Olive oil characterization using excitation-emission fluorescence spectroscopy and three-way methods of analysis

Doctoral Thesis

UNIVERSITAT ROVIRA I VIRGILI

Universitat Rovira i Virgili

Departament de Química Analítica i Química Orgànica

Olive oil characterization using excitation-emission fluorescence spectroscopy and three-way methods of analysis

Memòria presentada per
Francesca Guimet Vila
per optar al grau de
Doctora en Química
Tarragona, 2005

UNIVERSITAT ROVIRA I VIRGILI
Departament de Química Analítica
i Química Orgànica

El Dr. RICARD BOQUÉ MARTÍ i el Dr. JOAN FERRÉ BALDRICH, Professors Titulars del Departament de Química Analítica i Química Orgànica de la Facultat de Química de la Universitat Rovira i Virgili,

CERTIFIQUEM:

Que la present Tesi Doctoral, que porta per títol: "OLIVE OIL CHARACTERIZATION USING EXCITATION-EMISSION FLUORESCENCE SPECTROSCOPY AND THREE-WAY METHODS OF ANALYSIS", presentada per FRANCESCA GUIMET VILA per optar al grau de Doctora en Química, ha estat realitzada sota la nostra direcció, a l'Àrea de Química Analítica del Departament de Química Analítica i Química Orgànica d'aquesta universitat, i que tots els resultats presentats són fruits d'experiències realitzades per l'esmentada doctoranda.

Tarragona, setembre de 2005

Dr. Ricard Boqué Martí   Dr. Joan Ferré Baldrich

Voldria dedicar aquesta pàgina a donar les gràcies a tota la gent que d'una manera o d'una altra ha contribuït a la realització d'aquesta tesi.

En primer lloc, agraeixo als meus directors de tesi, el Dr. Ricard Boqué i el Dr. Joan Ferré, que m'hagin donat l'oportunitat de realitzar aquesta tesi doctoral. També els agraeixo els coneixements que m'han transmès, la seva ajuda i el seu suport durant aquests anys.

També voldria agrair al Prof. F. Xavier Rius, director del grup de Quimiometria, Qualimetria i Nanosensors la confiança que ha dipositat en mi i els ànims que m'ha donat durant la realització de la tesi.

Dono les gràcies també a la resta de membres del grup de Quimiometria, Qualimetria i Nanosensors, tant els actuals com els antics. Són moltes les persones que han passat pel grup durant els anys en els que he realitzat la tesi i no voldria oblidar-me de ningú. A tots els agraeixo la seva ajuda i amistat. No vull oblidar-me tampoc dels membres del grup de Química Analítica Enològica i dels Aliments, amb els quals també he compartit bons moments i han estat disposats a ajudar-me quan ha calgut.

Agraeixo al Josep Garcia i molt especialment a la Marta Vidal del Laboratori Agroalimentari de Cabrils la seva col·laboració en aquesta tesi durant el darrer any.

Dono les gràcies a tots els amics que durant aquests anys m'han acompanyat i que, potser sense saber-ho, han posat el seu granet de sorra a la realització d'aquesta tesi. Especialment, vull recordar-me de la Cristina i la Cori, que sempre han estat al meu costat.

Finalment, dono les gràcies als meus pares i als meus germans, Jordi i Eugeni, per la seva ajuda i el suport que m'han donat en tot moment.

STRUCTURE OF THE THESIS

This thesis is based on publications in various international journals. These papers have been edited to give uniform format and uniform mathematical notation along the thesis.

The thesis is structured in seven chapters:

Chapter 1 *Introduction and objectives* describes olive oil and the analytical techniques commonly used for its characterization. This chapter also points out the suitability of fluorescence spectroscopy for olive oil analysis. The potential of using three-dimensional fluorescence spectroscopy and chemometric methods is commented. Finally, the objectives of this thesis are stated.

Chapter 2 *Theoretical background* is divided into three parts. The first part introduces excitation-emission fluorescence spectroscopy (EEFS). The fluorescent behaviour of the oil types used in this thesis is also commented. The second part explains the arrangement of EEMs in three-way arrays. The last part of the chapter describes the chemometric methods applied throughout this thesis. They are divided into methods of exploratory analysis and methods of classification.

Chapter 3 *Exploratory analysis of olive oils* contains two papers about the application of chemometric methods to the fluorescence excitation–emission matrices (EEMs) of olive oils for exploratory analysis. These papers are *Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy, F. Guimet, J. Ferré, R. Boqué, and F.X. Rius, Anal. Chim. Acta 515 (2004) 75-85,* and *Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy, F. Guimet, R. Boqué, and J. Ferré, J. Agric. Food Chem. 52 (2004) 6673-6679.*

In both papers, different spectral ranges and preprocessing methods for distinguishing between various types of oils are compared. The main advantages of using second-order data over first-order (multivariate) data are also commented.

Chapter 4 *Fluorescence-quality relationships in olive oils* contains the paper *Excitation-emission fluorescence spectroscopy combined with three-way methods of analysis as a complementary technique for olive oil characterization, F. Guimet, J. Ferré, R. Boqué, M. Vidal, and J. Garcia, J. Agric. Food Chem. (accepted for publication)*, which shows the potential of EEFS for olive oil characterization. In this paper, the relationship between the EEMs of a set of commercial Spanish olive oils and some of their quality parameters (peroxide value, K_{270}) is studied.

Chapter 5 *Olive oil classification* includes three papers. In the first one, *Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis, F. Guimet, J. Ferré, and R. Boqué, Anal. Chim. Acta 544 (2005) 143-152*, various chemometric methods are applied to detect olive-pomace oil adulteration in extra virgin olive oils at low levels. The adulteration level is also quantified.

The second paper of this chapter, *Study of oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis, F. Guimet, R. Boqué, and J. Ferré, Grasas y Aceites 56 (4) (2005) 292-297*, shows the suitability of EEFS and chemometric methods to discriminate between extra virgin olive oils from the two Spanish regions of the protected denomination of origin "Siurana" production area.

In the third paper of this chapter, *Application of non-negative matrix factorization combined with Fisher's linear discriminant analysis for classification of olive oil excitation-emission fluorescence spectra, F. Guimet, R. Boqué, and J. Ferré, Chemom. Intell. Lab. Syst. (accepted for publication)*, the suitability of the non-negative matrix factorization (NMF) algorithm applied to olive oil EEMs is studied. This paper shows that NMF decomposes the spectral data into meaningful parts, which may be related to the fluorescence species present in oils. The paper also shows the potential of NMF in conjunction with a linear discrimination method for classification purposes.

Chapter 6 *Conclusions and suggestions for future research* contains the conclusions of the thesis and some ideas for future research.

The last chapter is an *Appendix* containing a list of the abbreviations used in this thesis, the list of papers and meeting presentations in which the author took part during the period of development of this thesis and a summary of the thesis.

TABLE OF CONTENTS

Chapter 1

Introduction and Objectives

This chapter is organised as follows. Section 1.1 explains the importance of olive oil authentication and discusses the lack of rapid analytical techniques in olive oil quality control laboratories. This section also highlights the suitability of excitation-emission fluorescence spectroscopy (EEFS) for olive oil characterization and reports that classification methods have been little explored in the field of three-way analysis. These points justify the objectives of this thesis (section 1.2).

1.1 OLIVE OIL

1.1.1 A little history

The origin of the olive tree (Fig. 1) is unknown. It is said to have appeared in prehistoric times in southern Asia Minor where there are now abundant forests of wild olive trees and appears to have spread from Syria towards Greece through Anatolia. However, other theories claim it originated in the Mediterranean basin or in Lower Egypt [1].

Figure 1. Olive tree.

The claim that the olive tree originated in Europe could arise from the Greek myth about the fight between Athena and Poseidon for the control of Athens [2]. According to ancient Greek history, Poseidon, god of the sea, and Athena, goddess of peace and wisdom, argued over whose name should be given to the newly built city in the land of Attica. To end this dispute, it was decided that the city would be named after the one who offered the most precious gift to the citizens. Poseidon struck his trident on a rock and spring water began to flow. Athena struck her spear on the ground and it turned into an olive tree (Fig. 2). It was decided that the

Figure 2. Representation of the Greek myth about the origin of the olive tree.

olive tree was more valuable to the people of Attica, so the new city was named Athens in honour of Athena. Even today, an olive tree stands where the legend took place. It is said that all the olive trees in Athens are descended from the first olive tree offered by Athena.

Around 600 B.C. olive tree cultivation spread from Greece or Northern Africa to Italy, Spain and the other Mediterranean countries (Fig. 3). In Spain, the olive tree was probably introduced by the Greeks, the Romans and the Arabs [2].
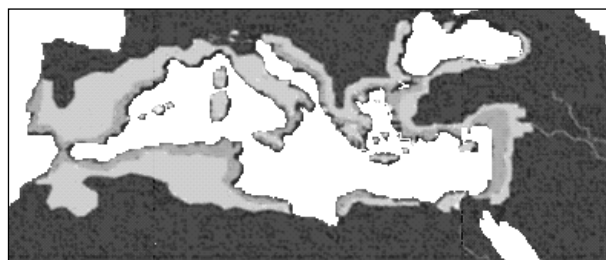
Figure 3. Olive tree expansion to the Mediterranean countries.

The scientific name for olive tree is *Olea europea L.*, which comes from the Greek word *elea*, which is used to designate the olive tree. In Latin this becomes *olea*.

1.1.2 Composition and classification

Olive oil is obtained from the fruit of the olive tree (the olive) (Fig. 4). It has a fine aroma, a pleasant taste and a high nutritional and health value. Olive oil mainly comprises triacylglycerols (triglycerides), and contains small quantities of free fatty
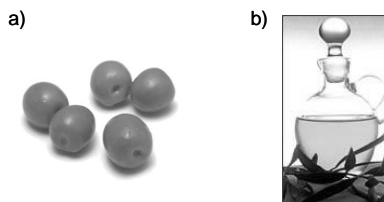
Figure 4. (a) Olives (*arbequina* variety) and (b) olive oil.

acids, glycerol, phosphatides, pigments, flavour compounds, sterols, unidentified resinous substances and other constituents. The structures of some of the components of olive oil are shown in Figure 5. These components can be divided into two categories [3,4]:

- The saponifiable fraction (triacylglycerols, free fatty acids, phosphatides).
- The unsaponifiable fraction (hydrocarbons, sterols, tocopherols, fatty alcohols, phenolic compounds, pigments, phospholipids, volatile compounds, diacylglycerols, waxes), which covers a small percentage (0.5-1.5%).

The main triacylglycerol present in olive oil is triolein (40% of the total composition). It consists of a glycerol with three ester linkages. The major fatty acids present as glycerides in olive oil are oleic acid (C18:1), linoleic acid (C18:2), palmitic acid (C16:0), and stearic acid (C18:0). Oleic acid is present at higher concentrations than the other acids (55-83%).

The International Olive Oil Council (IOOC (1985, 1995, 1997)) proposed a classification and definition of olive oil and olive-pomace oil, which in general agree with those of European Union (EU Commission Regulations 1991, 1995) [3]. The classification and main characteristics of olive oils are summarized in Table 1.

Figure 5. Chemical structures of some olive oil components: (a) triolein, (b) oleic acid, (c) linoleic acid, (d) steraric acid, (e) linolenic acid, (f) α-tocopherol, (g) chlorophyll *a*, (h) pheophytin *a*, and (i) hydroxytyrosol.

Table 1. Olive oil classification according to the IOOC

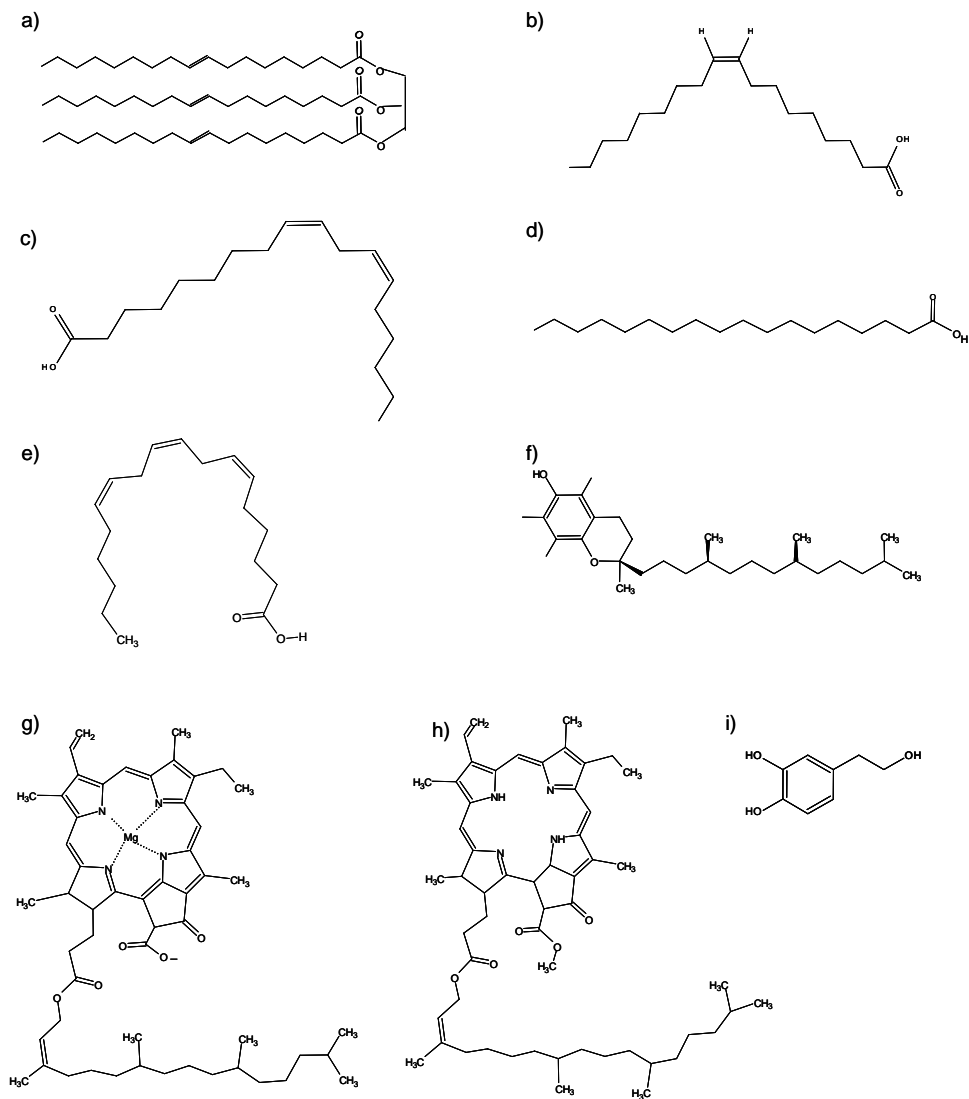| Olive oil type		Characteristics
Virgin olive oil	*Extra virgin olive oil (EVOO)*	- Obtained by mechanical or other physical means that do not lead to alterations of the oil. Has not undergone treatment other than washing, decantation, centrifugation and filtration. - Maximum acidity expressed as oleic acid, of 1.0g/100g. Meets the requirements for the sensory characteristics and other quality criteria for this oil category.
	Virgin olive oil (VOO)	- Maximum acidity of 2.0g/100g. Meets the requirements for the sensory characteristics and other quality criteria for this oil category. May be named "fine olive oil" at the production stage.
	Ordinary virgin olive oil (OVOO)	- Maximum acidity of 3.3g/100g. Meets the requirements for the sensory characteristics and other quality criteria for this oil category.
	Lampante virgin olive oil (LVOO)	- Acidity of over 3.3g/100g. Meets the requirements for sensory characteristics and other quality criteria for this oil category. Cannot be used for consumption as is. Must undergo refining or be used for technical purposes.
Refined olive oil (ROO)		- Obtained from virgin olive oil by a refining process that does not lead to alterations in the initial glycerol structure.
Olive oil (or pure olive oil, POO)		- A blend of virgin olive oil (except lampante) and refined olive oil.
Olive-pomace oil	*Crude olive-pomace oil (COPO)*	- Extracted from olive-pomace using solvents. Does not include oils obtained by re-esterification processes or any mixture of other oils. - Intended for refinement for human consumption or for technical purposes as is.
	Refined olive oil (ROPO)	- Obtained from crude olive-pomace oil by a refining process that does not lead to alterations in the initial glyceride structure.
	Olive-pomace oil (OPO)	- A mixture of refined olive-pomace oil and virgin olive oil (except lampante). This blend should not be called "olive oil".

1.1.3 Olive oil and health

Olive oil is a source of lipids, which provide chemical energy to the cells. Lipids are also a source of insoluble vitamins that play an important role in the development of the reproductive and nervous systems and have benefits for the skin and liver. Olive oil is also a source of vitamins A (β-carotene), E (tocopherols), and F (linoleic and linolenic acids) (Fig. 5). Vitamin F cannot be synthesized by the human organism but it is essential for regulating the fluency and permeability of the cellular membrane structure. For this reason, it must be present in the diet [2].

Among the free fatty acids present in olive oil, monounsaturated fatty acids, such as oleic acid, offer protection against heart disease by reducing low-density lipoprotein (LDL) cholesterol levels while raising high-density lipoprotein (HDL) cholesterol levels.

The accumulation of free radicals, due to oxidation processes in the body, causes serious problems for human health. Specifically, free radicals destroy the polyunsaturated fatty acids of the membranes and DNA. This facilitates the ageing process, damages the liver and can even lead to the formation of cancer. The human organism is protected from the free radicals by free radical scavengers such as vitamin E and phenols. Phenols, present in significant amounts in EVOO, prevent the destruction of human cells.

Olive oil is greatly assimilated by the human body. This is mainly attributed to the high percentage of triolein. Also, the pigments chlorophylls *a* and *b*, pheophytins *a* and *b* and the aroma components present in olive oil (i.e. aldehydes, ketones, hydrocarbons, esters, ethers, terpene alcohols, furan and thiophene derivates) facilitate its absorption because they produce changes in the gastric fluid composition of the stomach and increase digestive activity. The great assimilation of olive oil also helps the absorption of vitamin E and phenols [2].

1.1.4 Characterization

Olive oil is an economically important product, especially in Mediterranean countries, where it is one of the basic components of the diet. According to the Spanish Ministry of Agriculture, Fishing and Food [5], in 2001 the total production of olive oil in Spain was of 1,422,000 tons. The huge importance of this food product makes the quality control of olive oil and the detection of possible fraud issues of great interest. Olive oil authentication requires the measurement of several parameters. These are summarized in Table 2, which shows the legal limits established by the EU (Regulation (EEC) No 2568/91) [6].

The EU has established Official Methods of Analysis for these parameters. For example, stigmastadienes, fatty acids, erythrodiol and uvaol are analysed by gas chromatography with a flame ionization detector (GC-FID). Ultraviolet (UV) spectroscopy is used to determine K_{232}, K_{270} and ΔK (eq. 1). Finally, free acidity and

Table 2. Characteristics of olive oil types (Regulation (EEC) No 2568/91)

Category	Acidity (%)	Peroxide value (mEq O_2/kg)	Stigmastadienes (mg/kg)	K_{232}	K_{270}	ΔK	Fatty acid content (%)						Total sterols (mg/kg)	Erythrodiol and uvaol (%)
							myristic	linolenic	arachidic	eicosenoic	behenic	lignoceric		
EVOO	≤0,8	≤20	≤0,15	≤2,50	≤0,22	≤0,01	≤0,05	≤1,0	≤0,6	≤0,4	≤0,2	≤0,2	≥1000	≤4,5
VOO	≤2,0	≤20	≤0,15	≤2,60	≤0,25	≤0,01	≤0,05	≤1,0	≤0,6	≤0,4	≤0,2	≤0,2	≥1000	≤4,5
LOO	>2,0	-	≤0,50	-	-	-	≤0,05	≤1,0	≤0,6	≤0,4	≤0,2	≤0,2	≥1000	≤4,5
ROO	≤0,3	≤5	-	-	≤1,10	≤0,16	≤0,05	≤1,0	≤0,6	≤0,4	≤0,2	≤0,2	≥1000	≤4,5
POO	≤1,0	≤15	-	-	≤0,90	≤0,15	≤0,05	≤1,0	≤0,6	≤0,4	≤0,2	≤0,2	≥1000	≤4,5
COPO	-	-	-	-	-	-	≤0,05	≤1,0	≤0,6	≤0,4	≤0,3	≤0,2	≥2500	>4,5
ROPO	≤0,3	≤5	-	-	≤2,00	≤0,20	≤0,05	≤1,0	≤0,6	≤0,4	≤0,3	≤0,2	≥1800	>4,5
OPO	≤0,1	≤15	-	-	≤1,70	≤0,18	≤0,05	≤1,0	≤0,6	≤0,4	≤0,3	≤0,2	≥1600	>4,5

peroxide value (PV) are determined by titration [6]. Sensory analysis is also carried out to characterize olive oils [7].

$$\Delta K = K_m - \frac{K_{m-4} + K_{m+4}}{2} \qquad (1)$$

In eq. (1), K_m is the specific extinction of a solution of 1% of oil in isooctane using a length path of 1 cm at the *m* wavelength (i.e. the wavelength with maximum absorption at around 270 nm).

Of the techniques used to determine olive oil authenticity and quality, chromatographic techniques are the most widely used [8-14]. UV spectroscopy is also commonly used, especially for detecting adulterations [3,15]. Recently, however, other techniques in the field of olive oil analysis have emerged. The most noteworthy of these are headspace-mass spectrometry (HS-MS) [16,17], near-infrared (NIR), mid-infrared (MIR), and Fourier transform-Raman (FT-Raman) spectroscopy [4,18-27], nuclear magnetic resonance (NMR) spectroscopy [23,28-32], and electronic olfactometry [33,34].

Studies on olive oil stability against oxidation are also found in the literature. Psomiadou and Tsimidou [35] studied the changes in the lipid substrate due to oxidation by PV and K_{232} measurements. These authors also evaluated the changes in the α-tocopherol, pigment and squalane contents by HPLC and measured total polar phenol content colorimetrically. Deiana et al. [36] studied the activity of α-tocopherol in EVOOs by chromatographic analysis of α-tocopherol and hydroperoxides and took PV measurements.

Some of the species present in olive oils, e.g. vitamin E (tocopherols), chlorophylls, pheophytins and phenolic compounds, emit fluorescence (Fig. 5). Therefore, fluorescence spectroscopy can also be used to analyse these species.

Using fluorescence to analyse olive oils was first proposed in 1925 by Frehse, who studied the possibility of detecting the presence of ROOs in VOO by examining the oils under a quartz lamp with a Wood filter [37]. From these results, fluorescence was applied to detect adulterations in VOOs [3,37]. Because of the high sensitivity of this technique, fluorescence detectors have often been coupled to HPLC instruments for olive oil analysis [9,14]. The main applications are the determination of tocopherols [9,38] and phenolic compounds [14,39,40]. However, chromatographic techniques have several disadvantages – for example, they require large amounts of chemicals and are time consuming.

Engelsen [41] published an interesting study that evaluated the deterioration of frying oil (consisting of a blend of rapeseed and palm oil) from a commercial Chinese spring roll plant using fluorescence, near infrared-visible (NIR-VIS), FT-IR and FT-Raman spectroscopy. To resemble on-line measurements as closely as possible, all the sample spectroscopic evaluations were performed without pretreatments such as dilution or filtering. Several chemical-physical parameters (viscosity, glyceride content, free fatty acids, anisidine value, PV, iodine value, and vitamin E) were also measured in accordance with standard methods of analysis. This study showed that fluorescence used together with multivariate analysis is a powerful tool for monitoring the deterioration of frying oil.

Kyriakidis and Skarkalis showed that fluorescence spectra measured directly to undiluted oils distinguish between VOO and other vegetable oils [42]. Fluorescence spectroscopy can therefore take non-invasive measurements of olive oils without using chemicals and speed up the analysis. Fluorescence spectroscopy was also used to determine the amount of VOO in commercial POOs [43] and to detect hazelnut oil adulteration in VOOs [44].

Though fluorescence spectroscopy is a selective technique, when applied to complex natural systems, the selectivity of conventional techniques appears to be insufficient. Multidimensional fluorescence techniques (e.g. synchronous luminescence and total luminescence spectroscopy (TLS)) are therefore used in

such applications and provide additional information about the samples [45,46]. TLS is also called excitation-emission fluorescence spectroscopy (EEFS) and involves the simultaneous acquisition of multiple excitation (λ_{ex}) and emission (λ_{em}) wavelengths. As a result, a total fluorescence profile of the sample over the range scanned is obtained. This fluorescence landscape is called the excitation-emission matrix (EEM) (Fig. 6) because it is a data matrix of dimensions (number of $\lambda_{em} \times$ number of λ_{ex}). The arrangement of the data in a matrix array is known as second-order data.
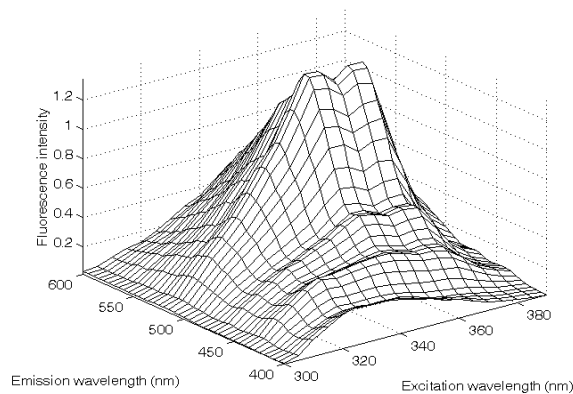
Figure 6. EEM of an EVOO between λ_{ex}= 300-390 nm and λ_{em}= 400-600 nm.

In 1984 Wolfbeis and Leiner [47] used EEFS to characterize various types of edible oils and this technique has recently been applied to olive oil characterization [46,48,49]. However, it is not well established in the field of olive oil analysis and, since its capabilities are still being explored, it is not normally used in olive oil quality control laboratories.

To extract useful information from the data, chemometric methods must be used. Such methods when applied to a set of second-order data are called three-way methods of analysis because the data from several samples are arranged in a three-dimensional structure. When several samples are measured with EEFS, a set of EEMs is obtained that can be arranged into a cube of dimensions (*samples* $\times$ number of $\lambda_{em} \times$ number of λ_{ex}).

Three-way methods of analysis have been widely used for exploratory analysis and for quantitative purposes. These methods include unfold principal component analysis (unfold-PCA), parallel factor analysis (PARAFAC) and three-way partial least squares (PLS) regression methods. They have been widely applied to EEFS [50-60]. Classification methods applied to EEFS have been little explored, however. Wittrup [61] compared several classification methods (soft independent modelling of class analogy (SIMCA) and discriminant PLS methods) for classifying fungal extracts from three-way fluorescence data. In the field of food analysis, studies on the application of EEFS together with classification methods are scarce. Recently, Scott et al. [48] classified edible oils using EEFS and various chemometric methods. The lack of a fast technique for olive oil characterization and the presence of fluorescence species in this food product make olive oil analysis an interesting research area for exploring the capabilities of EEFS combined with three-way methods of analysis, particularly classification methods.

1.2 OBJECTIVES OF THE THESIS

The main objective of this thesis is to develop new methods based on EEFS combined with three-way methods of analysis for olive oil characterization. The chemometric methods applied in this thesis can be divided into two large groups: exploratory methods and classification methods.

The first objective is to apply exploratory methods to the fluorescence EEMs of different types of oils to visualize the samples, discover sample groups, find relationships between them and between their fluorescence spectra and detect outliers.

The second objective is to study the capabilities of EEFS combined with three-way methods as a fast complementary technique for assessing olive oil quality.

The final objective is to develop new methods based on EEFS and three-way methods for olive oil authentication. Several aspects, including the detection of adulterations and authentication based on oil type and origin, are considered at this stage.

1.3 REFERENCES

[1] International Olive Oil Council USA. http://www.internationaloliveoil.org/, last access: May 3, 2005.

[2] A.K. Kiritsakis, *El aceite de oliva*, A. Madrid Vicente Ed., Madrid, 1992.

[3] A.K. Kiritsakis, *Olive oil. From the tree to the table*, Food and Nutrition Press, Inc., 2nd Ed., Trumbull, 1998.

[4] I. Montoliu, *Aplicació i desenvolupament de tècniques quimiomètriques a l'anàlisi agroalimentària*, UAB, Bellaterra, 2001.

[5] Spanish Ministry of agriculture, fishing and food. http://www.mapya.es/, last access: June 3, 2005.

[6] EU, Official Journal of the European Union, Commission Regulation (EC) No 1989/2003. Amending Regulation (EEC) No 2568/91, L295 (2003) 57, http://europa.eu.int/eur-lex/en/, last access: June 21, 2005.

[7] R. Aparicio, M.T. Morales, V. Alonso, *J. Agric. Food Chem.* 45 (1997) 1076-1083.

[8] R. Aparicio, R. Aparicio-Ruíz, *J. Chromatogr. A* 881 (2000) 93-104.

[9] A. Cert, W. Moreda, M.C. Pérez-Camino, *J. Chromatogr. A* 881 (2000) 131-148.

[10] B. Gandul-Rojas, M.R.L. Cepero, M.I. Mínguez-Mosquera, *JAOCS* 77 (2000) 853-858.

[11] D-S. Lee, B-S Noh, S-Y Bae, K. Kim, *Anal. Chim. Acta* 358 (1998) 163-175.

[12] T. Rezanka, H. Rezanková, *Anal. Chim. Acta* 398 (1999) 253-261.

[13] M. Tasioula-Margari, O. Okogeri, *Food Chem.* 74 (2001) 377-383.

[14] M. Brenes, A. García, J.J. Rios, P. García, A. Garrido, *Int. J. Food Sci. Tech.* 37 (2002) 615-625.

[15] M. Lees, *Food authenticity and traceability*, Woodhead publishing limited, Cambridge, 2003.

[16] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, *J. Chromatogr. A* 945 (2002) 221-230.

[17] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, L.R. Henriques, M.F. Peres, M.P. Simões, P.S. Lopes, *Anal. Bioanal. Chem.* 374 (2002) 1205-1211.

[18] I.J. Wesley, R.J. Barnes, A.E.J. McGill, *JAOCS* 72 (1995) 289-292.

[19] V. Baeten, M. Meurens, M.T. Morales, R. Aparicio, *J. Agric. Food Chem.* 44 (1996) 2225-2230.

[20] G. Downey, P. McIntyre, A.N. Davies, *J. Agric. Food Chem.* 50 (2002) 5520-5525.

[21] H. Yang, J. Irudayaraj, *JAOCS* 78 (2001) 889-895.

[22] Y.W. Lai, E.K. Kemsley, R.H. Wilson, *Food Chem.* 53 (1995) 95-98.

[23] M.J. Dennis, *Analyst* 123 (1998) 151R-156R.

[24] L. Küpper, H.M. Heise, P. Lampen, A.N. Davies, P. McIntyre, *Appl. Spectrosc.* 55 (2001) 563-570.

[25] V. Baeten, P. Dardenne, R. Aparicio, *J. Agric. Food Chem.* 49 (2001) 5098-5107.

[26] E.C. López-Díez, G. Bianchi, R. Goodacre, *J. Agric. Food Chem.* 51 (2003) 6145-6150.

[27] H.S. Tapp, M. Defernez, E.K. Kemsley, *J. Agric. Food Chem.* 51 (2003) 6110-6115.

[28] R. Sacchi, L. Mannina, P. Fiordiponti, P. Barone, L. Paolillo, M. Patumi, A. Segre, *J. Agric. Food Chem.* 46 (1998) 3947-3951.

[29] L. Mannina, A.P. Sobolev, A. Segre, *Spectrosc. Eur.* 15 (2003) 6-13.

[30] G. Vigli, A. Philippidis, A. Spyros, P. Dais, *J. Agric. Food Chem.* 51 (2003) 5715-5722.

[31] D.L. García-González, L. Mannina, M. D'Imperio, A.L. Segre, R. Aparicio, *Eur. Food Res. Technol.* 219 (2004) 545-548.

[32] A.D. Shaw, A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, D.B. Kell, *Anal. Chim. Acta* 348 (1997) 357-374.

[33] Y. González, M.C. Cerrato, J.L. Pérez, C. García, B. Moreno, *Anal. Chim. Acta* 449 (2001) 69-80.

[34] J. Brezmes, P. Cabré, S. Rojo, E. Llobet, X. Vilanova, X. Correig, *IEEE Sensors Journal* 5 (2005) 463-470.

[35] E. Psomiadou, M. Tsimidou, *J. Agric. Food Chem.* 50 (2002) 716-721.

[36] M. Deiana, A. Rosa, C.F. Cao, F.M. Pirisi, G. Bandino, M.A. Dessì, *J. Agric. Food Chem.* 50 (2002) 4342-4346.

[37] J. Gracián, *Analysis and characterization of oils, fats and fat products,* Vol. II, John Wiley & Sons, London, 1968.

[38] S. Lo Curto, G. Dugo, L. Mondello, G. Errante, M.T. Russo, *Ital. J. Food Sci.* 13 (2001) 221-228.

[39] D. Ryan, K. Robards, S. Lavee, *Int. J. Food Sci. Tech.* 34 (1999) 265-274.

[40] G.P. Cartoni, F. Coccioli, R. Jasionowska, D. Ramires, *Ital. J. Food Sci.* 12 (2000) 163-173.

[41] S.B. Engelsen, *JAOCS* 74 (1997) 1495-1508.

[42] N.B. Kyriakidis, P. Skarkalis, *J. AOAC Int.* 83 (2000) 1435-1439.

[43] D. Marini, L. Grassi, F. Balestrieri, E. Pascucci, *Riv. Ital. Sostanze Grasse* 67 (1990) 95-99.

[44] A. Sayago, M.T. Morales, R. Aparicio, *Eur. Food Res. Technol.* 218 (2004) 480-483.

[45] T.T. Ndou, I.M. Warner, *Chem. Rev.* 91 (1991) 493-507.

[46] E. Sikorska, A. Romaniuk, I.V. Khmelinskii, R. Herance, J.L. Bourdelande, M. Sikorski, J. Koziol, *J. Fluorescence* 14 (2004) 25-35.

[47] O.S. Wolfbeis, M. Leiner, *Mikrochim. Acta* 1 (1984) 221-233.

[48] S.M. Scott, D. James, Z. Ali, W.T. O'Hare, F.J. Rowell, *Analyst* 128 (2003) 966-973.

[49] M. Zandomeneghi, L. Carbonaro, C. Caffarata, *J. Agric. Food Chem.* 53 (2005) 759-766.

[50] L. Munck, L. Norgaard, S.B. Engelsen, R. Bro, C.A. Andersson, *Chemom. Intell. Lab. Syst.* 44 (1998) 31-60.

[51] R. Bro, *Chemom. Intell. Lab. Syst.* 46 (1999) 133-147.

[52] D. Baunsgaard, L. Norgaard, M.A. Godshall, *J. Agric. Food Chem.* 48 (2000) 4955-4962.

[53] D.K. Pedersen, L. Munck, S.B. Engelsen, *J. Chemometr.* 16 (2002) 451-460.

[54] R.D. Jiji, G.A. Cooper, K.S. Booksh, *Anal. Chim. Acta* 397 (1999) 61-72.

[55] R.D. Jiji, G.G. Andersson, K.S. Booksh, *J. Chemometr.* 14 (2000) 171-185.

[56] L. Moberg, G. Robertsson, B. Karlberg, *Talanta* 54 (2001) 161-170.

[57] K.S. Booksh, A.R. Muroski, M.L. Myrick, *Anal. Chem.* 68 (1996) 3539-3544.

[58] J.L. Beltrán, R. Ferrer, J. Guiteras, *Anal. Chim. Acta* 373 (1998) 311-319.

[59] M.L. Nahorniak, K.S. Booksh, *J. Chemometr.* 17 (2003) 608-617.

[60] J. Christensen, E. Miquel Becker, C.S. Frederiksen, *Chemom. Intell. Lab. Syst.* 75 (2005) 201-208.

[61] C. Wittrup, *J. Chemometr.* 14 (2000) 765-776.

Chapter 2

Theoretical Background

2.1 FLUORESCENCE SPECTROSCOPY

2.1.1 Background

Luminescence is the emission of light from any substance and occurs from electronically excited states. Molecular fluorescence is a type of luminescence in which the molecules emit light from excited singlet states after the absorption of visible or UV radiation (around 200-900 nm). In excited singlet states, the electron in the excited orbital is paired (of opposite spin) to the second electron in the ground-state orbital. Consequently, the return to the ground state is spin-allowed and occurs rapidly by the emission of a photon. Therefore, the emission rates of fluorescence are typically of 10^8 s^{-1}, so the typical fluorescence lifetime of a fluorophor (i.e. the average time between excitation and the return to the ground state) is around 10 ns [1,2].

The processes that occur between the absorption and the emission of light are illustrated by a Jablonski diagram [1] (Fig. 7). The singlet ground, first and second electronic states are depicted by S_0, S_1, and S_2, respectively. At each of these electronic energy levels, the fluorophors can exist in a number of vibrational energy levels (v = 0,1,2). Absorption typically occurs from molecules with the lowest vibrational energy. After light absorption, several processes usually occur. A fluorophor is usually excited to some higher vibrational level of either S_1 or S_2. Then there is a relaxation to the lowest vibrational level of S_1. This process, called internal conversion, generally occurs in 10^{-12} s or less. As fluorescence lifetimes are typically near 10^{-8} s, internal conversion is generally complete before the emission. Fluorescence emission therefore generally results from the lowest-energy vibrational state of S_1.

Molecules in the S_1 state can also undergo a spin conversion to the first triplet state, T_1 (i.e. the electron has the same spin as that in the ground-state orbital). Emission from T_1 is named phosphorescence and is generally shifted to longer wavelengths (lower energy) than fluorescence. Conversion of S_1 to T_1 is called intersystem crossing. Transition from T_1 to the singlet ground state is forbidden, so rate constants for triplet emission are several orders of magnitude smaller than those for fluorescence.
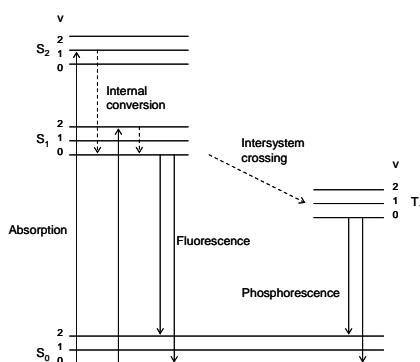
Figure 7. A simplified scheme of the Jablonski diagram.

Fluorescence typically occurs from aromatic molecules, due to low energy $\pi \rightarrow \pi^*$ transitions. Compounds containing the carbonyl group or structures with many conjugated double bonds can also present fluorescence.

Fluorescence intensity is a simultaneous function of the excitation and emission wavelengths. The intensity values of a fluorescence emission spectrum are obtained by keeping the excitation wavelength (λ_{ex}) constant while the emission wavelengths (λ_{em}) are scanned. Similarly, the intensity values of a fluorescence excitation spectrum are obtained by keeping the λ_{em} constant while the λ_{ex} are scanned [3]. When measuring several emission spectra at different λ_{ex} (or vice versa), a three-dimensional fluorescence landscape, the so-called fluorescence excitation-emission matrix (EEM), is obtained.

Several factors affect fluorescence. A rise in temperature leads to a decrease in fluorescence because there are more collisions between the molecules. A decrease in the viscosity of the solvent has the same effect. Solvents containing heavy atoms (e.g. carbon tetrabromide or ethyl iodide) also diminish fluorescence. When the fluorophor contains acid or basic substituents, its fluorescence may be affected by the pH of the solution. Another important factor is concentration. For highly concentrated solutions there are many collisions between the molecules. These collisions produce self-absorption phenomena (primary and secondary inner filter effects) [4]. Quenching is also a frequent problem in fluorescent measurements. This involves a reduction in fluorescence by a competing deactivating process resulting from the interaction between a fluorophor and another substance present

in the system. Oxygen is the best-known fluorescence quencher, but many other species may also act as such [1].

2.1.2 Spectra correction

The main components of a luminescence spectrometer are the source of light (or lamp) used to excite fluorescence, two grating monochromators (one in the path of the exciting light and the other in the path of the emitted fluorescence) and a detector to measure the intensity of fluorescence (Fig. 8). Because of their wavelength-dependence, these three components cause distortions in the fluorescence spectra. The light emitted by the lamp therefore cause distortions in the excitation spectra because the distribution of the intensity of this light depends on the wavelength. In addition, the efficiency of the monochromators and the sensitivity of the detector depend on the wavelength, which distorts the emission spectra [1,5]. To collect comparable spectra to those collected with other instruments, these distortions must be corrected. The instrument software itself includes correction factors to correct such deviations.
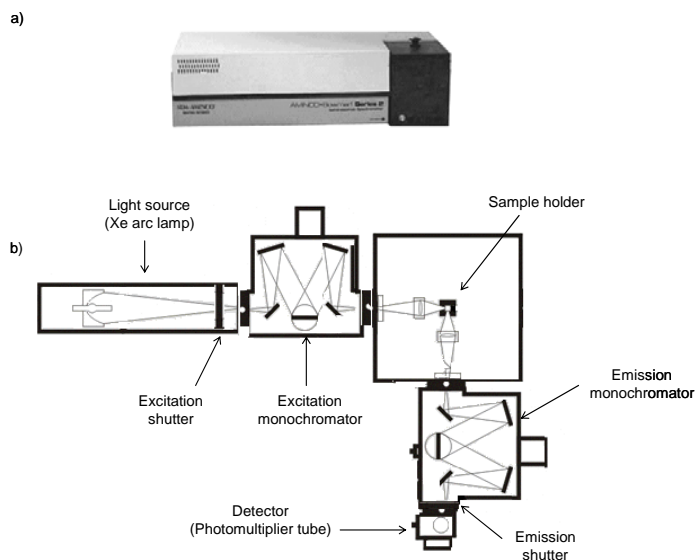

Figure 8. (a) Luminescence spectrometer Aminco Bowman Series 2, which has been used in this thesis. (b) Schematic representation of the instrumentation of a luminescence spectrometer.

We shall now describe the procedure for correcting excitation spectra. The output R of a photomultiplier tube is directly proportional to the flux of photons emitted by the excited sample. This relationship can be written as $R \propto I_0 \varepsilon \phi_f$, where I_0 is the excitation intensity, ε is the molar absorptivity, and ϕ_f is the *quantum yield* or *fluorescence efficiency* of the fluorophor (i.e. the ratio between the number of photons emitted and the number of photons used to excite the system). A plot of $\varepsilon \phi_f$ vs. the scanned wavelength is the true excitation spectrum. For many compounds, the quantum yield does not depend on the excitation wavelength. Therefore, the true excitation spectrum is a function of ε alone and is generally identical to the absorption spectrum. However, I_0 depends on the excitation source. The first step to correcting the excitation spectrum is to record the variation of I_0 with wavelength. I_0 is monitored using a reference channel (detector) equipped with a quantum counter. A tiny fraction of the excitation beam is directed to the reference detector. The quantum counter absorbs all this light and converts it to fluorescence with an efficiency of 100%. Any changes in the lamp output or monochromator will cause alterations in the output of the reference channel. Usually, because it covers a wide spectral range (220-600 nm), a solution of Rhodamine B in ethylene glycol (0.3%) is used as quantum counter. The intensity of the fluorescence of the uncorrected spectrum is then divided by the intensity of excitation I_0 at all the wavelengths over which the excitation spectrum is acquired, and a corrected excitation spectrum is obtained [5].

