

INVESTIGACIÓN

Application of artificial neural networks to determine the authentication of fattening diets of Iberian pigs according to their triacylglycerol profiles

By M. Narváez-Rivas¹, E. Gallardo¹, J.M. Jurado², I. Viera-Alcaide¹, F. Pablos²
and M. León-Camacho^{1,*}

¹ Food Characterization and Quality Department, Instituto de la Grasa (C.S.I.C.), Spain

² Department of Analytical Chemistry, Faculty of Chemistry, University of Seville,
c/ Profesor García González 1, E-41012 Seville, Spain

* Corresponding author: mleon@cica.es

RESUMEN

Aplicación de redes neuronales para resolver la autenticación de la alimentación de cerdos ibéricos de acuerdo a su perfil de triglicéridos

Se ha determinado mediante cromatografía de gases con detector de ionización de llama los triglicéridos de la grasa subcutánea de cerdos ibéricos, cebados con cuatro tipos de alimentación: montanera, recebo, cebo extensivo y cebo intensivo. Los análisis se realizaron en una columna con una fase estacionaria ligada químicamente (50% fenil-50% metilpolisiloxano) usando hidrógeno como gas portador. La grasa subcutánea se extrajo por fusión en horno de microondas, posteriormente se filtró y se disolvió en hexano. Un total de 2.783 muestras de varias campañas fueron analizadas. Usando los triglicéridos como descriptores químicos se ha llevado a cabo un estudio sobre la capacidad de discriminación de éstos para diferenciar el tipo y régimen de alimentación de los cerdos. A tal fin, se han empleado técnicas de reconocimiento de patrones, tales como análisis discriminante lineal (LDA) y redes neuronales artificiales de perceptrones multicapa (ANN-MLP). Las ANN presentan mejores resultados que el LDA, con una capacidad de predicción media de aproximadamente 97% en la diferenciación del tipo de alimentación entre *Montanera*, *Cebo* extensivo y *Cebo* intensivo. Al incluir el recebo, el modelo presenta un rendimiento promedio de 82%. La diferenciación del régimen de cebado también se ha llevado a cabo por medio de la ANN, con un rendimiento promedio del 96%.

PALABRAS CLAVE: Cerdo ibérico – Cromatografía de gases – Grasa Subcutánea – Reconocimiento de patrones – Redes neuronales – Triglicéridos.

SUMMARY

Application of artificial neural networks to determine the authentication of fattening diets of Iberian pigs according to their triacylglycerol profiles

The triacylglycerols in the subcutaneous fat from Iberian pigs reared on four different feeding types, *Montanera*, *Recebo*, extensive *Cebo* and intensive *Cebo*, have been determined by gas chromatography with a flame ionization detector. Analyses were performed in a column coated with a bonded stationary phase (50% phenyl-50%

methylpolisiloxane) with hydrogen as the carrier gas. Lipids were extracted by melting the subcutaneous fat in a microwave oven and then filtering and dissolving it in hexane. A total amount of 2783 samples from several campaigns were considered. Using the triacylglycerols as chemical descriptors, a study on the discriminating power to differentiate samples according to the pig feeding type and system was performed. With this aim, pattern recognition techniques, such as linear discriminant analysis (LDA) and multilayer perceptron artificial neural networks (MLP-ANN), have been used. ANN performed better than LDA, with a mean prediction ability of approximately 97% in the differentiation of fattening diets such as *Montanera*, extensive *Cebo* and intensive *Cebo*. In the case of including the *recebo* fattening diet, the model presents a mean performance of 82%. The differentiation of fattening systems has also been achieved by means of ANN, with a mean performance of 96%.

KEY-WORDS: Gas chromatography – Iberian pig – Neural networks – Pattern recognition – Subcutaneous fat – Triacylglycerols.

1. INTRODUCTION

The Iberian pig is an autochthonous porcine breed from the southwestern Iberian Peninsula traditionally fattened in a free-range system (extensive). According to the different type of feeding in the final stage of fattening, different qualities of Iberian dry-cured products are obtained. This final fattening diet of the animals determines their prices in the market. Therefore, products from pigs fattened according to an extensive system, based on *montanera*, which consists of the exploitation of acorn and grass, reach higher prices in markets than those fed mixed diets because they are more appreciated by consumers. Quality control measures have been proposed and established to avoid fraud in the marketing of the three commercial types of meat products from Iberian pigs depending on their production background: *montanera*, *recebo* (fed with acorn, pasture and formulated feed) or

cebo (fed with mixed feed) (López-Bote, 1998). In 2004, the Spanish government established a regulation to fix these three Iberian pig products according to the fatty acid composition (palmitic, stearic, oleic and linoleic) from the total lipids of the pig's subcutaneous adipose tissue (BOE, 2004). However, because of the use of feeds enriched with fatty acids similar to those found in acorns and the increase in formulated feeds in an extensive regimen during the final fattening period, the Spanish government has set a new regulation to fix the quality standards of Iberian pig products taking into account the extensive and intensive systems, collecting four different types of meat products: *Montanera*, *Recebo Extensive cebo* and *Intensive cebo* (BOE, 2007). But in this latest regulation, no analytical method is indicated although an efficient method is needed to prevent the wrongful use of the commercial name of higher quality products.

Several approaches have been applied in order to determine the final fattening diet received by the Iberian pigs. The profiles of fatty acids (Cava *et al.*, 1997, Flores *et al.*, 1988, Ruiz *et al.*, 1998), triacylglycerols (Díaz *et al.*, 1996, Narváez-Rivas *et al.*, 2009, Viera-Alcaide *et al.*, 2007, 2008), hydrocarbons (Viera-Alcaide *et al.*, 2009; Gamero-Pasadas *et al.*, 2006; Narváez-Rivas *et al.*, 2008) and volatile compounds (Narváez-Rivas *et al.*, 2010, 2011) have been related with the fattening diets of Iberian pigs and can be used as chemical descriptors to differentiate among the different feeding backgrounds of the animals. Near-infrared spectrometry (NIR) is another method that has been used for the authentication of this animal fattening diet (Arce *et al.*, 2009; Hervás *et al.*, 1994; Zamora-Rojas *et al.*, 2011, 2012). This is a very simple, fast, cheap and nondestructive technique, but it does not differentiate perfectly among the different fattening systems and does not give information about the sample chemical composition.

