DIFFERENTIATION OF FARM TYPES IN SPAIN: PROCEDURE AND RESULTS

José Antonio Riveiro-Valiño

Emilio Rafael Díaz-Varela

Manuel Francisco Marey-Pérez

Carlos José Álvarez-López

Universidad de Santiago de Compostela

Departamento de Ingeniería Agroforestal - GI- 1716 "Proyectos y planificación"

Abstract

In this paper, a methodology for the classification of farms into different, representative farm types is presented. This methodology improves a combinatorial method developed in our previous research that was limited to the region of Galicia, Spain. Agricultural census microdata was used as the only source of information, particularly the 1999 census (the latest census available). The combinatorial method was improved by extending its scope to the entire Spanish country and by integrating Principal Component Analysis. The improved procedure and the results for a reference land use type are presented. Applying such a methodology to various censuses would provide, in addition to a static picture for each year, an analysis of the temporal evolution of agricultural production structures.

Keywords: *Agricultural census, agricultural production structures, Principal Component Analysis*

1. Introduction

The Treaty of Rome (the Treaty establishing the European Economic Community in 1957) laid the foundations of the CAP. According to the Treaty, the specific objectives of the EC, named EU after the Treaty on European Union or Maastricht Treaty (1992) were to increase agricultural productivity, to ensure a fair standard of living for farmers, to stabilize markets, to assure the availability of supplies and to ensure reasonable prices for consumers.

The management of the CAP requires an information flow from Member States. The Statistical Office of the European Communities (Eurostat) is responsible for providing the European Union with statistical data at European level and for promoting the harmonization of the statistical methodology among Member States. The statistics provided by Eurostat are organized into nine main themes, among which agriculture and fisheries.

In Spain, statistical data are compiled by the National Statistics Institute (*Instituto Nacional de Estadística-INE*). The INE compiles an agricultural census every ten years, thus providing objective and comparable information about agriculture and its evolution. In addition, the information contained in the agricultural census can be used to monitor the situation, which is essential to guide the CAP.

The methodology of the Spanish Agricultural Census allows for the definition of relatively homogeneous types of farms based on two main characteristics: the gross margin of farms (SGM), which determines the economic size of farms, and the type of farming (TF), which is defined in terms of the relative importance of the different enterprises on the farm, measured as a proportion of each enterprise's SGM to the farm's total SGM.

Each TF defines a farm type and is designed as a statistical analysis tool that helps meet the needs of the CAP insofar as it allows for: a) the analysis of the situation of farms based on economic criteria, b) the comparison of the situation of farms based on farm type, period, and Member State or region, and c) the identification of relationships between the economic size of farms and farm physical size, farm labour or other factors (INE-Metodología, 1999).

Such a typology of farms can meet the needs of the CAP but is not suited to other technical analyses or other scales of analysis because the classification of TFs is not derived from the production structure of farms. Furthermore, the farms included in one TF can be subjected to constraints derived from different production systems.

The methodology of the Agricultural Census for the classification of farms is not the only methodology available. Other methodologies, aimed at different purposes, work with a number of data sources and scales of analysis. Generally, the scale of analysis and the methodology are related to the data source used and the objectives pursued.

Perrot and Fraysse (2002) used agricultural census data to characterize and classify livestock production systems in France at the national level based on the analysis of the combination of productions, whereas Zaldivar and Menacho (1991) and Gaspar et al. (2007, 2008) used Principal Components Analysis (PCA).

Many authors used cluster analysis to define homogeneous groups of farms based on data sampling or data collection through surveys. Most of them combined cluster analysis with other techniques that helped improve the results (Hardiman et al. 1990), (Kobrich et al. 2003), (Milan et al. 2006), (Tiffin 2006), (Usai et al. 2006).

Yet, using such techniques to classify farms into groups can be problematic because of multicollinearity among variables (Iraizoz et al., 2007). To overcome such problems, Ketchen and Shook (1996) proposed combining principal component analysis with orthogonal rotation, a technique used by Pardos et al. (2008).

Other less common statistical analysis techniques, such as multiple correspondence analysis, allow for the identification of groups of farms that show similar characteristics for the parameters considered (Castell et al. 2003).

