

A combined multi-variate statistical analysis to establish dairy farm typologies in Cantabria

Ibán Vázquez-González^{a,*}, Elena García-Suárez^b, Francisca Ruiz-Escudero^b, Ana Isabel García-Arias^a

^a University of Santiago de Compostela, Escola Politécnica Superior de Enxeñaría (Department of Applied Economics), Campus Universitario s/n, 27002 Lugo, Spain

^b Centre for Agricultural Research and Training (CITA), Government of Cantabria, 39600 Muriedas-Cantabria, Spain

ARTICLE INFO

Keywords:

Dairy farm categorisation
Principal component factor analysis
Hierarchical cluster analysis
Northern Spain
Complex sample

ABSTRACT

In the last few decades, dairy farms have undergone an intense process of structural adjustment. Despite this, dairy farming remains the most important agricultural activity in Cantabria (northern Spain), with many different types of dairy farm existing. However, there are few studies that have characterised and established typologies to understand this diversity. This study aimed to develop a method for characterising and categorising all the dairy farms in Cantabria (farm population) from a productive, economic and social point of view using combined multi-variate analysis techniques, including principal component factor analysis (PCFA) and hierarchical cluster analysis (HCA). For this purpose, 86 surveys were conducted on dairy cattle farms in Cantabria from 2016 to 2017 using stratified random sampling optimised with Neyman's minimum variance allocation. The results, which relate to all the dairy farms in Cantabria (SPSS complex sample module), have enabled us to characterise and categorise these farms. The sector is mostly made up of farms with low-production levels. Their main characteristics are the importance of rented land, the use of pasture, good production management, a notable absence of young owners, a strong family link and moderate economic viability. The PCFA synthesised 22 production and socio-economic variables into 8 factors that reproduce 77.7 % of variance, half of these factors being economic in nature. The HCA, which used a double decision criterion to define the optimum number of clusters, has classified the dairy farms in Cantabria into four production typologies, which differ in the level of agricultural activity, profitability and diversification. In addition, a fifth group of farms was identified as singular cases.

1. Introduction

The classification of farms into typologies is a common practice that has become widespread in recent decades (Ferreira-Golpe et al., 2021). It enables a series of individuals to be placed into homogeneous groups based on a set of variables and criteria (Schwering et al., 2022). Several classification criteria exist, and the most common are related to the production system and other socio-economic aspects (Sraïri and Lyoubi, 2003). The main advantages associated with the categorisation of agricultural production systems are that it reveals their structure, functioning and strengths and enables the identification of those aspects of production that need to be improved (Mađry et al., 2013). Furthermore, the establishment of typologies minimises the difficulties in understanding production systems and provides knowledge regarding their differences (Cortez-Arriola et al., 2015; Hassall et al., 2023). It is also

useful for defining improvement strategies in the framework of future policy plans (Graskemper et al., 2021), promoting sustainable development (Castel Genís et al., 2010) and advising on sustainability (Andersen et al., 2007; Bánkuti et al., 2020).

Most of the existing classifications have employed multi-variate statistical analysis techniques, which enable variables of different natures to be used together; Ferreira-Golpe et al. (2021) used cluster analysis with three productive and two socio-economic variables to characterise the honey production sector in northwest Spain. Different techniques and methods exist, which vary depending on the objectives pursued and the nature of the data (Andersen et al., 2007). For example, Connell et al. (2007) used descriptive statistics for a technical-productive characterisation of cattle systems in Anzoategui State (Venezuela). Silva et al. (2007) adopted frequency distribution to detect critical nodes of management processes in dual-purpose cattle

* Corresponding author.

E-mail address: iban.vazquez.gonzalez@usc.es (I. Vázquez-González).

<https://doi.org/10.1016/j.compag.2024.109007>

Received 19 October 2023; Received in revised form 25 April 2024; Accepted 1 May 2024

Available online 10 May 2024

0168-1699/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

production systems in Zulia State (Venezuela). [Siegmund-Schultze and Rischkowsky \(2001\)](#) used and compared three different methods (cluster analysis, logistic regression and correspondence analysis) in the identification of the socio-economic characteristics of urban sheep keeping in west Africa. In these cases, the determination of the typology is the result of a single analysis; however, the typology can also be determined in several steps, for example, [Perrot \(1990\)](#) used a directed iterative method to aggregate with the help of experts the farms into typologies. Others, however, established typologies of livestock farm by combining multi-variate statistical analyses, first a factor analysis and then a cluster analysis. The latter has been used for the studies conducted by [Serrano-Martínez et al., \(2004b\)](#), who identified homogeneous groups of cattle farms in León (Spain). [Cortez-Arriola et al. \(2015\)](#), who captured the diversity of family-based dairy farms in Michoacan (Mexico). [Blanco-Penedo et al. \(2019\)](#), who classified the diversity of organic dairy farms in four European countries. [Gökdağ et al. \(2020\)](#), who classified and characterised dairy goat farms in Italy and Turkey. [Ruiz et al. \(2020\)](#), who characterised extensive livestock farming systems (cattle, goats and sheep) in a protected area of Spain (Sierra Nevada). [Schwering et al. \(2022\)](#), who studied the typologies and motivations of farm management information systems in Germany. Other studies, such as those by [Mateus Silveira et al. \(2022\)](#), who represented the diversity of smallholder dairy production systems in a Brazilian semi-arid region, and [Riveiro-Valiño et al. \(2009\)](#), who validated the dairy farm types in Galicia (Spain), went further and incorporated a third discriminant-type analysis.

Dairy farming is a strategic part of the Spanish agri-food sector owing to its economic and social significance, with an annual turnover of around 13 billion euros and creation of more than 60,000 direct jobs ([López Iglesias and Lainez Andrés, 2022](#)). In recent decades, cattle farms have undergone an intense process of structural adjustment, characterised by a sharp reduction in the number of farms and employees, alongside an increase in production and the degree of intensification ([Riveiro-Valiño et al., 2009](#); [González-Mejía et al., 2018](#)).

Milk production in Spain can be divided into two different areas ([Flores-Calvete et al., 2017](#)). The northern strip, known as the Cantabrian Coast (Galicia, Asturias, Cantabria and the Basque Country), is the most important in terms of production (accounting for 79 % of the dairy farms and 56 % of milk production) ([MAPA, 2023](#)). Smaller production levels (on average 400 tonnes per year) and a diet based on fresh and preserved forage produced on the farm, which includes, to a varying extent, maize silage, and concentrate characterise the Cantabrian Coast farms ([Flores-Calvete et al., 2017](#)). Meanwhile, the rest of the country features farms with higher production levels and irrigated cropping systems and, in some cases, a greater reliance on concentrates and other purchased forage crops.

Our study focused on Cantabria, a small autonomous community located in the north of Spain, where bovines constitute the economic, social and regional mainstay of the agricultural sector ([Calcedo, 2013](#); [Vázquez-González et al., 2023](#)). Furthermore, it occupies a prominent position in the dairy cattle sector of the country ([Ruiz-Escudero et al., 2023](#)). In 2022, it was the third most important region in Spain by number of milk producers (977 units, equivalent to 8.41 % of the total)

and sixth by number of cows (54,324 units, equivalent to 6.81 % of the total) and milk production (408,130 tonnes, equivalent to 5.57 % of the total) ([MAPA, 2023](#)).

As each agricultural production system is different and faces specific problems whose solutions may be unique ([Mađry et al., 2013](#)), it is necessary to capture the heterogeneity and diversity of farms before making any decisions ([Mateus Silveira et al., 2022](#)). In Cantabria, there is a wide variety of dairy farms in relation to their production, socio-economic and management characteristics ([García-Suárez, 2021](#)). However, there are few studies that have characterised them, even fewer that have established typologies and none that focused on all the farms in the territory; for instance, [Blanco-Penedo et al. \(2019\)](#) classified the diversity of organic dairy farms in 14 European regions (one Cantabria); [Calcedo \(2013\)](#) analysed the evolution of the milk production sector in Cantabria during the quota period; [Celorio et al. \(2011\)](#) performed a farm structural characterisation of the Pasiega local bovine breed in Cantabria; [Doltra et al. \(2018\)](#) identified and characterised four forage management systems in 40 Cantabria dairy farms (grazing, zero grazing, conserved forages and maize silage); and [Salcedo et al. \(2022\)](#) calculated the hydric footprint of 53 dairy farms in the north of Spain according to six food typologies (organic, conventional grazing, manger grazing, grass silage, maize silage, grass and maize silage). The above classifications have the limitations of referring to an individual farm ([Doltra et al., 2018](#); [Salcedo et al., 2022](#)), a larger geographic scope ([Blanco-Penedo et al., 2019](#); [Salcedo et al., 2022](#)) or a specific species ([Celorio et al., 2011](#)) and do not use multi-variate statistical analysis to establish typologies or the typologies established are according to feed/forage management ([Doltra et al., 2018](#); [Salcedo et al., 2022](#)).

Due to the lack of knowledge presented above, this study aimed to: first, develop a method that characterises and categorises all the dairy farms in a given territory from a productive, economic and social point of view using combined multi-variate analysis techniques; second, apply this method to Cantabria to complete the available knowledge about its dairy sector. For this purpose, this study had three specific objectives: to obtain factors representative of variables of different nature (factorial characterisation), to obtain a representative productive and socio-economic typology (farm typologies) and to characterise all dairy farms in Cantabria and the resulting typologies.

We present results that refer to all dairy farms in Cantabria, not just those included in the surveys, using an unusual and effective method for that purpose (SPSS complex sample). In addition, we employed combined multi-variate statistical analysis to identify dairy farm typologies (non-existent in this area), considering the decision on the optimal number of groups in the cluster analysis using a twofold criterion. The obtained results are expected to expand the use of this methodological approach, provide further understanding of the diversity of the existing systems and help policymakers in formulating policies. It could also be used to establish links with livestock farm strategies.