One of the methods that would be a useful tool to authenticate Iberian pig products from different fattening diets is the analysis of triacylglycerols because this shows several advantages: i) it is faster than others that require a more or less long time for the derivatization or recovery; ii) it is carried out without a previous treatment or loss of them since it is performed by direct injection of the sample (fat) dissolved in hexane; and iii) it allows for differentiating among the three fattening diets (*montanera*, *recebo* and *cebo*) (Viera-Alcaide *et*

al., 2008) and between extensive and intensive systems (Gallardo *et al.*, 2012). Previously, the triacylglycerol profile has been extensively used for the characterization and authentication of several foodstuffs such as olive oil, vegetable oils, animal fats, fish oils, milk and dairy products, cocoa and coffee (Bosque-Sendra *et al.*, 2012; González *et al.*, 2001).

In this work triacylglycerols in a large number of Iberian pig subcutaneous fat samples from different campaigns have been determined by gas chromatography with flame ionization detection (GC-FID). Differences in the composition of the triacylglycerol profiles have been used to differentiate among four fattening diet types (*montanera*, *recebo*, *extensive cebo* and *intensive cebo*). Using them as chemical descriptors, pattern recognition (PR) techniques, such as principal component analysis (PCA), linear discriminant analysis (LDA) and artificial neural networks (ANN) have been applied.

2. MATERIALS AND METHODS

2.1. Chemicals and reagents

Trilinolein (LLL), triolein (OOO), tripalmitin (PPP) and tristearin (SSS) were obtained from Sigma Aldrich Fluka (Steinheim, Germany). Standard solutions of 1.0% (m/v) of the triacylglycerols were prepared by dissolving them in n-hexane (Romil, Cambridge, UK) of analytical reagent grade.

2.2. Samples

A total of 2783 samples of subcutaneous fat from castrated pure Iberian male pigs from the campaigns corresponding to the years 2002, 2003, 2004, 2005, 2009, 2010 and 2011 (Table 1) were analyzed. 1169 samples were from animals fed with a fattening diet based exclusively on acorn and pasture, usually called *montanera* (M). Another group consisted of 448 samples from animals fed with commercial feed and pasture, usually called *extensive cebo* (EC). 134 samples corresponded to animals fed with concentrated feed in an intensive system, *intensive cebo* (IC) and the last group (1032 samples) was made up of animals reared in *montanera* but also included commercial feeds in the final fattening period,

Table 1
Analyzed Iberian pig subcutaneous fat samples

Fattening diet	Campaigns						
	2002	2003	2004	2005	2009	2010	2011
<i>Montanera</i>	292	244	242	170	73	55	93
<i>Recebo</i>	187	229	289	181	67	50	29
<i>Extensive cebo</i>	107	119	63	16	36	75	32
<i>Intensive cebo</i>	–	3	4	–	32	25	70

which is usually known as *recebo* (R). The outdoor rearing (*montanera*, *recebo* and extensive *cebo*) are called extensive systems (ES). When they are fed only with concentrated feed in farms (intensive *cebo*) it is called intensive system (IS). Samples were kindly provided by the Designations of origin "Jamón de Huelva", and INIA (Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria). They were taken from different zones of production located in the southwest of Spain (Huelva, Cádiz, Sevilla, Cáceres, Badajoz, Salamanca, Córdoba and Málaga). The animals were classified into the different groups according to the field notes taken by the veterinary inspector of the Designations of origin and INIA during the final fattening period.

2.3. Apparatus and methods

2.3.1. Extraction of the subcutaneous fat

Two samples were taken from each sacrificed animal according to a normalized procedure described in the literature (BOE, 2004). Briefly, the procedure consists of cutting a piece of approximately 3 × 3 cm with at least 6 mm of thickness from the rump, at about 10 cm of the tail following the line of the back and containing skin, adipose tissue and some loin. At the laboratory, the skin and loin were carefully removed (BOE, 2004). All the chunks corresponding to each animal of a sacrifice lot were punctured and homogenized before extraction. The representative sample of the lot was then obtained by melting the fat in a microwave oven for 3 minutes at 360 W (De Pedro *et al.*, 1997). The fat samples were then filtered and 0.05 mg of fat were dissolved in 1.5 mL of n-hexane for the gas chromatography analysis. Three replicates were determined for each sample.

2.3.2. Determination of the triacylglycerols by GC

Triacylglycerols were analyzed and identified as described previously by gas chromatography (Viera-Alcaide *et al.*, 2007) in a Varian 3800 gas chromatograph (Varian Co, Palo Alto, CA, USA) using a DB-17HT (Agilent J&W, Loveland, CO, USA) fused silica capillary column (30 m long × 0.25 mm i.d. 0.15 mm film thickness). The oven temperature was kept at 320 °C, and was then raised to 350 °C at a rate of 2.0 °C min⁻¹ and held isothermally for 10 min. The injector temperature was kept at 360 °C, while the detector temperature was 370 °C. Hydrogen (2.1 mL min⁻¹ column constant flow) was used as carrier gas and the make-up gas was nitrogen. Aliquots of 2 µL were injected.

The identification of seventeen triacylglycerol species was carried out by means of standards of trilinolein (LLL), triolein (OOO), tripalmitin (PPP) and tristearin (SSS) and by comparison with the relative retention times described previously in the literature (Viera-Alcaide *et al.*, 2007). Table 2 includes the

Table 2
Triacylglycerols identified in fat samples

Peak	Triacylglycerol	T _{RR}
1	PPP	0.65
2	MOP	0.67
3	PPS	0.79
4	POP	0.81
5	POP _o +PLP	0.84
6	PLP _o +MLO	0.87
7	PSS	0.95
8	PSO	0.97
9	POO	1.00
10	PLO	1.03
11	PLL+PoLO	1.07
12	SOS	1.14
13	SOO	1.17
14	OOO	1.19
15	SOL	1.21
16	OOL	1.24
17	OLL	1.29

T_{RR}, relative retention time. M, Miristic; P, Palmitic; Po, Palmitoleic; S, Stearic; O, Oleic; L, Linoleic. See Figure 1 for peak numbers

different triacylglycerol species identified in the chromatograms shown in Figure 1. Palmitindioleine was used as a reference to calculate the relative retention times.

2.3.3. Data analysis

The seventeen triacylglycerols were used as chemical descriptors and their peak areas as analytical signals. The quantification of each one was carried out by evaluating the corresponding relative percentage according to the normalization area procedure, assuming an equal factor response for any species. A data matrix, whose rows are the samples and columns the variables, was built. Each element of this matrix *x_{ij}* corresponds to the content of the triacylglycerol *j* for the sample *i*. The PR calculations were made using the statistical package CSS:STATISTICA from Statsoft™ (Tulsa, OK, USA).