In the agricultural sector, making decisions at different levels (political, technical or administrative, among others) requires information on the structure and operation of farms, such as data of farm evolution or of the products and resources used on the farm. In Galicia, Spain, Alvarez et al. (2008) and Riveiro et al. (2008) used agricultural census data to better know agricultural production systems and their evolution. Their research focused on the development of flexible methodologies to identify homogeneous farm groups based on the production structure of farms. Farm groups were identified at a regional level and proved useful in agricultural planning research. In contrast to other research works cited, the data source used here was the total population of farms included in the census. Agricultural census data disaggregated at farm level were used. Moreover, such methodologies were validated using statistical models such as discriminant analysis (Riveiro et al. 2009).

2. Objectives

This paper presents the progress in the development of a methodology for the classification of agricultural production systems at the national level. The clustering procedure is based on the production structure of farms. To this end, our methodology takes advantage of the flexibility of the combinatorial model proposed by Riveiro et al, 2008, which showed suitable at the regional level, and of the potentialities of multivariate statistics. We present an innovative combination of techniques that reduces the subjectivity of the clustering methods and allows for easier and more reliable grouping. Another innovation introduced in our methodology is that the procedure is based exclusively on agricultural census data. Using agricultural census data allows for the analysis of the whole population of farms without sampling.

The research presented in this paper is driven by the need to determine the agricultural production structure of the territory as a basis for decision-making at different levels. Because the homogeneous groups of farms identified show similar characteristics, decision-makers will be able to use a common approach for each group of farms.

The methodology presented in this paper assumes that the presence of similar combinations of land uses or activities on farms is equivalent to similar production systems and, therefore, to similar farm types. Therefore, the specific objective of this paper is to define large farm types based on a given reference activity using a public data source–the Agricultural Census–in order to provide a solid basis for the assessment of the different production systems and to facilitate decision-making in different areas.

3. Methods

Spain has an area of almost 506 000 km². The climate of Spain is a consequence of its latitude (34º - 44º N), its mountainous terrain, and the Atlantic and Mediterranean influences. More than 20 types of climate are found in Spain, among which the Sub-tropical climate of the Canary Islands, the Continental Mediterranean climate, which is the most common type of climate in the peninsula, the Oceanic climate of the westernmost areas of the peninsula, the Mountain climate or the Semi-arid climate of some areas of the peninsula. Such a variety of climates, characterized by a wide temperature range (average temperatures range from 3 to 19ºC) and a marked contrast in annual rainfall values (from 190 to 2400 mm), allows for the presence of a wide variety of agricultural crops (forage crops, vegetables, cereals, fruit trees or even tropical crops), and particularly potato, over a Utilized Agricultural Area (UAA) of 25 million HA (INE-EEEA, 2007).

The agricultural census is the main data source for farm statistics and is compiled every ten years. The most recent agricultural census in Spain is the 1999 Agricultural Census (INE, 1999). The microdata included in the agricultural census provide information pertaining to the geographical location of the farms, the land uses or farming activities present on farms, and the size of each land use or farming activity, among other data that are not used in this paper.

According to the 1999 Agricultural Census, there are 1 790 162 farms in Spain, among which 77 306 have at least one dairy cow. The population of farms considered in this paper included dairy farms with at least 18 dairy cows, which amount to 22 740 farms all over the country.

The agricultural census identifies each farm through a unique code that allows for the identification of the municipality in which the farm is located. In addition to other variables, each farm is characterized by the agricultural land uses or activities present on the farms. Data of the area used for each crop or of the herd size of each farm is derived from the agricultural census, which includes 163 different land uses or activities. Some of these land uses or activities are negligible, while other can be grouped based on their common characteristics. By grouping the land uses or activities that share some characteristics, the number of land uses can be reduced to 54. This data provides the basis for the analysis performed and presented below.

The methodology of analysis presented here integrates some of the aspects of the combinatorial model (Riveiro et al. 2008) with PCA.

The combinatorial model is based on an algorithm that generates combinations of land uses or activities according to a number of constraints by recording the number of cases (farms) for which the same combination is repeated. The algorithm generates combinations of different order (number of land uses or activities in common) in successive stages, such that a combinatorial tree is obtained. The analysis of the tree, which is based on technical knowledge of the farming systems that incorporate each land use or activity, provides the different types of farms.

The combinatorial methodology loses effectiveness at larger spatial scales because it is based on the knowledge of the characteristics and performance of the production systems involved, and such knowledge is poorer at larger scales.

The different types of farms are defined based on the presence of strong correlations between the variables that characterize each type of farm, which allows for the definition of homogeneous groups. For this reason, an interdependence technique, namely Principal Component Analysis (PCA) was used. PCA is defined as a linear transformation that transforms the data to a new coordinate system to reduce the dimensionality of a data set (number of variables), without much loss of information. Such a transformation generates a reduced number of new factors or ‘principal components’ that are a linear combination of the original variables but independent of each other.