2. Materials and methods

The information used was obtained from a survey of dairy cattle farms in Cantabria conducted as follows.

2.1. Surveys

A representative sample (surveys) of dairy farms in Cantabria was determined based on five strata of milk production in the 2015–2016 season¹. The milk production strata were selected according to the population size distribution ([Table 1](#)) and the usual strata used in other

Table 1

Sampling information (population size, number of surveys and elevation factor), based on milk production strata (tonne) in the 2015–2016 season.

Milk production strata 2015–2016 (tonne)	Population size (N)	Number of surveys (n)	Elevation factor (N/n)
<100	364	8	45.5
[100;250]	482	17	28.4
[250;500]	325	19	17.1
[500;1000]	164	15	10.9
≥1000	57	27	2.1
TOTAL	1392	86	16.2

Source: own elaboration.

¹ The 1392 farm data used for sampling (monthly milk production) came from the mandatory cow's milk declarations, provided by the regional administration, with continuous deliveries in the period from 1 April 2015 to 31 March 2016.

works (Flores-Calvete et al., 2017; Doltra et al., 2018). The sampling method employed was stratified random sampling optimised using Neyman's minimum variance allocation for a sampling error of 5 % and confidence level of 95 % (Vázquez-González et al., 2023). The sample size (n) was given by the Equation 1, which includes the following parameters: Nh the population size, Sh the standard deviation, E the maximum sampling error (5 %), K the coefficient associated (1.96) with confidence level of 95 %, Y the population value of milk production in the 2015–2016 season (388,465 tonnes), Sh^2 the variance. The distribution by strata (nh) was given by the Equation 2, that uses any of the above parameters.

Sample size (n)

$$n = \frac{\sum_{h=1}^L (Nh \times Sh)^2}{\frac{E^2 \times Y^2}{K^2} + \sum_{h=1}^L Nh \times Sh^2} \quad (1)$$

Distribution by strata (nh)

$$nh = nx \frac{Nh \times Sh}{\sum_{h=1}^L Nh \times Sh} \quad (2)$$

n = sample size.

nh = sample size by stratum.

Nh = population size in stratum h.

E = maximum sampling error.

Y = population value of variable i (Cantabria milk production in the 2015–2016 season).

K = coefficient associated with confidence level.

Sh = standard deviation of variable i (Cantabria milk production in the 2015–2016 season) in stratum j.

Sh^2 = variance of variable i (Cantabria milk production in the 2015–2016 season) in stratum j.

L = last stratum.

The farms interviewed were selected through a random selection of cases function of the SPSS programme (v.21). This selection of farms was conducted separately for each of the five milk production strata and in two stages (initial and reserve farms), randomly and without replacement.

The questionnaire, which was designed between September and October 2016, consisted of 623 items of information structured in six blocks: ownership, production, family structure, economic structure², recent evolution and prospects (García-Suárez, 2021).

Between November 2016 and February 2017, 86 surveys were conducted on dairy cattle farms in Cantabria through direct personal interviews lasting approximately 1 h. The farmers were previously contacted via telephone and re-contacted after the transcript for missing information. If one initial farm refused to be interviewed, it was replaced by a reserve belonging to the same strata and in a nearby area. Around half of the initial farms were replaced by reserves as they did not answer the telephone or did not agree to the interview for reasons of mistrust, fear and lack of time or interest. The questionnaire was mainly answered by the owners if they were the strategic decision-makers or the person who made such decisions.

2.2. Data analysis

A double multi-variate analysis method was employed to classify the farms, first principal component factor analysis (PCFA) and second

hierarchical cluster analysis (HCA).

2.2.1. Principal component factor analysis

PCFA is a statistical dimension reduction method that represents a wide range of relationships between random variables through a subset of dimensions called factors (Vilela-Ferreira et al., 2021). The principal component method was employed to obtain the factors, and eigenvalues greater than 1 were selected as a criterion for the number of factors. To facilitate the interpretation of the factors, a rotation was performed using the Varimax method, whereas the adequacy of the sample was evaluated using Bartlett's sphericity test and the Kaiser–Meyer–Olkin test (KMO). The stages carried out in the PCFA were as follows:

A-Selection of variables³: a detailed selection was made based on the literature consulted and considering other issues: availability, quality and relevance (Kobrich et al., 2003; Mateus Silveira et al., 2022) (Table 2).

B-Correlation study⁴: highly correlated variables were eliminated ($R^2 \geq 0.9$) (Kobrich et al., 2003).

C-Carrying out successive factor analyses until a valid one was found: the final valid analysis must be substantial (KMO > 0.5 and significance < 0.05); furthermore, all the variables must have high commonality (>0.5) and be interpreted in the rotated component matrix (correlation $|r|$ > 0.5).

2.2.2. Hierarchical cluster analysis

The classification of the livestock farms into typologies has been performed using multi-variate HCA, the most common type of analysis, with the factor scores of the valid PCFA as the variables. Ward's method was employed, and the dissimilarity measures used were the squared Euclidean distance (Caruso, 1997). A double decision criterion was considered to define the optimal number of clusters: a dendrogram and the calculation of the rates of variation of the clustering coefficients or 'elbow rule' (Graskemper et al., 2021; Mateus Silveira et al., 2022). The former, which is more commonly used, is graphical and is based on the interpretation of the dendrogram; the latter, which is analytical, is based on the calculation of the rates of change of clustering coefficients between successive stages (Kobrich et al., 2003; Pérez and Santin, 2007).

2.2.3. Descriptive statistics and post-hoc tests

The results section provides information on the descriptive statistics (mean values, coefficient of variation and confidence intervals at the 95 % level) for a total of 34 non-highly correlated production and socio-economic variables. The values refer to the total population, to which the SPSS complex sample module has been applied. It is essential to use this module to build a sampling plan (csa plan file), with the specification of five milk production strata and their elevation factor,⁵ using the sampling assistant for complex samples. Significant differences in the resulting typologies (Wald F-test) and multiple comparisons in the mean values of the groups are also determined via *post-hoc* tests⁶ (5 % level) using the SPSS complex sample general linear model analysis (Zou et al., 2020).

³ 52 variables were selected (7 land base, 10 livestock, 9 milk production, 12 facilities and machinery, 4 family and work, 10 economic).

⁴ Bivariate Pearson's or Spearman's correlation was used, depending on the distribution of the data (Spearman's if the distribution was not normal). A total of 18 highly correlated variables were eliminated.

⁵ Ratio for each stratum that measures the number of dairy farms in Cantabria (population size) divided by the number of surveys.

⁶ A *post hoc* test was calculated for significant differences or trends in the Wald F-test (p -value < 0.1).

² The economic data collected in the survey correspond to 2016 and the rest of the variables to the time at which the survey was conducted. We are aware of the difficulty and limitations involved in collecting reliable economic information through a survey, so the results obtained are an approximation of the economic reality and should be interpreted with caution.

Table 2

List of 52 variables initially selected for the PCFA, grouped according to categories (milk production, land base, livestock, facilities and machinery, family and work, economic).

CATEGORY	Variable Id.	VARIABLES	SIG. KS ¹	Communality Ext. ⁵
MILK PRODUCTION	1	Milk production 15–16 per farm ²	0.007	
	2	Average milk price 15–16 (€/L) ^{3,4}	0.720	0.700
	3	Milk quota purchase since 1992 (% of quota 15) ²	0.379	
	4	Milk production 15–16 per cow (L/cow) ²	0.728	
	5	Milk production 15–16 per UAA (L/ha UAA) ²	0.075	
	6	Milk production 15–16 per AWU (L/UTA) ²	0.423	
	7	Variation in milk production 16–17 (% of production 15–16) ³	0.000	
	8	Production offset 14–15 (% of quota 14–15) ³	0.030	
	9	Milk quota purchase since 1992 (kg) ^{3,4}	0.000	0.787
LAND BASE	10	Average plot size (ha) ³	0.000	
	11	UAA maize (% of UAA total) ^{3,4}	0.000	0.668
	12	UAA grazing pastures dairy herd (% of UAA total) ³	0.000	
	13	UAA green-cut pasture manger dairy herd (% of UAA total) ³	0.092	
	14	UAA total (ha) ^{3,4}	0.099	0.838
	15	UAA used for pasture (ha) ²	0.023	
	16	Rented UAA (% of UAA total) ^{3,4}	0.380	0.722
LIVESTOCK	17	Livestock units (LU) (milk + other cattle) ²	0.006	
	18	Production milk cows (% of total cows) ^{3,4}	0.188	0.771
	19	Rearing rate (% heifers ≥ 12 months of total cows) ³	0.288	
	20	Beef cows (% of total livestock units) ^{3,4}	0.000	0.699
	21	Stocking rate (LU/UAA) ^{3,4}	0.826	0.793
	22	Productivity per annual work unit (LU/AWU) ³	0.580	
	23	Total dairy cows (production + dry) ²	0.007	
	24	Average longevity before culling (lactations) ^{3,4}	0.024	0.621
	25	Cows monitored for milk yield (% of total cows) ^{3,4}	0.000	0.751
	26	Primiparous dairy cows (% of production cows) ³	0.306	
FACILITIES AND MACHINERY	27	N° of headlocks/N° of beds ³	0.000	
	28	N° of headlocks/dairy cow ²	0.248	
	29	N° of beds/dairy cow ²	0.251	
	30	Slurry pit capacity per cow (m ³ /cow) ²	0.022	
	31	Year of construction most recent cowshed ³	0.010	
	32	Slurry pit capacity (m ³) ^{3,4}	0.000	0.753
	33	Most powerful tractor power per UAA (horsepower/ha) ^{3,4}	0.087	0.812
	34	Total surface area of cowshed (m ²) ²	0.000	
	35	Distance from cowshed to housing (m) ^{3,4}	0.000	0.810
	36	Maximum storage time slurry pit (months) ²	0.001	
	37	Most powerful tractor power (horsepower) ²	0.666	
	38	Cowshed surface area per cow (m ² /cow) ^{3,4}	0.075	0.857
FAMILY AND WORK	39	Number of family members ³	0.000	
	40	Age of the owner (years) ³	0.422	
	41	Total AWU ^{3,4}	0.000	0.819
	42	AWU paid labour (% of total AWU) ^{3,4}	0.000	0.732
ECONOMIC	43	Total gross product (GP) (euros) ²	0.003	
	44	Net margin (euros) ²	0.000	
	45	Net margin per AWU (euros/AWU) ^{3,4}	0.254	0.780
	46	Net margin per 1000 L (euros/1000 L) ^{3,4}	0.191	0.868
	47	Net margin per UAA (euros/ha UAA) ²	0.155	
	48	Total cost (TC) (euros) ²	0.003	
	49	Milk revenues (% of GP) ^{3,4}	0.001	0.805
	50	Purchased food (% of TC) ^{3,4}	0.816	0.914
	51	General cost (% of TC) ^{3,4}	0.080	0.771
	52	External factor cost (% of TC) ^{3,4}	0.030	0.824

