Available online at www.sciencedirect.com
ScienceDirect
Chemometrics and Intelligent Laboratory Systems 86 (2007) 60–67
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Multivariate statistical methods for Port Salut Argentino cheese analysis based on ripening time, storage conditions, and sampling sites

R.A. Verdini [a,b,c,*], S.E. Zorrilla [b], A.C. Rubiolo [b], S. Nakai [a]

[a] *University of British Columbia, Faculty of Agricultural Sciences, Food, Nutrition and Health Program, 6640 N.W. Marine Drive, Vancouver, British Columbia, Canada, V6T 1Z4*
[b] *Instituto de Desarrollo Tecnológico para la Industria Química (INTEC)–Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Litoral (UNL)–Güemes 3450, (3000) Santa Fe, Argentina*
[c] *Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario (UNR), Suipacha 531, (2000) Rosario, Argentina*

Received 15 May 2006; received in revised form 15 August 2006; accepted 18 August 2006
Available online 6 October 2006

Abstract

The objective of the present work was to compare multivariate statistical methods for the classification of Port Salut Argentino cheese samples based on ripening time (1, 6, 13, 27, and 56 days), storage conditions (traditionally ripened and ripened after frozen storage) and sampling sites (internal and external zones) using the contents of caseins, peptides and amino acids measured by chromatographic analysis as well as textural and physical parameters. In particular, two linear methods, principal component analysis (PCA) and principal component similarity (PCS), and a non-linear method, the Kohonen self-organizing artificial neural network (Kohonen ANN), were compared. The two linear methods showed the same grouping of cheese samples according to ripening time, sampling site and storage condition. These methods are closely related in their mathematical basis and the similar grouping showed by both methods can be explained by the fact that the first three principal components explained 89.3% of the data set variation. The non-linear Kohonen ANN uses a mathematical procedure completely different from PCA; however, only slight differences were observed in the grouping of cheese samples. Those differences may be related to the weight that each model gives to every variable. One interesting feature of Kohonen ANN is that weight maps (contour plots) sometimes are superior to principal component loadings (vectors) for the understanding of relationships between the groups and the original variables.

Keywords: Multivariate analysis; Neural networks; Cheese; Ripening; Freezing

1. Introduction

In a recent review, Coker et al. [1] discussed classification of cheese variety and maturity, based on the contents of caseins, peptides and amino acids measured by electrophoretic and/or chromatographic analysis. When a large amount of data must be considered, an objective assessment is essential.

A wide spectrum of multivariate methods is available in order to extract information from the data set. Pripp et al. [2] considered that multivariate analysis of proteolytic profiles was a powerful approach to discriminate cheese varieties, cheese quality and starter strains. Moreover, it also helps to better understand proteolysis during cheese ripening and how a change in cheese technology affects ripening and cheese quality.

One of the simplest and most frequently used tasks in handling complex multivariate data is the mapping of objects and variables from an *m*-dimensional into a two-dimensional space [3]. The main concern is the visual representation of objects (e.g., cheese samples) and variables (e.g., peak areas or chemical results). The mapping of objects and variables in a two-dimensional space allows the grouping of objects and the

* Corresponding author. Instituto de Desarrollo Tecnológico para la Industria Química (INTEC)–Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Litoral (UNL)–Güemes 3450, (3000) Santa Fe, Argentina. Tel.: +54 342 4559174/77; fax: +54 342 4550944.
E-mail address: rverdini@intec.unl.edu.ar (R.A. Verdini).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2006.08.006

understanding of relationships between the groups and the original variables.

Several authors have applied principal component analysis (PCA) for mapping multivariate data of cheese samples in two-dimensional plots, and in most cases, the first principal component (PC) was related to ripening time [4–11]. However, when more than two PCs account for most of the data variation, other multivariate methods can be more useful than PCA or at least complementary.

Principal component similarity (PCS), which derives from PCA, was introduced by Vodovotz et al. [12]. Furtula et al. [13] studied proteolytic profiles of Cheddar cheese samples aged by fast ripening process. When six PCs accounted for most of the data variability, PCS was very useful for classifying cheese samples as well as for assessing the effectiveness of the fast ripening process. The authors also applied PCS to data obtained from Cheddar cheeses that were accidentally exposed to ambient conditions and the abused samples belonged to a distinctive group in the PCS scattergram [14].

Recently, artificial neural networks (ANN) have been used for the multivariate analysis of cheese samples. European Emmental cheeses were classified depending on the geographic origin using 25 analytical parameters for ANN computation [15]. The authenticity of Ossolano cheese according to the Protected Designation of Origin labelling was validated using the composition of fatty acids and proteolytic profiles analysed by ANN [16].

Port Salut Argentino cheese is one of the most popular soft cheese varieties in Argentina. Expanding commercialisation of Port Salut Argentino cheeses has increased interest in preserving its characteristics for a longer storage period [8]. Freezing is one of the most effective treatments to ensure long-term preservation. The effects of the freezing process on cheese attributes, namely, flavour, aroma and texture, depend on whether cheeses are frozen before or after the ripening period. Particularly, when cheeses are frozen prior to ripening, the freezing process may influence some of the physicochemical transformations during cheese maturation [10].

