Available online at www.sciencedirect.com
SCIENCE DIRECT®
Food Chemistry 97 (2006) 438–446
Food Chemistry
www.elsevier.com/locate/foodchem

Polyphenolic compositions of Basque natural ciders: A chemometric study

Rosa M. Alonso-Salces [a], Carlos Herrero [b], Alejandro Barranco [a], Diana M. López-Márquez [a], Luis A. Berrueta [a,*], Blanca Gallo [a], Francisca Vicente [a]

[a] *Departamento de Química Analítica, Facultad de Ciencias, Universidad del País Vasco/Euskal Herriko Unibertsitatea, P.O. Box 644, E-48080 Bilbao, Spain*

[b] *Departamento de Química Analítica, Nutrición y Bromatología, Facultad de Ciencias, Universidad de Santiago de Compostela, Augas Férreas s/n, Campus Universitario, 27002 Lugo, Spain*

Received 13 December 2004; received in revised form 10 May 2005; accepted 10 May 2005

Abstract

Polyphenolic compositions of Basque natural ciders were determined by high-performance liquid chromatography, with diode array detection following thiolysis, in order to differentiate ciders according to the geographical origin of the main raw material used for their elaboration. Fifty percent of the apples used for cidermaking in the Basque Country are imported from France or Galicia (N.W. Spain); this gives beverages of different chemical compositions and sensory qualities. A data set, consisting of 64 cider samples and 33 measured variables, was evaluated using multivariate chemometric techniques. A preliminary study of data structure was performed by cluster analysis and principal component analysis. Different classification systems for the two categories were obtained on the basis of the chemical data by applying several supervised pattern recognition procedures, such as linear discriminant analysis (LDA), K-nearest neighbours (KNN), soft independent modelling of class analogy (SIMCA), and multilayer feed-forward artificial neural networks (MLF-ANN). KNN, SIMCA and the MLF neural network provided complementary results: KNN allowed the correct classification of almost all the ciders of the Galician category, SIMCA provided a model for the ciders of the French category that excluded all ciders made with Galician apples (50% of raw material), and the neural network achieved a level of hits for the classification of the ciders obtained from French apples (50% of raw material) above 95%. Polyphenolic profiles of the ciders provide enough information to develop classification rules for identifying ciders according to the geographical origin of the raw material used for cidermaking.

Keywords: Polyphenol; Cider; HPLC; Thiolysis; Chemometrics; Pattern recognition

1. Introduction

Polyphenols play an important role in cider quality, being related to several aspects. They are responsible for the colour and the balance of bitterness and astringency, which defines the "overall mouthfeel" of ciders (Lea, 1995). Moreover, they can be involved in the alcoholic and malolactic fermentations as metabolites, providing cider aroma, and as inhibitors of microorganism development (Salih, Le-Quéré, & Drilleau, 2000), controlling fermentation rates and cider spoilage (Cowan, 1999; Sponholtz, 1993). Phenolic compounds are also involved in the colloidal stability of cider, as well as in the inhibition of pre-fermentative clarification enzymes (Lea, 1990). In addition, polyphenols, as natural antioxidants constituents of human diet, are receiving increasing attention due to their health protective

* Corresponding author. Tel.: +34 94 601 5505; fax: +34 94 4648 500.
E-mail address: qapbesil@lg.ehu.es (L.A. Berrueta).

0308-8146/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2005.05.022

properties (Hertog, Feskins, Hollman, Katan, & Kromhout, 1993).

Polyphenolic composition of ciders mainly depends on the mixture of apple varieties and the cidermaking procedures used for their elaboration. In apples, a wide diversity of polyphenols is present: flavan-3-ols (catechins and procyanidins), hydroxycinnamic acids, dihydrochalcones, flavonols and anthocyanins (Amiot, Tacchini, Aubert, & Nicolas, 1992). Polyphenolic profiles of apples depend, not only on cultivars, but also on agronomic (Awad & De Jager, 2002) and climatologic factors (Lea, 1990), that can change between regions and/or seasons. Definitively, all these matters may influence polyphenolic composition of ciders. During manufacture of ciders, polyphenols are considerably modified, defining the sensory properties of ciders. Hence, knowledge of the precise composition of ciders may contribute to a better understanding of their implication in cider quality and diversity.

Traditional technology used in the Basque Country for cidermaking consists of the following steps: milling of cider apples; maceration of the pulp in the press for 24 h; pressing for 2 or 3 days; spontaneous must clarification and a natural fermentation to dryness. The stages of maceration and pressing can lead to certain problems, as uncontrolled proliferations of undesirable microorganisms (acetic and lactic bacteria, weakly fermentative and oxidative yeasts), as well as an important oxidation of the phenolic compounds of the milled apples take place during these steps of fruit processing. These phenomena can influence the development of fermentation because of the competition for the must nutrients between bacteria and non-fermentative yeast with fermentative yeast (Cabranes, Moreno, & Mangas, 1991). This may result in the generation of high levels of acetic acid, responsible for the typical high volatile acidity of Basque ciders (Irastorza, 1988). Traditional Basque practices for natural cidermaking are still maintained, but the ancient vertical screw presses of great mass and the wooden vats are slowly being replaced by more modern technologies. Thus, pneumatic presses and stainless steel vats are being introduced. With these kinds of presses, pressing times are shortened (2–3 h) and working conditions are more aseptic, allowing a better control of chemical and microbiological contaminations. In this sense, the new kinds of vats contribute to hygiene because they are easier to clean and disinfect. Moreover, in some cases, stainless steel vats are equipped with devices for controlling temperature and establishing inert atmospheres. These improvements allow a better control of the fermentative process, preventing the development of undesirable microflora. Furthermore, racking of ciders from yeast lees, once fermentation is complete, is becoming a common practice, thus guaranteeing a proper physicochemical and microbial stability of the ciders.

