

## Characterisation of apple cider cultivars by chemometric techniques using data from high-performance liquid chromatography and flow-injection analysis

D. Blanco-Gomis<sup>a</sup>, I. Herrero-Sánchez<sup>a</sup> and J. J. Mangas Alonso<sup>b</sup>

<sup>a</sup> Departamento de Química Física y Analítica, Facultad de Química, Universidad de Oviedo, E-33006 Oviedo, Spain

<sup>b</sup> Centro de Investigación Aplicada y Tecnología Agroalimentaria, E-33300 Villaviciosa, Spain

**Analytical techniques (HPLC and flow-injection analysis) for determining sugars, organic acids, polyphenols and pectins in apples, were employed along with chemometrics in the ripening and classification studies of cider apples. The use of principal component analysis allowed the authors to reduce the dimensionality of the data matrix; three new variables were obtained that accounted for 76% of variance. The projection of the apple cultivars in the reduced space allowed us to visualize the data structure on the basis of the degree of ripening and technological characteristics of the cider apple varieties monitored. Linear discriminant analysis computed a canonical variable with a prediction capacity of 93%, using three groups for cancellation in order to validate the method. The use of modelling techniques, such as SIMCA and partial least squares made an adequate grouping of apple cultivars feasible on the basis of their degree of ripening.**

**Keywords:** High-performance liquid chromatography; flow-injection analysis; chemometrics; ripening; apples

Cider is one of the most popular drinks all around the world, and especially in France, the United Kingdom, Germany and Spain.

In these countries, the apple crop and its subsequent transformation in order to obtain derivatives (brandy, vinegar, apple juice, *etc.*), is of enormous commercial, economic as well as social relevance. Therefore, any study related to the improvement of the conditions of production and more profitable farming practice, is of real importance and utility.

During the apple growing process, different biochemical transformations are produced. These changes exert a special influence on the contents of different compounds such as sugars, acids, polyphenols and pectins.<sup>1–10</sup>

The importance, incidence and evolution of these compounds in the growing and ripening process have been under constant study in recent years.<sup>11–15</sup>

Due to the high number of variables which are used in fruit ripening studies,<sup>16–18</sup> mathematical multivariate methods are needed to help us analyse and interpret the information that is generated.<sup>19–21</sup>

Current technological advances, especially in the field of microelectronics, have led to the manufacture of faster computers which simplify mathematical calculations, thus enabling the development of chemometrics as a multivariate analysis technique.<sup>22,23</sup>

In the present study, pattern recognition techniques were employed to establish discrimination rules that enable us to differentiate apples into technological and ripening groups from the values of certain chemical descriptors.<sup>24</sup> These parameters were selected according to their relevance in the ripening process as well as for their capacity as quality indicators.

All these studies will allow the cider producer to have more control over the raw material and to obtain higher quality products.

### Experimental

#### Samples

Traditional cider in Spain is made from the juice of a mixture of cider apples which have different sensory properties, the result being a natural, acidic juice.

The apple varieties used in this study were grown in the Agricultural Experiment Station orchards in Villaviciosa (Asturias, North of Spain), a village which is well known for the excellent quality of its cider apples and derivatives. Four different types of apples, namely Collaos (mild sharp), Durón Arroes and Picona Rayada (sweet), Meana (bittersharp) and Raxao (sharp), were chosen in this study for their agronomical and technological suitability. Twenty-eight samples of these cider apple varieties were taken at different moments of their ripening process.

#### Sample preparation

The apple extracts were prepared using Richmond's method<sup>25</sup> with some modifications<sup>26</sup> so as to ensure the complete extraction of the apple compounds. In order to achieve the said goal, 1 kg of apples were randomly chosen, of which a quarter was mashed into pieces and introduced into a mixer. The mixture was then covered with an adequate volume of ethanol and hydrochloric acid, reaching a final concentration of about 80% (m/m) and 0.1% (v/v), respectively. The fruit was crushed and vigorously mixed with these solvents for 2 min. The mush obtained was flow-back shaken in an inert atmosphere (nitrogen) for 2 h. The extract was subsequently filtered through a cellulose filter (Whatman, Maidstone, UK) and was washed with 80% (v/v) ethanol. The eluate was then reduced to a final volume of approximately 250 ml in a rotary evaporator at 37 °C. After this, the concentrate was diluted with 80% (v/v) ethanol to a final volume of 500 ml. Finally, the apple extracts were filtered through a 0.45 µm filter (Millex, Millipore Watford, UK) before analysis.