The methodology developed to identify farm groups (types) at the national level used a Combinatorial Method and PCA, and consisted of three basic steps:

1. The combinatorial model was applied to the population of farms in order to obtain the possible first-order combinations.
2. PCA was performed on the results obtained for first-order combinations.
3. The groups (types) of farms were selected based on the analysis and interpretation of factor loadings.

The combinatorial model was applied in the first step in order to obtain the first-order combinations, i.e. the combinations of two land uses or activities, according to the methodology reported by Riveiro et al. (2008). The reference land use or activity was combined with each secondary land use or activity. In this case, the reference land use was dairy cattle and the secondary land use was each of the rest of land uses or activities included in the agricultural census.

Each variable (land use or activity) allowed for the determination of the total number of farms comprised in each combination within a defined area. Each combination must satisfy some restrictions: 1) the secondary land use or activity in the combination must have a minimum size (cropping area or herd size). Such a restriction is particular to each land use or activity and is represented by a value used to distinguish production for own-consumption from commercial production. 2) Each combination must include a minimum number of farms in order to be recorded. The minimum number of farms included in a combination was set at 1% of the total number of farms.

By using the procedure described above, a list of combinations was generated and the farms included in each combination were counted. A table with 50 rows and a variable number of columns was created. Each row corresponded to one of the 50 territorial spaces considered (provinces) and each column corresponded to the combinations generated from the secondary land uses or activities. The values at each row/column intersection suggested the number of farms that showed each combination of land uses or activities in each province.

Under the assumption that a similar combination of land uses or activities on farms was an indicator of similar farm types, PCA was performed on the data generated by the combinatorial model.

First, a correlation matrix was generated and analyzed for all the variables. The variables used in the analysis were the combinations generated by the combinatorial method. PCA is useful only if there are strong correlations between variables. In this case, the determinant of the matrix was computed because a very low value of the determinant usually indicates correlations between variables. In addition, a Bartlett's test of sphericity was performed to reject the null hypothesis.

Second, the eigenvalues were computed in order to derive the factors (principal components) that accounted for a significant percentage of variability. The analysis of the correlation matrix allowed for the selection of the homogeneous groups and the definition of distinct farm types.

Because each factor loading represented a relative coordinate of each variable in the new reference coordinate system, the nearest values of the factors suggested proximity to a cluster. Thus, the variables were grouped based on minimum distances, interpreted according to the graphs in figure 1. The groups of variables (land uses or activities) obtained represented groups of farms.

After having defined the land uses that characterized each group, we identified the farms that belonged to each group and determined their spatial distribution. The territory that corresponded to each group had to meet a number of characteristic environmental parameters depending on the demands and requirements of the land uses or activities involved.

4. Results and discussion

The population of farms analyzed in this paper comprises 22 740 dairy farms with at least 18 dairy cows. The analyzed farms are heterogeneously distributed across Spain.

Table 1 shows the results for the first-order combinations obtained by using the combinatorial model.

Variable	No. Farms	Percentage
Grasslands	16 775	73.8%
Forage corn	6332	27.8%
Barley and Oats	3479	15.3%
Other pastures	2873	12.6%
Alfalfa and other legumes	2363	10.4%
Green fodder	2163	9.5%
Corn	2141	9.4%
Wheat	1878	8.3%
Sheep and Goats	1236	5.4%
Beef cattle	1010	4.4%
Breeding sows	692	3.0%
Horses	488	2.1%
Pig fattening	458	2.0%

Table 1. First-order combinations

From among a total of 53 complementary land uses, ‘dairy farming’ is combined with only the thirteen land uses or activities included in table 1, which correspond to the land uses or activities present in at least 1% of farms (228 farms). The most frequent land use or activity with the minimum size that is combined with dairy farming is forage production for feeding cattle (grasslands, silage corn and, to a lesser extent, alfalfa or other forage crops). The second most frequent land use or activity is cereal production (barley, oats, corn or wheat), either for feeding cattle or as a complementary land use or activity. Finally, the third most frequent land uses or activities are complementary activities that are independent of dairy farming (sheep and goat farming, beef cattle farming and pig farming, among others). In addition, Table 1 includes (V.14) the farms in which the area used for each possible forage crop is below a preset minimum value, i.e. the farms with no land available for forage production.

