

Use of high-performance liquid chromatographic–chemometric techniques to differentiate apple juices clarified by microfiltration and ultrafiltration

D. Blanco-Gomis^{*a}, P. Fernandez-Rubio^a, Ma. D. Gutiérrez-Alvarez^a and J. J. Mangas-Alonso^b

^a Departamento de Química Física y Analítica, Facultad de Química, Universidad de Oviedo, E-33006 Oviedo, Spain

^b Centro de Investigación Aplicada y Tecnología Agroalimentaria, E-33300 Villaviciosa, Spain

HPLC of amino acids and riboflavin in apple juices clarified by means of cross-flow membrane technology was used to characterise the juices. The chromatographic information was subjected to pattern recognition methods such as principal component analysis, K-nearest neighbour (KNN), linear discriminant analysis, Bayes analysis, soft independent modelling of class analogy and partial least squares.

Keywords: Amino acids; riboflavin; apple juice; high-performance liquid chromatography; chemometrics

Apples are amongst the most widely grown and widely consumed of temperate crops, taking second place only to grapes. The annual world apple crop is in the order of 40 million tons,¹ of which at least 25% is processed into juice.² The raw juice is nearly always turbid, brown and tends to sediment on storage. A clear juice can be obtained by the use of pectolytic enzymes, fining or filtration before bottling.³ Nowadays, inorganic membranes of γ -alumina– α -alumina and zirconia–carbon are preferred to microfiltration (MF) or ultrafiltration (UF) because of their resistance to corrosive cleaning and sterilisation chemicals, longer service life and ability to be autoclaved.⁴ Moreover, these techniques offer many advantages to the apple juice processor, since they combine both fining and filtration steps in one operation and are competitive in cost with respect to conventional fining and filtration.

The use of multivariate methods for analysing large amounts of data has become remarkably widespread over the last 30 years.⁵ Multivariate data analysis (MDA) can be divided into different parts, exploratory data analysis (EDA), factor analysis (FA), classification analysis (the first and second levels of pattern recognition) and correlation analysis (the third and fourth levels of pattern recognition); in addition, procedures for feature selection of relevant variables are also available within the aforementioned statistical techniques.⁶

By means of pattern cognition methods (exploratory techniques), we can visualise the information within the data matrix, select variables and carry out clustering of variables and observations; a typical exploratory technique used in the field of ecology, for instance, is correspondence analysis.⁷

On the other hand, factor analysis allows us to reduce the dimensionality of data set and noise, to detect data structuring and to ascertain the relationship between the original variables, factors and observations; pure display methods, such as principal component analysis (PCA) and non-linear mapping (NLM), are typically used to do this.⁸

Pattern recognition techniques have been employed for detecting the adulteration of fruit juices;⁹ factor analysis is used in the selection of the most discriminating parameters, so that

the adulteration of citrus fruit juices can be ascertained. Classification and correlation techniques have also been used for typifying distillates,^{10–14} almonds,¹⁵ wines,^{16,17} and biological samples.¹⁸

With the purpose of evaluating the nutritional quality of the apple juices obtained by MF and UF, we determined the sugar, organic acid, amino acid and riboflavin contents by means of high-performance liquid chromatography (HPLC) before and after the filtration processes. Neither the sugars nor the organic acids seem to change significantly as a consequence of the treatment by MF and UF. Conversely, significant changes in the contents of amino acids and riboflavin were observed. Consequently, these latter compounds were employed in the characterisation of apple juices obtained by MF and UF. The analytical data obtained were subjected to multivariate classification techniques and modelling techniques.

Experimental

Raw material: clarified apple juice

Apple juice was prepared from a mixture of apples with different sensory properties, endowing the resulting juice with an overall acidic nature. The juice was extracted in a pilot plant by milling the apples with a grating mill and pressing with a rack and frame press. Clarification of the apple juice was achieved using a back-pulsing cross-flow filtration technique. Two size-exclusion membrane types made of single-channel tubular modules of zirconium oxide (hydraulic diameter 6 mm; filtration area 0.16 m²) with a graphite support were employed, namely UF membranes with a nominal molecular mass cut-off (MMCO) of 50 000 Da and MF membranes of mean pore size 0.14 μ m. The UF and MF characteristics were as follows: pressure drop, 3.0×10^5 – 3.5×10^5 N m⁻²; superficial velocity, 5.0–6.5 m s⁻¹; and temperature, 20 and 50 °C.