3. RESULTS AND DISCUSSION

3.1. Fat samples analysis

Table 3 shows the mean, standard deviation, median and range values of all the samples and groups of different fattening type for the triacylglycerols determined in the analyzed fat samples corresponding to the seven different campaigns considered and

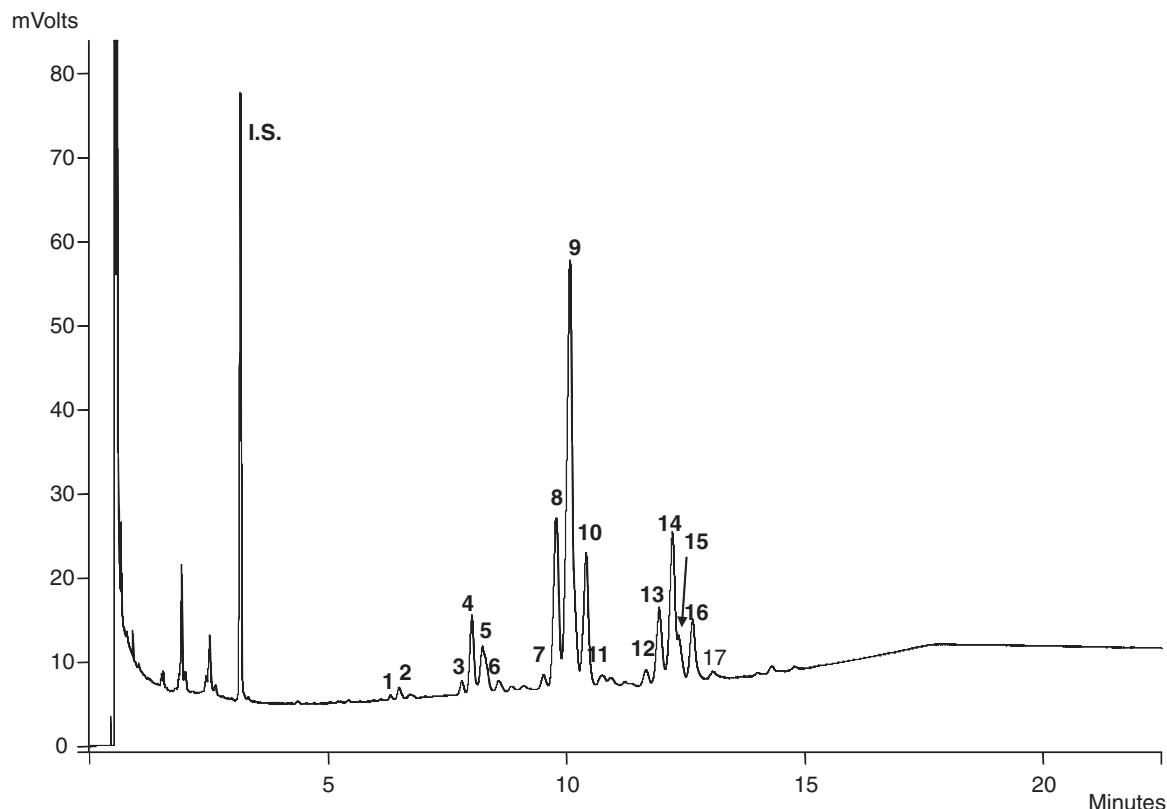


Figure 1
Chromatogram of the triacylglycerol profiles of Iberian pig subcutaneous fat sample. Peak identification: see table 2.

grouped according to the four fattening types. The results for the different species of triacylglycerols are in good agreement with those reported by other authors (Viera-Alcaide *et al.*, 2008). As can be seen, the most abundant triacylglycerol is POO, with median values of 31.87%. Other major triacylglycerols are PSO, OOO and PLO, with median values of 14.05, 10.39 and 10.27%, respectively. POP, POPo+PLP, SOO, SOL and OOL present medians that range between 3.67 and 6.63%. The remaining compounds range from 0.21%, for PPP, to 1.30%, for SOS. In the case of PLO, SOO, OOO, SOL, OOL and OLL, samples obtained from pigs fed in the extensive system (*montanera*, *recebo* and *extensive cebo*) present higher mean and median values than those from intensive fattening (*intensive cebo*). The mean and median contents in the remaining triacylglycerols are higher in intensive fattening samples (*intensive cebo*).

3.2. Classification of the different fattening types

The Kruskal-Wallis test was applied in order to find out significant differences among the triacylglycerol contents of samples from *montanera*, *extensive cebo* and *intensive cebo* fattening diets. This test calculates the H parameter for comparison with the chi-squared distribution for $n-1$ degrees

of freedom and $\alpha=0.05$, where n is the number of groups being considered (Muth, 1999). In order to highlight significant differences among the three types of fattening diets a *post hoc* comparison was also carried out. This test was applied to samples from campaigns from 2002 to 2005, because these cases were initially considered to build a first model to be extended to the campaigns 2009-2011. Table 4 shows the H -value obtained for the seventeen considered variables and the results of the *post hoc* analysis. It can be seen that the largest H -values are obtained for OOO, PSO, OOL, POP and OLL. These variables present significant differences between *montanera* and the two type of *cebo* fattening systems as well as PPP, MOP, PPS, POPo+PLP, PPS and SOS. *Intensive cebo* is statistically different from *extensive cebo* and *montanera* considering the contents of PLL+PoLO and SOL, but lower H -values are computed. POO and POPo+MLO show significant differences between *montanera* and *extensive cebo*. Only for SOO there are positive differentiations among all the possible comparisons. PLO does not present a significant H -value and it is not considered for further calculations.

