Talanta 81 (2010) 255–264

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Analysis of protein chromatographic profiles joint to partial least squares to detect adulterations in milk mixtures and cheeses

N. Rodríguez [a], M.C. Ortiz [a,*], L. Sarabia [b], E. Gredilla [c]

[a] Department of Chemistry, Faculty of Sciences, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain
[b] Department of Mathematics and Computation, Faculty of Sciences, University of Burgos, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain
[c] Subdirección de Calidad y Seguridad Alimentaria, Estación Tecnológica de la Leche, Instituto Tecnológico Agrario de Castilla y León, Consejería de Agricultura y Ganadería, Ctra. Autilla s/n, 34071 Palencia, Spain

ARTICLE INFO

Article history:
Received 16 July 2009
Received in revised form
18 November 2009
Accepted 29 November 2009
Available online 5 December 2009

Keywords:
Dairy products
Authenticity
HPLC-DAD
Partial least squares (PLS)
Principal components analysis (PCA)
False non-compliance or false compliance risks

ABSTRACT

To prevent possible frauds and give more protection to companies and consumers it is necessary to control that the types of milk used in the elaboration of dairy products correspond to those appearing in their label. Therefore, it is greatly interesting to have efficient, quick and cheap methods of analysis to identify them. In the present work, the multivariate data are the protein chromatographic profiles of cheese and milk extracts, obtained by high-performance liquid chromatography with diode-array detection (HPLC-DAD). These data correspond to pure samples of bovine, ovine and caprine milk, and also to binary and ternary mixtures. The structure of the data is studied through principal component analysis (PCA), whereas the percentage of each kind of milk has been determined by a partial least squares (PLS) calibration model. In cheese elaborated with mixtures of milk, the procedure employed allows one to detect 3.92, 2.81 and 1.47% of ovine, caprine and bovine milk, respectively, when the probability of false non-compliance is fixed at 0.05. These percentages reach 7.72, 5.52 and 2.89%, respectively, when both the probability of false non-compliance and false compliance are fixed at 0.05.

1. Introduction

Dairy products have been traded for many years and they represent a large proportion of the food industry. For legal, medical and ethical reasons, cheeses should be correctly labeled, and the substitution or omission of valuable compounds in milk products for less costly ingredients, the addition of ingredients to make products appear to be better, and the false or misleading labeling of dairy products is considered fraudulent. In many European countries, it is mandatory to state the type of milk used for manufacturing cheese or other dairy products [1] especially in the case of protected denomination of origin (PDO) cheeses [2]. However, differences in price and seasonal availability might make it attractive for farmers to adulterate expensive ewe and goat milk with cheaper cow milk. Protection against such frauds is of importance to warrant fairness in food trade, and also to protect consumers. Thus, there is an ultimate need of rapid, efficient, sensitive, and reliable control methods that can determine the composition of milk in cheese and other dairy products. The analytical techniques used for this task can be divided in two different groups: the ones based on the detection of different types of milk proteins and the ones focused on DNA analysis or genetic techniques [3]. Isoelectric focusing (IEF) of β-casein [4] is the present European Union reference method for cow milk detection.

To carry out the analysis of dairy products, the most employed electrophoretic techniques are cationic polyacrylamide gel electrophoresis (PAGE) [5], capillary isoelectric focusing combined to mass spectrometry [6] and capillary electrophoresis–mass spectrometry [7]. Among the chromatographic techniques, high-performance liquid chromatography with reverse phase columns (RP-HPLC) is employed in the detection and quantification of bovine, ovine and caprine milk percentages in Portuguese protected denomination cheeses [8]. Also, several references about the analysis of caseins [9–11] or β-lactoglobulins [12] in milk and cheeses by means of HPLC can be found. In order to take profit of the advantages of both electrophoretic and chromatographic techniques, sometimes they are combined [13–15]. Further, immunological essays are a widely used tool to identify the kind of milk used in milk mixtures and cheeses, especially immunoenzimatic techniques like ELISA [16–18]. On the other hand, Haasnoot et al. [19] use a fast biosensor immunoassay to detect cow milk in milk of ewe and goat, whereas Haasnoot and Du Pré [20] employed luminex-based triplex immunoassay for the simultaneous detection of soy, pea, and soluble wheat proteins in milk powder. Chromatographic

* Corresponding author. Fax: +34 947258831.
E-mail address: mcortiz@ubu.es (M.C. Ortiz).

0039-9140/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2009.11.067

and immunological techniques are also sometimes combined [21] again in the detection of cow milk in ewe milk cheeses. The latest advances made on molecular biology techniques have promoted the fast development of several genetic techniques successfully applied in the identification of animal species in food. Despite their higher cost and the need of more training than for those techniques based on the analysis of proteins, they offer significant advantages [22]. Especially interesting are their use with products exposed to strong thermical treatments due to the high stability of DNA in these procedures. Polymerase chain reaction (PCR) has become a very useful tool in the quality control of food industry. For example, PCR is used in [23–25] for detecting cow milk in different kinds of cheese, and López-Calleja et al. [26,27] use it to detect the presence of caprine milk. Other examples based on the DNA analysis can be found in [28,29]. The combination of electrophoretic, chromatographic and PCR techniques is used in [30] to identify the kind of milk used in the elaboration of cheese. In addition, several works describe the analysis of milk and dairy products by means of near infrared (NIR) spectroscopy [31,32].

Chemometric techniques for analytical data play a fundamental role in the characterization of foods and in the detection of adulteration. In particular, the chemometric analysis of digitalized profiles (chromatograms and electrophoretograms) allows one the characterization of foods without need of identifying all the detected compounds. This approach is used for the characterization of cheese or milk extracts with electrophoretic [33–35] or chromatographic techniques [36]. Buchgraber et al. [37] determine cocoa butter in milk chocolate by means of triacylglycerol profiling. The protein content in milk powder is also studied through infrared spectra, using least squares support vector machine (LS-SVM) in [38], and comparing the results to the ones obtained by means of PLS. A review of analytical methods coupled with chemometric tools for the determination of the quality and authentication of dairy products was published in 2007 [39]. Among the references where multivariate methods are employed [33–39], techniques such as cluster analysis, soft independent modeling of class-analogy (SIMCA), principal component analysis (PCA) or linear discriminant analysis (LDA) are found for qualitative studies. If the determination is carried out from a quantitative point of view, the employed regression models include LS-SVM, principal component regression (PCReg) and PLS. In the rest of references cited at the beginning of the introduction [5–29], quantification is always carried out by means of univariate regression models.

In the present work, PCA is used for the analysis of protein chromatographic profiles, obtained by means of HPLC-DAD, of mixtures of milk and different cheese samples, and then quantification (percentages of each kind of milk) is made by means of PLS. In order to evaluate the performance of the procedure to detect a specific kind of milk, the probability, α, of false non-compliance (saying that this kind of milk has been used in the sample when it is not) and also the probability, β, of false compliance (saying that the kind of milk is not in the sample when it is false) have been both fixed to 0.05. After that, the decision limit, CCα, and the capability of detection, CCβ, have been obtained. CCα is the percentage at and above which it can be concluded with an error probability of false non-compliance, α, that a sample has a specific kind of milk (in a milk mixture or in cheese). CCβ is the smallest percentage of the milk that may be quantified in a sample with a false compliance probability equal to β and a false non-compliance probability equal to α. The procedure to calculate these figures of merit can be consulted in [40] and [41] for univariate and multivariate cases, respectively. Neither in any of the works previously cited, nor in revision [39] the probabilities of false compliance and false non-compliance have been evaluated in the detection of the kind of milk used in a mixture or in the preparation of cheeses, independently of using univariate or multivariate techniques. The authentication of dairy products is an important issue for food processors and consumers, so minimizing risks of false compliance and false non-compliance in misleading labeling is relevant to evaluate the performance of a proposed procedure.