HPLC analysis of amino acids and riboflavin

Amino acids and riboflavin standards and the other chemicals used were purchased from Sigma–Aldrich (Madrid, Spain) and the solvents (HPLC grade) from Romil (Loughborough, UK). Ultra-pure water was obtained through a Milli-Q system (Millipore, Milford, MA, USA).

The liquid chromatograph (LKB, Bromma, Sweden) was equipped with two Model 2150 pumps and a Model 2152 controller for the generation of elution gradients, a Rheodyne (Cotati, CA, USA) Model 7125 injection valve with a 20 μ l loop and a Shimadzu (Kyoto, Japan) RF-535 fluorescence detector fitted with a 12 μ l flow cell. Chromatograms were plotted and integrated by means of a Shimadzu C-R3A integrator.

For monitoring the changes in the major amino acids present in apple juices [aspartic acid (Asp), glutamic acid (Glu),

asparagine (Asn), serine (Ser), glutamine (Gln) and alanine (Ala)], we chose 6-aminoquinolyl-*N*-hydroxysuccinimidyl carbamate¹⁹ (Waters AccQ.Fluor reagent kit; Millipore) as a fluorescent derivatizing reagent. The procedure followed for amino acids derivatization was that recommended by the manufacturer.

Amino acid separations were carried out on a 25 cm × 4.6 mm id column packed with 5 µm Spherisorb ODS-2, which was kept at 38 °C throughout the experiments. Elution was performed in gradient mode by using two solvent mixtures. Solvent A consisted of 6 × 10⁻² M sodium acetate (pH 5.2) and solvent B was acetonitrile–water (60 + 40). The flow rate was 0.8 ml min⁻¹ and before being used both solvents were vacuum filtered and degassed with helium. The gradient programme used was as follows: a linear step from 10% to 15% B in 5 min, anisocratic elution step at 15% B for 5 min, a linear step to 35% B in 3 min and finally a linear step to 80% B in 7 min.

The AQC derivatives of amino acids were monitored at an excitation wavelength of 250 nm and an emission wavelength of 395 nm.

Under these conditions, strong overlap between asparagine and serine was observed, and consequently both amino acids were quantified together. The average variation in the retention times of the amino acids assayed <1% (RSD), the average recoveries ranged between 91% and 109% and the RSD for sample quantification (*n* = 10) was always <7%. Quantification was achieved by the external standard method.

To determine the riboflavin in apple juice, acid hydrolysis of the sample was necessary. Thus, 5 ml of apple juice were placed in a 10 ml calibrated flask and 0.5 ml of 1 M H₂SO₄ was added. The flask was immersed in a water-bath at 100 °C for 30 min. After cooling, the solution was adjusted to pH 5–6 with NaOH and levelled off with water to 10 ml, and 20 µl of this solution were then injected into the chromatographic system. The detection limit, based on a signal-to-noise ratio of 3 : 1, was 0.6 µg l⁻¹. Recovery studies showed good results (98.0–100%) and the RSD (*n* = 10) was 3.2%. Quantification was carried out by the external standard method.

If the amount of riboflavin was lower than the quantitation limit (five times the detection limit), the recommended method for sample preconcentration was as follows: the hydrolysed sample was filtered and, using a Visiprep vacuum manifold (Supelco, Bellefonte, PA, USA) together with a membrane pump (Vacuum Brand, Wertheim, Germany), an aliquot (5 ml) of the sample was passed through a C₁₈ cartridge, previously conditioned by passing 5 ml of methanol and 2 × 5 ml of water. The cartridge was washed with water and riboflavin was eluted using 2 × 0.5 ml of methanol. The eluate was evaporated under nitrogen until dry and the residue was reconstituted in 0.2 ml of water before being injected into the chromatograph.

Separations were carried out on a 150 × 3.9 mm id NovaPack C₁₈ column (4 µm). Analyses were carried out in the isocratic mode by using methanol–water (35 + 65) as eluent at room temperature (20 ± 2 °C) at a flow rate of 0.7 ml min⁻¹. The column effluent was monitored at a fluorescence wavelength of 522 nm (excitation wavelength 422 nm). The retention time of riboflavin was 3.6 min.

