labour factors can reflect the social structure and composition of farms and they could be determinants for taking decisions regarding diversification (Birthal *et al.*, 2006; Manos *et al*., 2009). Off- farm work may influence the decision to diversify. A farmer who works also outside the farm will probably be less disposed to be involved in many different production activities due to lack of time. Results of an empirical study suggest that farmers more occupied in other activity than agriculture are less expected to include high value crops because of lack of time and skills (Birthal *et al.*, 2006).

Labour will be used more efficiently if it can be allocated all along the year in a combination of activities (Hardaker *et al*., 1997). Economies of scope can arise from sharing labor for different outputs. Empirical data suggest that diversification in high-value crops is concentrated among households having enough labour supply (Birthal *et al.,* 2006). If labour supply is a problem, substitution of a high-value and labour intensive crop as tobacco by lower-value and lower labour crops can be a solution (Manos *et al*., 2009).

Data and methods

Study area

This study focuses on three departments with tobacco production in Valle de Lerma (24° 30’and 25° 38’ Southern latitude and 65° 22’and 65° 37’ Western longitude), in Salta, in the Northwest of Argentina. The valley is an extended plain between mountains and it has a temperate climate and the annual rainfall varies from 500 to 1,000 mm. Tobacco is grown on irrigated land (Baudino, 1996; Bravo *et al*., 1999). Next to tobacco as the main crop, vegetables, bean, corn, fruits, pastures, beef and milk cattle are products of the area. The departments are Cerrillos, Chicoana and Rosario de Lerma.

Description of data

The source of data for this study was the Agricultural Census carried out by the National Institute of Statistics and Census (INDEC, 2002). Although the following census was held in 2008, at the moment of submitting the final version of this paper, results from this census were not available yet. The reference period of the census comprises July 1st, 2001 to June 30th, 2002. To summarize, the variables show general information about the farm and the farmer, use of land, agronomic practices, stock of livestock, inventory of buildings, facilities, machinery, equipment and vehicles, permanent and temporary labour, forms of management and marketing channels.

The total number of farms in the study area was 641. Only farms that grow tobacco in Cerrillos, Chicoana and Rosario de Lerma departments were included in this study. After checking important missing values the final usable number of observations was 278.

Selected variables

The selected variables are developed from the original variables in the database that concern determinants for diversification. In total, 16 variables are included to identify types of tobacco farms to explore potential diversification (Table 1).

Principal components analysis

The objective of principal components analysis is the reduction of the dimensionality of the selected data. Data have to be correlated to successfully apply principal components analysis. Two tests are used in this article to verify the feasibility of the data for the analysis: the sphericity test and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The sphericity test developed by Bartlett (Lattin *et al*., 2003; SPSS, 2005) tests the null hypothesis that the correlation matrix of the population is the identity matrix (a perfectly spherical set of data). If so, data are inde-

Table 1. Selected variables to be used for principal components analysis

Name of the variable[a]	Description	Mean	St. deviation	Maximum
Land area				
— Suitable land[b]	Hectares	91.37	169.8	1,990
Irrigation				
— Irrigated area	Hectares	41.14	58.4	480
General capital goods				
— Tractors	Number	3.09	2.7	16
— Tilling tools	Number	3.53	2.7	17
— Trucks and other vehicles	Number	3.48	4.6	37
— Fertilizer drill	Number	0.52	0.7	6
— Sprayers	Number	0.83	1.4	14
— Barns	Number	2.64	2.0	12
Specific capital goods				
— Tobacco curing barns	Number	10.68	11.6	86
— Backpacks for spraying	Number	4.56	4.6	30
— Grains machinery	Number	0.49	0.8	5
— Pastures machinery	Number	0.24	0.9	6
Ownership of land				
— Land in property	Hectares	73.47	171.1	2,000
Education[c]				
— Education level of the farmers	= 1 more educated = 0 less educated	0.58	0.5	1
Off-farm work[d]				
— Farmers with work outside the farm	= 1 works = 0 does not work	0.05	0.2	1
Labor availability[e]				
— Permanent workers	Number	5.08	6.9	52

[a] Minimum value for all variables = 0. [b] Suitable land includes not only the cultivated land, but also natural forests and pastures land and apt but not used land. [c] The binary variable level of education of farmers takes the value 1 in when farmers have at least graduated from secondary school. It takes the value of 0 in case farmers have not graduated from secondary school. [d] The binary variable of farmers working outside the farm takes the value of 1 when farmers work outside the farm and 0 when farmers work in the farm exclusively. [e] The variable of permanent workers includes the number of workers that work every day during six or more months per year in the farm.

pendent. If the null hypothesis can be rejected it may be justified to use principal components for data reduction. The KMO test indicates the amount of variance in the variables that might be caused by principal factors. High values, close to 1, suggest that a factor analysis may be useful, and values less than 0.5 indicate the analysis is not helpful (SPSS, 2005).