#### Analytical procedures

##### Sugars

Fructose, glucose, sucrose and sorbitol were determined according to the method described by Blanco *et al.*<sup>26</sup> by high-performance liquid chromatography using a Sugar Pak-I column (300 mm × 6.0 mm id, 10 µm) and an aqueous solution of 50 µg ml<sup>-1</sup> of calcium salt of EDTA as mobile phase; flow rate, 0.5 ml min<sup>-1</sup>; temperature, 90 °C. A refractive index detector was used as the detection system.

### Malic acid

This component was analysed using the following chromatographic method optimised by Blanco *et al.*:<sup>27</sup> Spherisorb column (ODS-2 250 mm × 4.5 mm id, 3 μm); mobile phase, phosphate buffer, pH 2.25 and 10<sup>-2</sup> M ionic strength; flow rate, 0.5 ml min<sup>-1</sup>; temperature, 36 °C and UV detection at 206 nm.

### Polyphenols

An FIA (flow injection analysis) system was used in the determination of polyphenols according to the method described by Mangas *et al.*,<sup>28</sup> the following configuration being used: carrier, Folin Ciocalteu reagent (1/100) in a 1% Na<sub>2</sub>CO<sub>3</sub> solution; carrier flow rate, 0.2 ml min<sup>-1</sup>; reagent, 1% Na<sub>2</sub>CO<sub>3</sub> solution; reagent flow, 0.6 ml min<sup>-1</sup>; channel dimension, 300 × 0.5 mm, 300 × 0.7 mm; process temperature, 25 °C; colorimetric detection at 673 nm; injection volume, 40 μl.

### Pectins

The different pectin fractions [water-soluble pectins (WSP), chelator-soluble pectins (CSP) and hydrochloric acid-soluble pectins (HASP)], were estimated as total galacturonic acid by incubation with sulfuric acid and subsequent colour development with alkaline *m*-hydroxydiphenyl.<sup>29</sup>

### Reagents

All the reagents, solvents and standards that were used were of analytical quality (99% minimal purity) and were supplied by Fluka (Buchs, Switzerland), Merck (Darmstadt, Germany) and Sigma-Aldrich (Madrid, Spain).

## Results and discussion

### Data processing

The data was processed by means of the PARVUS statistical package.<sup>30</sup> A matrix was constructed with rows (28) representing apple samples and columns (9) corresponding to sugars (sucrose, glucose, fructose), sorbitol, malic acid, total polyphenols and pectins (water-soluble pectin, chelator-soluble pectin and hydrochloric acid-soluble pectin). Samples were categorised as Ripe (R, 9) and Unripe (U, 19) according to the starch-iodine test criterion.<sup>31</sup> The data were subjected to autoscaling before statistical analysis.

### Univariate analysis

Table 1 shows the concentration of the variables monitored during the ripening of the apples. Univariate characterisation of the apples was carried out on the basis of Fisher's weight (FW). Fructose and water-soluble pectin were the two most discriminant variables ( $FW_{\text{fructose}} : 1.31$ ;  $FW_{\text{water-soluble pectin}} : 1.02$ ). However, the use of the most discriminant variables, fructose, did not allow us to differentiate between these two categories (see Fig. 1), so multivariate analysis was needed.

### Factor analysis

A principal component analysis (PCA) was performed in order to establish the relationship between variables and observations, as well as to recognise the data structure.

The application of this display method meant that we were able to observe the data structure using three principal components which accounted for 75.7% of the variance. An oblique axis rotation facilitates the interpretation of the data. The non-orthogonal varivectors were obtained from a linear

**Table 1** Concentration (g kg<sup>-1</sup>) of sugars, total polyphenols, pectin fractions and malic acid in apple cultivars