The production of cider apples in the Basque Country is not sufficient for supplying cidermakers the quantities of apples required. Therefore, significant amounts (about 50% of the total of the apples used for cidermaking) are brought from France (Brittany and Normandy) or Galicia (N.W. Spain). Basque ciders are made with 32% of Basque apples, and the rest of the apples used (18%) are from other European regions (e.g., Asturias (Spain), Czech Republic). Regarding the great influences that the raw material and polyphenols have on cider quality, it is interesting to distinguish ciders elaborated with apples cultivated in different regions, in order to select the most interesting apples for producing cider with the organoleptic properties and quality desired.

In this work, polyphenolic profiles of Basque ciders were analysed by chemometric techniques with the aim of differentiating ciders according to the geographical origin of the apples used for their elaboration: (France and Galicia). In this sense, classification rules were developed in order to predict the origin of the apples mainly used for making a Basque cider.

2. Materials and methods

2.1. Reagents and standards

Methanol (Romil Chemical Ltd., Heidelberg, Germany) was of HPLC grade. Water was purified on a Milli-Q system from Millipore (Bedford, MA, USA). Glacial acetic acid, toluene-α-thiol and fuming hydrochloric acid (37%), provided by Merck (Darmstadt, Germany), and ascorbic acid, provided by Panreac (Barcelona, Spain), were of analytical quality. All solvents used were previously filtered through 0.45 μm nylon membranes (Lida, Kenosha, WI, USA).

Polyphenol standards were supplied as follows: (+)-catechin, (−)-epicatechin, rutin, phloridzin, 5-caffeoylquinic acid, *p*-coumaric acid and caffeic acid by Sigma-Aldrich Chemie (Steinheim, Germany); hyperoside, isoquercitrin, avicularin and quercitrin by Extrasynthèse (Genay, France). (−)-Epicatechin 4R-benzylthioether and 4-p-coumaroylquinic acid were kindly provided by Dr. Guyot, and phloretin-2′-*O*-xyloglucoside and procyanidin B2 by Dr. F.A. Tomás-Barberán and Dr. C. Santos-Buelga, respectively. Stock standard solutions of (+)-catechin, (−)-epicatechin, (−)-epicatechin 4R-benzylthioether, rutin, phloridzin, 5-caffeoylquinic acid, *p*-coumaric acid and caffeic acid, at a concentration of 1 mg/ml and hyperoside, isoquercitrin, quercitrin and ideain at 0.6 mg/ml, were prepared in methanol and stored at 4 °C in darkness. The other standards were prepared in approximate concentrations and only used for chromatographic peak identification.

2.2. Samples

Sixty-four samples of natural ciders made in the Basque Country in the 2000 and 2001 seasons were analyzed. Thirty-two of the ciders had been made with 50% of French apples, and thirty-two ciders, with 50% of Galician apples. The other 50% of the apples used for cidermaking, were from the Basque Country (32%), and from other European regions [e.g., Asturias (Spain) (10%), Czech Republic (5%)], in both kind of ciders. Ciders were commercially available or supplied by cidermakers from Guipúzcoa and Vizcaya, and by the Diputación Foral de Bizkaia.

Ciders were degassed for 10 min in an ultrasonic bath and aliquots of 1 ml were sampled, freeze-dried and stored at room temperature in a desiccator prior to analysis.

2.3. Analytical procedures

2.3.1. Thiolysis and direct solvent extraction

Two freeze-dried aliquots (1 ml) were analysed for each cider sample: one aliquot was submitted to thiolysis as described by Guyot, Marnet, Sanoner, and Drilleau (2001), and the other, to direct solvent extraction with 2 ml of methanol–water–acetic acid (30:69:1, v/v/v) with ascorbic acid (2 g/l) in an ultrasonic bath during 15 min. Then, both the thiolysis reaction mixture and the crude solvent extract were filtered through a 0.45 μm PTFE filter (Waters, Milford, CA, USA) prior to injection into the HPLC system.

2.3.2. Reversed-phase HPLC analysis

Chromatographic analyses were performed in a Hewlett–Packard Series 1100 system, equipped with a vacuum degasser, a quaternary pump, a thermostatted autosampler, a thermostatted column compartment and a DAD detector, connected to an HP ChemStation software. A reversed-phase Nova-Pak C18 (300 × 3.9 mm I.D., 4 μm) column and a Nova-Pak C18 (10 × 3.9 mm I.D., 4 μm) guard column (Waters, Barcelona, Spain) were used. Solvents that constituted the mobile phase were A (acetic acid–water, 10:90, v/v) and B (methanol). The elution conditions applied were: 0–10 min, 0% B isocratic; 10–40 min, linear gradient 0–15% B; 40–60 min, 15% B isocratic; and finally, washing and reconditioning the column. The flow rate was 0.8 ml min^{-1} and the injection volume was 50 μl for the crude extracts or 10 μl for the thiolysis media. The chromatographic separation was carried out at 25 °C, maintaining the vials in the autosampler at 4 °C. Flavan-3-ols and dihydrochalcones were monitored and quantified at 280 nm, hydroxycinnamic acids at 320 nm and flavonols at 370 nm.