In principal component analysis, the original variables are linearly combined in new variables which are called components. The first components explain as much of the available information as possible. Each component is uncorrelated with each other. There are different criteria that can be followed to decide the number of components to be retained. In this research, Kaiser's rule is followed. This criterion suggests keeping principal components with eigenvalues (variance of each component) larger than one (Köbrich *et al.*, 2003; Lattin *et al.*, 2003). The retained components are used in cluster analysis to determine types of tobacco farms to explore potential diversification. Statistical analyses were performed with SPSS 14.0 and 15.0 (SPSS, 2005, 2006).

Cluster analysis

Cluster analysis entails the division of a large group of observations into smaller and more homogeneous groups. A combination of a hierarchical method and a partitioning method for clustering is applied in this study. The hierarchical method is applied in an exploratory way and the solution is used in a partitioning method to improve the cluster solution (Sharma, 1996; Valeeva *et al*., 2005; Hair *et al*., 2006).

First, Ward's method, a hierarchical agglomerative method is applied. Ward's method seeks to achieve clusters with the smallest sum of squares within the cluster. This approach starts with each observation in a single cluster and in the following steps clusters are joined, until only one cluster contains all the observations. The graphical result of these steps is called dendrogram, which is a hierarchical tree structure (Köbrich *et al*., 2003; Lattin *et al*., 2003). The agglomeration schedule is another result of the hierarchical method. It shows the two clusters that are combined at each stage and the increase in heterogeneity that happens when two clusters are combined (Byrne, 1998; SPSS, 2005; Hair *et al.,* 2006).

The partitioning method following the hierarchical method is the K-means clustering. The goal of K-means method is to split the total number of observations into a prearranged number of *K* homogeneous groups based on preferred characteristics. The method can deal with big number of cases and it seeks to make distances within the group as short as possible. In this study, the prearranged number of clusters comes from the previous step. Kruskal-Wallis non parametric test was performed to examine whether the values of the selected variables vary between the groups (Lattin *et al*., 2003; SPSS, 2005; Valeeva *et al*., 2005).

A variable used to show current diversification of the farms in each cluster is the Simpson diversity index. The Simpson diversity index (SID) is a scalar number, ranging from 0 to 1, built from the area shares allocated to crops (including those crops devoted to livestock production, natural forests and pastures) and it shows both the number of crops and their relative presence (Benin *et al*., 2004). The value of the index is 0 in case of complete specialisation and approaches to 1 as the number of crops increases. The SID is calculated as follows:

$$SID = 1 - \sum_{i=1}^{n} P_i^2$$

where P_i is the proportionate area of i^{th} crop in the total cropped land (Joshi *et al*., 2003). Crops include cereals, tobacco, crops for seed production, pulses, annual pastures, perennial pastures, vegetables, flowers, aromatics, fruits, other crops, cultivated forests, nurseries, natural pastures and natural forests.

Results

Principal component analysis

KMO test and Bartlett's test were performed to test the suitability of the data to apply principal components analysis. KMO test result is 0.839 and Bartlett's test result is highly significant ($p = 0.000$) to reject the hypothesis of sphericity of multivariate data.

Principal components analysis was applied on the 16 selected variables as shows Table 1. Following Kaiser's rule, four components were selected. Table 2 shows the variance explained by the four extracted components.

Table 3 presents the rotated component matrix. This matrix shows the correlations (loadings) between each of the extracted four components and the original variables. It facilitates to establish what each component represents.

