

## Definite influence of the extraction methods on the chemical composition of virgin olive oil.

Por R. Aparicio\*, M. Sánchez Navarro y M. S. Ferreiro

Instituto de la Grasa y sus Derivados  
Avda. Padre García Tejero, 4. 41012 - SEVILLA

### RESUMEN

**Influencia de los métodos de extracción en la composición química del aceite de oliva virgen.**

Los métodos de extracción del aceite de oliva virgen afectan a los valores de algunos de sus componentes químicos. Se han aplicado métodos estadísticos a un conjunto de 142 muestras de Andalucía caracterizadas por 55 parámetros químicos. El mayor nivel de significación para la diferenciación de los sistemas de extracción se obtuvo con dos alcoholes alifáticos, un esteroles y un ácido graso. Se ha utilizado el análisis discriminante por pasos para caracterizar los aceites según su sistema de extracción. 205 muestras de distinta procedencia, variedad y cosecha han sido utilizadas para conocer la bondad de los resultados.

*PALABRAS-CLAVE: Aceite de oliva virgen - Método estadístico - Sistema de extracción (influencia de).*

### SUMMARY

**Definite influence of the extraction methods on the chemical composition of virgin olive oil.**

The extraction methods of virgin olive oil affect the quantification values of some chemical compounds. Statistical tests and procedures have been applied to a set of 142 samples collected from all over Andalusia and characterized by 55 chemical compounds. It has been estimated that two aliphatic alcohols, one sterol and one fatty acid have the best significance levels identifying the extraction methodology (pressure, centrifugation). Stepwise Lineal Discriminant procedure has been used to characterize olive oils by the extraction systems. 205 samples from other provinces, varieties or harvests were used to verify these decision equations.

*KEY-WORDS: Extraction system (effect of) - Statistical method - Virgin olive oil.*

### 1. INTRODUCTION

A current line of research is the characterization of the extraction process of olive oil, perhaps as a prior step in its interrelationship with quality. This research involves numerous parameters –organoleptic, physical, biochemical, etc.– and

needs a great deal of different information to avoid randomness in the results, due to non-quantifiable external influences or errors.

In 1981, Martel and Alba studied, only in olives of the variety hojiblanca, the differences between the values of the aliphatic alcohols in oils obtained by centrifugation and pressure in the laboratory. Their conclusions point to the possibility that the temperature of beating (malaxation), during centrifugation, causes the heavier aliphatic alcohols to have higher values in the oils obtained by this method compared with those from the press. However, the work does not mention the statistical method applied, nor the probability on which the conclusions are based.

Later, Tacchino and Borgogni (1983) reached similar conclusions but widening the study to oils from two varieties, different states of ripeness, and extractions obtained both in the laboratory and in industrial processes, although the latter only as complementary information. These authors also did not state the tests used to reach their conclusions.

In 1986 Lanzón et al. published a work concluding that the aliphatic and triterpene alcohols can be used to identify whether and oil is from a pressure or centrifugation extraction system. The work follows the line of Martel and Alba (1981), and studies only the variety hojiblanca. However, the experimental design only allows conclusions based on a study of the repeatability of the quantification method of the alcohols. Thus, in respect of differentiation between the extraction systems, the study is reduced to only two samples –the others being repetitions. Very recently, Nergiz and Ünal (1990) concluded that there are no differences between Turkish olive oils obtained by pressure and centrifugation, using the data from the triperpene alcohols. However, they do not describe the mathematical test applied.

This work studies the influence of the extraction method from two different aspects: on one hand, identification of the most significant chemical parameters in differentiation of the extraction systems, and on the other, characterization of the extraction systems using multivariate statistical techniques.

The work does not start from any initial hypothesis on which some chemical compounds may have most influence on the differentiation or characterization. A total of 347 samples from different regions, varieties, and harvests have been quantified by 55 chemical compounds.

The paper describes the tests and statistical algorithms used, in both differentiation and characterization. Finally, the mathematical equations proposed are verified with samples from other harvests and/or sources to confirm the usefulness of the method. The results of these studies are shown in numerical or graphical form.