The aim of the present work was to compare multivariate statistical methods for the classification of samples (Port Salut Argentino cheese) based on ripening time (1, 6, 13, 27, and 56 days), storage conditions (traditionally ripened and ripened after frozen storage) and sampling sites (internal and external zones) using the contents of caseins, peptides and amino acids measured by chromatographic analysis as well as textural and physical parameters. In particular, two linear methods, PCA and PCS, and a non-linear method, the Kohonen self-organizing artificial neural network (Kohonen ANN), were compared.

2. Materials and methods

2.1. Cheese samples

Data from our preceding papers on Port Salut Argentino cheeses (from the same batch) were used in this study [8–10,17,18]. Cheeses were manufactured at a local factory, salted in a brine solution for 3 h at 3 °C, stored for 20 h and packed in heat-shrinkable plastic bags. Port Salut Argentino cheeses were 3.55±0.11 kg weight, 23.2±0.3 cm diameter, 7.7±0.3 cm height, 28.7±0.7% w/w fat, 20.4±0.9% w/w total protein, 48.8±2.6% w/w moisture, and 5.2±0.1 pH.

Thirty cheeses were transported in insulated boxes with ice from the factory to our laboratory and randomly separated in two groups. Fifteen cheeses were frozen in a Tabai Comstar PR 4GM chamber (Tabai Espec Corp., Osaka, Japan) at −30 °C until the centre reached −22 °C, held in frozen storage at −22 °C for 30 days, thawed at 5 °C and held for ripening at 5 °C (cheeses F). Fifteen cheeses were held at 5 °C for ripening and were used as control (cheeses C). Cheeses were sampled at different ripening times (1, 6, 13, 27, and 56 days) in triplicate.

Cubic pieces of 2.5 cm each side were cut from two different cheese zones, internal and external zone. The geometric centre of the cubic pieces of internal and external zones were separated approximately 5.5 cm in the radial direction and 2.0 cm in the axial direction [17].

2.2. Moisture and chloride analysis

Moisture content was measured with a microwave oven CEM AVC 80 (CEM, Mattheus, NC, USA). Chloride concentration was determined with an Automatic Titrator model DL40RC (Mettler Instrumente AG, Greifensee, Switzerland) as proposed by Fox [19].

2.3. Extraction and chromatographic analysis of nitrogenous compounds

Grated cheese (10 g) mixed with three times the sample weight of water was homogenised using an Ultra-Turrax® T25 (IKA® Werke, Janke & Kunkel GmbH & Co KG, Staufen, Germany) for 2 min [20]. The fractionation scheme proposed by Verdini et al. [9] was followed for the separation of three nitrogenous fractions: the water-insoluble fraction, the water-soluble fraction (WSF), and the sulfosalicylic acid-soluble fraction (SSASF). The nitrogenous fractions were stored in a freezer at −22 °C for further analysis.

The chromatographic analysis of the three nitrogenous fractions was performed as described in Verdini et al. [9]. Two peaks were analysed in the water-insoluble fraction of Port Salut Argentino cheese samples: the α_{s1}-casein peak was identified using a standard of α_s-casein (Sigma-Aldrich, St. Louis, MO, USA) and the α_{s1}-CN (f24-199) fragment, known as α_{s1}-I-casein, was assigned as described by Verdini et al. [9]. Sixteen peaks of the WSF that characterised Port Salut Argentino cheese ripening were selected for further analysis [9,10]. Fifteen free amino acids were detected and quantified in the SSASF of Port Salut Argentino cheese samples [9–11].

2.4. Compression and stress relaxation tests

Samples were put into plastic bags to prevent dehydration and left in the test room for 3 h to reach test temperature. Experiments were carried out at 21±1 °C. Samples were compressed by a Universal Testing Machine (Schimadzu DSS

10 T-S, Tokyo, Japan) with a 5-kg load cell. Relaxation curves were recorded for 8 min as suggested by Peleg [21]. Compression ratio of 40% and cross-head speed of 1 cm min^{-1} were used. Data were collected with a personal computer throughout an analogical output of the Universal Testing Machine. Stress relaxation data were normalised using Peleg's model with an empirical linear equation and the asymptotic equilibrium modulus (EA) of the normalised stress relaxation curve was obtained [21,22].

2.5. Data set

Three factors were considered for this study: storage condition, ripening time and sampling site. The nomenclature used for identifying the studied factors was storage condition (control: C, frozen: F), sampling site (internal: I, external: E), and ripening time (1, 6, 13, 27, and 56). Consequently, a cheese sample named FI1 belongs to a frozen cheese, sampled in the internal zone, and ripened for 1 day.

Thirty-six variables were considered for this study: EA (Pa), moisture (%w/w), NaCl (%w/w), α_{s1}-casein and α_{s1}-I-casein (peak areas per 100 g cheese), sixteen peaks of the WSF (peak areas per 100 g cheese) and 15 free amino acids in the SSASF (mg amino acid per 100 g cheese).