To determine the true emission spectrum, the following effects must be taken into account: the quantum efficiency of the detector, the bandwidth of the monochromator and the transmission efficiency of the monochromator. The dependence of the uncorrected emission spectrum on these effects can be expressed as

$$\frac{dF}{d\lambda} = \left(\frac{dI}{d\lambda}\right) P_\lambda B_\lambda M_\lambda = \left(\frac{dI}{d\lambda}\right) S_\lambda \tag{2}$$

where $dF/d\lambda$ is the apparent or observed intensity of fluorescence emission at wavelength λ; $dI/d\lambda$ is the true intensity at λ; P_λ, B_λ, and M_λ represent the relative quantum efficiency of the photomultiplier, the relative bandwidth of the monochromator, and the fraction of light transmitted by the monochromator, respectively, at λ. These last three factors have been combined into a single factor in S_λ, which is called spectral sensitivity factor of the monochromator-

photomultiplier combination. The true emission spectrum $dI/d\lambda$ can be calculated from the apparent emission spectrum by dividing each ordinate $dF/d\lambda$ by the corresponding value of S_λ. To determine S_λ, a lamp of known intensity distribution is used as excitation source. This intensity distribution is a function of wavelength and is represented as $(dI/d\lambda)_{lamp}$. The emission monochromator is illuminated with a fully collimated beam from this lamp and the response of the photomultiplier $(dF/d\lambda)_{lamp}$ is recorded as a function of wavelength λ. Since the spectral distribution of the lamp $(dI/d\lambda)_{lamp}$ is known, the quantity S_λ can be determined according to eq. (3) [5].

$$S_\lambda = \frac{(dF/d\lambda)_{lamp}}{(dI/d\lambda)_{lamp}} \qquad (3)$$

All the spectra presented in this thesis have been corrected as described above. Figure 9 shows the effect of correcting one excitation and one emission spectrum of an EVOO. We can see that correction changes the intensity of the fluorescence peaks considerably.
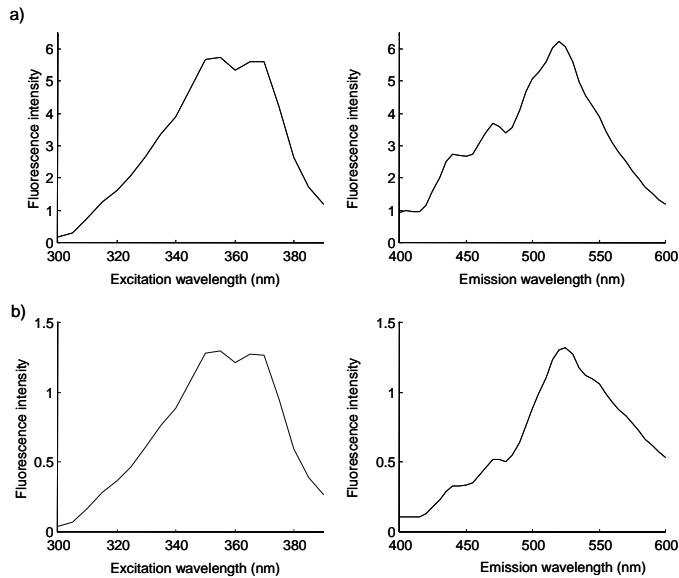

Figure 9. (a) Raw and (b) corrected spectra of an EVOO. Left: excitation spectra, right: emission spectra.

2.1.3 Instrumental settings

The function of a monochromator is to isolate a narrow wavelength range of light from a wide range. The excitation monochromator therefore selects a narrow wavelength range to excite the sample, and the emission monochromator selects a narrow wavelength range to detect the luminescent emission and reject unwanted light. The band-pass is the instrument setting used to simultaneously adjust both the range of wavelengths and the amount of light that passes through a monochromator. As the band-pass is increased, so is the sensitivity of measurement, because more light passes through the sample. However, selectivity (i.e. the resolution of the peaks) is reduced because the wavelength range of light is widened [1].

Besides the band-pass, another parameter to set up is the step size of the monochromators (i.e. the frequency of the measurements). The step size determines the number of points to be recorded and influences the time of analysis. Figure 10 shows the EEMs of the same EVOO obtained using different step sizes. Both matrices were recorded in the λ_{ex} = 300-390 nm; λ_{em} = 400-600 nm range but in Figure 10a it was recorded every 2 nm in the excitation domain and every 1 nm in the emission domain and in Figure 10b it was recorded every 5 nm in both dimensions.
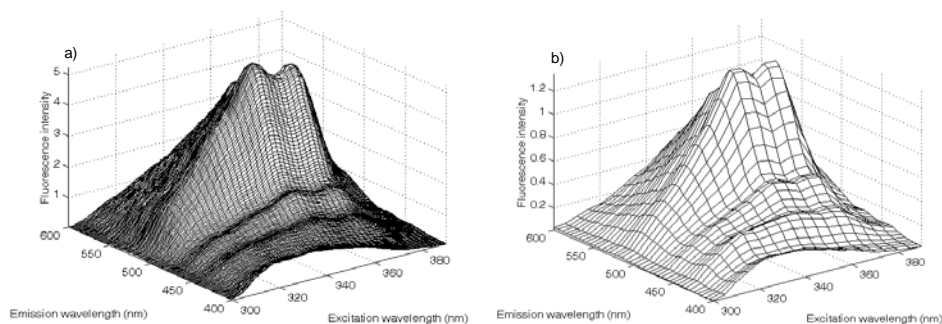

Figure 10. EEMs of an EVOO recorded at different step sizes.

Obviously, some information is always lost when the step size is increased. However, the main sources of variation are kept in the data (see Figure 10). Also, the lower number of points to be recorded helps to speed up the analysis. As an example, Table 3 shows the time of analysis needed at different step sizes (range recorded and scan rate kept constant).

Table 3. Time of analysis at different step sizes (range: λ_{ex} = 300-390 nm; λ_{em} = 400-600 nm, scan rate = 30 nm/s).

excitation increment (nm)	emission increment (nm)	time of analysis
1	1	11 min 38.8 s
1	2	5 min 53.3 s
5	5	2 min 28.8 s

2.1.4 Range selection and Rayleigh scatter

When deciding which range to include in an EEM, one has to bear in mind that emission wavelengths below the excitation wavelength do not exhibit any fluorescence because the emitted energy is always lower than the excitation energy.

Scattering is produced by small particles in the samples and makes the light deviate from its original path and spread in all directions. Scattering provides no information about the fluorescence properties of the sample and must be removed. Rayleigh scatter is often the most difficult type of scattering to deal with. In fluorescence EEMs, Rayleigh scatter produces a diagonal line across the landscapes [2] (see Figure 11a-b). This line always appears at emission wavelengths equal to excitation wavelengths (first-order Rayleigh). Rayleigh scatter appears again at emissions obtained at wavelengths around twice the excitation wavelength (second-order Rayleigh).

There are different ways of handling Rayleigh scatter e.g. inserting missing values or zeros on the scattered area or plainly avoiding the part of the matrix that includes the scatter. However, this can only be done when the removed wavelengths contain little or no information [6]. With olive oils, most of the

information is placed at emissions above 400 nm. For this reason, we have not considered emission wavelengths below 400 nm in this thesis (Fig. 11c).

2.1.5 Fluorescence of olive oils

The fluorescence that olive oils emit in their native form has been related to species such as vitamin E, oxidation products and chlorophylls [7] (Fig. 11a and 11c). As Figure 11a shows, in EVOOs, the chlorophyll peak (λ_{em} = 600-700 nm) is much more intense than those of the other fluorescent species. The different magnitude of this peak may cause problems when handling the data and, for this reason, the fluorescent region of chlorophylls has been often removed in this thesis. When included, preprocessing methods were applied so as to reduce these differences of magnitude.
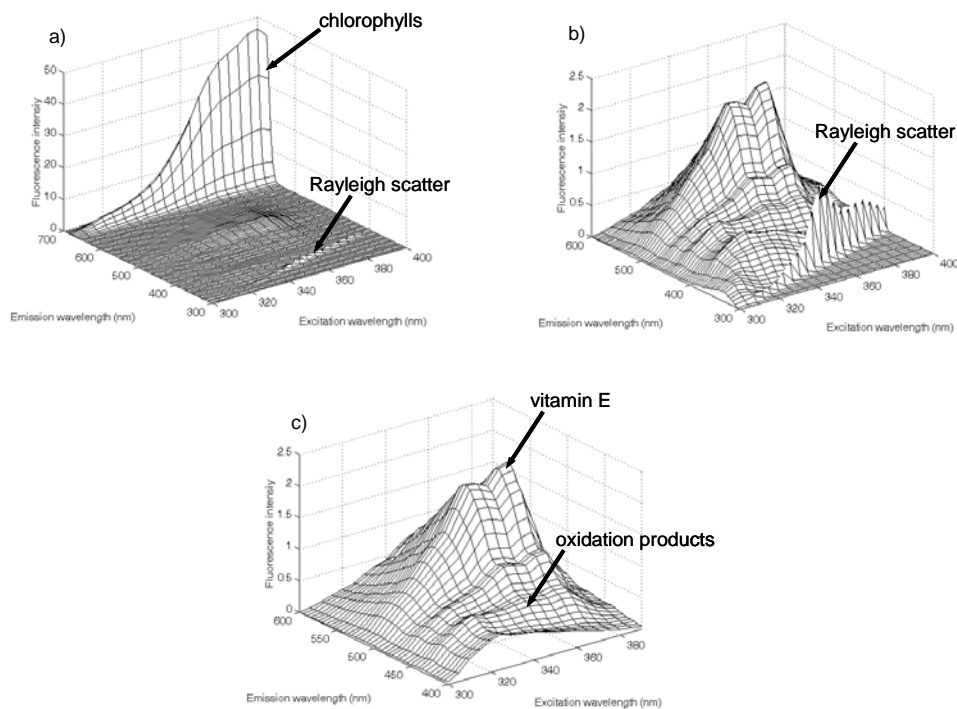

Figure 11. EEMs of an EVOO at different ranges: (a) λ_{ex} = 300-400 nm, λ_{em} = 300-695 nm, (b) λ_{ex}= 300-400 nm, λ_{em}= 300-600 nm and (c) λ_{ex}= 300-390 nm, λ_{em}= 400-600 nm. (a) and (b) present Rayleigh scatter.

The various grades of olive oils have compositional differences that can be detected from their fluorescence EEMs. There are also great differences between the EEMs of olive oils and those of other vegetable oils. These EEMs of vegetable oils can therefore be used as fingerprints for oil characterization. As an example, Figure 12 shows the fluorescence EEMs of an EVOO, a POO, an OPO and a sunflower oil, in the λ_{ex} = 300-390 nm; λ_{em} = 400-600 nm range. The larger vitamin E content in EVOOs (Fig. 12a) causes the higher fluorescence intensity of the spectra of these samples around λ_{em} = 525 nm. On the other hand, oils that underwent refining processes (Fig. 12 b-d) present a broad peak at around λ_{em} = 450 nm, which is due to large quantities of oxidation products.
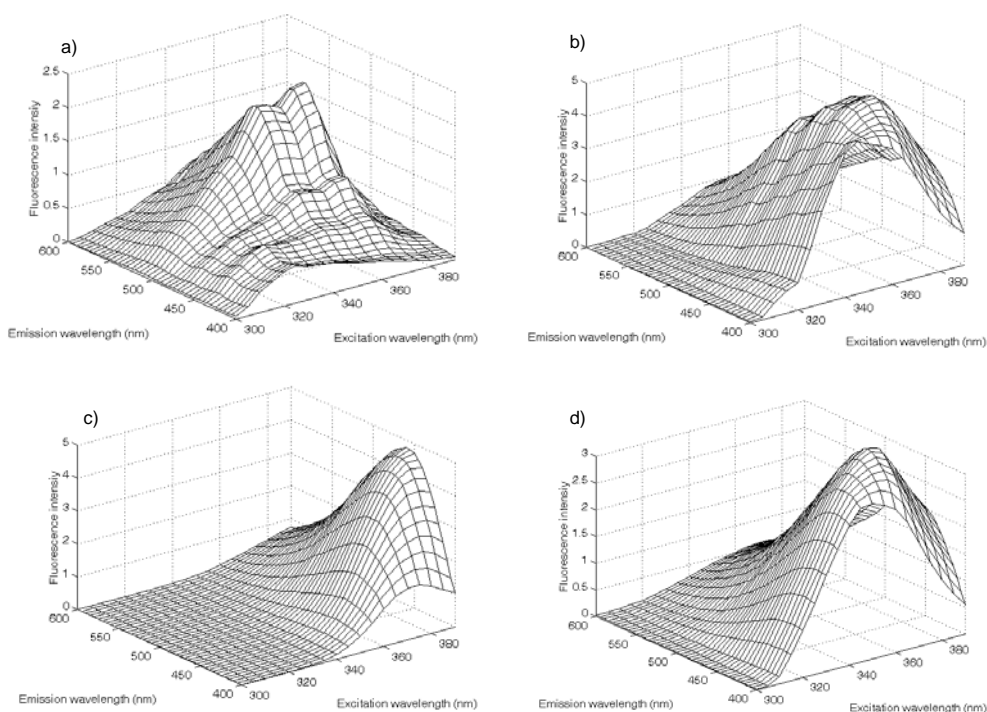

Figure 12. Fluorescence EEM of an EVOO (a), a POO (b), an OPO (c), and a sunflower oil (d) in the λ_{ex} = 300-390 nm; λ_{em} = 400-600 nm range.

As we stated earlier, measuring highly concentrated samples may cause the appearance of inner filter effects. This especially occurs when working with right-

angle geometries (i.e. when the detector is oriented 90º with respect to the lamp) [1,8]. Inner filter effects involve the attenuation of the emission intensity due to the absorption of the incident excitation light and to the absorption of the emitted light [1,9]. This leads to distortions in the intensity and shape of the spectra. To overcome this problem, front-face geometries are often used to measure the fluorescence spectra of undiluted olive oils [9,10] because front-face fluorescence spectra are much less affected by inner filter effects. Another way to reduce these distortions is to dilute the sample with an appropriate solvent. As an example, Figure 13 shows the changes to the spectrum of an EVOO upon dilution with n-hexane at different proportions. Despite their possible distortions, however, the spectra of undiluted oils may be of interest for the differentiation and characterization of oils. Another advantage of working with undiluted oils is that quenching (i.e. reduction in fluorescence due to the interaction between the fluorophor and another substance) is reduced because there are fewer collisions between the fluorophor and oxygen or other quenchers in viscous samples than in diluted solutions [9]. As we can see in Figure 12, directly measuring EEMs from undiluted oils provides useful information about the samples and may help to detect differences between different types of oils. For this reason, all the results in this thesis were obtained from undiluted oils.
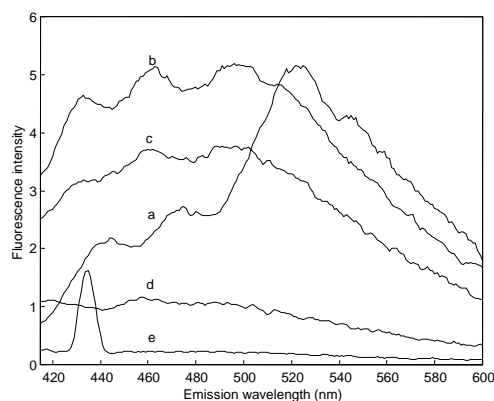

Figure 13. Fluorescence spectra of an EVOO between λ_{em} = 415-600 nm at λ_{ex} = 345 nm. (a) undiluted oil, and dilutions (b) 1:3, (c) 1:6, (d) 1:30, and (e) 1:150 in n-hexane.

2.2 THREE-WAY DATA

2.2.1 Notation

In this section we summarise the mathematical notation used in this thesis, which is the one commonly accepted by the scientific community [11].

Italic lowercase letters (e.g. x) indicate scalars (i.e. zero-order data), bold lowercase letters (e.g. $\mathbf{x}$) indicate vectors (i.e. first-order data), bold uppercase letters (e.g. $\mathbf{X}$) indicate matrices (i.e. second-order data), and underlined bold uppercase letters (e.g. $\underline{\mathbf{X}}$) indicate three-way arrays. Transposition of a vector or matrix is symbolized by a superscripted "T" (e.g. $\mathbf{X}^{\mathrm{T}}$). A superscripted "-1" (e.g. $\mathbf{X}^{-1}$) indicates the inverse matrix. Finally the symbol $|\otimes|$ symbolizes the Khatri-Rao product of two matrices ($\mathbf{A}$ and $\mathbf{B}$) with the same number of columns (J). The Khatri-Rao product is defined as

$$\mathbf{A}|\otimes|\mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{a}_J \otimes \mathbf{b}_J] \tag{4}$$

where $\otimes$ is the Kronecker product, which is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ \cdots & \cdots & \cdots \\ a_{I1}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{pmatrix} \tag{5}$$

2.2.2 Three-way arrays

Any set of data for which the elements can be arranged as

$$x_{ijk\dots} \qquad i=1,\dots,I,\ j=1,\dots,J,\ k=1,\dots,K,\ \dots \tag{6}$$

where the number of indices may vary, is a multi-way array. When three dimensions are used, the data can be arranged in a three-way array, where the *ith* index refers to the rows, the *jth* index refers to the columns and the *kth* index refers to the third dimension (or tubes) (Fig. 14). Each dimension of this array is named way or mode and the number of levels in the mode is named the dimension of the mode [12]. Any set of second-order data can be arranged in a three-way array.

Thus, a set of EEMs obtained by measuring different samples can be arranged in a three-way array where, for example, samples are in the first mode, λ_{em} are in the second mode and λ_{ex} are in the third mode (Fig. 14).
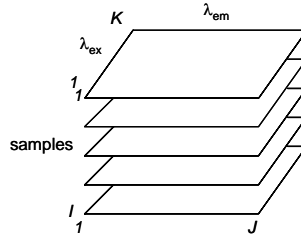
Figure 14. Three-way array.

2.2.3 Preprocessing

Unfolding

Unfolding (or matricization) involves rearranging a three-way array to form a matrix [11,12]. This is done by combining two of the original modes while keeping the other fixed. Unfolding a three-way structure of $I \times J \times K$ dimensions can be done in three different ways, depending on which modes are combined: the second and the third ($I \times JK$), the third and the first ($J \times KI$) or the first and the second ($K \times IJ$). Figure 15 shows graphically the process of unfolding a three-way array by concatenating the second and third modes. This type of unfolding has been used in this thesis in order to keep the sample mode (mode 1) and to combine the spectral modes (modes 2 and 3).

Figure 15. Unfolding of a three-way array by combining the second and the third modes.

Several authors have used unfolding as a method to handle three-way data [12-24]. However, although unfolding may often be helpful, it has several disadvantages. For example, the three-w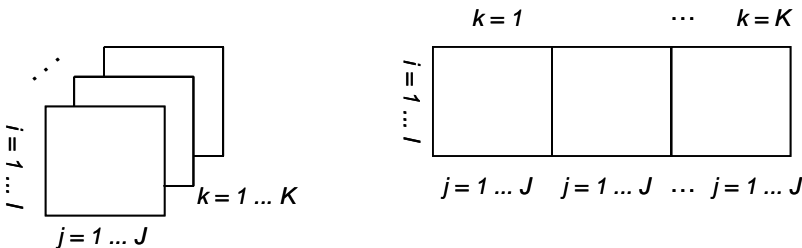ay structure of the data is lost, the number of variables of the resulting matrix is high because of the combination between the variables of two of the modes, and the resulting models are more difficult to interpret [12].

Centering

To remove a constant offset, the data can be translated along the coordinates' origin. In two-way data (matrices), the most common procedure is column mean-centering, where every datum x_{ij} is centered by subtracting the column mean $\bar{x}_j$ according to:

$$x_{ij(cent)} = x_{ij} - \bar{x}_{ij} \tag{7}$$

where i is the row index, j is the column index, and $\bar{x}_j$ is the column mean calculated from:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \tag{8}$$

Centering a three-way array may be applied across any of the modes. For example, centering the first mode can be done by unfolding the array to a $I \times JK$ matrix (Fig. 16), and then column mean-centering as described above. If centering is to be performed across more than one mode, one has to do this by first centering one mode and then centering the outcome of this centering [13].

Scaling

Scaling is done to eliminate differences of magnitude between variables or samples. There are various scaling methods. One of the most common methods is
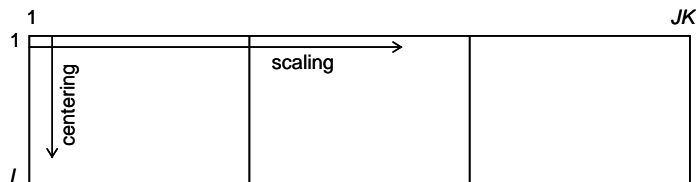

Figure 16. Scheme of a three-way unfolded array, showing the directions along which scaling and centering should be performed.

standardization, which involves dividing each element of a matrix by its column standard deviation:

$$x_{ij(st)} = \frac{x_{ij}}{s_j} \tag{9}$$

Another scaling method is autoscaling. In this case, the columns are mean-centered and then divided by their standard deviation. Autoscaled data have column mean zero and column unit variance [25]:

$$x_{ij(autosc)} = \frac{x_{ij} - \overline{x}_i}{s_j} \tag{10}$$

$$\text{where} \qquad s_j = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}{n-1}} \tag{11}$$

Another scaling method is the normalization of a data vector to length one. This is done by dividing the vector by the Euclidean norm ($\left\| \mathbf{x}_j \right\|$). This procedure is described below:

$$x_{ij(norm)} = \frac{x_{ij}}{\left\| \mathbf{x}_j \right\|} \tag{12}$$

$$\text{where} \qquad \left\| \mathbf{x}_j \right\| = \sqrt{x_{1j}^2 + x_{2j}^2 + ... + x_{nj}^2} \tag{13}$$

Normalization is performed in order to remove systematic variation, which is usually associated with the total amount of sample [26]. In this thesis, this method has been used to reduce differences in intensity between the fluorescence spectra of different oils.

In three-way analysis, scaling must be done on the rows of a matrix and not across the columns, as is the case with centering (i.e. one has to scale whole matrices

instead of columns) (Fig. 16). For example, to remove differences in magnitude between the λ_{em} of a set of EEMs, one can scale the three-way array within the emission mode. To do this, the three-way array has to be unfolded in such a way that λ_{em} (variable j) is in the rows and then the rows are scaled. This is because all the columns where variable j occurs have to be scaled [13]. Not all combinations between centering and scaling are possible when working with three-way data. Only centering across arbitrary modes or scaling within one mode is straightforward [13].

2.3 CHEMOMETRIC METHODS

In this section we provide an overview of the chemometric methods used in this thesis. These methods are grouped into two blocks. The first block contains methods used for exploratory analysis (unfold principal component analysis (unfold-PCA or U-PCA), non-negative matrix factorization (NMF), parallel factor analysis (PARAFAC) and hierarchical cluster analysis (HCA)). The second block contains methods used for discrimination and classification purposes: Fisher's linear discriminant analysis (LDA), discriminant unfold partial least squares (DU-PLSR) regression and discriminant multi-way partial least squares (DN-PLS) regression.

2.3.1 Exploratory analysis

Exploratory analysis involves applying chemometric methods to find patterns in the data, i.e. to better understand the structure of the data. No previous knowledge of the data is used to find such patterns. For this reason, these methods are often called unsupervised pattern recognition methods [27]. Some typical aims are to reveal groups and trends in the data, to find relationships between samples and variables and to detect outliers. The results of exploratory analysis are often presented in graphical displays, which makes it easy to interpret the information.

Unfold Principal Component Analysis (unfold-PCA)
Unfold-PCA involves computing principal component analysis (PCA) on a matrix ($\mathbf{X}$) obtained after unfolding a three-way array ($\underline{\mathbf{X}}$). This method is also named multi-way principal component analysis [18,19] or Tucker1 [14]. The aim of PCA is

to retain the main information contained in the original variables in a smaller number of variables, called principal components (PCs), which describe the main variations in the data. These PCs are linear combinations of the original variables [26]. Some properties of PCs are that they are orthogonal (i.e. uncorrelated to each other), hierarchical (i.e. the first PC retains the maximum information of the data, the second PC retains the maximum information that is not included in the first one, and so on), and are calculated sequentially (i.e. the *F-1* component model is a subset of the *F* component model).

Unfold-PCA involves first unfolding a three-way array in any of the three modes and then computing the PCA model, which makes a bilinear decomposition of the $\mathbf{X}$ matrix into a score matrix ($\mathbf{T}$) and a loading matrix ($\mathbf{P}$), which describe the original data in a more condensed way (eq. 14). If unfolding has been done by keeping the sample mode, $\mathbf{T}$ will contain information about the samples and $\mathbf{P}$ will contain information about the variables (see Fig. 15). The residuals (i.e. the difference between the original and the reconstructed matrix with the calculated PCs) are collected in the $\mathbf{E}$ matrix (Fig. 17):

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \tag{14}$$

The *I*, *J* × *K* and *F* indices in Figure 17 stand for samples, variables and PCs, respectively.

Unfold-PCA is a powerful tool for obtaining an overview of second-order data, such as fluorescence EEMs [15]. After reducing dimensionality, it allows us to find groups and trends in the data and provides information about what variables are responsible for the main variations in the data. In this thesis, unfold-PCA has been used as a method for exploratory analysis. It has also been used to reduce dimensionality and so that the scores can be used as inputs for subsequent discrimination methods.

One of the main disadvantages of the methods used after unfolding is that the loadings are very difficult to interpret because the variables in the unfolded modes get mixed up. To overcome this problem, we can fold back the loadings of each PC into a matrix with the same dimensions of the matrices that formed the original

Figure 17. Principal component decomposition.

three-way array. This process is often called refolding [18] and provides more interpretable loadings. However, though refolding makes it easier to visualize the loadings, note that such loadings do not correspond to the initial three-way structure because the combination of two of the modes in the unfolding process causes the loss of the three-way structure. Figure 18 shows an example of how refolding improves the understanding of the unfold-PCA loadings. In this example, the dimensions of $\underline{\mathbf{X}}$ were $49 \times 41 \times 21$ (samples $\times$ number of λ_{em} $\times$ number of λ_{ex}). After unfolding the array by combining the spectral modes, an $\mathbf{X}$ matrix of size 49×861 was obtained. Unfold-PCA was then computed on the unfolded matrix and each PC was a row vector with 861 elements. Figure 18a shows the loadings of the first PC. If the PC loadings are reshaped into the same dimensions of the original EEMs (41×21), they can be plotted as shown in Figure 18b. As we can see, the refolded loadings are easier to interpret.
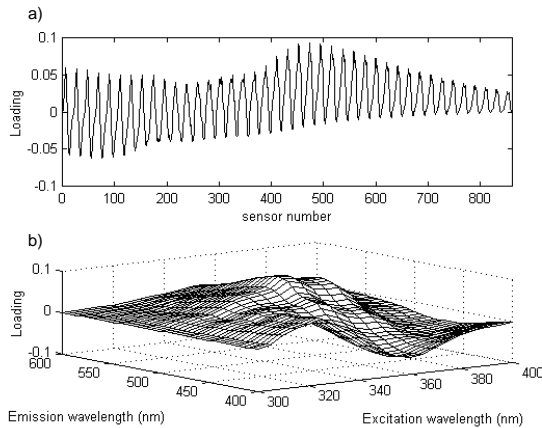

Figure 18. (a) Unfolded and (b) refolded loadings of unfold-PCA.

From any PCA model, two statistics can be defined: Hotelling T^2 and Q. The Hotelling T^2 statistic is calculated for the systematic part of the variation of the data. The Hotelling T^2 plot therefore represents the projection of each measurement onto the hyperplane defined by the PCs of the model. For each sample i, T^2 is defined as

$$T_i^2 = \mathbf{t}_i \boldsymbol{\lambda}^{-1} \mathbf{t}_i^T = \mathbf{x}_i \mathbf{P} \boldsymbol{\lambda}^{-1} \mathbf{P}^T \mathbf{x}_i^T \tag{15}$$

where $\mathbf{t}_i$ refers to the *ith* row of the score matrix $\mathbf{T}$ from eq. 14, $\mathbf{P}$ is the loading matrix from eq. 14 and $\boldsymbol{\lambda}^{-1}$ is the diagonal matrix containing the inverse of the eigenvalues associated with the f PCs retained in the model. Alternatively, the Q statistic is calculated for the residual part (i.e. the part not included in the model). Q is the sum of squares of each sample of $\mathbf{E}$ from eq. 14, i.e., for the *ith* sample in $\mathbf{X}$, $\mathbf{x}_i$:

$$Q_i = \mathbf{e}_i \mathbf{e}_i^T = \mathbf{x}_i (\mathbf{I} - \mathbf{P}_f \mathbf{P}_f^T) \mathbf{x}_i^T \tag{16}$$

where $\mathbf{e}_i$ is the *ith* row of $\mathbf{E}$, $\mathbf{P}_f$ is the matrix of the f loading vectors retained in the PCA model and $\mathbf{I}$ is the identity matrix. The Q plot represents the squared distance of each new measurement perpendicular to the plane defined by the PCs.
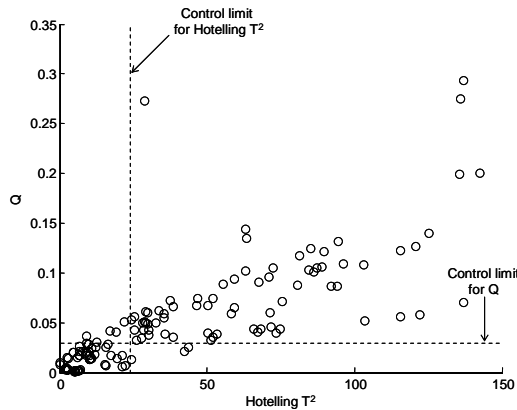

Figure 19. Plot of Hotelling T^2 and Q statistics.

These statistics have been widely applied in the statistical process control of multivariate processes [17,28]. Confidence limits for both parameters can be defined from a training set and then the Hotelling T^2 and Q values for new observations can be computed. Those lying outside the control limits are considered to be out of control in the space spanned by the PCs and the residuals, respectively (Fig. 19). In this thesis, the Hotelling T^2 and Q statistics have been applied to unfold-PCA as a fast screening method for discrimination purposes.

Model selection

In any PCA model, one has to decide how many PCs to include in the model. We shall now briefly describe the methods used in this thesis for choosing the number of PCs in the PCA models.

- Cross-validation (CV)

In this procedure, a subset of samples (the cancellation group) is removed from the data set and a PCA model is built. The residuals of the left out samples (i.e. the difference between the measured and the estimated values) are then calculated. The subset of samples is returned to the data set and the process is repeated for different subsets of samples until each sample has been excluded from the data set once. Finally, the root mean square error of cross-validation (RMSECV_PCA) is computed [26]. This error is calculated for different numbers of PCs:

$$\text{RMSECV\_PCA}_f = \sqrt{\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}(residual_f(i,j))^2}{I}} \tag{17}$$

In eq. (17), $RMSECV\_PCA$ is a measure of the magnitude of the residuals of the new samples projected onto the PCA with f PCs. The term $residual_f(i,j)$ is the residual for the *ith* sample and the *jth* variable.

Leave-one-out cross-validation (or full cross-validation) is a particular case of CV in which one sample is left out at a time. This procedure has been widely used in this thesis.

- Double cross-validation (DCV)

In 1978 Wold introduced DCV as an alternative method to CV for determining the number of significant components of PCA models [29]. In DCV the cancellation groups are not samples, as in CV, but points in the data matrix [30]. DCV uses the NIPALS (non-linear iterative partial least squares) algorithm [31], since it can compute loadings even in the case of missing data. The DCV procedure is described below [29,30]:

1. Take $\mathbf{X}$, which is the original data matrix properly preprocessed.

2. Compute the sum of squares of $\mathbf{X}$ (or residual squared error (RSE)):

$$RSE^f = \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}^2 \tag{18}$$

where I and J indicate the number of cancellation groups and variables, respectively, and f indicates the number of PCs (at the beginning, $f = 0$).

3. Create a cancellation matrix, with as many rows and columns as $\mathbf{X}$, and divide it into *G* groups by numbering it from 1 to *G*. (Normally, to reduce computer time, *G* is chosen to be lower than the total number of data). A common approach used to create the cancellation matrix is to delete values on the diagonals of $\mathbf{X}$ [30], but random number schemes may also be employed [27]. Figure 20 shows an example of a cancellation matrix with 5 randomly chosen cancellation groups (numbered from 1 to 5). Data in $\mathbf{X}$ whose corresponding position in the cancellation matrix is filled with **1** will make up the first cancellation group.

$$
\begin{pmatrix}
1\,2\,4\,2\,\mathbf{1}\,\mathbf{1}\,2\,5\,\mathbf{1}\,3\,4\,\mathbf{1}\,2\,5\,4\,2\,3\,3\,\mathbf{1}\,2\,4 \\
3\,5\,3\,2\,2\,3\,4\,2\,5\,4\,3\,\mathbf{1}\,2\,3\,\mathbf{1}\,4\,5\,3\,4\,\mathbf{1}\,5 \\
\mathbf{1}\,4\,3\,3\,4\,5\,\mathbf{1}\,\mathbf{1}\,3\,5\,2\,2\,\mathbf{1}\,4\,3\,5\,\mathbf{1}\,5\,2\,3\,4 \\
5\,4\,4\,5\,\mathbf{1}\,2\,2\,4\,2\,\mathbf{1}\,5\,4\,3\,2\,\mathbf{1}\,5\,2\,3\,4\,\mathbf{1}\,2 \\
\mathbf{1}\,2\,3\,4\,4\,3\,\mathbf{1}\,3\,4\,5\,\mathbf{1}\,3\,2\,\mathbf{1}\,4\,3\,5\,4\,2\,\mathbf{1}\,3
\end{pmatrix}
$$
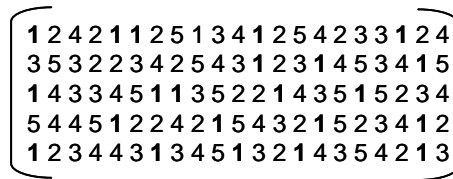
Figure 20. Example of a cancellation matrix for DCV with 5 cancellation groups. The numbers indicate the membership of the data points in a cancellation group (e.g. all positions labelled **1** constitute the first cancellation group).

4. Remove the first cancellation group ($\mathbf{X}_1$) of $\mathbf{X}$ to obtain $\mathbf{X}_M$, which is the $\mathbf{X}$ matrix with missing values in the positions marked with **1**. Then compute PC1 from $\mathbf{X}_M$.

5. Predict $\mathbf{X}_1$ from PC1 ($\mathbf{X}_{1p}$) .

6. Compute the prediction error (residual matrix)

$$\mathbf{E}_1 = \mathbf{X}_1 - \mathbf{X}_{1p} \tag{19}$$

and the sum of squares of the residual matrix to give the partial predicted residual error sum of squares ($PRESS_P$):

$$PRESS_P = \sum \mathbf{E}_1^2 \tag{20}$$

7. Restore $\mathbf{X}$ and successively delete the G cancellation groups, each time calculating $PRESS_P$ as in step 6.

8. When all G groups have been deleted once, and so all elements in $\mathbf{X}$ have been deleted once and only once, sum the partial PRESS to obtain the total PRESS ($PRESS_T$).

9. Compute the ratio $R = PRESS_T / RSE^f$. $R < 1$ shows that the predictions are improved by including the latest PC.

10. Compute PCA on the complete matrix $\mathbf{X}$, reconstruct it from PC1 and compute the new residual matrix $\mathbf{E}$. Increase f by one and go back to step 2.

11. $R > 1$ indicates that the latest PC ($f + 1$) did not improve the prediction errors and the best value is f .

Non-negative matrix factorization (NMF)

Although PCA is the most common method for compressing multivariate data, the mathematical property of orthogonality implies having both negative and positive loadings and scores. This does not have much chemical or physical sense for signals that are inherently non-negative.

The NMF algorithm [32] is another method for compressing multivariate data that uses non-negativity constraints. These constraints mean that only positive values can be obtained for the decomposed data. Also, with NMF we can obtain parts-based representations of the data. This is because of the non-negative constraints of the algorithm, which allow only additive, not subtractive, combinations. In some cases, therefore, NMF may be more suitable than PCA because it provides a more realistic interpretation of the data.

The NMF algorithm works as described below. To apply the method, the $\mathbf{X}$ matrix of dimensions $I \times J$ (samples $\times$ variables) has to be transposed to obtain a $\mathbf{V}$ matrix of dimensions $J \times I$ (i.e. each column of $\mathbf{V}$ is the vector data corresponding to one sample). This matrix is then approximately decomposed into a $J \times F$ matrix $\mathbf{W}$ and a $F \times I$ matrix $\mathbf{H}$. The NMF algorithm builds approximate decompositions of the form:

$$\mathbf{V}_{ji} \approx (\mathbf{WH})_{ji} = \sum_{f=1}^{F} \mathbf{W}_{jf} \mathbf{H}_{fi} \tag{21}$$

where the f columns of $\mathbf{W}$ are called the basis set, each column of $\mathbf{H}$ is called an encoding, and F is the rank of the factorization. NMF does not allow negative entries in the matrices $\mathbf{W}$ and $\mathbf{H}$.

The NMF algorithm proposed by Lee and Seung [32] starts by randomly initializing matrices $\mathbf{W}$ and $\mathbf{H}$, which are iteratively updated to minimize the objective function

$$D = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[\mathbf{V}_{ji} \log(\mathbf{WH})_{ji} - (\mathbf{WH})_{ji} \right] \tag{22}$$

subject to non-negative constraints. Different update rules can be applied to minimize the objective function [33]. In this thesis we have used the divergence-based update equations [34]:

$$\mathbf{H}_{fi} \leftarrow \mathbf{H}_{fi} \frac{\sum_{j} \mathbf{W}_{jf} \mathbf{V}_{ji} / (\mathbf{WH})_{ji}}{\sum_{k} \mathbf{W}_{kf}} \qquad \mathbf{W}_{if} \leftarrow \mathbf{W}_{if} \frac{\sum_{i} \mathbf{H}_{fi} \mathbf{V}_{ji} / (\mathbf{WH})_{ji}}{\sum_{v} \mathbf{H}_{fv}} \tag{23}$$

To represent a new vector using a predefined set of basis functions, the same algorithm is iterated without modifying the matrix factor $\mathbf{W}$. So, fixing $\mathbf{W}$ and starting with positive random $\mathbf{H}$ obtains a representation of a new data vector according to the basis defined in $\mathbf{W}$.

Some researchers have modified the original algorithm by, for example, adding constraints [35] and modifying the initialization step [36].

NMF has been successfully applied to many areas, such as image [32] and text analysis [32,37,38], genetics [34], spectral analysis [36,39], and sound analysis [40,41]. It has also been used to resolve overlapped chromatograms and mass spectra [35].

Parallel Factor Analysis (PARAFAC)

PARAFAC is a decomposition method for multi-way data that works directly on the multi-way structure. This decomposition is made into triads or trilinear components. Each component consists of three loading vectors that are treated mathematically in the same way [13]. However, the vector associated with the sample mode is often named the score vector. A PARAFAC model of a three-way array is therefore given by three loading matrices, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, with elements a_{if}, b_{jf} and c_{kf}, where the i, j and k indices refer to the dimensions of the array and f is the number of factors of the model. Figure 21 is a graphical representation of the PARAFAC decomposition. The cube $\underline{\mathbf{E}}$ contains the residuals.
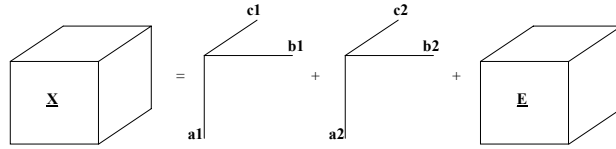

Figure 21. Representation of a two-component PARAFAC model of the data array $\underline{\mathbf{X}}$.

Mathematically, the PARAFAC model with *F*-components is expressed as [13]:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{24}$$

The model can also be expressed in matrix form as [12]:

$$\mathbf{X_a} = \mathbf{A}(\mathbf{C}|\otimes|\mathbf{B})^{\mathrm{T}} + \mathbf{E_a} \tag{25}$$

where $\mathbf{X_a}$ $(I \times JK)$ is the unfolded matrix along mode A from the array $\underline{\mathbf{X}}$, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the loading matrices and $\mathbf{E_a}$ $(I \times JK)$ is the unfolded residual matrix.

Unlike bilinear methods such as PCA, PARAFAC provides unique solutions. This means that, apart from trivial variations of scale and column order, no restrictions are needed to estimate the model [12,13]. Another difference with respect to PCA is that PARAFAC is a non-nested model (i.e. it cannot be calculated sequentially). This means that the *F-1* factor model is not a subset of the *F* factor model.

There are several ways to determine the correct number of components in PARAFAC [13]. In this thesis, we used split-half analysis and residual analysis.

Split-half analysis involves dividing the data into two halves and then making a PARAFAC model on both halves. Because of the uniqueness of the PARAFAC model, if the correct number of components is used, the same loadings in the non-split modes on both data sets should be obtained [13].

Residual analysis can be performed as in bilinear models. Systematic variation left in the residuals indicates that more components can be extracted. If a plot of the residual sum of squares against the number of components sharply flattens out for a certain number of components, this indicates that the right number of components has been reached [13].

PARAFAC can be used for quantitative purposes because the score matrix **A** contains information about the differences of composition between the samples. In this thesis we have applied multiple linear regression (MLR) to the PARAFAC scores in order to build regression models between them and some of the quality parameters of olive oils. MLR is a generalisation of univariate linear regression that is used when several independent predictor variables x are to be used to predict y [42]. The model can be written as:

$$y = b_0 + \sum_{k=1}^{K} b_k x_k + e \qquad (26)$$

where b_0 and b_k are the regression coefficients and e is the random error.

The MLR model can also be expressed in matricial form as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \qquad (27)$$

The regression coefficients of the model are calculated as

$$\mathbf{b} = \mathbf{X}^+\mathbf{y} \tag{28}$$

where $\mathbf{X}^+$ is the pseudoinverse matrix ($\mathbf{X}^+ = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$), in the case that $\mathbf{X}$ is non-singular and $I > J$.

In recent years, PARAFAC has been widely used in food analysis [12,13,15,43-48].

Hierarchical cluster analysis (HCA)

Cluster analysis (CA) is an unsupervised pattern recognition method used to find natural groups in multivariate data according to the similarity of the variables of the samples [25,42]. To perform CA, one must measure the similarity (or dissimilarity) of the samples. There are several measures of similarity [49]. In this thesis, we have used Euclidean distance. The equation of Euclidean distance between samples *i* and *i'* is

$$D_{ii'} = \sqrt{\sum_{j=1}^{J}(x_{ij} - x_{i'j})^2} \tag{29}$$

where J is the number of variables.