Results and discussion

Data

Table 1 presents the average concentrations of amino acids (aspartic, glutamic, glutamine, alanine and asparagine + serine) and riboflavin. Data were processed by means of the PARVUS statistical package.²⁰ We constructed a data matrix in which the rows (35) represented apple juices obtained by the two membrane techniques (UF and MF) at different temperatures (20 and 50 °C) and clarification times (1, 2.5 and 4 h) and the columns (six) corresponded to amino acids and riboflavin, so

that each object was a data vector of six variables. Data were categorized as category U, ultrafiltered apple juice, and category M, microfiltered apple juice. Eleven unclarified apple juices were used as the evaluation set.

Univariate treatment

Before multivariate analysis, univariate characterization of the apple juices was carried out on the basis of the Fisher weight. Riboflavin showed the highest value (1.47005) and glutamic acid showed the lowest (0.02374). In consequence, the most discriminatory variable (riboflavin) was used to explore its capacity for typifying both categories. As can be seen in Fig. 1, where two Box–Whisker diagrams are shown, differentiation of ultrafiltered and microfiltered apple juices is not possible using the variable with the greatest discriminatory power, so it was necessary to use multivariate techniques to achieve this objective.

Cluster analysis

Once the variables had been standardised, cluster analysis was carried out. A 35 × 35 matrix consisting of similarities of correlation coefficients where each observation was represented by a six-dimensional vector was used for hierarchical clustering on the basis of the average linkage-unweighted pair group method. As can be seen in Fig. 2, two clusters were found at a similarity level of 0.363; the first includes all ultrafiltered apple juices with clarification times <4 h; on the other hand, the

Table 1 Average concentrations of amino acids (µg ml⁻¹) and riboflavin (ng ml⁻¹)

Sample	Asp	Glu	Asn + Ser	Gln	Ala	Ribo
M _{20a}	22.03	11.07	40.27	1.94	2.14	18.95
M _{20b}	32.20	10.80	50.87	1.77	1.78	21.93
M _{20c}	20.53	8.70	23.73	1.16	1.57	35.97
M _{50a}	35.10	13.10	24.03	1.09	1.12	31.21
M _{50b}	39.33	16.00	54.43	1.94	1.25	27.32
M _{50c}	37.70	16.20	46.37	1.79	1.13	31.66
U _{20a}	41.10	15.80	67.67	1.08	1.30	1.90
U _{20b}	29.00	8.63	37.70	0.86	0.52	4.82
U _{20c}	18.83	11.17	54.97	1.64	0.58	6.22
U _{50a}	11.33	10.94	66.40	0.58	3.33	10.98
U _{50b}	14.72	11.96	81.03	0.73	6.07	24.04
U _{50c}	39.90	17.10	74.50	1.80	0.38	22.52

Asp, aspartic acid; Glu, glutamic acid; Asn + Ser, asparagine + serine; Gln, glutamine; Ala, alanine; Ribo, riboflavin; U, ultrafiltered; M, microfiltered; 20 and 50: process temperature (°C); a, b and c: clarification time, 1, 2.5 and 4 h, respectively.

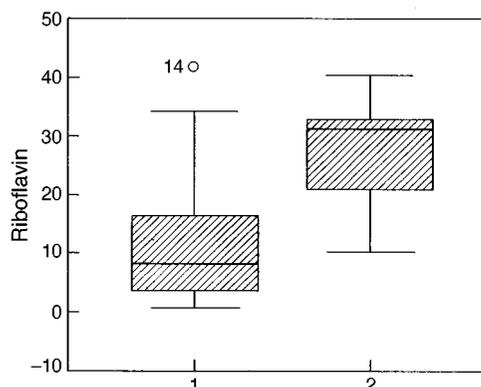


Fig. 1 Box–Whisker diagrams constructed for riboflavin. Category 1, ultrafiltered; category 2, microfiltered.

second cluster includes all the microfiltered apple juices. At the same time, at a similarity level of 0.454, the first cluster (including ultrafiltered apple juices) is split into two new clusters on the basis of treatment temperature, 20 and 50 °C. Finally, apple juices microfiltered at 20 °C with a clarification time of <4 h are grouped at a similarity level of 0.743, and those microfiltered at 50 °C with a clarification time of >1 h are included in a new cluster at a similarity level of 0.780.