As a result, the data set consisted of 60 cheese samples and 36 input variables. However, data from nitrogenous compounds in the WSF and free amino acids in the SSASF were large and highly correlated. Consequently, to reduce dimensionality, data subsets from nitrogenous compounds in the WSF and free amino acids in the SSASF were analysed for data reduction.

2.6. Data reduction

The term data reduction in the context of data mining is usually applied when the goal is to aggregate or amalgamate the information contained in large data sets into manageable (smaller) information nuggets (Statistica 7 electronic user manual, StatSoft Inc., Tulsa, OK, USA). PCA was used to reduce dimensionality of the data from nitrogenous compounds in the WSF and free amino acids in the SSASF. The fundamentals of PCA are outlined in Section 2.7.

2.7. Data analysis

The reduced data set was analysed using PCA, PCS, and Kohonen ANN. Principal component analysis was performed with Minitab 13.20 using the correlation matrix (Minitab Inc., State College, PA, USA); PCS was conducted using the algorithm proposed by Vodovotz et al. [12] and the Kohonen ANN was implemented using Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA).

2.7.1. Principal component analysis

Principal component analysis is based on the linear combination of the measured variables to produce derived variables, principal components (PCs), which are mutually orthogonal in the principal component space [23]. The PCs are numbered in order of the amount of variation in the original data set that they explain, so that the PC1 accounts for the most variation, and each subsequent principal component accounts for as much of the remaining variation as possible. An adequate condensation of the information is achieved when no more than two or three PCs can explain at least 80–90% of the total variability [23–25]. It is often possible to view the structure in the data by plotting two or three of the most important PCs. The analysis of the PC scores gives evidence of sample grouping in the PC space according to similarities in their characteristics while the examination of the PC loading considers the influence of the original variables in the sample arrangement [23]. Principal component analysis was applied to the mean centred data matrix as suggested by Verdini and Rubiolo [7].

Table 1
Average of three values of the scaled variables used for data analysis

Storage	Sampling site	Ripening time (days)	Variables						
			α_{s1}-Casein	α_{s1}-I-Casein	PC1-WSF	PC1-SSASF	EA	Moisture	NaCl
Control	Internal	1	87.82	18.37	5.44	2.93	70.63	31.32	3.23
		6	84.41	11.43	1.95	3.73	60.10	20.23	12.76
		13	82.38	21.30	9.82	8.71	51.70	38.64	31.39
		27	55.73	30.90	37.86	33.67	23.17	20.36	31.22
		56	29.96	60.86	78.22	76.03	9.49	54.41	62.41
	External	1	63.77	21.29	2.17	0.96	55.58	74.40	51.60
		6	67.54	13.32	5.17	4.87	36.34	73.09	53.34
		13	66.54	20.18	15.07	6.76	20.57	79.71	75.57
		27	46.68	42.53	43.23	38.37	5.28	74.57	51.86
		56	22.95	66.37	73.23	50.39	0.45	92.86	67.72
Frozen	Internal	1	69.18	5.45	4.36	2.82	82.13	16.21	3.11
		6	80.66	5.99	10.42	8.91	48.06	43.30	13.86
		13	25.84	60.77	32.75	20.90	24.02	25.15	29.16
		27	30.49	47.45	37.50	65.41	39.02	27.98	43.92
		56	8.71	26.37	42.90	98.21	9.50	54.44	48.74
	External	1	87.39	2.91	2.47	2.05	40.27	72.57	59.11
		6	91.78	14.08	11.72	7.66	20.64	82.51	56.88
		13	24.22	77.56	30.61	23.31	7.47	75.12	58.98
		27	18.35	43.74	32.29	50.85	10.80	70.94	58.15
		56	0.64	25.29	37.44	74.86	0.00	92.86	52.76

2.7.2. Principal component similarity

Principal component similarity is an algorithm that derives from PCA and can also be categorised as unsupervised learning. After application of PCA to the original data, linear regression analysis is carried out using the PC scores with eigenvalues higher than 1. The algorithm for PCS computation was proposed by Vodovotz et al. [12]. The independent variables for the linear regression are the accumulated proportion of variability computed from the eigenvalues. A reference sample is selected and the dependent variables for the linear regression are the deviations from the corresponding reference PC scores. The result of PCS analysis is a plot of slope vs. coefficient of determination derived from the linear regression analysis. PCS plots do not allow the differentiation of the contribution of each PC to the grouping of the samples. Subsequently, additional information can be obtained from PCS analysis throughout adjusted factor score deviation plots that show the deviations of PC scores from a reference line [12]. Then, PCs with higher deviations from the reference line can explain the relationships between the groups and the original variables.