Hierarchical agglomerative methods are widely used in CA. These start from as many clusters as there are samples. Gradually, samples are joined into clusters, up to the final cluster with all the samples [50]. HCA is represented graphically by a tree called a dendrogram (Fig. 22).

Computing the distances between all pairs of samples yields the dissimilarity matrix, which is a symmetrical matrix containing these distances. There are several algorithms to compute the hierarchical clustering from the dissimilarity matrix [49,51]. In this thesis we used the average linkage algorithm, which involves computing the average of the distances of the samples to be merged into a new cluster.
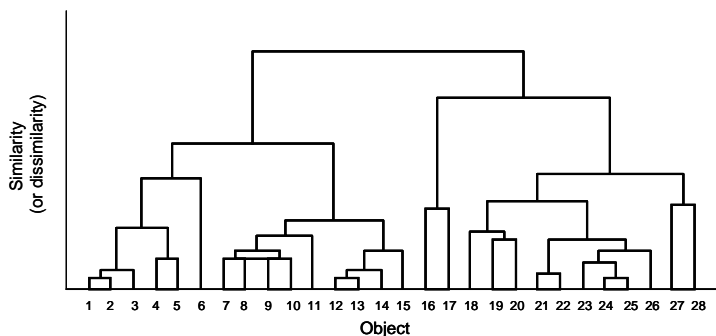
Figure 22. Hierarchical clustering representation (dendrogram).

One of the most important issues in CA is the evaluation of clustering results in order to find the partitioning that best fits the underlying data. This is the main subject of cluster validity [52]. One way to measure the validity of the cluster is to compare the information generated by the linkage algorithm and the original proximity between samples. If the clustering is valid, the linking samples in the cluster tree should have a strong correlation with the distance between samples in the dissimilarity matrix. The *cophenetic correlation coefficient* computes the correlation between these two sets of values. The closer the *cophenetic correlation coefficient* is to one, the better the clustering solution [53].

Exploratory analysis and outlier detection
Exploratory analysis methods are useful for detecting outlier samples because they provide several tools that can be used for this purpose [26].

One easy and simple way to detect outliers is to study the score plots obtained from methods such as unfold-PCA or PARAFAC. Score plots reveal how the samples are related to each other on the basis of the measurements made [26]. Samples with similar measurements will therefore be placed close to each other on the score plots. If a sample is placed a long way from the others, it may be an outlier.

Residual analysis is another widely used strategy for detecting outliers [26]. If an outlier is projected into a model built with "normal" samples, its residuals will be larger than those of the other samples because part of the variability of the outlier sample has not been included in the model.

Grouping methods such as HCA can also be used to reveal the presence of outliers [26] since samples with similar characteristics should be close to each other in the cluster tree. If not, this is an indication that something unusual has happened with the samples.

2.3.2 Classification

In this section we describe the classification methods used in this thesis. Classification methods use a training set (i.e. a set of samples belonging to a known class) to derive a classification rule that enables us to classify new samples with unknown origin in one of the known classes. For this reason, these methods are often called supervised pattern recognition methods [49]. The validity of the classification rule is assessed with a test set.

To validate classification methods we distinguish between *recognition* and *prediction ability*. *Recognition* (or *classification*) *ability* is characterized by the percentage of the members of the training set that are correctly classified. *Prediction ability* is determined by the percentage of new samples not included in the training set that are correctly classified. Both *recognition* and *prediction ability* are usually expressed as correct classification rate [49]. It is always necessary to determine *prediction ability* because it illustrates how well the model will work with new samples. If only *recognition ability* is determined, this information is not obtained and one runs the risk of taking an overoptimistic view of the classification results.

The ideal situation for validating a method is to have enough samples available to create independent training and test sets. In this case, the training set is used to develop the classification rule and to compute the *recognition ability*, whereas the test set is used to determine the *prediction ability*. When few samples are available, CV is usually used to determine the *prediction ability*. The procedure is similar to the one described in section 2.3.1, in that the left out samples are used as a test set.

Fisher's linear discriminant analysis (LDA)
Fisher's LDA or canonical variate analysis (CVA) is a discrimination method that provides classification rules for unknown samples by maximizing the differences between the classes. This is because it seeks directions in multivariate space that

separate the groups as much as possible and uses information along these directions in simple scatter plots [42].

The method begins by computing two matrices. First, the within-classes sum of squares and products (SSP) matrix, which measures the variability within classes, is calculated. It is defined as

$$\mathbf{W} = \sum_{j=1}^{G} (\mathbf{X}_j - \mathbf{1}\bar{\mathbf{x}}_j^{\mathrm{T}})^{\mathrm{T}} (\mathbf{X}_j - \mathbf{1}\bar{\mathbf{x}}_j^{\mathrm{T}}) \tag{30}$$

where $\mathbf{X}$ is the matrix of observations for class j, and G is the number of classes. The between-class SSP matrix, which measures the variability between classes, is then calculated:

$$\mathbf{B} = \sum_{j=1}^{G} N_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^{\mathrm{T}} \tag{31}$$

where N_j is the number of samples in each class and $\bar{\mathbf{x}}$ is the average vector for the whole data set.

The algorithm then seeks a direction defined by the vector $\mathbf{a}$ that maximizes the ratio

$$\frac{\mathbf{a}^{\mathrm{T}}\mathbf{B}\mathbf{a}}{\mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{a}} \tag{32}$$

This means that the algorithm finds the direction for which the difference between the means of the classes is as large as possible compared to the within-class variance. The vector $\mathbf{a}$ can be found as the eigenvector of $\mathbf{W}^{-1}\mathbf{B}$, which corresponds to the largest eigenvalue. This is called the first canonical variate or Fisher's linear discriminant function (LDF). The value of LDF for a particular sample is called the score on the LDF.

With more than two classes, it will be necessary to compute more than one discriminant axis. They are defined in the same way as in (32), but in such a way that the variability along each vector is uncorrelated with variability along already computed vectors. The maximum number of canonical variates that can be

computed to discriminate between G groups is $G-1$, but fewer are usually enough.

Some examples of Fisher's LDA applied in food analysis are found in the literature [42,54,55].

Discriminant unfold partial least squares regression (DU-PLSR)

Partial least squares (PLS) regression is a method for building regression models between independent ($\mathbf{X}$) and dependent ($\mathbf{Y}$) variables. It is based on a bilinear decomposition of the calibration matrix and establishing a relationship (regression) between $\mathbf{X}$ and $\mathbf{Y}$ [21]. The decomposition is achieved by maximizing the covariance between $\mathbf{X}$ and $\mathbf{Y}$. Thus, an optimal regression is achieved with respect to the prediction error. Factors of $\mathbf{X}$ and $\mathbf{Y}$ are computed simultaneously by successively substituting the scores of the $\mathbf{X}$ matrix (called $\mathbf{t}$) by the scores of the $\mathbf{Y}$ matrix (called $\mathbf{u}$), and vice versa, up to convergence [56]. The PLS decomposition can be expressed in matricial form as:

$$\mathbf{X} = \sum_{f=1}^{F} \mathbf{T}\mathbf{W}^{\mathrm{T}} + \mathbf{E} \tag{33}$$

$$\mathbf{Y} = \sum_{f=1}^{F} \mathbf{U}\mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{34}$$

$\mathbf{T}$ and $\mathbf{W}^{\mathrm{T}}$ are the score and loading matrices of $\mathbf{X}$, respectively, while $\mathbf{U}$ and $\mathbf{Q}^{\mathrm{T}}$ are the score and loading matrices of $\mathbf{Y}$. The columns of $\mathbf{W}$ are often called the loading *weights* because they indicate how the $\mathbf{t}$-scores are to be computed from $\mathbf{X}$ to obtain an orthogonal decomposition. $\mathbf{E}$ and $\mathbf{F}$ are the corresponding error matrices.

There are several ways to calculate the PLS model parameters. The most intuitive method is known as NIPALS for non-iterative PLS [57] and is described below.

The PLS decomposition begins by selecting a column of $\mathbf{Y}$, $\mathbf{y}_{j,}$ as the starting estimate for $\mathbf{u}_1$ (usually the column of $\mathbf{Y}$ with the greatest variance is chosen). In the case of univariate $\mathbf{y}$, $\mathbf{u}_1 = \mathbf{y}$. For the $\mathbf{X}$ data block, the algorithm works as follows:

$$\mathbf{w}_1 = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{u}_1}{\left\|\mathbf{X}^{\mathrm{T}}\mathbf{u}_1\right\|} \tag{35}$$

$$\mathbf{t}_1 = \mathbf{X}\mathbf{w}_1 \tag{36}$$

In the $\mathbf{Y}$ data:

$$\mathbf{q}_1 = \frac{\mathbf{Y}^{\mathrm{T}}\mathbf{t}_1}{\left\|\mathbf{Y}^{\mathrm{T}}\mathbf{t}_1\right\|} \tag{37}$$

$$\mathbf{u}_1 = \mathbf{Y}\mathbf{q}_1 \tag{38}$$

If the $\mathbf{Y}$-block is univariate, eq. (37) and (38) can be omitted and $\mathbf{q}_1 = 1$. Convergence is checked by comparing $\mathbf{t}_1$ in eq. (36) with the value from the previous iteration. If they are not equal (within rounding error), the algorithm returns to eq. (35) using the $\mathbf{u}_1$ from eq. (38). When convergence is achieved, the $\mathbf{X}$ data block loadings ($\mathbf{P}$) are calculated and the scores ($\mathbf{T}$) and weights ($\mathbf{W}$) are rescaled:

$$\mathbf{p}_1 = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{t}_1}{\left\|\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1\right\|} \tag{39}$$

$$\mathbf{p}_{1\mathrm{new}} = \frac{\mathbf{p}_{1\mathrm{old}}}{\left\|\mathbf{p}_{1\mathrm{old}}\right\|} \tag{40}$$

$$\mathbf{t}_{1\mathrm{new}} = \mathbf{t}_{1\mathrm{old}}\left\|\mathbf{p}_{1\mathrm{old}}\right\| \tag{41}$$

$$\mathbf{w}_{1\mathrm{new}} = \mathbf{w}_{1\mathrm{old}}\left\|\mathbf{p}_{1\mathrm{old}}\right\| \tag{42}$$

Then the regression coefficient (b_1) is calculated:

$$b_1 = \frac{\mathbf{u}_1^{\mathrm{T}}\mathbf{t}_1}{\mathbf{t}_1^{\mathrm{T}}\mathbf{t}_1} \tag{43}$$

After the scores and loadings have been calculated for the first factor, the residuals of the $\mathbf{X}$ and $\mathbf{Y}$ blocks are calculated as follows:

$$\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} \tag{44}$$

$$\mathbf{F}_1 = \mathbf{Y} - b_1\mathbf{t}_1\mathbf{q}_1^{\mathrm{T}} \tag{45}$$

The entire procedure is then repeated for the next factor starting from eq. (35). $\mathbf{X}$ and $\mathbf{Y}$ are replaced with their residuals $\mathbf{E}_1$ and $\mathbf{F}_1$, respectively, and all subscripts are incremented by one.

When the property to be predicted ($\mathbf{Y}$) is membership to one class, the method is called discriminant PLS and can be used for classification. If there are only two classes, $\mathbf{y}$ is a vector of ones and zeros depending on the class membership of the sample in $\mathbf{X}$ (i.e. ones indicate membership to the class and zeros indicate non-membership). For more than two classes, there are as many $\mathbf{y}$ column vectors as classes. Each vector has ones for samples belonging to the class and zeros for the other samples. In this case, PLS regression must be done separately between $\mathbf{X}$ and each $\mathbf{y}$ column.

Discriminant unfold partial least squares regression (DU-PLSR) has been used in the field of sensors [20], in EEFS for classifying fungal extracts [58], and to discriminate between snack groups by measuring their chewing sounds [59]. In this thesis we applied DU-PLSR to EEFS to discriminate between oils of different origins.

Discriminant multi-way partial least squares regression (DN-PLSR)

The basis of multi-linear or multi-way partial least squares (N-PLS) regression is the same as that of bilinear PLS but extended to multi-way data. For three-way data, the method is named tri-PLS. This refers to the decomposition of a cube $\underline{\mathbf{X}}$ into a set of triads. A triad is the trilinear equivalent of a bilinear factor. It consists of one score vector $\mathbf{t}$ and two weight vectors - one in the second order, called $\mathbf{w}^J$, and one in the third order, called $\mathbf{w}^K$. The model of $\underline{\mathbf{X}}$ can be expressed by eq. (46) [21]. The term e_{ijk} is the error.

$$x_{ijk} = t_i w_j^J w_k^K + e_{ijk} \tag{46}$$

The $\underline{\mathbf{X}}$ decomposition can also be expressed in matricial form:

$$\mathbf{X} = \mathbf{T}\left(\mathbf{W}^K \big| \otimes \big| \mathbf{W}^J\right)^{\mathrm{T}} + \mathbf{E} \tag{47}$$

where $\underline{\mathbf{X}}$ has been previously unfolded to an $\mathbf{X}$ matrix. Figure 23 is a representation of this decomposition.
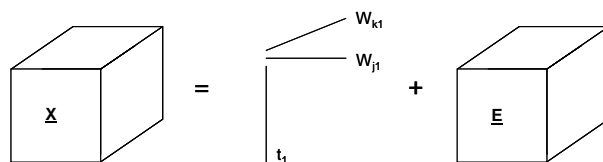
Figure 23. Graphical representation of $\underline{\mathbf{X}}$ decomposition using one N-PLS factor.

If the dependent variable $\mathbf{Y}$ is a two-way matrix as is the multivariate case, its decomposition is performed as in eq. (34).

When the N-PLS method is used for discrimination purposes, it is called DN-PLSR. In this case the $\mathbf{Y}$ matrix is made up of a set of columns. Each column is a class and has ones for samples that belong to this class and zeros for those that do not. When there are only two classes, $\mathbf{y}$ is a column vector with ones for the samples that belong to one of the classes and zeros for the samples that belong to the other.

Some applications of DN-PLS are found in the literature [20,58,59]. In this thesis we used this method to discriminate between adulterated and non-adulterated oils and between oils from two different geographical origins.

N-PLS has several advantages over unfold-PLS because the trilinear model is much simpler, easier to interpret and less prone to noise, since the information across all orders is used for the decomposition [21]. However, the choice of method depends on the behaviour of the data and no method is definitive.

2.4 REFERENCES

[1] J.R. Lakowicz, *Principles of fluorescence spectroscopy*, Kluwer Academic/Plenum Publishers, New York, 1999.

[2] A. Townshend, *Encyclopedia of analytical science*, vol. 3, Academic Press, London, 1995.

[3] W.R.G. Baeyens, D. De Keukeleire, K. Korkidis, *Luminescence techniques in chemical and biochemical analysis*, Marcel Dekker, Inc. New York, 1991.

[4] D.A. Skoog, J.J. Leary, *Análisis instrumental*, 4th Ed., McGraw-Hill, Madrid, 1994.

[5] SLM AMINCO, *Technical Note* No. 101, Urbana.

[6] L.G. Thygesen, A. Rinnan, S. Barsberg, J.K.S. Moller, *Chemom. Intell. Lab. Syst.* 71 (2004) 97-106.

[7] N.B. Kyriakidis, P. Skarkalis, *J. AOAC Int.* 83 (2000) 1435-1439.

[8] D. Rendell, *Fluorescence and phosphorescence*, John Wiley & sons, Chichester, 1987.

[9] E. Sikorska, A. Romaniuk, I.V. Khmelinskii, R. Herance, J.L. Bourdelande, M. Sikorski, J. Koziol, *J. Fluorescence* 14 (2004) 25-35.

[10] M. Zandomeneghi, L. Carbonaro, C. Caffarata, *J. Agric. Food Chem.* 53 (2005) 759-766.

[11] H.A.L. Kiers, *J. Chemometr.* 14 (2000) 105-122.

[12] R. Bro, *Multi-way analysis in the food industry: models, algorithms and applications*. Ph.D. Thesis, University of Amsterdam, Amsterdam, 1998.

[13] R. Bro, *Chemom. Intell. Lab. Syst.* 38 (1997) 149-171.

[14] R. Henrion, *Chemom. Intell. Lab. Syst.* 25 (1994) 1-23.

[15] D. Baunsgaard, L. Norgaard, M.A. Godshall, *J. Agric. Food Chem.* 48 (2000) 4955-4962.

[16] P. Nomikos, J.F. MacGregor, *AIChE Journal* 40 (1994) 1361-1375.

[17] P. Nomikos, J.F. MacGregor, *Technometrics* 37 (1995) 41-59.

[18] J.A. Westerhuis, T. Kourti, J.F. MacGregor, *J. Chemometr.* 13 (1999) 397-413.

[19] B.M. Wise, N.B. Gallagher, S. Watts Butler, D.D. White Jr, G.G. Barna, *J. Chemometr.* 13 (1999) 379-396.

[20] R.E. Shaffer, S.L. Rose-Pehrsson, R.A. McGill, *Field Anal. Chem. Tech.* 2 (1998) 179-192.

[21] R. Bro, *J. Chemometr.* 10 (1996) 47-61.

[22] C. Demir, R.G. Brereton, *Analyst* 122 (1997) 631-638.

[23] M. Azubel, F.M. Fernández, M.B. Tudino, O.E. Troccoli, *Anal. Chim. Acta* 398 (1999) 93-102.

[24] K.D. Zissis, R.G. Brereton, S. Dunkerley, R.E.A. Escotte, *Anal. Chim. Acta* 384 (1999) 71-81.

[25] M. Otto, *Chemometrics: statistics and computer application in analytical chemistry*, Wiley-VCH, Weinheim, 1999.

[26] K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics. A practical guide*, John Wiley & sons, Inc., New York, 1998.

[27] R.G. Brereton, *Chemometrics. Applications of mathematics and statistics to laboratory systems*, Ellis Horwood Limited, Chichester, 1990.

[28] T. Kourti, J.F. MacGregor, *Chemom. Intell. Lab. Syst.* 28 (1995) 3-21.

[29] S. Wold, *Technometrics* 20 (1978) 397-405.

[30] M. Forina, S. Lanteri, R. Boggia, E. Bertran, *Química Analítica* 12 (1993) 128-135.

[31] W. Wold, Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. In: *Perspectives in probability and statistics*, J. Gani (Ed.), Academic Press, London, 1975, 117-142.

[32] D.D. Lee, H.S. Seung, *Nature* 401 (1999) 788-791.

[33] D.D. Lee, H.S. Seung, *Adv. Neural Info. Proc. Syst.* 13 (2001) 556-562.

[34] J-P. Brunet, P. Tamayo, T.R. Golub, J.P. Mesirov, *PNAS* 23 (2004) 4164-4169.

[35] H-T. Gao, T-H. Li, K. Chen, W-G. Li, X. Bi, *Talanta* 66 (2005) 65-73.

[36] P. Sajda, S. Du, T.R. Brown, R. Stoyanova, D.C. Shungu, X. Mao, L.C. Parra, *IEEE T. Med. Imaging* 23 (2004) 1453-1465.

[37] C.W. Lee, H. Kang, K. Jung, H.J. Kim, *Lect. Notes Comput. Sc.* 2756 (2003) 470-477.

[38] D. Guillamet, J. Vitrià, *Proceedings of the 11th International Conference on Image Analysis and Processing*, Palermo, 2001, 256-261.

[39] G. Buchsbaum, O. Bloch, *Vision Res.* 42 (2002) 559-563.

[40] H. Asari, *unpublished Technical Note* (2004). http://zadorlab.cshl.edu/asari/nmf.html, last access April 19, 2005.

[41] Y-C. Cho, S. Choi, *Pattern Recogn. Lett.* 26 (2005) 1327-1336

[42] T. Naes, T. Isaksson, T. Fearn, T. Davies, *A user-friendly guide to multivariate calibration and classification*, NIR Publications, Chichester, 2002.

[43] L. Munck, L. Norgaard, S.B. Engelsen, R. Bro, C.A. Andersson, *Chemom. Intell. Lab. Syst.* 44 (1998) 31-60.

[44] R. Bro, *Chemom. Intell. Lab. Syst.* 46 (1999) 133-147.

[45] D. Baunsgaard, L. Noorgaard, M.A. Godshall, *J. Agric. Food Chem.* 49 (2001) 1687-1694.

[46] D.K. Pedersen, L. Munck, S.B. Engelsen, *J. Chemometr.* 16 (2002) 451-460.

[47] V. Pravdova, C. Boucon, S. de Jong, B. Walczak, D.L. Massart, *Anal. Chim. Acta* 462 (2002) 133-148.

[48] J. Christensen, E. Miquel Becker, C.S. Frederiksen, *Chemom. Intell. Lab. Syst.* 75 (2005) 201-208.

[49] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of chemometrics and qualimetrics*, Vol 20B, Elsevier, Amsterdam, 1998.

[50] M. Forina, C. Armanino, V. Raggio, *Anal. Chim. Acta* 454 (2002) 13-19.

[51] D.L. Massart, L. Kaufman, *The interpretation of analytical chemical data by the use of cluster analysis*, Wiley, New York, 1983.

[52] M. Halkidi, Y. Batistakis, M. Vazirgiannis, *J. Intell. Inf. Syst.* 17 (2001) 107-145.

[53] Statistics Toolbox Tutorial, Matlab. The Mathworks, South Natick, MA, http://www-ccs.ucsd.edu/matlab/pdf_doc/stats/stats_tb.pdf; last access June 20, 2005.

[54] M.R. Ellekjaer, K.I. Hildrum, T. Naes, T. Isaksson, *J. Near Infrared Spec.* 1 (1993) 65-75.

[55] G. Downey, *J. Near Infrared Spec.* 4 (1996) 47-61.

[56] K. Esbensen, T. Midtgaard,T, S. Schönkopf, *Multivariate Analysis*, Camo AS, Trondheim, 1996.

[57] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, *PLS-Toolbox 3.0 for use with Matlab. Manual*, Eigenvector Research Inc., 2003.

[58] C. Wittrup, *J. Chemometr.* 14 (2000) 765-776.

[59] N. De Belie, M. Sivertsvik, J. De Baerdemaeker, *J. Sound Vib.* 266 (2003) 625-643.

Chapter 3

**Exploratory Analysis
of Olive Oils**

3.1 INTRODUCTION

Olive oil authentication is of primary importance in order to ensure the quality and the authenticity of this food product. The lack of a fast analytical technique to carry out such authentication and the knowledge that olive oil contains fluorescent species [1-3] was the starting point of this thesis.

Kyriakidis and Skarkalis [2] compared the fluorescence spectra of a set of vegetable oils. They found that the spectra of virgin olive oil have considerable differences to those of the rest of oils, which have undergone refining processes during their manufacturing. The spectra of virgin olive oil display four peaks at λ_{em} = 445, 475, 525 and 681 nm. The peak at 681 nm was attributed to the chlorophylls. The assignment of the other peaks was less obvious. It was suggested that the peak at 525 nm may partially derive from vitamin E. The peaks at 445 and 475 nm were related to oxidation products. However, a study about vitamin E stability that was also made in this work showed that the fluorescence spectra of oxidized vitamin E also displayed a peak around 445 nm. All refined oils displayed only one intense peak between λ_{em} = 400-550 nm. This peak was attributed to oxidation products formed as a result of the larger content of polyunsaturated fatty acids in these oils. In this work, first-order data were used (i. e. a single fluorescence spectrum for sample) and individual fluorescence values at one emission wavelength were related to some of the oil parameters by means of univariate analysis.

From the results reported by Kyriakidis and Skarkalis [2], the first aim of this thesis was to study whether fluorescence EEMs (second-order data) are more advantageous compared to single fluorescence spectra (multivariate or first-order data) to distinguish between oil types. Thus, EEFS was applied to various sets of commercial Spanish olive oils. The fluorescent second-order peaks were related to chemical species present in oils. Different spectral ranges and preprocessing methods were applied and compared.

The results of this study are presented in two papers that show the application of several three-way methods for the exploratory analysis of olive oils. In the first paper (*Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy, Anal. Chim. Acta 515 (2004) 75-85*), two types of oils (29 virgin olive oils

(including 28 EVOOs and 1 VOO) and 20 POOs) were studied using unfold-PCA and PARAFAC. Both methods enabled to distinguish between the two types of oils. In addition, PARAFAC had the advantage of providing good approximations of the excitation and emission profiles of the fluorescence species present in oils.

In the second paper (*Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy, J. Agric. Food Chem. 52 (2004) 6673-6679*), the same groups of oils plus a third group of 7 OPOs were used. In this case, the ability of HCA to discriminate between the three types of oils was studied. HCA was applied to the unfolded EEMs of oils. In addition, an algorithm for selecting the most discriminatory variables was developed. The best results were obtained by applying HCA to the unfolded EEMs after normalization and column autoscaling of the data. In this case, all the samples were clustered into the correct groups.

References

[1] J. Gracián, *Analysis and characterization of oils, fats and fat products*, Vol. II, John Wiley & Sons, London, 1968.

[2] N.B. Kyriakidis, P. Skarkalis, *J. AOAC Int.* 83 (2000) 1435-1439.

[3] D. Marini, L. Grassi, F. Balestrieri, E. Pascucci, *Riv. Ital. Sotanze Grasse* 67 (1990) 95-99.

3.2 PAPER

F. Guimet, J. Ferré, R. Boqué, F. X. Rius
Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy

Application of Unfold Principal Component Analysis and Parallel Factor Analysis to the Exploratory Analysis of Olive Oils by means of Excitation-emission Matrix Fluorescence Spectroscopy

Francesca Guimet, Joan Ferré, Ricard Boqué, F. Xavier Rius

Department of Analytical Chemistry and Organic Chemistry, Rovira i Virgili University
Plaça Imperial Tàrraco 1, E-43005 Tarragona, Catalonia, Spain

ABSTRACT

Discrimination between virgin olive oils and pure olive oils is of primary importance for controlling adulterations. Here, we show the potential usefulness of two multi-way methods, unfold principal component analysis (U-PCA) and parallel factor analysis (PARAFAC), for the exploratory analysis of the two types of oils. We applied both methods to the excitation-emission fluorescence matrices (EEM) of olive oils and then compared the results with the ones obtained by multivariate principal component analysis (PCA) based on a fluorescence spectrum recorded at only one excitation wavelength. For U-PCA and PARAFAC, the ranges studied were λ_{ex} = 300-400 nm, λ_{em} = 400-695 nm and λ_{ex} = 300-400 nm, λ_{em} = 400-600 nm. The first range contained chlorophylls, whose peak was much more intense than those of the rest of species. The second range did not contain the chlorophylls peak but only the fluorescence spectra of the remaining compounds (oxidation products and Vitamin E). The three-component PARAFAC model on the second range was found to be the most interpretable. With this model, we could distinguish well between the two groups of oils and we could find the underlying fluorescent spectra of three families of compounds.

Keywords: U-PCA; PARAFAC; EEM; Olive oils; Three-way analysis

1. INTRODUCTION

Olive oil, obtained from the fruit of the olive tree (*Olea europea L.*), is an economically important product. The International Olive Oil Council (IOOC) has established criteria for classifying of olive oil into various grades: namely virgin, refined and pure. The best quality oil is called "extra virgin" and is derived from the first, cold pressing of the olive. Refined olive oil is obtained from virgin olive oil using refining methods that do not lead to alterations in the initial glyceridic structure, whereas pure olive oil (or simply olive oil) consists of a blend of virgin and refined oil [1-4].

Because of its high price, fraudulent practices have led to olive oil being adulterated with small amounts of seed oils or olive-pomace oil [3,5-7]. Ultraviolet spectroscopy is widely used to detect the adulteration of extra virgin olive oil with refined oil [1,5-7]. Other analytical techniques used are gas chromatography (GC) and liquid chromatography (HPLC), but these are time consuming and involve the use of solvents [1,3-7]. More recently, the use of spectroscopic techniques such as Fourier transform infrared (FT-IR) [1,5-7], Raman spectroscopy [5-7] and nuclear magnetic resonance (NMR) spectroscopy [5-7] combined with multivariate techniques, have been shown to have potential for discriminating between extra virgin olive oils and seed oils.

Olive oils exhibit strong fluorescence [8-10] and it is possible to distinguish between virgin and pure olive oils on the basis of their fluorescence emission spectra. Kyriakidis and Skarkalis [8] pointed out that the fluorescent spectra of virgin olive oils at excitation wavelength 365 nm have a peak around 681 nm, due to chlorophylls, and three other peaks (two of low intensity at 445 and 475 nm, and one more intense at 525 nm), which can be attributed to Vitamin E, whereas refined oils have one wide peak at 400-500 nm, produced by oxidation products.

Factors such as oxygen, temperature, light, ionising radiations and metals accelerate oxidation, which involves the addition of oxygen to the double bonds of unsaturated fatty acids and formation of hydroperoxides that later are degraded to aldehydes and ketones [2,11]. Hence, the fluorescence emission spectra of olive oils will be related to its composition and stability. Virgin olive oils are quite stable against oxidation because of their low fatty acid unsaturation and the antioxidant

activity of phenolic compounds and Vitamin E (α-tocopherol) [2,12,13]. In addition, chlorophylls protect oils in the darkness. They have a synergic effect with Vitamin E as a free radical scavenger, but act as a photosensitiser in the presence of light (photooxidation) [2,11]. So the fluorescence emission spectra of virgin olive oils contain peaks related to Vitamin E and chlorophylls, as indicated above [8]. However, refining processes decrease the antioxidants, such as Vitamin E, and pigments, such as chlorophylls. As a result, refined oils are more liable to undergo oxidation processes. These changes are reflected in their fluorescence spectra, in which these oxidation products give a wide peak between 400 and 500 nm [8]. This is why the fluorescence spectra of virgin and pure olive oils are quite different, even though in some cases virgin olive oil is added to pure olive oils to improve their quality.

The aim of this paper is to show that excitation-emission matrix (EEM) fluorescence spectroscopy and three-way methods of analysis, concretely unfold principal component analysis (U-PCA) and parallel factors analysis (PARAFAC), can be used for distinguishing between commercial samples of virgin and pure olive oils. These methods provide more information about the fluorescent species in these oils than the fluorescence emission spectra measured at only one excitation wavelength.

2. EXPERIMENTAL

2.1. Samples

Forty-nine olive oils (29 virgin and 20 pure) were acquired in a shopping centre. The samples were stored at room temperature and protected from light until they were analysed. The oils were analysed without any prior treatment.

2.2. Instrumentation and software

EEMs were measured with an Aminco Bowman series 2 luminescence spectrometer equipped with a 150 W xenon lamp and 10 mm quartz cells. The detector was operated with a high voltage of 600 V and in a ratio mode, i.e., the fluorescence intensity was related to the lamp source signal in order to minimise the effect of lamp fluctuations. The excitation and emission ranges were λ_{ex} = 300-400 nm and λ_{em}= 400-700 nm. The step size and band-pass of both monochromators were set to 5 and 4 nm, respectively. Scan rate was set to 30 nm

s^{-1}. The instrument software was used to correct the EEMs from deviations of ideality of the lamp, monochromators and detector [14,15].

Data were processed with Matlab software (version 6.0) [16] and the PARAFAC algorithm was obtained from the N-way toolbox [17]. The NIPALS algorithm for double cross-validation (DCV) was obtained from the Multi-block Toolbox [17].

3. RESULTS AND DISCUSSION

3.1. EEMs and preprocessing

Fig. 1 shows the EEMs of virgin and pure olive oils in the range λ_{ex}= 300-400 nm, λ_{em} = 400-695 nm. The very intense peak at λ_{em} = 600-695 nm is attributed to chlorophylls [8,9] while the range λ_{ex} = 300-400, λ_{em} = 400-600 nm mainly shows peaks due to oxidation products and Vitamin E [8] (Fig. 2).
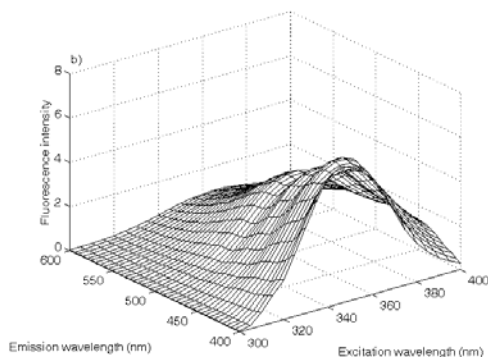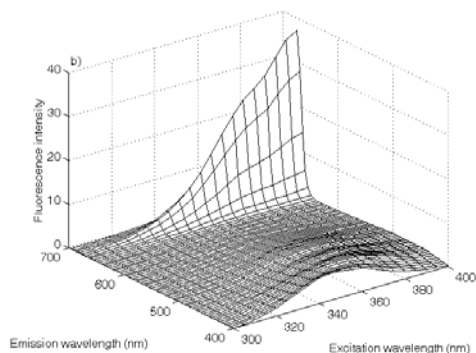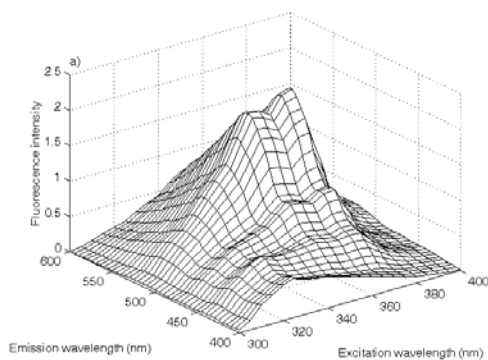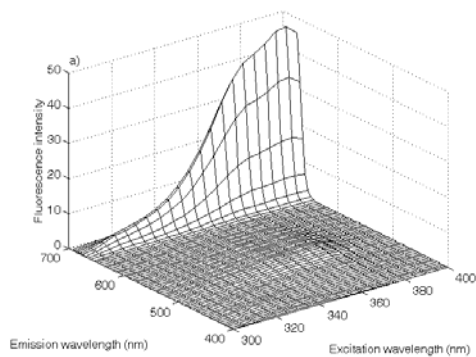

Fig. 1. EEMs of a virgin (a) and a pure (b) olive oil between λ_{ex} = 300 and 400 nm, λ_{em} = 400 and 695 nm.

Fig. 2. EEMs of a virgin (a) and a pure (b) olive oil between λ_{ex} = 300 and 400 nm, λ_{em} = 400 and 600 nm.

Studies were carried out first on the full measured range. In order to avoid the decomposition being dominated by the chlorophylls peak, autoscaling was used in PCA and U-PCA, and scaling in PARAFAC. Secondly, a more focused study about the contribution of oxidation products and Vitamin E was carried out without considering the chlorophylls peak. In this case, column mean-centering was used in PCA and U-PCA. No preprocessing was used in PARAFAC.

3.2. Deterioration of virgin olive oil

In order to determine whether the peaks in the fluorescence spectra of virgin olive oils were related to Vitamin E and chlorophylls, as indicated by Kyriakidis and Skarkalis [8], the effect of oxidation on the fluorescence spectra of virgin olive oils was studied. Fifty millilitres of virgin olive oil were placed in a beaker and heated at 70-95°C for 3 h. A stream of air from an air source was directed to the oil surface through a Pasteur pipette at regular intervals so as to accelerate oxidation. EEMs were measured in the range λ_{ex} = 300-400 nm, λ_{em} = 400-695 nm, before and after deterioration. Fig. 3 shows the total component spectra of the oil, i.e., the sum of all the fluorescence emission spectra that constituted the total EEMs.
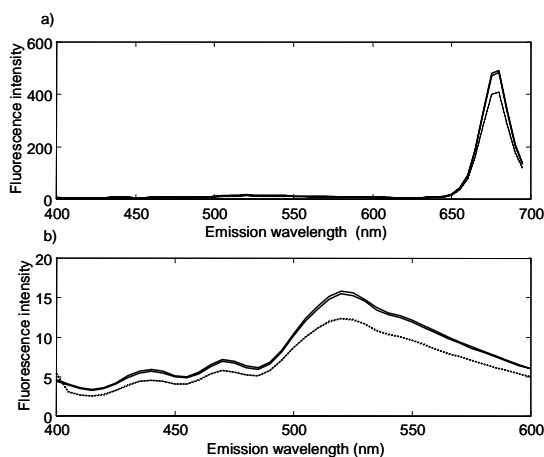

Fig. 3. Total component spectra of a virgin olive oil. (a) Fluorescence emission spectra between 400 and 695 nm; (b) fluorescence emission spectra between 400 and 600 nm. Continuous line: before oxidation; dotted line: after 3 h heating. Two replicates were measured.

It can be seen that oil degradation caused by heat and oxygen decreased the peak between 650 and 700 nm (Fig. 3a), which is due to chlorophylls [8], and those around 440, 475 and 525 nm (Fig. 3b), which can be attributed to Vitamin E [8]. Because of the high stability of virgin olive oils to oxidation, peaks formed by oxidation products were not detected.

3.3. Principal component analysis (PCA)

PCA was calculated on the emission spectra of oils between λ_{em} = 400 and 695 nm measured at λ_{ex} = 365 nm, on the basis of the work of Kyriakidis and Skarkalis [8] (Fig. 4). Two preprocessing methods were tested. Column mean-centering led to PCA being dominated by the chlorophylls peak, which is much more intense than the peaks of the rest of species. This decomposition was outperformed by PCA after column autoscaling, which could also make use of the contributions of the other species. For autoscaled data, the number of significant principal components (PCs) was five (99.1% of explained variance), when DCV [18] with random cancellation matrix with 11 cancellation groups was applied. Only the first PC (81.8% of explained variance, Table 1) could differentiate the two classes of oils (Fig. 5a). This PC combines the contribution of the chlorophylls (λ_{em} = 650-695 nm) and oxidation products (λ_{em} = 400-550 nm) (Fig. 5b). It must be noted, however, that both groups of oils overlapped slightly. Virgin olive oils 1, 25, 26 and 28 are similar to pure olive oils. In particular, sample 25 was commercially labelled as a virgin olive oil. However, its fluorescence emission zone of chlorophylls (λ_{em} = 650-695 nm) is much less intense than for the rest of virgin olive oils. It also has abnormal high intensity at the oxidation products zone (λ_{em} = 400-550 nm). Hence, this sample was more deteriorated than virgin olive oils usually are.
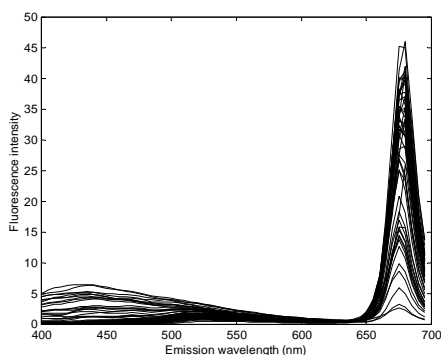

Fig. 4. Fluorescence emission spectra (λ_{em} = 400-695 nm) of the 49 olive oils, measured at λ_{ex} = 365 nm.
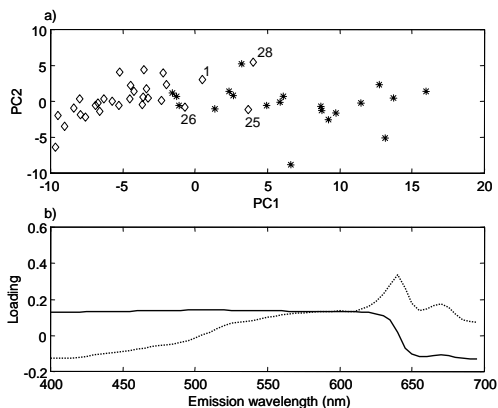
Fig. 5. (a) Score plot from PCA calculated on the emission spectra of the oils (λ_{em} = 400-695 nm) at λ_{ex} = 365 nm. (◊) Virgin olive oils; (∗) pure olive oils. (b) Loading plot. Continuous line: PC1; dotted line: PC2. Column autoscaled data.

In order to avoid the influence of chlorophylls, we applied PCA on the fluorescence emission spectra between λ_{em} = 400 and 600 nm. In this case, data were column mean-centered. The two first PCs accounted for 99.6% of variance (Table 1). Again, only PC1 contributed to differentiate between virgin and pure olive oils (Fig. 6a). This separation is mainly due to oxidation products [8], as it can be seen from the large loadings around λ_{em} = 400-550 nm (Fig. 6b). The pure olive oil samples are more scattered than the virgin olive oils. This is to be expected because pure olive oils are mixtures of different olive oils and their compositional variability is much larger than for virgin oils. Moreover, the virgin olive oils group is now less disperse than when chlorophylls are considered in the model (Fig. 5a). This might be due to the varying amount of chlorophyll in these samples. Vitamin E is the main contribution in PC2, because of the similarity between PC2 loadings and the fluorescence spectrum of this compound [8]. The content of Vitamin E is not related to the type of oil. This explains why the two types of olive oils cannot be distinguished along PC2. It is to note that sample 25 still remained close to pure olive oils. In order to check if this sample could have a strong influence on the PCA, we calculated the model again without sample 25. The results were almost identical than when the sample was included.
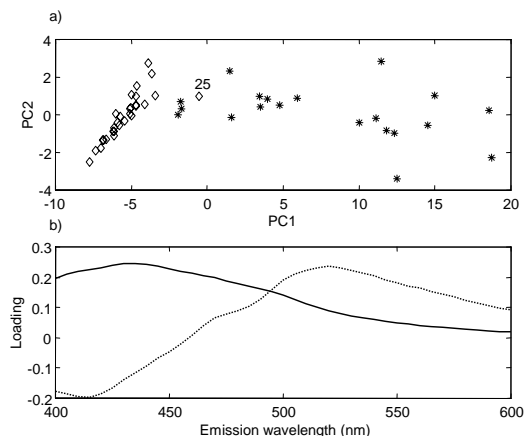
Fig. 6. (a) Score plot from PCA calculated on the emission spectra of the oils (λ_{em} = 400-600 nm) at λ_{ex} = 365 nm. ($\Diamond$) Virgin olive oils; ($*$) pure olive oils. (b) Loading plot. Continuous line: PC1; dotted line: PC2.

Table 1. Percentage of explained variance of the PCA, U-PCA and PARAFAC models

λ_{ex} (nm)	λ_{em} (nm)	PCA (λ_{ex} = 365 nm)		U-PCA		PARAFAC (non-negativity constraints on all modes)
		PC1	PC2	PC1	PC2	
300-400	400-695	81.8	11.6	(a) 60.5	18.7	(c) 98.7 (four components)
300-400	400-600	97.0	2.6	(b) 94.9	2.3	(d) 99.0 (three components)

Matrix and cube dimensions: (a) 49×1260; (b) 49×861; (c) 49×21×60; (d) 49×21×41.
In (c) data were scaled within the emission mode.