In light of Kruskal-Wallis test results, triacylglycerol contents could be considered as potential chemical descriptors to differentiate feed types for Iberian pigs. To corroborate this thought, principal component analysis (PCA) was applied with the aim of finding trends in the

Table 3
Triacylglycerol contents (%) of Iberian pig subcutaneous fat samples analyzed during seven campaigns and grouped by fattening diets

		PPP	MOP	PPS	POP	POPo+PLP	PLPo+MLO	PSS	PSO	POO	PLO	PLL+PoLO	SOS	SOO	OOO	SOL	OOL	OLL
M (n=1169)	Mean	0.19	0.72	0.64	4.96	4.13	0.80	0.90	13.32	32.71	10.31	1.49	1.24	6.67	11.22	3.66	5.91	1.13
	S.D.	0.07	0.15	0.16	0.68	0.94	0.15	0.21	1.52	2.52	0.91	0.33	0.19	0.59	1.40	0.48	0.76	0.24
	Median	0.19	0.72	0.63	4.94	4.41	0.80	0.89	13.34	31.87	10.29	1.43	1.27	6.73	11.18	3.68	5.87	1.10
	Minimum	0.03	0.16	0.03	0.39	1.06	0.16	0.23	6.98	0.77	7.22	0.54	0.49	4.34	6.24	0.60	3.61	0.41
	Maximum	0.77	1.52	1.33	7.75	6.89	1.46	1.64	21.63	40.88	17.94	3.24	1.87	9.38	17.02	7.17	9.70	2.18
R (n=1032)	Mean	0.21	0.78	0.71	5.32	4.25	0.82	0.98	13.94	32.43	10.47	1.55	1.26	6.54	10.30	3.73	5.62	1.09
	S.D.	0.08	0.18	0.21	0.84	1.05	0.17	0.27	1.84	2.38	1.10	0.40	0.21	0.66	1.75	0.42	0.93	0.28
	Median	0.20	0.78	0.69	5.25	4.55	0.82	0.96	13.95	31.70	10.38	1.44	1.29	6.62	10.38	3.74	5.60	1.04
	Minimum	0.04	0.16	0.10	3.25	1.10	0.25	0.19	7.65	20.87	6.97	0.58	0.55	4.15	4.65	1.68	3.20	0.45
	Maximum	0.70	1.54	1.59	8.49	6.40	1.51	2.09	21.74	42.22	15.28	3.18	1.98	8.68	15.66	6.97	8.48	2.26
EC (n=448)	Mean	0.28	0.96	0.95	6.46	4.74	0.86	1.35	17.05	32.31	10.29	1.60	1.49	6.34	7.08	3.52	3.91	0.80
	S.D.	0.08	0.20	0.23	0.95	1.08	0.18	0.32	2.04	2.78	1.09	0.49	0.26	0.74	1.56	0.55	0.79	0.28
	Median	0.28	0.96	0.94	6.44	5.05	0.87	1.35	17.20	31.38	10.14	1.47	1.53	6.49	6.96	3.56	3.85	0.73
	Minimum	0.08	0.38	0.39	3.43	1.34	0.30	0.47	10.55	27.83	6.87	0.80	0.51	2.94	4.00	1.99	1.88	0.35
	Maximum	0.57	1.59	1.57	8.74	7.54	1.59	2.45	22.49	42.85	14.24	3.67	2.17	7.72	15.00	7.24	6.62	1.91
IC (n=134)	Mean	0.35	1.06	1.21	7.31	4.64	0.81	1.57	18.37	34.93	9.73	1.74	1.43	5.45	5.34	2.86	2.70	0.50
	S.D.	0.12	0.26	0.35	1.27	1.17	0.14	0.45	2.71	1.47	1.43	0.54	0.26	0.80	1.63	0.51	0.95	0.37
	Median	0.33	1.07	1.18	7.18	4.79	0.79	1.52	18.37	35.03	9.47	1.57	1.41	5.30	4.98	2.92	2.45	0.32
	Minimum	0.13	0.52	0.50	4.50	1.83	0.37	0.34	11.04	30.39	6.99	1.00	0.91	3.95	3.05	1.81	1.26	0.09
	Maximum	0.66	1.71	2.08	10.07	6.86	1.30	2.67	24.72	40.21	13.94	3.27	2.39	7.52	10.03	4.40	5.36	1.52
Total (n=2783)	Mean	0.22	0.80	0.74	5.45	4.29	0.82	1.03	14.39	32.65	10.34	1.54	1.30	6.51	9.93	3.62	5.33	1.03
	S.D.	0.09	0.20	0.25	1.06	1.04	0.16	0.33	2.39	2.53	1.05	0.40	0.23	0.71	2.36	0.51	1.24	0.32
	Median	0.21	0.77	0.70	5.25	4.53	0.82	0.98	14.05	31.87	10.27	1.44	1.30	6.63	10.39	3.67	5.49	1.03
	Minimum	0.03	0.16	0.03	0.39	1.06	0.16	0.19	6.98	0.77	6.87	0.54	0.49	2.94	3.05	0.60	1.26	0.09
	Maximum	0.77	1.71	2.08	10.07	7.54	1.59	2.67	24.72	42.85	17.94	3.67	2.39	9.38	17.02	7.24	9.70	2.26

M: *montanera*; R: *recebo*; EC: *extensive cebo*; IC: *intensive cebo*. n, number of samples of each type

sample distribution. PCA sequentially obtains new variables, called principal components (PCs), as a linear combination of the original ones and condenses the variance of the data set as much as possible for each successive PC. Thus, a graphical representation of the cases in the space of the first two PCs allows for an easier visualization of data trends with a lower dimensionality (Jolliffe, 2002). In this case, the first three PCs account for 52.53%, 13.78% and 10.64% of total variance, respectively. The most contributing variables to PC1 were PPP, MOP, PPS, POP, PPS, PSO, SOS, OOO, OOL and OLL. PC2 is highly correlated to PLO and PLL+PoLO and PC3 to SOO. The distribution of the data in the plane of the two first PCs is shown in Figure 2. Some separation can be observed for samples from *montanera* and *extensive cebo* according to their scores of PC1, as might be expected taking into account the most contributing

variables to this PC and the results of the Kruskal-Wallis test. Samples from *intensive cebo* appear completely separated from *montanera* in the space of these two PCs. In light of such results, linear discriminant analysis was applied to build an adequate classification model.

LDA computes linear combinations of the data to obtain discriminant functions (DFs) as linear combinations of the original variables. These DFs separate the considered categories by minimizing the within-class and between-class ratio of the sum of squares. The model can be constructed through a stepwise approach by selecting only the most discriminating variables and reducing the number of variables used as chemical descriptors (Massart, 1998). In this case, a forward stepwise approach was used, obtaining two DFs in order to differentiate *montanera*, *extensive cebo* and *intensive cebo* from campaigns 2002 to 2005. This model eliminates the

Table 4
Kruskal-Wallis test results

Variable	H ^a	Significant difference in <i>post hoc</i> comparison		
		M-EC	M-IC	EC-IC
PPP	358.7	x	X	
MOP	490.9	x	X	
PPS	491.1	X	X	
POP	629.5	X	X	
POPo+PLP	364.6	X	X	
POPo+MLO	119.3	X		
PPS	513.6	X	X	
PSO	657.7	X	X	
POO	142.5	X		
PLO	2.3			
PLL+PoLO	11.5		X	X
SOS	433.9	X	X	
SOO	50.6	X	X	X
OOO	711.0	X	X	
SOL	22.6		X	X
OOL	647.6	X	X	
OLL	577.5	X	X	

^a Significant difference for $H > 5.99$

variables POO and PLL+PoLO, using the remaining 14 variables to solve the classification problem. The recognition ability of the model, defined as the percentage of samples from those used to compute the DFs correctly assigned to their class, was 98.1%, 92.8% and 100% for *montanera*, *extensive cebo* and *intensive cebo*, respectively. As can be seen in the distribution of the samples in the plane of the two computed DFs (Figure 3), the considered classes appear separately.