The thiolysis reaction prior to HPLC analyses allows the assay of procyanidins. However, direct HPLC analyses of the crude solvent extracts without thiolysis was also necessary to assay monomeric catechins. Identification of polyphenols was possible by comparison of the retention times and UV–vis spectra with those of the standards that were available. Some other chromatographic peaks were assigned to a particular polyphenol class according to the UV–vis spectra and bibliographic sources. In this sense, those unknown chromatographic peaks that exhibit flavan-3-ol spectra were nominated as CAT-*n*, and those with a spectrum of 5-caffeoylquinic acid as CAA-*n*, of *p*-coumaric as CMA-*n*, of dihydrochalcone as PLD-*n* and of flavonol as QG-*n* (where "*n*" is a number). Quantification was performed by reporting the measured integration areas in the calibration equation of the corresponding standards. Thus, procyanidin B2 and the unknown flavan-3-ols were quantified as (+)-catechin; phloretin, phloretin-2′-*O*-xyloglucoside and the unknown dihydrochalcones as phloridzin, aviculin, quercetin and the unknown flavonols as rutin, CAA-*n* species as 5-caffeoylquinic acid, 4-*p*-coumaroylquinic acid and CMA-*n* species as *p*-coumaric acid.

2.4. Data analysis and chemometric procedures

The data set consisted of a 64 × 33 matrix, in which rows represented the Basque ciders analysed (64 objects), and columns the concentrations of 31 individual polyphenols determined by HPLC-DAD, the total concentrations of procyanidins and the average degrees of polymerisation of procyanidins (DPn) (33variables). Each cider sample was represented in the 33-dimensional space by a data vector which is an assembly of the 33 features. Data vectors belonging to the same category (apple origin: France or Galicia) were first analysed by univariate procedures (ANOVA, Fisher index and Box-Whisker plots), and afterwards, by the following chemometric techniques, that have been described in the literature (Lattore, Peña, García, & Herrero, 2000; Padín et al., 2001): cluster analysis (CA), linear discriminant analysis (LDA), K-nearest neighbours (KNN), soft independent modelling of class analogy (SIMCA), and multilayer feed-forward artificial neural networks (MLF-ANN). Statistical and chemometric data analyses were performed by means of the statistical software packages Statgraphics (1994–2000), Parvus (Forina, Lanteri, & Armanino, 2000), SPSS (1989–1999) and WinNN32 (1993–1996).

Cluster analysis is a pattern recognition technique that is used to reveal the structure residing in a data set, and to disclose the natural groupings existing between samples characterised by the values of a set of measured variables. It is commonly applied before other multivariate techniques owing to its unsupervised character. CA was performed on the autoscaled data. Sample similarities were calculated on the basis of the squared Euclidean distance, and the Ward hierarchical

agglomerative method was used to establish clusters (Massart & Kaufman, 1983).

Principal component analysis, performed on the autoscaled data matrix, allows reduction of the number of variables retaining the maximum amount of variability present in data, in order to provide a partial visualisation of data structure in a reduced dimension.

The supervised pattern recognition techniques (LDA, KNN, SIMCA, and a MLF neural network) were used in order to attain classification rules for predicting the (French or Galician) origin of the apples mainly used for the elaboration of Basque ciders. These techniques were applied to the autoscaled data matrix of the polyphenolic profiles of ciders. The classification rules achieved were validated by means of a cross-validation procedure, which was performed by dividing the complete data set into a training set and an evaluation set. Samples were assigned randomly to a training set, consisting of 75% of them, and a test set, composed of the remaining 25% of the samples. Such a division allows for a sufficient number of samples in the training set and a representative number of members among the test set. The same process was repeated four times with different constitutions of both sets to ensure that all the samples had the possibility of being included in the evaluation set at least once. The different pattern recognition techniques were applied to the four training-test sets obtained. The reliability of the classification models achieved was studied in terms of recognition ability (percentage of the members of the training set correctly classified) and prediction ability (percentage of the members of the test set correctly classified by using the rules developed in the training step).

When using KNN, the inverse square of the Euclidean distance was used as the criterion for calculating the distance between samples, and the number of neighbours (K) was selected after studying the success in classification with different K values, applying this technique to a training set with all the samples.

The SIMCA version applied was the normal range model. The stop criterion used by SIMCA for the calculation of the principal components was the minimum percentage of retained variance (95%, $\alpha = 0.05$). The model achieved by SIMCA for each category was also evaluated in terms of sensitivity and specificity. The sensitivity of the model is known as the percentage of objects belonging to the category correctly identified by the mathematical model, and its specificity, as the percentage of objects foreign to the category classified as foreign (Meléndez, Ortíz, Sánchez, Sarabia, & Iñiguez, 1999).