2.7.3. Kohonen self-organising artificial neural networks

The Kohonen self-organising artificial neural network creates a two-dimensional map from a series of high-dimensional feature vectors. Kohonen mapping is mathematically even simpler than the PCA [26]. A Kohonen ANN has only two layers: the input layer and an output layer of radial units known as the topological map (top-map). The most characteristic feature of the Kohonen ANN is its implementation of corrections [27,28]. The number and the extension of the corrections change during the learning. A time-decaying learning rate, which is used to perform the weighted sum, ensures that the alterations become subtler as the epochs pass. The neighbourhood, a set of neurons surrounding the winning neuron, also decreases over time. Once the network has been trained to recognise structure in the data, it can be used as a visualisation tool to examine the data [27]. The top map of an $n \times n$ Kohonen ANN has $n \times n$ entries, each of which corresponds exactly to one neuron and the number of weights in each neuron is equal to the dimension of the input vector. One interesting feature of the Kohonen ANN is that each neuron has the same number of weights and in each level of weights only data of one specific variable are handled [27,28]. Before training starts, data are randomized in the interval [0, 1], being the initial weight, and at the end of learning in each level a map showing the distribution of values of the particular variable is formed.

3. Results and discussion

3.1. Data reduction

It is well known that classification improves when more discriminative variables are used. In general, the network sacrifices classification performance in an attempt to reconstructing all data, and consequently, data for modelling purposes can be inefficiently used. In other words, redundant data make the network inefficient for classification or sample grouping. Therefore, it is very important to recode data into fewer meaningful categories, which are more likely to yield meaningful results. So data mining consists not only of reducing the number of variables in a black-box approach, but also of reducing the number of variables applying domain-specific knowledge.

In this case, principal component analysis applied to 60 cheese samples × 16 variables (WSF) and to 60 cheese samples × 15 variables (free amino acids in the SSASF), yielded one PC that explained 86.5% and 93.7% of the data set variation, respectively (Verdini et al. [10]). In both cases, the first PC accounted for more than 85% of the variability so one input variable PC1-WSF replaced the 16 original variables (peak areas of WSF), and one input variable, PC1-SSASF, replaced the 15 original variables (free amino acids in the SSASF).

The new input variables were EA, moisture, NaCl, α_{s1}-casein, α_{s1}-I-casein, PC1-WSF, and PC1-SSASF. Because the input variables differed in magnitude, all values of a given variable were scaled from 0 to 100 with respect to the range between the smallest and the largest variable value [3]. Values of the scaled variables are shown in Table 1.

3.2. Principal component analysis

Principal component analysis applied to the 60 samples and 7 variables yielded three PCs that explained 89.3% of the data set variation (PC1 60.8%, PC2 20.6%, and PC3 7.9%).

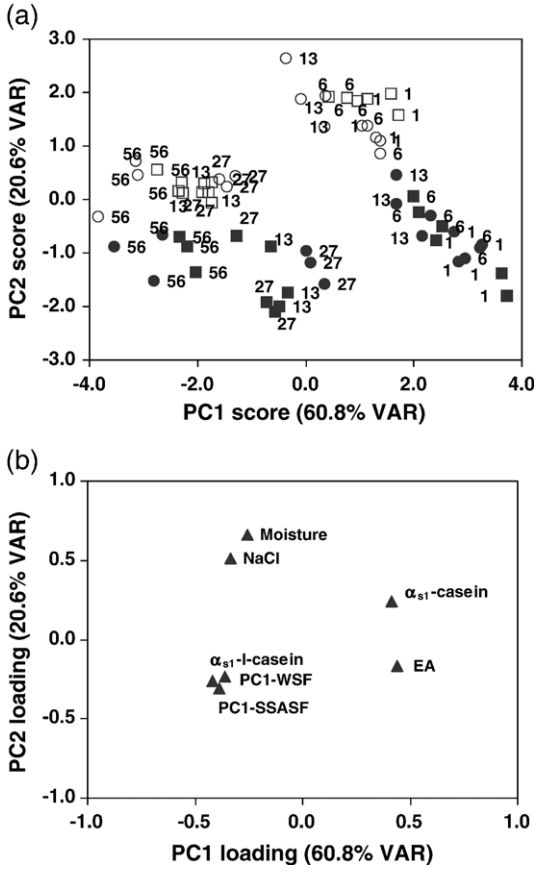

Fig. 1. Plots of the first two principal components (PC1 vs. PC2): (a) PC scores plot. (●) Cheeses C–zone I, (○) cheeses C–zone E, (■) cheeses F–zone I, (□) cheeses F–zone E. Numbers indicate the ripening time of the samples; (b) PC loadings plot.

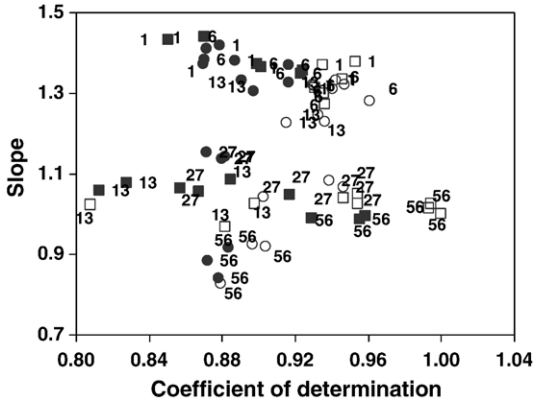

Fig. 2. Principal component similarity plot when cheese samples were analysed against FE56 (frozen cheese, external zone, 56 days of ripening) as a reference. (●) Cheeses C–zone I, (○) cheeses C–zone E, (■) cheeses F–zone I, (□) cheeses F–zone E. Numbers indicate the ripening time of the samples.