According to these results, the contents of PPP, MOP, PPS, POP, POPo+PLP, PLPo+MLO,

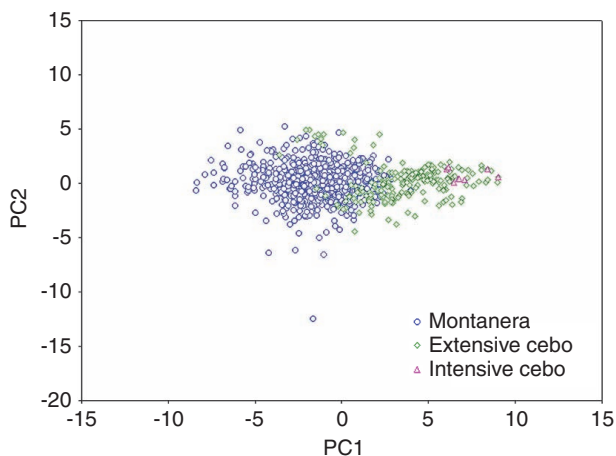


Figure 2
Score plot in the plane of the two first PCs.

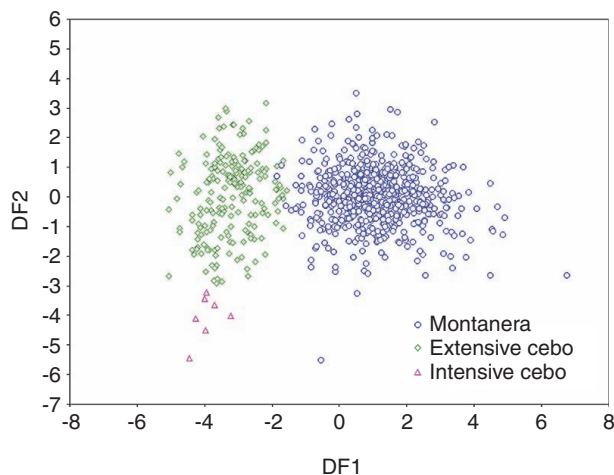


Figure 3
Distribution of the samples according to their scores for the discriminant functions.

PSS, PSO, SOS, SOO, OOO, SOL, OOL and OLL, could be used as discriminant variables to differentiate the three considered classes. Using the listed variables, different models were built in order to classify samples from the campaigns 2009, 2010 and 2011. The models were obtained using a training set of cases to find out the relationship between the discriminant functions and the original variables and a test set to study the model performance. Two parameters, sensitivity (SENS) and specificity (SPEC) were computed from the classification matrix of the test set. SENS of a class represents the percentage of cases correctly classified in that class and, SPEC corresponds to the percentage of objects not belonging to a certain class and subsequently classified as pertaining to another (Forina *et al.*, 1991). The results of all models are shown in Table 5.

The first model (LDA1a) was built by using samples of *montanera*, *extensive cebo* and *intensive cebo* from campaigns 2002-2005 as training cases and samples from 2009 as test ones. In this case the prediction ability was 84.9% for *montanera*, 41.7% for *extensive cebo* and 65.6% for *intensive cebo*. SPEC varies from 84.4% to 88.6% for *intensive cebo* and *extensive cebo*, respectively. In order to improve these results, a second model (LDA1b) was constructed by adding 2/3 of the samples from campaign 2009 to the training set. These samples were selected randomly. When the model was applied to the test set, which consisted of 1/3 of samples from 2009, SENS was increased to 91.8% and 100% for *montanera* and *intensive cebo*. The class *extensive cebo* presented the lowest SENS value (44.4%). SPEC of the three classes was also improved. The model LDA2a was obtained using samples from campaigns 2002-2005 and 2009 as training cases and was applied to test cases from 2010. In this case, SENS of *montanera* and *extensive cebo* were 96.4% and 74.7%, respectively. Although no samples were incorrectly classified as intensive

Table 5
Sensitivities and specificities (%) obtained for the models built

Model	Montanera		Extensive cebo		Intensive cebo		Recebo		Overall	
	SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA1a	84.9	79.4	41.7	88.6	65.6	84.4			64.1	84.1
LDA1b	91.8	82.4	44.4	97.1	100	89.9			78.7	89.8
LDA2a	96.4	81	74.7	67.5	4	100			58.4	82.8
LDA2b	94.4	93.9	88	96.2	100	97.7			94.1	95.9
LDA3a	78.5	95.1	46.9	62.6	38.6	88.8			54.7	82.2
LDA3b	77.4	94.1	27.3	85.2	78.3	76.2			61	85.2
ANN1a	98.6	79.4	61.1	90.5	71.9	100			77.2	90
ANN1b	100	95.7	91.7	100	100	100			97.2	98.6
ANN2a	92.7	98	97.3	63.8	0	100			63.3	87.3
ANN2b	100	97	96	100	100	100			98.7	99
ANN3a	88.2	88.2	75	73	47.1	100			70.1	87.1
ANN3b	96.8	97.1	90	96.4	95.8	100			94.2	97.8
LDA1ar	79.5	74.8	33.3	86.6	68.8	76.7	19.4	96.5	50.3	83.7
LDA1br	79.2	71.1	41.7	91.2	100	84.5	18.2	93.6	59.8	85.1
LDA2ar	56.4	94.7	57.3	79.2	4	100	84	65.8	50.4	84.9
LDA2br	61.1	96	92	93	100	100	76.5	84.3	82.4	93.3
LDA3ar	46.2	96.2	34.4	62	38.6	87.7	31	81	37.6	81.7
LDA3br	51.6	93.2	18.2	90.6	87	69.2	30	86.2	46.7	84.8
ANN1ar	83.6	84.4	55.6	88.4	37.5	99.4	85.1	88.7	65.5	90.2
ANN1br	91.7	100	75	100	100	100	100	89.4	91.7	97.4
ANN2ar	85.5	84	66.7	76.9	4	100	46	80.6	50.6	85.4
ANN2br	63.2	94	96	97.7	100	100	82.4	86.5	85.4	94.6
ANN3ar	83.9	87	56.3	78.6	32.9	99.4	44.8	83.1	54.5	87.0
ANN3br	93.5	86	60	90.6	87.5	100	33.3	95.4	68.6	93.0