¹ Statistical significance of the Kolmogorov–Smirnov test (normal distribution contrast). If the p-value is equal or greater than 0.05 (normal distribution).

² 18 variables highly correlated ($R^2 \geq 0.9$), not analysed in PCFA.

³ 34 variables not highly correlated included in the PCFA.

⁴ 22 variables used in PCFA final valid analysis.

⁵ Extraction of communalities for 22 variables used in PCFA final valid analysis.

Source: own elaboration.

3. Results

3.1. Factor characterisation

The final valid PCFA used 22 variables (2 milk production, 3 land

base, 5 livestock, 4 facilities and machinery, 2 family and work, 6 economic) and generated 8 factors reproducing 77.7 % of the original variance. The analysis was satisfactory in statistical terms as the measure of adequacy (KMO) was high (0.708), Bartlett's Sphericity test was significant (0.000) and the chi-squared value was equal to 95.1.

Table 3

Interpretation of the factors resulting from final valid dairy cattle PCFA (explained variance, significance, identification of variables and correlation coefficient in the rotated components matrix).

Factor→ Eigenvalue→ %Variance→ (accumulated)	Name → Meaning factor	Variables (Variable Id.) → name	Correlation with factor ¹
F1 →6.32→ 28.7 %→ (28.7 %)	Production dimension and intensification → Higher milk prices, dimension of the facilities, investments, labour needs, paid labour, maize cultivation and stocking rate and lower cow longevity	(2) →Average milk prize 15–16 (€/L)	0.525
		(24) →Average longevity before culling (lactations)	−0.570
		(32) →Slurry pit capacity (m ³)	0.836
		(41) →Total AWU	0.861
		(42) →AWU paid labour (% of total AWU)	0.651
		(11) →UAA maize (% of UAA total)	0.736
		(21) →Stocking rate (LU/UAA)	0.518
		(9) →Milk quota purchase since 1992 (kg)	0.785
		(45) →Net margin per AWU (euros/AWU)	0.800
		(46) →Net margin per 1000L (euros/1000 L)	0.922
F2 →2.52→ 11.4 %→ (40.2 %)	Economic profitability → Higher farm income per annual work unit and per volume of milk	(14) →UAA total (ha)	0.747
F3 →2.02→ 9.2 %→ (49.3 %)	Extensification → Higher utilised agricultural area and lower stocking rate and power per utilised agricultural area	(21) →Stocking rate (LU/UAA)	−0.641
F4 →1.5→ 6.8 %→ (56.2 %)	High-production-cost structure → Higher specific cost of purchased food and lower general cost	(33) →Most powerful tractor power per UAA (horsepower/ha)	−0.820
F5 →1.38→ 6.3 %→ (62.5 %)	Production specialisation → Greater relevance of rented land and cattle management (milk yield monitoring regime, production milk cows)	(50) →Purchased food (% of TC)	0.929
		(51) →General cost (% of TC)	−0.566
		(16) →Rented UAA (% of UAA total)	0.629
		(25) →Cows monitored for milk yield (% of total cows)	0.653
F6 →1.22→ 5.5 %→ (68.0 %)	Economic specialisation towards milk → Higher revenues from milk sales and lower relevance of beef cows.	(18) →Production milk cows (% of total cows)	0.649
		(49) →Milk revenues (% of GP)	0.708
		(20) →Beef cows (% of total livestock units)	−0.803
F7 →1.09→ 5.0 %→ (73.0 %)	External factors → Greater distance from cowshed to house and the percentage of external factor costs (land rent, paid labour and interest on loans)	(35) →Distance from cowshed to housing (m)	0.861
F8 →1.03→ 4.7 %→ (77.7 %)	Animal welfare →More space for livestock	(52) →External factor cost (% of TC)	0.650
		(38) →Cowshed surface area per cow (m ² /cow)	0.893

¹ Only highly correlated variables ($|r| > 0.5$) are presented.

Source: own elaboration.

Furthermore, all the variables used exhibited high communalities and high correlation coefficients in the rotated component matrix (see [Tables 2 and 3](#)). The factors have been defined according to the nature of the highly correlated variables ($|r| > 0.5$) in the rotated component matrix ([Table 3](#)).

Factor 1, called ‘production dimension and intensification’, explains 28.7 % of the total variance. It is positively correlated with the following variables: milk price, slurry pit capacity, total annual work unit (AWU) and percentage of paid labour, percentage of utilised agricultural area (UAA) under maize cultivation, stocking rate (total livestock units [LU]/UAA) and amount of quota purchased since 1992. It is also negatively correlated with average cow longevity.

Factor 2, called ‘economic profitability’, accounts for 11.4 % of the total variance. The best correlated variables, all of them being of an economic nature, are the economic profitability indicators, such as the net margin (NM) per AWU (NM/AWU) and the NM per volume of milk produced (NM/1000 L), both positively correlated.

Factor 3, called ‘extensification’, explains 9.2 % of the total variance. The variable positively correlated with this factor is the total UAA and the variables negatively correlated with it are the stocking rate and the power in horsepower (hp) of the most powerful tractor on the farm (less than 10 years old) expressed as hectares (hp/ha).

Factor 4, labelled ‘high-production-cost structure’, explains 6.8 % of

the total variance. The variables positively correlated with this factor are the percentage of purchased food over total costs and the variable negatively correlated with it is the percentage of general costs over total costs.

Factor 5, defined as ‘production specialisation’, explains 6.3 % of the total variance. The variables positively correlated with this factor are the percentage of UAA rented as well as the percentage of cows under a milk yield monitoring regime and cows producing milk.

Factor 6, defined as ‘economic specialisation towards milk’, explains 5.5 % of the total variance. The variable positively correlated with it is the percentage of revenues from milk sales compared with gross product⁷ and the variable negatively correlated with it is the percentage of beef cows over LU.

Factor 7, defined as ‘external factors’, explains 5 % of the total variance. The most important variables, which are positively correlated with it, are the distance from the cowshed to the house and the percentage of external factor costs (land rent, paid labour and interest on loans) over total costs.

⁷ The following five revenues are considered: sale of milk, subsidies, cattle, other livestock and other agricultural revenues (processing of agricultural products, fertilizers, crops, insurance, etc.).

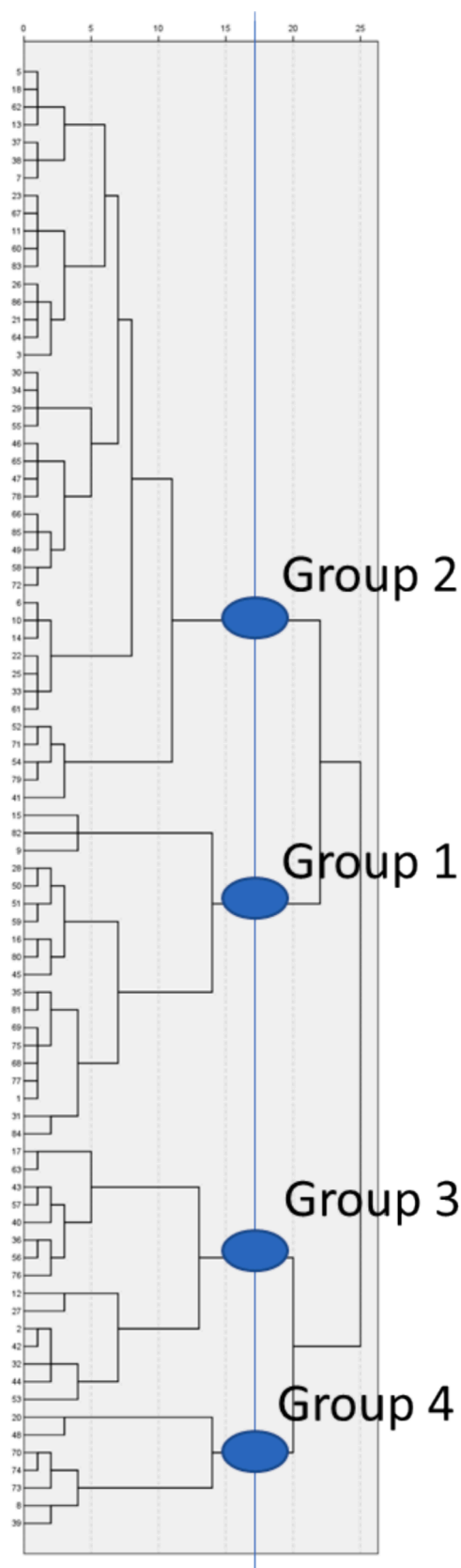


Fig. 1. Dendrogram.