The PC scores and the PC loadings plots mapping cheese samples and variables in the two-dimensional space are shown in Fig. 1a and b, respectively.

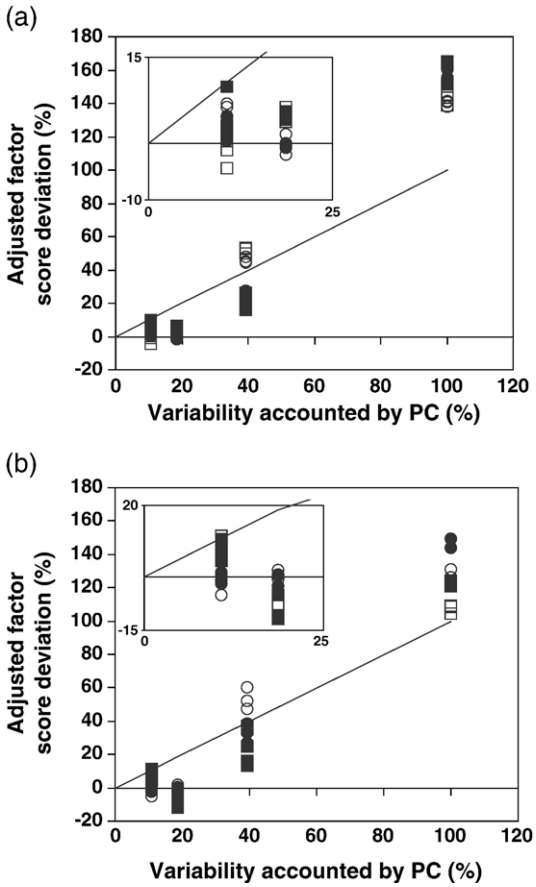

Fig. 3. Adjusted factor score deviations for cheeses of 1 and 13 days of ripening when cheese samples were analysed against FE56 (frozen cheese, external zone, 56 days of ripening) as a reference. Deviations from the reference lines are plotted as differences from the reference line; the data points are first, second, third and fourth adjusted factor scores from right to left: (a) 1 day of ripening, (b) 13 days of ripening. (●) Cheeses C–zone I, (○) cheeses C–zone E, (■) cheeses F–zone I, (□) cheeses F–zone E. Expanded area corresponds to third and fourth adjusted factor scores.

Cheese samples were grouped according to ripening time, sampling site and storage condition. Unripened control cheeses (from 1 to 13 days) and frozen cheeses (from 1 to 6 days) were grouped towards the right-hand side of Fig. 1a, while all ripened cheeses (27 and 56 days) and frozen cheeses of 13 days of ripening were grouped towards the left-hand side of Fig. 1a, showing an "early ripening" for frozen cheeses. In addition, cheeses sampled in the internal zone were grouped towards the bottom region of Fig. 1a, while cheeses sampled in the external zone were grouped towards the upper region of Fig. 1a.

Variables were separated into four groups: group 1 (α_{s1}-I-casein, PC1-WSF and PC1-SSASF), group 2 (α_{s1}-casein), group 3 (EA), and group 4 (moisture and NaCl) as shown in Fig. 1b.

To achieve a better understanding of the relationships between the groups and the original variables, the two-dimensional PC plots (Fig. 1a and b) and raw data (Table 1) were discussed. The variables PC1-WSF and PC1-SSASF (group 1) represented compounds that, with the exception of a few cases, increased during ripening, but with a higher rate in the cases of frozen cheeses [9,10]. However, α_{s1}-I-casein increased between 13 and 56 days during the ripening of control cheeses, but increased between 6 and 13 days and then dramatically decreased between 13 and 56 days during the ripening of frozen cheeses [10,18]. On the other hand, α_{s1}-casein (group 2) and EA (group 3) decreased during cheese ripening but EA also showed differences between sampling sites [9,17,18]. The freezing process significantly increased α_{s1}-casein hydrolysis, but not the decay rates of EA [10,18]. In particular, there was an earlier decrease in the α_{s1}-casein content during the ripening of frozen cheeses. That decrease was observed between 13 and 27 days in control cheeses, while it was detected between 6 and 13 days in frozen cheeses [10,18].

Moisture (group 4) showed differences between sampling sites but not according to ripening. On the other hand, NaCl (group 4), showed differences between both ripening time and

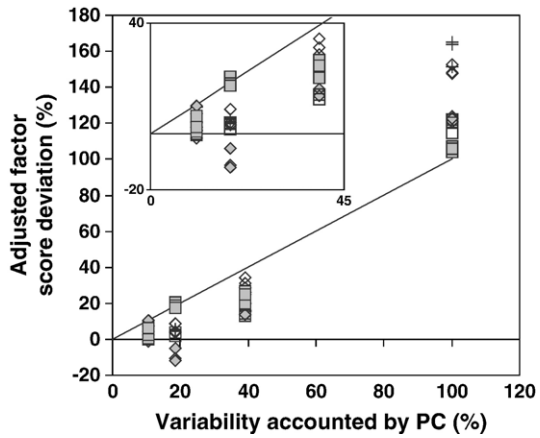

Fig. 4. Adjusted factor score deviations for internal zone of frozen cheeses were analysed against FE56 (frozen cheese, external zone, 56 days of ripening) as a reference. Deviations from the reference lines are plotted as differences from the reference line; the data points are first, second, third and fourth adjusted factor scores from right to left. (+) 1 day of ripening, (◊) 6 days of ripening, (◈) 13 days of ripening, (□) 27 days of ripening, and (▣) 56 days of ripening. Expanded area corresponds to second, third and fourth adjusted factor scores.