Table 2
Composition of cows feeding from which milk has been taken for milk analysis or for preparation of cheeses analysed.

	Bovine	
	Milk	Cheese
Forage (60%)		
Grass silo (%)	–	77
Sorghum silo (%)	65	–
Alfalfa hay (%)	17.5	11.5
Vetch-oats hay (%)	17.5	11.5
Feed (40%)		
Barley (%)	22	24
Maize (%)	40	38.8
Oats (%)	4.4	8
Soy (%)	15	16
Bran (%)	13.2	8
Corrector (%)	1.8	2
Bicarbonate (%)	1.4	1.6
Palm soap (%)	1.8	1.6
Salt (%)	0.4	0

2. Experimental

2.1. Samples and chemicals

On one hand, bovine and ovine raw milk was directly obtained from producers in "Centro de Formación Agraria de Viñalta, Junta de Castilla y León" from Palencia (Spain) in two different days, therefore slight differences in their composition could be found. Cows belong to "frisona" breed and are between 2 and 9 years old whereas ewes are from "churra" breed and are between 2 and 8 years old, respectively. Caprine raw milk always came from Ávila (Spain) and belonged to "murciano-granadina" goats. Table 1 describes % fat, % protein, % sec extract, somatic cellules and bacteria per mL for these three kinds of milk and Tables 2 and 3 contain details about animals feeding (cows and ewes, respectively).

Cheese samples analysed were prepared in the "Estación Tecnológica de la Leche" from Palencia (Spain). They were "fresh cheeses" elaborated with milk from "churras" ewes, "frisonas" cows and "murciano-granadina" goats. Their percentage of fat, protein and sec extract can also be seen in Table 1 meanwhile Tables 2 and 3

Table 1
Composition of ovine, bovine and caprine milk used in milk samples studied and in the preparation of the analysed cheeses.

	Ovine		Bovine		Caprine
	Milk	Cheese	Milk	Cheese	Milk
Fat (%)	5.80	6.81	3.57	3.84	4.65
Protein (%)	4.92	5.35	3.14	3.25	3.36
Sec extract (%)	16.61	18.00	8.59	8.83	13.37
Somatic cellules (×1000/mL)	679	1029	374	187	–
Bacteria 50 °C (×1000/mL)	103	146	14	13	–

Table 3
Composition of ewes feeding from which milk has been taken for milk analysis or for preparation of cheeses analysed.

	Ovine	
	Milk	Cheese
Feed		
Barley (%)	91.6	28.6
Oats (%)	–	28.5
Soy (%)	4	40
Corrector? (%)	3	3
Bicarbonate	1	–
Salt	0.4	–
Hay sainfoin (kg/animal)	1.25	1.15
Straw	At discretion	At discretion

contain animals feeding (cows and ewes) in the different periods of year where milk was taken to be analysed or to prepare cheeses. Cheeses were prepared at low pasteurization (63 °C and 30 min) and a ripening time of 3 days. Furthermore, they were vacuum packed in portions and frozen at a temperature between −10 and −20 °C.

In this work, a total of 31 milk samples are prepared and measured in three different times (two consecutive days and 1 month earlier): the first 10 samples are measured the first day, the next 7 the second one and the last 14 for the third day (see Table 4). The percentages of each kind of milk in the mixtures are also described in Table 4: six pure milk samples, nine ovine and bovine mixtures, six ovine and caprine mixtures, six caprine and ovine mixtures and four ternary mixtures.

Twelve cheeses, whose composition is shown in Table 5, are provided. Samples to be analysed are prepared in duplicate and the 24 final samples are also measured in three different days: the first six samples from the first replicate for the first day, the last six samples from the first replicate for the second day and the 12 samples from the second replicate for the third day. It must be highlighted that between the 12 cheeses, there are three which have been elaborated with raw milk whereas the other eight contain pasteurized milk.

Trichloroacetic acid (TCA) employed in proteins precipitation was purchased from Fluka (Steinheim, Germany) and the working solution was prepared at 24% (w/v). Deionised water was obtained from a Milli-Q water purification system (Millipore).

Acetonitrile (ACN) and trifluoroacetic acid (TFA) used in the preparation of the gradient elution in chromatographic analyses were, respectively, obtained from Scharlau and Panreac (Barcelona, Spain). All reagents used in this work were of analytical grade for HPLC. To minimize the loss of peptides all the material used is made of glass.

Table 4
Percentage of each kind of milk contained in the 31 milk mixtures and day of the analysis.

Sample number		Day	Ovine (%)	Bovine (%)	Caprine (%)
1	2	3			
1	–	18	100	–	–
2	–	19	–	100	–
3	–	21	85	15	–
4	–	22	90	10	–
5	–	23	95	5	–
6	–	–	80	20	–
7	–	–	70	30	–
8	–	–	60	40	–
9	–	20	–	–	100
10	–	24	85	–	15
–	11	25	90	–	10
–	12	27	–	15	85
–	13	28	–	10	90
–	14	30	80	10	10
–	15	–	70	15	15
–	16	–	80	–	20
–	17	–	–	20	80
–	–	26	95	–	5
–	–	29	–	5	95
–	–	31	90	5	5

Table 5
Percentage of each kind of milk contained in the 12 cheeses analysed.

Sample number		Ovine (%)	Bovine (%)	Caprine (%)
First replicate	Second replicate			
1	13	–	100	–
2	14	100	–	–
3	15	–	–	100
4	16	98	2	–
5	17	95	5	–
6[a]	18[a]	98	2	–
7	19	98	–	2
8	20	95	–	5
9[a]	21[a]	98	–	2
10	22	–	2	98
11	23	–	5	95
12[a]	24[a]	–	2	98

[a] Cheese samples which contain raw instead of pasteurized milk.

2.2. Preparation of samples for HPLC analysis

Whey protein fractions are obtained from 15 mL of milk or 5 g of cheese to which 15 mL of water have been added before leaving them for 30 min in a sonicator. Proteins are precipitated by the addition of TCA (24%, w/v) until pH 4.6. As a consequence, caseins are eliminated and only soluble proteins (whey proteins) remain. So, the profile studied consists of four groups of proteins: albumines (β-lactoglobulin, α-lactoalbumin and serum albumine), globulins (immunoglobulin: IgG, IgA and IgM), proteoso-peptones and others (lactoferrin).

After heating samples in a water bath at 40 °C for 5 min, they are centrifuged at 4300 rpm and 26 °C for 10 min. Once the resulting supernatant is filtered, a new centrifugation step is carried out with less sample volume and higher intensity (9000 rpm, 26 °C and 10 min). Finally, in order not to damage the chromatographic column, samples are passed through 0.45 μm acetate filters before HPLC analyses.

2.3. Instrumentation and conditions in the HPLC analysis

In order to get the whey a Heraeus Megafuge 1.0 R centrifuge (Thermo Scientific, Milan, Italy) is used.