## 2. MATERIALS AND METHODS

### 2.1. Chemical analysis

Every sample is described by 55 chemical parameters grouped into five series of chemical compounds (fatty acids, methyl-sterols, sterols, aliphatic and triterpenic alcohols, hydrocarbons) plus oleanolic aldehyde, phytol and erythrodiol.

The whole chemical parameters were measured by gas chromatography using a variety of procedures. Thus, fatty acids have been measured as their methyl-esters produced by inter-esterification, neutralization and a subsequent methylation using a solution of HCl in methanol. A solution of water saturated with NaCl was then added and the final solution was extracted with hexane. The organic phase was analyzed by a Perkin-Elmer Sigma 2 Gas Chromatograph (EGS 2.5%, Chromosorb W 80-100, 2M\*2mm). An equal response factor was considered for the whole set of fatty acids.

The other variables were determined by saponification of the oils, fractionation by preparative thin-layer chromatography (TLC) on a silica gel layer (0.5mm) of the unsaponifiable matter using n-hexane/ethyl acetate (85:15, v/v) as developer. Three great bands were scratched from the TLC plate, the first one containing triterpenic alcohols and a part of phytol and aliphatic alcohols, the second one containing the rest of phytol and aliphatic alcohols, methylsterols and hydroxialdehyde-triterpene, and the third one sterols and erythrodiol.

The three separated bands and the fraction recovered therefrom were extracted separately with isopropyl ether; the solutions were silanated with a mixture of pyridine, hexamethyldisilane and trimethyl-chlorosilane (9:3:1) and independently analyzed by two Hewlett-Packard 5670 GC (methyl-phenyl-silicon HP-5, capillary column 25m ± 0.3 mm).

Phytol and aliphatic alcohols were determined adding the data of the first and second portions, using the cholestane peak as weigh and quantifying the results with heneicosane.

Triterpenic alcohols, methyl-sterols and hydroxi-aldehyde-triterpene were also quantified with heneicosane whilst sterols and erythrodiol with betulin.

The quantification of the hydrocarbons was quite different. To a previous saponification of the oils is added water, and the new solution was many times extracted with ether. The residue fractionation was done in a silica gel column; only the two first fractions were taken. The solutions were separately evaporated and analysed by a gas chromatography (Hewlett-Packard 5670, wide capillary column 30m x 0.75 mm, glass wool with SPB-1). The first fraction allowed to quantify the whole group of hydrocarbons excepting Farnesene and Allofarnesene, which were determined in the second.

### 2.2. Data

The database of the expert system SEXIA (Aparicio et al. 1991c) has been used. This stores information on virgin olive oils from different harvests from cooperatives in Spain, Italy and Portugal. Each oil analysed is a proportional mixture of the samples taken in each tank of the cooperative.

To arrive at the equations, 142 samples were selected, obtained with pressure or centrifugation systems, from different cooperatives of the provinces of Córdoba, Jaén, Málaga and Seville, harvest 1987-88.

The same samples, but from the 1988-89 harvest, were used as verification. Reliable 63 samples from other areas of Spain were also used to know the universality of the equations described.

The samples are of different varieties, climates, and even extraction method conditions. Therefore the results obtained are not biased by the selection of an appropriate variety or by laboratory extraction methods.

### 2.3. Statistical procedures

The statistical libraries BMDP y SPSS-X+ have been used for the work, using a VAX 750 computer.

## 3. RESULTS AND DISCUSSION

Firstly, the normality of data distribution was analysed using kurtosis and skewness. A mathematical transformation was made on those data of each chemical parameter whose distribution was far from normal. Kolmogorov-Smirnov test (BMDP: 3S and SPSS-X: NPAR TEST procedures) was later applied to select the refined set of chemical parameters conform to normality. Thus, 18 from the initial total of 55 chemical parameters were selected (Table I).

Table I  
Skewness and kurtosis of the accepted chemical variables. List of those removed.