Finally, Factor 8, defined as ‘animal welfare’, explains 4.7 % of the total variance. The only single variable that is highly positively correlated with it is cowshed area per cow.

3.2. Farm typologies

Out of the whole sample (86 surveys), we decided not to include three farms in the classification analysis as they always formed a different group in the different HCAs conducted due to the very different values (outliers) of some of the 22 variables used in the PCFA. This group of three farms, hereafter referred to as singular cases (SC), will be considered in the characterisation of the resulting typologies.

3.2.1. Determination of the number of clusters

The dendrogram shows that the optimum number of clusters is 4, drawing a vertical line at a distance between 14 and 20 points. This interval gives the maximum distance for the horizontal lines separating the clusters for all the clusters with respect to the later (right) or earlier (left) stages (Fig. 1) (Kobrich et al., 2003).

The graphical representation of the analytical solution, relative to the clustering rates at the different stages, suggests, as does the dendrogram, that the optimal number of clusters is 4. It is noticeable how the slope increases at a lower number of clusters, e.g. 3 (Fig. 2).

3.2.2. Characterisation of dairy farms in Cantabria

In relation to the land base, the holdings have an average of 28.6 ha of UAA, with relatively large plots (2.7 ha), where renting is the main tenancy regime (54.7 % of the UAA). The amount of land dedicated to fodder maize is low, accounting for less than 5 % of the UAA; contrarily, pasture under grazing is the main use, with about half of the UAA (Table 4).

The average herd size per farm is 46 dairy cows, of which a high percentage is in production or being monitored for milk yield (82 % and 63 %, respectively). The average stocking rate is 2.9 LU/ha, the average longevity before becoming cull cows is 4.2 lactations and the rearing rate is 41.1 %, which might seem high for these herd longevity values. The diversification of production into other livestock activities, such as beef cows, is not widespread, accounting for less than 5 % of the total LU.

In terms of milk production, the establishment of milk quotas in Spain in April 1992 and their subsequent abolition in March 2015 have conditioned the evolution of dairy farms. Since 1992, farms bought an average quota of 153,000 kg of milk. Furthermore, during the period immediately before (2014–2015 season) and after (2016–2017 season) the elimination of quotas, an opposite productive behaviour was detected. In the first period, milk production was almost 10 % lower than the available quota, whereas in the second period, it slightly

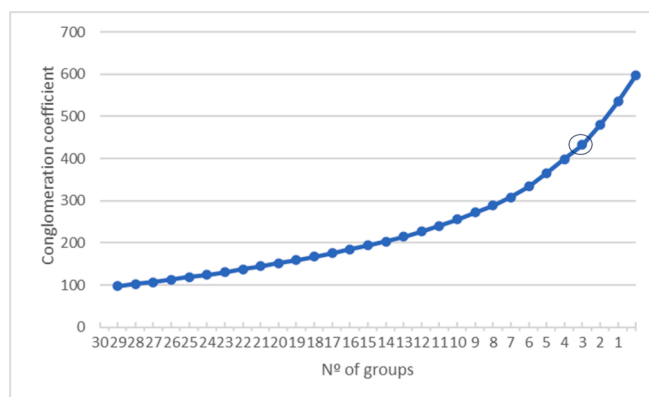


Fig. 2. Graphical representation of the variation rates of the clustering coefficients (elbow rule).

Source: own elaboration

Table 4

Production and socio-economic characterisation of the resulting clusters. Results elevated to the population as a whole, taking as a reference the 34 non-highly correlated variables used in the PCFA.

Variable Id.	GROUPS VARIABLES							Desc. St. Complex Sample ⁴			GLM	
		G1	G2	G3	G4	CS	Total	Coeff. Var.	LCI	UCI	F Wald	Sig ³ .
	N° Surveys	19	42	15	7	3	86					
	Population (n° farms) ⁵	67	559	488	222	56	1392					
MILK PRODUCTION												
7	Variation in milk production 16–17 (% of production 15–16)	1.88a ²	0.70b	1.48ab	2.58a	0.00c	1.3	0.432	0.18	2.42	2.87	0.028
2	Average milk price 15–16 (€/1000L) ¹	313a	295b	270c	259c	264c	280	0.012	279	287	10.8	0.000
8	Production offset 14–15 (% of quota 14–15)	15.7a	0.5ab	−19.4bc	−15.5b	−31.6c	−9.6	−0.304	−15	−3.8	4.42	0.030
9	Milk quota purchase since 1992 (1000 kg) ¹	605a	213b	53.4c	34.1c	352ba	153	0.085	127	179	13.8	0.000
LAND BASE												
14	UAA total (ha) ¹	55.0a	35.8b	20.7c	19.7c	30.6bc	28.6	0.092	23.4	33.9	9.21	0.000
16	Rented UAA (% of UAA total) ¹	68.8b	61.6b	49.4bc	37.5c	83.1a	54.7	0.086	45.3	64	5.84	0.000
11	UAA maize (% of UAA total) ¹	35.4a	3.3b	4.0b	0.0c	10.3ab	4.8	0.208	2.8	6.8	7.89	0.000
12	UAA grazing pastures dairy herd (% of UAA total)	26.1b	34.3b	63.6a	59.4ab	56.2ab	49.1	0.092	40.1	58.1	3.04	0.022
13	UAA green-cut pasture for manger dairy herd (% of UAA total)	9.1	27.4	28.6	28.7	45.5	27.9	0.183	17.7	38.1	1.37	0.251
10	Average plot size (ha)	3.0ab	3.2a	2.7ab	1.8b	1.1b	2.7	0.058	1.9	3.5	2.19	0.077
LIVESTOCK												
18	Production milk cows (% of total cows) ¹	86.6a	85.4a	76.7b	84.4a	86.7a	82.3	0.012	80.3	84.3	4.34	0.003
26	Primiparous dairy cows (% of production cows)	36.9	31.4	31.1	39.8	39.6	33.2	0.073	28.4	38.1	1.0	0.410
25	Cows monitored for milk yield (% of total cows) ¹	100a	100a	24.4b	33.3b	100a	62.9	0.099	50.5	75.2	26.7	0.000
19	Rearing rate (% heifers ≥ 12 months of total cows)	36.1	35.2	38.8	61.9	42.7	41.1	0.079	34.6	47.5	1.96	0.107
24	Average longevity before culling (lactations) ¹	3.1b	4.3ab	4.8a	2.9b	4.6a	4.2	0.048	3.8	4.6	12.54	0.000
21	Stocking rate (LU/UAA) ¹	3.6a	2.8b	2.6bc	3.6a	2.1c	2.9	0.056	2.6	3.2	5.25	0.001
20	Beef cows (% of total livestock units) ¹	0.0b	2.2ab	1.9ab	16.8ab	6.8ab	4.5	0.365	1.2	7.7	3.24	0.016
22	Productivity per annual work unit (LU/AWU)	56.8a	35.3b	28.8b	45.4ab	36.9b	35.7	0.069	30.8	40.6	8.15	0.000
FACILITIES AND MACHINERY												
31	Year of construction the most recent cowshed (over 1900)	99a	95a	74b	71b	90ab	84	0.002	77	91	3.99	0.005
35	Distance from cowshed to housing (1000 m) ¹	0.97a	0.21b	0.092b	0.96ab	11.5a	0.784	0.459	0.07	1.49	2.43	0.055
32	Slurry pit capacity (100 m ³) ¹	12.4a	3.69b	2.72bc	1.57c	4.71b	3.47	0.084	2.89	4.06	9.09	0.000
33	Most powerful tractor power per UAA (hp/ha) ¹	2.9b	4.3ab	5.3a	4.2ab	5.1ab	4.6	0.093	3.75	5.46	3.18	0.018
38	Cowshed surface area per cow (m ² /cow) ¹	13.4a	14.1a	11.9a	7.5b	25.3a	12.7	0.086	10.5	14.9	7.6	0.000
27	N° of headlocks/N° of beds	1.00	1.06	1.03	1.00	0.98	1.03	0.024	0.98	1.08	1.21	0.314
FAMILY AND WORK												
39	Number of family members	3.7	3.6	3.5	3.0	3.5	3.5	0.053	3.1	3.83	0.39	0.816
40	Age of the owner (years)	49.8	52.6	51.8	51.0	46.6	51.7	0.032	48.4	55.0	1.49	0.213
41	Total AWU ¹	3.6a	2.3b	1.8bc	1.6c	1.8bc	2.1	0.046	1.86	2.23	4.61	0.002
42	AWU paid labour (% of total AWU) ¹	37.9a	7.5b	1.0c	0.0	21.6ab	6.1	0.190	3.8	8.4	9.31	0.000
ECONOMIC												
45	Net margin per AWU (1000 euros/AWU) ¹	38.9a	9.6b	2.9b	12.0b	11.9b	9.1	0.197	5.55	12.7	3.94	0.006
46	Net margin per 1000 L ¹	124	51	17	115	62	53	0.424	8.3	98.5	1.7	0.158
49	Milk revenues (% of GP) ¹	83.5a	81.9a	71.4b	58.2c	70.8bc	74.1	0.023	70.7	77.4	8.82	0.000
50	Purchased food (% of TC) ¹	41.0b	50.0a	38.7b	47.1ab	34.1b	44.7	0.039	41.3	48.1	5.68	0.000
51	General cost (% of TC) ¹	20.7c	22.3bc	35.2a	28.9ab	24.7b	27.9	0.045	25.4	30.4	5.82	0.000
52	External factor cost (% of TC) ¹	11.9b	6.2c	6.3c	8.2bc	19.9a	7.4	0.163	5.0	9.8	3.76	0.007

¹ 22 variables used in PCFA final valid analysis.