Apple	Class	S	G	F	So	TP	WSP	CSP	HASP	M
C1	U	9.9	19.7	46.0	4.57	3.46	0.11	0.24	3.00	3.40
C2	U	8.3	18.8	42.9	3.51	2.89	0.06	0.10	3.48	2.45
C3	U	13.9	17.1	45.1	3.38	2.97	0.08	0.14	5.64	3.42
C4	U	15.4	16.7	43.9	3.06	2.66	0.06	0.15	4.40	2.61
C5	R	15.0	18.0	44.0	3.56	2.80	0.09	0.19	4.31	1.81
C6	R	18.7	19.1	46.8	3.93	3.25	0.28	0.46	4.13	1.28
D1	U	7.3	13.5	33.3	4.41	3.23	0.03	0.17	4.51	3.96
D2	U	11.7	12.5	35.6	4.53	3.15	0.03	0.10	2.92	2.78
D3	U	11.7	12.0	36.3	13.99	2.60	0.03	0.12	3.17	2.70
D4	U	12.7	13.4	34.8	4.38	2.75	0.03	0.08	2.45	3.72
D5	U	18.3	15.5	38.3	6.89	1.87	0.04	0.20	5.08	2.31
D6	U	24.0	15.1	36.2	4.87	2.66	0.04	0.18	5.52	2.51
D7	U	30.1	17.4	41.1	7.31	3.13	0.08	0.20	5.91	1.71
M1	U	21.4	6.5	41.3	4.49	4.16	0.08	0.14	3.90	5.65
M2	U	27.2	11.9	49.7	3.02	4.22	0.04	0.08	3.81	4.90
M3	U	26.6	10.0	47.8	2.61	3.69	0.06	0.17	4.17	3.53
M4	U	33.0	6.4	47.7	3.02	3.39	0.06	0.19	4.18	4.50
M5	R	29.6	11.0	56.6	3.65	4.15	1.11	1.26	7.28	3.31
P1	U	9.1	17.8	56.2	5.44	3.30	0.27	0.30	16.09	1.74
P2	U	8.0	21.7	49.8	2.43	1.92	0.15	0.43	18.42	4.60
P3	R	15.8	20.6	69.0	5.00	2.82	0.18	0.24	13.60	2.43
P4	R	16.1	20.1	75.0	5.73	3.32	0.91	0.62	18.18	2.14
P5	R	9.5	15.9	56.7	5.53	3.38	1.29	0.55	12.54	1.72
R1	U	7.4	26.5	50.0	3.86	2.59	0.13	0.23	8.43	5.37
R2	U	10.7	22.5	48.1	3.59	2.40	0.15	0.30	13.47	3.62
R3	R	15.1	22.4	63.0	6.38	3.03	0.11	0.10	6.97	2.14
R4	R	11.4	20.9	49.5	3.29	2.17	0.25	0.57	8.53	5.66
R5	R	9.4	20.3	50.0	3.58	2.01	0.49	0.58	7.48	3.41

Abbreviations used for variables: S, sucrose; G, glucose; F, fructose; So, sorbitol; TP, total polyphenols; WSP, water-soluble pectin; CSP, chelator-soluble pectin; HASP, hydrochloric acid-soluble pectin; M, malic acid. Abbreviations used for apple samples: R, Raxao; M, Meana; P, Picona Rayada; C, Collaos; D, Durón Arroes. The subindex shows the degree of ripening within each variety, the highest subindex corresponding to the ripest apples. U, unripe class; R, ripe class.

transformation of the original orthogonal factors. As can be seen in Fig. 2, where samples and variables are projected onto the plane of the first two factors, the said factors differentiate the apple varieties on the basis of their degree of ripening. Thus, the ripest apples (higher subindex) have, in general, smaller scores for the first factor than unripe ones.

The variables most correlated with the first factor were fructose and pectins. Consequently, we can establish that the processes of fructose accumulation and apple softening are very important stages in fruit senescence when determining the degree of ripening. We can also see in this representation how the samples are reasonably well structured in variety groups.

At the same time, the technological characteristics of the apples are visualised when the observations are projected onto the map formed by the second and third variectors. As can be seen in Fig. 3, bittersharp apple varieties, such as Meana, had lower scores for the second component and negative values for the third variector. On the other hand, more acidic apples, such as Raxao, had positive scores for the second principal component and negative values for the third variector. Sweet and mild sharp cultivars, such as Picon Rayada, Durón Arroes and Collaos, did not exhibit a differentiated distribution when they were projected onto this map (Fig. 3). In this figure, it can also be seen that the second variector is close to phenolic compounds (negative correlation) whereas malic acid is correlated with the third variector (negative correlation).