Table 2. Variance explained by four components using principal components analysis

Component	Initial eigenvalues		
	Total	% of variance	Cumulative %
1	7.034	43.964	43.964
2	1.462	9.136	53.099
3	1.235	7.718	60.817
4	1.105	6.907	67.724

The total column shows the amount of variance in the original variables accounted for by each component (eigenvalue). The column of percentage of variance presents the ratio of the variance accounted for by each component to the total variance of the entire variables. The cumulative column explains the percentage of variance accounted for by *n* components. The four components explain 67.724% of the total variance in the original variables. These components can be used to reduce the complexity of the data losing 32.276% of the information.

Table 3. Correlation of four components with initial variables using principal components analysis

Variables	Component			
	1[a]	2[b]	3[c]	4[d]
Suitable land	0.253	0.180	**0.919**	0.029
Irrigated area	**0.644**	0.422	0.389	0.040
Tractors	**0.810**	0.426	0.201	0.008
Tilling tools	**0.555**	**0.623**	0.169	–0.036
Trucks and other vehicles	**0.788**	–0.032	0.193	0.110
Fertilizer drill	**0.650**	0.261	0.072	0.235
Sprayers	0.303	**0.532**	0.178	0.218
Barns	**0.717**	0.138	0.214	–0.071
Tobacco curing barns	**0.835**	0.092	0.263	0.001
Backpacks for sparring	**0.684**	0.151	–0.148	0.059
Grains machinery	0.403	**0.578**	0.188	–0.098
Pastures machinery	0.009	**0.834**	0.053	0.100
Land in property	0.172	0.146	**0.946**	0.064
Education level of the farmers	0.029	0.257	0.066	**0.583**
Farmers with work outside the farm	0.073	–0.116	–0.003	**0.826**
Permanent workers	**0.658**	0.365	0.332	0.071

Extraction method: Principal Component Analysis. Rotation method: Varimax with Kaiser Normalization. Correlations above 0.5 are in bold. [a] The first component explains 43.964% of the variance and it is positively and highly correlated with tobacco curing barns and tractors. Since tobacco curing barns is less correlated with the other two components it represents better the component. [b] The second component (9.136% of variance) is related to pastures machinery, tilling tools and grains machinery and it can represent production activities that different from tobacco. [c] The third component (7.718% of variance) is correlated with suitable land and land in property and it represents the size and ownership of the farm. [d] The fourth component (6.907) is correlated with education level and work outside the farm and it represents characteristics of the farmer.

Cluster analysis

The four components were used for cluster analysis. First, Ward's method was applied. From this method a preliminary cluster solution was identified. The agglomeration coefficient and the dendrogram were used as stopping rules to choose the number of clusters. A large increase of the agglomeration coefficient suggests that two rather different clusters were combined. In Table 4 the agglomeration coefficients of the last stages of Ward's method are presented.

Table 5 presents the farm types that arise from the 4 clusters (K-means method). All the selected variables are significant at 0.001 level (Kruskal-Wallis non parametric test), suggesting that farm size, irrigation, general capital goods, specific capital goods, ownership of land, education, off-farm work and labor availability are useful for discriminating clusters with respect to determinants for diversification. Mann-Whitney Test was performed to compare clusters. Bonferroni adjustment for a 0.05 significance level was utilized. Results are given in the description of the clusters. Other variables are used for cluster description next to the initially selected variables, like cultivated area of main cash crops and annual and perennial pastures, number of heads of different livestock and the Simpson diversity index.

Cluster 1. Farms specialized in tobacco growing with a more educated farmer

This cluster represents 44% of the total farms. The education level of farmers is the highest of all. This group shows a high level of specialization, since the mean value for the SID is 0.17. Farms produce an average of 23 ha of tobacco, which is the lowest of the four clusters. Fifty seven per cent of the farms are farms specialized in tobacco growing (SID = 0). Those farms that are not specialized present also production of pulses, pastures, cereals, vegetables, other crops and livestock.