●1 ●1 ●6 ●6 ■1 ■1	●13 ■1 ■6	■13 ■13 ■13 ●27 ●27 ●27	■ 27 ■27
●1 ●6 ■6 ■6	●13 ●13	■ 27	●56 ■ 56
○1 ○1 ○1 ○6 ○6	○13 □6	○27 ○27 ○27 □27	■ 56 ■ 56 □ 56 □56 □56
□1 □1 □1 □6 □6	○6 ○13 ○13	□13 □13 □13 □27 □27	●56 ●56 ○56 ○56 ○56

Fig. 5. Distribution of the cheese samples into the 4×4 Kohonen map. (●) Cheeses C–zone I, (○) cheeses C–zone E, (■) cheeses F–zone I, (□) cheeses F–zone E. Numbers indicate the ripening time of the samples.

sampling site. However, the freezing process had affected neither moisture nor salt contents at the beginning of the ripening nor moisture and salt redistribution during the studied ripening period [10].

As a result, the grouping of the variables could be related to the predominant source of variability, namely, ripening time, sampling site or storage condition.

3.3. Principal component similarity

Four PCs were used for PCS computation and sample FE56 was used as reference. The result of PCS analysis is a plot of slope against coefficient of determination derived from the linear regression, PCS plot (Fig. 2).

The grouping of cheese samples in the PCS plot was similar to the PC1 vs. PC2 score plot, which is in agreement with the fact that two PCs accounted for more than 80% of the data variation. Adjusted factor score deviations from the reference line of some selected samples are shown in Figs. 3 and 4. Adjusted factor score deviations from the reference line for cheeses of 1 day of ripening showed that although the PC1 and PC2 contributed to the separation according to sampling site, PC2 is much more relevant (Fig. 3a). Adjusted factor score deviations from reference line for cheeses of 13 days of ripening demonstrated that PC2 contributed to the separation according to sampling site, while PC1, PC3 and PC4 contributed to the separation according to storage condition that indicates the early ripening of frozen cheeses (Fig. 3b). Adjusted factor score deviations from reference line for the internal zone of frozen cheeses showed that PC1 was the leading factor in the overall mapping of frozen cheeses according to ripening time (Fig. 4).

3.4. Kohonen self-organizing artificial neural network

In this work, the training was conducted in two phases: a relatively short phase with 100 epochs, a high learning rate (from 0.1 to 0.01), and a large neighbourhood (from 3 to 1); along with a second phase, with 1000 epochs, a low learning rate (throughout 0.01), and a zero neighbourhood.

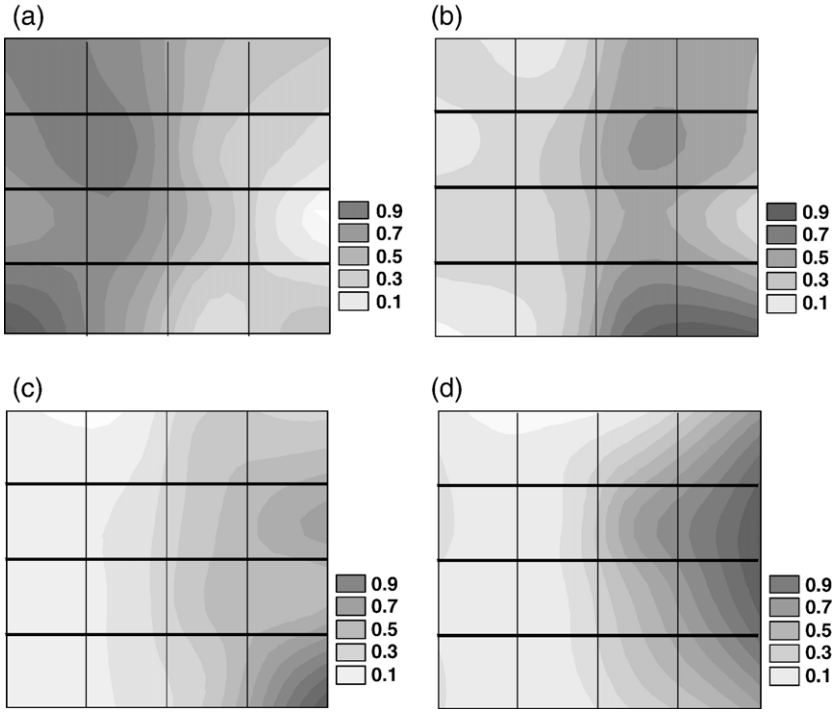

Fig. 6. Weight maps corresponding to the original variables related to proteolysis: (a) α_{s1}-casein, (b) α_{s1}-I-casein, (c) PC1-WSF, and (d) PC1-SSASF. Weight maps were presented as contour plots of the weights (in the interval [0, 1]) at the end of learning process.