Variable	Transformed	Kurtosis	Skewness	Kolmogorov	
				Smirnov	2-Tail
Palmitoleic		-0.147	0.192	0.515	0.954
Stearic		-0.456	0.200	0.521	0.814
Linolenic	ln(X+1.0)	-0.778	0.054	0.695	0.719
Obtusifolios		0.180	-0.241	0.377	0.999
Citrostadienol	ln(X+1.0)	0.197	0.335	0.469	0.980
Docosanol	ln(X+1.0)	-0.167	-0.130	0.607	0.855
Tetracosanol	ln(X+1.0)	0.560	0.150	0.585	0.884
Hexacosanol	ln(X+1.0)	-0.097	0.027	0.563	0.909
Octacosanol	ln(X+1.0)	0.176	0.081	0.574	0.897
Stigmasterol	ln(X+1.0)	-0.252	0.375	0.576	0.894
Delta5_Avenasterol	ln(X+1.0)	0.566	0.396	0.612	0.848
Dammaradienol	ln(X+1.0)	-0.195	-0.004	0.698	0.715
Butyrospermol	sqrt(X)	0.470	0.220	0.541	0.931
24-Methylenlanosterol	ln(X+1.0)	-0.544	0.256	0.664	0.770
Cycloartenol	ln(X+1.0)	-0.166	-0.222	0.450	0.987
24-Methylencycloartenol	ln(X+1.0)	-0.161	0.237	0.650	0.791
Tridecene	ln(X+1.0)	-0.103	-0.075	0.586	0.882
Tricosane	ln(X+1.0)	0.118	0.299	0.577	0.893

Chemical variables not considered:

Palmitic, Margaric, Margaroleic, Oleic, Linoleic, Arachid, Gadoleic, Behenic, Squalene, Methyl 4 (unknown), Gramisterol, Cycloencalenol, Ethyl-lo-phenol, Phytol, Erithrodiol, Oleanolic Aldehyde, Beta-Sitosterol, Campesterol, Taraxerol, Beta-Amyrin, Triter8 (unknown), Copaene, Valencene, Muurolene, Heptadecene, Heneicosane, Tetracosane, Pentacosane, Hexacosane, Heptacosane, Octacosane, Nonacosane, Tricontane, Hentriacontane, Dotriacontane, Tritriacontane, Pentatriacontane, Farnesene, AlloFarnesene.

Tabachnick and Fidell (1983) suggest to verify other conditions before applying multivariate procedures. These are: 1) multivariate homoscedasticity; 2) interrelationship among variables; 3) a number of samples greater than thirty.

The multivariate homoscedasticity hypothesis was verified by Box's M test (SPSS : DSC / STATISTICS=7). The result 6.01, transformed in  $F=0.96$ ,  $df=6$ ,  $p=0.444$ , accepts the null hypothesis that indicates the variance-covariance matrices of the two extraction system are equals one other.

The result of applying Bartlett's sphericity test (SPSS: FACTOR/PRINT = KMO), 2958.99  $p<0.00001$ , shows the correlation matrix is not an identity one and then, there are significative relationships among chemical parameters. Complementary, Kaiser-Meyer-Olkin test (SPSS : FACTOR / PRINT = KMO), 0.76087, points out the relationship between pairs of chemical parameters can be "fine" explained by the others (Kaiser, 1974).

Known multivariate algorithms can be applied, once the chemical parameters had been selected, those most significant in differentiation of the extraction systems and those involved in characterization equations were studied independently.

### 3.1. Chemical parameters related to differentiation of the extraction systems

Hotelling's T-squared (BMDP : 3D) is an appropriate statistical test to know whether a group of chemical parameters (henceforth variables or DVs) is able to differentiate the samples by their extraction system. The test also allows calculation of the significance level of each of the variables in extraction system differentiation, (Harris, 1985).

Concerning the first calculation, T-squared value 254.30, equal to  $F = 2.188$   $p = 0.0037$ , rejected the null hypothesis and consequently the two groups are different enough. In relation to the second calculation, the maximum significance levels are for the aliphatic alcohols hexacosane and octacosane, in agreement with the results of Martel and Alba (1981) plus stearic fatty acid.