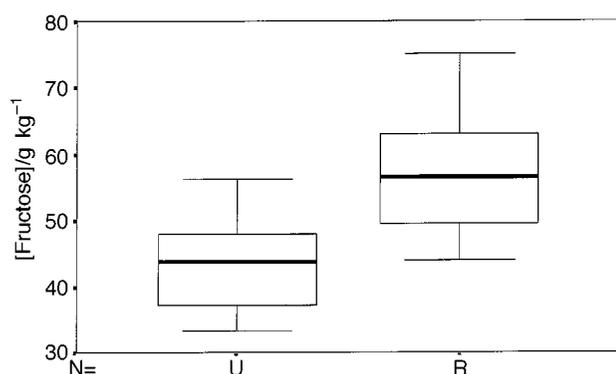


Fig. 1 Box-Whisker plots for the fructose variable. U: unripe class; R: ripe class.

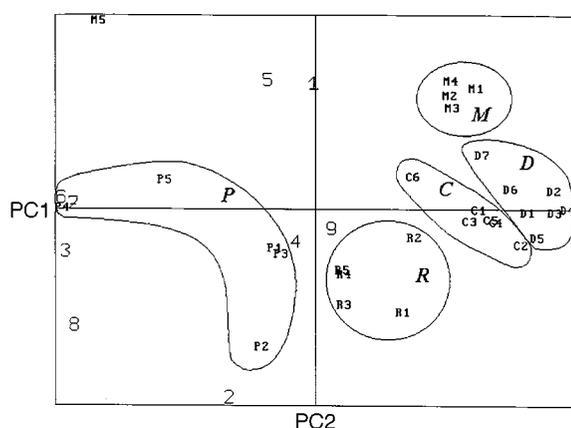


Fig. 2 Projection of the variables and samples onto the plane formed by the two first factors. R, Raxao; M, Meana; P, Picon Rayada; C, Collaos; D, Durón Arroes. The subindex shows the degree of ripening within each variety, the highest subindex corresponding to the ripest apples. 1, Sucrose; 2, glucose; 3, fructose; 4, sorbitol; 5, total polyphenols; 6, water-soluble pectin; 7, chelator-soluble pectin; 8, hydrochloric acid-soluble pectin; and 9, malic acid.

Hence, we may state that PCA has allowed us to visualise the data structure, since apple cultivars were reasonably well grouped in accord with their technological characteristics and degree of ripening.

### Linear discriminant analysis (LDA)

LDA is a supervised method that can be used for visualising the data contained in complex data-bases. Mathematical decision rules obtained from a training set are used to classify unknown samples. As can be seen in Table 2, 92.9% of correct predictions were obtained using an internal cross-validation method with three groups for cancellation.

A basic problem in LDA is deciding which variables should be included in the analysis. This may be achieved using Wilks' lambda ( $\lambda$ ) as selection criterion, and an  $F$  statistic to determine the significance of the changes in lambda when a new variable is tested. Generally, values of 3.84 as  $F$ -to-enter and 2.71 as  $F$ -to-remove are used, which correspond to a confidence level of 90%.

In accord with the results obtained in the univariate treatment, fructose and water-soluble pectin were the most relevant variables. The Wilks' lambda value obtained was 0.49, which means that 51% of total variance is explained by within-group differences; 82.1% of correct predictions were obtained using these two variables to perform the cross-validated LDA method.

### Soft independent modelling of class analogy (SIMCA)

The SIMCA technique is used for constructing an enclosure for each category using a principal component model. A reduced model with four significant principal components, obtained from a single-cross full validation, was employed, which accounted for 88.2% of the variance for the unripe model, and 95.8% of the variance for the ripe model. The first and second class errors ( $\alpha$  and  $\beta$ ), related to model sensitivity ( $1-\alpha$ ) and specificity ( $1-\beta$ ), were also taken into account. As can be seen in Table 3, the unripe class is more sensitive and specific than

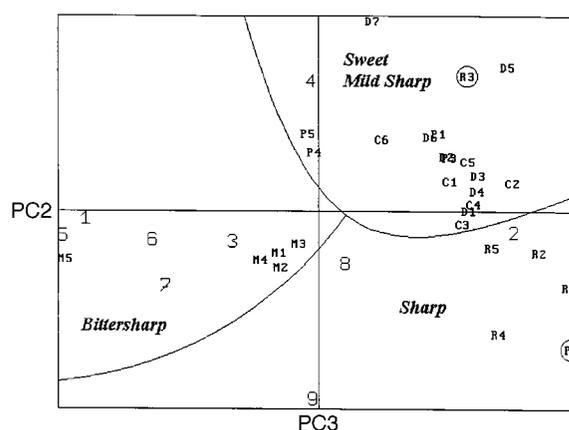


Fig. 3 Projection of the variables and samples onto the plane formed by the second and third factors. Abbreviations used, see Fig. 2.