Chromatographic analyses are carried out at room temperature in a liquid chromatograph from Agilent Technologies including a G1379A vacuum degasser, a G1310A pump, a G1313A injector and a G1315B diode-array detector. Twenty-five microliters are injected into the system and gradient elution is carried out with a mixture of two solvents. Solvent A consisted of 0.1% TFA in water and solvent B consisted of 0.1% TFA in 80% aqueous ACN (v/v). Proteins are eluted with a series of linear gradients increasing the proportion of solvent B, from 36 to 56% in 20 min, from 56 to 60% in 10 min and from 60 to 36% in the last 5 min.

Chromatographic compounds separation is achieved with a reversed-phase column Bio-Rad High Pore RP-318 with dimensions 250 mm × 4.6 mm. The flow rate is 1.0 mL min^{-1} with column temperature equal to 45 °C and the detection is made at a wavelength of 210 nm according to method employed in Ref. [8].

2.4. Software

Chromatographic data acquisition is performed with the aid of Chemstation software from Agilent incorporated in the HPLC equipment. The PLS Toolbox [42] for MATLAB is employed to carry out the PCA and build the calibration models based on PLS. STATGRAPHICS [43] is employed to statistically validate the linear regression models, and a home-made program NWAYDET is used to obtain CCα and CCβ values [40,41].

3. Results and discussion

3.1. Chromatographic signals

3.1.1. Milk samples

Once chromatograms from the 31 milk samples described in Section 2.1 are recorded, the chromatographic profile is constituted by the nine peaks selected in this work. Their retention times are shown in Table 6. It is checked that there are two specific peaks for bovine milk, another two for caprine milk and three for ovine milk. It must be highlighted that the other two peaks are shared between ovine and caprine milk (peak numbers 3 and 7 from Table 6). Chromatograms for pure milk samples (ovine, bovine and caprine) can be seen in Fig. 1 whereas Fig. 2a shows a ternary mixture chromatogram (80% ovine, 10% bovine and 10% caprine milk, respectively). The standardized profile has been obtained by transforming into percentages the areas of the nine selected peaks.

Table 6
Retention times corresponding to the nine selected peaks and their presence in the pure milk samples chromatogram.

Peak number	Retention time (min)	Milk
1	14.9	Caprine
2	15.1	Ovine
3	15.4	Caprine/ovine
4	17.8	Ovine
5	18.3	Ovine
6	19.2	Caprine
7	19.6	Caprine/ovine
8	20.1	Bovine
9	20.8	Bovine

3.1.2. Cheese samples

Chromatographic profile for cheese samples is made up of the sixteen peaks whose retention times are shown in Table 7. It is checked that there are four specific peaks for ovine, four for caprine and one for bovine milk, respectively. In addition, another specific peak, at retention time of 8.5 min, is related to raw milk (samples 6, 9, 12, 18, 21 and 24 in Table 5). Fig. 2b and c shows the chromatograms of two cheese samples prepared with caprine milk mixed with 5% bovine milk (Fig. 2b) and 2% bovine milk (Fig. 2c).

3.2. Principal components analysis (PCA)

With the aim of making an exploratory data analysis in a qualitative way, PCA is carried out both for milk and cheese samples sets separately. In order to have a first structure of data as clear as

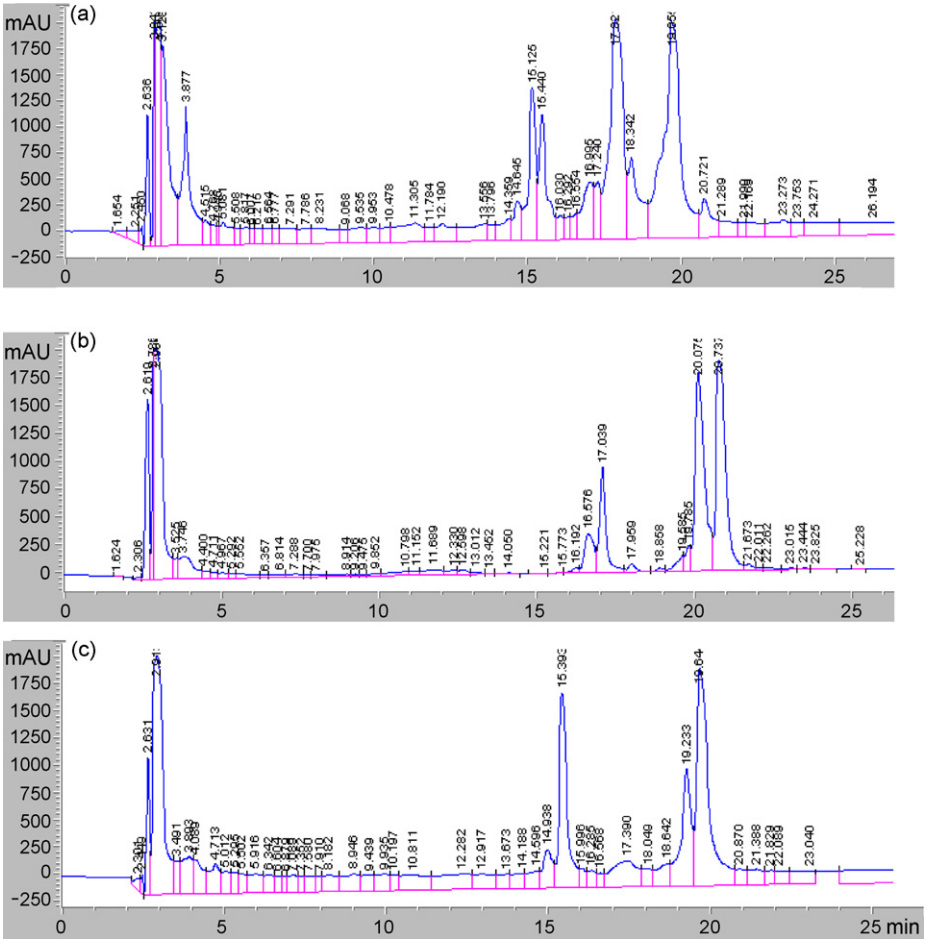

Fig. 1. Chromatograms obtained for pure milk samples: (a) ovine, (b) bovine and (c) caprine.