WinNN32 MLF-ANN was applied on an input pattern consisting of the autoscaled data matrices of each apple material. The target output was assigned as 0 or (0,1) for unripe fruits and 1 or (1,0) for ripe ones, and a sigmoidal function $f(x) = 1/(1 + [\exp(-x)])$ was used as the transfer function. The neural network was trained by means of an algorithm that combined the use of an adaptative learning rate parameter (η) and a momentum (μ) which have been described previously (Padín et al., 2001). The initial values of the weights associated with the connections between neurons were selected randomly in the range -3 to 3. The maximum number of epochs was 2000 and the initial values of η and μ were 0.2 and 0.5, respectively; the target error was 0.1.

3. Results and discussion

3.1. Polyphenolic profiles of Basque ciders

Cider analysis results are summarised in Table 1. Total polyphenol contents of Basque ciders, determined by HPLC-DAD, differ notably between both kinds of ciders: those made mainly with French apples and those with Galician apples. Indeed, they are higher, even 10-fold, in ciders obtained from Galician apples. Polyphenol concentrations ranged from 0.28 to 217 mg/l in the ciders from French apples, and from 21 to 512 mg/l in ciders from Galician apples. The ciders made with 50% of Galician apples presented higher contents of most polyphenols (EC, PC, CAF, PCM, CAA-6, CAA-8, CMA-2, PL, PLXG, PLG, HYP, IQC and QCI) or similar concentrations (PB2, CAT-1, PCQ, CAA-1, CAA-5, CMA-1 and QCE) to the ciders made with 50% of French apples. Some phenolic compounds were not detected in ciders from French apples: (+)-catechin (CAT), CAT-3 and CAT-4, 5-caffeoylquinic acid (CQA), PLD-1 and QG-1. However, in these ciders, other species with the spectrum of a flavan-3-ol, named CAT-2, and aviculin (AVI), were detected, whereas they were not in ciders from Galician apples.

Taking into account that ciders were made using the same technologies and traditional procedures, the differences observed in the polyphenolic contents were probably due to the raw material used: the apples. The authors concluded in a previous work (Alonso-Salces et al., 2004) that polyphenolic compositions of Basque cider apples harvested in the 2000 and the 2001 seasons did not present significant differences. In addition, ciders were made using the same percentage of Basque cider apples (32%). Hence, the difference in the phenolic composition between the two kinds of ciders has to be due, mainly, to the 50% of apples that were imported from France or Galicia.

In a first approach, it is interesting to use apples with high phenolic contents for the elaboration of cider following the Basque tradition, in order to compensate for the great polyphenol loss that takes place by enzymatic and chemical oxidation (Lea & Timberlake, 1978).

Table 1
Polyphenol concentrations (mg/l) in Basque ciders[a]

	French category ($n = 32$)				Galician category ($n = 32$)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Flavan-3-ols	51	31	nd	116	71	44	15	189
CAT	nd				0.8	1.3	nd	7.0
EC	0.2	0.7	nd	3.6	2	4	nd	15
PB2	3	5	nd	14	3	2	nd	12
PC	50	31	nd	116	68	41	15	185
CAT-1	57	27	nd	128	56	32	nd	137
CAT-2	12	12	nd	38	nd			
CAT-3	nd				10	8	nd	35
CAT-4	nd				4	3	nd	10
DPn[b]	2.2	0.5	1.0	2.7	2.3	0.4	1.2	2.9
Hydroxycinnamic acids	16	8	1	28	32	29	8	147
CQA	nd				3	8	nd	35
CAF	0.1	0.3	nd	1.7	9	24	nd	124
PCM	0.2	0.4	nd	1.3	0.6	1.0	nd	4.3
PCQ	8	5	nd	16	9	5	2	21
CAA-1	5	2	nd	9	7	2	3	12
CAA-5	0.3	0.4	nd	1.5	0.4	0.7	nd	2.8
CAA-6	0.3	0.4	nd	1.3	0.8	0.4	nd	1.9
CAA-7	0.9	0.4	nd	1.6	0.4	0.4	nd	2.0
CAA-8	0.1	0.2	nd	0.6	0.8	0.4	0.3	1.6
CAA-9	0.7	0.3	nd	1.1	0.07	0.21	nd	0.92
CMA-1	0.5	1.1	nd	5.8	0.4	0.3	nd	1.2
CMA-2	0.4	0.3	nd	1.3	0.6	0.3	nd	1.6
Dihydrochalcones	11	8	nd	37	19	18	1	67
PLD-1	nd				0.6	0.8	nd	2.7
PLD-2	2	1	nd	6	1.0	0.9	nd	3.3
PLXG	3	2	nd	10	7	7	1	30
PLG	5	6	nd	27	10	11	nd	43
PL	0.3	0.7	nd	3.0	1	2	nd	7
Flavonols	3	2	0.3	8	4	2	0.7	10
HYP	0.3	0.6	nd	2.1	0.5	0.6	nd	2.7
IQC	0.02	0.04	nd	0.18	0.06	0.13	nd	0.51
QG-1	nd				0.2	0.2	nd	0.9
QG-4	0.2	0.2	nd	0.7	0.10	0.09	nd	0.32
AVI	0.1	0.1	nd	0.6	nd			
QCI	0.6	1.0	nd	3.1	1	1	nd	4
QCE	1.6	0.8	0.3	3.3	1.5	0.7	nd	2.8