Based on table 1, which summarizes the total number of farms included in each combination, we calculated the distribution of the farms across the 50 spatial units considered (provinces), and generated a matrix with 50 rows and 13 columns that was subjected to principal component analysis (PCA). Rows represent cases and columns represent the variables that characterize each case.

Once PCA is performed, the results obtained for the correlation matrix suggest the occurrence of strong correlations between variables for many combinations. In addition, the determinant of this matrix takes a very small value (3.212x10-10), which indicates the occurrence of such strong correlations. Moreover, the Bartlett’s test of sphericity shows that the null hypothesis must be rejected because the level of significance is 0.000 (below 0.050, which suggests that the cloud of points does not fit a perfect sphere and, therefore, the variables are correlated. Consequently, PCA is the appropriate technique to analyze the data available.

Table 2 shows the initial eigenvalues obtained to explain up to 100% of the variance. The use of the first three factors can explain 81.862% of the variance. A fourth factor explains 8.917% of the variance but makes the interpretation of results difficult. For this reason, we have considered the first three factors as principal components.

Factor	Initial eigenvalues		
	Total	% Variance	Cumulative %
1	5.834	44.875	44.875
2	3.304	25.418	70.292
3	1.504	11.570	81.862
4	1.159	8.917	90.779
5	0.482	3.707	94.486
6	0.295	2.271	96.757
7	0.158	1.219	97.976
8	0.108	0.828	98.804
9	0.064	0.491	99.295
10	0.052	0.404	99.699
11	0.025	0.192	99.890
12	0.009	0.069	99.960
13	0.005	0.040	100.000

Table 2. Eigenvalues

Table 3 shows the factor matrix computed by maximum likelihood from the correlation matrix for three principal components.

Variable	Extracted factors		
	F1	F2	F3
Grasslands	0.740	0.672	-0.001
Forage corn	0.714	0.566	0.079
Barley and Oats	0.360	-0.401	0.792
Other pastures	0.300	-0.524	0.749
Alfalfa and other legumes	0.374	0.193	0.715
Green fodder	0.782	-0.015	0.290
Corn	0.674	0.603	-0.007
Wheat	0.281	-0.307	0.812
Sheep and Goats	0.805	0.398	0.225
Beef cattle	0.377	-0.121	0.201
Breeding sows	0.735	-0.677	-0.014
Horses	0.722	-0.567	-0.093
Pig fattening	0.553	0.407	0.220

Table 3. Factor matrix

The graphs in figure 1 allow for the definition of the groups of farms from the variables used to characterize the groups (land uses or activities).

Gr1 is composed of dairy farms strongly associated with forage production, with presence of grasslands or even other pastures and forage corn, and hardly any complementary land use or activity. This type of farm is located mainly in northwest and north Spain, which is the most suitable area for forage production.