To test these results, other statistical test were applied. For example, the initial  $F_{to\_Enter}$  (BMDP: 7M, SPSS: DSC) considers the aliphatic alcohols (hexacosanol y octacosanol) most significant. The non-parametric Kruskal-Wallis test (BMDP: 3S, SPSS: CROSSTABS) adds a sterol (delta5-avenasterol) and a fatty acid (stearic) (Table II).

Table II

Results of the parametric and non-parametric tests.

Variable	Significance Levels		
	Hotellig's T-squared	Kruskall Wallis	F_to_Enter
Stearic	0.0164	0.0281	0.020
Tetracosanol	0.1197	0.0955	0.100
Hexacosanol	<0.0001	<0.0001	<0.001
Octacosanol	0.0005	0.0002	<0.001
Delta5-Avenasterol	0.0698	0.0464	0.080

All the tests used show the little -or non-significance of the triterpene alcohols in differentiation of the extraction systems, in contrast with the conclusions of Lanzón et al. (1986). This is

due to the fact that in that work, a type II error was made in applying the statistical test, the use of missing values, the non-suitability of the data, etc.

However, the acceptance of the aliphatic alcohols is in agreement with the studies of Martel and Alba (1981) although, as seen in table III, the ratio of the alcohols values obtained by centrifugation and pressure is in some cases greater than zero and in others less, depending on the variety analysed. An ANOVA (SPSS: ANOVA) pointed the main effect is due to the varieties rather than extraction systems, for every chemical parameter described in table III. Providing only two varieties (hojiblanca, picual) are considered, the Octacosanol

Table III

Mean values of the alcohols and Delta5\_Avenasterol classified by extraction system.

Centrifugation system					
Variety	docosane	tetracosane	hexacosane	octacosane	delta5_avenasterol
Hojiblanca	1.727	3.600	4.604	1.986	12.479
Picual	3.573	7.129	11.405	2.968	7.782
Empeltre	1.961	3.682	4.553	2.133	12.591
Cornicabra	4.514	4.977	5.039	1.989	8.470
Farga	5.916	11.018	11.855	5.756	10.764
Manzanillo*					
Nevadillo	3.182	6.075	9.444	3.982	
Pressure system					
Variety	butyrospermol	beta-amyrin	24-methylen-lanosterol	cycloartenol	24-methylen-cycloartanol
Hojiblanca	2.360	4.970	0.888	15.196	94.193
Picual	2.658	5.043	0.836	15.472	86.527
Empeltre	1.084	3.427	0.550	9.550	36.073
Cornicabra	2.335	2.627	0.538	10.403	44.554
Farga	5.968	4.173	1.337	20.063	59.861
Manzanillo*					
Nevadillo	2.135	4.913	0.693	13.690	80.411
Centrifugation system					
Variety	docosane	tetracosane	hexacosane	octacosane	delta5_avenasterol
Hojiblanca	3.194	4.889	5.542	2.534	12.728
Picual	3.562	6.250	8.053	2.229	7.258
Empeltre	2.716	5.334	5.539	2.783	11.042
Cornicabra	7.439	10.848	8.170	2.441	9.687
Farga	6.302	11.657	12.495	5.791	7.716
Manzanilla	4.112	6.233	7.346	3.431	14.383
Nevadillo	4.795	9.837	12.127	3.834	8.586
Pressure system					
Variety	butyrospermol	beta-amyrin	24-methylen-lanosterol	cycloartenol	24-methylen-cycloartanol
Hojiblanca	3.419	4.562	0.854	16.226	92.821
Picual	2.334	4.709	0.823	14.527	90.243
Empeltre	3.317	4.092	0.815	16.872	50.687
Cornicabra	3.215	3.475	0.705	13.819	58.630
Farga	5.347	4.438	1.157	15.836	46.972
Manzanilla	3.916	3.576	0.853	13.543	66.682
Nevadillo	3.138	5.166	0.984	16.133	82.030

\* No information on this variety.

nol has the greatest source of variation in extraction systems. Thus, it can be deduced that a study of this type cannot be restricted to analysis of only one variety, particularly when this is hojiblanca.