Factor Analysis of the Internal Structure

The principal components technique was applied in order to define the structure of the data, and the number of principal components (PC) was determined. Three PCs that accounted for 84.05% of the variance were chosen on the basis of Kaiser's criterion (eigenvalues > 1.0 are chosen). The search for a latent structure is carried out by means of rotations, which may be orthogonal or oblique. In our case, an orthogonal rotation (raw varimax) was used. In Fig. 3, apple juices (microfiltered and ultrafiltered) are represented on the plane formed by the first and third varivectors; as can be seen, a suitable separation of the two categories is obtained. In general, the scores for microfiltered apple juices are the lowest for the third rotated component.

When the original variables are projected on to the two-dimensional structure of the first and third varivectors (Fig. 4), close linking was detected between riboflavin and the third

varivector, and between alanine, aspartic acid and glutamine and the first varivector.

Classification: linear discriminant analysis (LDA) and K nearest neighbour (KNN) methods

Linear discriminant analysis

A pooled variance-covariance matrix was used, so that variance-covariance matrices for both groups are assumed to be not significantly different. The classification and prediction matrices are shown in Table 2, three groups for cancellation having been established for validating the discriminant method. As can be seen, the classification and prediction capacities were 100%.

In order to ascertain the relevant variables for classification purposes, a stepwise linear discriminant analysis, based on the minimisation of the Wilks' lambda, was also employed, taking into account an *F*-to-enter of 3.84 and an *F*-to-remove of 2.71. The most relevant variables found were riboflavin, asparagine + serine, glutamine and alanine; the proportion of the total variance in the discriminant scores not explained by differences among groups was 18.14%. Classification and prediction matrices (using three groups for cancellation) found when these variables were considered in LDA analysis show an adequate prediction capacity (97.14% correct predictions) and an excellent classification capacity (100% hits).

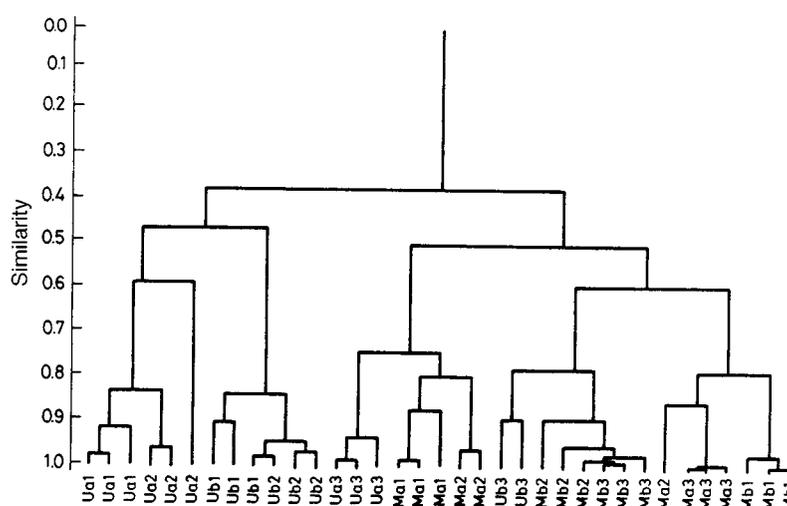


Fig. 2 Dendrogram obtained by cluster analysis. $U_{T,t}$, ultrafiltered; $M_{T,t}$, microfiltered; *T*, temperature of process (a, 20 °C; b, 50 °C); *t*, clarification time (1, 1 h; 2, 2.5 h; 3, 4 h).

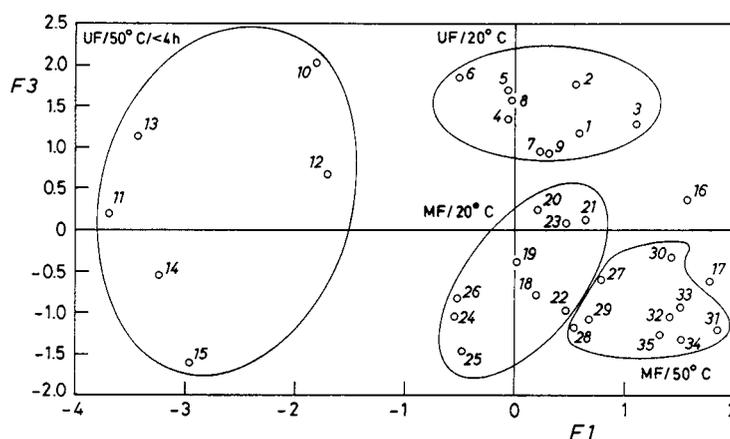


Fig. 3 Eigenvector projections of apple juices.