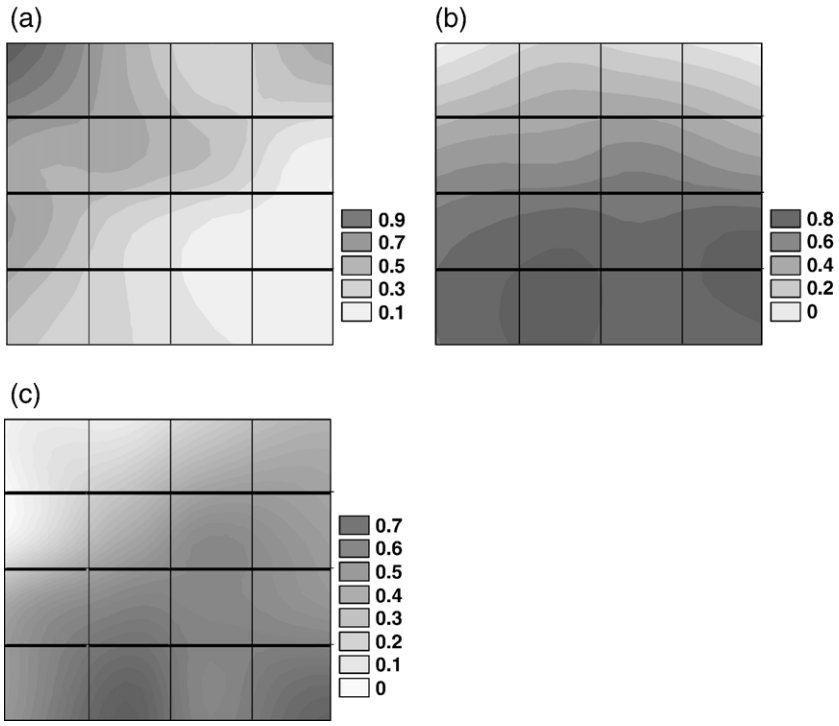

Fig. 7. Weight maps corresponding to the original variables: (a) EA, (b) moisture, and (c) NaCl. Weight maps were presented as contour plots of the weights (in the interval [0, 1]) at the end of the learning process.

Several architectures of Kohonen ANN were tested: 3×3, 4×4, 5×5, and 6×6. The sample grouping in the top map, except only in a few cases, was similar for all tested architectures (data not shown). The distribution of the cheese samples into the 4×4 Kohonen map is shown in Fig. 5. Unripened cheeses, control cheeses (from 1 to 13 days) and frozen cheeses (from 1 to 6 days), were grouped together on the left of Fig. 5, while ripened cheeses (27 and 56 days) and frozen cheeses of 13 days of ripening were grouped together on the right of Fig. 5, showing the "early ripening" pattern of frozen cheeses in agreement with PCA and PCS. Cheeses, except only in a few cases, were also separated in two subgroups corresponding to internal and external zones (upper and lower region, respectively).

The weight maps of the original variables are shown in Figs. 6 and 7. Weight maps were presented as contour plots of the weights (in the interval [0, 1]) at the end of the learning process and each weight map showed the contribution of one original variable to the distribution of cheese samples in the top map of Fig. 5.

Variables related to proteolysis, α_{s1}-casein, α_{s1}-I-casein, PC1-WSF and PC1-SSASF, are shown in Fig. 6. Variables that increased during cheese ripening like α_{s1}-I-casein, PC1-WSF, and PC1-SSASF; and the variable that decreases during ripening, α_{s1}-casein, showed left–right (increasing or decreasing, respectively) patterns related to the top map (Figs. 5 and 6). The weight map of α_{s1}-I-casein showed that higher weights are near the areas where frozen cheeses of 13 days of ripening are located, underlining that this variable was strongly related to the "early ripening" pattern of frozen cheeses (Figs. 5 and 6b). The weight maps of α_{s1}-I-casein and PC1-WSF showed that weights at the end of the studied ripening period (56 days) were higher in control cheeses than in frozen ones (Figs. 5 and 6b, c). No differences according to sampling sites are shown in the weight maps of α_{s1}-I-casein, PC1-WSF and PC1-SSASF.

Equilibrium modulus (Fig. 7a) showed differences according to ripening time (a left-right decreasing pattern) and sampling site (upper-bottom decreasing pattern). Moisture content shows a pattern related to differences between sampling sites, with higher weights in the lower region (Fig. 7b) that is the region in which cheese samples from the internal zone are grouped in Fig. 5. Consequently, moisture showed a close relation to the distribution of cheese samples in the top map according to sampling site.

Salt content (Fig. 7c) shows an almost homogeneous pattern on the right side, where ripened cheeses that have reached equilibrium in their NaCl content are located in Fig. 5, while on the left side, there is a gradient related to high NaCl values in the external zone and low NaCl values in the internal zone at the beginning of the ripening.