cebo, as can be concluded from the SPEC value (100%), this class had a very low SENS (4%). When 2/3 of samples from 2010 were added to the training set, model LDA2b was obtained with 94.1% and 95.9% of overall SENS and SPEC, respectively. It could be thought that this is a good solution to the classification problem, but models LDA3a, constructed with samples from 2002-2005 and 2009-2010 and tested with samples from 2011 present bad results. The overall prediction ability hardly improves by using the same training cases and 2/3 of samples from 2011 (LDA3b).

In view of the fact that a linear approach did not provide an adequate solution to our classification problem, a non-linear model was applied. Artificial neural networks (ANN) mimic the biological nervous system and can be used where other modeling techniques cannot predict complex phenomena (Zupan *et al.*, 1993). In this case, multilayer perceptrons (MLP) ANN have been applied. MLP-ANN are feed forwarded networks consisting of neurons arranged in layers

(an input layer, various hidden layers and an output layer), with unidirectional connections from input to output. Just as LDA, ANNs learn from the data and a training set is necessary to find out the relationships between inputs and outputs, whilst a test set is used to show the prediction ability. The basic structure of the models was a 14:8:3 MLP-ANN, or 14 input neurons (one for each variable used in LDA models), 8 neurons in the hidden layer and 3 outputs (one for each of the considered classes). The training method was back propagation, minimizing the error made by the network. To avoid over-fitting problems, a third set of cases, the verification set, is required to cross-validate the models (Tetko *et al.*, 1995). Model ANN1a uses the same data matrix as LDA1a but, in this case, samples pertaining to campaigns from 2002 to 2005 were randomly divided into two subsets. 2/3 of the samples were used as training cases and 1/3 as verification cases. All samples from campaign 2009 were used as test cases. SENS of *montanera*, *extensive cebo* and *intensive*

cebo were 98.6%, 61.1% and 71.9%, respectively. These results improve when data from 2009 was divided into 1/3 training cases, 1/3 verification cases and 1/3 test cases. This model (ANN1b) leads to 100% prediction ability for *montanera* and intensive *cebo*, and 91.7% for extensive *cebo*. SPEC was 100% for the two last classes. ANN2a, using data from 2010 as test and the previous years as training (2/3) and verification (1/3), was unable to correctly classify any intensive *cebo* samples. However, the inclusion of samples from 2010 into training and verification sets solved this problem (ANN2b), leading to overall SENS and SPEC of 98.7% and 99.0%, respectively. The performance declines again when using data from 2002-2005 and 2009-2010 to train and verify the model ANN3a and samples from 2011 as test cases. But this performance also improves when dividing the samples from this year into training, verification and test (model ANN3b), with overall SENS and SPEC of 94.2% and 97.8%, respectively. From these results it can be concluded that non-linear models, such as MLP-ANN, perform better than linear ones in the classification of samples of Iberian pig fat according to the fattening diet. In addition, it is necessary to consider the annual variability between different campaigns to establish relationships between the triacylglycerol contents and the fattening diet. For this reason, every new model must include training samples from the campaign to be studied.

Once the classification problem of *montanera*, extensive *cebo* and intensive *cebo* is solved, it is important to pay attention to samples of *recebo*. If the developed models were used to classify samples of *recebo*, they are mostly assigned to *montanera* and extensive *cebo* classes, as the models do not include samples of *recebo* in the training set. To include *recebo* samples a new forward stepwise LDA model was computed using campaigns 2002-2005 to find out the most discriminant variables. The selected variables were the same fourteen used in the previous models. This initial LDA

presents recognition abilities of 65%, 92.1%, 100% and 51.7% for *montanera*, extensive *cebo*, intensive *cebo* and *recebo*, respectively. As can be seen in figure 4, samples of *recebo* appear overlapped with *montanera* and extensive *cebo* in the plane formed by the first two DFs. On the basis of these results, twelve new models have been developed including samples of *recebo* in the different subsets of cases. Codes of these models are the same as the twelve explained above because the same campaigns and percentage of training, verification and test cases were used, but the letter *r* is added to indicate that *recebo* is also included. For instance, LDA1ar implies that samples from 2002 to 2005 were used as training and samples from 2009 were used as test. The overall results of this model were SENS of 50.3% and SPEC of 83.7. Samples of *recebo* were confused with the other three classes. This situation does not improve by including samples from 2009 into the training set (LDA1br), with a SENS of 18.2% for the *recebo* class. Model LDA2ar, using campaigns 2002-2005 and 2009 as training set, improves the prediction ability for samples of *recebo* from 2010 to 84%, but intensive *cebo* falls to 4%. The best results from those models using LDA were obtained from LDA2br, with overall SENS and SPEC of 82.4% and 93.3%, respectively. The improvement is reversed by models using data from 2011, LDA3ar and LDA3br, which showed overall SENS values of 37.6 and 46.7 respectively.

ANN models including *recebo* improve LDA results, especially when samples of the tested campaign were included to build the model (series *br*). The best result was obtained by ANN1br, where SENS was 91.7% and SPEC 97.4%. SENS decreases slightly to 85.4% when samples from 2010 are incorporated (ANN2br). The same happens when using samples from 2011 (ANN3br), showing a SENS of 68.6%. This last model only correctly classified 33.3% of *recebo* and 60% of extensive *cebo*. In conclusion, the use of non-linear models as well as the inclusion of samples from the tested campaign into the training set improves LDA results, but the fact of considering *recebo* as a class lead to unsatisfactory results.

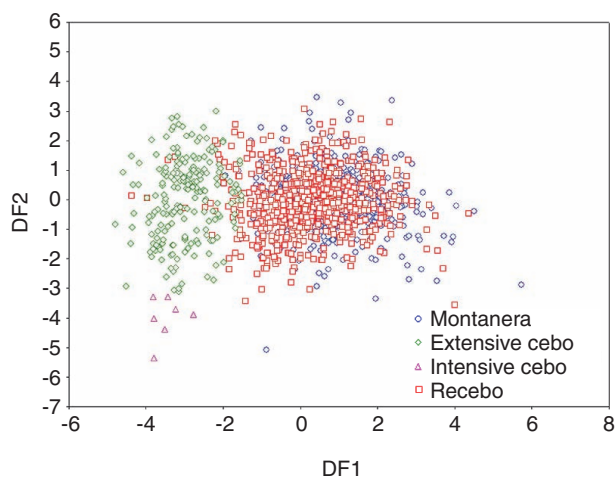


Figure 4

Distribution of the samples (including recebo) according to their scores for the discriminant functions.