The variables of education level of farmers and off-farm work help to discriminate cluster 1 from Cluster 2 ($p = 0.000$). Farmers in Cluster 1 are much better educated and work outside the farm in some cases. With respect to Clusters 3 and 4, main differences arise when suitable land, irrigated area, capital goods, land in property,

Table 4. Agglomeration coefficient of Ward's cluster analysis of the last 10 stages

Number of clusters	Stage	Agglomeration coefficients	Percentage of change in agglomeration coefficient
10	268	225.043	10.19
9	269	255.011	13.32
8	270	288.635	13.19
7	271	334.451	15.87
6	272	393.106	17.54
5	273	455.182	15.79
4	274	596.078	30.95
3	275	756.743	26.95
2	276	927.096	22.51
1	277	1,108	19.51

The last column gives insight about the increase in cluster heterogeneity. The highest change in heterogeneity happens between stages 273 and 274. The agglomeration coefficient of 596.078 represents the heterogeneity when five clusters are reduced to four clusters. The significant jump when five clusters are combined in four clusters suggests the five-cluster solution as a potential cluster solution to be examined in the K-means cluster analysis. The dendrogram (not shown here because of its huge length) also suggests a possible solution of five clusters. Then, the number of clusters used in K-means method was 5. A single farm cluster was deleted from the description, ending with 4 clusters.

education level of farmers and permanent workers are compared ($p \leq 0.003$), except for pastures machinery when it is compared with Cluster 3.

Cluster 2. Farms specialized in tobacco growing with a less educated farmer

This is the largest cluster, representing 45% of the total number of farms. This group is the smallest in terms of suitable land. Farmers have the lowest level of education of all the clusters. All farmers in the group work exclusively at the farm. The SID is the second lowest of all (0.20). The mean value for tobacco area is 24 ha. Fifty six per cent of the farms in this cluster show a SID = 0 and they only grow tobacco. This cluster also produces cereals, pulses, pastures, vegetables, other crops and livestock.

In general, this cluster shows differences with Clusters 3 and 4 in terms of suitable land, irrigated area, availability of capital goods, land in property, education level of farmers and permanent workers ($p \leq 0.005$), except for pastures machinery when it is compared with Cluster 3 and for trucks and other vehicles when it is compared with Cluster 4. It is similar to Cluster 4 with respect to a full time devotion to the farm work.

Cluster 3. Large diversified tobacco farms

This cluster accounts for 3% of the total number of farms. It has the highest average values for many of the variables selected for sorting out the clusters. The mean value for the SID is 0.51. This group is the largest in tobacco production of the four. The mean value for the tobacco cultivated area is 176 ha. Full tobacco specialization is not found within the cluster. It is also the largest pulse and vegetable producer of all. They produce also calves and fatten livestock.

Differences with Clusters 1 and 2 mainly follow from suitable land, irrigated area, capital goods, land in property, education level of farmers and permanent workers ($p \leq 0.003$), except for pastures machinery. This cluster also differs from Cluster 2 with respect to off-farm work. The variables that show a higher power to discriminate this cluster from Cluster 4 include general capital goods like tractors, barns, trucks and other vehicles, and specific capital goods like tobacco curing barns and pastures machinery ($p \leq 0.002$).

Cluster 4. Highly diversified farms with important livestock production

This cluster comprises 8% of the total number of farms. This cluster shows the highest value in pastures machinery. Besides, farms grow annual and perennial pastures and present the highest number of heads of fatten and dairy livestock. The mean value of the SID is 0.67. The average value for tobacco cultivated area is 43.50 ha, being the second biggest tobacco producers of all the clusters. Full tobacco specialization is not found in this group.

This cluster differs from Cluster 1 and Cluster 2 in almost all the variables selected to discriminate groups (p-value 0.005 or lower), except for off-farm work and trucks and other vehicles when it is compared with Cluster 2.

Discussion

The combination of principal components analysis and cluster analysis was useful to discriminate four