3.4. Unfold principal component analysis (U-PCA)

Two three-dimensional structures (cubes) of data were built with the EEMs of the 49 samples considering the two ranges shown in Figs. 1 and 2 (with and without chlorophylls). As the signal had been measured every 5 nm, the dimensions of the cubes were 49×21×60 and 49×21×41 (samples × number of λ_{ex} × number of λ_{em}). Later the cubes were unfolded by combining the spectral modes (Fig. 7). Hence, two matrices of dimensions 49×1260 (with chlorophyll peaks) and 49×861 (without chlorophyll peaks) were obtained. Then PCA was calculated on the unfolded matrices [19, 20].
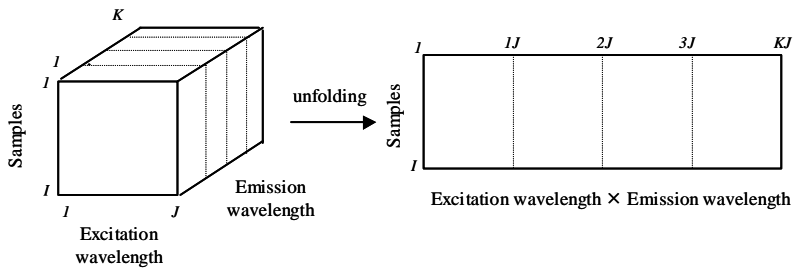
Fig. 7. Arrangement of the EEMs in a cube and unfolding by combining the spectral modes.

As in Section 3.3, when the fluorescence emission zone of chlorophylls was included in the model, the most remarkable results were obtained for column autoscaled data. DCV was carried out as in Section 3.3. The first eight PCs were significant and accounted for 97.6% of variance. Again, the first PC (60.5%, Table 1) was the most important for separating the two types of oils. The score plot (Fig. 8) is similar to the one obtained when a single emission spectrum for each sample is used (Fig. 5a), but the two types of oils are less overlapped. Hence, considering additional emission spectra in the analysis had a positive effect on differentiating between the two groups. Again sample 25 has an abnormal behaviour since it is closer to pure olive oils. For a better visualisation, loadings were refolded, i.e., each loading vector was reshaped to the same dimensions as the measured EEM (Fig. 9).
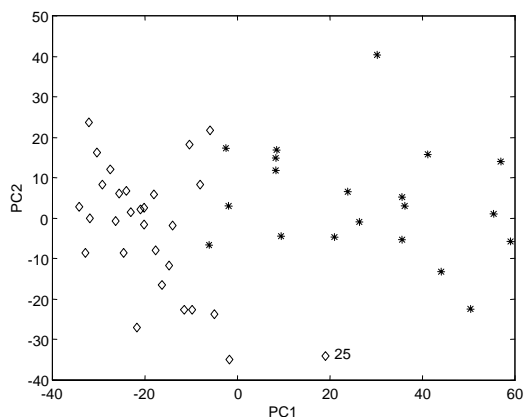

Fig. 8. Score plot from U-PCA of the 49×1260 matrix (column autoscaled). ($\Diamond$) Virgin olive oils; ($\ast$) pure olive oils.
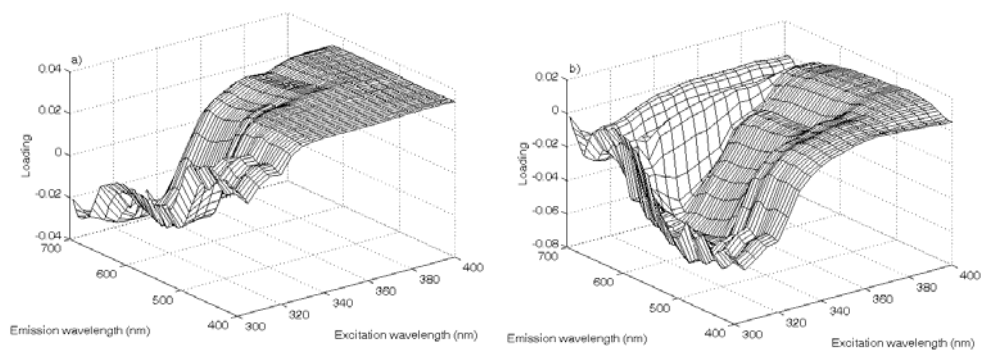
Fig. 9. Refolded loadings from U-PCA of the 49×1260 matrix (column autoscaled). (a) PC1; (b) PC2.

The loadings of PC1 and PC2 are similar to those found when PCA was applied to the fluorescence spectra at λ_{ex} = 365 nm. However, a larger contribution of oxidation products can be observed on PC1 (the peaks at low emission wavelengths in Fig. 9a). As we have explained above, these species enabled a good differentiation between the two types of oils. On the other hand, PC2 has less influence of chlorophylls (the small peak between λ_{em} = 600 and 695 nm, Fig. 9b).

In order to avoid the contribution of chlorophylls, we applied PCA on the 49×861 unfolded matrix. As in Section 3.3, data were column mean-centered. The two first PCs accounted for 97.2% of the variance (Table 1). In the score plot (Fig. 10), the group of virgin olive oils was less scattered than when chlorophylls were included in the model and the two groups appeared more separated. Sample 25 remained closer to the group of pure olive oils because of its PC1 value. Hence, the same pattern observed in PCA is repeated in U-PCA. The refolded loadings (Fig. 11) show that the wavelengths that most influence PC1 were between λ_{ex} = 300-400 and λ_{em} = 400-500 nm (Fig. 11a). The peak observed in this region was attributed to oxidation products and hydrocarbons formed during the refining process of olive oils [8,9]. The low scores in the PC1 of virgin olive oils indicate that these samples have a low content of oxidation products, unlike pure olive oils. The position and shape of loadings indicate that Vitamin E was the most influential species in PC2 [11]. Unlike PC1, PC2 does not distinguish well between the two types of oils, meaning that the model cannot distinguish the oils on the basis of their Vitamin E content.
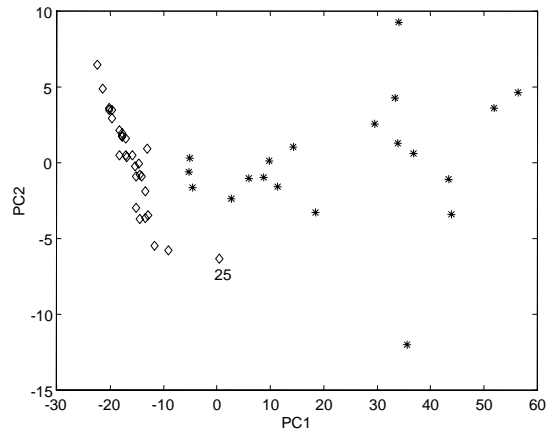
Fig. 10. Score plot from U-PCA of the 49×861 matrix (column centered). (◊) Virgin olive oils; (∗) pure olive oils.
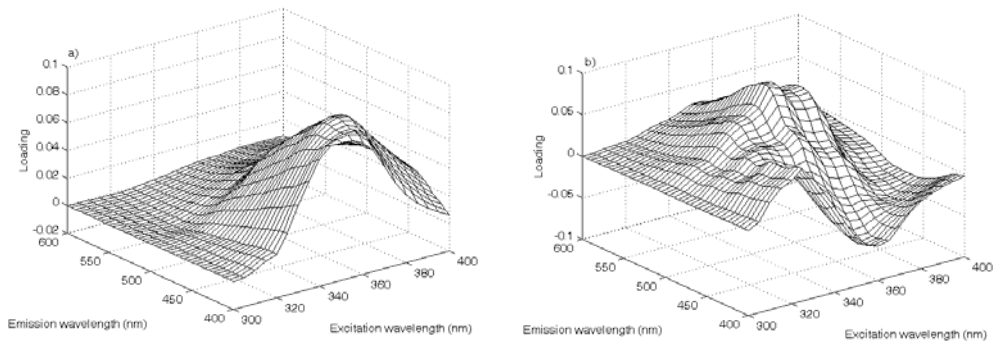

Fig. 11. Refolded loadings from U-PCA of the 49×861 matrix (column centered). (a) PC1; (b) PC2.

3.5. Parallel factor analysis (PARAFAC)

PARAFAC models with a different number of components and non-negativity constraints on all modes were calculated on EEMs in the range λ_{ex} = 300-400 nm; λ_{em} = 400-695 nm. Data were scaled within the emission mode: EEMs were unfolded to a 60×1029 matrix (number of $\lambda_{em} \times$ (samples $\times$ number of λ_{ex})) and each row was then divided by its standard deviation [21]. Residual and split-half analysis [21] pointed out that the four-component model (explained variance 98.7%, Table 1) was the optimal. The score and loading plots are shown in Fig. 12a-

d. Emission loadings were rescaled, i.e., loadings at each emission wavelength were multiplied by the standard deviation calculated above. The separation between the two types of oils was accomplished mainly along the third and the fourth component (Fig. 12a). This separation was clearer than the one obtained from PCA and U-PCA applied to the same range (Figs. 5a and 8). Again sample 25 was closer to the pure olive oil group. The first component was mainly related to Vitamin E, because of the peaks at λ_{em} = 440, 475 and 525 nm (Fig. 12d). The second component is due to chlorophylls (Fig. 12c) and its emission profile is much more intense than the rest. The third component has contribution of oxidation products (λ_{em} = 400-600 nm) and chlorophylls. Finally, the fourth component is related to oxidation products.
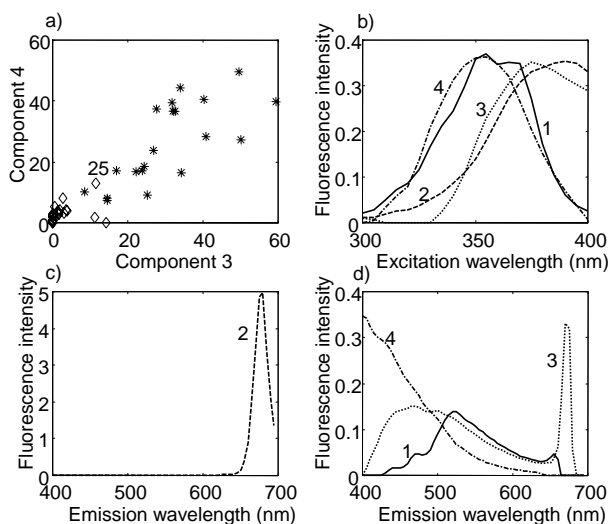

Fig. 12. (a) Score plot on component 3 and 4 of the four-component PARAFAC model calculated on the $49 \times 21 \times 60$ cube of EEM spectra, (b) excitation profiles (λ_{ex} = 300-400 nm), (c) and (d) rescaled emission profiles (λ_{em} = 400-695 nm). Continuous line: component one; broken line: component two; dotted line: component three; dash-dotted line: component four. Data scaled within the emission mode.

The same procedure was repeated but this time on the raw EEMs in the range λ_{ex} = 300-400 nm, λ_{em} = 400-600 nm. Residual and split-half analysis [21] suggested that a three-component model (explained variance 99.0%, Table 1) was the optimal. Again the group of pure olive oils was more scattered than the virgin olive oils (Fig. 13a-c), especially in the first two components (Fig. 13a). Most virgin olive oils do not contain component one, which is related to oxidation products, since its

emission profile has a wide peak around 450 nm (Fig. 13d) [8,9]. The reason why component one is zero for most of virgin olive oils is due to the non-negativity constraint applied on the concentration mode. Without this constraint, scores would be slightly negative although very little scattered. Therefore, the low value of virgin olive oils on this component is consistent with the knowledge that this type of oil has very low amounts of oxidation products. The second component is thought to be another family of oxidation products, because it was obtained from the decomposition of the component related to oxidation products in a two-component PARAFAC model. Virgin olive oils have low values on this component as well (Fig. 13a and c). On the contrary, virgin olive oils vary considerably along the third component. This component was attributed to Vitamin E, because of its shape and position [8]. All PARAFAC score plots in this range provided a rather good distinction between the two classes of oils studied. However, sample 25 was still close to pure olive oils, which confirmed that it is an outlier.
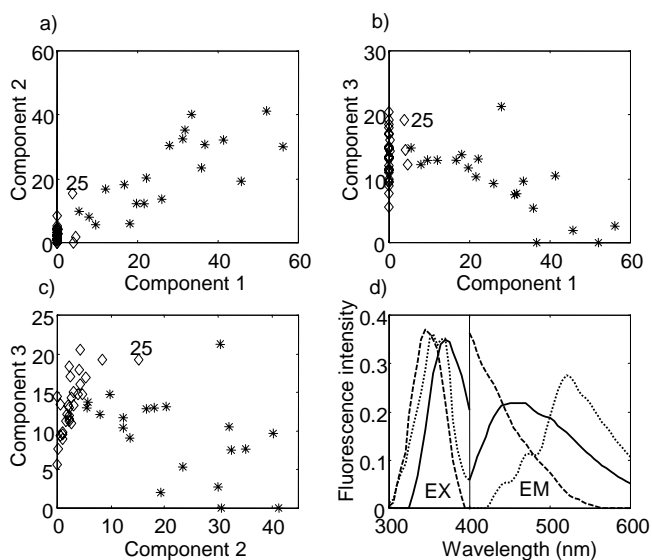

Fig. 13. (a-c) Score plots of the three-component PARAFAC model calculated on the $49 \times 21 \times 41$ cube of EEM spectra, (d) excitation (λ_{ex} = 300-400 nm) and emission (λ_{em} = 400-600 nm) profiles. Continuous line: component one; broken line: component two; dotted line: component three. Raw data.

4. CONCLUSIONS

EEM fluorescence spectroscopy has been shown to be a very useful technique for discerning composition differences between olive oils. Both U-PCA and PARAFAC applied to the EEMs of the two main groups of olive oils (virgin and pure) show clear differences between these types of oils. Chlorophylls had a strong influence on the models because of their high fluorescence intensity. As a result, data have to be scaled when its fluorescence region is included in the models. If the chlorophyll peak is not considered, column mean-centering is enough for U-PCA and no preprocessing is needed in PARAFAC. Differentiation between the two types of oils is better when the chlorophylls fluorescence region is not included in the models. In this case, oxidation products are the species that most contribute to the separation between the two groups.

The main advantage of using PARAFAC instead of U-PCA is that the output loadings are more interpretable, since they correspond to the underlying spectra of the fluorescent compounds or mixtures of compounds.

The encouraging results of this exploratory analysis suggest that the study could be extended to the development and application of three-way clustering and classification methods to EEM fluorescence and other second-order data.

ACKNOWLEDGEMENTS

REFERENCES

[1] Y.W. Lai, E.K. Kemsley, R.H. Wilson, Food Chem. 53 (1995) 95.

[2] A.K. Kiritsakis, Olive Oil: From the Tree to the Table, Food and Nutrition Press, Trumbull, 1998, p. 155.

[3] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, J. Chromatogr. A 945 (2002) 221.

[4] B. Gandul-Rojas, M.R.L. Cepero, M.I. Mínguez-Mosquera, JAOCS 77 (2000) 853.

[5] K. Papadopoulos, T. Triantis, C.H. Tzikis, A. Nikokavoura, D. Dimotikali, Anal. Chim. Acta 464 (2002) 135.

[6] M.J. Dennis, Analyst 123 (1998) 151R.

[7] V. Baeten, M. Meurens, J. Food Chem. 44 (1996) 2225.

[8] N.B. Kyriakidis, P. Skarkalis, J. AOAC Int. 83 (2000) 1435.

[9] D. Marini, L. Grassi, F. Balestrieri, E. Pascucci, Riv. Ital. Sostanze Grasse 67 (1990) 95.

[10] J. Gracian, Analysis and Characterisation of Oils, Fats and Fat Products, Wiley, London, 1968, p. 346.

[11] F. Gutiérrez-Rosales, J. Garrido-Fernández, L. Gallardo-Guerrero, B. Gandul-Rojas, M.I. Mínguez-Mosquera, JAOCS 69 (1992) 866.

[12] M. Deiana, A. Rosa, C. Falqui-Cao, F.M. Pirisi, G. Bandino, M.A. Dessì, J. Agric. Food Chem. 50 (2002) 4342.

[13] N. Pellegrini, F. Visioli, S. Buratti, F. Brighenti, J. Agric. Food Chem. 49 (2001) 2532.

[14] SLM AMINCO, Technical Note No. 101, Urbana, p. 1.

[15] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, Kluwer Academic/Plenum Publishers, New York, 1999, p. 25.

[16] Matlab, The Mathworks, South Natick, MA, USA, http://www.mathworks.com, last access: 17/07/03.

[17] http://www.models.kvl.dk/source/ (Website with algorithms for Matlab), last access: 28/11/03.

[18] S. Wold, Technometrics 20 (1978) 397.

[19] R. Henrion, Chemom. Intell. Lab. Syst. 25 (1994) 1.

[20] H.A.L. Kiers, J. Chemometr. 14 (2000) 105.

[21] R. Bro, Multi-way Analysis in the Food Industry: Models, Algorithms and Applications, Ph.D. Thesis, University of Amsterdam, 1998, p. 110.

3.3 PAPER

F. Guimet, R. Boqué, J. Ferré

Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy

Cluster Analysis Applied to the Exploratory Analysis of Commercial Spanish Olive Oils by Means of Excitation-Emission Fluorescence Spectroscopy

Francesca Guimet, Ricard Boqué, Joan Ferré

Department of Analytical Chemistry and Organic Chemistry, Rovira i Virgili University
Plaça Imperial Tàrraco 1, E-43005 Tarragona, Catalonia, Spain

ABSTRACT

Olive oil fluorescence is related to oil composition. Here it is shown that the natural clustering of different types of commercial Spanish olive oils depends on their fluorescence excitation-emission matrices (EEMs). Fifty-six commercial samples of olive oil (29 virgin olive oils, 20 pure olive oils, and 7 olive-pomace oils) were used. The clustering method was hierarchical agglomerative clustering using the Euclidean distance as a similarity measure and the average linkage. Two spectral ranges were considered (which either contained the fluorescence peak of the chlorophylls or did not), and various methods for preprocessing the fluorescence spectra were compared. The oils were clearly distinguished using the unfolded EEMs measured between λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm. The optimal preprocessing was normalization of the unfolded spectra followed by column autoscaling. Also shown are the advantages of using second-order data (EEMs) instead of first-order data (a single fluorescence spectrum) for each sample.

Keywords*: Fluorescence spectroscopy; Olive oil; Cluster analysis; Second-order data

INTRODUCTION

Food science is receiving more and more attention because of its close relationship with health. Chemometric techniques applied to analytical data have proved to be important tools in food analysis, because they can be used for exploratory analysis and classification (*1-4*). Olive oil is an economically important product. It is obtained from the fruit of the olive tree (*Olea europea L.*). There are different types of olive oils, with virgin olive oil being the one with the best quality. The characteristic odor and flavor of virgin olive oil are due to the olives being mechanically pressed and the lack of any refining processes. Refined olive oil is obtained from virgin olive oil with refining methods that do not lead to alterations in the initial glyceridic structure. This oil does not have adequate organoleptic properties. Hence, it is blended with virgin olive oil to form pure olive oil (or simply olive oil). By extracting the olive-pomace (i.e., the olive residue remaining from previous pressings) with authorized solvents, refined olive-pomace oil is obtained. This oil is improved with edible virgin oil to obtain the oil known as olive-pomace oil (*4, 5*).

Olive oil has been analyzed with such techniques as chromatography (*6, 7*), mass spectrometry (MS) (*4*), and a variety of spectroscopic techniques: infrared (IR) and Fourier transform (FT)-Raman (*8-11*), nuclear magnetic resonance (NMR) (*9, 12, 13*), fluorescence (*14-17*), and chemiluminescence (*18*). The advantages of fluorescence spectroscopy are its speed of analysis and the fact that solvents and reagents are not required, because olive oil exhibits natural fluorescence (*14, 15*). Other interesting advantages are that only a small amount of sample is needed and that it is a nondestructive technique. However, fluorescence applied to olive oil has been explored only very little, basically using a fluorescence emission spectrum recorded at one excitation wavelength (first-order data) (*14, 15*). Nevertheless, it is also possible to record entire fluorescence excitation-emission matrices (EEMs), which consist of emission spectra measured at different excitation wavelengths (second-order data). Wolfbeis and Leiner (*16*) used EEMs to characterize four types of edible oils. Scott et al. (*17*) applied pattern recognition methods to fluorescence EEMs to discriminate between four types of vegetable oils and to detect adulterations in extra virgin olive oils.

Cluster analysis (CA) is a pattern recognition technique used to form groups of objects having variables of similar values. Several similarity measures can be applied to form such groups (*19, 20*). The main advantage of CA over visualization techniques such as principal component analysis (PCA) is that it provides numerical values of the similarity between objects. As a result, the information is more objective (*19, 20*). In addition, when a large number of principal components (PCs) are required to visualize the information, CA has the advantage of reducing dimensionality while keeping the information. In many cases, the joint use of both visualization and clustering techniques is recommended (*19*). In the field of olive oil analysis, clustering has been applied to many different forms of data for many different reasons. It has been used with chromatographic data to characterize the stage of ripeness of virgin olive oil by determining the content of volatile compounds (*3*). With IR and FT-Raman spectroscopic data, it has been used to detect the adulteration of virgin olive oil (*10, 11*), and with visible and near-IR spectroscopic data, it has made a geographic classification of Mediterranean extra virgin olive oils (*21*). With NMR spectroscopic data it has been used to study the effects of climatic conditions on olive oil (*12*) and to geographically categorize virgin olive oils (*13*). Finally, it has been used with sensory, chromatographic, and MS data to differentiate virgin olive oils on the basis of the extraction methodology adopted during industrial olive oil processing (*22*). However, there are no references to applications of CA to discriminate between different types of oils using fluorescence data.

The objective of this paper is to test the ability of CA to discriminate between the three main types of commercial Spanish olive oils used for human consumption (virgin, pure, and olive-pomace oil) using EEMs as fingerprints. We applied the hierarchical agglomerative clustering (HAC) method with the Euclidean distance as a similarity measure and the average linkage method (*19, 20*) to the unfolded EEMs. Different preprocessing methods and EEM ranges were tested to find the most appropriate way of handling data and optimizing the sample grouping into clusters. Then we compare the results obtained from the EEMs to those obtained from a single fluorescence spectrum. This spectrum is selected as the one that maximizes the differences among samples. We show that the oils are best grouped into types when EEMs are used instead of single fluorescence spectra.

MATERIALS AND METHODS

Samples

A set of 56 olive oil bottles containing three different types of edible Spanish olive oils (29 virgin, 20 pure, and 7 olive-pomace oils) (**Table 1**) were purchased in a shopping center. Although they were not reference samples, they were suitable for the exploratory purposes of this study. Most of them were well-known Spanish brands, and some even prestigious. The samples were stored in the dark at room temperature until the moment of analysis. The samples were analyzed without any prior treatment.

Table 1. Samples

sample[a]	olive variety	origin
V1	Arbequina	La Palma d'Ebre (Tarragona)
V2	Arbequina	Reus (Tarragona)
V3	Arbequina	La Serra d'Almos (Tarragona)
V4	Arbequina	Llorenç del Penedès (Tarragona)
V5		Jaén
V6		Granada
V7		Jaén
V8	Hojiblanca, Arbequina	Córdoba
V9	Hojiblanca	Málaga
V10	Picual, Hojiblanca	Córdoba
V11	Picual, Hojiblanca, Picuda	Córdoba
V12		Sevilla
V13	Arbequina	Tàrrega (Lleida)
V14	Arbequina	Tàrrega (Lleida)
V15	Arbequina	Reus (Tarragona)
V16	Arbequina	Córdoba
V17	Arbequina	Les Borges Blanques (Lleida)
V18	Arbequina, Cornicabra, Hojiblanca	Jaén
V19	Hojiblanca	Tàrrega (Lleida)
V20	Hojiblanca	Tàrrega (Lleida)
V21	Picual	Tàrrega (Lleida)
V22	Picual	Tàrrega (Lleida)
V23		Jaén
V24		Córdoba
V25		Málaga
V26		Sevilla
V27		Tortosa (Tarragona)
V28		Córdoba
V29	Morrut, Farga, Sevillano	Montsià (Tarragona)
P30-49		different areas of Spain
OP50-56		different areas of Spain

[a]V, virgin olive oil; P, pure olive oil; OP, olive-pomace oil.

Instrumentation and Software

Oil EEMs were measured with an Aminco Bowman series 2 luminescence spectrometer equipped with a 150 W xenon lamp and 10 mm quartz cells. The instrument detector was operated using the EmL/Ref channel and applying a 600 V voltage for virgin and pure olive oil samples. However, P32 and olive-pomace oil samples were measured at 580 and 560 V, respectively, to avoid detector saturation. Excitation and emission ranges were λ_{ex} = 300-400 nm and λ_{em} = 400-700 nm, respectively. Measuring emission wavelengths above excitation wavelengths prevented Rayleigh scatter. The step size and band-pass of both monochromators were 5 and 4 nm, respectively. The scan rate was 30 nm s^{-1}. The instrument software was used to correct the EEMs for deviations in the ideality of the lamp, monochromators, and detector (*23, 24*). As a result, the last emission wavelength was lost and the final ranges were λ_{ex} = 300-400 nm and λ_{em} = 400-695 nm.

Data were processed with Matlab software (version 6.0) (*25*), and dendrograms were constructed from the Matlab statistics toolbox (*25*). PCA models were validated using Unscrambler software (version 8.0) (*26*).

Algorithm for Selecting the Most Discriminatory Wavelengths

Olive oil can be characterized by means of fluorescence spectroscopy. This has usually been done with emission spectra measured at one excitation wavelength (*14, 15*). Here we checked whether EEMs would make the clustering better than when a single fluorescence spectrum was used for each sample. Because we had recorded the EEMs, we applied the following algorithm to select the most discriminatory excitation wavelength (λ_{ex}):

1. Take the fluorescence emission spectrum at λ_{ex} = 300 nm of every sample from the EEMs and make them the rows in a matrix $\mathbf{Q}$.

2. Normalize each row of $\mathbf{Q}$ ($\mathbf{q}_i$) to length one (*27*) to obtain $\mathbf{Q}_N$, that is, $\mathbf{q}_{Ni} = (\mathbf{q}_i/(\mathbf{q}_i\mathbf{q}_i^T)^{1/2})$, where T means transposed and $\mathbf{q}_{Ni}$ is the *i*th row of $\mathbf{Q}_N$.

3. Column mean-center $\mathbf{Q}_N$ to obtain $\mathbf{Q}_{NC}$, that is, calculate the mean of each column of $\mathbf{Q}_N$ and subtract it from every value in the column.

4. Calculate the sum of squares (SS) of all the elements of $\mathbf{Q}_{NC}$. This SS value represents the differences between each fluorescence emission spectrum at λ_{ex} = 300 nm and the average spectrum.

5. Repeat steps 1-4 for the rest of λ_{ex}.

Hence, for each λ_{ex}, a value of SS was obtained. The λ_{ex} giving the highest SS was considered to be the most discriminatory of all the excitation wavelengths tested. To find the emission wavelength (λ_{em}) that best distinguished the types of oils, a similar procedure was used, but this time **Q** contained the fluorescence spectrum of each sample at one λ_{em}.

RESULTS AND DISCUSSION

EEMs

Figure 1 shows the average EEMs of the three types of oils studied in the ranges λ_{ex} = 300-400 nm and λ_{em} = 400-695 nm. Virgin and pure olive oil spectra look similar due to their high peak between λ_{ex} = 300-400 nm and λ_{em} = 650-695 nm, which is attributed to chlorophylls (*14, 15*). In the olive-pomace oil EEMs, this peak is much less intense and an intense peak appears between λ_{ex} = 300-400 nm and λ_{em} = 400-550 nm, which is attributed to oxidation products (*14*). Oxidation products are formed when oil comes in contact with oxygen. The process of oxidation of olive oil involves radical reactions between oxygen and double bonds of unsaturated fatty acids. Light accelerates this process. As a result, conjugated hydroperoxides are formed. These compounds are unstable, and they quickly decompose into aldehides and ketones (*5*).

Because the chlorophyll peak may hamper oil differentiation (*28*), the EEMs without this peak (λ_{ex} = 300-400 nm, λ_{em} = 400-600 nm) were also considered (**Figure 2**). In this range, the EEMs of the studied oils have considerable differences (**Figure 2**). The shape of virgin olive oil EEMs (**Figure 2a**) is mainly due to vitamin E, which emits between λ_{ex} = 300-400 nm, λ_{em} = 500-600 nm, whereas pure and olive-pomace oil samples have a larger content of oxidation products, which give rise to a broad peak at lower emission wavelengths (*14*). The different position of this peak on pure and olive-pomace EEMs enables us to distinguish between them (**Figure 2b,c**).
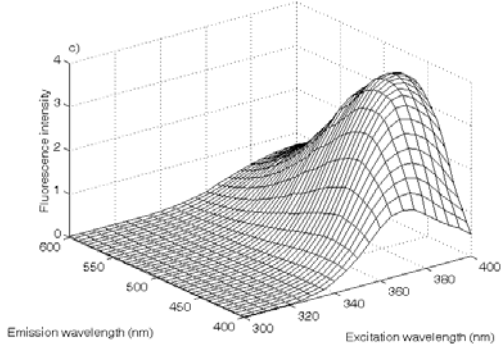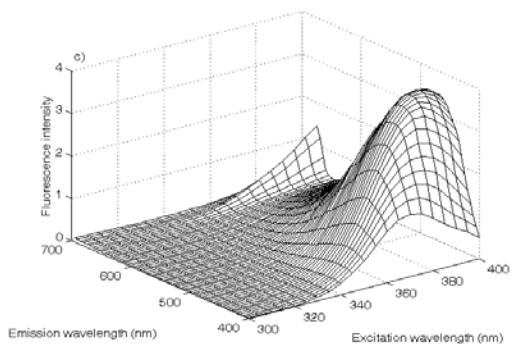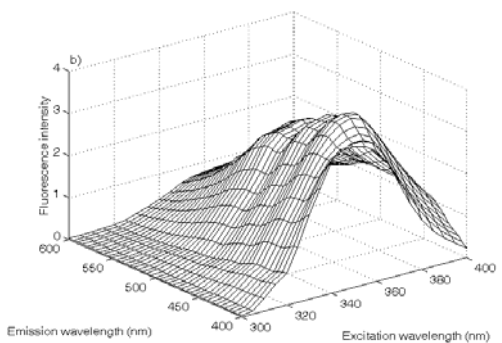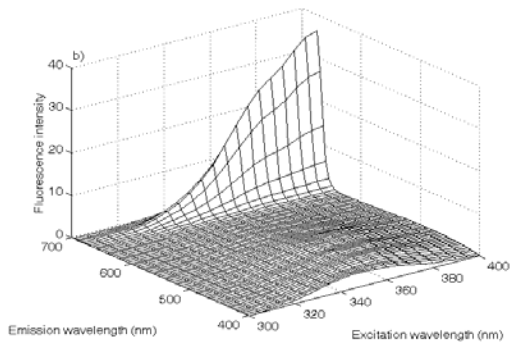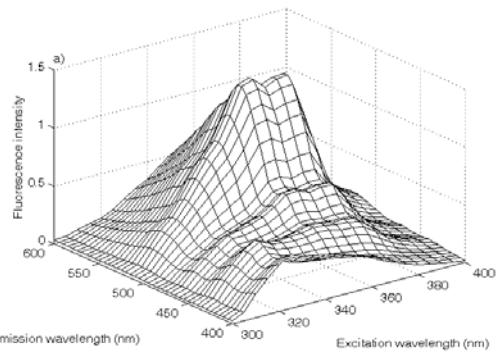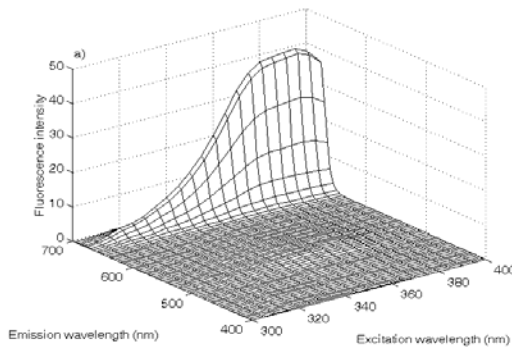
Figure 1. Average EEMs between λ_{ex}=300-400 nm and λ_{em} = 400-695 nm: (**a**) virgin olive oil; (**b**) pure olive oil; (**c**) olive-pomace oil.

Figure 2. Average EEMs between λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm: (**a**) virgin olive oil; (**b**) pure olive oil; (**c**) olive-pomace oil.

To perform CA on the full range of EEMs, the matrices were stacked in a three-way array of $56 \times 21 \times 60$ (samples $\times$ number of λ_{ex} $\times$ number of λ_{em}) and the EEMs without the chlorophylls in a three-way array of $56 \times 21 \times 41$ (samples $\times$ number of λ_{ex} $\times$ number of λ_{em}).

Cluster Analysis of Unfolded EEMs

The HAC method was applied to the EEMs in the two ranges indicated above (with and without the chlorophyll peak). In both cases, each three-way array was first unfolded to a matrix of size [samples $\times$ (number of λ_{ex} $\times$ number of λ_{em})], where each row contained the unfolded EEM of a sample. These rows were then normalized to length one (*27*) to avoid variations due to differences in intensity, and the resulting matrix was column autoscaled (i.e., every column of the matrix was set to zero mean and unit variance).

EEMs Containing the Chlorophyll Peak

After unfolding of the full-range three-way array, a 56 × 1260 matrix was obtained; the rows were samples and the columns were the excitation and emission wavelengths. Neither the raw matrix nor the matrix preprocessed as indicated above provided a good distinction between the three types of oils. Two other preprocessing methods were also tested: row normalization to length one only (*27*) and normalization to length one followed by scaling within emission mode (*29*). However, the results did not improve. The reason was that the chlorophyll peak had large variations, even between samples of the same type. A previous study based on unfolded principal component analysis and parallel factor analysis (*28*) had already shown that the chlorophyll peak hampered oil separation. The inclusion of the chlorophyll peak caused a larger overlap between the groups. For this reason, the results for the EEMs will henceforth be shown in the ranges λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm, without the chlorophyll peak.

EEMs without the Chlorophyll Peak

The unfolding of the three-way array not containing the chlorophyll peak led to a 56 × 861 matrix. The same three preprocessing methods used for the full-range matrices were tested. Then HAC was applied. Results were best when the matrix was normalized and column autoscaled. PCA was also calculated for this matrix to determine the number of variability sources in the data and to find what wavelengths caused the main differences between the types of olive oils. The PCA model was validated by leave-one-out cross-validation, and seven PCs (98.9% of explained variance) were found to be significant. However, the greatest differences between the types of oils could be seen from the score plot using the first two PCs (84.7% of explained variance (**Figure 3**)). The three types of oils are separated mainly along PC1. Olive-pomace oils have the most negative scores and virgin

olive oils the most positives scores. The loadings in **Figure 4** indicate that a high score in PC1 is related to a high content in vitamin E and a low content in oxidation products. This is concluded from the two main regions in **Figure 4a**. The region with positive values (region 1, around λ_{ex} = 300-380 nm and λ_{em} = 500-600 nm) is related to vitamin E, which emits fluorescence in this range (see **Figure 2a**). The region with negative loadings (region 2, around λ_{ex} = 340-400 nm and λ_{em} = 400-480 nm) is related to oxidation products, mainly those present in olive-pomace oils (see **Figure 2c**). Thus, the positive scores of virgin olive oils on PC1 indicate that they have high vitamin E content and fewer oxidation products than the other oils. PC2 does not separate the groups as well as PC1, but pure olive oils tend to have negative scores, whereas the rest of the oils tend to have positive scores. The wavelengths that most influence PC2 are in region 3, around λ_{ex} = 300-380 nm and λ_{em} = 400-550 (**Figure 4b**). This region has negative loadings and is related to the oxidation products of pure olive oils (see **Figure 2b**). Hence, a sample with negative scores on PC2 has a higher content on this type of products. Olive-pomace oil samples are very different from the rest on both PCs, and in general their scores are the lowest on PC1 and the highest on PC2. Only a few samples (V25, V28, V29, and OP52) appear separated from their oil group. This indicates that they are probably extreme examples of their sample type.
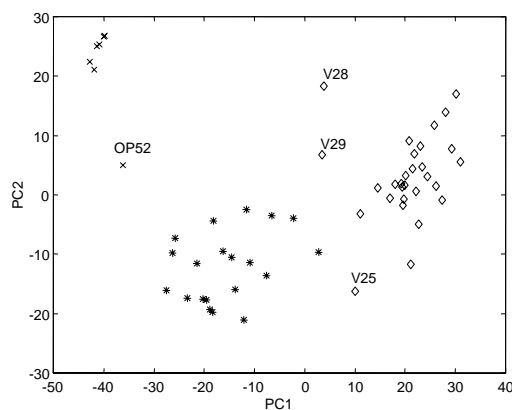
Figure 3. Score plot from PCA on the 56×861 unfolded matrix [samples $\times$ (number of $\lambda_{ex} \times$ number of λ_{em})], containing the spectra measured between λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm: (◊) virgin olive oil; (✳) pure olive oil; (×) olive-pomace oil (normalized and autoscaled spectra).

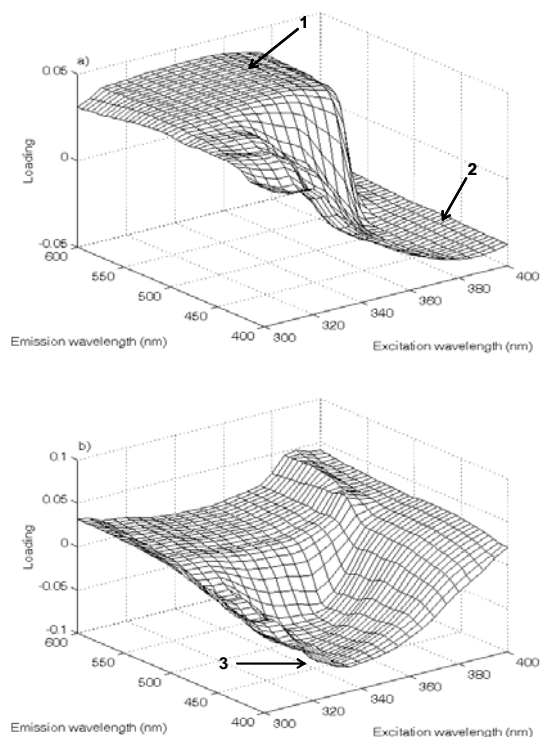**Figure 4.** Loading plot from PCA on the 56×861 unfolded matrix [samples $\times$ (number of λ_{ex} $\times$ number of λ_{em})], containing the spectra measured between λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm: (**a**) PC1; (**b**) PC2 (normalized and autoscaled spectra).

The CA dendrogram showed the three types of oils perfectly separated (**Figure 5**). The cophenetic correlation coefficient (Coph. *r*) was used as an indication of the cluster validity (*25, 30*). It measures the correlation between the linked objects in the cluster tree and the distances between objects. A Coph. *r* close to one indicates that the cluster solution reflects the similarity between objects before the tree is built. In this case, Coph. *r* = 0.7. Thus, the solution was quite good. As in PCA, the largest differences are between olive-pomace oils and the rest. These two clusters merge at a dissimilarity level of 75%. Slightly below this level (dotted line), the dendrogram shows three well-differentiated clusters, each of which contains all of the samples of one of the types of oils studied. Even samples V25, V28, and V29 are grouped in the expected cluster. However, because the number of samples of olive-pomace oil is much smaller than the others, this result must be seen as a trend and not as a general conclusion about the application of this method to olive-pomace

oils. The sample source is also a limitation. Hence, the relevance of these clustering results must be considered taking into account the facts that the samples were purchased in shopping centers and that they are not reference samples. Thus, we have only the information given by the suppliers on the label of the sample bottles.
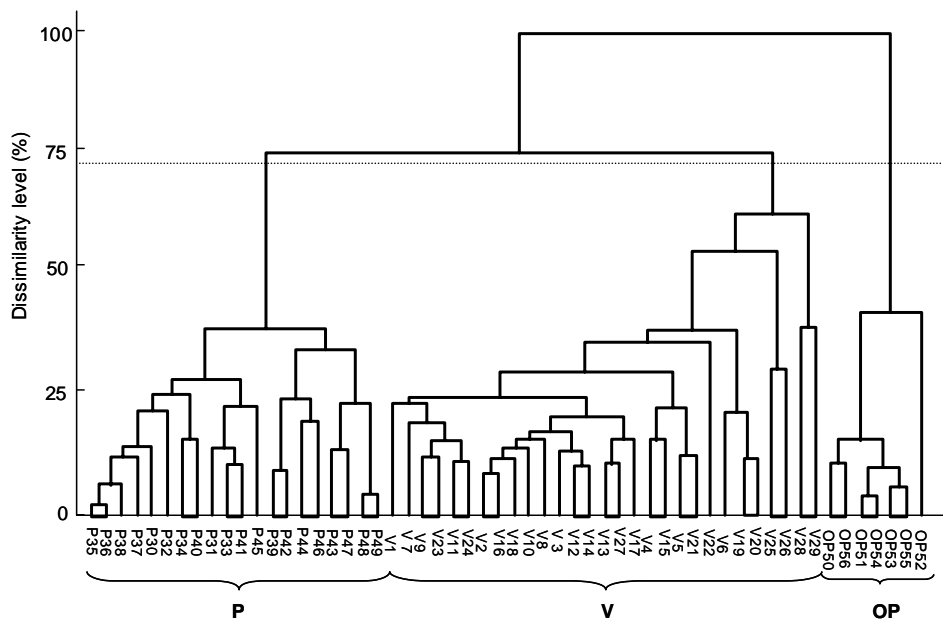
Fig. 5. Dendrogram of the 56×861 unfolded matrix (λ_{ex} = 300-400 nm, λ_{em} = 400-600 nm) using the Euclidean distance as similarity measure and the average linkage method. The distance is expressed as a percentage of dissimilarity (normalized and autoscaled spectra).

Clustering Based on Selected Fluorescence Spectra

To determine whether the same information could be obtained by using a single fluorescence spectrum for each sample instead of the whole EEM, we applied the variable selection algorithm described under Materials and Methods to find the most discriminatory fluorescence excitation and emission spectra.

The most discriminatory variables were selected from the ranges λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm. In all cases, after the variables had been selected, two preprocessing methods were applied to spectra taken from raw EEMs: normalization to length one (*27*) and normalization to length one plus column autoscaling. Then CA was applied.

Clustering Based on Emission Spectra

To find the excitation wavelength that provided the main differences between fluorescence emission spectra, the variable selection algorithm was applied for each λ_{ex}. Hence, for each λ_{ex} a 56×41 (samples $\times$ number of λ_{em}) matrix was built, and the algorithm was run on each matrix. The SS value was maximum for λ_{ex} = 345 nm. Hence, CA was applied to the emission spectra between λ_{em} = 400 and 600 nm at λ_{ex} = 345 nm (**Figure 6a**). The results were worse than when the 56×861 unfolded matrix was used. In both cases, the dendrograms displayed two large clusters, one of which contained all pure and olive-pomace oil samples and V25. Therefore, it was not possible to distinguish the oil types on the basis of the selected wavelengths.
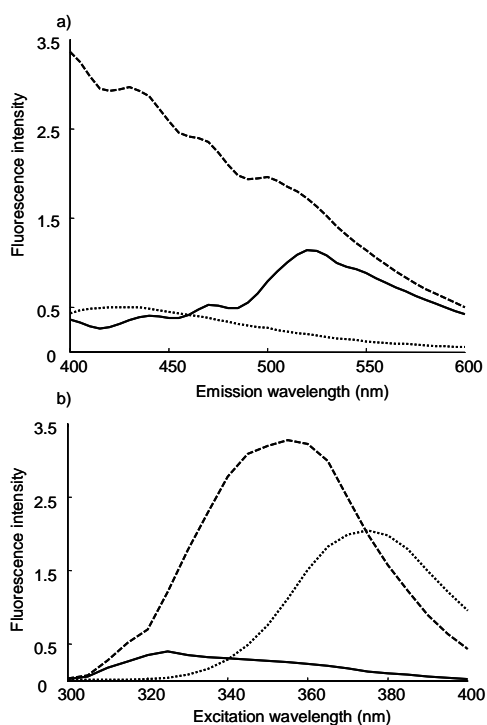
Figure 6. Average fluorescence spectra of each type of oil: (**a**) emission spectra between λ_{em} = 400 and 600 nm at λ_{ex} = 345 nm; (**b**) excitation spectra between λ_{ex} = 300 and 400 nm at λ_{em} = 410 nm; (—) virgin olive oil; (– – –) pure olive oil; (· · ·) olive-pomace oil.