KNN method

This technique is a non-parametric classification procedure. The classification matrices for $K = 3, 5$ and 8 are shown in Table 3. In general, small values of K are preferred since, as can be seen in Table 3, an increase in the number of nearest samples selected promotes a decrease in the method's classification capacity. However, the number of correct classifications obtained with this method is lower than when the discriminant technique is employed using the most relevant variables.

Modelling: the bayes, soft independent modelling of class analogy (SIMCA) and partial least-squares (PLS) methods

Bayes analysis

In this modelling technique, a variance-covariance matrix is evaluated for each class. The classification and prediction matrices, using four cancellation groups for validating, are shown in Table 4; as can be seen, the classification and prediction capacities are very high.

At the same time, probabilities of first-class error (α) and second-class error (β) for both classes were also estimated. In the ultrafiltered model, values of $\alpha = 0$ and $\beta = 0.22$ were obtained, demonstrating that this category had a high sensitivity (all ultrafiltered apple juices being included in the ultrafiltered model) and a lower specificity, since 22% of microfiltered apple juices were accepted in the ultrafiltered box, as shown in the Coomans' diagram (Fig. 5). On the other hand, the microfiltered

model showed a very high sensitivity ($\alpha = 0$) and specificity ($\beta = 0$), all microfiltered apple juices being included in the microfiltered box along with any ultrafiltered apple juice which was also included in this model.

Finally, when 11 unclarified apple juices were projected on to the Coomans' diagram, 72.7% of them were considered as outliers (Fig. 5). In consequence, the models established could be considered to discriminate reasonably the unclarified apple juices from those processed by means of membrane technology.

SIMCA method

This modelling technique considers each category separately. The modelling properties of PCA are used to define a SIMCA box for each class.

A reduced model with three principal components for each category with a percentage retained variance of 87.07% for the ultrafiltered model and 95.32% for the microfiltered model was established. The first-class and second-class errors found for the two models were as follows: model U, $\alpha = 0.058$ and $\beta = 0.166$; and Model M, $\alpha = 0.11$ and $\beta = 0$. The sensitivity estimated from the SIMCA method for both models is lower than that with the Bayes analysis; on the other hand, the specificity obtained for the ultrafiltered model with the SIMCA analysis is higher than that with the Bayes modelling technique.

At the same time, as is shown in Fig. 6, all unclarified apple juices were considered as outliers of the two SIMCA boxes.

Partial least squares

The PLS method was also used in order to distinguish between the apple juices processed through membrane technology. According to the chemometric studies developed by Ortiz

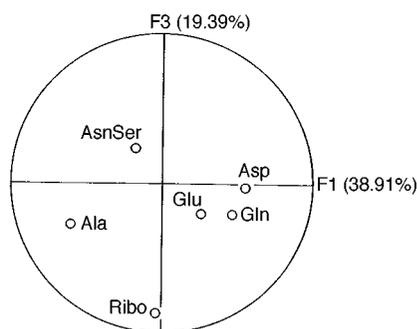


Fig. 4 Eigenvector projections of original variables. Ribo, riboflavin; Gln, glutamine; Asp, aspartic acid; Glu, glutamic acid; Ala, alanine; AsnSer, asparagine + serine.

Table 2 LDA method: classification and prediction matrices (three cancellation groups)

True category	Classification assigned category			Prediction assigned category		
	U	M	Hits (%)	U	M	Hits (%)
Ultrafiltered (U)	17	0	100	17	0	100
Microfiltered (M)	0	18	100	0	18	100
Overall			100			100

Table 3 Classification matrices for KNN method

True category	$K = 3$			$K = 5$			$K = 8$		
	U	M	Hits (%)	U	M	Hits (%)	U	M	Hits (%)
Ultrafiltered (U)	15	2	88.2	12	5	70.6	10	7	58.8
Microfiltered (M)	0	18	100	0	18	100	0	18	100
Overall			94.3			85.7			80

Table 4 Bayes analysis: classification and prediction matrices (four cancellation groups)