² Subscript with a different letter represents the statistical significance between the groups (p -value < 0.05).

³ Significance level in black (p -value < 0.05).

⁴ Descriptive statistics: coefficient of variation (standard deviation/mean value) and confidence intervals at the 95 % level (lower LCI and upper UCI).

⁵ This variable represents the number of all dairy farms in the territory (Cantabria) that belongs to each group.

Source: own elaboration.

increased.

In relation to the facilities and machinery available, it should be noted that on average, the most recent cowsheds are amortised, as they are more than 30 years old. These cowsheds are located at an average distance from the housing of 784 m and have other characteristics such

as an average stall area per cow of 12.7 m² and 1.03 headlocks per bed. The mechanisation index, measured as the ratio of the power (hp) of the most powerful tractor on the farm (less than 10 years old) per hectare of UAA, is 4.6 hp/ha.

Regarding the characteristics of the owner, family and work, almost

Table 5

Percentage distribution of dairy cattle typologies by milk production strata (tonne).

GROUPS	Milk production strata 2015–2016 (tonne)					Total
	<100	[100;250]	[250;500]	[500;1000]	≥1000	
G1. Farms with high production and profitability	0.0	0.0	0.0	49.3	50.7	100
G2. Farms with medium production and profitability	0.0	30.4	45.9	19.5	4.2	100
G3. Farms with low production and profitability	55.9	34.8	7.0	2.2	0.0	100
G4. Diversified farms with low production	41.1	51.2	7.7	0.0	0.0	100
G5. Singular cases	0.0	50.3	30.3	19.4	0.0	100
Total	26.1	34.6	23.3	11.8	4.1	100

Source: own elaboration.

all the dairy farms in Cantabria are family farms (98.6 %), mainly with the legal status of the owner as a natural person (58.9 %), followed by single-family companies (21.7 %), where all the partners are family members living in the same household and multi-family companies (18 %) (García-Suárez, 2021). The average number⁸ of family members is 3.5, with the average age of the owner being 51.7 years. The average number of AWU is 2.1 AWU, of which 6.1 % is paid labour.

In relation to the economic information for 2016, the average NM per AWU is 9,153 €/AWU, and the NM per unit of milk produced is 0.053 €/L. In terms of revenues, milk is by far the main income source (74.1 %), and in terms of expenditure, purchased feed is the main cost (44.7 %).

Finally, the production structure of the sector is mostly made up of farms with low production, with 60.7 % of dairy farms producing less than 250 tonnes (Table 5).

3.2.3. Characterisation of farm typology

Of the 34 variables analysed in Table 4, 9 do not substantially differ in mean values between typologies (p -value > 0.05); in these cases, the mean value of most of the groups is within the confidence intervals (LCI and UCI). The variables, grouped in categories, are as follows: land base (percentage UAA devoted to green-cut pasture for in-manger feeding, average size of the plots), livestock (percentage of primiparous dairy cows, rearing rate), facilities and machinery (distance from the cowshed to the house, headlocks/bed ratio), family and work (number of family members, age of the owner) and economic (MN/1000 L).

In the HCA, a total of four groups have been obtained, which are described below, together with the group of SC.

Group 1–‘Farms with high production and profitability’. This group is made up of 19 of the farms surveyed, equivalent to 4.8 % of the dairy farms in Cantabria; all of them belong to the strata with the highest milk production (500 tonnes or more).

Their land base is characterised by a higher UAA, a higher level of forage maize cultivation and rented land. Their livestock is characterised by a larger herd size, increased stocking rate, higher milk productivity per head, more controlled management (percentage of cows in production and under milk yield monitoring) and specialisation in milk production (absence of beef cows); however, the average longevity before culling is among the lowest. In terms of production, these farms, which have the highest milk production volume, received a higher price per litre of milk, have bought a larger amount of the milk quota and, in the last milking season (2014–2015) before the milk quota was abolished (March 2015), registered a considerable excess of production, which differs from the overall situation for all the farms in Cantabria.

In relation to the facilities and machinery, their cowsheds were constructed more recently and are located at a greater distance from the house (significant trend); they also have a larger surface area per cow and a greater slurry pit capacity. The power ratio of the most powerful tractor per hectare of UAA has the lowest values, which may be due to

the greater UAA available. Regarding labour, this group has higher labour requirements and a greater percentage of paid work (six times higher than the average for the sector). On the economic side, Group 1 has a greater economic profitability per worker (MN/AWU) as it combines higher milk productivity per AWU and unitary NM (NM/1000 L). Furthermore, a strong economic specialisation exists towards dairy, with this group having the highest percentage revenues from the sale of milk; however, general costs are lower due to the higher proportion of specific costs.

Group 2–‘Farms with medium production and profitability’. This is the largest group, comprising 42 of the farms surveyed and representing 40.2 % of the dairy farms in Cantabria. Group 2 is made up of farms with lower production, all of them with a production of at least 100 tonnes; those with production between 250 and 500 tonnes (46 %) are the largest.

The land base is characterised by an UAA that is substantially smaller than that of Group 1, and renting is the main land tenure regime, as in Group 1. However, the importance of forage maize is less, and the area under green-cut pasture or grazing is greater. As in Group 1, there is a high percentage of cows in production, all of them being monitored for milk yield; however, the longevity of the herd is higher, which is possibly associated with a lower stocking rate and milk productivity. The characteristics related to milk production (purchased quota, milk price and production variation), annual labour requirements and the number of paid workers are also lower than in group 1. The cowsheds are recently built and well-sized buildings, as in Group 1, but the slurry pit capacity is much lower.

Group 2 has an NM/AWU that is much lower than that of Group 1 and similar to the average for the sector. Similar to Group 1, it specialises in milk production, as can be seen in the high dependence on milk revenues and the low percentage of general costs. However, the percentage of external factors is the lowest of all groups and indicates lower dependence on external resources (land, labour and capital).

Group 3–‘Farms with low production and profitability’. Group 3 is made up of 15 of the farms surveyed, representing 35.1 % of the dairy farms in Cantabria. This group is composed almost entirely of farms producing less than 250 tonnes, of which the majority (56 %) produce less than 100 tonnes.

Group 3 is characterised by a reduced surface, mainly dedicated to grazing. The herd is smaller in size and productivity (LU/AWU), as is the percentage of cows in production and under milk yield monitoring. Furthermore, the longevity of the cows is the highest, close to 5 lactations. The average price received for the milk sold, together with the negative evolution of production in the last quota campaign and the volume of quota purchased, has lower values different (p -value < 0.05) from the two previous groups. The same is true for the age of the cowsheds. The power ratio of the most powerful tractor per hectare of UAA is the highest among all the groups due to the smaller size of the farms. Moreover, the annual labour requirements are lesser, and there is virtually no paid labour. In the economic side, Group 3 has the lowest profitability per worker, less than €3,000 (MN/AWU), which shows the economic constraints they are experiencing and the need to supplement their incomes with other revenues. The percentage of expenditure on

⁸ In the case of multi-family companies and non-family companies, the number of members considered is the number of partners.

purchased feed is lower than the average, indicating the greater use of feed produced on the farm (greater autonomy), although the percentage of general costs is higher.

Group 4-‘Diversified farms with low production’. This group is made up of 7 of the farms surveyed, representing 15.9 % of the dairy farms in Cantabria. It is composed of farms with low production; however, it has more farms with higher production than Group 3, since the majority (51 %) of these farms produce between 100 and 250 tonnes.

Although Group 4 is similar to Group 3, some differences exist. The size, with 20 ha of UAA, is small, as is the case for Group 3; however the degree of renting is lower. In terms of livestock, despite having a similar livestock size, Group 4 has a higher percentage of cows in production and a higher stocking rate. Furthermore, the presence of beef cattle and a high rate of rearing seem to indicate a diversification of activity. The price received for the milk and the amount of quota purchased are the lowest, exhibiting no significant differences from those of Group 3. This also happens with the age of the herd, capacity of the slurry pit, annual labour requirement or number of paid workers. However, the available surface area per cow is substantially lower. Despite its small production size, Group 4 has a much higher NM, both per litre and per worker, than Group 3, and even higher than the average for dairy farms in Cantabria. It has the lowest percentage of revenues from milk sales among all the groups (58.2 %), which is compensated for by the money obtained from the diversification of agricultural activity (subsidies 22.3 %, calves 11.2 %, rearing heifers 4.2 %, cull cows 4.1 %).

Group 5-‘SC’. This group is made up of three of the farms surveyed, representing 4 % of the dairy farms in Cantabria. The SC have a larger production structure than those of Groups 3 and 4; they are farms belonging to the intermediate-size strata (from 100 to 1000 tonnes), with half of them producing between 100 and 250 tonnes.

Group 5 presents some unique production data, such as the high degree of renting (83 % of the UAA), small plots (1.1 ha), greater longevity before culling (4.6 lactations) and lower stocking rate (2.1 LU/ha). In relation to milk production, they have purchased a substantial proportion of the quota since 1992, but the recent evolution has been marked by a much lower production than their potential; furthermore, the average price received for their milk is the lowest (264 €/1000 L). The most remarkable aspect in relation to the facilities is the greater distance of the cowshed from the house (11 km) and the greater capacity of the cowshed (25 m²/cow). On the economic side, their profitability is close to the average, but they have higher external factor costs (20 % of the total), which seems to be conditioned by the higher percentage of rented land, paid labour and investments made.