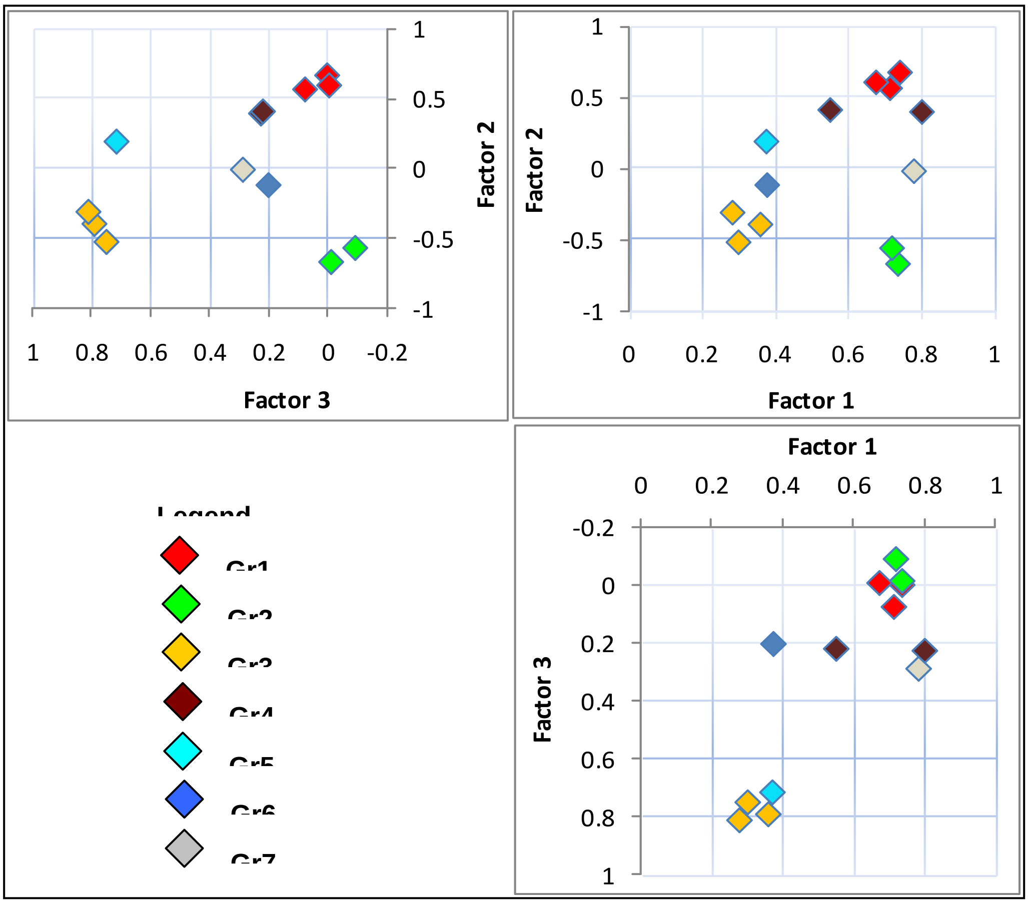

Figure 1. Interpretation of groups

Gr2 is composed of dairy farms associated with pig breeding or fattening. The spatial location of the farms in Gr2 corresponds to two distinct areas: rangelands and areas of Cataluña and Galicia with intensive farming. Actually, this group could be divided into two subgroups at other scales of analysis.

Gr3 includes dairy farms associated mainly with cereal production (wheat, barley, oats) and, in some cases, with the production of alfalfa or other leguminous plants. The farms in Gr3 are located in wide plains with low precipitation levels and a long cold period (Castilla, a part of Cataluña or Aragón, among others).

Gr4 is composed of dairy farms associated with beef cattle and horse breeding. Similarly as for the farms in Gr1, the farms in Gr4 are associated with forage production, and are located mainly in northwest and north Spain and in some areas of northeast Spain. Depending on the scale of analysis, groups Gr1 and Gr4 could even be merged. Although their location is very similar, merging both groups would not be appropriate because of the differences in the production structure characteristic of both types of farms.

Group Gr5 includes all the dairy farms that have corn production in common. Corn production shares some requirements with cereal production (sun exposure requirements), but shows stronger crop water requirements. Because of crop requirements, Gr3 and Gr5 are near each other. The farms in Gr5 are located in Northwest Castilla-León and A Coruña. Although we were not aware of the presence of corn in A Coruña, the agricultural census data reveals the presence of corn crops in A Coruña, probably because forage corn is considered as corn grain.

Gr6 is composed of dairy farms with presence of sheep or goats. This type of farm is located in two distinct areas: humid and mountainous areas in the Cantabrian coast and rangelands in Extremadura and West Castilla-La Mancha with low precipitation levels and capacity for forage production, and low numbers of livestock units per ha. Sheep and goats use resources that dairy cattle would not use.

Gr7 is composed of dairy farms that have the production of green fodder in common. The distance between Gr7 and groups Gr1 and Gr4 is short, while the distance between Gr7 and the rest of groups is long. The spatial distribution of the farms in Gr7 coincides mostly with the spatial distribution of farms in Gr1 and Gr4, and partly coincides with other groups. The farms in Gr7 are located mainly in two opposite areas: Northwest Spain, where green fodder production is complementary to other types of forage production, and Northeast Spain, where green fodder replaces other types of forage production.

5. Conclusions

We obtained seven clearly differentiated groups of farms (types) with different levels of structural complexity. Only Gr4 tends to be confused with Gr1 in terms of both their production structure and their spatial location. The rest of groups show distinct preferential spatial locations.

Based on these results, we can conclude that different farm types matching different production structures can be defined using agricultural census microdata as the sole source of information and jointly considering the combination of land uses or activities on the farms and their spatial location.

The possibility to define differentiated types of farms allows us to analyze groups of farms and to adopt common solutions, useful in decision-making in different areas.

Since agricultural censuses are compiled every ten years, the types of farms and the analysis of these farms can be updated, and the historical evolution of the different types of farms can be analyzed.

References

Álvarez C.J., Riveiro J.A., Marey M.F. “Typology, classification and characterization of farms for agricultural production planning”. *Spanish Journal of Agricultural Research*. Vol. 6 (1), 2008, pp. 125-136.