[a] Abbreviations: AVI, avicularin; CAA-1, -5, -6, -7, -8, -9 unknown hydroxycinnamic acids with UV spectrum of CQA; CAF, caffeic acid; CAT, (+)-catechin; CAT-1, -2, -3, -4, unknown flavan-3-ols; CMA-1, -2, unknown hydroxycinnamic acids with UV spectrum of PCM; CQA, 5-caffeoylquinic acid; EC, (−)-epicatechin; HYP, hyperoside; IQC, isoquercitrin; PB2, procyanidin B2; PC, total procyanidins; PCM, *p*-coumaric acid; PCQ, 4-*p*-coumaroylquinic acid; PL, phloretin; PLD-1, -2, unknown dihydrochalcones; PLG, phloridzin; PLXG, phloretin-2′-*O*-xyloglucoside; QCE, quercetin; QCI, quercitrin; QG-1, -4, unknown flavonols; SD, standard deviation; max, maximum; min, minimum; nd, not detected; *n*, number of samples.
[b] DPn, average degree of polymerisation of procyanidins.

3.2. Univariate data analysis

The analysis of variance (ANOVA) performed on the data matrix consisting of the individual polyphenol concentrations, total procyanidin content and DPn disclosed significant differences for all variables between ciders made with 50% of French apples (French category) and those made with 50% of Galician apples (Galician category). The Fisher index was calculated to establish the discriminant capacity of the variables one by one (Sharaf, Illman, & Kowalski, 1986). The most discriminant variables were those that presented the highest Fisher weights ($p < 0.001$): some hydroxycinnamic acids (CAA-8, CAA-9, CAA-7 and CAA-6) and some flavan-3-ols (CAT-4, CAT-3, CAT-2), but their Box-Whisker plots showed an overlap between the concentration ranges of the two classes. Thus, none of the variables measured was able to discriminate between French and Galician categories by itself. Therefore, it was necessary to move on to multivariate data analysis in order to differentiate ciders according to the geographical origin (France and Galicia) of the apples mainly used for cidermaking.

3.3. Multivariate data analysis

3.3.1. Cluster analysis

The results achieved by CA are presented as a dendrogram in Fig. 1. Two clusters are observed: cluster A contains only ciders made with French apples (50% of raw material), and cluster B is made up with all the ciders obtained from Galician apples (50% of raw material) and a cider sample from French apples. These results disclosed that there were notable differences among ciders of both categories and that the polyphenolic data may contain adequate information to attain cider differentiation according to the established classes.

3.3.2. Principal component analysis

In the bidimensional plot of the sample scores in the space defined by the two first principal components, a natural separation of the ciders made with 50% of French apples (French category) and those made with 50% of Galician apples (Galician category), is observed (Fig. 2). The two first principal components accounted for 39% of total system variability. From the loadings of the variables (Table 2), the most influential features on the first principal component (PC1, accounting for 25% of total variability) were polyphenols that were found in higher concentrations in ciders of the Galician category: the procyanidins (PC), CAA-1, CAA-8, 4-*p*-coumaroylquinic acid (PCQ), phloretin-2′-*O*-xyloglucoside (PLXG), PLD-1, isoquercitrin (IQC) and QG-1. Major contributions to PC2, which accounted for 14% of total variability, were due to phenolic compounds that were present in higher contents in ciders of the French category: CAT-2, PLD-2, CAA-9 and avicularin (AVI), or were negatively related to PC2, as CAA-8.

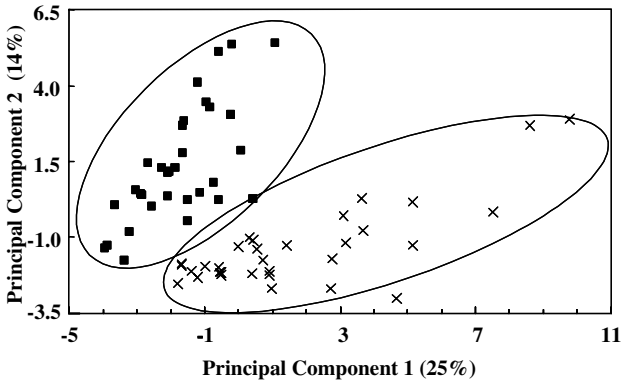

Fig. 2. Eigenvector projection of cider samples. Sample codes: ■, French category: ciders made with 50% of French apples; ×, Galician category: ciders made with 50% of Galician apples. The amount of variability explained by each principal component is shown on the axis.

PCA results agree with those obtained by CA, since the samples of the two categories are grouped in well-defined areas in the bidimensional plot of the two first principal components (Fig. 2). These observations can be explained by the fact that polyphenolic profiles of apples depend on agronomic (nitrogen fertilisation, pruning, chemical agents) and climatologic factors, which can change among regions and/or seasons (Awad & De Jager, 2002).