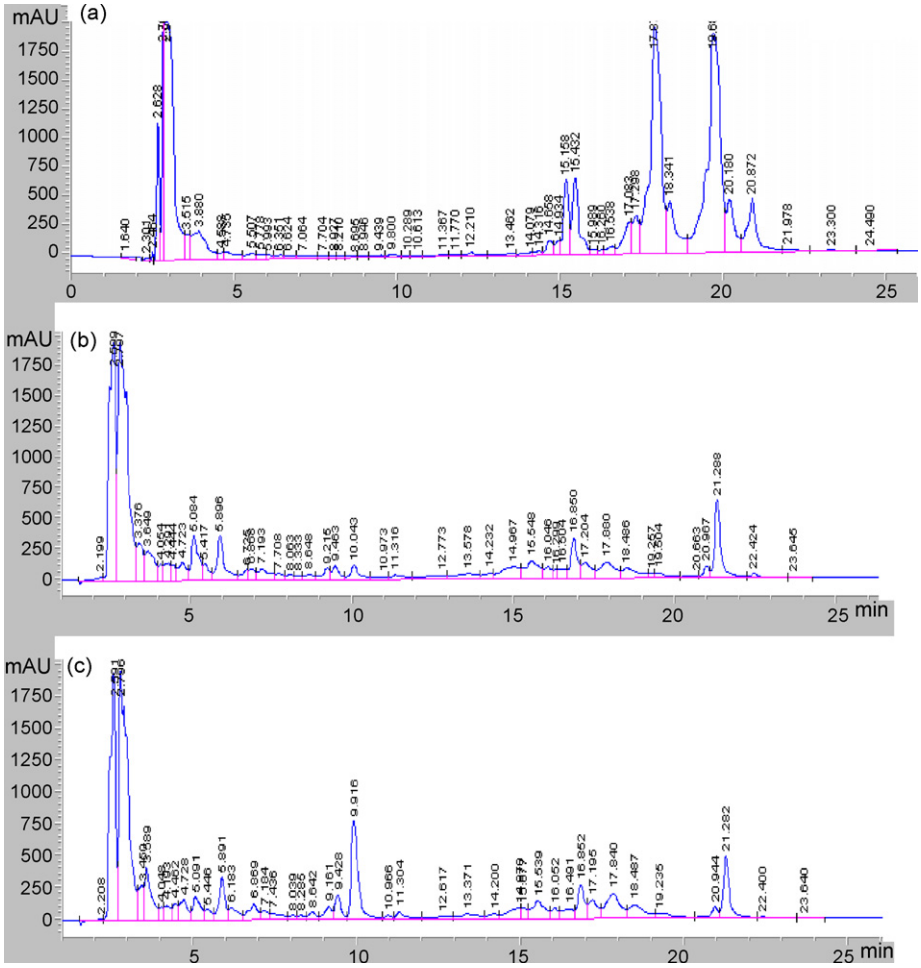

Fig. 2. (a) Chromatograms of a milk mixture with 80, 10 and 10% ovine, bovine and caprine milk, respectively. (b) Chromatograms of a cheese sample made with 95% caprine and 5% bovine milk. (c) Chromatograms of a cheese sample made with 98% caprine and 2% bovine milk.

possible, only samples measured on the same day are chosen: 14 milk samples and 12 cheese samples, arranged in two matrices **X** and **Y** with dimensions (14 × 9) and (12 × 16), where 14 and 12 correspond to the number of milk and cheese samples, respectively, and 9 and 16 are the number of variables or chromatographic peaks chosen in each case.

Table 7
Retention times corresponding to cheese samples elaborated with ovine, caprine or bovine pure milk.

Peak number	Retention time (min)	Milk
1	5.8	Ovine
2	8.5	Raw[a]
3	8.7	Caprine
4	9.3	Ovine/caprine/bovine
5	15.5	Caprine
6	15.9	Caprine
7	16.5	Ovine
8	16.8	Ovine/caprine
9	17.1	Caprine
10	17.7	Ovine/caprine
11	18.4	Caprine/bovine
12	19.2	Ovine
13	19.5	Ovine
14	20.9	Ovine/caprine
15	21.3	Ovine/caprine
16	22.4	Bovine

[a] Specific peak for raw milk.

3.2.1. Milk samples

A PCA model is built with data matrix **X** previously centred by column. The number *k* of principal components will be obtained by means of cross-validation with leave one out procedure.

Q and Hotelling's T^2 statistics are used to identify outliers. The *Q* statistic indicates how well each sample conforms to the model. It is a measure of the difference, or residual, between a sample and its projection into the *k* principal components (or latent variables) retained in the model. The sum of normalized squared scores, known as Hotelling's T^2 statistic, is a measure of the variation of each sample within the model. That is, the distance of each sample to the centroid of the hyper-ellipsoid that makes up the space with the *k* principal components (or latent variables) and with the desired confidence level. Confidence limits can be calculated for *Q* and T^2 at the desired confidence level. Data for which both statistics result higher than the threshold value (usually at 95 or 99% of confidence) are removed and the model will be redone.

By following this procedure, three principal components are chosen. With this model, none of the objects was removed since they did not present *Q* residual or Hotelling T^2 statistics values higher than 95% threshold. Percentages of variance corresponding to the three principal components are 55.26, 39.44 and 4.67%, respectively, which together explain 99.37% of the variability of predictors. Loadings for the nine variables and the two principal components are shown in Fig. 3a for the first principal component and Fig. 3b for the second one, respectively.

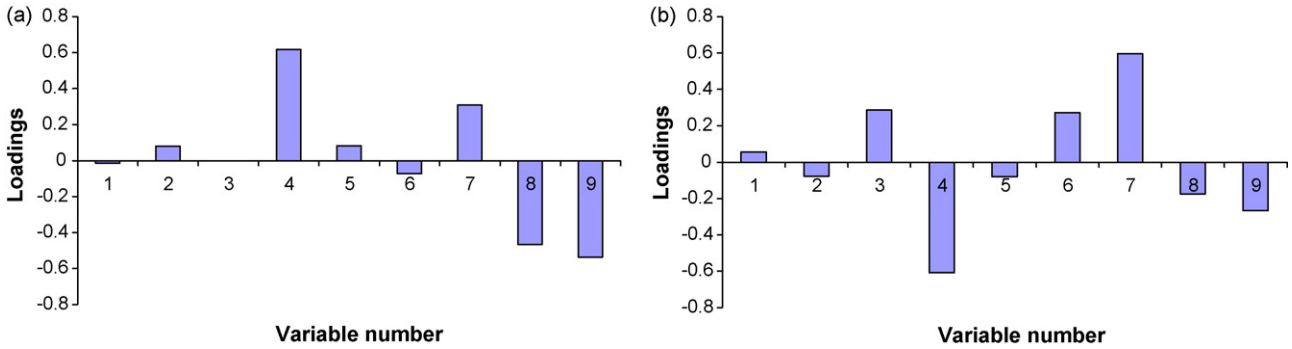

Fig. 3. Loadings on the first (a) and second (b) principal components from PCA models built with milk samples.

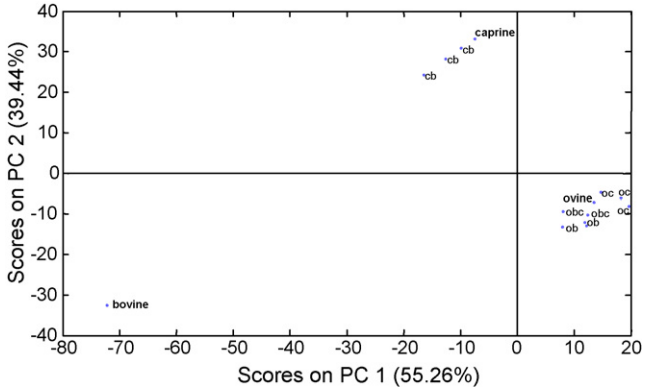

Fig. 4. Scores on the plane second *vs.* first principal components from PCA model with milk samples. "b", "o" and "c" mean bovine, ovine and caprine milk, respectively.

It can be seen that in the first principal component, Fig. 3a, loadings are negative for variables 8 and 9, and positive for variables 4 and 7. The opposite behaviour is also observed between variables 4, 8 and 9 with respect to variables 3, 6 and 7 in the second principal component (see Fig. 3b). Table 6 shows how peaks 8 and 9 are related to the presence of bovine milk, whereas peak 4 is related to the presence of ovine milk, peak 6 to the presence of caprine milk and peaks 3 and 7 to the presence of both ovine or caprine milk in the mixtures, respectively. This explanation has been detailed for first and second principal components since they are the ones which explain the maximum percentage of the variability (94.70%) found in predictors.