True category	Classification assigned category			Prediction assigned category		
	U	M	Hits (%)	U	M	Hits (%)
Ultrafiltered (U)	51	0	100	17	0	100
Microfiltered (M)	0	54	100	3	15	83.3
Overall			100			91.4

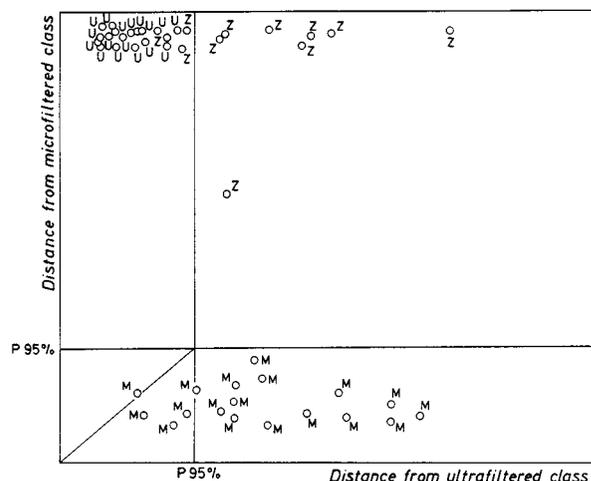


Fig. 5 Coomans' diagram for Bayes analysis. Z, unclarified apple juices.

*et al.*¹⁴ for typifying alcoholic distillates, we define the binary response variable $Y_{\text{ultrafiltered}} = 2$ and $Y_{\text{microfiltered}} = 1$, PLS regression then being carried out between the new variable (Y) and the original variables (amino acids plus riboflavin). The model constructed using the PLS technique consisted of four latent variables estimated by cross-validation with three deletion groups; the percentages of cross-validation explained variance and explained variance were 80.31 and 81.85%, respectively, and the calculated correlation coefficient (r^2) was 84%. Fig. 7 shows the calculated values for the Y variable by means of Box-Whisker plots. As can be seen, both classes are well characterised since different scores are obtained from ultrafiltered and microfiltered apple juices.

Conclusion

The analytical methodology described, amino acid and riboflavin analysis and chemometric techniques such as grouping, classification and modelling methods, have allowed us to differentiate correctly apple juices processed by membrane technology, UF and MF, in addition to unclarified apple juices.

Univariate analysis, using the most discriminant variable (riboflavin), did not allow us to separate the two categories satisfactorily.

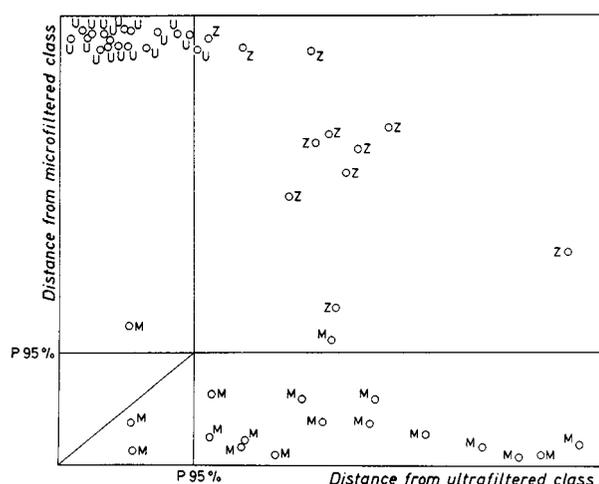


Fig. 6 Coomans' diagram for SIMCA analysis. Z, unclarified apple juices.

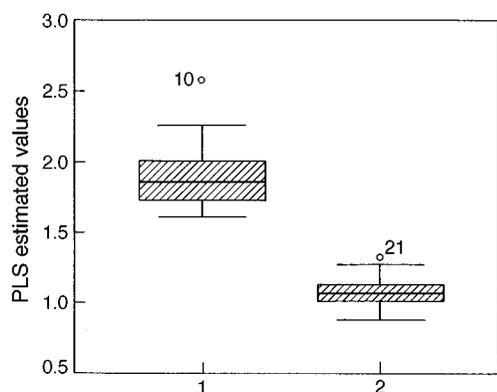


Fig. 7 Box-Whisker plots constructed for PLS-calculated values. Category 1, ultrafiltered; category 2, microfiltered.