Table 5. Mean values of variables to compare the different clusters

Variables	Cluster 1 N = 122	Cluster 2 N = 126	Cluster 3 N = 8	Cluster 4 N = 21
Land area				
Suitable land (ha)	71.61	49.64	402.38	247.64
Irrigation				
Irrigated area (ha)	26.39	29.94	208.88	128.26
General capital goods				
Tractors (n°)	2.33	2.71	11.13	6.76
Tilling tools (n°)	2.65	3.24	7.38	8.95
Trucks and other vehicles (n°)	2.65	2.93	22.00	4.67
Fertilizer drill (n°)	0.49	0.36	2.00	1.10
Sprayers (n°)	0.75	0.58	2.25	2.33
Barns (n°)	2.05	2.66	8.25	3.76
Specific capital goods				
Tobacco curing barns (n°)	7.80	9.66	56	16.48
Backpacks for spraying (n°)	3.80	4.25	14.63	7.19
Grains machinery (n°)	0.29	0.37	1.50	2.00
Pastures machinery (n°)	0.07	0.02	0.00	2.62
Ownership of land				
Land in property (ha)	59.15	32.34	329.38	214.21
Education				
Education level of the farmers	0.98	0.16	0.63	0.81
Off-farm work				
Farmers with work outside the farm	0.11	0.00	0.13	0.00
Labor availability				
Permanent workers (n°)	3.39	4.22	24.88	12.57
Current level of diversification				
Index of diversification	0.17	0.20	0.51	0.67
Crop production				
Cereals (ha)	1.66	2.58	6.25	14.76
Tobacco (ha)	22.87	24.30	176.25	43.50
Pulses (ha)	9.35	8.01	130.38	52.95
Pastures (ha)	2.69	3.31	23.63	78.57
Vegetables (ha)	0.52	0.82	8.25	2.38
Other crops (ha)	1.73	0.13	0.0	0.95
Livestock production				
Calves (n°)	0.78	4.90	32.00	26.52
Fatten livestock (n°)	0.57	1.25	30.25	62.14
Dairy livestock (n°)	1.61	1.33	0.00	83.38

clusters with respect to determinants of diversification. The results reveal that there is heterogeneity among tobacco farms regarding variables that define the possibilities of a farm for diversification in Valle de Lerma.

The results of this study provide a framework to analyze the problems of tobacco production and the possibilities for diversification. Besides the classification of farms according to determinants of diversification, this typology provides insight into the needs for diversification. The clusters recognized in this study will be useful to develop mathematic programming models concerning the analysis of diversification possibilities in the region. Developments from this work include the exploration of the impact of different production alternatives on farm income, production risk and soil quality of the identified types of tobacco farms in Valle de Lerma. Further research includes the development of diversification plans adapted to the different groups identified in this research.

Cluster 1 and Cluster 2 present the lowest level of diversification of the four clusters and they are highly specialized in tobacco production. Therefore they show the highest need for diversification. They differ mainly in the characteristics of farmer. Farmers in Cluster 1 are much better educated and, in some cases, they have another work in addition to the work in the farm. According to the literature, a more educated farmer will be in better conditions to pick up information regarding different crops and production activities (Bravo-Ureta *et al.*, 2006; Minot *et al.*, 2006). In contrast, off-farm work farm may prevent farmer to be involved in new and different activities (Birthal *et al.*, 2006). Both clusters have a good availability of suitable land, this being higher for Cluster 1. Availability of land may encourage diversification (Benin *et al.*, 2004; Bravo-Ureta *et al.*, 2006). Ownership of the land would encourage diversification of crops (Bravo-Ureta *et al.*, 2006). Cluster 1 has a higher availability of own land. Farms in Clusters 1 and 2 seem to have an acceptable level of general capital goods and specific capital goods for tobacco. General capital goods may contribute to diversification of outputs, while specific capital goods may encourage output specialization (Fernández-Cornejo *et al.*, 1992; Hardaker *et al.*, 1997).

Cluster 3 is the smallest in terms of number of farms and, in this sense, it is not very representative of the farms in the sample. However, the size of the farm, the highest level of tobacco cultivated area of all and a relative high level of diversification make this group interesting to analyze. Diversification in this group may contribute to reduce risk (Hardaker *et al.*, 1997; Upton, 2004; Minot *et al.*, 2006). Irrigated land is devoted mainly to grow tobacco. The intense tobacco production may imply a decrease in soil organic matter content and soil fertility (Corvalán, 1997). The problems of soil fertility may have an impact on the farm income. Consequently, there is scope to explore diversification to improve soil quality in Cluster 3.

Cluster 4 is the most diversified cluster and in this sense they may be reducing risk. They grow perennial pastures, suggesting that they have a better management of the soil than the others. Therefore, this group looks appealing to analyze alternatives of diversification for other clusters.

The selected variables were useful to discriminate clusters of tobacco farms. Nevertheless, it is worth noting some limitations and consequences of the selection of those variables. For example, it is inferred from the literature that irrigation does not encourage tobacco farmers to diversify (Panchamukhi, 2000). This statement is reasonable for Valle de Lerma, because tobacco is a profitable crop and farmers with more availability of water will try to grow more tobacco instead of other crops. This, however, does not imply that the provision of irrigation facilities would prohibit shifting away from tobacco. An encouraging plan, taken by the government and/or cooperatives is required to persuade farmers to shift away to other crops (Panchamukhi, 2000). Labour supply can motivate farmers to diversify to alternative production activities that are less labor demanding than tobacco. If this would be the case then some social consequences may arise. In this sense, Manos *et al.* (2009) found an increase of unemployment when tobacco was replaced by less labor demanding and more mechanized crops.