Fig. 4 shows the projection of the objects (its scores) on the plane formed by the first two principal components. Letters 'b', 'o' and 'c' in names of Fig. 4 mean the presence of bovine, ovine or caprine milk, respectively, in the mixture. Hence, this Fig. 4 shows a distribution of the objects according to a mixture diagram where pure samples are placed in the vertices of a triangle. Those objects with only caprine or caprine and bovine milk have high scores of the second component and therefore they are placed on the top of Fig. 4 with big areas for peaks 3, 6 and 7 (related to the presence of caprine milk in the mixture) and small for peak 4 (related to ovine milk); it will be also small area from peaks 8 and 9 (bovine milk) due to the little quantity of bovine milk present in the mixtures. In contrast, samples with ovine and bovine milk or ternary mixtures have high scores of the first component (right side of Fig. 4) but lower than pure ovine sample (as it is to be expected because they are mixtures which also have bovine milk). This is in agreement with the fact that the first principal component has positive loadings for variables 4 and 7 (both related to the presence of ovine milk in the mixtures). Also in that side of Fig. 4, but now with higher scores than pure ovine sample for the second component, they are found all samples with ovine and caprine milk (second component has positive loadings for variables 3, 6 and 7, all related to the presence of caprine milk in samples). Furthermore, pure bovine milk sample has the highest and the most negative value of score for the first and second principal components (on the bottom left side of Fig. 4) where variables 8 and 9, the ones related to the presence of bovine milk in mixtures, have negative loadings.

From PCA, it can be concluded that chromatographic profiles gather the information about the kind of milk mixture from each sample.

3.2.2. *Cheese samples*

An analogous PCA as the one carried out in previous section with milk samples is now going to be considered for cheese samples. Thus, a PC model is built with data matrix **Y** also centred by column and two principal components (chosen by cross-validation with leave one out procedure). Once checked the absence of outlier data (Q residual or Hotelling T^2 statistics at 95% threshold), the percentages of variance explained by first and second princi-

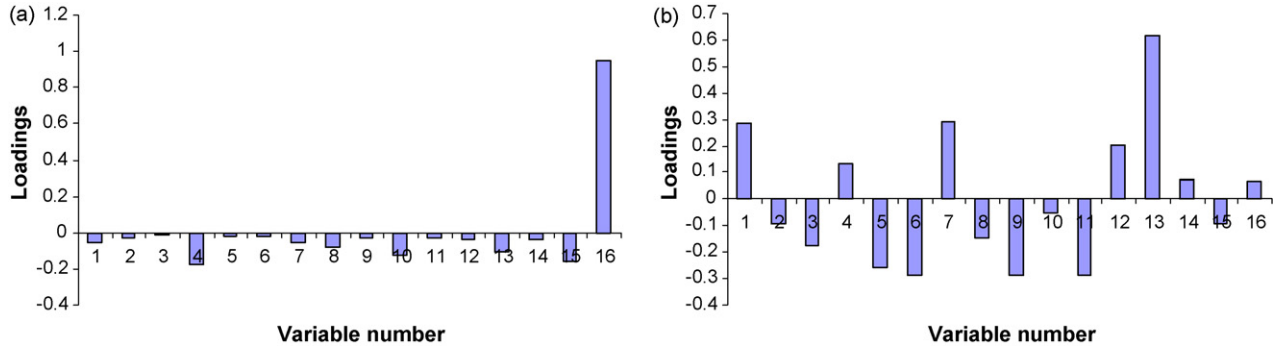

Fig. 5. Loadings on the first (a) and second (b) principal components from PCA models built with cheese samples.

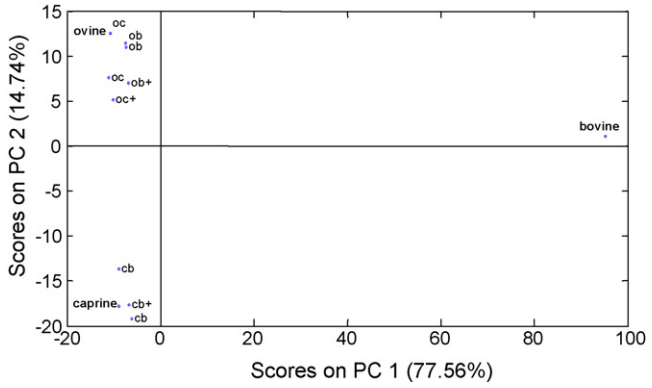

Fig. 6. Scores on the plane second *vs.* first principal components from PCA model with cheese samples. "b", "o" and "c" mean bovine, ovine and caprine milk, respectively.

pal components in the model are 77.56 and 14.74%, respectively, that is a total of 92.30% of variance.

Loadings for the sixteen variables and the two principal components are shown in Fig. 5 in a similar way as in milk case. The first component marks the opposition between variable 16 (peak related to the presence of bovine milk) in contrast to the rest of the variables. The second component is not related to variable 16. It marks the opposition between variables 1, 4, 7, 12, 13, and 14 with positive loadings and related to the presence of ovine milk *vs.* variables 3, 5, 6, 8, 9, 10, 11 and 15 with negative loadings and all related to the presence of caprine milk.

Fig. 6 shows the scores on the plane formed by the first two components similarly to Fig. 4. The same nomenclature is also chosen: letters 'b', 'o' and 'c' mean the presence of bovine, ovine or caprine milk in cheese production and the plus sign (+) indicates that the cheese is prepared with raw milk. The distribution of the objects in this plane has the same shape as in milk study: a triangle with pure cheese samples in the vertices and binary mixtures on its sides. Samples are all very close to the vertices because the kind of milk which is in mayor proportion is always near 100 (95 or 98% according to Table 5). With respect to the situation of the rest of the objects in Fig. 6, the explanation is similar than the one for Fig. 4. Pure bovine cheese sample has high scores for the first principal component (positive loading for variable 16, related to the presence of bovine milk). Regarding the distribution of the objects according to the second principal component, it can be said that samples with ovine milk have high scores (positive loading for variables 1, 4, 7, 12, 13, and 14 related to the presence of ovine milk) whereas samples with caprine milk have low scores for this second principal component (negative loading for variables 3, 5, 6, 8, 9, 10, 11 and 15 related to the presence of caprine milk).

Furthermore, as it happened with milk samples, the principal component analysis shows that the information contained in the chromatographic profile allows describing the milk mixture used in the production of each cheese. Comparing both PC models, less number of components is obtained to describe the samples of cheese, what indicates the existence of fewer sources of variability.

3.3. Partial least squares analysis (PLS)

In order to be able to quantify the percentage of each kind of milk present in a mixture or in the composition of a cheese, PLS calibration models are performed with samples prepared and measured in different days. The summary of all samples is described in Tables 4 and 5 for milk and cheese respectively. If PLS models are

Table 8
True and calculated by means of PLS and univariate models percentages for ovine, caprine and bovine milk with 21 milk samples in the training set and the 10 milk samples in the prediction set.