Table 2 Prediction matrix for the LDA method (three groups for cancellation)

True Category	Assigned Category		
	U	R	Hits (%)
U	18	1	94.7
R	1	8	88.9
Overall			92.9

the ripe class. These results can be easily seen on a Coomans' plot. As is shown in Fig. 4, three samples (marked with an arrow) are incorrectly classified (belonging to the unripe class); furthermore three outliers (marked with an asterisk) were detected, one of them being the misclassified, unripe sample.

### Partial least squares (PLS)

The determination of the mathematical relationship between two variables groups (predictor variables and criterion variables), is carried out using multivariate regression techniques. In particular, the PLS method is specially recommended when the number of observations is small in relation to the number of variables.

We established a binary answer (assigning value 1 to the unripe category and 2 to the ripe category), and then carried out a multivariate regression between this criterion variable (ripening) and the predictor variables (fructose, sorbitol, glucose, sucrose, malic acid, polyphenols, water-soluble pectin, chelator-soluble pectin, and hydrochloric acid-soluble pectin).

The model constructed using the PLS regression consisted of four latent variables estimated by cross-validation with three groups for cancellation. The percentages of cross-validated explained variance, explained variance and correlation coefficient were 55, 66 and 71%, respectively. Fig. 5 shows two Box-Whisker plots for each category using the PLS estimated value. As we can see, the PLS technique enables correct discrimination of apple varieties based on their degree of ripening.

### Conclusions

The use of factor analysis and classification and modelling methods to analyse the data obtained by means of the analytical techniques of HPLC and FIA have enabled us to discriminate between apples on the basis of their technological characteristics and their state of ripening. Whereas fructose and pectins were found to be the most significant variables in the apple

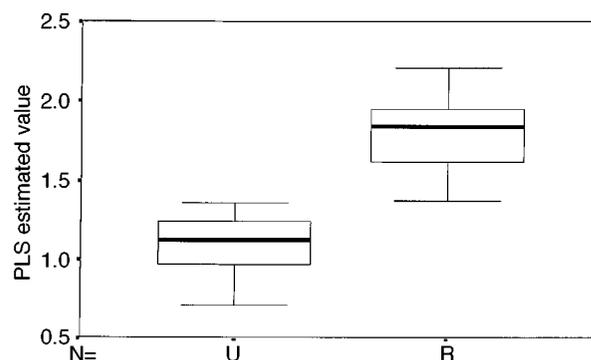


Fig. 5 Box-Whisker plots for the PLS estimated value. U, Unripe class; R, ripe class.

ripening process, malic acid and polyphenols enabled the technological characterisation of the apple varieties.

This work was made possible by financial support from the CICYT (ALI 92-1027-C03).