4. Discussion

4.1. PCFA

Quantitative variables were used in the PCFA, which is the most common type of analysis (Mądry et al., 2013; Vázquez-González et al., 2022). The same authors, who were aware that the selection of variables depends on the production system analysed and the objective pursued in the characterisation, found that the most used technical variables were surface area, size and stocking rate, paid labour, feed supply and productivity; in terms of economic variables, the most used variables were revenues, expenditure and margins (income). Kaouche-Adjlane et al. (2015), in establishing a typology of dairy farms in Algeria, used variables related to ownership (age, education), structure (land base, livestock and equipment), management (feeding, production and reproduction) and economy. Maseda et al. (2004), in categorising family dairy farms in Galicia (northwest Spain), used 94 variables related to the location of the farm, family structure, sources of income, production, characteristics of the cowshed, characteristics of the facilities and transit routes. In our case, the production and socio-economic variables used that were related to milk production, land base, livestock, facilities and machinery, family and work and economy largely coincided with those

used in the literature.

The results obtained from the PCFA regarding the number of factors (8) and the percentage of variance (77.7 %) were consistent with those obtained in other studies characterising dairy farms with variance ranging from 50 % in England and Wales (González-Mejía et al., 2018) and other European regions (Blanco-Penedo et al., 2019) to 84.2 % in Michoacan (Mexico) (Cortez-Arriola et al., 2015). The number of factors may vary from 3 (Kaouche-Adjlane et al., 2015; González-Mejía et al., 2018) to 16, as reported by Maseda et al. (2004).

The eight factors obtained have a decreasing importance in the percentage of variance explained, something that also occurs in the works consulted and is a characteristic of the analysis. In terms of the nature of the factors, the economic aspect is of greatest importance as it is present in four of the eight factors obtained (Factors 2, 4, 6 and 7). Thus, Factor 2, called economic profitability, is defined by two very important variables in the economic viability of a farm (NM/AWU and NM/1000 L). Bach et al. (2020) argued that in the economic analysis of a dairy farm, it is necessary to pay attention to milk production and to the unit margin, variables that were considered in our study. The Factor 4, referred to as the high-production-cost structure, is defined by a higher percentage for purchased feed costs; Salinas-Martínez et al. (2020) attributed higher feed costs to large farms owing to their greater dependence on concentrates. Factor 6, called economic specialisation, corresponds to farms whose income mainly depends on dairy farming. The European Commission applies a similar concept when defining the concept of technical economic specialisation, when at least 66 % of the gross margin is directly associated with this activity (EC, 2012). Pérez-Méndez et al. (2020), who investigated how health and reproduction affect the technical efficiency of dairy farms in Asturias (northwest Spain), have defined as specialised dairy farms where milk accounts for more than 90 % of sales revenues. Finally, Factor 7, referred to as external factor cost, is related to those farms that have fewer of their own resources (land, labour and capital).

The number of factors obtained, their interpretation and the position they occupy exhibit similarities to those of other studies. We considered that the resulting factors can be grouped into two categories, basic and complementary, as reported by Serrano-Martínez et al., (2004a). The basic factors are ‘dimension and intensified production’ and ‘economic profitability’, which reproduces a higher percentage of the variance and is related to size, intensified production and economic profitability. The variables that define these factors, such as size, production, management, labour and economic factors, are considered to be necessary when defining production systems to characterise and classify farms (Cortez-Arriola et al., 2015; González-Mejía et al., 2018).

There are numerous complementary factors that reproduce a smaller percentage of the original variance (Serrano-Martínez et al., 2004b). These factors are defined by specific variables, for example, Factor 3, called extensification, is defined by three variables that take into account surface area (UAA, stocking rate and power/area ratio). Mateus Silveira et al. (2022) reported that surface area is a key variable when characterising extensive production systems. Factor 5, referred to as production specialisation, corresponds to farms with more controlled management. Pérez-Méndez et al. (2020) stated that all the specialised dairy farms in the study were monitoring milk yield. Finally, Factor 8, called animal welfare, is defined by a single variable, cowshed area per cow, which is considered in some studies to be a measure of animal welfare (García-Pérez et al., 2022).

4.2. Identification and characterisation of farm typologies

In our study, before the classification, we decided not to consider three farms (SC) in the HCA as they had very different values (outliers) in some of the 22 variables used in the PCFA. The same decision has been made in the previous literature; thus, Sraïri and Lyoubi (2003) and Cortez-Arriola et al. (2015) decided not to consider these farms in the classification analysis (2 and 1 farms respectively) and treated them as a

complementary group of SC.

In this study, four production typologies were obtained, which represent the diversity among dairy farms in relation to their production, socio-economic and management characteristics. The groups were defined on the basis of the production level, profitability and degree of diversification, according to the names of the first two basic factors of the PCFA (F1–production dimension and intensification, F2–economic profitability). Therefore, our study highlights the importance of the basic factors (Serrano-Martínez et al., 2004a) resulting from the PCFA in classifying and defining the resulting groups of farms when HCA is employed in combination with the PCFA.

Cortez-Arriola et al. (2015) stated that farm diversity can be broadly described in terms of size, as we do, and the intensity of production. In our case, the diversity of farms identified in these four groups may be representative of three different forage management systems according to land uses: Group 1 can also be identified as maize-based because maize is the main land use; Group 2 is an intermediary system, in which grazing, zero grazing and forage conservation have relevance; finally, Groups 3 and 4 can be identified as grazing. These results seem to be consistent with those in the literature as they match with four forage management systems (grazing, zero grazing, conserved forage and maize silage) identified and characterised by Doltra et al. (2018) in Cantabria (Spain). Hereinafter, Salcedo et al. (2022), classified farms into six feeding typologies, and our results seem to coincide with three of them (conventional grazing, manger forage and maize). Evidently, this diversity of dairy farms observed in Cantabria is related to its climatic conditions, such as a temperate Atlantic climate with abundant rainfall, which enables Cantabria to have one of the highest levels of forage production in Europe (Vázquez-González et al., 2023).

In the characterisation of the resulting typologies, a direct relationship was observed between a higher production level and profitability with a greater surface area, degree of renting and area under fodder maize (Ruiz-Escudero et al., 2023; Vázquez-González et al., 2023). Regarding livestock, activity and profitability are associated with an increase in the percentage of cows in production, cows under milk yield monitoring, degree of intensification and productivity (Vázquez-González et al., 2023). Furthermore, the longevity of the cows before culling and diversification into other livestock activities decreases as farm size increases. Todde et al. (2016) also observed this connection between larger farm size and productivity, which Álvarez-López et al. (2008) attributed to mechanisation and Pérez-Méndez et al. (2020) to health and reproductive conditions. Regarding facilities and machinery, the buildings are more recently built and have greater capacity, as are the investments made. However, labour needs increase, and so does the number of paid workers. Bánkuti et al. (2020) associated small farms with family labour. Finally, the economic profitability per worker is also higher on large farms, as is the proportion of dairy farming income, and the relative weight of general costs is less due to the dilution effect of economies of scale.

The results obtained in the categorisation of dairy cattle farms in other areas are similar to ours. González-Mejía et al. (2018) observed a lower diversity of dairy systems in the UK, considering two typologies (extensive and intensive). Furthermore, a convergence between the two types is observed over time due to intensification on the extensive farms. Álvarez-López et al. (2008) proposed a classification of Galician (Spain) dairy farms into three groups, two of them being differentiated according to maize cultivation and the third showing diversification activities in both livestock and crops. Castel Genís et al. (2010) determined the typologies of dairy farms in western Poland, obtaining five groups: two have high productivity, the third has low production and productivity diversified towards off-farm activities and the remaining two are characterised by soil quality. Cortez-Arriola et al. (2015) obtained four typologies based on size (LU and UAA), land ownership and labour; two have a large dimension, and the other two are small and differ in their diversification activities, similar to our third and fourth groups. Graskemper et al. (2021) also found a group of small livestock farms

seeking to diversify the farm's activities. Mateus Silveira et al. (2022) considered livestock species diversity as a means to diversify production so as to ensure food supply security and business opportunities. This diversity of species can be observed in our fourth group, defined as diversified farms with low production, where beef cows have increased relevance. Furthermore, other economic diversification activities have been observed: farm resources (labour, facilities, land) are used to produce agricultural products such as calves (for rearing and for slaughter), to rear heifers and to cull cows. All the diversified activities observed in our fourth group are what García-Arias et al. (2015) called agricultural diversification (producers choosing to develop new products, crops, animal breeds and new productions methods or to begin differentiation quality strategies) a few years ago in Galicia.

4.3. Limitations, novelties and strengths

One of the major limitations of this study is the timing of data collection. Because data were collected in 2017, there is a risk of presenting an outdated picture. In fact, the process of structural adjustment (Vázquez-González et al., 2023) may have continued during this period, leading to a series of changes that may have distorted the characteristics and typologies obtained. Although this limitation is considered to be not so important as the objective of this work follows a methodological approach to characterise and categorise all dairy farms in Cantabria (farm population); we considered it interesting to provide an overview of the main structural changes that have occurred in the sector between 2016 and 2022 (MAPA, 2020,2023) to determine whether this research reflects the current situation. The main change was a sharp decrease of 32.6 % in the number of dairy farms, concentrated in those with an annual production lower than 500 tonnes. There has also been an increase in the productive dimension (cows per farm 19.1 % and milk production per farm 33.6 %) and in milk yield per cow (14 %). Thus, there has been a concentration of milk production in the largest farms (≥ 500 tonnes), which went from representing 16 % of the dairy farms and 52 % of the milk production in 2016 to 23.5 % and 63.2 %, respectively. Although the process of structural adjustment is underway, it is unlikely that these changes will affect the number and denomination of the resulting typologies. However, changes in the representativeness of the groups are expected; thus, Group 1 (≥ 500 tonnes) will increase its relevance and Group 3 would be the group most affected in terms of farm losses.