Samples set	True ovine milk (%)	Calculated ovine milk (%)		True caprine milk (%)	Calculated caprine milk (%)		True bovine milk (%)	Calculated bovine milk (%)		
		PLS	Univariate peak 4		PLS	Univariate peak 6		PLS	Univariate peak 8	Univariate peak 9
Training	100	97.07	103.13	0	2.27	4.58	0	0.83	−2.79	−1.33
	0	−0.01	−1.28	0	1.34	4.58	100	98.23	97.44	100.81
	85	84.93	85.95	0	−1.31	4.58	15	17.52	14.12	14.47
	95	86.63	89.72	0	6.38	4.58	5	11.14	6.61	6.94
	70	74.42	90.29	0	1.10	4.58	30	23.04	22.67	21.27
	0	0.91	−1.28	100	103.95	100.26	0	−4.14	−2.79	−1.33
	85	90.71	95.84	15	10.05	4.58	0	0.27	−2.79	−1.33
	90	95.51	104.60	10	3.89	4.58	0	0.62	−2.79	−1.33
	0	−0.44	−1.28	90	87.16	87.29	10	16.10	19.95	12.13
	70	74.03	78.52	15	8.47	4.58	15	19.29	15.98	17.62
	0	1.31	−1.28	80	70.91	62.01	20	28.43	30.51	30.47
	100	102.75	77.71	0	−2.04	4.58	0	1.03	−2.79	−1.33
	0	−0.40	−1.28	0	2.20	4.58	100	98.07	100.16	98.60
	0	−1.64	−1.28	100	105.22	122.69	0	−2.83	−2.79	−1.33
	90	84.69	90.53	0	6.05	4.58	10	6.33	6.50	7.20
	95	91.32	85.99	0	2.22	4.58	5	5.79	3.51	5.46
	85	86.74	88.96	15	13.83	4.58	0	−3.39	−2.79	−1.33
	90	91.95	77.98	10	9.64	4.58	0	−1.15	−2.79	−1.33
	0	0.59	−1.28	85	82.50	79.61	15	13.32	25.72	19.53
	0	−0.27	−1.28	95	96.45	99.38	5	2.15	10.76	5.81
	90	84.19	85.97	5	9.71	4.58	5	4.35	3.41	5.27
Prediction	90	88.01	90.64	0	−0.91	4.58	10	13.99	10.95	10.61
	80	83.37	98.54	0	−4.46	4.58	20	19.99	16.87	17.83
	60	64.65	74.23	0	3.93	4.58	40	31.66	29.08	30.60
	0	0.85	−1.28	85	79.45	72.62	15	20.98	23.64	21.39
	80	77.28	87.57	10	8.41	4.58	10	14.94	12.84	13.61
	80	85.07	99.94	20	15.79	4.58	0	−2.48	−2.79	−1.33
	85	84.91	83.34	0	3.26	4.58	15	10.36	9.29	10.45
	95	87.57	94.31	5	12.66	4.58	0	−3.43	−2.79	−1.33
	0	−0.98	−1.28	90	89.83	100.95	10	7.43	19.80	12.98
	80	84.10	77.07	10	7.16	4.58	10	8.01	6.64	8.30

able to suitably predict percentages of each kind of milk present in samples measured in different days, it will not be necessary to build calibration models daily.

3.3.1. Milk samples

Data are arranged in a matrix **Z** with dimensions (31 × 9) where 31 correspond to the number of milk samples and 9 to the number of variables recorded (chromatographic peaks in Table 6). The same procedure was followed with cheese samples, arranging their data in a matrix **K** (24 × 16) where 24 correspond to the number of cheese samples and 16 to the number of chromatographic peaks chosen (Table 7).

To evaluate the capability of prediction of the PLS model, samples from matrix **Z** are divided into training set and prediction set so that the 3 days and the different kinds of milk mixtures are represented in both sets and approximately a third of the samples are in prediction set. The training and prediction sets with 21 and 10 samples are shown in Table 8. Three PLS models are performed, data previously centred by column, and taking as response the ovine, bovine or caprine milk percentage in each sample. The number of latent variables in the PLS regression is determined by cross-validation with venetian blinds procedure, reaching the minimum root mean square error in cross-validation (RMSECV) with four latent variables when ovine or caprine milk percentage is used as response and two latent variables when the response is the bovine milk percentage (see Table 9). None object is outlier according to Q residual and T^2 Hotelling statistics. For PLS models, more than 98.28% of the response is explained with four or two latent variables, respectively, depending on the response in the model. The RMSECV varies from 4.59 to 6.57 in these three calibration models.

Values of true and calculated percentages of ovine, caprine and bovine milk in the mixtures for calibration samples are shown in Table 8. To determine the trueness, regression lines between percentages of milk calculated by PLS *vs*. true percentages of milk present in a mixture have been carried out. It is checked whether the regression models have slope and intercept statistically equal to 1 and 0, respectively, and suitable results are got for the three cases because p-values are greater than 0.05 as it can be seen in Table 10.

The capability of prediction of the method is evaluated with samples from the external test set. Percentages of ovine, caprine and bovine milk obtained for them can also be seen in Table 8. By comparing the percentages of ovine, caprine and bovine obtained in calibration and prediction, it can be said that models are stable.

Other figures of merit such as CCα and CCβ have been evaluated to guarantee the quality of PLS models.

In our case, values of CCα for probability of false non-compliance fixed at 0.05, and values of CCβ for probabilities of false non-compliance and false compliance fixed both at 0.05, have been obtained and they are shown in Table 10. As an example, when ovine milk percentage is taken as response in the suitable PLS calibration, CCα is 6.97% and CCβ 13.97%. By following the same methodology, similar values are provided by PLS models when they are the caprine or ovine milk percentages which are, respectively, taken as response in the calibration model and they can again be appreciated in Table 10.

All these results have been compared with the ones obtained through an univariate procedure. For so, the percentages of peak area from peak 4 (related to ovine milk), 6 (related to caprine milk) and 8 or 9 (related to bovine milk) are taken. Three regression lines are built with these percentages *vs*. the true percentage of ovine, caprine and bovine milk respectively present in each sample. Calculated percentages of ovine, caprine and bovine milk are obtained for all samples, their values can be seen in Table 8. Their corresponding CCα and CCβ values are shown in Table 10, as an example CCα and CCβ are equal to 16.93 and 33.43%, respectively, when the percentage of ovine milk is determined. These values are much worse than the PLS ones. Finally, the trueness of the method is checked. It can be concluded that in general, better results are got by means of PLS regression models except for the bovine milk when taking into account information from peak 9.

Table 9
Parameters of the PLS calibration models with data matrix **Z** and **K**. The first 10 raws correspond to milk samples models, whereas the other three belong to cheese samples models. In both cases, ovine, caprine and bovine milk percentage are, respectively, taken as response.

Kind of milk taken as response (%)	Explained variance (%)		L.V.[a]	RMSECV
	X-block	*Y*-block		
Milk samples: data matrix **Z**				
Ovine	53.15	92.67	1	12.76
	96.04	96.91	2	8.35
	97.56	98.63	3	6.22
	99.60	**99.31**	**4**	**4.63**
Caprine	38.45	95.62	1	13.43
	96.02	96.18	2	8.39
	97.40	98.22	3	8.20
	99.59	**98.89**	**4**	**6.57**
Bovine	57.31	96.73	1	6.04
	95.99	**98.28**	**2**	**4.59**
Cheese samples: data matrix **K**				
Ovine	79.72	43.78	1	44.29
	95.19	98.48	2	8.30
	96.28	99.61	3	8.58
	98.21	**99.84**	**4**	**6.72**
Caprine	66.79	44.71	1	39.36
	95.19	98.10	2	8.69
	95.98	99.69	3	9.11
	97.16	**99.90**	**4**	**6.99**
Bovine	84.07	99.90	1	1.17
	94.34	**99.95**	**2**	**1.05**

The chosen models are in bold.
[a] L.V.: number of latent variables.