Castel , J.M., Menaa Y., Delgado-Pertíñez M., Camúñez J., Basulto J., Caravaca F., Guzmán-Guerrero J.L., Alcalde M.J. “Characterization of semi-extensive goat production systems in southern Spain”. *Small Ruminant Research*. Vol. 47, 2003, pp. 133–143.

Gaspar P., Mesías F.J., Escribano M. Rodriguez de Ledesma A., Pulido F. “Economic and management characterization of dehesa farms: implications for their sustainability”. *Agroforestry Systems*. Vol. 71, 2007, pp. 151–162.

Hardinan R.T., Lacey R., Yang M.Y. “Use of cluster-analysis for identification and classification of farming systems in Qingyang county, central North China”, *Agricultural Systems*. Vol. 33 (2), 1990, pp. 115-125.

INE-EEEA. *“Encuesta sobre la Estructura de las Explotaciones Agrícolas de 2005*”, [On-line], Available at: Internet <http//www.ine.es> (Accessed 10 January 2007).

INE-Met. *“Metodología del Censo Agrario*” [On-line], Available at: Internet <http//www.ine.es> (Accessed 31 March 2009).

Iraizoz B., Gorton M., Davidova S. “Segmenting farms for analysing agricultural trajectories: A case study of the Navarra region in Spain”, *Agricultural Systems*. Vol. 93 (1-3), 2007, pp. 143–169.

Ketchen D.J., Shook C.L. “The application of cluster analysis in strategic management research: an analysis and critique”. *Strategic Management Journal*, Vol. 17, 1996, pp. 441-458.

Köbrich C., Rehman T., Khan M. “Typification of farming systems for constructing representative farm models: two illustrations of the application of multi-variate analyses in Chile and Pakistan”, *Agricultural Systems*. Vol. 76 (1), 2003, pp. 141-157.

Milán M.J., Bartolome J., Quintanilla R., García-Cachán M.D., Espejo M., Herráiz P.L., Sánchez-Recio J.M., Piedrafita J. “Structural characterisation and typology of beef cattle farms of Spanish wooded rangelands (dehesas)”. *Livestock Science*. Vol. 99, 2006, pp- 197– 209.

Pardos L., Maza M.T., Fantova E., Sepúlveda W. “The diversity of sheep production systems in Aragón (Spain): characterisation and typification of meat sheep farms”. Spanish Journal of Agricultural Research. Vol. 6(4), 2008, pp. 497-507.

Perrot C, Fraysse J.L. “Diversity of livestock (ruminants) production systems: main factors and elements of quantification from the agricultural census 2000”. *9th Meeting on Ruminant Research, Paris, France*. 2002, pp. 165-168.

Riveiro J.A., Álvarez C.J., Marey M.F., Marco J.L., Alvarez C.J. “Procedure for the classification and characterization of farms for agricultural production planning: Application in the Northwest of Spain”, *Computers and Electronics in Agriculture*. Vol. 6 (1), 2008, pp. 169–178.

Riveiro-Valiño J.A., Álvarez-López C.J., Marey-Pérez M. F. “The Use of Discriminant Analysis to Validate a Methodology for Classifying Farms Based on a Combinatorial Algorithm” *Computer and electronics in agriculture.* Vol. 66 (2), 2009, pp. 113-120.

Tiffin R. ‘‘Bayesian Clustering of Farm Types Using the Mixtures Model”. Journal of *Agricultural Economics.* Vol. 57 (3), 2006, pp. 547–562.

Usai M.G., Casu S., Molle G., Decandia M., Ligios S., Carta A. “Using cluster analysis to characterize the goat farming system in Sardinia”. *Livestock Science*. Vol. 104, 2006, pp. 63–76.

Zaldivar M., Menacho C. “Statistical methodology for characterization of guinea-pig farms (Cavia-Porcellus)”. *Turrialba*. Vol. 41(1), 1991, pp. 15-21.

Acknowledgements

The authors are grateful to CICYT for funding the project ‘Integration of a methodological model applied to decision making in the management of agricultural production planning processes’ (AGL2006-04789/AGR).

Correspondence author (For further information, please contact):

José Antonio Riveiro Valiño
Escuela Politécnica Superior – Universidad de Santiago de Compostela
Campus Universitario s/n 27002 Lugo
Phone: +34 982 285 900
Fax: + 34 982 285 926
E-mail : joseantonio.riveiro@usc.es