3.3.3. Supervised pattern recognition methods

Table 4 shows recognition and prediction abilities afforded with each multivariate technique. LDA achieved recognition percentages of 100% for both categories; however, prediction abilities were less satisfactory, especially for ciders made with Galician apples

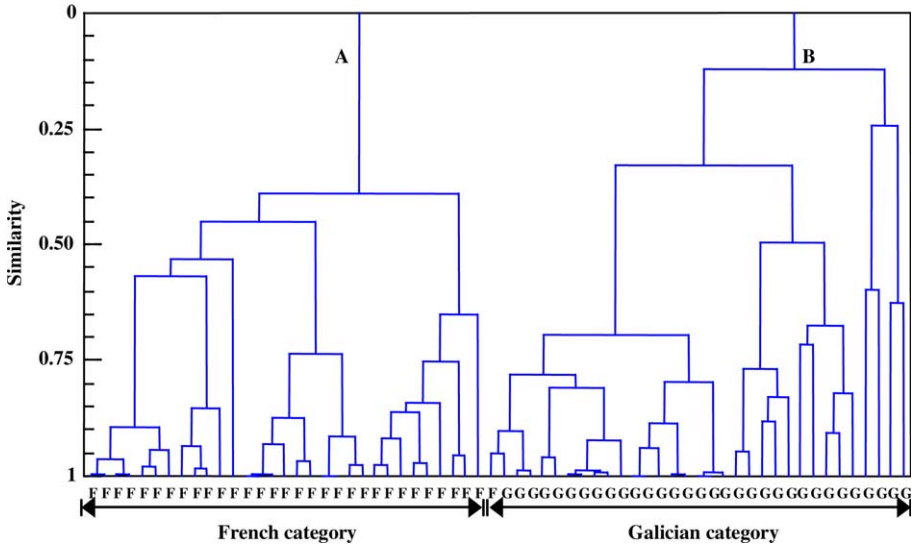

Fig. 1. Dendrogram of cluster analysis. Sample codes: F, French category: ciders made with 50% of French apples; G, Galician category: ciders made with 50% of Galician apples.

Table 2
Loadings of the features in the first three principal components for the ciders made with 50% of French apples (French category) or 50% of Galician apples (Galician category)[a]

Variables	PC1	PC2
CAT	0.170	−0.112
EC	0.185	−0.056
PB2	0.093	0.074
CAT-1	0.113	0.157
CAT-2	−0.090	0.357
CAT-3	0.141	−0.190
CAT-4	0.154	−0.197
PC	**0.253**	0.153
CQA	0.179	0.036
CAF	0.106	−0.113
PCM	0.163	−0.075
PCQ	**0.239**	0.200
CAA-1	**0.256**	0.082
CAA-5	−0.029	−0.113
CAA-6	0.150	−0.222
CAA-7	−0.083	0.205
CAA-8	**0.226**	**−0.258**
CAA-9	−0.165	**0.303**
CMA-1	0.058	0.103
CMA-2	0.150	0.013
PLD-1	**0.241**	−0.098
PLD-2	0.089	**0.341**
PLG	**0.270**	0.147
PLXG	**0.247**	0.007
PL	0.130	−0.067
HYP	0.198	0.236
IQC	**0.260**	0.153
QG-1	**0.293**	0.005
QG-4	−0.001	0.151
AVI	−0.024	**0.278**
QCI	0.175	−0.009
QCE	0.056	0.203
DPn	0.167	0.128

[a] Abbreviations: See Table 1; PC1, first principal component; PC2, second principal component.

(50% of raw material), for which only 83.8% of the predictions were correct. Prediction ability of LDA for ciders obtained from French apples (50% of raw material) was 90.9%.

In KNN, the numbers of neighbours (K) assayed, in the preliminary study using a training set with all the samples, were 3, 4, 5, 6, 7, 8, 9 and 11. For K values of 3 and 5, two samples were misclassified by the model whereas, for the rest of the K-values, only one sample was wrongly classified, therefore $K = 4$ was selected. KNN afforded excellent results for both categories: recognition abilities of 96.9% (French category) and 99.9% (Galician category), and prediction abilities of 100% for both classes. These results were consistent with sample distribution and the multidimensional space visualised by PCA (Fig. 2) and CA (Fig. 1), where a sample of the French class was included in the group that contained the samples of the Galician class.

Table 3
MLF-ANN architectures assayed and their prediction abilities for the ciders made with 50% of French apples (French category) or 50% of Galician apples (Galician category)

MLF ANN architecture	Prediction ability (%)	RMSE[a]
33-3-1	89.8	0.084
33-5-1	92.2	0.073
33-3-2	**94.5**	**0.088**
33-4-2	90.6	0.080

[a] RMSE, root mean square error.

SIMCA attained better results for the Galician category, with 100% of correct classification. SIMCA recognition and prediction abilities for ciders made with French apples (50% of raw material) were 95.8% and 90.6%, respectively. The Galician model presented a sensitivity of 76% and a specificity of 94%, which meant that it accepted 76% of the samples of Galician class and 6% of the samples of the French class. However, the French model was really selective, since it recognised 75% of the ciders made with French apples (50% of raw material) and rejected 100% of those made with Galician apples (50% of raw material), these percentages being the sensitivity and specificity of the model, respectively. Hence, any cider made with 50% of Galician apples was classified as belonging to the French category, whereas there existed a small probability (0.06) that ciders made with 50% of French apples were misclassified by this technique as ciders of the Galician class. Fig. 3 represents SIMCA results as a Coomans plot for the squared SIMCA distances obtained from the complete data set.