### 3.2. Characterization of the extraction systems

To obtain decision rules in a process in which there are no expert rules, multivariate statistical methods have been used (Aparicio, 1988) (Aparicio et al., 1991b). The algorithms used are from principal component procedures, stepwise linear discriminant analysis, and multiple regression, whose study and results are described below.

#### 3.2.1. Principal component analysis

The procedure of principal components allows knowing what the structure of the data set is like

before applying other multivariate statistical algorithms.

The statistical criterion CHISQ/DF was chosen to detect the existence of outliers. None were found, as the values of the criterion for each sample were lower than 1.69. The values of the statistic SMC (Squared Multiple Correlation) were much lower than 0.99, which is the limit indicating the existence of multicollinearity in the data set (Tabachnick and Fidell 1983).

Analysis of the values of the communalities shows that these are greater in the samples obtained by centrifugation than in those by pressure. This difference could be due to the lower influence of the press on the chemical composition of the oils, or, from another point of view, on the presumed homogeneity of the oils due to the beating (Martel and Alba, 1981).

The interrelationships between variables are disclosed by principal component analysis. The

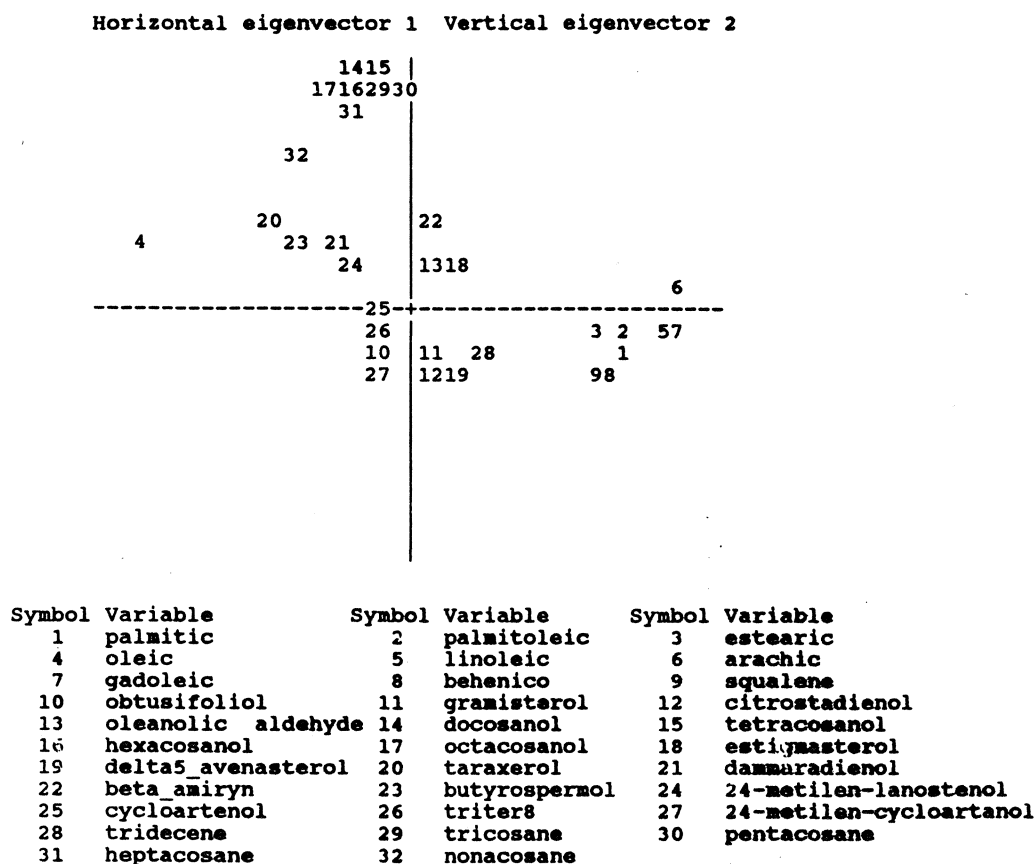


Figure 1  
Loading plot on eigenvector 1 - eigenvector 2.