3.3. Classification of different fattening systems

The same strategy has been used to study the differentiation capability of LDA and ANN according to the fattening system (extensive and intensive). Only SENS was computed because with only two classes, SPEC of a class has the same value of SENS of the other considered class. Results are shown in Table 6. An initial LDA model was built by using samples from campaigns 2002-2005 and a forward stepwise approach. The selected variables were PPP, MOP, PPS, POP, POP_o+PLP, SOS, SOL, OOL and OLL. Figure 5 shows the sample distribution according to the scores obtained for the computed DF. The recognition ability was 100% and 96.1% for intensive and extensive systems,

Table 6
Sensitivities (%) obtained for the models applied to the differentiation of extensive and intensive fattening systems

LDA models			ANN models		
Model	Extensive system	Intensive system	Model	Extensive system	Intensive system
LDA4a	74.4	84.4	ANN4a	100	61.5
LDA4b	100	81.0	ANN4b	100	100
LDA5a	100	12	ANN5a	100	4
LDA5b	96.7	100	ANN5b	100	100
LDA6a	91.6	50	ANN6a	99	60
LDA6b	71.2	82.6	ANN6b	98	79.2

respectively. This model was applied by using samples from 2009 as test set (LDA4a), with 74.4% and 84.4% of sensitivity for extensive and intensive systems, respectively. When 2/3 of the samples from 2009 are included in the training set, the SENS of the intensive fattening system rises to 100% and the value for the extensive system increases to 81.0%. Using samples from 2002-2005 and 2009 as training (LDA5a), samples from 2010 belonging to extensive systems are 100% correctly classified but a SENS of 12% is obtained for the intensive system. This value grows to 100% by using 2/3 of samples from this campaign as training cases, falling to 96.7% for the extensive system. Model LDA6a was built to test 2011 samples using cases from 2002-2005 and 2009-2010 as training set, with 91.6% and 50% of SENS for extensive and intensive fattening systems, respectively. These values improve to 71.2% and 82.6% when samples from 2011 are used as training cases.

ANN models were built in the same way as LDA, but using 1/3 of training cases as the verification set. Model ANN4a used samples from 2002-2005 to obtain a 9:10:2 network with 100% prediction ability

for extensive system samples from 2009. SENS for the intensive fattening system was of 61.5%. Using samples from the campaign 2009 as training (1/3) and verification (1/3) cases, results of 100% were obtained for both fattening systems (ANN4b). Model ANN5a, trained with samples from 2002-2005 and 2009, correctly classified all extensive system samples, but only 4% of the intensive system. These results improve to 100% for both classes with model ANN5b, which incorporates samples from 2010 to training and verification sets. Model ANN6a classified samples from 2011 with SENS of 99% and 60% for extensive and intensive fattening systems, respectively. By incorporating samples from 2011 to the training and verification sets, SENS remains approximately the same for extensive systems (98%) and improves to 79.2% for the intensive system. Generally, models based on ANN improve the results obtained by LDA and the incorporation of samples from the campaign to be tested also lead to the best results.

4. CONCLUSIONS

LDA and MLP-ANN have been applied to study the classification of Iberian pig fat samples according to their fattening diet. Fourteen variables were selected by forward stepwise LDA to obtain the models, using the contents of PPP, MOP, PPS, POP, POP_o+PLP, PLP_o+MLO, PSS, PSO, SOS, SOO, OOO, SOL, OOL and OLL. The models were built using samples from different campaigns and testing with cases from new campaigns. From this chemometric study it can be concluded that a non-linear approach, such as MLP-ANN, obtains better classification efficiencies than linear ones. On the other hand, due to the between-years variability, the use of training samples from the same campaign to be tested is essential. This fact can be due to different factors, such as the pasture composition or amounts of acorn consumed by the animals, and these are intrinsically related to the climate and rainfall of the particular year (Narváez-Rivas *et al.*, 2009 and Viera-Alcaide *et al.*, 2008). In addition to these conclusions, *recebo* samples

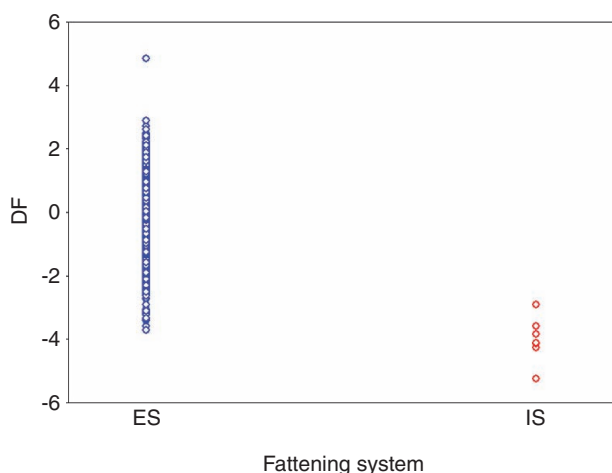


Figure 5

Distribution of samples from extensive (ES) and intensive (IS) fattening systems according to the scores obtained for the computed discriminant functions.

appeared mixed with *montanera* and extensive *cebo*. This fact depends on the procedure and sequence of feeding, by alternating feed and acorn, which is usually performed to mimic products of higher quality.

LDA and ANN models were also trained to carry out the differentiation of fattening systems. In this case the contents of PPP, MOP, PPS, POP, POP_o+PLP, SOS, SOL, OOL and OLL were selected by forward stepwise LDA. In view of the results, the same conclusions can be extracted: ANN models perform better than LDA and samples of the tested campaign must be included to take into account between-year variability.

ACKNOWLEDGMENTS

The authors are grateful to the INIA, especially to Dr. J. García Cascos (INIA) for his collaboration and help. This study was supported by INIA contract 2009X0983.