The statistical method employed to obtain the typologies for the double multi-variate analysis, first the PCFA and then HCA, is the most commonly used method (Kobrich et al., 2003; Ferreira-Golpe et al., 2021). However, it has the limitation that the variables considered *a priori* to be relevant in the classification could not be used, either because they were discarded in the previous stages of variable selection or because they were not significant in the PCFA (Vázquez-González et al., 2022). Mateus Silveira et al. (2022) reported that it is important in typology studies to include variables related to farm size, such as the number of cows. However, in our case, we were not able to include this variable and some others, such as the volume of milk production, total LU, number of family members or the age of the owner, because they have been left out in the previous stages of variable selection (see Table 2). Ruiz et al. (2020), in characterising extensive livestock systems (cattle, goats and sheep) in Sierra Nevada (Spain), experienced something similar; starting from 35 variables, they were left with 7. Furthermore, Gökdağ et al. (2020), in studying dairy goat systems in Turkey and northern Italy, used 18 out of 44 variables.

The strength and novelty of this work is the double criterion used to determine the optimal number of groups in the HCA (dendrogram and the rates of variation of the clustering coefficients), which has facilitated our decision-making and removed some of the subjectivity that was implicit in the use of a dendrogram alone (Hassall et al., 2023). For instance, how to trace the cutting line on the dendrogram seems to be unclear. In our case and also that of Riveiro et al. (2013), the most

suitable cutting height was when the distance coefficient with respect to the later (right) or earlier (left) stages was higher; however, in the case of Blanco-Penedo et al. (2019), the cutting line seemed to attend to the goals of researchers (number of relevant clusters and farms included in them), which Kobrich et al. (2003) recognised as valid but that adds more subjectivity to the decision.

Another strength of our study is that the results are referred to all the dairy farms in Cantabria, not only to those included in the surveys. This was made possible by using the SPSS complex sample module, which requires an appropriate sampling plan, and then applying the specificities of the sampling plan (elevation factor) to the output of the results. The advantage of using this method is that the results refer to the whole territory without having to survey all the existing farms, thus saving resources. The same strategy is used in electoral surveys to see vote estimation. In all the literature reviews we conducted, we found only one study (Flores-Calvete et al., 2017) that employed a similar methodology. In other cases where the results were specific to a group of farms, it is not appropriate to generalise and refer to them as representative of the whole territory.

5. Conclusion

The present study characterises and categorises the milk production sector in a given territory from a productive, economic and social point of view using a combination of multi-variate analysis techniques and unusual methods. The results in this study refer to all the dairy farms in Cantabria as opposed to the usual practice of referring to surveys, which increase the efficiency of resource management. This has been made possible by a survey sample that was representative of the study population and the use of the SPSS complex sample module.

The double statistical analysis method (first PCFA and then HCA) used to define the typologies has proved to be effective in capturing the diversity of the dairy farms in Cantabria. The first two factors of the PCFA, which are related to production, intensification and profitability, are considered basic and determining factors in the delimitation of the production typologies. Furthermore, the double criterion used to determine the optimal number of groups (dendrogram and the rates of variation of the clustering coefficients) in HCA has facilitated our decision and eliminated some subjectivity.

The dairy farms in Cantabria were classified into four production typologies, which differ in terms of production level, profitability and diversification. The observed heterogeneity of the dairy farms corroborates the usefulness of the double multi-variate analysis for establishing typologies. The production structure of the sector in Cantabria is formed, in their majority, by farms with low and medium production. In addition, a fourth group consisting of low production but diversified farms was discovered; these group of farms are an example of how farms with low-production size can improve their economic viability by diversifying their activities.

The limitation of this study is the timing of the data collection. However, it is unlikely that the structural adjustment has affected the number and names of the resulting typologies, although changes in the representativeness of the groups can be expected.

This study has achieved its objective of developing a methodology that can characterise and categorise the dairy sector in Cantabria, obtaining information that was previously not readily available and is necessary to understand the diversity of the system and to adopt, with clearer criteria, policies aimed at promoting the sustainability of livestock farms. We expect to continue this work by analysing the resulting typologies and the relationships that the livestock farms in the region maintain, with their evolving strategies and an analysis of their sustainability.

CRedit authorship contribution statement

Ibán Vázquez-González: Writing – review & editing, Writing –

original draft, Visualization, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elena García-Suárez:** Investigation, Formal analysis, Data curation. **Francisca Ruiz-Escudero:** Writing – original draft, Visualization, Project administration, Investigation, Formal analysis. **Ana Isabel García-Arias:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by pre-doctoral contract FPI-INIA 2015-00016 (García-Suárez, 2021).

References

- Álvarez-López, C.J.A., Riveiro-Valiño, J.A.R., Marey-Pérez, M., 2008. Typology, classification and characterization of farms for agricultural production planning. *Span. J. Agric. Res.* 1, 125–136. <https://doi.org/10.5424/sjar/2008061-299>.
- Andersen, E., Elbersen, 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 (3), 353–362. <https://doi.org/10.1016/j.jenvman.2006.04.021>.
- Bach, A., Terré, M., Vidal, M., 2020. Symposium review: Decomposing efficiency of milk production and maximizing profit. *J. Dairy Sci.* 103 (6), 5709–5725. <https://doi.org/10.3168/jds.2019-17304>.
- Bánkuti, F.I., Prizon, R.C., Damasceno, J.C., De Brito, M.M., Pozza, M.S.S., Lima, P.G.L., 2020. Farmers' actions toward sustainability: A typology of dairy farms according to sustainability indicators. *animal*. 14(S2), 417–423. [10.1017/S1751731120000750](https://doi.org/10.1017/S1751731120000750).
- Blanco-Penedo, I., Sjöström, K., Jones, P., Krieger, M., Duval, J., Van Soest, F., Sundrum, A., Emanuelson, U., 2019. Structural characteristics of organic dairy farms in four European countries and their association with the implementation of animal health plans. *Agr. Syst.* 173, 244–253. <https://doi.org/10.1016/j.agsy.2019.03.008>.
- Calcedo, V., 2013. Cantabria en el sector lácteo español: un radical cambio estructural. *Revista De Estudios Agro-Sociales y Pesqueros*. 234, 13–48. <http://hdl.handle.net/10902/5895>.
- Caruso, J.C., Cliff, N., 1997. Empirical Size, Coverage, and Power of Confidence Intervals for Spearman's Rho. *Educ. Psychol. Meas.* 57, 637–654. <https://doi.org/10.1177/0013164497057004009>.
- Castel Genís, J.M., Roszkowska-Madra, B., Mena Guerrero, Y., Lupa, W., Dabrowski, M., Madry, W., Gozdowski, D., 2010. Family dairy farms in the Podlasie province, Poland: farm typology according to farming system. *Span. J. Agric. Res.* 8 (4), 946–961.
- Celorio, S., Argüello, S.D., Barquín, F., Crespo, M.J., Chomón, N., Cimadevilla, C., Calderón, L., Cañón, J., 2011. Caracterización estructural de las explotaciones de ganado bovino de raza Pasiéga. *Archivos De Zootecnia* 60 (231), 409–412.
- Connell, J., Navarro, L., Torrealba, M., Rodríguez, I., Guevara, E., Ramírez, E.M., Coromoto, A., Tirado, H., 2007. Caracterización técnica productiva de los sistemas ganaderos del sur del estado Anzoátegui: Manejo del recurso pastizal. *Zootec. Trop.* 25 (3), 201–204.
- Cortez-Arriola, J., Rossing, W.A., Massiotti, R.D.A., Scholberg, J.M., Groot, J.C., Tittonell, P., 2015. Leverages for on-farm innovation from farm typologies? An illustration for family-based dairy farms in north-west Michoacán, Mexico. *Agric. Syst.* 135, 66–76. <https://doi.org/10.1016/j.agsy.2014.12.005>.
- Doltra, J., Villar, A., Moros, R., Salcedo, G., Hutchings, N.J., Kristensen, I.S., 2018. Forage management to improve on-farm feed production, nitrogen fluxes and greenhouse gas emissions from dairy systems in a wet temperate region. *Agr. Syst.* 160, 70–78. <https://doi.org/10.1016/j.agsy.2017.11.004>.
- EC (European Commission), 2012. Agriculture, farm accounting data network, sample selection. Retrieved April 6, 2012, from http://ec.europa.eu/agriculture/rca/methology2_en.cfm.
- Ferreira-Golpe, M.A.F., García-Arias, A.I.G., Pérez-Fra, M.P., & Vázquez-González, I. V., 2021. Impacto socioeconómico de la actividad apícola y características productivas del sector en la provincia de A Coruña. *ITEA*. 117(3), 295–310. [10.12706/itea.2020.031](https://doi.org/10.12706/itea.2020.031).
- Flores-Calvete, G., Martínez-Fernández, A., Doltra, J., García, A., Eguino, P., 2017. Encuesta sobre estructura y sistemas de alimentación de las explotaciones lecheras