To test whether fluorescence emission spectra measured at other excitation wavelengths could improve oil clustering, the HAC method was again applied to

the emission spectra at λ_{ex} = 365 nm [see Kyriakidis et al. (*14*)] and λ_{ex} = 390 nm, which was close to the excitation wavelength proposed by Marini et al. (λ_{ex} = 392 nm) (*15*). However, both preprocessing methods applied to these sets of spectra led to dendrograms that contained clusters with mixtures of samples of different types.

Clustering Based on Excitation Spectra

To find the emission wavelength that provided the main differences between fluorescence excitation spectra, the variable selection algorithm was applied again, this time for each λ_{em}. Hence, for each λ_{em} a 56×21 (samples $\times$ number of λ_{ex}) matrix was built, and the algorithm was run on each matrix. The difference was maximum for λ_{em} = 410 nm. Hence, CA was applied to the excitation spectra between λ_{ex} = 300 and 400 nm at λ_{em} = 410 nm (**Figure 6b**). When only normalization was carried out, virgin olive oil samples V12, V14, and V29 could not be distinguished from the pure olive oil cluster and samples V28 and OP52 were very different from the rest of samples, joining the olive-pomace oil cluster. Autoscaling provided a dendrogram similar to the previous one. The most noteworthy difference is that V29 was the only virgin olive oil sample that remained in the pure olive oil cluster.

Conclusions

This work has shown the potential of fluorescence spectroscopy to distinguish three types of commercial Spanish olive oils (virgin, pure and olive-pomace). Visual inspection of the EEMs of samples makes it possible to assign a large number of samples to the expected type. However, special characteristics, such as different olive oil varieties or the high deterioration of some virgin olive oil samples, may cause considerable variations in the shape of EEMs and thus hamper type differentiation. We used samples whose membership is stated in order to study the ability of the HAC method to cluster olive oil samples of different types. Different preprocessing methods and wavelength ranges were compared, and the results were best with the normalized and autoscaled unfolded matrix obtained from the EEMs between λ_{ex} = 300-400 nm and λ_{em} = 400-600 nm. Under these conditions, the samples were perfectly grouped. We also developed an algorithm for variable selection to find the most discriminatory emission and excitation spectra. However, the use of a single fluorescence spectrum worsened the grouping. These results indicate that working with second-order data is more

advantageous than working with first-order data, because the larger number of variables used contains additional information that increases discrimination power.

The results obtained here might lay the foundations for developing and applying classification methods to the EEMs of olive oils. Possible applications are olive oil characterization or the detection of frauds.

ABBREVIATIONS USED

CA, cluster analysis; EEM, excitation-emission matrix; FT, Fourier transform; HAC, hierarchical agglomerative clustering; IR, infrared; MS, mass spectrometry; PC, principal component; PCA, principal component analysis.

LITERATURE CITED

(1) Munck, L.; Norgaard, L.; Engelsen, S. B.; Bro, R.; Andersson, C. A. Chemometrics in food science – a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 31-60.

(2) Forina, M.; Armanino, C.; Raggio, V. Clustering with dendrograms on interpretation variables. *Anal. Chim. Acta* **2002**, *454*, 13-19.

(3) Aparicio, R.; Morales, M. T. Characterization of olive ripeness by green aroma compounds of virgin olive oil. *J. Agric. Food Chem.* **1998**, *46*, 1116-1122.

(4) Marcos, I.; Pérez, J. L.; Fernández, M. E.; García, C.; Moreno, B. Detection of adulterants in olive oil by headspace-mass spectrometry. *J. Chromatogr. A* **2002**, *945*, 221-230.

(5) Kiritsakis, A. K. International Olive Oil Council – Quality Criteria and Classification of Olive Oil. In *Olive Oil: from the Tree to the Table*, 2nd ed.; Food and Nutrition Press: Trumbull, CT, 1998; pp 237-260.

(6) Cert, A.; Moreda, W.; Pérez-Camino, M. C. Chromatographic analysis of minor constituents in vegetable oils, review. *J. Chromatogr. A* **2000**, *881*, 131-148.

(7) Gandul-Rojas, B.; Cepero, M. R. L.; Mínguez-Mosquera, M. I. Use of chlorophyll and carotenoid pigment composition to determine authenticity of virgin olive oil. *J. Am. Oil Chem. Soc.* **2000**, *77*, 853-858.

(8) Lai, Y. W.; Kemsley, E. K.; Wilson, R. H. Quantitative analysis of potential adulterants of extra virgin olive oil using infrared spectroscopy. *Food Chem.* **1995**, *53*, 95-98.

(9) Dennis, M. J. Recent developments in food authentication. *Analyst* **1998**, *123*, 151R-156R.

(10) Downey, G.; McIntyre, P.; Davies, A. N. Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern Mediterranean by visible and near-infrared spectroscopy, *J. Agric. Food Chem.* **2002**, *50*, 5520-5525.

(11) Baeten, V.; Meurens, M. Detection of virgin olive oil adulteration by Fourier transform Raman spectroscopy. *J. Agric. Food Chem.* **1996**, *44*, 2225-2230.

(12) Mannina, L.; Sobolev, A. P.; Segre, A. Olive oil as seen by NMR and chemometrics. *Spectrosc. Eur.* **2003**, *15*, 6-14.

(13) Sacchi, R.; Mannina, L.; Fiordiponti, P.; Barone, P.; Paolillo, L.; Patumi, M.; Segre, A. Characterization of Italian extra virgin olive oils using ^{1}H-NMR spectroscopy. *J. Agric. Food Chem*. **1998**, *46*, 3947-3951.

(14) Kyriakidis, N. B.; Skarkalis, P. Fluorescence spectra measurement of olive oil and other vegetable oils. *J. AOAC Int.* **2000**, *83*, 1435-1438.

(15) Marini, D.; Grassi, L.; Balestrieri, F.; Pascucci, E. Analisi spettrofotofluorimetrica dell'olio di oliva. Possibilità di applicazione. *Riv. Ital. Sostanze Grasse* **1990**, *67*, 95-99.

(16) Wolfbeis, O. S.; Leiner, M. Characterization of edible oils via 2D-fluorescence. *Mikrochim. Acta* **1984**, *1*, 221-233.

(17) Scott, S. M.; James, D.; Ali, Z.; O'Hare, W. T., Rowell, F. J. Total luminescence spectroscopy with pattern recognition for classification of edible oils. *Analyst* **2003**, *128*, 966-973.

(18) Papadopoulos, K.; Triantis, T.; Tzikis, C. H., Nikokavoura, A.; Dimotikali, D. Investigations of the adulteration of extra virgin olive oils with seed oils using their weak chemiluminescence. *Anal. Chim. Acta* **2002**, *464*, 135-140.

(19) Massart, D. L.; Kaufman, L. Hierarchical clustering methods. In *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley: New York, 1983; pp 75-99.

(20) Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; de Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. Cluster analysis. In *Handbook of Chemometrics and Qualimetrics*; Elsevier: Amsterdam. The Netherlands, 1998; Vol. 20B, Part B, pp 57-86.

(21) Downey, G.; McIntyre, P., Davies, A. N. Geographic classification of extra virgin olive oils from the eastern Mediterranean by chemometric analysis of visible and near-infrared spectroscopic data. *Appl. Spectrosc*. **2003**, *57*, 158-163.

(22) Angerosa, F.; Mostallino, R.; Basti, C.; Vito, R.; Serraiocco, A. Virgin olive oil differentiation in relation to extraction methodologies. *J. Sci. Food Agric.* **2000**, *80*, 2190-2195.

(23) SLM AMINCO, Technical Note 101, Urbana, Italy.

(24) Lakowicz, J. R. Instrumentation for fluorescence spectroscopy. In *Principles of Fluorescence Spectroscopy*, 2nd ed.; Kluwer Academic/Plenum Publishers: New York, 1999; pp 25-61.

(25) Matlab. The Mathworks, South Natick, MA; http://www.mathworks.com; last access July 17, 2003.

(26) http://www.camo.com/rt/products/uns/, Website of the CAMO Co.; last access Dec 18, 2003.

(27) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Preprocessing. In *Chemometrics. A Practical Guide*; Wiley: New York, 1998; pp 26-55.

(28) Guimet, F.; Ferré, J.; Boqué, R.; Rius, F. X. Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy. *Anal. Chim. Acta* **2004**, *515*, 75-85.

(29) Bro, R. Validation. In Multi-way Analysis in the Food Industry. Models, Algorithms and Applications. Ph.D. Thesis, University of Amsterdam, Amsterdam, The Netherlands, 1998; pp 99-134.

(30) Halkidi, M.; Batistakis, Y. B.; Vazirgiannis, M. On clustering validation techniques. *J. Intell. Inf. Syst.* **2001**, *17*, 107-145.

Chapter 4

**Fluorescence-Quality
Relationships
in Olive Oils**

4.1 INTRODUCTION

Olive oil authentication is based on the evaluation of many parameters. Some of them are acidity, major fatty acids composition, peroxide value (PV), UV absorbance, trinolein content, sterol composition [1-3] and sensory analysis [3]. Thus, quality assessment of olive oils requires several analyses, which sometimes are tedious and time-consuming. Hence, a complementary fast technique capable of providing information about the state of the oils would be very helpful, especially for doubtful samples. Recently, Brezmes et al. [4] proposed an electronic nose system and modified fuzzy artmap neural techniques for discriminating between defective and defect-free olive oils. In addition, total luminescence spectroscopy (TLS) has been applied to characterize and differentiate between several vegetable oils (soybean, rapessed, corn, sunflower, linseed and olive oils). This technique has also been applied to control the effect of thermal and photo-oxidation in oils [5].

Besides detecting defective samples, it is also important to ensure the genuiness of the olive oil (i.e. the oil type and the origin). Thus, techniques capable of providing oil fingerprints are very valuable. Spectral nephelometry has been proposed for this purpose [6].

Some quality parameters measured in olive oil characterization are related to their degradation. High PV and UV absorbance at 232 nm (K_{232}) indicate large content of primary oxidation products (i.e. conjugated hydroperoxides), meaning that oils are highly degraded. High absorbance at 270 nm (K_{270}) indicates large amounts of secondary oxidation products (aldehydes, ketones and low molecular weight acids), and hence, an advanced degradation stage of the oil [7].

It was shown in the previous chapter that different types of commercial olive oils can be distinguished with EEFS. However, for samples whose spectra do not follow the general trend of their type, it is difficult to decide if this is due to a major degradation of these oils or if they have been subjected to some fraudulent practice. In order to find out the cause of this special behaviour, data about their composition must be available, namely, the typical parameters analysed according to Official Methods of Analysis.

This chapter is devoted to study the possibilities of EEFS for olive oil characterization. We show that EEFS may be a complementary technique to the routine methods. Concretely, the relationship between the EEMs of a set of Spanish olive oils and the quality parameters commented above (PV, K_{232} and K_{270}) is studied. In addition, we tried to associate the chemical species responsible of these parameters to specific regions of the fluorescence landscapes. We also show that data modelling carried out by PARAFAC enables to obtain fluorescence profiles that can be used as fingerprints of the oil types.

The results of this study are presented in the paper *Excitation-emission fluorescence spectroscopy combined with three-way methods of analysis as a complementary technique for olive oil characterization, J. Agric. Food Chem. (accepted for publication).* The sample set used for this study included 13 EVOOs, 2 VOOs, 16 POOs and 2 OPOs. The quality parameters of oils were analysed according to official methods (Regulation (EEC) No 2568/91) of the EU [8]. The relationship between the fluorescence spectra and the quality parameters was evaluated in terms of correlation. Various regression methods were applied and compared. They included N-PLS regression and MLR computed to the PARAFAC scores. N-PLS provided the best results. This method showed that there is a strong correlation between the fluorescence spectra of olive oils and the peroxide value and K_{270}, which are key quality parameters of oils.

References

[1] R. Aparicio, R. Aparicio-Ruíz, *J. Chromatogr. A* 881 (2000) 93-104.

[2] C. Armanino, R. Leardi, S. Lanteri, G. Modi, *Chemom. Intell. Lab. Syst.* 5 (1989) 343-354.

[3] F. Marini, F. Balestrieri, R. Bucci, A.D. Magrì, A.L. Magrì, D. Marini, *Chemom. Intell. Lab. Syst.* 73 (2004) 85-93.

[4] J. Brezmes, P. Cabré, S. Rojo, E. Llobet, X. Vilanova, X. Correig, *IEEE Sensors Journal 5* (2005) 463-470.

[5] E. Sikorska, A. Romaniuk, I.V. Khmelinskii, R. Herance, J.L. Bourdelande, M. Sikorski, J. Koziol, *J. Fluorescenc*e 14 (2004) 25-35.

[6] A.G. Mignani, P.R. Smith, L. Ciaccheri, A. Cimato, G. Sani, *Sensor Actuat. B-Chem.* 90 (2003) 157-162.

[7] A.K. Kiritsakis, *Olive oil. From the tree to the table*, Food and Nutrition Press, Inc., 2nd Ed., Trumbull, 1998.

[8] EU, Official Journal of the European Union, Commission Regulation (EC) No 1989/2003. Amending Regulation (EEC) No 2561/91, L295 (2003) 57, http://europa.eu.int/eur-lex/en, last access: June 21, 2005

4.2 PAPER

F. Guimet, J. Ferré, R. Boqué, M. Vidal, J. Garcia

Excitation-emission fluorescence spectroscopy combined with three-way methods of analysis as a complementary technique for olive oil characterization

Journal of Agriculture and Food Chemistry (accepted for publication).

Excitation-emission Fluorescence Spectroscopy Combined with Three-way Methods of Analysis as a Complementary Technique for Olive Oil Characterization

Francesca Guimet[a], Joan Ferré[a], Ricard Boqué[a], Marta Vidal[b], Josep Garcia[b]

[a]*Department of Analytical Chemistry and Organic Chemistry, Rovira i Virgili University C/ Marcel·lí Domingo s/n, E-43007 Tarragona, Catalonia, Spain*

[b]*Alimentary section, Laboratori Agroalimentari, D.A.R.P. Generalitat de Catalunya Crta/ Vilassar de Mar-Cabrils s/n, E-08348 Cabrils, Catalonia, Spain*

ABSTRACT

This paper shows the potential of excitation-emission fluorescence spectroscopy (EEFS) and three-way methods of analysis (parallel factor analysis (PARAFAC) and multi-way partial least squares (N-PLS) regression) as a complementary technique for olive oil characterization. The fluorescence excitation-emission matrices (EEMs) of a set of Spanish extra virgin, virgin, pure and olive-pomace oils were measured and the relationship between them and some of the quality parameters of olive oils (peroxide value, K_{232} and K_{270}) was studied. N-PLS was found to be more suitable than PARAFAC combined with multiple linear regression (MLR) for correlating fluorescence and quality parameters, yielding better fits and lower prediction errors. The best results were obtained for predicting K_{270}. EEFS allowed detecting extra virgin olive oils highly degraded at early stages (with high peroxide value) and little oxidized pure olive oils (with low K_{270}). The proposed methodology may be used as an aid to analyze doubtful samples.

Keywords: Olive oils; Characterization; Fluorescence; Three-way methods

INTRODUCTION

Olive oil is obtained from the fruit of the olive tree *(Olea europea L.)*. There are different grades of olive oils (e.g. extra virgin (EV), virgin (V), pure (or simply olive oil) (P) and olive-pomace (OP) oil, among others). Each of these grades must fulfill some specifications. Due to its nutritional and economic importance, olive oil authentication is an issue of great interest in the manufacturing countries. Authenticity covers many aspects, including adulteration, mislabeling, characterization and misleading origin (*1*). Olive oil authentication is usually based on chemical parameters (acidity, major fatty acids composition, peroxide value (PV), ultraviolet absorbance, trinolein content and sterol composition) (*1-3*) and sensory analysis (*4*).

Olive oils are oxidized in the dark in contact with oxygen. As a result, essential fatty acids are destroyed and the fat soluble vitamins E (tocopherols) and A (β-carotene) disappear (oxidation). Oxidation products have an unpleasant flavor and odor and may affect the nutritional value of the oil. Nevertheless, the low content on polyunsaturated fatty acids and the natural antioxidants (phenolic compounds, tocopherols and β-carotene) present in olive oils protect them against oxidation. The four pigments contained in olive oils (chlorophyll *a* and *b*, and their derivates pheophytin *a* and *b*) also act as antioxidants in the dark, but have and oxidizing effect in presence of light (photo-oxidation) (*5*). As a result of oxidation of polyunsaturated fatty acids, conjugated hydroperoxides are formed (primary oxidation products). These compounds have high absorbance in the ultraviolet (UV) region at 232 nm (K_{232}) and they are also detected by measuring the peroxide value (PV) of the oils. Due to their low stability, hydroperoxides decompose rapidly into aldehydes, ketones and low molecular weight acids (secondary oxidation products). These compounds have high absorbance at 270 nm (K_{270}) (*5*). Due to the role of chlorophylls as sensitizers in the photo-oxidation mechanism, the longer the oils are exposed to light the more rapidly will be the conversion of conjugated hydroperoxides into secondary oxidation products. This implies an increase on K_{270}. Evaluation of the oxidation state of oils should not be done only on the basis of the peroxide value. This is because the oxidation products present in much degraded oils are not detected by measuring the peroxide value and this parameter may actually give normal values. Thus, other parameters must be considered, especially sensory analysis. The processes involved in olive oil

production have also influence on their stability. Thus, refining processes remove almost totally phenolic compounds. As a result, P and OP oils, which undergo refining processes during their manufacturing, are more prone to degradation than EV or V olive oils. In addition, refining processes produce conjugated dienes and trienes. These compounds increase K_{232} and K_{270} values, respectively.

Despite the determination of quality and purity parameters of olive oils is done according to official methods of analysis, samples that do not fulfill the reglamentation are usually analyzed again to ensure the results. Sometimes this implies much work, since some of the determinations are tedious and time-consuming, as in the case of the PV, which involves several steps. For this reason, a complementary technique capable of providing rapid information for doubtful samples would be very helpful.

Fluorescence spectroscopy has been used in the past for determining olive oil authenticity (*6*). The advantages of this technique are its speed of analysis, no use of solvents and reagents and that small amounts of sample are required. In addition, it is a non-invasive technique. Kyriakidis and Skarkalis (*7*) showed that useful information can be extracted from the fluorescence spectra of native vegetable oils. They showed that the fluorescence spectra of virgin olive oils between 400-700 nm measured at excitation wavelength (λ_{ex}) 365 nm have clear differences compared to the spectra of other vegetable oils. Virgin olive oils present two low peaks at 445 and 475 nm, one intense peak at 525 nm and another peak at 681 nm. Kyriakidis and Skarkalis suggested that the peaks at 445 and 475 nm were related to fatty acid oxidation products and the one at 525 nm was derived from vitamin E. However, they also showed that addition of vitamin E acetate to a virgin olive oil not only increased fluorescence intensity at 525 nm but also at 445 and 475 nm. They stated that this was due to oxidized vitamin E, which emits fluorescence approximately in this region. Finally, the peak at 681 nm was related to the chlorophylls. The very low intensity of the peaks at 445 and 475 nm of virgin olive oils is due to their large content on monounsaturated fatty acids and phenolic antioxidants, which provide them more stability against oxidation. All refined oils show only one intense and wide peak at around 400-560 nm, which is due to a larger oxidation state of these oils as a result of their large content on polyunsaturated fatty acids. Fluorescence of native olive oils has also been used for detecting adulterations (*8*).

Besides measuring one fluorescence spectrum at one λ_{ex}, a set of fluorescence spectra at different λ_{ex} can also be recorded. Thus, for each sample, a three-dimensional landscape is obtained, the so-called fluorescence excitation-emission matrix (EEM). The main advantage of EEMs is that more information about the fluorescent species can be extracted, because the bands arising in a wider area are considered. There are some examples in the literature of the application of excitation-emission fluorescence spectroscopy (EEFS) to native olive oils (*9-12*).

The aim of this paper is to study the potential of EEFS combined with three-way methods of analysis (parallel factor analysis (PARAFAC) and multi-way partial least squares (N-PLS) regression) as a complementary technique for olive oil characterization. The relationship between the fluorescence EEMs of EV, V, P and OP oils and some of the quality parameters of oils (peroxide value, K_{232}, and K_{270}) is studied. In addition, the PARAFAC factors provide fingerprints for the different oil types.

MATERIALS AND METHODS

Samples and reagents

A set of 33 olive oil samples, consisting of 13 EV, 2 V, 16 P and 2 OP oils were kindly supplied by the official laboratory of the Catalan government, in Spain. All oils came from Spanish cultivars and were obtained during the same harvesting year (final 2003-2004). The chemical analyses were performed by this laboratory, according to official methods of analysis (Regulation (EEC) No 2568/91) and included quality parameters (acidity, peroxide value, K_{232}, K_{270}, ΔK) and purity parameters (individual fatty acids, *trans* isomers, sterols, stigmastadienes, erythrodiol and uvaol). Table 1 shows some of the parameters analyzed. The peroxide value and K_{232} of some of the samples were not available. They are marked with a hyphen in Table 1. Oils were stored in amber glass bottles. The fluorescence EEMs were measured directly from the samples, without any prior treatment. All samples were measured in duplicate and the mean value of each sample was always used.

(+/-)-α-tocopherol acetate was purchased from Sigma-Aldrich Chimie (Alcobendas, Spain) and was stored at 7°C.

Table 1. Analytical parameters of oils
Legal limits (Regulation (EEC) No 2568/91): PV (mEq O_2/kg): 20 (EV and V), 15
(P and OP); K_{232}: 2.50 (EV), 2.60 (V); K_{270}: 0.22 (EV), 0.25(V), 0.90 (P), 1.70 (OP)

Sample	Peroxide value (mEq O_2/kg)	K_{232}	K_{270}
EV1	10	1.74	0.15
EV2	9	1.81	0.18
EV3	6	2.05	0.11
EV4	8	2.21	0.11
EV5	7	1.80	0.09
EV6	11	1.79	0.14
EV7	9	1.96	0.10
EV8	12	2.11	0.15
EV9	7	1.93	0.23
EV10	-	2.30	0.16
EV11	-	2.30	0.14
EV12	19	3.21	0.13
EV13	9	1.97	0.21
V1	11	2.29	0.19
V2	11	2.22	0.21
P1	5	-	0.20
P2	6	-	0.24
P3	5	-	0.29
P4	9	-	0.28
P5	4	-	0.27
P6	5	-	0.32
P7	3	-	0.32
P8	3	-	0.35
P9	5	-	0.26
P10	6	-	0.59
P11	5	-	0.41
P12	6	1.78	0.15
P13	6	1.91	0.18
P14	3	-	0.34
P15	6	-	0.42
P16	7	-	0.34
OP1	2	-	1.34
OP2	3	-	1.23

Instrumentation and software

EEMs were measured with an Aminco Bowman series 2 luminescence
spectrometer equipped with a 150 W xenon lamp and 10 mm quartz cells. The
instrument detector was operated using the EmL/Ref channel and applying a 600
V voltage. Excitation and emission ranges were λ_{ex} = 300-390 nm and λ_{em} = 415-600
nm, with 5 nm intervals in both dimensions. The band-pass of both
monochromators were set at 4 nm. The scan rate was 30 nm s^{-1}. The instrument

software was used to correct the EEMs for deviations in the ideality of the lamp, monochromators and detector (*13, 14*).

Data were exported to ASCII code and processed with Matlab software (version 6.5) (*15*). The chemometric models were calculated with the PLS-Toolbox (*16*).

RESULTS AND DISCUSSION

Exploratory analysis
Fluorescence EEMs of oils

Fig. 1 shows the EEMs in the range between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm of one sample of each type studied (EV, V, P and OP). Most of the samples analyzed displayed the same pattern and thus, in general, the types of oils can be differentiated from their fluorescence landscapes. EV and V oils present their maximum fluorescence intensity at emissions above 500 nm (Fig. 1a, b). On the contrary, P and OP oils exhibit much more fluorescence intensity below 500 nm (Fig. 1c, d). The main difference between these two types of oils is that EEMs of OP oils have very little fluorescence when exciting below λ_{ex} = 340 nm.
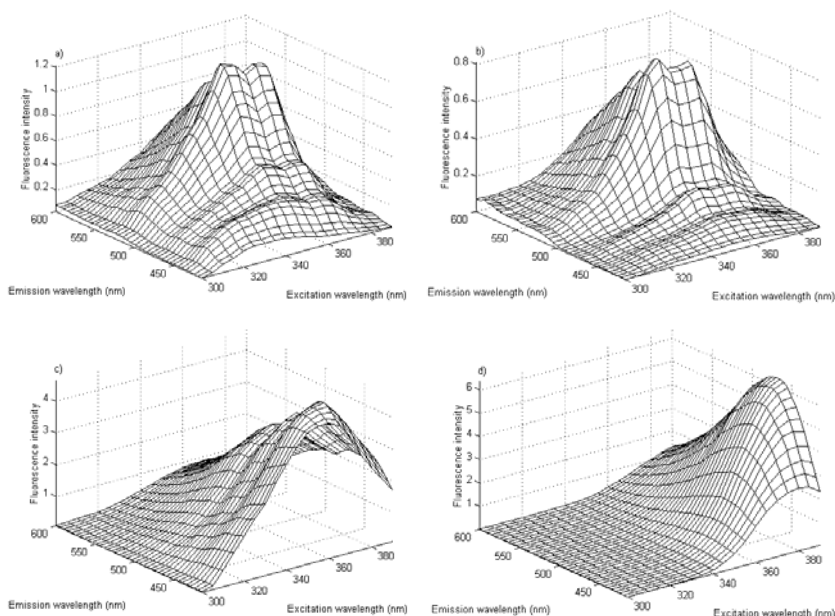

Fig. 1. EEMs between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm of an EV (a), a V (b), a P (c), and (d) an OP oil.
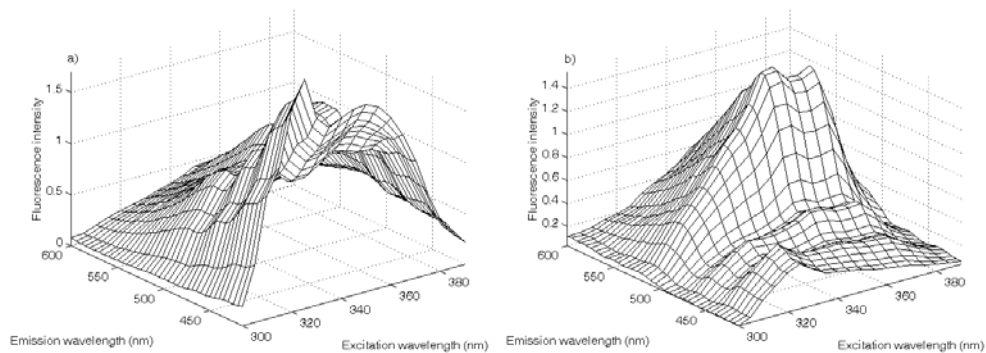
Fig. 2. EEM between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm of EV12 (a), and P12 (b).

In spite of the general trend, there are some samples (EV12, P12, and P13) with special fluorescence landscapes. The EEM of sample EV12 is very different from the other EV olive oils. It exhibits strong fluorescence at emissions below 500 nm (Fig. 2a). On the other hand, the EEMs of samples P12 and P13 are very similar to those of EV oils. Fig. 2b shows the EEM of sample P12. It is well-known that P oils consist of a blend of EV and refined oils. We suggest that the similarity between P12, P13 and EV oils is due to a high ratio EV/refined oils in P12 and P13, which would explain the shape of their EEMs.

Parallel factor analysis (PARAFAC)

In order to look into the whole set of fluorescence data, the EEMs of the 33 samples were arranged in a three-dimensional structure of size $33 \times 38 \times 19$ (samples $\times$ number of $\lambda_{em} \times$ number of λ_{ex}). The array was decomposed by PARAFAC (*17*) using different number of factors. In all cases, non-negativity constraints for the resolved profiles in all modes were applied. This was done in order to obtain a realistic solution, because the concentrations and the spectra should be positive. Residual analysis indicated that the optimal number of factors was three (98.65% of explained variance).

Figures 3-4 show the spectral profiles and the sample projection plots obtained from the PARAFAC model. The emission profile of factor 1 (Fig. 3a) is very similar to the fluorescence spectra of EV olive oils, whereas that of factor 2 is very similar to the fluorescence spectra of refined oils (*7*). As it was said above, the peak at λ_{em} = 525 nm is thought to be related to vitamin E (Fig. 3a, factor 1) and the peak

between λ_{em} = 415-560 nm to oxidation products (Fig. 3a, factor 2). The sample projection plots (Fig. 4) show that the oil types are quite differentiated on the basis of the PARAFAC factors. OP oils are very different from the rest of samples, having the highest values on factor 2 and the lowest on factor 1. This indicates that factor 2 describes mainly the oxidation products contained in OP oils and that OP oils are the ones having the lowest content on vitamin E. EV and V oils have the lowest values on factor 2 because of their stability against oxidation. These two types of oils cluster together because acidity, which is the parameter that distinguishes between the two grades, is not captured by fluorescence measurement. EV and P oils have similar values on factor 1. This means that these two types of oils have similar vitamin E content. However, P oils have larger values on factor 3. This suggests that factor 3 may be related to the presence of degradation products of oils produced during the manufacturing processes. Note that EV12 is very similar to P oils as far as factor 3 is concerned, and that P12 and P13 cluster with EV oils in all the plots. The special characteristics of these three samples have been commented above.
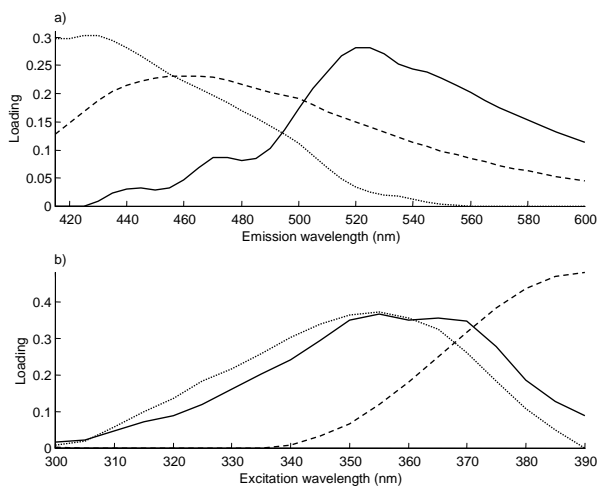

Fig. 3. Emission (a) and excitation (b) profiles obtained from the three-factor PARAFAC model calculated on the EEMs of the 33 oils in the range λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm. Factor 1 (——), factor 2 (– – –), factor 3 (…).

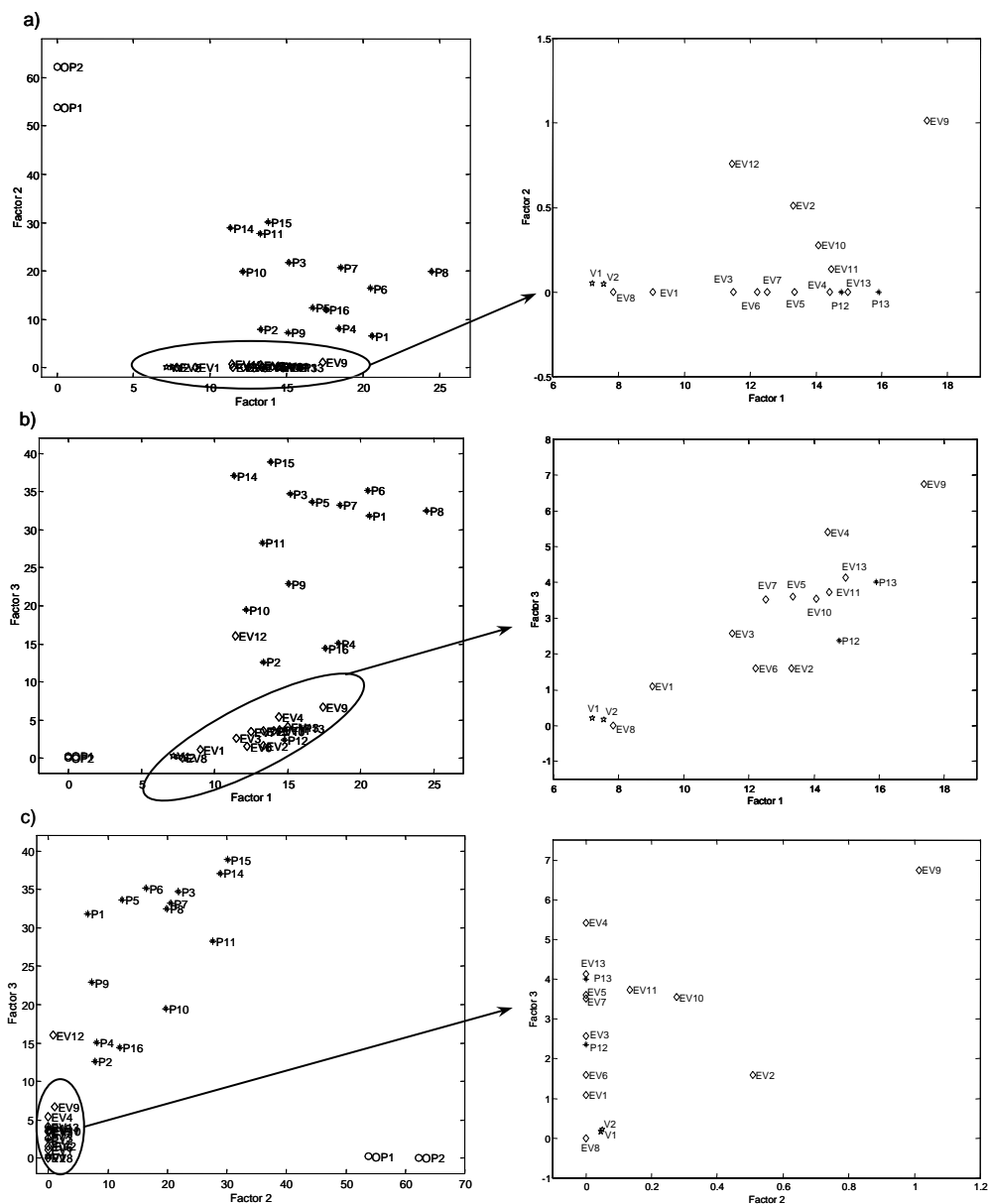**Fig. 4.** Sample projection plots of the three-factor PARAFAC model calculated on the EEMs of the 33 oils in the range λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm. Oil types: EV (◊), V (☆), P (✳), OP (o). The region containing EV and V oils has been enlarged (plots on the right).
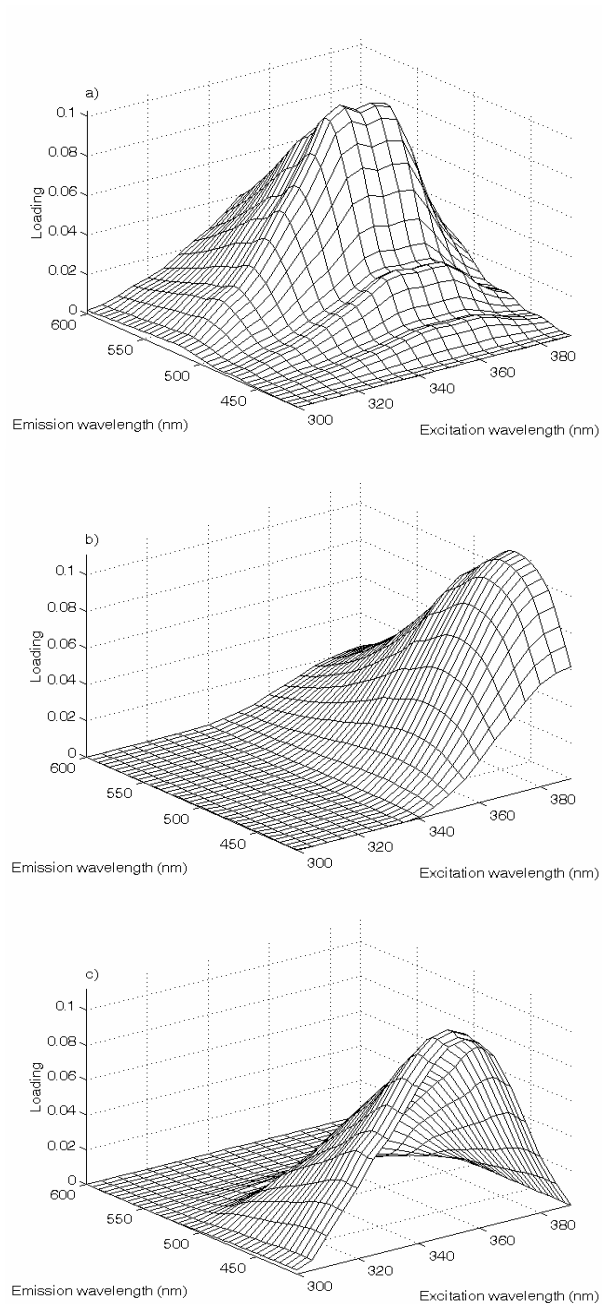
Fig. 5. Three-dimensional structure of factor 1 (a), factor 2 (b) and factor 3 (c) from PARAFAC.

Fig. 5 shows the three-dimensional structure of the PARAFAC factors, obtained after multiplying each pair of excitation and the emission profiles plotted in Fig. 3. Note the resemblance between the PARAFAC factors and the EEMs plotted in Fig. 1. Factor 1 (Fig. 5a) describes well EV and V oils (Fig. 1a), factor 2 (Fig. 5b) describes OP oils (Fig. 1c) and factor 3 (Fig. 5c) describes P oils (Fig. 1b). Hence, the PARAFAC factors may be used as a fingerprint of the types of olive oils studied.

Vitamin E and fluorescence

In order to confirm the hypothesis that the fluorescence peak at λ_{em} = 525 nm corresponds to vitamin E, we added vitamin E acetate ((+/-)-α-tocopherol acetate) to an EV olive oil sample. We used vitamin E acetate and not pure vitamin E because the latter is more unstable and is quickly oxidized by atmospheric oxygen (*7*). Vitamin E acetate was added directly to the oil so as to avoid solvent interferences and to obtain spectra directly comparable to the raw oil. The addition was made at two concentration levels. The first one consisted on adding 160 ppm of vitamin E acetate, which is equivalent to 146 ppm of pure vitamin E. The second addition consisted on 320 ppm of vitamin E acetate, which is equivalent to 292 ppm of pure vitamin E. Then, the EEMs of the spiked samples and the raw oil were recorded at the same range that the previous oils. In order to avoid detector saturation its sensitivity was set to 60% of full scale using the more concentrated sample. Each sample was measured in duplicate and the mean value of each pair of EEMs was calculated. For a better visualization of the changes produced after the addition of vitamin E acetate, we extracted the fluorescence spectra at λ_{ex} = 350 nm from the entire EEMs (Fig. 6). This λ_{ex} was selected because it provides the most intense fluorescence spectra. The plot confirms that addition of vitamin E to an EV olive oil increases fluorescence intensity at 525 nm. However, the peaks at 445 and 475 nm also increase. This was already observed by Kyriakidis and Skarkalis (*7*) who explained that this may be due to the fluorescence that oxidized vitamin E emits near this region. The hypothesis that fatty acid oxidation products are the main responsible for the peaks at 445 and 475 nm in the fluorescence spectra of EV oils could not be confirmed at this stage.
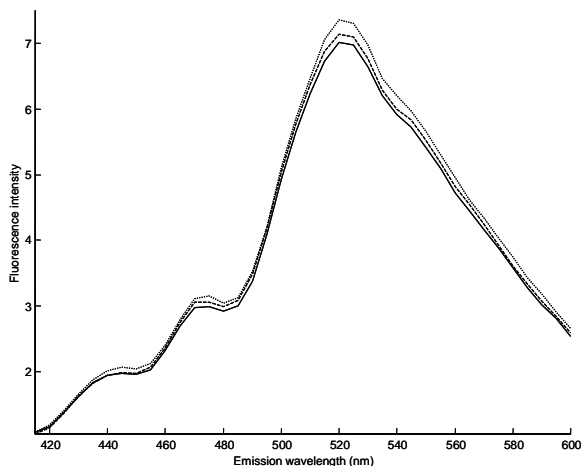
Fig. 6. Mean fluorescence spectra between λ_{em} = 415-600 nm at λ_{ex} = 350 nm. Raw EV olive oil (—), EV olive oil with 146 ppm of vitamin E added (– – –), EV olive oil with 292 ppm of vitamin E added (…).

Relationship between fluorescence and primary oxidation products

Fluorescence, PV and K_{232}

PV and K_{232} indicate the presence of primary oxidation products (i.e. conjugated hydroperoxides) in olive oils. We studied the relationship between fluorescence EEMs of oils and these parameters. All samples have values within the limits established by the Regulation (EEC) No 2568/91 (Table 1), with the exception of EV12, which has a K_{232} above the limit (2.50). This sample also has a PV very superior to those of the rest of EV oils and very close to the maximum allowed (20 mEq O_2/kg). The high values of these parameters pointed out that this sample has been much degraded and thus it presents rancidity. This may explain its special fluorescence. However, further analysis confirmed that this sample had not been adulterated, because the content on stigmastadienes was below the maximum allowed (0.15 mg/kg).