Table 4
Classification results for the supervised pattern recognition techniques applied to cider samples[a]

Technique	Category[b]	Recognition ability (%)	Prediction ability (%)
LDA	French	100.0	90.9
	Galician	100.0	83.8
KNN ($K = 4$); inverse squared Euclidean distance	French	96.9	100.0
	Galician	99.0	100.0
SIMCA; normal range; $\alpha = 0.05$	French	95.8	90.6
	Galician	100.0	100.0
MLF-ANN ($33 \times 3 \times 2$); $\eta = 0.2$; $\mu = 0.5$; sigmoidal transfer function	French	100.0	95.4
	Galician	100.0	93.7

[a] Validation was performed by a cross-validation in four steps (75% of samples as training set, 25% of samples as test set).
[b] French category, ciders made with 50% of French apples; Galician category, ciders made with 50% of Galician apples.

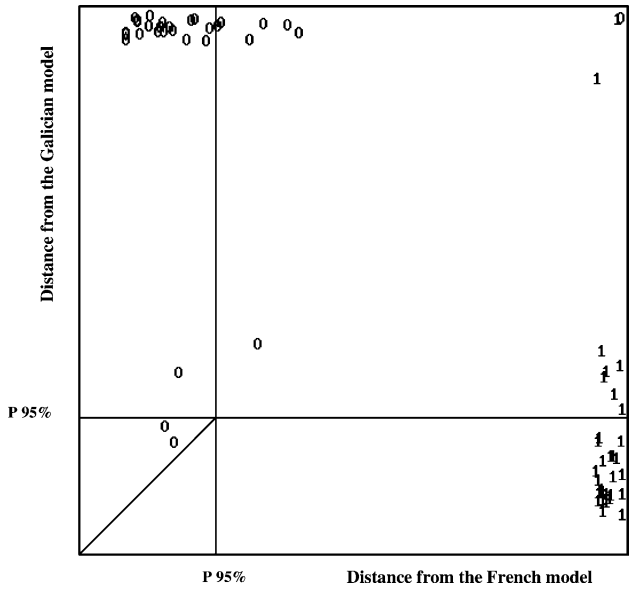

Fig. 3. Coomans plot for the squared SIMCA distances. Codes: 0, French category: ciders made with 50% of French apples; 1, Galician category: ciders made with 50% of Galician apples.

Regarding the neural network, some empirical preliminary trials were performed to determine an adequate MLF-ANN structure. The neural architecture which gave better results was a MLF-ANN with three layers: an input layer with 33 neurons, one hidden layer with 3 neurons, and an output layer consisting of two neurons with binary output (Table 3). The MLF neural network $(33 \times 3 \times 2)$ achieved satisfactory prediction abilities: 95.4% and 93.7% for the French and Galician categories, respectively.

Thus, the pattern recognition techniques: KNN, SIMCA and the MLF neural network, led to complementary results. KNN allowed the detection of almost all the ciders of the Galician category, SIMCA provided a model for the ciders of the French category that excluded all the ciders made with Galician apples (50% of raw material), and the neural network achieved a level of hits for the classification of ciders obtained from French apples (50% of raw material) above 95%.

4. Conclusion

This study revealed the great influence that apples have on the characteristics of natural ciders. In this particular case, the geographical origin of the apples used for cidermaking was evaluated: France or Galicia. In this sense, several aspects, such as apple varieties, agronomic procedures, climatology, were being considered. These aspects are different, to a larger or lesser extent, in each geographical region, and can be responsible for the differences found in the final products obtained. From the results obtained in this work, it is concluded that the polyphenolic profiles of ciders provide adequate and sufficient information to allow classification rules to be developed by pattern recognition techniques for identifying ciders according to the geographical origin of the apples used for their elaboration. Thus, an authentication control of the origin of the raw material used for cidermaking can be achieved. In this sense, ciders made with at least 50% of apples imported from France and ciders made with at least 50% of apples from Galicia were distinguished, using complementary pattern recognition techniques.

Abbreviations used

Polyphenols: AVI, avicularin; CAA-1, -5, -6, -7, -8, -9, unknown hydroxycinnamic acids with caffeic acid UV spectra; CAF, caffeic acid; CAT, (+)-catechin; CAT-1, -2, -3, -4, unknown flavan-3-ol; CMA-1, -2, unknown hydroxycinnamic acid with *p*-coumaric acid UV spectra; CQA, 5-caffeoylquinic acid; DPn, average degree of polymerisation of procyanidins; EC, (−)-epicatechin; HYP, hyperoside; IQC, isoquercitrin; PB2, procyanidin B2; PC, total procyanidins; PCM, *p*-coumaric acid; PCQ, 4-*p*-coumaroylquinic acid; PL, phloretin; PLD-1, -2, unknown dihydrochalcones; PLG, phloridzin; PLXG, phloretin-2′-*O*-xyloglucoside; QCE, quercetin; QCI, quercitrin; QG-1, -4, unknown flavonols.

Others: CA, cluster analysis; DAD, diode array detector; HPLC, High Performance Liquid Chromatography; KNN, *K*-nearest neighbours; LDA, linear discriminant analysis; MLF-ANN, multilayer feed-forward-artificial neural network; nd, not detected; PC1, first principal component; PC2, second principal component; PCA, principal component analysis; RMSE, root medium square error; SD, standard deviation; SIMCA, soft independent modelling of class analogy.