To summarize, this research found four different types of tobacco farms in Valle de Lerma regarding variables that can help to define possibilities for diversification. Two large clusters in terms of number of farms were identified. These farms are highly specialized in tobacco production. The fact that, in general, that these farmers do not have a work outside the farm suggests that they are highly dependent on farm income for living. Both groups are interesting for further research regarding diversification alternatives. Those alternative plans have to focus on maintaining or improving income and at the same time on reducing soil degradation. The other two clusters are smaller in terms of number of farms. The smallest cluster is formed by the largest tobacco producers. Intensive tobacco production makes this group appealing for exploring diversification alternatives to improve soil conditions. The last cluster is the most diversified. The presence of perennial pastures suggests a better soil management than the others clusters. This cluster looks appealing to investigate what can be done regarding diversification and to analyze the possibilities to introduce those alternatives to the other groups.

Acknowledgements

This research was supported by the Instituto Nacional de Tecnología Agropecuaria (INTA) from Argentina. The authors want to thank two anonymous reviewers and the journal's editor for their valuable comments to improve the first version of the article.

References

ALTMAN D.G., LEVINE D.W., HOWARD G., HAMILTON H., 1996. Tobacco farmers and diversification: opportunities and barriers. Tob Control 5, 192-198. doi: 10.1136/tc.5.3.192.

ALTMAN D.G., ZACCARO D.J., LEVINE D.W., AUSTIN D., WOODELL C., BAILEY B., SLIGH M., COHN G., DUNN J., 1998. Predictors of crop diversification: a survey of tobacco farmers in North Carolina (USA). Tob Control 7, 376-382. doi: 10.1136/tc.7.4.376.

ANDERSEN E., ELBERSENB B., GODESCHALK F., VERHOOG D., 2007. Farm management indicators and farm typologies as a basis for assessments in a changing policy environment. J Environ Manage 82, 353-362. doi: 10.1016/j.jenvman.2006.04.021.

BARRETT C.B., REARDON T., WEBB P., 2001. Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications. Food Policy 26, 315-331. doi: 10.1016/S0306-9192(01)00014-8.

BAUDINO G., 1996. Hidrogeología del Valle de Lerma, provincia de Salta. Doctoral thesis. Universidad Nacional de Salta, Argentina. 166 p.[on line]. Available in: http://www.unsa.edu.ar/natura/GBaudino/ [10 June, 2009] [In Spanish].

BENIN S., SMALE M., PENDER J., GEBREMEDHIN B., EHUI S., 2004. The economic determinants of cereal crop diversity on farms in the Ethiopian highlands. Agr Econ 31, 197-208. doi: 10.1016/j.agecon.2004.09.007.

BIRTHAL P.S., JHA A.K., JOSHI P.K., SINGH D.K., 2006. Agricultural diversification in North eastern region of India: Implications for growth and equity. Indian J Agr Econ 61(3), 328-340.

BRAVO G., BIANCHI A., VOLANTE J., ALDERETE SALAS S., SEMPRONII G., VICINI L., FERNÁNDEZ M., LIPSHITZ H., PÍCCOLO A., 1999. Regiones agroeconómicas del noroeste argentino, 7 pp. [on line]. Available in: http://www.inta.gov.ar/prorenoa/zonadescarga/regiones_agroeco/informe_regiones_agroec.pdf [9 June, 2009] [In Spanish].

BRAVO-URETA B.E., SOLÍS D., COCCHI H., QUIROGA R.E., 2006. The impact of soil conservation and output diversification on farm income in Central American hillside farming. Agr Econ 35, 267-276. doi: 10.1111/j.1574-0862.2006.00161.x.

BYRNE D., 1998. An example of doing a cluster analysis in a simple way with continuous data. Environmental and Planning, 6 pp. [on line]. Available in: http://www.dur.ac.uk/case.2004/workshop%20programmes/Cluster%20Analysis%20examplepdf [10 June, 2009].