matrix of variance-covariance shows that some chemical compounds are highly intercorrelated with other of their serie. For instance, Docosanol-Tetracosanol (0.93,  $p < 0.0001$ ), Tetracosanol-Hexacosanol (0.85,  $p < 0.0001$ ), Hexacosanol-Octacosanol (0.93,  $p < 0.0001$ ), Dammaradienol-Butyrospermol (0.93,  $p < 0.0001$ ), Dammaradienol-Cycloartenol (0.8,  $p < 0.0001$ ).

On the other hand, the varivectors show the most significant chemical parameters in explaining the variance. The first four varivectors explain 83.9% of the variance in the factorial space. The first of these (the aliphatic alcohols) contributes 34.9% of the variance; the second (different compounds, among which there are some triterpenic alcohols) 23.3%; the third (fatty acids) 17.1%; and the fourth (the other triterpenic alcohols), 8.6%. Figure 1 shows the situation of the variables bidimensionally, considering only the first two varivectors. Only one clear group can be observed, formed by the aliphatic compounds, which is saturated on the first factor.

### 3.2.2. Stepwise lineal discriminant analysis

The procedure of SLDA (Stepwise Lineal Discriminant Analysis) allows obtaining canonical equations able to identify the extraction system of an unknown sample, within certain limits of probability.

SLDA was used on only those samples with values greater than F-to-enter (4.79) and F-to-remove (4.60) in each selection step. This avoids obtaining an equation by pure randomness. The final probability indicated corresponds to that obtained after applying the "jackknife" method.

At the end of the algorithm-selection process, the canonical equation (eq. 1) is formed by the two aliphatic alcohols, tetracosanol and hexacosanol, and a sterol, delta5-avenasterol. With these three variables SLDA explains 57% of the variance of the data, with a canonical correlation coefficient  $R=0.60$ , for  $CHISQ=44.49$ ,  $p < 0.00001$ ,  $df=3$ . The equation, verified by the jackknifed method, correctly classifies 75% of the samples extracted by pressure, and 81.5% of those extracted by centrifugation, with a mean of 78.4% correct classifications. Unfortunately, the discriminanting criterion, ratio between the variability intergroups and intragroups, shows the equation is not too efficient characterizing the extraction systems.

$$Y = 3.921 \cdot \text{Ln} (\text{Tetracosanol} + 1.0) - 5.757 \cdot \text{Ln} (\text{Hexacosanol} + 1.0) - 2.227 \cdot \text{Ln} (\text{Delta5-Avenasterol} + 1.0) + 10.298 \quad (\text{eq. 1})$$

The choice of Tetracosanol may be due to the interrelationship among aliphatic alcohols, in spite of its low initial level of significance, table II.

### 3.3. Verification of the results

The equations described above were verified using samples from the same places but other harvests, with samples from different places, and even other harvests and varieties.

For the oils extracted by pressure, 88% of the samples for 1989 from Andalusia, and 79.5% of those from other areas, were correctly classified. For the oils extracted by centrifugation, 74% of the Andalusian oils were well classified and 73% of the rest.

It can be seen in the samples from Andalusia for 1989 that the correct classifications for the pressure-extracted oils decreases markedly in comparison with those of the continuous systems. This leads to the deduction that the latter are less affected, from the chemical point of view, by variability of the data with time. This theory is supported by the fact that 34.5% of the errors are repeated in the same cooperatives during the two harvests. Of these, 70% are for pressure-extracted samples.

The process is reversed on applying the equations obtained with Andalusian varieties to samples from other regions. In general, the results are good, taking into account that 43% of the errors in classification of the continuous system are with the variety cornicabra from the province of Toledo, and 33% of the errors of pressure-extracted oils with the variety manzanillo from the province of Huelva, having analysed 63 samples from the provinces of Badajoz, Cáceres, Ciudad Real, Castellón, Huelva, Tarragona, Teruel, Toledo and Zaragoza.