4. Conclusion

The implementation of data mining previous to cheese classification mapping was very useful because the input variables initially considered in this study were reduced from 36 to 7.

The two linear methods, PCA and PCS, showed the same grouping of cheese samples according to ripening time, sampling site and storage condition. These methods are closely related in their mathematical basis and the similar grouping showed by both methods can be related to the fact that the first three PCs explained 89.3% of the data set variation. However, adjusted factor score deviations plots obtained from PCS

analysis helped to explain the relationships between the groups and the original variables more easily than PCA.

The non-linear Kohonen ANN uses a mathematical procedure completely different from PCA; however, only slight differences were observed in the grouping of cheese samples. Those differences may be related to the weight that each model gives to every variable.

In addition, Kohonen ANN always maps objects and variables in a two-dimensional space, while in the case of PCA, dimensionality cannot always be reduced to two dimensions. Another interesting feature of Kohonen ANN is that weight maps (contour plots) sometimes are superior to PC loadings (vectors) for the understanding of relationships between the groups and the original variables. However, as the number of factor increases, weight patterns become more difficult to examine.

For successful classification or differentiation, analytical and statistical methods should be carefully selected and the quality of the data set must be high. In addition, Kohonen ANN has the potential to operate as a class-modelling device, provided an adequate number of samples is used for the training procedure and certain modifications are introduced into the algorithm [28].

Acknowledgements

This work was done with the financial support of Universidad Nacional del Litoral (Santa Fe, Argentina), Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina), Agencia Nacional de Promoción Científica y Tecnológica (Argentina) and Natural Science and Engineering Council of Canada.

References

[1] C.J. Coker, R.A. Crawford, K.A. Johnston, H. Singh, L.K. Creamer, Int. Dairy J. 15 (2005) 631–643.
[2] H. Pripp, L. Stepaniak, T. Sørhaug, Int. Dairy J. 10 (2000) 249–253.
[3] J. Zupan, M. Noviè, X. Li, J. Gasteiger, Anal. Chim. Acta 292 (1994) 219–234.
[4] D. González de Llano, M.C. Polo, M. Ramos, P. Martin-Alvarez, Lebensm.-Wiss. Technol. A 193 (1991) 529–532.
[5] M. Frau, J. Massanet, C. Roselló, S. Simal, J. Cañelas, Food Chem. 60 (1997) 651–657.
[6] C. Benfeldt, J. Sørensen, Int. Dairy J. 11 (2001) 567–574.
[7] R.A. Verdini, A.C. Rubiolo, J. Food Sci. 67 (2002) 963–967.
[8] R.A. Verdini, S.E. Zorrilla, A.C. Rubiolo, J. Food Sci. 67 (2002) 3264–3270.
[9] R.A. Verdini, S.E. Zorrilla, A.C. Rubiolo, Int. Dairy J. 14 (2004) 445–454.
[10] R.A. Verdini, S.E. Zorrilla, A.C. Rubiolo, Int. Dairy J. 15 (2005) 363–370.
[11] J.M. Poveda, L. Cabezas, P.L.H. McSweeney, Food Chem. 84 (2004) 213–218.
[12] Y. Vodovotz, G.E. Arteaga, S. Nakai, Food Res. Int. 26 (1993) 355–363.
[13] V. Furtula, S. Nakai, G.F. Amantea, L. Laleye, J. Food Sci. 59 (1994) 528–532 and 567.
[14] V. Furtula, S. Nakai, G.F. Amantea, L. Laleye, J. Food Sci. 59 (1994) 533–538.
[15] L. Pillonel, U. Bütikofer, H. Schlichtherle-Cerny, R. Tabacchi, J.O. Bosset, Int. Dairy J. 15 (2005) 557–562.
[16] D. Barile, J.D. Coisson, M. Arlorio, M. Rinaldi, Food Control 17 (2005) 197–206.
[17] R.A. Verdini, A.C. Rubiolo, J. Food Sci. 67 (2002) 1808–1813.
[18] R.A. Verdini, S.E. Zorrilla, A.C. Rubiolo, J. Texture Stud. 34 (2003) 331–346.
[19] P.F. Fox, J. Food Sci. 72 (1989) 1379–1400.
[20] C.N. Kuchroo, P.F. Fox, Milchwissenschaft 37 (1982) 331–335.
[21] M. Peleg, J. Rheol. 24 (1980) 451–463.
[22] M. Peleg, J. Food Sci. 44 (1979) 277–281.
[23] W.P. Gardiner, Statistical analysis methods for chemists, A Software-Based Approach, The Royal Society of Chemistry, Cambridge, UK, 1997.
[24] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice Hall International, London, UK, 1998.
[25] R. Kellner, J.M. Mermet, M. Otto, H.M. Widmer, Analytical Chemistry, Wiley-VHC, Weinheim, Germany, 1998.
[26] J. Zupan, Croat. Chem. Acta 75 (2002) 503–515.
[27] J. Zupan, M. Noviè, I. Ruisánchez, Chemometr. Intell. Lab. Syst. 38 (1997) 1–23.
[28] F. Marini, J. Zupan, A.L. Magri, Anal. Chim. Acta 544 (2005) 306–314.