CABALLERO R., 2001. Typology of cereal-sheep farming systems in Castile-La Mancha (south-central Spain). Agr Syst 68, 215-232. doi: 10.1016/S0308-521X(01)00009-9.

CÁMARA DEL TABACO DE SALTA, 2008. La producción de tabaco en la República Argentina. Presentation. 16 pp. [In Spanish].

CORVALÁN E., 1997. Niveles de fertilidad de los suelos agrícolas de Salta. INTA. Panorama Agropecuario 50, 34-38 [In Spanish].

FERNÁNDEZ-CORNEJO J., GEMPESAW C.M., ELTERICH J.G., STEFANOU S.E., 1992. Dynamic measures of scope and scale economies: an application to German agriculture. Am J Agr Econ 74, 329-242. doi: 10.2307/1242487.

FITTIPALDI C.R., 2004. Análisis de una estrategia de diversificación de la empresa tabacalera. Master's thesis. Escuela de Negocios, Universidad Católica de Salta, Argentina. 153 pp. [In Spanish].

GEIST H.J., CHANG K., 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. doi: 10.1016/j.landusepol.2009.01.003.

HAIR J.F., BLACK W.C., BABIN B.J., ANDERSON R.E., TATHAM R.L., 2006. Multivariate data analysis, 6th ed. Person Prentice Hall. 899 pp.

HARDAKER J.B., HUIRNE R.B.M., ANDERSON J.R., 1997. Coping with risk in agriculture. CAB Intnl. 274 pp.

HENGSDIJK H., GUANGHUO W., VAN DEN BERG M.M., JIANGDI W., WOLF J., CHANGHE L., ROETTER R.P., VAN KEULEN H., 2007. Poverty and biodiversity trade-offs in rural development: A case study for Pujiang county, China. Agr Syst 94, 851-861. doi: 10.1016/j.agsy.2006.11.018.

ILBERY B.W., 1991. Farm diversification as an adjustment strategy on the urban fringe of the West Midlands. J Rural Stud 7(3), 207-218.

INDEC, 2002. Censo Nacional Agropecuario-Database. Instituto Nacional de Estadística y Censos. [In Spanish].

ITGA, 2008. Tobacco types. International Tobacco Growers' Association. [on line]. Available in: http://www.

tobaccoleaf.org/about_tobacco/index.asp?op=2. [10 November, 2009].

JOSHI P.K., GULATI A., BIRTHAL P.S., TEWARI L., 2003. Agriculture diversification in South Asia: patterns, determinants, and policy implications. International Food Policy Research Institute. Markets and Structural Studies Division. Discussion paper Nº 57, 51 pp. [on line]. Available in: http://www.ifpri.org/publication/agriculture-diversification-south-asia [5 November, 2009].

KÖBRICH C., REHMAN T., KHAN M., 2003. Typification of farming systems for constructing representative farm models: two illustrations of the application of multivariate analyses in Chile and Pakistan. Agr Syst 76, 141-157. doi: 10.1016/S0308-521X(02)00013-6.

LATTIN J.M., CARROLL J.D., GREEN P.E., 2003. Analyzing multivariate data. Duxbury Applied Series. Thomson. Brooks/Cole. 525 pp.

MANOS B., BOURNARIS T., PAPATHANASIOU J., CHATZINIKOLAOU P., 2009. Evaluation of tobacco cultivation alternatives under the EU common agricultural policy (CAP). J Policy Model 31, 225-238. doi: 10.1016/j.jpolmod.2008.07.001.

MILÁN M.J., BARTOLOMÉ J., QUINTANILLA R., GARCÍA-CACHÁN M.D., ESPEJO M., HERRAIZ P.L., SÁNCHEZ-RECIO J.M., PIEDRAFITA J., 2006. Structural characterisation and typology of beef cattle farms of Spanish wooded rangelands (dehesas). Livest Sci 99, 197-209. doi: 10.1016/j.livprodsci.2005.06.012.

MINAGRI, 2008. Tabaco. Producción y mercados. Mapas a nivel departamental por provincia 2006/2007. Ministerio de Agricultura, Ganadería y Pesca. Presidencia de la Nación. Agricultura [on line]. Available in: http://www.minagri.gob.ar/SAGPyA/agricultura/tabaco/02=produccion_y_mercados/01-interno/produccion/xprov_0607/mapa_ porprovincia0607.php [18 March, 2010] [In Spanish].