Fig. 7 shows a chart of the fluorescence EEMs arranged in increasing order of PV. The general trend is that OP oils have the lowest PV, whereas EV and V oils have the highest ones. We suggest that the low content of hydroperoxides in OP oils is due to the fact that they have been further oxidized into carbonyl compounds (secondary oxidation products), which do not contribute to PV. As it was commented in the introduction section, most of the natural antioxidants in OP oils are removed during the manufacturing processes. This makes these oils be very

prone to oxidation. Therefore conversion of primary oxidation products into secondary oxidation products is more probable in these oils. This makes that most of the oxidation products present in OP oils are not detectable by indicators of primary oxidation products, such as PV or K_{232}, but they should be detected by indicators of secondary oxidation products, such as K_{270}. The samples with the lowest PV display strong fluorescence around λ_{ex} = 340-390 nm; λ_{em} = 415-600 nm (Fig. 1d). On the contrary, the sample with the highest PV (EV12) has its maximal fluorescence around λ_{ex} = 315-370 nm; λ_{em} = 415-460 nm (Fig. 2a). Therefore, oils at early degradation stages (i.e. with large amounts of primary oxidation products) can be detected by fluorescence, since they exhibit strong fluorescence between λ_{ex} = 315-370 nm; λ_{em} = 415-460 nm, which does not occur for well conserved samples.
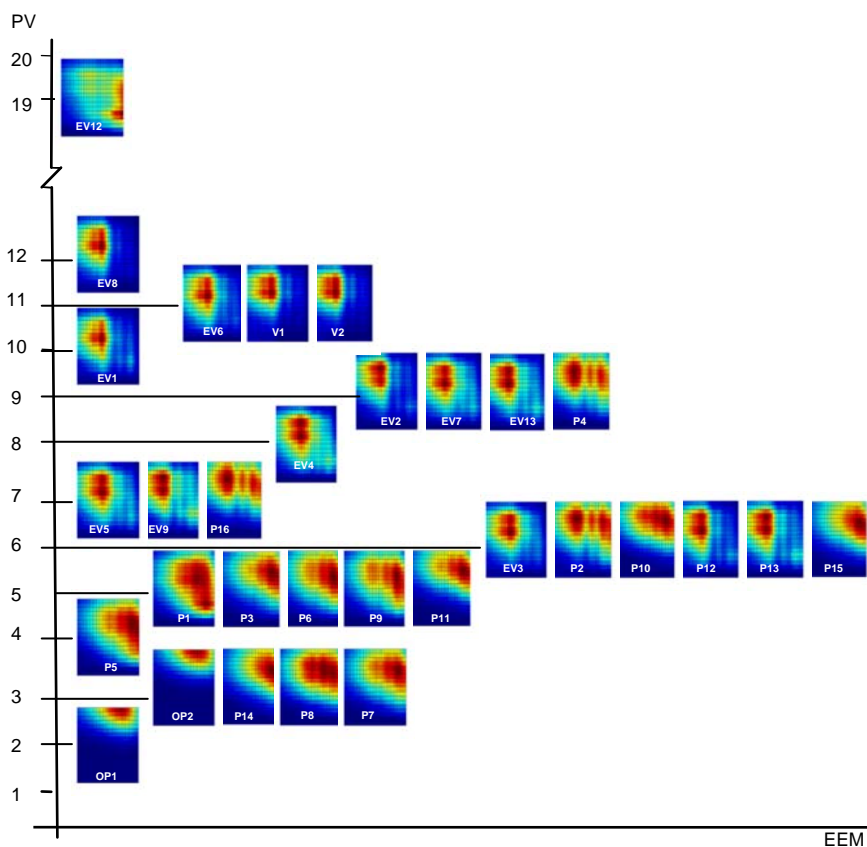

Fig. 7. EEMs of oils arranged in increasing order of PV (in mEq O_2/kg). Horizontal axis: λ_{em} = 415-600 nm (from right to left); vertical axis: λ_{ex} = 300-390 nm (from bottom to top).

Correlation between fluorescence and PV

We studied the correlation between fluorescence and PV. The EEMs of samples with known PV (Table 1) were arranged in a three-dimensional array of size $30 \times 38 \times 19$ (samples $\times$ number of $\lambda_{em} \times$ number of λ_{ex}). EV12 was excluded for its high PV.

A three-factor PARAFAC model (98.76% explained variance) was calculated on the array applying again non-negativity constraints on all modes. The spectral profiles and sample projection plots were very similar to those plotted in Figs. 3-4. A multiple linear regression (MLR) model was then applied to correlate the values of the projected samples (i.e. scores) obtained from PARAFAC with the PV. The prediction error calculated by means of the leave-one-out cross-validation procedure was 1.7 mEq O_2/kg and the correlation coefficient of the MLR model obtained in the validation step was r_{val} = 0.78. Despite of the global correlation observed, no good predictions were obtained for some samples. For example, EV3, P12, P13, P2, P10 and P15 have similar measured PV, but their predicted values differ considerably. This may indicate that the model is sensitive to some variations in fluorescence that occur between these samples that are not captured when measuring the PV.

We checked if correlation between the fluorescence EEMs and PV could be improved by using the N-PLS regression method. N-PLS is a generalization of PLS to multi-way data (*18*). This method has some nice properties, since it models both the independent ($\underline{\mathbf{X}}$) and the dependent ($\mathbf{Y}$) variables simultaneously to find the latent variables in $\underline{\mathbf{X}}$ that will best predict the latent variables in $\mathbf{Y}$. The model was applied on the centered data (across the sample mode). The optimum number of factors, selected by leave-one-out cross-validation was nine (99.92% of explained variance ($\underline{\mathbf{X}}$), 88.15% of explained variance ($\mathbf{Y}$)). The high number of factors obtained is probably due to the presence of some samples that are not very well fitted by the model. This would force the model to require more factors so as to reduce the error. However, as the number of samples available for doing this study was not very high, we decided not to remove any sample so as not to lose robustness. Figs. 8-9 show the spectral profiles and sample projection plots of the first two factors. For a better visualization, the region containing EV and V oils in Fig. 9 has been enlarged. The types of oils appear quite separated on the basis of the N-PLS factors (Fig. 9). OP oils have the highest values on factor 2. This factor is related to fluorescence at λ_{em} around 430 nm and λ_{ex} around 320 and 350 nm. P oils

have the lowest values on factor 2 and EV and V oils have the highest values on factor 1. This factor is related to a wide fluorescence peak around λ_{em} = 460 nm and λ_{ex} = 370 nm. Again, P12 and P13 are grouped with EV and V oils.
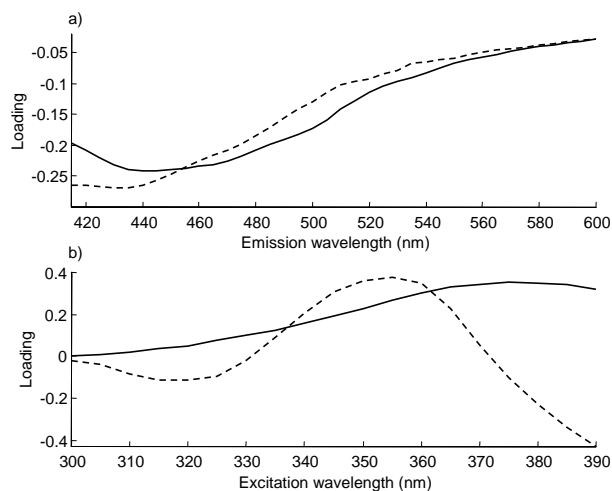

Fig. 8. Emission (a) and excitation (b) profiles of the first two factors of the nine-factor N-PLS model calculated to correlate the fluorescence EEMs with the PV. Factor 1 (——), factor 2 (– – –).

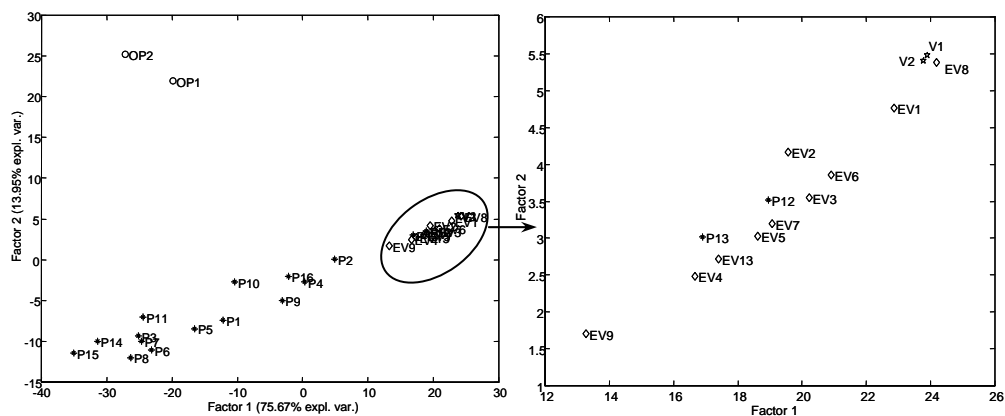**Fig. 9.** Sample projection plot of the first two factors of the nine-factor N-PLS model calculated to correlate the fluorescence EEMs with the PV. EV (◊), V (☆), P (✳), OP (o).

Fig. 10 shows the predicted vs. measured PV obtained from the nine-factor N-PLS model. Using N-PLS, correlation between fluorescence and PV was improved compared to MLR on the PARAFAC scores. In addition, the prediction errors were

lower. A similar procedure was carried out to correlate fluorescence and K_{232} (not shown). However a poor correlation was observed between these two parameters. This may indicate that some of the species that contribute to K_{232} do not emit fluorescence in the range studied.
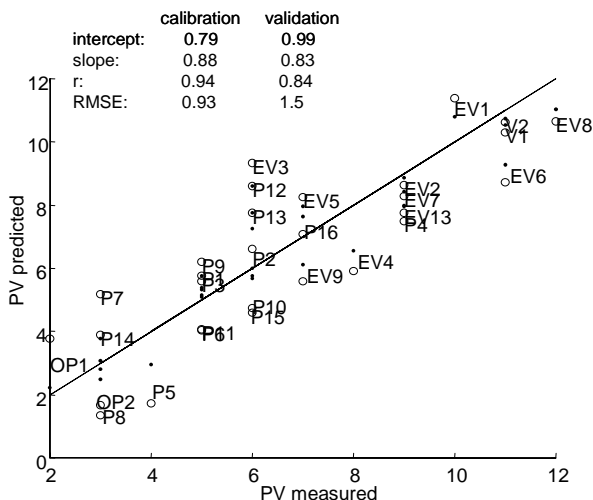

Fig. 10. Predicted vs. measured PV from the nine-factor N-PLS model in the range between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm. Calibration (•), validation (o). RMSE (calibration) is the root mean square error of calibration and RMSE (validation) is the root mean square error of cross-validation.

From the results obtained, we can state that EEFS is capable to detect samples highly degraded at early stages, because they emit strong fluorescence around λ_{ex} = 315-370 nm; λ_{em} = 415-460 nm. Thus, samples having high PV (as EV12), can be detected rapidly by recording their EEM. Hence EEFS is proposed as a rapid complementary technique for samples with high PV.

Relationship between fluorescence and secondary oxidation products

Fluorescence and K_{270}

As it has been commented previously, K_{270} is also an indicator of the oxidation state of oil, because secondary oxidation products (aldehides, ketones and other carbonyl compounds) absorb at 270 nm. Table 1 shows the K_{270} values of the 33 samples. As it can be seen, in general there is a relationship between the oil type and its K_{270}. OP oils have the highest K_{270}, whereas EV oils tend to have the lowest

values of this parameter. This indicates that OP oils are the most deteriorated and thus they contain more secondary oxidation products, whereas EV oils are the most preserved, which was expected due to their larger content on natural antioxidants. Note that, as a general trend, the higher K_{270} the lower the PV is. So there is an inverse relationship between the amount of primary and secondary oxidation products in oils. As it was explained above, this is due to the conversion of primary oxidation products into secondary oxidation products. The variation of K_{270} values in oils is also captured by the fluorescence EEMs. The oils having the highest K_{270} (OP) exhibit a wide peak between λ_{ex} = 340-390 nm and λ_{em} = 415-600 nm with a maximum fluorescence at λ_{ex} = 390 nm and λ_{em} = 470 nm (Fig. 1d). For oils having the lowest K_{270} (EV) the main fluorescence peaks appear above λ_{em} = 500 nm (Fig. 1a), with the exception of EV12.

Correlation between fluorescence and K_{270}

We applied MLR to correlate the PARAFAC scores of the 33 samples with K_{270}. The PARAFAC model chosen was that depicted in Figs. 3-4. In this case, sample EV12 was included because it was not an outlier regarding K_{270}. Validation was performed by leave-one-out cross-validation following the same procedure as above. The correlation coefficients were r_{cal} = 0.97 and r_{val} = 0.95 and the calibration and prediction errors were root mean square error of calibration RMSEC = 0.07 and root mean square error of cross-validation RMSECV = 0.08. In spite of the high correlation coefficients obtained for the whole set of samples, little correlation was observed for EV oils.

We tried to improve the correlation between fluorescence and K_{270} by applying N-PLS. Data were centered across the first mode (i.e. a matrix where each row consisted of all the emission spectra of one sample concatenated was created. Then each column of this matrix was centered by subtracting its mean value). The optimum number of factors was selected by leave-one-out cross-validation. Six factors were found to be significant (99.75% of explained variance ($\underline{\mathbf{X}}$), 97.32% of explained variance ($\mathbf{Y}$)). Figs. 11-12 show the spectral profiles and sample projection plots of the first two factors. In Fig. 12, the region containing EV and V oils has been enlarged so as to make easier visualization. Note that the profiles of factor 1 (Fig. 11) were very similar to those of Fig. 8. On the contrary, factor 2 presented more differences. When N-PLS is applied to correlate fluorescence and K_{270}, the excitation profile has only one wide peak with a minimum at 350 nm. OP

oils have the lowest values on factor 1, whereas EV and V oils tend to have the highest (Fig. 11). P oils have the highest values on factor 2.
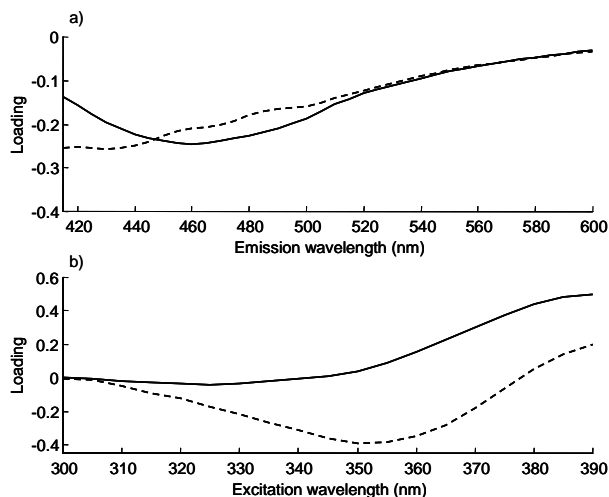

Fig. 11. Emission (a) and excitation (b) profiles of the first two factors of the six-factor N-PLS model calculated to correlate the fluorescence EEMs with K_{270}. Factor 1 (—), factor 2 (– – –).
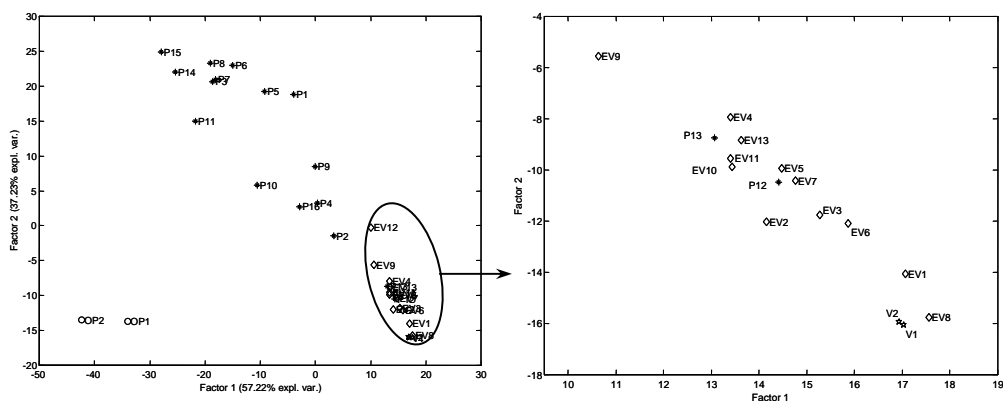

Fig. 12. Sample projection plot of the first two factors of the six-factor N-PLS model calculated to correlate the fluorescence EEMs with K_{270}. EV (◊), V (☆), P (✳), OP (o).

Fig. 13 shows the predicted vs. measured K_{270} values obtained from the six-factor N-PLS model. The group of samples not including OP oils has been enlarged. Using N-PLS enabled to obtain a better fit compared to that obtained from MLR applied to the PARAFAC scores, especially for EV samples. As it can be seen from the plot, the N-PLS factors are highly correlated with K_{270} (r_{cal} = 0.99, r_{val} = 0.96), and the prediction errors are lower compared to MLR on the PARAFAC scores.
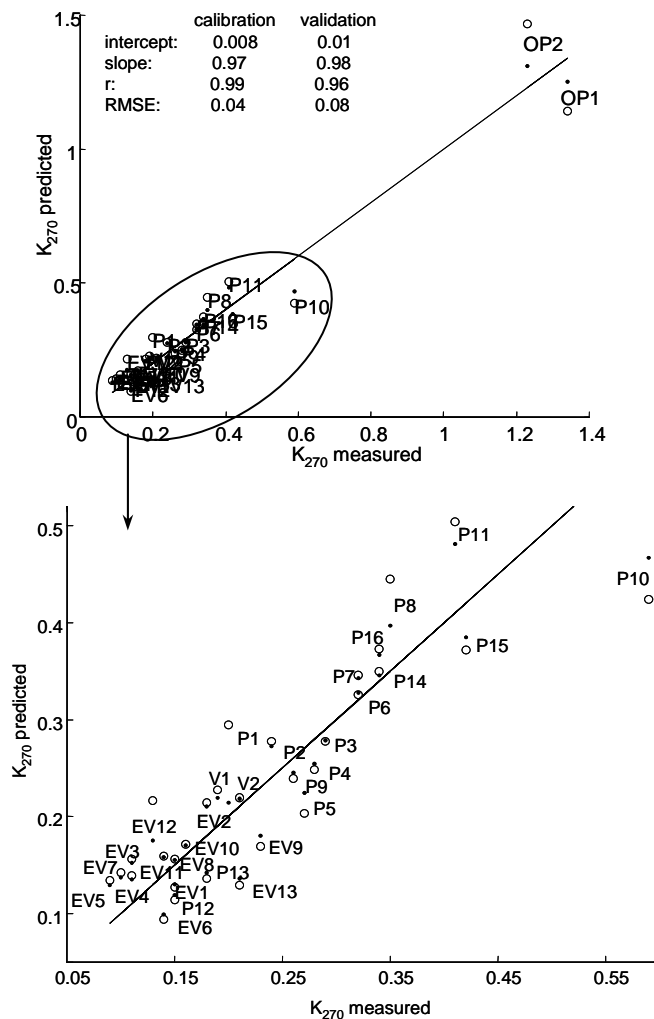

Fig. 13. Predicted vs. measured K_{270} from the six-factor N-PLS model in the range between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm. Calibration (•), validation (o). RMSE (calibration) is the root mean square error of calibration and RMSE (validation) is the root mean square error of validation.

Throughout this study, we have shown that the fluorescence data of samples P12 and P13 are very similar to those of EV oils. Nevertheless, the analytical parameters shown in Table 1 confirm that these samples belong to the P grade, because all the parameters are within the limits established. As far as K_{270} is concerned, P12 and P13 have the lowest values among the P group (Table 1 and Fig. 13). K_{270} is an indicator of the quality of oils. Low K_{270} values indicate low content of secondary oxidation products, which is due to a high stability of oils. As it was commented above, the appearance of the EEMs of P12 and P13 seems indicate a high ratio EV/refined oils in these samples. This would imply having a high level of natural antioxidants such as phenolic compounds, which would explain their stability. Hence, EEFS may be useful for studying the quality of P oils.

Conclusions

This paper has shown the potential of EEFS and three-way methods of analysis (PARAFAC and N-PLS) as a complementary technique for olive oil characterization. This methodology is proposed as an aid to determine the quality of olive oils and may be especially helpful for doubtful samples. Concretely, it has been shown that EEFS enables to detect EV oils highly deteriorated at early stages (with high PV) and little oxidized P oils (with low K_{270}). However, the results reported have to be interpreted taken into account the scarce number of samples analyzed. In further studies we aim to enlarge the data set and include more doubtful samples so as to confirm these results.

The methodology presented here is somewhat innovative. Previous studies on olive oils had already been reported about studying the correlation between fluorescence intensity and quality parameters. However, none of them used EEFS and three-way chemometric methods. The latter has some additional advantages. For instance, EEFS enables to obtain an overview of the fluorescence of various chemical species from the same analysis, which may be of interest for finding trends or patterns into the data. In addition, the chemometric analysis of these data enables to extract the spectral profiles related to the fluorescent species of oils. These profiles may be later used to make a comparative study of the contribution of the fluorescent species in the oil samples.

ACKNOWLEDGEMENTS

LITERATURE CITED

1. Aparicio, R.; Aparicio-Ruíz, R. Authentication of vegetable oils by chromatographic techniques. *J. Chromatogr. A* **2000**, *881*, 93-114.

2. Armanino, C.; Leardi, R.; Lanteri, S.; Modi, G. Chemometric analysis of Tuscan olive oils. *Chemom. Intell. Lab. Syst.*, **1989**, *5*, 343-354.

3. Marini, F.; Balestrieri, F.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; Marini, D. Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemom. Intell. Lab. Syst.*, **2004**, *73*, 85-93.

4. Aparicio, R.; Morales, M.T.; Alonso, V. Authentication of European virgin olive oils by their chemical compounds, sensory attributes, and consumers' attitudes. *J. Agric. Food Chem.*, **1997**, *45,* 1076-1083.

5. Kiritsakis, A.K. Deterioration of olive oil. In *Olive oil: from the tree to the table*, 2nd. ed.; Food & Nutrition Press Inc.: Trumbull, CT, 1998; p.155-190.

6. Gracian, J. The chemistry and analysis of olive oil. In *Analysis and characterization of oils, fats and fat products*; John Wiley & Sons: London, U.K., 1968; Vol. 2, pp. 315-366.

7. Kyriakidis, N.B.; Skarkalis, P. Fluorescence spectra measurement of olive oil and other vegetable oils. *J. AOAC Int.*, **2000**, *83*, 1435-1439.

8. Sayago, A.; Morales, M.T.; Aparicio, R. Detection of hazelnut oil in virgin olive oil by a spectrofluorimetric method. *Eur. Food Res. Technol.*, **2004**, *218*, 480-483.

9. Scott, S.M.; James, D.; Ali, Z.; O'Hare, W.T.; Rowell, F.J. Total luminescence spectroscopy with pattern recognition for classification of edible oils. *Analyst*, **2003**, *128*, 966-973.

10. Guimet, F.; Ferré, J.; Boqué, R.; Rius, F.X. Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy. *Anal. Chim. Acta*, **2004**, *515*, 75-85.

11. Guimet, F.; Boqué, R.; Ferré, J. Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy. *J. Agric. Food Chem.*, **2004**, *52*, 6673-6679.

12. Guimet, F.; Ferré, J.; Boqué, R. Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis. *Anal. Chim. Acta*, **2005**, *544*, 143-152.

13. SLM AMINCO, Technical Note N°101, Urbana, Italy.

14. Lakowicz, J.R. Instrumentation for fluorescence spectroscopy. In *Principles of Fluorescence Spectroscopy*; 2nd. ed.; Kluwer Academic/Plenum Publishers: New York, 1999; pp. 26-55.

15. Matlab. The Mathworks, South Natick, MA, USA (2002). http://www.mathworks.com, last access: March 30, 2005.

16. Toolbox for Matlab by Eigenvector Research, Inc. (2003). http://www.eigenvector.com/, last access: March 30, 2005.

17. Bro, R. PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst.*, **1997**, *38*, 149-171.

18. Bro, R. Multi-way calibration. Multi-linear PLS. *J. Chemometr.*, **1996**, *10*, 47-62.

Chapter 5

Olive Oil Classification

5.1 INTRODUCTION

Three-way classification methods had been hardly explored when this thesis was started. Some examples of classification methods with three-way data are shown in the literature [1-3]. Recently, Scott et al. [4] applied several chemometric methods to the fluorescence EEMs of edible oils for classification purposes. However, this field is still to be explored and more research should be done so as to establish new methodologies.

As it was shown in chapters 3 and 4, EEFS allows to distinguish between different types of olive oils. This chapter is devoted to study the possibilities of EEFS together with three-way classification methods for classification of olive oils. The first study presented in this chapter shows the suitability of various three-way classification methods for detecting adulterations. In the second study, EVOOs from two close geographical regions were discriminated. Finally, the third study of this chapter proposes a new classification method based on NMF combined with Fisher's LDA. The method was applied for classifying oils according to their category, their origin and for detecting adulterations.

An important aspect of olive oil authenticity is detection of adulterations. Adulteration has negative economic and nutritional implications. For this reason, constant controls for detecting such a practice are carried out.

Most of the detection of edible oil adulteration is based on chromatographic analysis [5]. Other techniques such as HS-MS [6], NMR spectrometry [7-9], UV [10,11], NIR, MIR, and FT-Raman spectroscopy [7,12-17], and chemiluminescence [18] have also been applied. García-González and Aparicio [19] have proposed metal oxide semiconductor sensors combined with neural networks for detecting *lampante* olive oils in virgin olive oils. Fluorescence spectroscopy has also been applied to detect hazelnut oil in virgin olive oils [20]. Moreover, EEFS has been applied to detect sunflower and rapessed oils in EVOOs [4]. In this case, simplified fuzzy adaptive resonance theory mapping (SFAM), traditional back propagation (BP) and radial basis function (RBF) neural networks were compared to detect such adulterations.

OPO is one of the adulterants commonly found in EVOOs. This is because OPOs are the less expensive oils obtained from olives, whereas EVOOs are the most expensive. Thus, several methodologies have been developed for detecting this type of adulterant [6,12,14,15]. None of these methodologies used EEFS.

The first paper of this chapter (*Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis, Anal. Chim. Acta 544 (2005) 143-152*) shows that EEFS and three-way methods of analysis enable to detect OPO adulteration in EVOOs at low levels (5% w/w). In a first approach, the Hotelling T^2 and Q statistics are applied to unfold-PCA as a fast screening method for detecting adulteration. Then, two supervised pattern recognition methods are applied and compared (Fisher's LDA and DN-PLSR). DN-PLS gave the best results. With this method the percentages of correct classification were 100% for the non-adulterated samples as well as for the adulterated. The level of adulteration was also quantified by means of the N-PLS method, which provided acceptable prediction errors.

Besides controlling adulterations, there is also the need for authenticating the geographical origin of EVOOs. This is usually done from data about the chemical composition of oils, such as free acidity, PV and fatty acids, among others. These parameters are usually determined according to Official Methods of Analysis [21-24]. Other techniques such as HS-MS [25], NMR spectroscopy [26,27], VIS and IR spectroscopy [28,29] and isotope ratio mass spectrometry (IRMS) [30] have also been applied. Chemometric methods have enabled to discriminate between oils from different origin [21-25,27,30,31].

In order to protect quality agriculture and food products of recognised origin from fraud, the EU created two distinctions under the EC Regulation 2081/92 [11], concretely, Protected Denomination of Origin (PDO) and Protected Geographical Indication (PGI). PDO is used to describe foodstuffs which are produced, processed and prepared in a given geographical area using recognised know-how [32]. Some quality olive oils are protected under the PDO label.

PDO "Siurana" olive oil includes the EVOOs produced within the Tarragona province, in Catalonia (Spain) (Fig. 24). These EVOOs come mainly from *arbequina*

olives and are characterized for having excellent sensory properties and typical physicochemical characteristics [33]. The PDO production area can be divided into two regions: Siurana-Montsant (SM) and Siurana-Camp de Tarragona (SC) (Fig. 24). Due to differences in orography, soil characteristics and climatic conditions, oils produced in the two regions have different composition, which gives them different stability. There are also other factors affecting the physicochemical composition of oils, such as the harvesting year and the ripening stage of the olives.
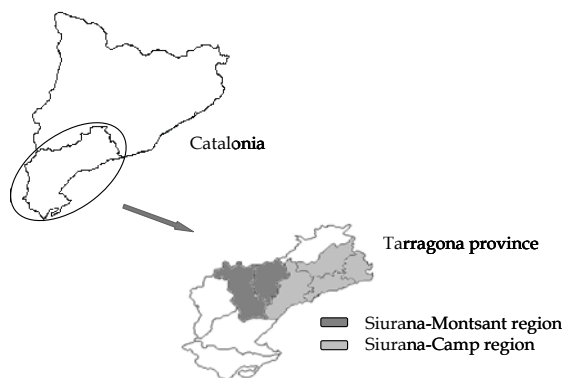

Figure 24. Map of the production area of PDO "Siurana" olive oils.

Tous et al. [33] differentiated between SM and SC oils on the basis of several parameters (free acidity, K_{270}, phenolic content, bitterness (K_{225}), stability, fatty acids composition, and sensory analysis). However, there are not references on application of fluorescence spectroscopy to distinguish between oils from different geographical origin.

The second paper of this chapter includes a study about the capability of fluorescence spectroscopy to distinguish between a set of SC and SM oils. The study is based on applying EEFS and DU-PLS for discriminating between them. The results are presented in the paper *Study of oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis, Grasas y Aceites 56 (4) (2005) 292-297.*

Lee and Seung [34] proposed the NMF algorithm for obtaining a parts-based representation from image data. They showed that this method has some

advantages respect to other decomposition methods, such as PCA, because of the non-negative constraints of the NMF algorithm. These constraints make that only positive solutions can be obtained, which provide a more realistic approximation of the data when working with non-negative data.

Besides image analysis, NMF has been successfully applied to text analysis [34-36], genetics [37], spectra [38,39] and sound analysis [40,41], and curve resolution [42].

NMF is able to handle spectroscopic data because of their non-negative characteristics. However it has not been applied to fluorescence spectroscopy. In this part of the thesis, application of NMF to the fluorescence EEMs of various sets of olive oils is studied. This method is applied with two objectives. First, to decompose the EEMs into meaningful parts and to relate them to the fluorescence species present in oils. Second, to study the capabilities of this method for oil classification. To achieve this second objective, NMF is used together with Fisher's LDA.

The study was carried out using three sets of olive oils. The first one contained three types of commercial olive oils (virgin, pure and olive-pomace). The second one contained EVOOs from the two PDO "Siurana" production area. The last set included the same "Siurana" oils and admixtures of them with 5% (w/w) of OPOs. In all the cases, good classifications were obtained (90-100%). In addition, some of the parts obtained from the NMF decomposition could be related to the fluorescent species of oils. The results of this study are presented in the paper *Application of non-negative matrix factorization combined with Fisher's linear discriminant analysis for classification of olive oil excitation-emission fluorescence spectra, Chemom. Intell. Lab. Syst. (accepted for publication).*

References

[1] R.E. Shaffer, S.L. Rose-Pehrsson, R.A. McGill, *Field Anal. Chem. Tech.* 2 (1998) 179-192.

[2] C. Wittrup, *J. Chemometr.* 14 (2000) 765-776.

[3] N. De Belie, M. Sivertsvik, J. De Baerdemaeker, *J. Sound Vib.* 266 (2003) 625-643.

[4] S.M. Scott, D. James, Z. Ali, W.T. O'Hare, F.J. Rowell, *Analyst* 128 (2003) 966-973.

[5] R. Aparicio, R. Aparicio-Ruíz, *J. Chromatogr. A* 881 (2000) 93-104.

[6] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, *J. Chromatogr. A* 945 (2002) 221-230.

[7] M.J. Dennis, *Analyst* 123 (1998) 151R-156R.

[8] G. Vigli, A. Philippidis, A. Spyros, P. Dais, *J. Agric. Food Chem.* 51 (2003) 5715-5722.

[9] D.L. García-González, L. Mannina, M. D'Imperio, A.L. Segre, R. Aparicio, *Eur. Food Res. Technol.* 219 (2004) 545-548.

[10] A.K. Kiritsakis, *Olive oil. From the tree to the table*, Food and Nutrition Press, Inc., 2nd Ed., Trumbull, 1998.

[11] M. Lees, *Food authenticity and traceability*, Woodhead publishing limited, Cambridge, 2003.

[12] I.J. Wesley, R.J. Barnes, A.E.J. McGill, *JAOCS* 72 (1995) 289-292.

[13] Y.W. Lai, E.K. Kemsley, R.H. Wilson, *Food Chem.* 53 (1995) 95-98.

[14] V. Baeten, M. Meurens, M.T. Morales, R. Aparicio, *J. Agric. Food Chem.* 44 (1996) 2225-2230.

[15] H. Yang, J. Irudayaraj, *JAOCS* 78 (2001) 889-895.

[16] G. Downey, P. McIntyre, A.N. Davies, *J. Agric. Food Chem.* 50 (2002) 5520-5525.

[17] E.C. López-Díez, G. Bianchi, R. Goodacre, *J. Agric. Food Chem.* 51 (2003) 6145-6150.

[18] K. Papadopoulos, T. Triantis, C.H. Tzikis, A. Nikokavoura, D. Dimotikali, *Anal. Chim. Acta* 464 (2002) 135-140.

[19] D.L. García-González, R. Aparicio, *J. Agric. Food Chem.* 51 (2003) 3515-3519.

[20] A. Sayago, M.T. Morales, R. Aparicio, *Eur. Food Res. Technol.* 218 (2004) 480-483.

[21] C. Armanino, R. Leardi, S. Lanteri, G. Modi, *Chemom. Intell. Lab. Syst.* 5 (1989) 343-354.

[22] S. Lanteri, C. Armanino, E. Perri, A. Palopoli, *Food Chem.* 76 (2002) 501-507.

[23] R. Boggia, P. Zunin, S. Lanteri, N. Rossi, F. Evangelisti, *J. Agric. Food. Chem.* 50 (2002) 2444-2449.

[24] F. Marini, F. Balestrieri, R. Bucci, A.D. Magrì, A.L. Magrí, D. Marini, *Chemom. Intell. Lab. Syst.* 73 (2004) 85-93.

[25] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, L.R. Henriques, M.F. Peres, M.P. Simões, P.S. Lopes, *Anal. Bioanal. Chem.* 370 (2002) 1205-1211.

[26] A.D. Shaw, A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, D.B. Kell, *Anal. Chim. Acta* 348 (1997) 357-374.

[27] R. Sacchi, L. Mannina, P. Fiordiponti, P. Barone, L. Paolillo, M. Patumi, A. Segre, *J. Agric. Food Chem. 46* (1998) 3947-3951.

[28] G. Downey, P. McIntyre, A.N. Davies, *Appl. Spectrosc.* 57 (2003) 158-163.

[29] H.S. Tapp, M. Defernez, E.K. Kemsley, *J. Agric. Food Chem.* 51 (2003) 6110-6115.

[30] F. Angerosa, O. Bréas, S. Contento, C. Guillou, F. Reniero, E. Sada, *J. Agric. Food Chem.* 47 (1999) 1013-1017.

[31] G. Bianchi, L. Giansante, A. Shaw, D.B. Kell, *Eur. J. Lipid Sci. Technol.* 103 (2001) 141-150.

[32] Website of the European Commission, http://europa.eu.int/comm/agriculture/foodqual/quali1_en.htm, last access June 10, 2005.

[33] J. Tous, A. Romero, J. Plana, L. Guerrero, I. Díaz, J.F. Hermoso, *Grasas y Aceites* 48 (1997) 415-424.

[34] D.D. Lee, H.S. Seung, *Nature* 401 (1999) 788-791.

[35] C.W. Lee, H. Kang, K. Jung, H.J. Kim, *Lect. Notes Comput. Sc.* 2756 (2003) 470-477.

[36] D. Guillamet, J. Vitrià, *Proceedings of the 11th International Conference on Image Analysis and Processing*, Palermo, 2001, 256-261.

[37] J-P. Brunet, P. Tamayo, T.R. Golub, J.P. Mesirov, *PNAS* 23 (2004) 4164-4169.

[38] G. Buchsbaum, O. Bloch, *Vision Res.* 42 (2002) 559-563.

[39] P. Sajda, S. Du, T.R. Brown, R. Stoyanova, D.C. Shungu, X. Mao, L.C. Parra, *IEEE T. Med. Imaging* 23 (2004) 1453-1465.

[40] H. Asari, *unpublished Technical Note* (2004). http://zadorlab.cshl.edu/asari/nmf.html, last access June 27, 2005.

[41] Y-C. Cho, S. Choi, *Pattern Recogn. Lett.* 26 (2005) 1327-1336.

[42] H-T. Gao, T-H. Li, K. Chen, W-G. Li, X. Bi, *Talanta* 66 (2005) 65-73.

5.2 PAPER

F. Guimet, J. Ferré, R. Boqué
Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis

Rapid Detection of Olive-pomace Oil Adulteration in Extra Virgin Olive Oils from the Protected Denomination of Origin "Siurana" Using Excitation-emission Fluorescence Spectroscopy and Three-way Methods of Analysis

Francesca Guimet, Joan Ferré, Ricard Boqué

Department of Analytical Chemistry and Organic Chemistry, Rovira i Virgili University
C/ Marcel·lí Domingo, s/n E-43007 Tarragona, Catalonia, Spain

ABSTRACT

Extra virgin olive oil (EVOO) is the highest-quality type of olive oil. This makes it also the most expensive. For this reason, it is sometimes adulterated with cheaper oils. One of these is olive-pomace oil (OPO). The protected denomination of origin (PDO) "Siurana" distinction is given to the EVOO produced in a specific area of the south of Catalonia. Here we study the potential of excitation-emission fluorescence spectroscopy (EEFS) and three-way methods of analysis to detect OPO adulteration in PDO "Siurana" olive oils at low levels (5%). First, we apply unfold principal component analysis (unfold-PCA) and parallel factor analysis (PARAFAC) for exploratory analysis. Then, we use the Hotelling T^2 and Q statistics as a fast screening method for detecting adulteration. We show that discrimination between non-adulterated and adulterated samples can be improved using Fisher's linear discriminant analysis (LDA) and discriminant multi-way partial least squares (N-PLS) regression, the latter giving a 100% of correct classification. Finally, we quantify the level of adulteration using N-PLS.

Keywords*: Olive oils; Adulteration; Fluorescence; Three-way methods; Discrimination

1. INTRODUCTION

Olive oil is an economically important product in the Mediterranean countries. It has a fine aroma and a pleasant taste, and it is also known for its health benefits. The quality of olive oil ranges from the high-quality extra virgin olive oil (EVOO) to the low-quality olive-pomace oil (OPO) (or raw residue oil). EVOO is obtained from the fruit of the olive tree (*Olea europea L.*) by mechanical press and without application of refining processes. Its acidity cannot be greater than 1%. Due to its high quality it is the most expensive type of olive oil. For this reason, it is sometimes mislabelled or adulterated. Mislabelling often consists in false labelling concerning the geographic origin or the oil variety of an olive oil [1-3]. Adulteration involves addition of cheaper oils. The most common adulterants found in virgin olive oil are refined olive oil, OPO, synthetic olive oil-glycerol products, seed oils (such as sunflower, soy, maize and rapessed) [4-8] and nut oils (such as hazelnut and peanut oil) [8-10]. In some cases, besides the economic fraud, adulteration may cause serious harm to health, such as happened in 1981 in the case of the Spanish toxic oil syndrome, which affected over 20000 people [11].

Olive pomace is one of the main by-products of oil fruit processing. It contains fragments of skin, pulp, pieces of kernels and some oil. The oil present in the olive pomace undergoes rapid deterioration due to the moisture content that speeds up trialylglycerol hydrolysis. Refined olive-pomace oil is obtained from olive pomace after an extraction with authorized solvents and a refining process, which includes neutralization, deodorization and decolorization. This oil is improved with virgin olive oil to obtain the oil known as OPO [12]. Owing to the low price of this oil, it is sometimes used for adulterating EVOO. For this reason, a rapid method to detect such a practice is important for quality control and labelling purposes [7,13].

Several techniques can be used to detect olive oil adulteration. These include colorimetric reactions, determination of iodine value, saponification value, density, viscosity, refractive index and ultraviolet absorbance [14]. Chromatographic techniques have also been widely applied [9,15]. However, they may be time-consuming and require sample manipulation. To overcome these handicaps, other techniques have been applied. The most noteworthy are headspace-mass spectrometry (HS-MS) [7,8] and spectroscopic techniques, such as near-infrared (NIR), mid-infrared (MIR), Fourier transform infrared (FT-IR), Fourier transform

Raman (FT-Raman) [4-6,13-19], nuclear magnetic resonance (NMR) [17,20] and chemiluminescence [21]. Recently, Mignani et al. have proposed spectral nephelometry for recording EVOO fingerprints [22]. Usually spectroscopic techniques are applied together with multivariate analysis. Hence, supervised pattern recognition methods are commonly used [4-6,10,13,16,18-20].

Fluorescence spectroscopy has been used in the past for determining the authenticity of olive oils [14]. However, few papers have been published in recent years on the use of fluorescence in vegetable oils. Sayago et al. [10] applied fluorescence spectroscopy for detecting hazelnut oil adulteration in virgin olive oils. This technique has some advantages such as its speed of analysis, that it is reagentless, and that small amounts of sample are required. Kyriakidis and Skarkalis [23] showed that emission fluorescence spectra of virgin olive oils between 400 and 700 nm measured at excitation wavelength 365 nm have clear differences compared to the spectra of other vegetable oils. Virgin olive oils present two low peaks at 445 and 475 nm (related to conjugated hydroperoxides), one intense peak at 525 nm (due to vitamin E) and another peak at 681 nm (due to chlorophylls). The very low intensity of the peaks at 445 and 475 nm is due to their large content on monounsaturated fatty acids and phenolic antioxidants, which provide more stability against oxidation. All refined oils show only one intense peak at 445 nm. It is due to fatty acid oxidation products formed as a result of the large percentage of polyunsaturated fatty acids present in these oils.

In recent years, instrumental improvements and the availability of software specially designed to extract the information contained in spectra have contributed to the development of fluorescence spectroscopy. Hence, it is possible to record one fluorescence excitation-emission matrix (EEM) for each sample, i.e. a set of emission spectra recorded at several excitation wavelengths. EEMs had already been used in 1984 by Wolfbeis and Leiner [24] to characterize four types of edible oils. More recently, Scott et al. [25] discriminated between different vegetable oils and detected adulterations in EVOO from their EEMs. They applied simplified fuzzy adaptive resonance theory mapping, traditional back propagation and radial basis function neural networks. We have already shown the potential of excitation-emission fluorescence spectroscopy (EEFS) and three-way methods of analysis to distinguish between commercial Spanish olive oils [26,27].

The aim of this paper is to develop a fast screening method based on EEFS and three-way methods of analysis for detecting adulterations of OPO at 5% level in EVOO from the protected denomination of origin (PDO) "Siurana". This is a prestigious distinction given to the EVOO produced in a specific area of the south of Catalonia, Spain. These oils come mainly from *arbequina* olives and they are guaranteed to be of high quality [28]. We first apply unfold principal component analysis (unfold-PCA) and parallel factor analysis (PARAFAC) for exploratory analysis. Then, we compare the ability of Hotelling T^2 and Q statistics, Fisher's linear discriminant analysis (LDA) and discriminant multi-way partial least squares (N-PLS) regression for discriminating between EVOO and adulterated samples. Finally, we quantify the level of adulteration by using the N-PLS regression method.

2. EXPERIMENTAL

2.1 Samples

In the present work, 29 EVOO from the PDO "Siurana" and 5 commercial OPO were analysed. A set of 13 samples was selected from the EVOO group and was used as a test set for the chemometric methods. The test set was used for preparing admixtures (AD) of EVOO containing 5% (w/w) of OPO. The AD were prepared as follows: each sample of the test set was used for preparing five AD, each one containing a different OPO. Thus, 65 AD were prepared. Samples were stored in amber glass bottles under nitrogen atmosphere. The samples were analysed without any prior treatment. All samples were measured in duplicate and the mean value of each sample was always used.