REFERENCES

- Arce L, Domínguez-Vidal A, Rodríguez-Estévez, López-Vidal S, Ayora-Cañada MJ, Valcárcel M. 2009. Feasibility study on the use of infrared spectroscopy for the direct authentication of Iberian pig fattening diet. *Anal. Chim. Acta* **636**, 183-189.
- Boletín Oficial del Estado (BOE). Orden PRE/3844/2004, por la que se establecen los métodos oficiales de toma de muestras en canales de cerdos ibéricos y el método de análisis para la determinación de la composición de ácidos grasos de los lípidos totales del tejido adiposo subcutáneo de cerdos ibéricos. BOE núm. 283, Madrid, Spain, 2004; 38770-38775.
- Boletín Oficial del Estado (BOE). REAL DECRETO 1469/2007, por el que se aprueba la norma de calidad para la carne, el jamón, la paleta y la caña de lomo ibéricos. BOE núm. Madrid, Spain, 2007; 45087-45104.
- Bosque-Sendra JM, Cuadros-Rodríguez L, Ruis-Sanblas C, De la Mata AP. 2012. Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data—A review. *Anal. Chim. Acta* **724**, 1-11.
- Cava R, Ruiz J, López-Bote C, Martín L, García C, Ventanas J, Antequera T. 1997. Influence of finishing diet on fatty acid profiles of intramuscular lipids, triglycerides and phospholipids in muscles of Iberian pig. *Meat Sci.* **45**, 263-170.
- De Pedro, E.; Casillas, M. & Miranda, C.M. 1997. Microwave Oven Application in the Extraction of Fat from the Subcutaneous Tissue of Iberian Pig Ham. *Meat Sci.* **45**, 45-51.
- Díaz I, García Regueiro JA, Casillas M, De Pedro E. 1996. Triglyceride composition of fresh ham fat from Iberian pigs produced with different systems of animal nutrition. *Food Chem.* **55**, 383-387.
- Flores J, Biron C, Izquierdo L, Nieto P. 1988. Characterization of green hams from Iberian pigs by fast analysis of subcutaneous fat. *Meat Sci.* **23**, 253-262.
- Forina M., Armanino C, Leardi R, Drava G. 1991. A class modelling technique based on potential functions. *J. Chemometr.* **5**, 435-453.
- Gallardo E, Narvaéz-Rivas M, Pablos F, Jurado JM, León-Camacho M. 2012. Subcutaneous fat triacylglycerol profile from Iberian pigs as a tool to differentiate between intensive and extensive fattening systems. *J. Agric. Food Chem.* **60** 1645-1651.
- Gamero-Pasadas A, Viera-Alcaide I, Ríos JJ, Graciani E, Vicario IM, León-Camacho M. 2006. Characterization and quantification of the hydrocarbons fraction of the subcutaneous fresh fat of Iberian pig by off-line combination of high performance liquid chromatography and gas chromatography. *J. Chromatogr. A* **1123**, 82-91.
- Gonzalez AG, Pablos F, Martín MJ, León-Camacho M, Valdenebro MS. 2001. HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chem.* **73**, 93-101.
- Jolliffe, I. T. 2002. *Principal components analysis* (2nd ed.), Springer, New York.
- López-Bote CJ. 1998. Sustained utilization of the Iberian pig breed. *Meat Sci.* **49**, S17-S27.
- Massart, D. L. 1998. Handbook of Chemometrics and Qualimetrics, Part B, Elsevier, Amsterdam.
- Muth, J. 1999. Basic statistic and pharmaceutical statistical applications, Marcel Dekker, New York.
- Narvaéz-Rivas M, León-Camacho M, Vicario IM. 2009. Fatty acid and triacylglycerol composition of the subcutaneous fat from Iberian pigs fattened on the traditional feed: "Montanera". Effect of anatomical location and length of feeding. *Grasas Aceites* **60**, 238-247.
- Narvaéz-Rivas M, Pablos F, Jurado JM, León-Camacho M. 2011. Authentication of fattening diet of Iberian pig according to their volatile compounds profile from raw subcutaneous fat. *Anal. Bioanal. Chem.* **399**, 2115-2122.
- Narvaéz-Rivas M, Ríos JJ, Arteaga JF, Quiles JF, Barrero AF, León-Camacho M. 2008. Determination of ent-kaurene in subcutaneous fat of Iberian pigs by gas chromatography multi-stage mass spectrometry with the aim to differentiate between intensive and extensive fattening systems. *Anal. Chim. Acta* **624**, 107-112.
- Narvaéz-Rivas M, Vicario IM, Alcalde MJ, León-Camacho M. 2010. Volatile hydrocarbon profile of Iberian dry-cured hams. A possible tool for authentication of hams according to the fattening diet. *Talanta* **81**, 1224-1228.
- Ruiz J, Cava R, Antequera T, Martín L, Ventanas J, López-Bote CJ. 1998. Prediction of the feeding background of Iberian pigs using the fatty acid profile of subcutaneous, muscle and hepatic fat. *Meat Sci.* **49**, 155-163.
- Tetko IV, Livingstone DJ, Luik AI. 1995. Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.* **35**, 826-833.
- Viera-Alcaide I, Vicario IM, Graciani E, León-Camacho M. 2007. Authentication of fattening diet of Iberian pig according to their triacylglycerols profile from subcutaneous fat. *Anal. Chim. Acta* **596**, 319-324.
- Viera-Alcaide I, Vicario IM, Graciani E, León-Camacho M. 2008. A multivariate study of the triacylglycerols composition of the subcutaneous adipose tissue of Iberian pig in relation to the fattening diet and genotype. *Grasas Aceites* **59**, 327-336.
- Viera-Alcaide I, Narvaéz-Rivas M, Vicario IM, Graciani E, León-Camacho M. 2009. Different fattening systems

- of Iberian pig according to 1-alkene hydrocarbon content in the subcutaneous fat. *Grasas Aceites* **60**, 68-76.
- Zamora-Rojas E, Garrido-Varo A, De Pedro-Sanz E, Guerrero-Ginel JE, Pérez-Marín D. 2011. Monitoring NIRS calibrations for use in routine meat analysis as part of Iberian pig-breeding programs. *Food Chem.* **129**, 1889-1897.
- Zamora-Rojas E, Per ez-Marín D, De Pedro-Sanz E, Guerrero-Ginel JE, Garrido-Varo A. 2012. In-situ Iberian pig carcass classification using a micro-electro-mechanical system (MEMS)-based near infrared (NIR) spectrometer. *Meat Sci.* **90**, 636-642.
- Zupan, J, Gasteiger, J. 1993. Neural networks for chemists: An introduction. VCH, Weinheim.

Recibido: 1/10/12
Aceptado: 26/1/13