MINAGRI, 2009. Tabaco. Producción y mercados. Ministerio de Agricultura, Ganadería y pesca. Presidencia de la Nación. Agricultura [on line]. Available in: http://www.minagri.gob.ar/SAGPyA/agricultura/tabaco/02=produccion_y_mercados/01-interno/ [18 March, 2010] [In Spanish].

MINOT N., EPPRECHT M., ANH T.T.T., TRUNG L.Q.T., 2006. Income diversification and poverty in the Northern Uplands of Vietnam. Washington, D.C. Intl Food Pol Res Inst. Research Report Nº 145. 149 pp. [on line]. Available in: http://www.ifpri.org/pubs/abstract/145/rr145.pdf [10 June, 2009].

NAHED J., CASTEL J.M., MENA Y., CARAVACA F., 2006. Appraisal of the sustainability of dairy goat systems in Southern Spain according to their degree of intensification. Livest Sci 101, 10-23. doi: 10.1016/j.livprodsci.2005.08.018.

OUDE LANSINK A., 2001. Long and short term economies of scope in Dutch vegetable production. J Agr Econ 52(1), 123-138. doi: 10.1111/j.1477-9552.2001.tb00913.x.

PANCHAMUKHI P.R., 2000. Agricultural diversification as a tool of tobacco control. The World Health Organization (WHO) International Conference on Global Tobacco Control Law: Towards a WHO Framework Convention on Tobacco Control. 48 pp [on line]. Available in: http://www.who.int/tobacco/media/en/PANCHIMUKHI2000X.pdf [10 June, 2009].

PARDOS L., MAZA M.T., FANTOVA E., SEPÚLVEDA W., 2008. The diversity of sheep production systems in Aragón (Spain): characterisation and typification of meat sheep farms. Span J Agric Res 6(4), 497-507.

SENTHILKUMAR K., BINDRABAN P.S., DE BOER W., DE RIDDER N., THIYAGARAJAN T.M., GILLER K.E., 2009. Characterising rice-based farming systems to identify opportunities for adopting water efficient cultivation methods in Tamil Nadu, India. Agr Water Manage 96, 1851-1860. doi: 10.1016/j.agwat.2009.08.007.

SHARMA S., 1996. Applied multivariate techniques. John Wiley & Sons, Inc. 493 pp.

SHARMA S.K., SHARMA S.N., 2005. Effect of crop diversification of rice-wheat cropping system on productivity and profitability. J Sustain Agr 26(1), 39-48. doi: 10.1300/J064v26n01_05.

SPSS, 2005. SPSS for Windows. SPSS Inc, Chicago, IL.

SPSS, 2006. SPSS for Windows. SPSS Inc, Chicago, IL.

TITTONELL P., VANLAUWE B., LEFFELAAR P.A., ROWE E.C., GILLER K.E., 2005. Exploring diversity in soil fertility management of smallholder farms in western Kenya I. Heterogeneity at region and farm scale. Agr Ecosyst Environ 110, 149-165. doi: 10.1016/j.agee.2005.04.001.

UPTON M., 2004. The role of livestock in economic development and poverty reduction. Pro-Poor Livestock Policy Initiative (PPLPI). Food and Agriculture Organization (FAO). Working paper Nº 10. 66 pp. [on line]. Available in: http://www.fao.org/ag/againfo/programmes/en/pplpi/docarc/wp10.pdf [10 June, 2009].

USAI M.G., CASU S., MOLLE G., DECANDIA M., LIGIOS S., CARTA A., 2006. Using cluster analysis to characterize the goat farming system in Sardinia. Livest Sci 104, 63-76. doi: 10.1016/j.livsci.2006.03.013.

VALEEVA N., MEUWISSEN M., BERGEVOET R., OUDE LANSINK A., HUIRNE R., 2005. Improving food safety at the dairy farm level: farmers' and experts' perceptions. Rev Agr Econ 4, 574-592. doi: 10.1111/j.1467-9353.2005.00265.x.

WHO, 2003. WHO Framework convention on tobacco control. World Health Oganization, Geneva, Switzerland [on line]. Available in: http://www.who.int/fctc/en/ [3 November, 2009].