## 4. CONCLUSIONS

Earlier works showed that aliphatic alcohols could distinguish oils by their extraction systems. This work, using a statistical study, confirms those tendencies, adding tetracosanol and hexacosanol to the group of aliphatic alcohols, together with the sterol delta5-avenasterol and the fatty acid stearic. The influence of the triterpene alcohols has not been detected in this differentiation.

The strong relationship among the compounds in the aliphatic alcohol series means that the other

compounds of that series do not have a high significance level. This relationship might explain the small discrepancy with the conclusion of Martel and Alba (1981) on the greater influence of the higher molecular weight aliphatic alcohols, hexacosanol and octacosanol.

Applying Hotelling's test separately to virgin olive oils of different varieties, it was seen that the hydrocarbons tetracosane ( $p=0.075$ ), hexacosane ( $p=0.076$ ) and heptacosane ( $p=0.0117$ ) could distinguish between the systems, although it has not been possible to verify with a large set of different samples.

The results from the characterization equations show that improved results may be obtained by means of independent analysis of the oils by variety. Methods such as those indicated by Aparicio et al., (1991a) could be used, followed by identification of the variety by means of SEXIA (Aparicio, 1988).

The work, from a strictly mathematical point of view, will follow the line noted by Aparicio (1988) using algorithms from possibility theory.

#### ACKNOWLEDGEMENT

This work has been supported by the CAYCIT, project ALI-88-0187-CO2-2, and the Dirección General de Investigación y Extensión Agraria of the Junta de Andalucía.

#### REFERENCES

- Aparicio, R. (1988).- "Characterization of foods by inexact rules: The SEXIA expert system".- *J. of Chemometrics*, **3**, 175-192.
- Aparicio, R.; Ferreiro, L.; Cert, A. and Lanzón, A. (1990).- "Andalusian Virgin Olive Oil Characterization".- *Grasas y Aceites*, **41**, 23-39 (in Spanish).
- Aparicio, R.; Ferreiro, L. and Rodriguez, J. (1991a).- "Foods characterization linking relational and lineal decision rules. An application on virgin olive oils of Malaga".- *Grasas y Aceites*, **42**, 2 (in Spanish).
- Aparicio, R.; Ferreiro, L.; Leardi, R. and Forina, M. (1991b).- "An Approach for building Decision Rules by Chemometric analysis. An Application on olive oil".- *Chemometrics and Intelligent Laboratories Systems*, **10**, 349-358.
- Aparicio, R.; Ferreiro, L. and Rodriguez, J. (1991c).- Characterization of Andalusian Virgin Olive Oil. Statistical results of SEXIA project".- Seville (Spain).- Agriculture Ministry of Andalusia (in Spanish).
- Harris R.J. (1985).- "A primer of Multivariate Statistics".- Academic Press, Orlando, Florida (U.S.).
- Kaiser, H. F. (1974).- "An index of factorial simplicity".- *Psychometrika* **39**, 31-36.
- Lanzón, A.; Albi, T. and Gracián, J. (1986).- "Influencia del sistema de extracción en algunos componentes del aceite de oliva. Primeros ensayos".- *Grasas y Aceites*, **37**, 254-258.
- Martel, J. and Alba, J. (1981).- "Influencia del método de obtención de aceites de oliva por centrifugación de pastas sobre su contenido en alcoholes grasos superiores".- *Grasas y Aceites*, **32**, 233-237 (in Spanish).
- Nergiz, C. and Ünal, K. (1990).- "The effect of extraction systems on triterpene alcohols and squalene content of virgin olive oil".- *Grasas y Aceites*, **41**, 117-121.
- Tabachnick, B. G. and Fidell, L.S. (1983).- "Using Multivariate Statistics".- Harper & Row, New York.
- Tacchino, E. and Borgogni, C. (1983).- "Indagine sul contenuto di alcoli alifatici dell'olio d'oliva di pressione e di estrazione".- *Riv. Ital. Sostanze Grasse*, **40**, 575-581.

(Recibido: Noviembre 1990)