Table 10
Performance characteristics calculated for the PLS and univariate calibration models. The first three raws correspond to milk samples models, whereas the other three belong to cheese samples models. In both cases, ovine, caprine and bovine milk percentage are, respectively, taken as response.

Kind of milk taken as response (%)	Precision (s_{yx}) (%)		CCα[a] (%)		CCβ[b] (%)		Trueness			
							Intercept (p-value)		Slope (p-value)	
	PLS	Univariate	PLS	Univariate	PLS	Univariate	PLS	Univariate	PLS	Univariate
Ovine	3.78	9.24	6.97	16.93	13.77	33.43	0.38 (0.78)	0.00 (1.00)	0.99 (0.72)	1.00 (0.99)
Caprine	4.40	8.79	7.96	15.76	15.73	31.13	0.33 (0.79)	0.00 (0.99)	0.99 (0.65)	0.99 (0.99)
Bovine	3.88	5.14 (peak 8)	7.04	9.17 (peak 8)	13.91	18.11 (peak 8)	0.27 (0.78)	0.00 (0.99)	0.98 (0.57)	0.99 (0.99)
		3.61 (peak 9)		6.44 (peak 9)		12.71 (peak 9)		0.00 (0.99)	0.99 (0.99)	
Ovine	2.08	9.60	3.92	18.08	7.72	35.56	0.09 (0.91)	0.00 (0.99)	0.99 (0.88)	1.00 (0.99)
Caprine	1.52	7.38	2.81	13.57	5.52	26.69	0.03 (0.95)	0.00 (0.99)	1.02 (0.91)	1.00 (0.99)
Bovine	0.80	0.94	1.47	1.71	2.89	3.37	0.01 (0.97)	0.00 (0.99)	0.99 (0.93)	1.00 (0.99)

[a] Probability of false non-compliance fixed at 0.05
[b] Both probabilities of false non-compliance and false compliance fixed at 0.05.

Table 11
True and calculated by means of PLS and univariate models percentages for ovine, caprine and bovine milk with 16 cheese samples in the training set and the 8 cheese samples in the prediction set.

Samples set	True ovine milk (%)	Calculated ovine milk (%)		True caprine milk (%)	Calculated caprine milk (%)		True bovine milk (%)	Calculated bovine milk (%)	
		PLS	Univariate peak 7		PLS	Univariate peak 9		PLS	Univariate peak 16
Training	0	0.05	0.19	0	−0.04	1.03	100	99.98	100.04
	100	101.84	114.72	0	−0.97	1.03	0	0.08	−0.06
	0	−4.15	0.19	100	102.82	100.12	0	0.11	−0.06
	95	94.09	74.59	0	0.47	1.03	5	5.06	5.50
	98	101.19	84.45	2	−0.39	1.03	0	−0.31	−0.06
	98	97.56	100.06	2	2.85	1.03	0	−0.14	−0.06
	0	−0.80	0.19	95	94.48	80.70	5	4.91	4.41
	0	1.24	0.19	98	98.63	115.60	2	3.22	2.73
	0	0.05	0.19	0	−0.04	1.03	100	99.98	100.04
	100	102.09	112.58	0	−0.73	1.03	0	0.04	−0.06
	0	0.33	0.19	100	99.16	100.95	0	−0.08	−0.06
	98	96.72	112.99	0	1.02	1.03	2	3.04	2.94
	98	95.05	99.03	0	2.05	1.03	2	3.01	3.59
	95	94.74	90.53	5	3.75	1.03	0	−0.67	−0.06
	98	97.00	89.89	2	3.44	1.03	0	−0.01	−0.06
	0	4.02	0.19	95	92.51	90.46	5	2.79	2.22
Prediction	98	96.08	122.42	0	−0.65	1.03	2	4.98	5.09
	98	97.99	85.94	0	−2.14	1.03	2	5.19	5.08
	95	101.72	90.83	5	−2.77	1.03	0	−0.04	−0.06
	0	9.51	0.19	98	84.84	88.67	2	2.39	2.85
	95	95.89	81.55	0	0.97	1.03	5	2.91	3.39
	98	100.54	86.64	2	0.26	1.03	0	0.02	−0.06
	0	16.24	0.19	98	80.32	90.55	2	0.07	0.94
	0	6.51	0.19	98	95.30	64.32	2	1.36	1.02

3.3.2. Cheese samples

The way to tackle the problem with cheese samples is analogous to the one described for milk samples. Firstly, 16 samples from matrix **K** are chosen to take part in the training set whereas the other 8 will be used to evaluate the capability of prediction of the model, so that, different days and kinds of milk mixtures are in both sets. The three PLS models (taking as response the ovine, bovine or caprine milk percentage in each cheese) are again performed over the training set from experimental matrix **K** data previously centred by column. In PLS regression four latent variables, when ovine or caprine milk percentage is the response or two when it is the percentage of bovine milk, are chosen by cross-validation with leave one out procedure in the three cases. None of the objects is outlier according to values of Q residual and T^2 Hotelling statistics at 95% threshold. For the three PLS models, more than 99.84% of the response is explained and the RMSECV values vary from 1.05 to 6.72. Parameters of PLS calibration models, just as the values of some performance characteristics calculated for them can be seen in Tables 9 and 10. The values of true and calculated percentages of ovine, caprine and bovine milk in cheeses for calibration and test samples are respectively shown in Table 11.

As it was to be expected, regressions built with percentages calculated by PLS models *vs*. true percentages have, in all cases, slope and intercept statistically equal to 1 and 0, being able to conclude that the method is trueness when calculated by PLS model *vs*. true milk percentage regressions are performed (see columns 8 and 10 from Table 10).

Percentages of peak area from peak 7 (related to ovine milk), 9 (related to caprine milk) and 16 (related to bovine milk) are taken with the aim of comparing these results with the univariate ones. Table 11 shows the calculated percentages of three kinds of milk obtained for calibration and test samples. Values of the corresponding performance characteristics can be seen in Table 10 by concluding the same as for milk samples: much better results are got by means of PLS regression models than the univariate ones.

By comparing values of CCα and CCβ calculated for PLS calibration models with cheese samples with the ones obtained for models with milk samples, it can be said that, better results are attaint in models with cheese samples. CCα values obtained are 3.92, 2.81 and 1.47% and those of CCβ 7.72, 5.52 and 2.89% for ovine, caprine and bovine milk percentage, respectively.

The adsorption of peptides is a well know phenomenon [44] and reduces the analyte concentration, increasing the risk of false negative, in other words, increasing CCβ. The adsorption is a concentration dependent surface phenomenon, and the lower concentrations are studied the higher peptides losses are observed [45]. This effect is important in pharmaceutical and proteomic research, but the concentration of peptides associated to a, say, 5% of milk in cheese is several orders of magnitude greater than the one reported in these references, so this effect is negligible in the analyses carried out in the present work. In the quantification stage, standard samples with known quantities of each kind of milk are used. That means that the PLS calibration models gather the possible loss (signal reduction) that could happen. Tables 8 and 11 show that the differences, in relative error for both milk and cheese samples, are very small in training and prediction samples.

It must be added, as final note, that PLS calibration models have the advantage of being able to identify samples which are different from the ones used in training set, avoiding, therefore, a wrong use of the PLS model when predicting the percentage of each kind of milk present in the analysed sample. When these models are used in a routine way, the non-similar samples are separately studied so that, later, samples with the 'new' structure are incorporated to the training set. In this way, the model can be actualized or extended including temporary milk changes or new industrial practices for cheese making.