- de Galicia, Cornisa Cantábrica y Navarra. Proyecto INIA-RTA2012-00065-C05. INTIA, 52 pp. <http://ciam.gal/pdf/informeinia.pdf>.
- García-Arias, A.I., Vázquez-González, I., Sineiro-García, F., Perez-Fra, M., 2015. Farm diversification strategies in northwestern Spain: Factors affecting transitional pathways. *Land Use Policy*. 49, 413–425. <https://doi.org/10.1016/j.landusepol.2015.08.011>.
- García-Pérez, C., Villalba-Mata, D., Casals-Maestre, R., Blanco-Penedo, I., 2022. Caracterización del bienestar animal en explotaciones de vacuno lechero de la raza Holstein del noreste de España. *ITEA*. 118(2), 239–261. 10.12706/itea.2021.023.
- García-Suárez, E., 2021. El sector bovino en Cantabria. Dinámica, caracterización y análisis de su sostenibilidad. Tesis Doctoral, Universidad De Santiago De Compostela, España. <http://hdl.handle.net/10347/26973>.
- Gökdağ, A., Magrin, L., Sakarya, E., Contiero, B., Gottardo, F., 2020. Characterization and typologies of dairy goat farms in the Mediterranean region: A case of Italy and Turkey. *Small Rumin. Res.* 191, 106–196. <https://doi.org/10.1016/j.smallrumres.2020.106196>.
- González-Mejía, A., Styles, D., Wilson, P., Gibbons J., 2018. Metrics and methods for characterizing dairy farm intensification using farm survey data. *PLOS ONE* 13(5): e0195286. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0195286> 10.1371/journal.pone.0195286.
- Graskemper, V., Yu, X., Feil, J.H., 2021. Farmer typology and implications for policy design—An unsupervised machine learning approach. *Land Use Policy* 103, 105328. <https://doi.org/10.1016/j.landusepol.2021.105328>.
- Hassall, K.L., Baudron, F., MacLaren, C., Cairns, J.E., Ndhlela, T., McGrath, S.P., Nyagumbo, I., Haefele, S.M., 2023. Construction of a generalised farm typology to aid selection, targeting and scaling of onfarm research. *Comput. Electron. Agric.* 212, 108074 <https://doi.org/10.1016/j.compag.2023.108074>.
- Kaouche-Adjlane, S.G.F.M.A., Ghoulane, F., Mati, A., 2015. Typology of dairy farming systems in the Mediterranean basin (Case of Algeria). *Biotechnol. Anim. Husbandry*. 31 (3), 385–396. <https://doi.org/10.2298/BAH1503385K>.
- Kobrich, C., Rehman, T., Khan, M., 2003. Typification of farming systems for constructing representative farming models: two illustrations of the application of multi-variate analyses in Chile and Pakistan. *Agric. Syst.* 76, 141–157. [https://doi.org/10.1016/S0308-521X\(02\)00013-6](https://doi.org/10.1016/S0308-521X(02)00013-6).
- López Iglesias, E., Lainez Andrés, M., 2022. El sector lácteo en España. Cajamar.
- Mađry, W., Mena, Y., Roszkowska-Mađra, B., Gozdowski, D., Hryniewski, R., Castel, J.M., 2013. An overview of farming system typology methodologies and its use in the study of pasture-based farming system: a review. *Span. J. Agric. Res.* 11 (2), 316–326. <https://doi.org/10.5424/sjar/2013112-3295>.
- MAPA (2020). Estructura del sector vacuno lechero en España 2016-2020. Obtenido de: Subdirección General de Producciones Ganaderas y Cinegéticas, Dirección General de Producciones y Mercados Agrarios, Ministerio de Agricultura, Pesca y alimentación.
- MAPA (2023). Estructura del sector vacuno lechero en España 2018-2022. Obtenido de: Subdirección General de Producciones Ganaderas y Cinegéticas, Dirección General de Producciones y Mercados Agrarios, Ministerio de Agricultura, Pesca y alimentación.
- Maseda, F., Díaz, F., Alvarez, C., 2004. Family dairy farms in Galicia (NW Spain): classification by some family and farm factors relevant to quality of life. *Biosyst. Eng.* 87 (4), 509–521. <https://doi.org/10.1016/j.biosystemseng.2004.01.00>.
- Mateus Silveira, R.M.F., da Silva, V.J., Ferreira, J., dos Santos Fontenelle, R.O., Vega, W. H.O., Sales, D.C., Sales, A.P., Castro, M.S.M., Toro-Mujica, P., de Vasconcelos, A.M., 2022. Diversity in smallholder dairy production systems in the Brazilian semiarid region: Farm typologies and characteristics of raw milk and water used in milking. *J. Arid Environ.* 203, 104774 <https://doi.org/10.1016/j.jaridenv.2022.104774>.
- Pérez, C., Santin, D., 2007. Minería de datos. Técnicas y herramientas. Madrid, España: Paraninfo Cengage Learning.
- Pérez-Méndez, J.A., Roibás, D., Wall, A., 2020. Somatic cell counts, reproduction indicators, and technical efficiency in milk production: A stochastic frontier analysis for Spanish dairy farms. *J. Dairy Sci.* 103 (8), 7141–7154. <https://doi.org/10.3168/jds.2019-17146>.
- Perrot, C., 1990. Typologie d'exploitations construite par agrégation autour de pôles définis à dire d'experts Proposition méthodologique et premiers résultats obtenus en Haute-Marne. *INRA. Prod. Anim.* 3 (1), 51–66. <https://doi.org/10.20870/productions-animales.1990.3.1.4360>.
- Riveiro, J.A., Mantecón, A.R., Álvarez, C.J., Lavín, P., 2013. A typological characterization of dairy Assaf breed sheep farms at NW of Spain based on structural factor. *Agr. Syst.* 120, 27–37. <https://doi.org/10.1016/j.agsy.2013.05.004>.
- Riveiro-Valiño, J.A., Álvarez-López, C.J., Marey-Pérez, M.F., 2009. The use of discriminant analysis to validate a methodology for classifying farms based on a combinatorial algorithm. *Comput. Electron. Agric.* 66 (2), 113–120. <https://doi.org/10.1016/j.compag.2008.12.001>.
- Ruiz, F.A., Vázquez, M., Camuñez, J.A., Castel, J.M., Mena, Y., 2020. Characterization and challenges of livestock farming in Mediterranean protected mountain areas (Sierra Nevada, Spain). *Spanish J. Agric. Res.* 18 (1) <https://doi.org/10.5424/sjar/2020181-14288>.
- Ruiz-Escudero, F., Vázquez-González, I., Búsque-Marcos, J., Villar-Bonet, A., 2023. Manejo productivo y territorial de las ganaderías lecheras de Cantabria según el contenido de forraje en la dieta. *ITEA*. Accepted 28 september 2023.
- Salcedo, G., Varsaki, A., Salcedo-Rodríguez, D., 2022. Huella hídrica de las explotaciones lecheras según tipología de alimentación. *Informacion Tecnica Economica Agraria* 118 (4). <https://doi.org/10.12706/itea.2021.040>.
- Salinas-Martínez, J.A., Posadas-Domínguez, R.R., Morales-Díaz, L.D., Rebollar-Rebollar, S., Rojo-Rubio, R., 2020. Cost analysis and economic optimization of small-scale dairy production systems in Mexico. *Livest. Sci.* 237, 104028 <https://doi.org/10.1016/j.livsci.2020.104028>.
- Schwering, D.S., Bergmann, L., Sonntag, W.I., 2022. How to encourage farmers to digitize? A study on user typologies and motivations of farm management information systems. *Comput. Electron. Agric.* 199, 107133 <https://doi.org/10.1016/j.compag.2022.107133>.
- Serrano-Martínez, E.S., González, P.L., García, F.G., Jal, A.B., Mantecón, A.R., 2004a. Classification variables of cattle farms in the mountains of León, Spain. *Span. J. Agric. Res.* 2 (4), 504–511.
- Serrano-Martínez, E.S., González, M.P.L., Mantecón, Á.R., Jal, A.B., García, F.J.G., 2004b. The identification of homogeneous groups of cattle farms in the mountains of León, Spain. *Span. J. Agric. Res.* 2 (4), 512–523.
- Siegmund-Schultze, M., Rischkowsky, B., 2001. Relating household characteristics to urban sheep keeping in West Africa. *Agr. Syst.* 67 (3), 139–152. [https://doi.org/10.1016/S0308-521X\(00\)00052-4](https://doi.org/10.1016/S0308-521X(00)00052-4).
- Silva, D., Rincón, N., Pérez, Y., Peña, M., Urdaneta, F., 2007. Nudos críticos de procesos gerenciales en unidades de producción de ganadería de doble propósito del municipio Machiques de Perijá, estado Zulia. *Revista De La Facultad De Agronomía De La Universidad Del Zulia*. 24 (3), 547–576.
- Srairi, M.T., Lyoubi, R., 2003. Typology of dairy farming systems in Rabat suburban region. *Morocco. Archivos De Zootecnia*. 52 (197), 47–58.
- Todde, G., Murgia, L., Caria, M., Pazzona, A., 2016. A multivariate statistical analysis approach to characterize mechanization, structural and energy profile in Italian dairy farms. *Energy Rep.* 2, 129–134. <https://doi.org/10.1016/j.egyr.2016.05.006>.
- Vázquez-González, I., García-Suárez, E., Ruiz-Escudero, F., Vilela-Ferreira, G.C., García-Arias, A.I., 2022. Construcción de un índice de sostenibilidad para las ganaderías de bovino lechero en Cantabria. *Economía Agraria y Recursos Naturales-Agricultural and Resource Economics*. 22 (2), 117–149. <https://doi.org/10.7201/earn.2022.02.06>.
- Vázquez-González, I., Ruiz-Escudero, F., Busqué, J., Villar, A., 2023. Potential land use of Cantabria for grass-fed milk production. *Span. J. Agric. Res.* 21 (4), e0608.
- Vilela-Ferreira, G.C.V., Miziara, F., Vazquez-González, I., 2021. Intensificação da pecuária em Goiás. *Rev. Econ. Sociol. Rural*. 60 <https://doi.org/10.1590/1806-9479.2021.242960>.
- Zou, D., Lloyd, J.E., Baumbusch, J.L., 2020. Using SPSS to analyze complex survey data: a primer. *Journal of Modern Applied. Stat. Methods*. 18 (1), 16. <https://doi.org/10.22237/jmasm/1556670300>.