Acknowledgements

This research was supported by Gobierno Vasco/Eusko Jaurlaritza (projects numbers PI-1997-19 and PI-1999-106) and Universidad del País Vasco/Euskal Herriko Unibertsitatea (project number 171.310-EB013/98). The authors express their gratitude to Diputación Foral de Bizkaia and to the Basque cidermakers for providing cider samples, and to Dr. S. Guyot (INRA, France), Dr. F.A. Tomás-Barberán (CEBAS-CSIC, Spain) and Dr. C. Santos-Buelga (Universidad de Salamanca, Spain) for providing polyphenol standards. Rosa M^a Alonso-Salces thanks Gobierno Vasco/Eusko Jaurlaritza for a Ph.D. grant.

References

Alonso-Salces, R. M., Barranco, A., Abad, B., Berrueta, L. A., Gallo, B., & Vicente, F. (2004). Polyphenolic profiles of Basque cider apple cultivars and their technological properties. *Journal of Agricultural Food Chemistry, 52*, 2938–2952.

Amiot, M. J., Tacchini, M., Aubert, S., & Nicolas, J. (1992). Phenolic composition and browning susceptibility of various apple cultivars at maturity. *Journal of Food Science, 57*, 958–962.

Awad, M. A., & De Jager, A. (2002). Relationship between fruit nutrients and concentrations of flavonoids and chlorogenic acid in "Elstar" apple skin. *Science Horticulturae, 92*, 265–276.

Cabranes, C., Moreno, J., & Mangas, J. J. (1991). Aspectos microbiológicos y bioquímicos de la elaboración de Sidra Natural en lagares asturianos. *Alimentaria, 4*, 63–69.

Cowan, M. M. (1999). Plant products as anti-microbial agents. *Clinical Microbiology Reviews, 12*, 564–582.

Forina, M., Lanteri, S., & Armanino, C. (2000). *Q-Parvus 3.0. An extendable package of programs for data explorative analysis, classification and regression analysis*. Genova (Italy): Dipartimento Chimica e Tecnologie Farmaceutiche ed Alimentari, University of Genova.

Guyot, S., Marnet, N., Sanoner, P., & Drilleau, J. F. (2001). Direct thiolysis on crude apple materials for the HPLC characterization and quantification of polyphenols in cider apple tissues and juices. *Methods in Enzymology, 335*, 57–70.

Hertog, M. G. L., Feskins, E. J. M., Hollman, P. C. H., Katan, M. B., & Kromhout, D. (1993). Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen elderly study. *Lancet, 342*, 1007–1011.

Irastorza, A. (1988). Estudio microbiológico y analítico de las sidras guipuzcoanas. Tésis Doctoral de la Universidad del País Vasco, San Sebastián, Spain.

Lattore, M. J., Peña, R., García, S., & Herrero, C. (2000). Authentification of Galician (N.W. Spain) honeys by multivariate techniques based on metal content data. *The Analyst, 125*, 307–312.

Lea, A. G. H. (1990). Bitterness and astringency: the procyanidins of fermented apple ciders. In R. L. Rousself (Ed.), *Bitterness in food and beverages* (pp. 123–143). Oxford: Elsevier.

Lea, A. G. H. (1995). Cidermaking. In A. G. H. Lea & J. R. Piggott (Eds.), *Fermented beverage production* (pp. 66–96). London: Blackie Academic & Professional.

Lea, A. G. H., & Timberlake, C. F. (1978). The phenolics of ciders: effect of processing conditions. *Journal of the Science of Food and Agriculture, 29*, 484–492.

Massart, L., & Kaufman, L. (1983). *The interpretation of analytical chemical data by the use of clusters analysis*. New York: Wiley Interscience.

Meléndez, M., Ortíz, M. C., Sánchez, M., Sarabia, L., & Iñiguez, M. (1999). Chemometric characterization of the clarets and rose wines of the certified denomination of origin Rioja using Cielab parameters. *Química Analítica, 18*, 119–126.

Padín, P. M., Peña, R. M., García, S., Iglesias, R., Barro, S., & Herrero, C. (2001). Characterization of Galician (N.W. Spain) quality brand potatoes: a comparison study of several pattern recognition techniques. *Analyst, 126*, 97–103.

Salih, A. G., Le-Quéré, J. M., & Drilleau, J. F. (2000). Action des acides hydroxycinnamiques libres et estérifiés sur la croissance des bactéries lactiques. *Sciences des Aliments, 20*, 537–560.

Sharaf, M., Illman, D., & Kowalski, B. R. (1986). *Chemometrics*. New York: Wiley.

Sponholtz, W. R. (1993). Wine spoilage by micro-organisms. In G. H. Fleet (Ed.), *Wine microbiology and biotechnology* (pp. 397–399). USA: Harwood Academic Publishers.

SPSS. (1989–1999). SPSS for Windows version 9.0, SPSS Inc., ©1989–1999.

Statgraphics Plus. (1994–2000). Statgraphics Plus 5.0, Statistical Graphics Corporation, ©1994–2000.

WinNN32. (1993–1996). WinNN32 1.2, Y. Danon, ©1993–1996.