2.2 Instrumentation and software

EEMs were measured with an Aminco Bowman series 2 luminescence spectrometer equipped with a 150 W xenon lamp and 10 mm quartz cells. The instrument detector was operated using the EmL/Ref channel. Excitation and emission ranges were λ_{ex} = 300-390 nm and λ_{em} = 415-600 nm, respectively. Measuring emission above excitation prevented Rayleigh scatter. The step size and band-pass of both monochromators were 5 and 4 nm, respectively. The scan rate was 30 nm s^{-1}. The instrument software was used to correct the EEMs for deviations in the ideality of the lamp, monochromators and detector [29,30]. EVOO

and the AD were measured applying a 600 V voltage, whereas OPO were measured at 570 V, in order to avoid detector saturation.

Data were exported to ASCII code and processed with Matlab software (version 6.5) [31]. Unfold-PCA, PARAFAC, Hotelling T^2 and Q statistics were calculated with the PLS-Toolbox [32]. The Fisher's LDA algorithm was built in-house.

3. RESULTS AND DISCUSSION

3.1. Exploratory analysis

3.1.1 Spectra

Fig. 1 shows some of the measured EEMs. They were obtained from two EVOO (a) and (d), an OPO (b) and a 5% AD (c). Most EVOO analysed had EEMs similar to that plotted in Fig. 1a. The most intense peaks of the spectra of this type of oil appear between λ_{ex} = 300-390 nm; λ_{em} = 500-600 nm and they are due to Vitamin E. At lower λ_{em} there are some less intense peaks, which are related to conjugated hydroperoxides formed as a result of oil oxidation. The low intensity of these peaks indicates that EVOO are highly stable against oxidation. However, sample 6 has an abnormally high fluorescence intensity below λ_{em} = 500 nm (Fig. 1d), being more similar to the AD in this range (Fig. 1c). All OPO showed only one wide peak between λ_{ex} = 340-390 nm; λ_{em} = 415-550 nm (Fig. 1b). This means that their content on conjugated hydroperoxides is higher as a result of a greater oxidation [12,23]. It can be seen that adulteration mainly causes an increase of fluorescence at emissions below 500 nm (Fig. 1c). However, it is not always easy to detect adulterations by visual inspection of the spectra. For this reason, chemometric methods must be used for discrimination. As it has been shown, recording entire EEMs enables to extract more information from the samples compared to when a single fluorescence spectrum is measured. In this case, only the changes produced at one λ_{ex} can be observed.
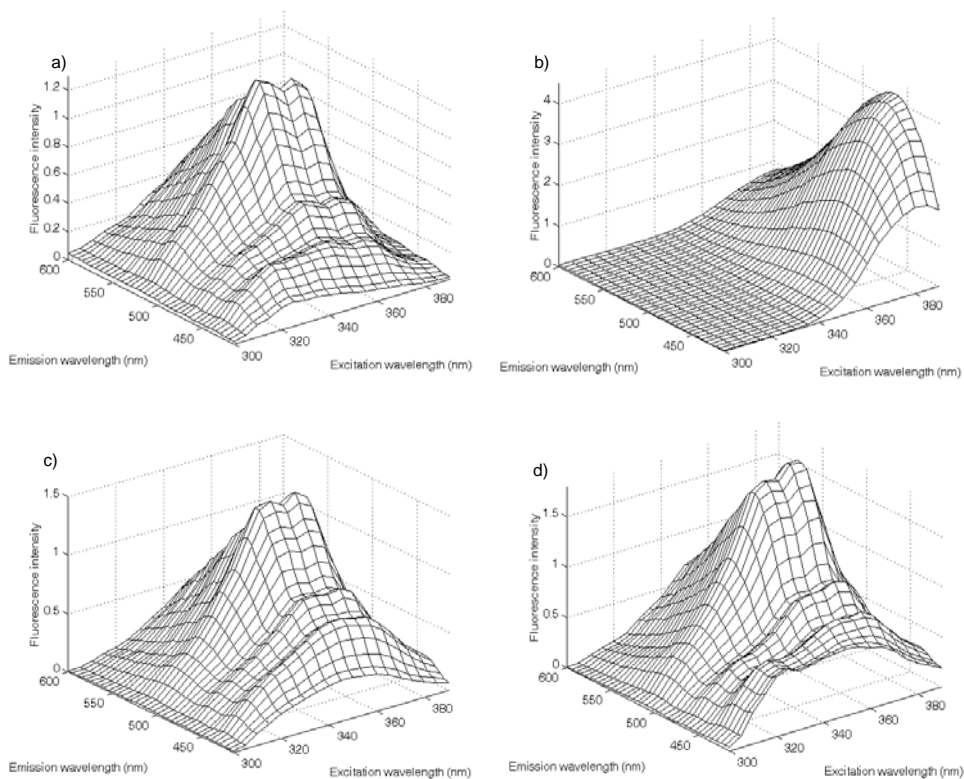
Fig. 1. EEMs of two EVOO (a) and (d), an OPO (b) and a 5% AD (c). Range: λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm.

3.1.2 Unfold principal component analysis

In order to do a preliminary study of the data and to detect possible outliers, we applied unfold-PCA to the EVOO group. The EEMs were first arranged in a $I \times J \times K$ three-way array ($29 \times 38 \times 19$), where the indexes I, J and K refer to samples, emission wavelengths (λ_{em}) and excitation wavelengths (λ_{ex}), respectively. Then the array was unfolded to a matrix of size $I \times JK$ (29×722). Before applying unfold-PCA, the matrix was column mean-centered. Figs. 2 and 3 show the score and loading plots of the first three PCs (94.57% expl. var.).
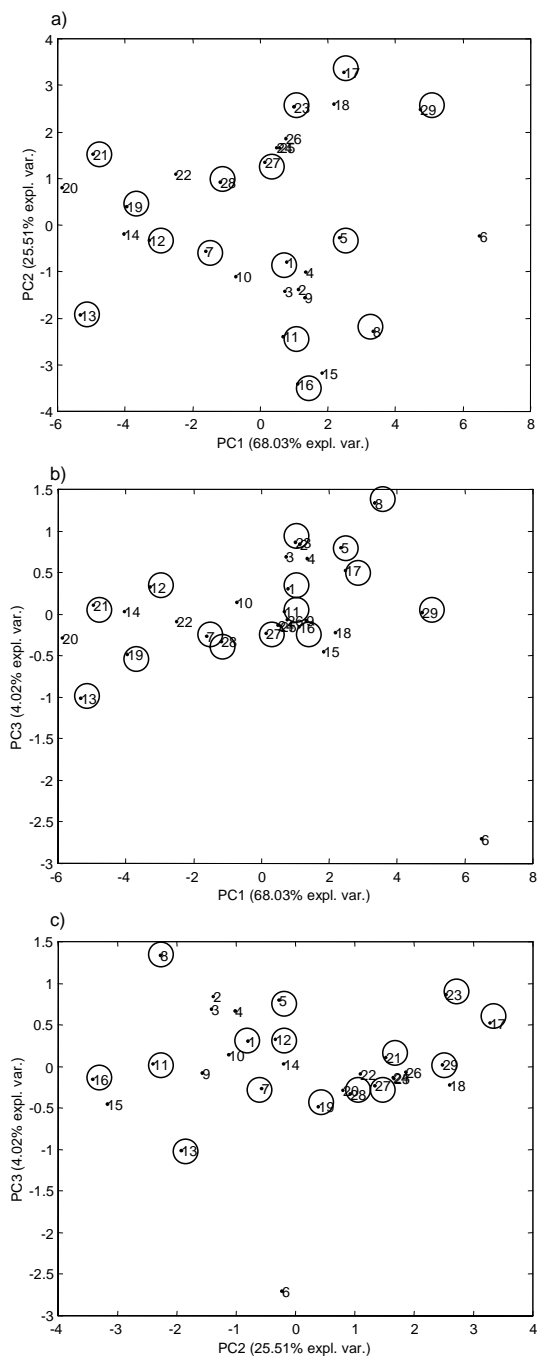
Fig. 2. Score plots of the first three PCs from unfold-PCA calculated to the 29 EVOO: PC1 vs. PC2 (a), PC1 vs. PC3 (b) and PC2 vs. PC3 (c). The samples marked with circles were used as training set for the discrimination models.
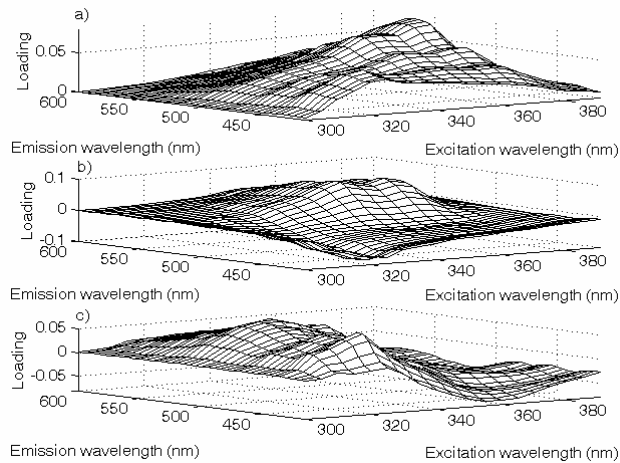
Fig. 3. Refolded loading plots of the first three PCs from unfold-PCA calculated from the 29 EVOO: PC1 (a), PC2 (b), and PC3 (c).

It can be seen that, in general, sample 6 appears quite separated from the rest of EVOO, especially along PC3. The special features of this sample have been commented above. Since it was very different from the rest of EVOO, sample 6 was removed from the sample set and not considered any longer.

As it has been commented, in the range considered (λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm), the main fluorescent species present in EVOO are Vitamin E and some oxidation products (depending on the state of degradation). This is captured on the loadings of the first three PCs (Fig. 3), because they have influence of the wavelengths where these species emit. Thus, the loadings above λ_{em} = 500 nm explain variations due to Vitamin E, whereas those below λ_{em} = 500 nm describe oxidation products.

3.1.3 Parallel factor analysis

To study the effect of adulteration on the oils, we decomposed the EEMs of each group (EVOO, OPO, and 5% AD) using PARAFAC. The sizes of the three-way arrays (samples $\times$ number of λ_{em} $\times$ number of λ_{ex}) were as follows: $28 \times 38 \times 19$ (EVOO), $65 \times 38 \times 19$ (AD) and $5 \times 38 \times 19$ (OPO).

All PARAFAC models were fitted using non-negativity constraints on all modes. The optimum number of factors was selected on the basis of residual analysis and split-half analysis [33]. In all cases, the best solution was obtained using three factors. Fig. 4 shows the emission and the excitation loadings related to the EVOO, 5% AD and OPO sample sets. The excitation loadings are more difficult to interpret than the emission loadings because of the lack of fluorescence excitation spectra in the literature.
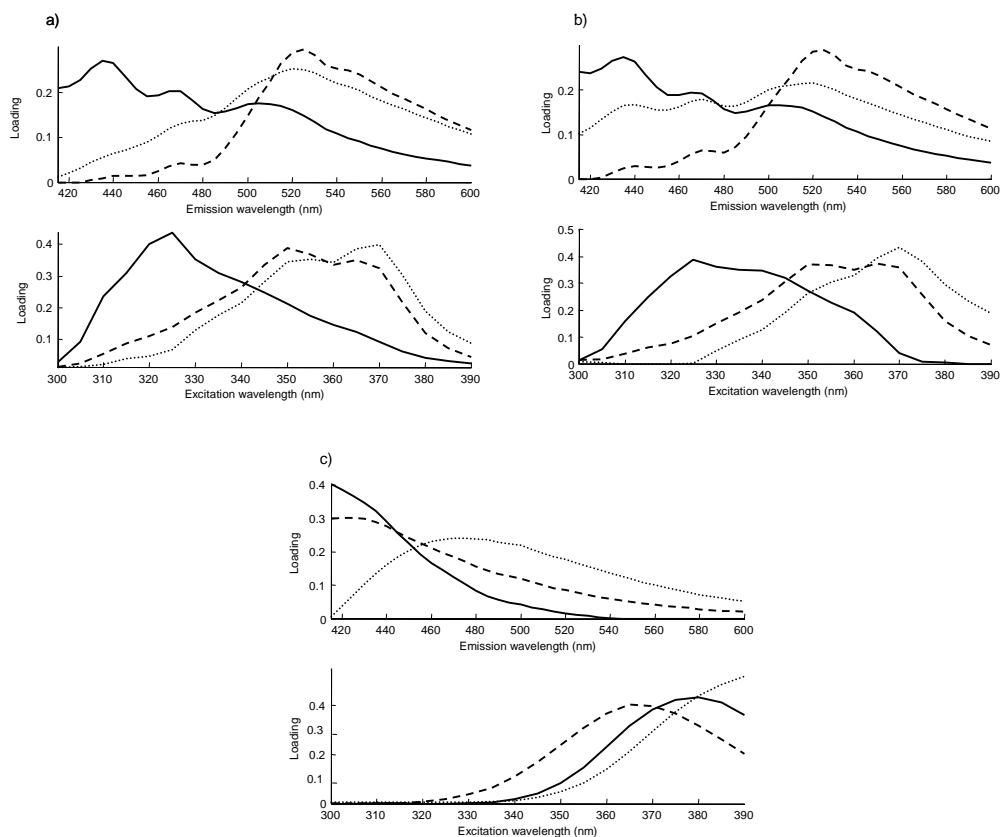

Fig. 4. PARAFAC excitation and emission loadings of EVOO (a), 5% AD (b) and OPO (c). Top: emission loadings, bottom: excitation loadings. Factor 1 (—), factor 2 (– – –), factor 3 (…). Var. expl.: 99.81% (a), 99.83% (b), and 99.98% (c).

Factor 1 is not much modified by adulteration and the greatest changes are produced on the shape and position of the excitation profile. For EVOO, factors 2 and 3 have contributions of Vitamin E (the peak around λ_{em} = 525 nm) and oxidation products (the peaks around λ_{em} = 445 nm and λ_{em} = 475 nm) (see Fig. 4a, emission loadings). Nevertheless, the influence of the oxidation products is less important on factor 2 and adulteration only produces a slight rise on these peaks (Fig. 4a and b). On the contrary, factor 3 undergoes more changes due to adulteration and there is a clear rise of the peaks related to oxidation products and a decrease of the Vitamin E peak. As far as OPO are concerned, the emission loadings of factors 2 and 3 have only one wide peak (see Fig. 4c) that is attributed to the very high content on oxidation products due to the advanced state of oxidation of these oils.

Next, we projected the 5% AD onto the PARAFAC model of the EVOO. The score plot of factor 1 versus factor 3 (Fig. 5) clearly shows the differences between the types of oils. The AD tend to have the highest scores on factor 3, meaning that adulteration is at least partially described by this factor. This agrees with the fact that factor 3 undergoes great changes as a result of adulteration, as it was explained above. In spite of the trend seen on the score plot, there is an overlap between the groups of oils.
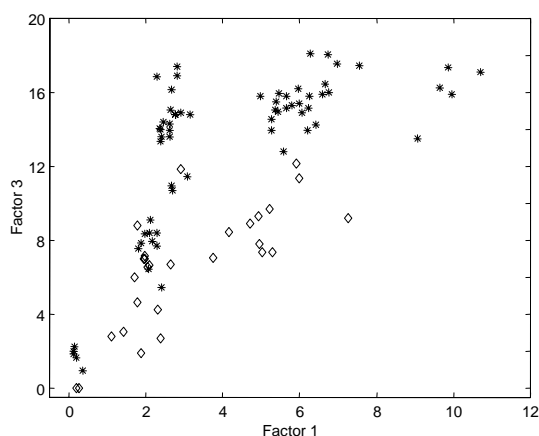

Fig. 5. Score plot of the PARAFAC model. Training set: EVOO (◊); projected set: 5% AD (✳).

3.2 Discrimination

3.2.1 Selection of the training set

In order to discriminate between EVOO and the 5% AD, we applied different chemometric methods. To select the training set to be used, we applied unfold-PCA to the mean-centered matrix of the EVOO group. The size of this matrix was 28×722. The number of PCs was selected by leave-one-out cross-validation (five PCs, 99.49% expl. var.). The loadings and score plots are very similar to those in Figs. 2 and 3. The selection of the training set was made on the basis of the score plots and the dendrogram obtained after clustering the scores of the first five PCs. The criterion was to cover all the domain of variability. The 15 selected samples are marked inside a circle in Fig. 2a-c. As explained in Section 2.1, the rest of EVOO were used for preparing the 5% AD and were used as test set.

3.2.2 Hotelling T^2 and Q statistics

The Hotelling T^2 statistic is calculated for the systematic part of the variation of the data, i.e. the Hotelling T^2 plot represents the projection of each new measurement onto the plane defined by the principal components (PCs). The Q statistic is calculated for the residual part, i.e. the Q plot represents the squared distance of each new measurement perpendicular to the plane defined by the PCs. In this paper we apply these statistics to establish a fast screening method for detecting OPO adulteration. EVOO are considered to be normal samples and a PCA model is built with them. Then, the adulterated samples are expected to fall out of the confidence limits of the PCA model. Thus, the Hotelling T^2 and Q statistics are easily applicable for detecting adulterations because only non-adulterated samples have to be measured for building the model.

We calculated unfold-PCA on the 15×722 column mean-centered matrix containing the training set selected above. We used five PCs (99.64% expl. var.), selected by leave-one-out cross-validation. Afterwards, we projected the rest of EVOO and their AD onto this model. Fig. 6 shows the score plot of the training set and the projected samples using the first two PCs (96.13% expl. var.). The loadings were almost identical to those in Fig. 3. In spite of the fact that the two groups of oils appear quite mixed up, there are more samples having negative scores on PC2 within the AD group. These values are related to oxidation products (the negative loadings below λ_{em} = 500 nm, Fig. 3b). This indicates again that adulteration is mainly detected by a rise on oxidation products content. On the contrary, EVOO

tend to have the highest scores on PC2, which are related to Vitamin E (the positive loadings above λ_{em} = 500 nm, Fig. 3b), and the lowest on PC1.
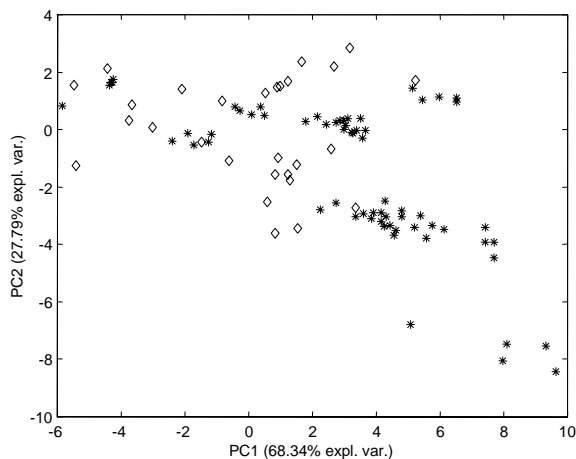

Fig. 6. Score plot from unfold-PCA calculated to the training set formed by the 15 EVOO selected: EVOO (◊), and 5% AD (✳).

We calculated the limits of the Hotelling T^2 and Q statistics from the training set (the fifteen selected EVOO) with a 95% of confidence. The limits were 23.28 for Hotelling T^2 and 0.12 for Q. The Q limit was calculated using the Jackson and Mudholkar's method [34]. Fig. 7 shows the enlarged joint plot of the two statistics for the projected samples (considering only the region near the confidence limits) and Table 1 summarizes the percentage of correct assignation to the groups. For EVOO, the best results were obtained for Hotelling T^2. This statistical test assigned all the EVOO to the correct group. In contrast, the AD had better percentages of Q. This means that EVOO are best described by the variation contained in the model and the AD are best described by the residual part, since adulteration is not included in the model. The percentage of correct classification of the AD was 97% considering the Q statistical and only 68% regarding to Hotelling T^2 (Table 1). Thus, other discriminant methods should be applied to improve the detection of OPO adulteration in EVOO.
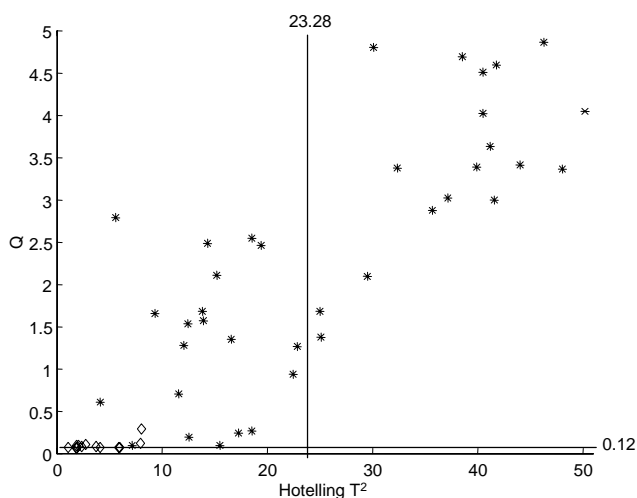
Fig. 7. Enlarged joint plot of the Hotelling T² and Q statistics (95% of confidence): EVOO (◊), and 5% AD (✳).

Table 1. Results of Hotelling T² and Q statistics

Samples	Correct classification (%)	
	Hotelling T²	Q
EVOO	100	85
5% AD	68	97

3.2.3 Fisher's linear discriminant analysis (LDA)

Using supervised pattern recognition methods usually improves discrimination, because the subgroups made up of the different types of samples are included in the training set and are used for developing the classification rules. These rules are later used for allocating new and unknown samples to the most probable subgroup (class). Fisher's LDA (or canonical variate analysis) is a supervised pattern recognition method. It finds the directions in multivariate space for which the difference between the groups' means is as large as possible compared to the within-group variance, i.e. the directions that discriminate as much as possible between all groups [35, 36]. One problem of applying this method to spectroscopic data is collinearity. If a large number of variables are used, it will not be possible to calculate the inverse of the covariance matrices involved. In order to overcome this

problem, we first applied unfold-PCA to the EEMs to reduce dimensionality. Then, we applied Fisher's LDA to the unfold-PCA scores.

We applied Fisher's LDA to test its ability for discriminating between EVOO and 5% AD. Before applying Fisher's LDA, unfold-PCA was computed to reduce dimensionality. It was computed on the same 15×722 unfolded and column mean-centered matrix used in Section 3.2.2. Again five PCs were considered for computing the model (99.64% expl. var.). The unfold-PCA scores of these 15 EVOO were used as class 1 of the training set. Since adulterated samples have also to be included in the training set, we prepared mixtures (at 5% adulteration level) of the oils in class 1 and four of the five OPO, making a total of 28 AD. They were prepared by adding the OPO randomly. These samples were projected onto the unfold-PCA model of the EVOO samples and their scores were used as class 2 of the training set. Afterwards, we projected the rest of EVOO (13 samples) and their AD (65 samples), whose scores formed the test set of Fisher's LDA. This procedure enabled to have independent samples on the two sets. In addition, one of the adulterants of the test set was not included in the training set. This was done in order to test the model against new adulterants not included in the calibration data.

The training set was used to compute the Fisher's linear discriminant function (LDF) or canonical variate. Then, the samples of the training and test sets were projected onto the LDF and their Fisher's LDA scores calculated. Fig. 8 shows the score plot of Fisher's LDA. The boundary between classes was set taking the centroid of each class and drawing a line half-way between the two centroids [36]. Table 2 shows the percentage of correct classification for the training set (recognition ability) and the test set (prediction ability). All the samples on the training set were classified in the correct class. For the test set, high percentages of correct classification were obtained (85% for EVOO and 98% for AD). The AD of the test set prepared from the new OPO not included in the training set were correctly classified. Comparing these results with those obtained from Hotelling T^2 and Q statistics (Table 1), it should be noticed that both methods have similar ability to recognize non-adulterated samples. In the case of the AD, Fisher's LDA is slightly better than Q and very superior than Hotelling T^2.
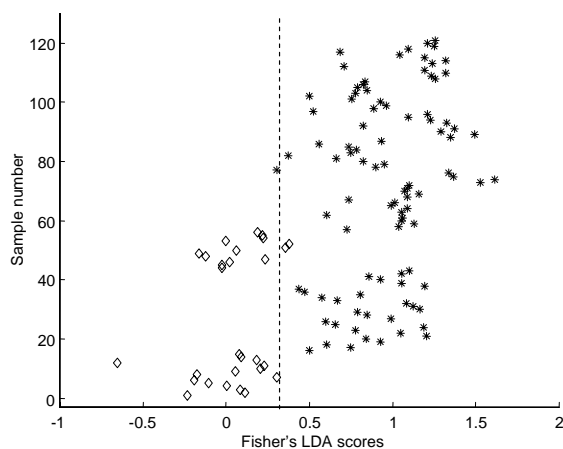
Fig. 8. Score plot from Fisher's LDA (training and test sets): EVOO (◊), and 5% AD (✳). Boundary between classes (– – –).

Table 2. Results of Fisher's LDA

Class	Correct classification (%)	
	Recognition ability	Prediction ability
EVOO	100	85
5% AD	100	98

3.2.4 Discriminant multi-way partial least squares regression

Another supervised pattern recognition method is discriminant N-PLS. We checked if discrimination between EVOO and 5% AD could be improved by using this method. The training set ($43 \times 38 \times 19$) (samples$\times$number of $\lambda_{em}\times$number of λ_{ex}) contained the 15 EVOO selected in section 3.2.1 and the same 28 AD made from these oils, used in Fisher's LDA. The test set ($78 \times 38 \times 19$) contained the rest of EVOO (13 EEMs) and their AD (65 EEMs). As in Fisher's LDA, one of the adulterants of the test set was not included in the training set. As we had only two classes, we used discriminant N-PLS1. The dependent variable **y** had 0 for objects belonging to the EVOO group (class 1) and 1 for objects belonging to the 5% AD group (class 2).

Prior to the application of the N-PLS model, the data were centered across the first mode (samples). After applying leave-one-out cross-validation, we selected six factors to be considered in the model (var. expl. $\mathbf{y}$ = 99%, var. expl. $\underline{\mathbf{X}}$ = 85%, calibration error (RMSEC = 0.2), cross-validation error (RMSECV = 0.2)). Then we projected the test set. The prediction error (RMSEP) was around 0.2. Fig. 9 shows the N-PLS scores of the training and the test sets of the first two factors (78.67% of expl. var.) and Fig. 10, the loadings of these factors. Discrimination between the two classes was achieved by a combination of factors 1 and 2. The loading plot shows that factor 1 has mainly contributions of oxidation products because of the presence of two peaks around λ_{em} = 440 nm and λ_{em} = 470 nm (Fig. 10a). Some contribution of Vitamin E can also be observed (λ_{em} = 525 nm). On the contrary, factor 2 has more influence of Vitamin E.

The criterion to assign one object to one class was that predicted $\mathbf{y}$ smaller than 0.5 was interpreted as belonging to the EVOO class and $\mathbf{y}$ predicted larger than 0.5 was interpreted as belonging to the AD class. On the basis of this criterion, a 100% of correct classification was obtained for the training and the validation sets. Fig. 11 shows the plot of the predicted vs. the measured values for the validation set. As it can be seen, there is no overlap between the two classes.
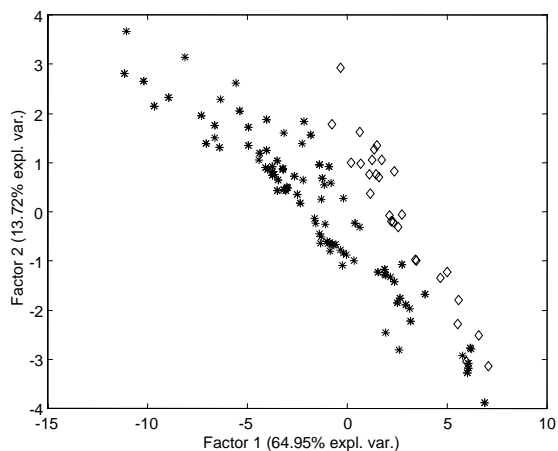

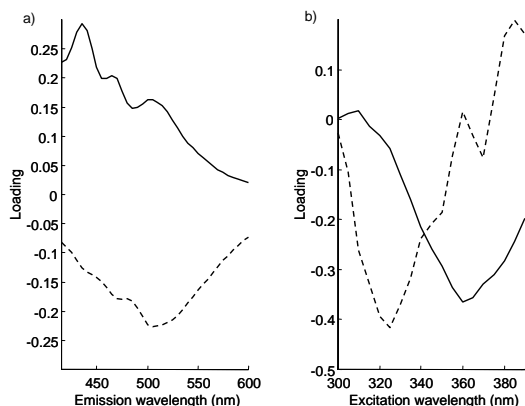Fig. 9. Score plot from discriminant N-PLS. Factor 1 vs. factor 2 (training and test set): EVOO (◊), and 5% AD (✳).

Fig. 10. Discriminant N-PLS emission (a) and excitation (b) loadings: factor 1 (—), factor 2 (− − −).
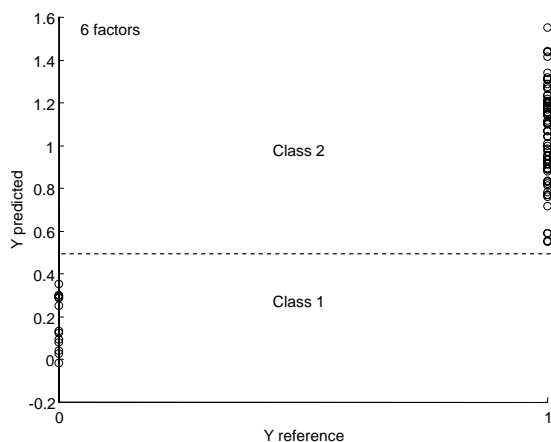
Fig. 11. Predicted vs. measured values for the test set obtained by discriminant N-PLS using six factors. Boundary between classes (− − −).

3.3 Quantification

We applied the N-PLS regression method to check if besides detecting OPO adulteration, N-PLS was also able to quantify it at the level of 5%. In order to ensure not having predicted values at the extremes of the model and thereby avoiding large prediction errors, we included 10% AD in the calibration set. Thus, the calibration set contained the same samples used as training set for Fisher's LDA and discriminant N-PLS plus a group of 10% AD. The latter was prepared

similarly to the 5% AD (i.e. from the 15 EVOO of the training set by adding the same four OPO). In this way, one of the OPO used for preparing the AD of the test set remained out of the calibration set. Hence, the calibration set was composed of 71 samples (15 EVOO, 28 AD at 5% and 28 AD at 10% level). The test set consisted of the rest of EVOO (13 samples) and their 5% AD (65 EEMs). Prior to the application of the N-PLS model, the data were centered across the first mode (samples). The optimum number of factors was selected by leave-one-out cross-validation. We found six factors to be significant (99.60% expl. var. ($\underline{\mathbf{X}}$), 87.36% expl. var. ($\mathbf{y}$)) and the RMSECV and RMSEC were 1.6% and 1.4%, respectively. The prediction error for the test set was RMSEP = 1.2%.

4. CONCLUSIONS

We have shown the potential of EEFS combined with three-way methods of analysis to detect OPO adulteration in EVOO from the PDO "Siurana" at 5% level. Unfold-PCA combined with Hotelling T^2 and Q statistics can be used as a fast screening method for discriminating between non-adulterated and adulterated oils. The Q statistic gave best classifications than Hotelling T^2 for adulterated samples (97%). The reason is that Q is calculated for the residual part of the projected samples and thus finds more easily variations that are not included in the model. The results were slightly improved by using Fisher's LDA. Subsequently, discriminant N-PLS was found to be superior for detecting adulteration and a 100% of correct classification was obtained. Finally, we quantified OPO adulteration around 5% level using the N-PLS regression method, obtaining a prediction error of 1.2%.

In future work we aim to build a robust methodology based on EEFS and three-way methods of analysis for detecting other adulterations in EVOO.

REFERENCES

[1] R. Aparicio, M.T. Morales, V. Alonso, J. Agric. Food Chem. 45 (1997) 1076.

[2] R. Sacchi, L. Mannina, P. Fiordiponti, P. Barone, L. Paolillo, M. Patumi, A. Segre, J. Agric. Food Chem. 46 (1998) 3947.

[3] I. Marcos Lorenzo, J.L. Pérez Pavón, M.E. Fernández Laespada, C. García Pinto, B. Moreno Cordero, L.R. Henriques, M.F. Peres, M.P. Simões, P.S. Lopes, Anal. Bioanal. Chem. 374 (2002) 1205.

[4] I.J. Wesley, R.J. Barnes, A.E.J. McGill, JAOCS 72 (1995) 289.

[5] V. Baeten, M. Meurens, M.T. Morales, R. Aparicio, J. Agric. Food Chem. 44 (1996) 2225.

[6] G. Downey, P. McIntyre, A.N. Davies, J. Agric. Food Chem. 50 (2002) 5520.

[7] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, J. Chromatogr. A 945 (2002) 221.

[8] R. Goodacre, S. Vaidyanathan, G. Bianchi, D.B. Kell, Analyst 127 (2002) 1457.

[9] G.P. Blanch, M.M. Caja, M.L. Ruiz del Castillo, M. Herraiz, J. Agric. Food Chem. 46 (1998) 3153.

[10] A. Sayago, M.T. Morales, R. Aparicio, Eur. Food Res. Technol. 218 (2004) 480.

[11] G.M. Wood, P.T. Slack, J.B. Rossell, P.J. Mann, P.J. Farnell, J. Agric. Food Chem. 42 (1994) 2525.

[12] A.K. Kiritsakis, Olive Oil: From the Tree to the Table, 2nd ed., Food & Nutrition Press Inc., Trumbull, 1998, p. 237.

[13] H. Yang, J. Irudayaraj, JAOCS 78 (2001) 889.

[14] J. Gracian, Analysis and Characterization of Oils, Fats and Fat Products, vol. 2, Wiley, London, 1968, p. 315.

[15] R. Aparicio, R. Aparicio-Ruiz, J. Chromatogr. A 881 (2000) 93.

[16] Y.W. Lai, E.K. Kemsley, R.H. Wilson, Food Chem. 53 (1995) 95.

[17] M.J. Dennis, Analyst 123 (1998) 151R.

[18] L. Küpper, H.M. Heise, P. Lampen, A.N. Davies, P. McIntyre, Appl. Spectrosc. 55 (2001) 563.

[19] E.C. López-Díez, G. Bianchi, R. Goodacre, J. Agric. Food Chem. 51 (2003) 6145.

[20] G. Vigli, A. Philippidis, A. Spyros, P. Dais, J. Agric. Food Chem. 51 (2003) 5715.

[21] K. Papadopoulos, T. Triantis, C.H. Tzikis, A. Nikokavoura, D. Dimotikali, Anal. Chim. Acta 464 (2002) 135.

[22] A.G. Mignani, P.R. Smith, L. Ciaccheri, A. Cimato, G. Sani, Sens. Actuators B 90 (2003) 157.

[23] N.B. Kyriakidis, P. Skarkalis, J. AOAC Int. 83 (2000) 1435.

[24] O.S. Wolfbeis, M. Leiner, Mikrochim. Acta 1 (1984) 221.

[25] S.M. Scott, D. James, Z. Ali, W.T. O'Hare, F.J. Rowell, Analyst 128 (2003) 966.

[26] F. Guimet, J. Ferré, R. Boqué, F.X. Rius, Anal. Chim. Acta 515 (2004) 75.

[27] F. Guimet, R. Boqué, J. Ferré, J. Agric. Food Chem. 52 (2004) 6673.

[28] Protected Denomination of Origin "Siurana", http://www.siurana.info/, last access: June 16, 2004.

[29] SLM AMINCO, Technical Note No. 101, Urbana.

[30] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 2nd ed., Kluwer Academic/Plenum Publishers, New York, 1999, p. 25.

[31] Matlab, The Mathworks, South Natick, MA, USA. http://www.mathworks.com, last access: August 20, 2004.

[32] Toolbox for Matlab by Eigenvector Research Inc. http://www.eigenvector.com/, last access: May 27, 2004.

[33] P. Nomikos, J.F. MacGregor, Technometrics 37 (1995) 41.

[34] J. Naes, T. Isaksson, T. Fearn, T. Davies, A User-Friendly Guide to Multivariate Calibration and Classification, NIR publications, Chichester, 2002, p. 221.

[36] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Vol 20, Part B, Elsevier, Amsterdam, 1998, p. 207.

5.3 PAPER

F. Guimet, R. Boqué, J. Ferré
Study of oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis

Study of Oils from the Protected Denomination of Origin "Siurana" using Excitation-emission Fluorescence Spectroscopy and Three-way Methods of Analysis

Francesca Guimet, Ricard Boqué, Joan Ferré

Department of Analytical and Organic Chemistry, Rovira i Virgili University
C/ Marcel·lí Domingo, s/n E-43007 Tarragona, Catalonia, Spain

SUMMARY

The production area of the Protected Denomination of Origin (PDO) "Siurana" olive oils can be divided into two regions: "Montsant" and "Camp de Tarragona", with different orography, soil characteristics and climatic conditions. As a result of these differences, oils have different stability. Here we show that excitation-emission fluorescence can detect this different stability. Oils from "Montsant" region (more stable) show more fluorescence intensity above 500 nm (due to vitamin E) and less below 500 nm (due to oxidation products) than those from "Camp de Tarragona". Discrimination between oils from the two PDO "Siurana" regions was achieved by means of discriminant unfold partial least squares regression, giving a percentage of correct classification of 94% for "Siurana-Camp" and 100% for "Siurana-Montsant" oils.

Keywords: Fluorescence spectroscopy; Partial least squares; PDO "Siurana"; Three-way methods of analysis; Virgin olive oil

1. INTRODUCTION

Olive oil is an important food product, especially in the Mediterranean countries, and its characterisation is an issue of current interest. At present, there is no single analytical index for characterising it, and numerous physical and chemical parameters (such as free acidity, peroxide value, K_{232} and K_{270} among others) must be measured to determine its geographical origin and/or olive variety (Aparicio, 1988; Tous *et al.*, 1997; Lanteri *et al.*, 2002; Marini *et al.*, 2004). This requires application of several techniques, such as chromatographic and spectroscopic, which have often the problem of being time-consuming. Other techniques (e.g. headspace-mass spectrometry (Marcos *et al.*, 2002) and nuclear magnetic resonance (Sacchi *et al.*, 1998)) have also been applied for this purpose. However some of them are not always present in the quality control laboratories. Fluorescence spectroscopy has also been applied to assess the authenticity of olive oils (Kyriakidis and Skarkalis, 2000; Sayago *et al.*, 2004). The main advantages of this technique are that no complex sample preparation is required and that it is very fast. With excitation-emission fluorescence spectroscopy (EEFS) a set of fluorescence spectra can be recorded at different excitation wavelengths (λ_{ex}) yielding an excitation-emission matrix (EEM). Such amount of data enables more information to be obtained compared to when a single spectrum is measured. There are some examples of the application of EEFS to olive oils in the literature (Scott *et al.*, 2003; Guimet *et al.*, 2004 a, b). Multivariate chemometric methods have been widely applied to the data obtained from analytical techniques in order to differentiate between olive oil cultivars and/or varieties (Armanino *et al.*, 1989; Boggia *et al.*, 2002; Lanteri *et al.*, 2002, Marini *et al.*, 2004).

Spain is one of the most important olive oil producers and currently has nine protected denominations of origin (PDO) (five in Andalusia, two in Catalonia, one in Castilla La Mancha and one in Aragon). The PDO label defines the origin of the oils and the varieties used and guarantees the production and transformation of the product in its geographic areas. The products included in the same PDO have some exceptional characteristics in common, including both analytical specifications and organoleptic properties. In this work we focus on olive oils from the PDO "Siurana". The PDO "Siurana" production area is a strip of the province of Tarragona, in Catalonia (Spain). "Siurana" oils, which are mainly obtained from *arbequina* olives (more than 90%), may also contain small proportions of *royal* and *morruda* varieties. The physicochemical characteristics of "Siurana" oils, such as

fatty acids, polyphenols, and stability vary depending on several factors. One of the most important is the state of ripeness of the olives. The geographical origin of olives (which includes orography, soil characteristics and climatic conditions) has also great influence (Tous *et al.*, 1997). The PDO "Siurana" production area can be divided into two regions: one inner and mountainous region, which is denominated "Montsant" and another closer to the sea, which is denominated "Camp de Tarragona". "Siurana-Montsant" (SM) oils have more content on oleic acid and phenolic compounds and less content on linoleic acid than Siurana-Camp (SC) oils. This makes SM oils more stable against oxidation than SC oils. These two types of oils have also different sensory properties. SM oils are in general bitterer and their colour is more intense, whereas SC oils are sweeter and more fluid. The differences between SM and SC oils are explained in more detail elsewhere (Tous *et al.*, 1997).

In this paper we show the potential of the fast technique EEFS to differentiate between oils from the two regions of PDO "Siurana". The method discriminant unfold partial least squares regression (DU-PLSR) was applied due to its ability for discriminating.

2. EXPERIMENTAL

2.1. Samples

29 extra virgin olive oils from the PDO "Siurana" were analysed. Samples 1-16 came from SC region and samples 17-29 came from SM region. All the samples analysed came mainly from *arbequina* olives and were obtained during the same harvesting year (2003). The ripeness stage of the two sets of olives was similar. Table 1 shows the locality of origin of the samples. The oils were stored in amber glass bottles under nitrogen atmosphere at 7°C and were analysed without any prior treatment. All samples were measured in duplicate.

2.2. Instrumentation and software

EEMs were measured with an Aminco Bowman series 2 luminescence spectrometer equipped with a 150 W xenon lamp and 10 mm quartz cells. The instrument detector was operated using the EmL/Ref channel and applying a 600 V voltage. Excitation and emission ranges were λ_{ex} = 300-390 nm and λ_{em} = 415-600 nm, respectively. Measuring emission above excitation prevented Rayleigh scatter. The step size and band-pass of both monochromators were 5 nm and 4 nm,

respectively. The scan rate was 30 nm s⁻¹. The instrument software was used to correct the EEMs for deviations in the idealiy of the lamp, monochromators and detector (Lakowicz, 1999).

Data were exported to ASCII code and processed with Matlab software (version 6.5) (Matlab, 2002) and the PLS-Toolbox (version 3.0) (PLS-Toolbox, 2003).