### References

- Pal, D. K., and Kumar, P. S., *J. Hort. Sci.*, 1995, **70**, 569.
- Karaoulanis, G. D., and Dilley, D., *Int. J. Refrig.*, 1993, **16**, 364.
- Wu, Q. D., Szakacsdozsi, M., Hammat, M., and Harazdina, G., *Plant Physiol.*, 1993, **102**, 219.
- Gussman, C. D., Goffreda, J. C., and Gianfagna, T. J., *Hortscience*, 1993, **28**, 135.
- Percy, A. E., Melton, L. D., and Jameson, P. E., *Plant Sci.*, 1997, **125**, 31.
- Mayr, U., and Treutter, D., *Erwerbsobstbau*, 1996, **38**, 164.
- Fischer, M., and Amado, R., *Carbohydr. Polym.*, 1994, **25**, 161.
- Fischer, M., Arrigoni, E., and Amado, R., *Carbohydr. Polym.*, 1994, **25**, 167.
- Percy, A. E., O'Brien, I. E. W., Jameson, P. E., Melton, L. D., Macrae, E. A., and Redgwell, R. J., *Physiol. Plantarum*, 1996, **96**, 43.
- Garriz, P. I., Alvarez, H. L., and Bartusch, A. M., *Turrialba*, 1995, **45**, 101.
- Chen, H., Duprat, F., Grotte, M., Loonis, D., and Pietri, E., *J. Texture Stud.*, 1996, **27**, 123.
- Yamada, H., Ohmura, H., Arau, C., and Terui, M., *J. Am. Soc. Hortic. Sci.*, 1994, **119**, 1208.
- Autio, W. R., Hayden, R. A., Micke, W. C., and Brown, G. R., *Fruit Varieties J.*, 1996, **50**, 45.
- Murata, M., Tsurutani, M., Tomita, M., Homma, S., and Kaneko, K., *J. Agric. Food Chem.*, 1995, **43**, 1115.
- Larrigaudiere, C., and Vendrell, M., *Sci. Hortic.*, 1993, **55**, 263.
- Golias, J., and Bataille, B., *Facultas Horticulturae*, 1994, **9**, 69.
- Brackmann, A., Streif, J., and Bangerth, F., *Gartenbauwissenschaften*, 1995, **60**, 1.
- Brackmann, A., Streif, J., and Bangerth, F., *Gartenbauwissenschaften*, 1994, **59**, 252.
- Dever, M. D., Cliff, M. A., and Hall, J. W., *J. Sci. Food Agric.*, 1995, **69**, 329.
- Defernez, M., Kemsley, E. K., and Wilson, R. H., *J. Agric. Food Chem.*, 1995, **43**, 109.
- Defernez, M., and Wilson, R. H., *J. Sci. Food Agric.*, 1995, **67**, 461.
- Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., and Kaufman, L., *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988.
- Martens, H., and Naes, T., *Multivariate Calibration*, John Wiley & Sons, Chichester, 1989.
- Seiden, P., Bro, R., Poll, L., and Munck, L., *J. Agric. Food Chem.*, 1996, **44**, 3203.
- Richmond, M. L., Brandao, S. C. C., Gray, J. I., Markais, P., and Stine, C. M., *J. Agric. Food Chem.*, 1981, **29**, 4.
- Blanco, D., Gutiérrez, M. D., and Mangas, J. J., *Chromatographia*, 1988, **25**, 701.
- Blanco, D., Morán, M. J., Gutiérrez, M. D., and Mangas, J. J., *Chromatographia*, 1988, **25**, 1054.

Table 3 Sensitivities (S), specificities (Sp) and classification capacity of the SIMCA technique

Model	Explained variance (%)	S (%)	Sp (%)	Hits (%)
U	88.2	89.5	88.9	84.2
R	95.8	88.9	78.9	100.0

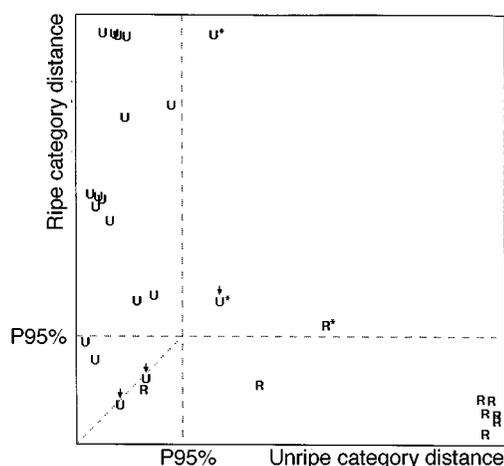


Fig. 4 SIMCA analysis. Coomans' plot. (\*), Outlier; (↓), misclassified; U; unripe class; R, ripe class.

- 
- 28 Mangas, J. J., Suárez, B., and Blanco, D., *Z. Lebensm. Unters. Forsch.*, 1993, **197**, 424.
- 29 Mangas, J. J., Dapena, E., Rodríguez, M. S., Moreno, J., Gutiérrez, M. D., and Blanco, D., *Hortscience*, 1992, **27**, 328.
- 30 Forina, M., Leardi, R., Armanino, C., and Lanteri, S., *PARVUS. An Extendable Package of Programs for Data Exploration, Classification and Correlation*, Elsevier, Amsterdam, 1988.
- 31 Blanco, D., Morán, M. J., Gutiérrez, M. D., Moreno, J., Dapena, E., and Mangas, J. J., *Z. Lebensm.-Unters. Forsch.*, 1992, **194**, 33.

*Paper 7/08534F*

*Received November 26, 1997*

*Accepted March 24, 1998*