Acknowledgements

Authors thank the Instituto Tecnológico Agrario de Castilla y León and the Consejería de Agricultura y Ganadería for their collaboration on the development of this work. They also acknowledge the Ministerio de Ciencia e Innovación (Project CTQ2008-02264/BQU), the Junta de Castilla y León (Project

BU024A07) and FEDER funds for financial support. N. Rodríguez gratefully acknowledges the Junta de Castilla y León for her FPI Grant.

References

[1] Commission Regulation (EC) No. 213/2001 of 9 January 2001 laying down detailed rules for the application of Council Regulation (EC) No. 1255/1999 as regards methods for the analysis and quality evaluation of milk and milk products and amending Regulations (EC) No. 2771/1999 and (EC) No. 2799/1999, Off. J. L037, 07/02/2001, 1–99.
[2] Council Regulation (EEC) No. 2081/92 of 14 July 1992 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs, Off. J. L208, 24/07/1992, 1–8.
[3] I. López-Calleja, I. González, V. Fajardo, I. Martín, M. Rojas, M.A. Pavón, T. García, R. Martín, CTC Aliment. 28 (2006).
[4] Commission Regulation (EC) No. 1081/96 of 14 June 1996 establishing a reference method for the detection of cows' milk and caseinate in cheeses from ewes' milk, goats' milk and buffalos' milk or mixtures of ewes' milk, goats' milk and buffalos' milk, Off. J. L142, 15/06/1996, 15–25.
[5] S.E. Kaminaridesa, P. Koukiassab, Food Chem. 78 (2002) 53.
[6] A. Somma, P. Ferranti, F. Addeo, R. Mauriello, L. Chianese, J. Chromatogr. A 1192 (2008) 294.
[7] L. Müller, P. Barták, P. Bednár, I. Frysová, J. Sevcík, K. Lemr, Electrophoresis 29 (2008) 2088.
[8] I.M.P.L.V.O. Ferreira, H. Caçote, J. Chromatogr. A 1015 (2003) 111.
[9] G. Moatsou, A. Hatzinaki, G. Psathas, E. Anifantakis, Int. Dairy J. 14 (2004) 219.
[10] E. Bramanti, C. Sortino, M. Onor, F. Beni, G. Raspi, J. Chromatogr. A 994 (2003) 59.
[11] S. Sforza, G. Aquino, V. Cavatorta, G. Galaverna, G. Mucchetti, A. Dossena, R. Marchelli, Int. Dairy J. 18 (2008) 1072.
[12] G. Enne, D. Elez, F. Fondrini, I. Bonizzi, M. Feligini, R. Aleandri, J. Chromatogr. A 1094 (2005) 169.
[13] A.C.A. Veloso, N. Teixeira, A.M. Peres, A. Mendonça, I.M.P.L.V.O. Ferreira, Food Chem. 87 (2004) 289.
[14] A.C.A. Veloso, N. Teixeira, I.M.P.L.V.O. Ferreira, J. Chromatogr. A 967 (2002) 209.
[15] H.K. Mayer, D. Heidler, C. Rockenbauer, Int. Dairy J. 7 (1997) 619.
[16] I.P. Hurley, H.E. Ireland, R.C. Coleman, J.H.H. Williams, Int. J. Food Sci. Technol. 39 (2004) 873.
[17] W. Richter, I. Krause, C. Graf, I. Sperrer, C. Schwarzer, H. Klostermeyer, Z. Lebensm. Unters. Forsch. A 204 (1997) 21.
[18] I.P. Hurley, R.C. Coleman, H.E. Ireland, J.H.H. Williams, Int. Dairy J. 16 (2006) 805.
[19] W. Haasnoot, N.G.E. Smits, A.E.M. Kemmers-Voncken, M.G.E.G. Bremer, J. Dairy Res. 71 (2004) 322.
[20] W. Haasnoot, J.G. Du Pré, J. Agric. Food Chem. 55 (2007) 3771.
[21] H. Colak, A. Aydin, B. Nazli, O. Ergun, Food Control 17 (2006) 905.
[22] I. López-Calleja, I. González, V. Fajardo, I. Martín, M. Rojas, M.A. Pavón, T. García, R. Martín, CTC Aliment. 28 (2006).
[23] I. López-Calleja, I. González Alonso, V. Fajardo, M.A. Rodríguez, P.E. Hernández, T. García, R. Martín, Int. Dairy J. 15 (2005) 1122.
[24] I. Mafra, A. Roxo, I.M.P.L.V.O. Ferreira, M.B.P.P. Oliveira, Int. Dairy J. 17 (2007) 1132.
[25] J. Bania, M. Ugorski, A. Polanowski, E. Adamczyk, J. Dairy Res. 68 (2001) 333.
[26] I. López-Calleja, I. González, V. Fajardo, I. Martín, P. Hernández, T. García, M. Martín, Food Res. Technol. 225 (2007) 345.
[27] I. López-Calleja, I. González, V. Fajardo, I. Martín, P. Hernández, T. García, M. Martín, Food Control 18 (2007) 1466.
[28] S. Reale, A. Campanella, A. Merigioli, F. Pilla, J. Dairy Res. 75 (2008) 107.
[29] C. Maudet, P. Taberlet, J. Dairy Res. 68 (2001) 229.
[30] H.K. Mayer, Int. Dairy J. 15 (2005) 595.
[31] S.N. Jha, T. Matsuoka, Int. J. Food Sci. Technol.-Mysore 41 (3) (2004) 313.
[32] J.L. Rodríguez-Otero, M. Hermida, J. Centeno, J. Agric. Food Chem. 45 (8) (1997) 2815.
[33] D. Wu, Y. He, S. Feng, D.W. Sun, J. Food Eng. 84 (2008) 124.
[34] E. Molina, P.J. Martín-Álvarez, M. Ramos, Int. Dairy J. 9 (1999) 99.
[35] B. Vallejo-Córdoba, J Capillary Electrophor. 5 (3–4) (1998) 133.
[36] P. Piraino, E. Parente, P.L.H. Mcsweeney, J. Agric. Food Chem. 52 (2004) 6904.
[37] M. Buchgraber, S. Androni, E. Anklam, J. Agric. Food Chem. 55 (2007) 3284.
[38] J.M. Herrero-Martínez, E.F. Simó-Alfonso, G. Ramis-Ramos, C. Gelfi, P.G. Righetti, Electrophoresis 21 (2000) 633.
[39] R. Karoui, J. De Baerdemaeker, Food Chem. 102 (2007) 621.
[40] M.C. Ortiz, L.A. Sarabia, Trends Anal. Chem. 13 (1994) 1.
[41] M.C. Ortiz, L.A. Sarabia, I. García, D. Giménez, E. Meléndez, Anal. Chim. Acta 559 (2006) 124.
[42] M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, PLS Toolbox 4.01, Eigenvector Research Inc., Manson, WA, 2006.
[43] STATGRAPHICS plus for Windows Ver. 5.1, Statistical Graphics Corp., Rockville, 1998.
[44] H. John, M. Walden, S. Schäfer, S. Genz, W.G. Forssmann, Anal. Bioanal. Chem. 378 (2004) 883.
[45] A. Pezeshki, V. Vergote, S. van Dorpe, B. Baert, C. Burvenich, A. Popkov, B. de Spiegeleer, J. Pharm. Biomed. Anal. 49 (2009) 607.