Cluster analysis revealed the occurrence of the groups of the samples on the basis of the clarification technology, the temperature of the process and the clarification time. PCA and orthogonal rotation allowed us to establish three PCs that account for the major percentage of variance, a latent structure of the system having been detected; riboflavin was linked to the third varivector and alanine, aspartic acid and glutamine were closely related to the first varivector.

Stepwise linear discriminant analysis permitted us to ascertain the most relevant variables, an excellent classification of both classes being obtained.

At the same time, modelling techniques such as Bayes analysis and SIMCA allowed us to obtain models which are sufficiently sensitive and specific for typifying apple juices clarified by membrane technology, the use of multivariate regression techniques (PLS) providing an adequate model for characterising the apple juices.

References

- 1 Lea, A. G. H., in *Production and Packaging of Non-carbonated Fruit Juices and Fruit Beverages*, ed. Ashurst, P. R., Chapman and Hall, Glasgow, 1995, ch. 6.
- 2 Binnig, R., and Possmann, P., *Fruit Juice Processing Technology*, Agscience, Auburndale, FL, 1993, ch. 8.
- 3 van Buren, J. P., in *Processed Apple Products*, ed. Downing, D. L., Van Nostrand Reinhold, New York, 1989, ch. 5.
- 4 Mir, L., Michaels, S. L., Goel, V., and Kaiser, R., in *Membrane Handbook*, ed. Winston Ho, W. S., and Sirkar, K. K., Van Nostrand Reinhold, New York, 1992, ch. 35.
- 5 Resurreccion, A. V. A., *Food Technol.*, 1988, **42**, 128.
- 6 Meloun, M., Militky, J., and Forina, M., in *Chemometrics for Analytical Chemistry*, ed. Masson, M., Ellis Horwood, Chichester, 1992, ch. 5.
- 7 Greenacre, M., and Hastie, T., *J. Am. Stat. Assoc.*, 1987, **82**, 437.
- 8 Derde, M. P., Massart, D. L., Ooghe, W., and De Waele, A., *J. Autom. Chem.*, 1983, **5**, 136.
- 9 Brown, M. B., Katz, B. P., and Cohen, E., in *Adulteration of Fruit Juice Beverages*, ed. Nagy, S., Attaway, J. A., and Rhodes, M. E., Marcel Dekker, New York, 1988, ch. 4.
- 10 Mangas, J. J., Rodríguez, R., Moreno, J., Suárez, B., and Blanco, D., *J. Agric. Food Chem.*, 1996, **44**, 3303.
- 11 Mangas, J. J., Rodríguez, R., Moreno, J., and Blanco, D., *J. Agric. Food Chem.*, 1996, **44**, 268.
- 12 Mangas, J. J., Rodríguez, R., Moreno, J., and Blanco, D., *Lebensm.-Wiss. Technol.*, 1996, **29**, 357.
- 13 Mangas, J. J., Rodríguez, R., Moreno, J., Suárez, B., and Blanco, D., *J. Agric. Food Chem.*, 1997, **45**, 4076.
- 14 Cruz Ortiz, M., Saez, J. A., and López Palacios, J., *Analyst*, 1993, **118**, 801.
- 15 García-López, C., Grané-Teruel, N., Berenguer-Navarro, J., García-García, E., and Martín-Carratalá, M. L., *J. Agric. Food Chem.*, 1996, **44**, 1751.
- 16 Pueyo, E., Martín-Alvarez, P., and Polo, C., *Am. J. Enol. Vitic.*, 1995, **46**, 518.
- 17 Etiévant, P., Schlich, P., Cantagrel, R., Bertrand, A., and Bouvier, J. C., *J. Sci. Food Agric.*, 1989, **46**, 421.
- 18 Wold, S., Albano, C., Dunn, W. J., III, Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., and Sjöström, M., in *Chemometrics, Mathematics and Statistics in Chemistry*, ed. Kowalski, B. R., Reidel, Dordrecht, 1983, pp. 17-95.
- 19 Cohen, S. A., and Michaud, D. P., *Anal. Biochem.*, 1993, **211**, 279.
- 20 Forina, M., Leardi, R., Armanino, C., and Lanteri, S., *PARVUS. An Extendable Package of Programs for Data Exploration, Classification and Correlation*, Elsevier, Amsterdam, 1988.

Paper 7/04081D
Received June 11, 1997
Accepted September 30, 1997