Table 1. Samples and their locality of origin

Sample number	Origin	Region
1, 10	La Selva del Camp	SC
2-4	Reus	SC
5	Llorenç del Penedès	SC
6	Valls	SC
7	Alcover	SC
8	Constantí	SC
9, 16	La Canonja	SC
11	Maspujols	SC
12	Alforja	SC
13	Les Borges del Camp	SC
14	Almoster	SC
15	Castellvell	SC
17, 18	Ulldemolins	SM
19	La Serra d'Almos	SM
20	La Bisbal de Falset	SM
21	Margalef	SM
22	Cabacés	SM
23	La Vilella Alta	SM
24	Marçà	SM
25	Capçanes	SM
26	Falset	SM
27	El Masroig	SM
28	El Molar	SM
29	La Palma d'Ebre	SM

3. RESULTS AND DISCUSSION

3.1. Spectra

In the range studied (λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm) (Fig. 1), the main fluorescent compounds present in virgin olive oils are vitamin E, which emits around λ_{em} = 525 nm and oxidation products (conjugated hydroperoxides formed as a result of fatty acids oxidation by oxygen), which give rise to two low peaks around λ_{em} = 445 and λ_{em} = 475 nm (Kyriakidis and Skarkalis, 2000). The inspection

of the EEMs of the samples reveals that the two types of "Siurana" oils have different fluorescence. SC oils tend to have less intensity at emissions above 500 nm (vitamin E) and more intensity at emissions below 500 nm (oxidation products) than SM oils. A large content on oxidation products indicates a major degradation of the oils, whereas vitamin E contributes to give them stability, because of its antioxidant effect. Thus, the trend seen on the EEMs agrees with the fact that SM oils are more stable than SC oils.
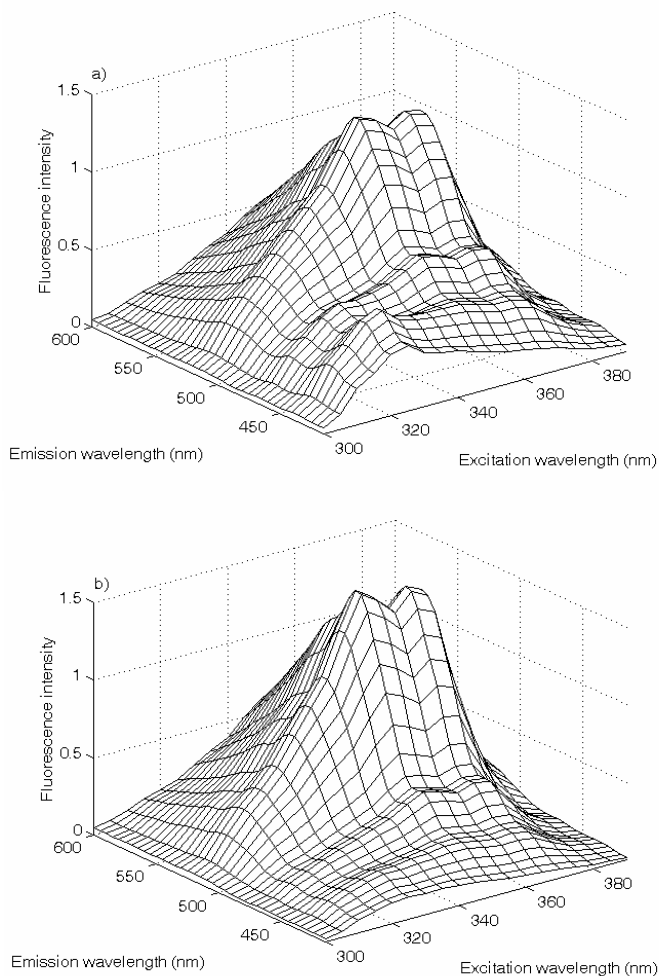

Fig. 1. Average EEMs between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm: (a) "Siurana-Camp"; (b) "Siurana-Montsant" oils.

In order to check if the differences between the mean spectra plotted in Figure 1 were statistically significant, we applied the Hotelling's T^2 statistic (Krzanowski, 2000). For applying this test, the mean values of the two replicates of each sample were used. The EEMs were unfolded in such a way that all emission spectra of each EEM were concatenated in a row. Thus, for each sample a row-vector of size (1×722) was obtained, where 722 is the product number of $\lambda_{ex} \times$ number of λ_{em}. As the number of variables was superior to the number of samples, principal component analysis (PCA) was computed to reduce dimensionality. The Hotelling's T^2 statistic was applied on the scores of the PCA model using five PCs. The calculated F was $F_{cal} = 18.21$ and it was compared to the critical F-value for a two-sided test at $\alpha = 0.05$ ($F_{tab} = 3.15$). As F_{cal} was higher than the critical F-value, we can conclude that the mean spectra of the two "Siurana" regions are statistically different. However, the differences between the EEMs of the two groups of oils were not obvious for all the samples. For this reason, chemometric methods were applied for discrimination.

3.2. Discriminant unfold partial least squares regression (DU-PLSR)

Partial least squares regression (PLSR) is a calibration method widely used in multivariate analysis. The regression is based on factors determined by employing both the independent variables ($\mathbf{X}$) and the variable to be predicted ($\mathbf{y}$). Discriminant partial least squares regression (D-PLSR) is a particular case of PLSR in which the variable to be predicted is the membership or no-membership of the samples to one class.

Before applying D-PLSR, an $\mathbf{X}$ matrix containing the unfolded EEMs of the samples was built (i.e. each row consisted of all the fluorescence spectra measured of one sample concatenated). Then, the method applied to this matrix is named DU-PLSR. The size of $\mathbf{X}$ was 58×722 (samples (in duplicate) $\times$ (number of $\lambda_{em} \times$ number of λ_{ex})). The matrix was column mean-centered. Then we built a regression model $\mathbf{y} = f(\mathbf{X})$ where $\mathbf{y}$ was a class variable made up of a column with zeros for SC oils (class 1) and ones for SM oils (class 2). The optimal number of factors, selected by leave-two-out cross-validation was two. The explained variance by the two factors was 93% of $\mathbf{X}$ and 76% of $\mathbf{y}$. Figures 2-3 show the score and loading plots of DU-PLSR. The loadings were folded back. The score plot (Fig. 2) shows that factor 1, which has 33% expl. var. of $\mathbf{X}$ and 71% expl. var. of $\mathbf{y}$, is the most discriminating between the two types of oils, because SM oils tend to

have scores higher than 1 on this factor and SC oils tend to have scores lower than 1. However, some of the samples are not well separated along factor 1 (12, 14, 19, 27, 28, and 29). As far as factor 2 is concerned, there is no discrimination between "Siurana" regions. These results can be interpreted by looking at the loadings (Fig. 3). Factor 1 loadings have two differentiated regions: a positive zone at emissions above 500 nm (due to the influence of vitamin E) and a negative zone at emissions below 500 nm (due to oxidation products). The larger value of the loadings in the negative zone indicates that oxidation products have more influence than vitamin E on this factor. A closer inspection of the EEMs revealed that the main differences between the two "Siurana" types of oils lie in the fluorescence below $\lambda_{em} = 500$ nm (the oxidation products emission region), whereas the differences on vitamin E (λ_{em} above 500 nm) are less obvious. This explains the greater weight of oxidation products on the most discriminant factor of DU-PLSR. Thus, the negative scores on factor 1 of most of SC oils indicate that they are more deteriorated than SM oils. On the contrary, high scores on factor 1 indicate more stability of the oils (high content on vitamin E and low on oxidation products).
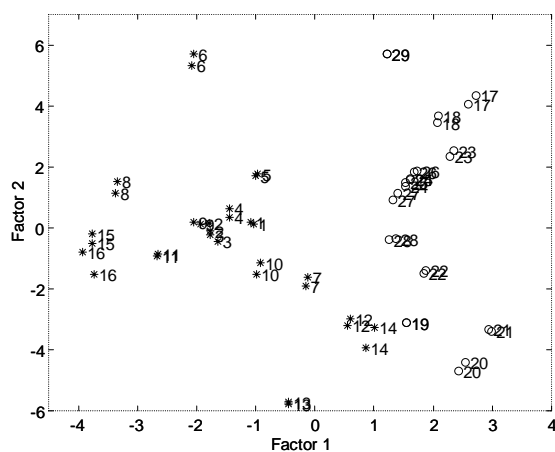

Fig. 2. Score plot from DU-PLSR to the EEMs of the "Siurana" olive oils. (✳) SC oils, (o) SM oils.
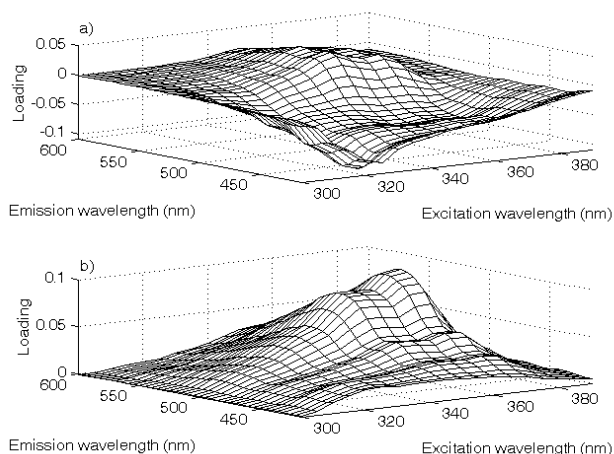
Fig. 3. Refolded loadings from DU-PLSR to the EEMs of the "Siurana" olive oils . (a) factor 1; (b) factor 2.

Samples 12 and 14 are located on the boundary between the two classes, having scores close to one on factor 2 (Fig. 2). The special behaviour of samples 12 and 14 is due to the fact that their EEMs are very similar to those of samples 19, 20 and 21, whose region below λ_{em} = 500 nm is almost flat. This indicates a very low content on oxidation products. This can be due to several factors, e.g., a greater stability or a better storage. In addition, the whole intensity of these EEMs was not very high. In spite of the membership of sample 12 to the SC region, it is very close to the SM region. This may explain its position on the score plot (Fig. 2). For sample 14, the reason is less obvious, but the shape of its fluorescence EEMs indicated that this sample has been very little deteriorated.

Factor 2 was seen to be correlated to the whole fluorescence intensity of the oils. Thus, the most intense EEMs (samples 6 and 29) had the largest scores of this factor and the less intense EEM (sample 13) had the lowest scores. From the score plot (Fig. 2), it can also be observed that repeatability of the measurements was quite good, because the duplicates of the samples appear very close each other. In order to check the precision of the method, we applied analysis of variance (ANOVA) to the score matrix. Var_{within} indicates the percentage of variance between the pairs of replicates, and $var_{between}$ indicates the percentage of variance between samples.

Both values are referred to the total variability. These values were $var_{within} = 0.2\%$ and $var_{between} = 99.8\%$ in relative terms. From these results, we can conclude that the precision of the method was acceptable because the variability between replicates is much smaller than the variability between samples.

Figure 4 shows the predicted vs. reference values obtained by cross-validation from DU-PLSR for the two-factor model. The oils were assigned to SC class when the predicted **y** was lower than 0.5. Oils with a predicted value higher than 0.5 were assigned to SM class. Under this criterion, the percentage of correct classification was 94% for SC oils and 100% for SM oils. Only sample 14 was misclassified. The root mean square error of cross-validation (RMSECV) was 0.27.
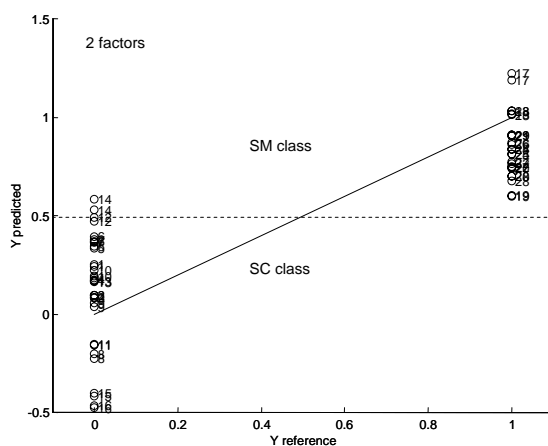

Fig. 4. Predicted vs. reference values obtained by cross-validation from DU-PLSR (2 factors). The straight line is the ideal regression line. The broken line (– – –) identifies the boundary between classes.

Besides of concatenating the data matrices to form a larger matrix, they can also be arranged in a three-dimensional structure. The D-PLSR method applied to the data arranged in this way is named discriminant multi-way partial least squares regression (DN-PLSR). We checked if the results obtained from DU-PLSR could be improved using DN-PLSR. Thus, the oil EEMs were stacked on a three-way array of size $58 \times 38 \times 19$ (samples (in duplicate) $\times$ number of $\lambda_{em} \times$ number of λ_{ex}). Then the three-way array was centered along the first mode (samples) and DN-PLSR was computed on it. The class variable was defined as in DU-PLSR. The optimal

171

number of factors, selected by leave-two-out cross-validation was two (90% expl. var. of $\mathbf{X}$ and 76% expl. var. of $\mathbf{y}$). The results did not improve compared to DU-PLS. The percentages of correct classification were 91% for SC oils and 100% for SM oils. The two replicates of sample 14 were misclassified and one replicate of sample 12 was in the boundary between the classes ($\mathbf{y}$ = 0.5).

4. CONCLUSIONS

This work has shown the potential of EEFS and three-way methods of analysis for discriminating between the two types of "Siurana" olive oils. The different orography, soil characteristics and climatic conditions of the SC and SM regions lead to olive oils with different chemical composition, which can be detected by fluorescence spectroscopy. The more stability of SM oils makes them to have less content on oxidation products. This implies that SM emit much less fluorescence at emissions below 500 nm compared to SC, which are more deteriorated. In addition, SM oils tend to have more content on vitamin E (emissions above 500 nm).

Using DU-PLSR enabled to discriminate between the two types of "Siurana" oils with a percentage of correct classification of 94% for SC and 100% for SM oils, with only one misclassified sample. This method also showed the good repeatability of the measurements.

ACKNOWLEDGEMENTS

REFERENCES

Aparicio, R. 1988. Characterization of foods by inexact rules: the SEXIA expert system. *J. Chemometr.*, **3**, 175-192.

Armanino, C., Leardi, R., Lanteri, S., Modi, G. 1989. Chemometric analysis of Tuscan olive oils. *Chemom. Intell. Lab. Syst.*, **5**, 343-354.

Boggia, R., Zunin, P., Lanteri, S., Rossi, N., Evangelisti, F. 2002. Classification and class modelling of "Riviera Ligure" extra-virgin olive oil using chemical-physical parameters. *J. Agric. Food Chem.*, **50**, 2444-2449.

Guimet, F., Ferré, J., Boqué, R., Rius, F.X . 2004, a. Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy. *Anal. Chim. Acta*, **515**, 75-85.

Guimet, F., Boqué, R., Ferré, J., 2004, b. Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy. *J. Agric. Food Chem.*, **52**, 6673-6679.

Krzanowski, W.J. 2000. Principles of Multivariate Analysis. A user's perspective, Oxford University Press, Oxford.

Kyriakidis, N.B., Skarkalis, P. 2000. Fluorescence spectra measurement of olive oil and other vegetable oils. *J. AOAC Int.*, **83**, 1435-1439.

Lakowicz, J.R. 1999. Principles of Fluorescence Spectroscopy, 2nd ed., Kluwer Academic/Plenum Publishers, New York.

Lanteri, S., Armanino, C., Perri, E., Palopoli, A. 2002. Study of oils from Calabrian olive cultivars by chemometric methods. *Food Chem.*, **76**, 501-507.

Marcos, I., Pérez, J.L., Fernández, M.E., García, C., Moreno, B., Henriques, L.R., Peres, M.F., Simões, M.P., Lopes, P.S. 2002. Application of headspace-mass spectrometry for differentiating sources of olive oil. *Anal. Bioanal. Chem..*, **374**, 1205-1211.

Marini, F., Balestrieri, F., Bucci, R., Magrì, A.D., Magrì, A.L., Marini, D. 2004. Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. *Chemom. Intell. Lab. Syst.*, **73**, 85-93.

Matlab. The Mathworks, South Natick, MA, USA 2002. http://www.mathworks.com, last access: 30/03/05.

PLS-Toolbox. Toolbox for Matlab by Eigenvector Research, Inc. 2003. http://www.mathworks.com, last access: 30/03/05.

Sacchi, R., Mannina, L., Fiordiponti, P., Barone, P., Paolillo, L., Patumi, M., Segre, A. 1998. Characterization of Italian extra virgin olive oils using ^{1}H-NMR spectroscopy. *J. Agric. Food Chem.*, **46**, 3947-3951.

Sayago, A. Morales, M.T., Aparicio, R. 2004. Detection of hazelnut oil in virgin olive oil by a spectrofluorimetric method. *Eur. Food Res. Technol.*, **218**, 480-483.

Scott, S.M., James, D., Ali, Z., O'Hare, W.T., Rowell, F.J. 2003. Total luminescence spectroscopy with pattern recognition for classification of edible oils. *Analyst*, **128**, 966-973.

Tous, J., Romero, A., Plana, J., Guerrero, L., Díaz, I., Hermoso, J.F. 1997. Características químico-sensoriales de los aceites de oliva "Arbequina" obtenidos en distintas zonas de España. *Grasas y Aceites*, **48**, 415-424.

5.4 PAPER

F. Guimet, R. Boqué, J. Ferré

Application of non-negative matrix factorization combined with Fisher's linear discriminant analysis for classification of olive oil excitation-emission fluorescence spectra

Application of Non-negative Matrix Factorization Combined with Fisher's Linear Discriminant Analysis for Classification of Olive Oil Excitation-emission Fluorescence Spectra

Francesca Guimet, Ricard Boqué, Joan Ferré

Department of Analytical and Organic Chemistry, Rovira i Virgili University
C/ Marcel·lí Domingo, s/n E-43007 Tarragona, Catalonia, Spain

ABSTRACT

Non-negative matrix factorization (NMF) is a technique that decomposes multivariate data into a smaller number of basis functions and encodings using non-negative constraints. These constraints make that only positive solutions can be obtained and thus this method provides a more realistic approximation to the original data than other factorization methods that allow positive and negative values. Here we show that NMF is a powerful technique for learning a meaningful parts-based representation of the fluorescence excitation-emission matrices (EEMs) of different sets of olive oils. The capabilities of NMF together with Fisher's LDA for discriminating between various types of oils were also studied. In all cases, good classifications were obtained (90-100%).

Keywords: Non-negative matrix factorization; Fisher's linear discriminant analysis; classification; olive oil; EEMs

1. INTRODUCTION

Olive oil authentication is an issue of great interest in the manufacturing countries of this food product. Authenticity covers many aspects, including adulteration, mislabelling, characterization, and misleading origin [1]. Edible olive oils include virgin (V), pure (P), and olive-pomace (OP) oil. V olive oil is the highest-quality type of olive oil. It is obtained from the fruit of the olive tree by mechanical processes that do not lead to alterations in the oil. It is divided into different categories regarding its acidity and its sensory characteristics. Thus, extra virgin (EV) olive oil has a maximum acidity allowed of 1.0g/100g, expressed as oleic acid, whereas fine virgin (FV) olive oil has a maximum acidity allowed of 2.0g/100g. P olive oil is a blend of edible V olive oil and refined olive oil. By extracting the olive-pomace (i.e. the olive residue remaining from previous pressings) with authorized solvents, refined olive-pomace oil is obtained. This oil is improved with edible V olive oil to obtain OP oil [2].

Owing to its higher price, V olive oil is susceptible of undergoing fraudulent practices concerning its authenticity. For this reason, analytical techniques must constantly be developed and improved to fight against fraud. Nowadays, olive oil authenticity is manly carried out by means of chromatographic techniques [1]. These techniques have some handicaps, such as being time-consuming and requiring sample manipulation. For this reason, other techniques have also been applied. They include headspace-mass spectrometry (HS-MS) [3], and spectroscopic techniques, such as near infrared (NIR), mid-infrared (MIR), Fourier transform infrared (FT-IR), Fourier transform-Raman (FT-Raman) [4-11], nuclear magnetic resonance (NMR) [9,12-15], chemiluminescence [16], and spectral nephelometry [17]. Fluorescence spectroscopy has also been applied for assessing olive oil authenticity [18-24]. This technique has some nice properties, such as its speed of analysis, that it is reagentless, and that small amounts of sample are required. When a set of fluorescence spectra at different λ_{ex} are recorded, a three-dimensional landscape is obtained, the so-called fluorescence excitation-emission matrix (EEM). Recording EEMs enables to obtain more information about the fluorescent species present in the oils, because the bands arising in a wider area are considered [22,23].

Several chemometric methods, including unsupervised pattern recognition [6,12,17,22,23], classification [3,6,14,15,19,20,24,25] and regression [4-6,8,10,11,24] methods have been widely applied for olive oil authentication.

Lee and Seung [26] introduced non-negative matrix factorization (NMF) in its modern formulation as a method to decompose images into parts-based representations. The method compresses the data using non-negativity constraints. These constraints make that only additive combinations of the parts are allowed to represent the original data. As a result, and unlike principal component analysis (PCA), NMF leads to decompositions with only positive values. Lee and Seung [26] applied this method to a set of face images and showed that the resulting basis functions represented localized features that corresponded with intuitive notions of the parts of faces (eyes, mouths, noses, …). By contrast, they noted that the application of PCA to image data yielded components with no obvious visual interpretation. Hence, in some cases, NMF may be more suitable than PCA, because it provides a more realistic interpretation of the data.

Besides image analysis, NMF has been applied to other many areas, including text data [26-28], genetics [29], colour spectra [30], and sound signals [31]. In some cases, NMF has been successfully applied for discrimination purposes [27,28].

In this paper we apply NMF to the EEMs of different sets of olive oils with a twofold objective. On one side, we study the ability of NMF to obtain a meaningful parts-based representation of the spectra, and look into the decomposed data to see if they can be related to the fluorescence species present in oils. On the other side, we check the capabilities of NMF used together with a discrimination method such as Fisher's linear discriminant analysis (LDA) for classification of different sets of oils.

2. NON-NEGATIVE MATRIX FACTORIZATION (NMF)

The NMF algorithm [26,32] is a method that compresses a set of objects in a smaller number of basis functions and their encodings. The basis functions are analogous to PCA loadings, because they contain information about the variables, whereas the encodings are analogous to PCA scores, because they are related to the objects. Given a set of multivariate j-dimensional data vectors, the vectors are placed in the columns of a $J \times I$ matrix $\mathbf{V}$ where I indicates the number of objects in the data

set. This matrix is then approximately factorized into a $J \times F$ matrix $\mathbf{W}$ and a $F \times I$ matrix $\mathbf{H}$. Then the NMF algorithm constructs approximate factorizations of the form:

$$\mathbf{V}_{ji} \approx (\mathbf{WH})_{ji} = \sum_{f=1}^{F} \mathbf{W}_{jf} \mathbf{H}_{fi} \, , \qquad (1)$$

where the f columns of $\mathbf{W}$ are called the basis set, each column of $\mathbf{H}$ is called and encoding, and F is the rank of the factorization. Although there is not a clear criterion for selecting F, it is generally chosen by $(J + I)F < JI$. NMF does not allow negative entries in the matrix factors $\mathbf{W}$ and $\mathbf{H}$.

The method starts by randomly initializing matrices $\mathbf{W}$ and $\mathbf{H}$ with positive values, which are iteratively updated to minimize the objective function

$$D = \sum_{j=1}^{J} \sum_{i=1}^{I} \left[\mathbf{V}_{ji} \log(\mathbf{WH})_{ji} - (\mathbf{WH})_{ji} \right] \qquad (2)$$

subject to the non-negativity constraints described above. Different update rules can be applied for minimizing the objective function [32]. Here we apply the divergence-based update equations [29]:

$$\mathbf{H}_{fi} \leftarrow \mathbf{H}_{fi} \frac{\sum_{j} \mathbf{W}_{jf} \mathbf{V}_{ji} / (\mathbf{WH})_{ji}}{\sum_{k} \mathbf{W}_{kf}} \qquad \mathbf{W}_{jf} \leftarrow \mathbf{W}_{jf} \frac{\sum_{i} \mathbf{H}_{fi} \mathbf{V}_{ji} / (\mathbf{WH})_{ji}}{\sum_{v} \mathbf{H}_{fv}} \qquad (3)$$

To represent a new vector using a predefined set of basis functions, the same algorithm is iterated without modifying the matrix $\mathbf{W}$. Thus, fixing $\mathbf{W}$ and starting with positive random $\mathbf{h}$, a representation of a new data vector according to the basis defined in $\mathbf{W}$ is obtained.

Some features of NMF that distinguish it from PCA are summarized below:

- The basis functions are non-orthogonal and they do not correspond to directions of maximal variance, but to physical or conceptual features in a non-negative space.

- NMF is not nested (i.e. it provides different basis functions depending on how many factors are computed).

- Implementation requires non-linear iterative optimization.

- Different objectives functions can be used.

- Data must not be centered before applying the algorithm, because they must remain non-negative.

3. EXPERIMENTAL

3.1. Sample sets

3.1.1. Commercial Spanish olive oils

In the present work, three sets of olive oils were considered. The first set was composed of 56 commercial olive oils from different Spanish regions and from different olive varieties. It included 29 V (28 EV and 1 FV), 20 P and 7 OP oils [23]. Figure 1 shows the average EEMs of each type of oil. They have noticeable differences. V oils have their maximum fluorescence above λ_{em} = 500 nm, whereas at lower emissions the fluorescence peaks are much lower. This is because of the high content of vitamin E and the low content of oxidation products in these oils [18]. On the contrary, P and OP oils have a larger content of oxidation products, which give rise to a broad peak around λ_{em} = 450 nm.

3.1.2. Extra virgin olive oils from the Protected Denomination of Origin "Siurana"

The second set was composed of 28 EV olive oils from the Protected Denomination of Origin (PDO) "Siurana". This production area is divided into two regions ("Siurana-Camp" (SC) and "Siurana-Montsant" (SM)). Oils from these regions differ on composition, which makes them have different stability and sensory properties [33]. The set of "Siurana" oils used in this study was composed of 15 SC and 13 SM oils. Figure 2 shows the average EEMs of the SC and the SM oils used. The main difference between them is that SC EEM displays a peak around λ_{ex} = 300-350 nm, λ_{em} = 415-470 nm, which does not appear in SM EEM. This indicates a more advanced degradation stage for SC oils.
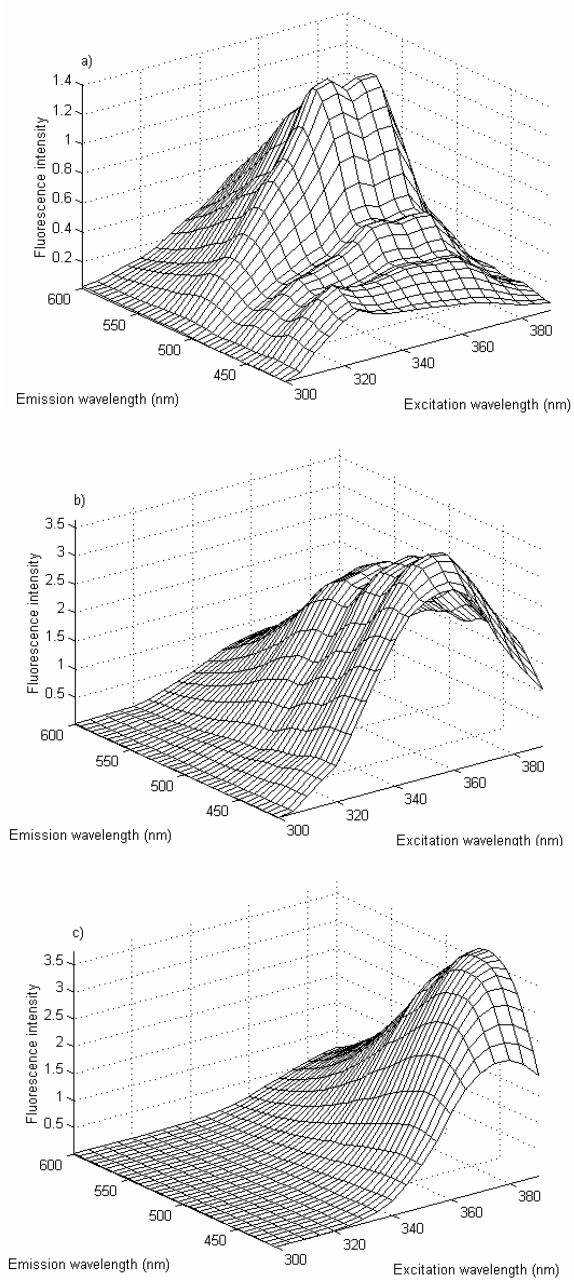
Figure 1. Average EEMs of data set 1 between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm: (a) V, (b) P, and (c) OP oils.

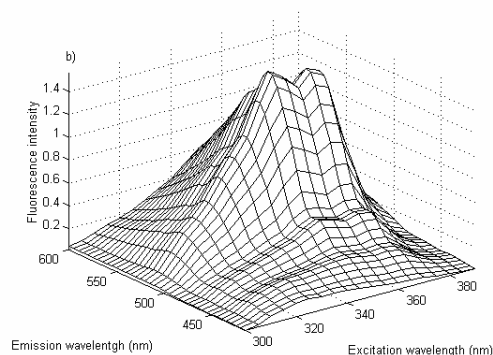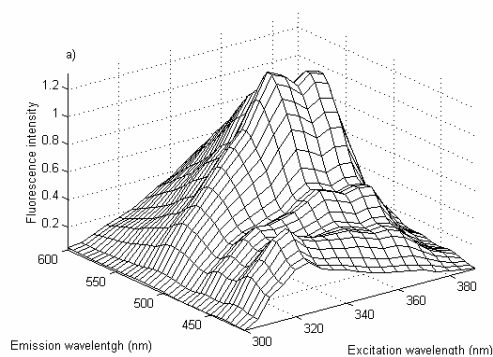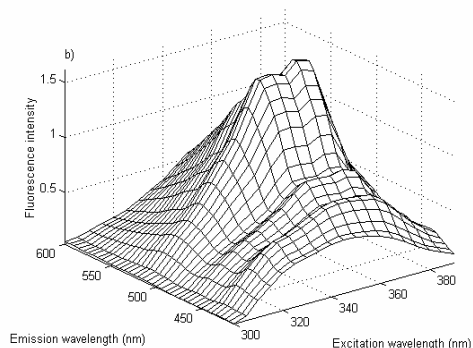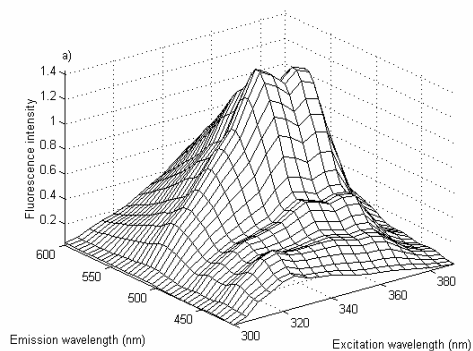**Figure 2**. Average EEMs of data set 2 between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm: (a) SC, (b) SM oils.

Figure 3. Average EEMs of data set 3 between λ_{ex} = 300-390 nm; λ_{em} = 415-600 nm: (a) PDO "Siurana" EV olive oils, (b) 5% (w/w) ADs with OP oils.

3.1.3. Extra virgin olive oils adulterated with olive-pomace oils

The third set of oils contained the same 28 PDO "Siurana" EV olive oils and admixtures (ADs) of them with 5% (w/w) of 5 different OP oils. The total number of ADs was 93. The procedure for preparing the ADs is explained elsewhere [24]. Figure 3 shows the average EEMs of the 28 EV olive oils and the 93 ADs. Note that adulteration mainly causes an increase of fluorescence at emissions below 500 nm. This is due to the presence of oxidation products in the OP oils added.

3.2. Instrumentation and software

EEMs were measured with and Aminco Bowman series 2 luminescence spectrometer equipped with a 150W xenon lamp and 10 mm quartz cells. The instrument detector was operated using the EmL/Ref channel and a voltage

around 600 V was applied for all samples. Excitation and emission ranges were λ_{ex} = 300-390 nm and λ_{em} = 415-600 nm, respectively. Measuring emission above excitation prevented Rayleigh scatter. The step size and band-pass of both monochromators were 5 and 4 nm, respectively. The scan rate was 30 nm s^{-1}. The instrument software was used to correct the EEMs for deviations in the ideality of the lamp, monochromators, and detector [34,35].

Data were exported to ASCII code and processed with Matlab software (version 6.5) [36]. The Matlab code for the NMF algorithm was obtained from the Broad Institute's website [37]. The Fisher's LDA algorithm was built-in-house.

4. RESULTS AND DISCUSSION

4.1. Selection of the training and tests sets
For all the sets of oils, the procedure for selecting the training and the test sets for the chemometric methods was the following. The EEMs were stacked in a three-way array of size (samples $\times$ number of $\lambda_{em}\times$ number of λ_{ex}). This array was then unfolded to a matrix of size (samples $\times$ (number of $\lambda_{em}\times$ number of λ_{ex})) by combining the spectral modes. PCA was computed on the centered matrix after selecting the optimal number of factors by means of the leave-one-out cross-validation method. The samples in the training set were chosen on the basis of the score plots. The criterion was to cover the entire variability domain. The rest of samples were included in the test set.

For the set of commercial Spanish olive oils a six-factor PCA model (99.78% expl. var.) was calculated on the unfolded matrix containing the spectra of the 56 samples. From the score plots, it was decided to split the sample set into a training set containing 28 oils (14 V, 10P and 4 OP) and a test set with 28 oils (15 V, 10 P and 3 OP).

In the case of oils from the two "Siurana" regions, a five-factor PCA model (99.49% expl. var.) was calculated. The set of 28 samples was split into a training set of 15 oils (8 SC and 7 SM) and a test set of 13 oils (7 SC and 6 SM).

For the set of PDO "Siurana" oils and their ADs, only non-adulterated samples were considered for the PCA model. The training set to be used for Fisher's LDA

contained 15 EV oils and the 28 ADs prepared from them. The test set included the remaining 13 EV oils and the 65 ADs prepared from them [24].

4.2. Application of non-negative matrix factorization and Fisher's linear discriminant analysis

4.2.1. Discrimination between commercial Spanish olive oils

NMF was applied to the unfolded matrix containing the spectra of the 28 oils of the training set. Different number of factors (2-6) were considered. After obtaining the basis functions and the encodings for the training set, the test set was projected onto the models and the encodings of the new samples were calculated. Afterwards, the NMF encodings of the training set were used to compute the Fisher's linear discriminant functions (LDFs) for discriminating between the three classes of oils. The NMF encodings of the test set were also used as test set for Fisher's LDA.

The best classification was obtained from the NMF model using 5 factors (99.18% expl. var.). Figure 4 shows the basis functions obtained from this model. They were folded back to the original dimensions of the EEMs.

Figure 5 shows the encodings of factor 3 vs. factor 5 for all the samples, which are those displaying the best separation between the three classes of oils. The basis functions obtained provided representations of typical fluorescent features of the oil types. Thus, basis functions 1 and 3 (Fig. 4a and 4c) are similar to the EEMs of V oils (Fig. 1a), specially basis function 3, having strong fluorescence around λ_{em} = 525 nm (due to vitamin E) and emitting little fluorescence below λ_{em} = 500 nm [23]. Basis function 4 (Fig. 4d) provides a good representation of the fluorescence landscapes of P oils (Fig. 1b), and basis function 5 (Fig. 4e) displays the typical fluorescence of OP oils (Fig. 1c). The broad peak at λ_{em} below 500 nm found in P and OP oils has been related to the presence of large amounts of oxidation products [23]. Basis function 2 does not match with the whole EEM of any type of oil, but the peak arising around λ_{ex} = 310-360 nm; λ_{em} = 415-480 nm may also be related to the presence of oxidation products in highly degraded oils [23].
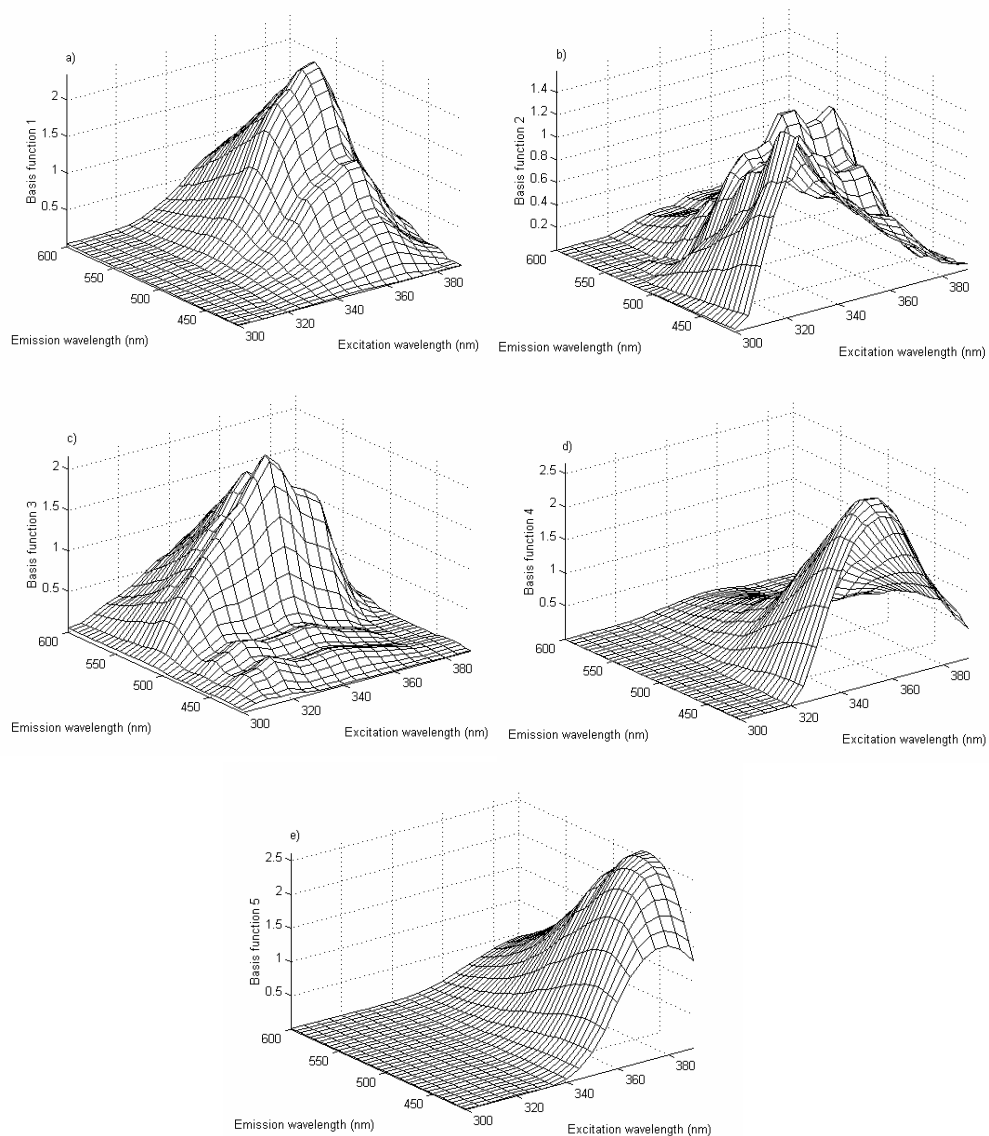
Figure 4. NMF basis functions obtained for the commercial Spanish olive oils (data set 1).

A quite good separation of the oil types was obtained from the encodings of factors 3 and 5 (Fig. 5). As it was expected from the shape of the basis functions, OP oils had the highest encodings of factor 5, and V oils had the highest of factor 3.
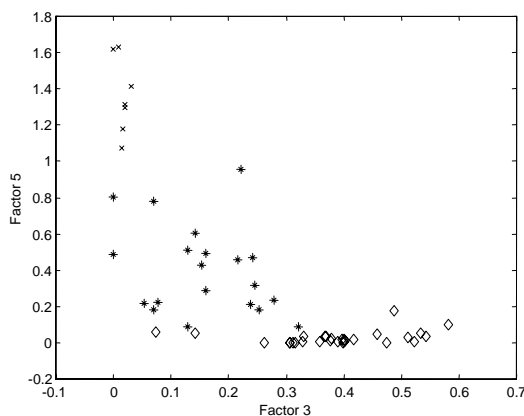
Figure 5. NMF encodings of factor 3 vs. factor 5 for data set 1 (training and test set). V (◊), P (✳), and OP (✕) oils.

Figure 6 shows the score plot of Fisher's LDA. The boundaries between classes were set taking the centroid of each class and drawing a line half-way between the pair of centroids [38]. This procedure was followed for all the Fisher's LDA models presented in this paper. Table 1 shows the percentage of correct classification for the training set (recognition ability) and the test set (prediction ability). All V and OP oils were classified in the correct class. Regarding to P oils, a 90% of correct classification was obtained for both the training and the test set.

Figure 6. Score plot from Fisher's LDA applied to data set 1 (training and test sets): V (◊), P (✳), and OP (✕) oils. Boundaries between classes (– – –).

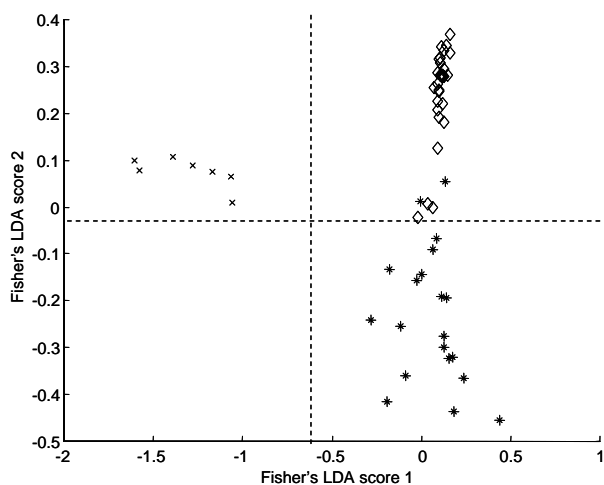**Table 1**. Results of Fisher's LDA applied to the commercial Spanish olive oils

Class	Correct classification (%)	
	Recognition ability	Prediction ability
1 (V)	100	100
2 (P)	90	90
3 (OP)	100	100

Fisher's LDA was computed again using PCA scores instead of NMF scores. The six-factor PCA model (99.74% expl. var.) gave the best results. Classification percentages were identical to those obtained from NMF.

4.2.2. Discrimination between EV oils from the two "Siurana" regions

NMF was applied to the unfolded matrix containing the spectra of the 15 oils of the training set (8 SC and 7 SM oils). Then the test set was projected.

In this case, the best classification was obtained with the NMF model using 6 factors (99.97% expl. var.). Figure 7 shows the refolded basis functions of this model, and Figure 8 shows the encodings of factors 1, 3 and 4, which are the most discriminating. Basis function 1 displays a peak around λ_{ex} = 300-360 nm, λ_{em} = 415-480 nm. As it was said above, high fluorescence in this region indicates degraded oils. In general, SC oils have the highest encodings on factor 1 (Fig. 8). This means that they are more degraded than SM oils. Actually, this feature can also be observed from their EEMs (Fig. 2). Basis functions 3 and 4 (Fig. 7c and 7d) are related to the fluorescence of vitamin E. This is inferred from their shape, having a maximum intensity around λ_{em} = 525 nm. The encoding plots (Fig. 8) indicate that SM oils have more content on vitamin E than SC, because their encodings on factors 3 and 4 tend to be higher. The shape of basis function 2 (Fig. 7b) indicates that it has also some influence of vitamin E. In addition, contributions of oxidation products of oils can also be found on this basis function (fluorescence around λ_{ex} = 300-380 nm; λ_{em} = 415-440 nm). Basis functions 5 and 6 (Fig. 7e and 7f) are not assignable to any particular species, because their profiles cover the whole range.
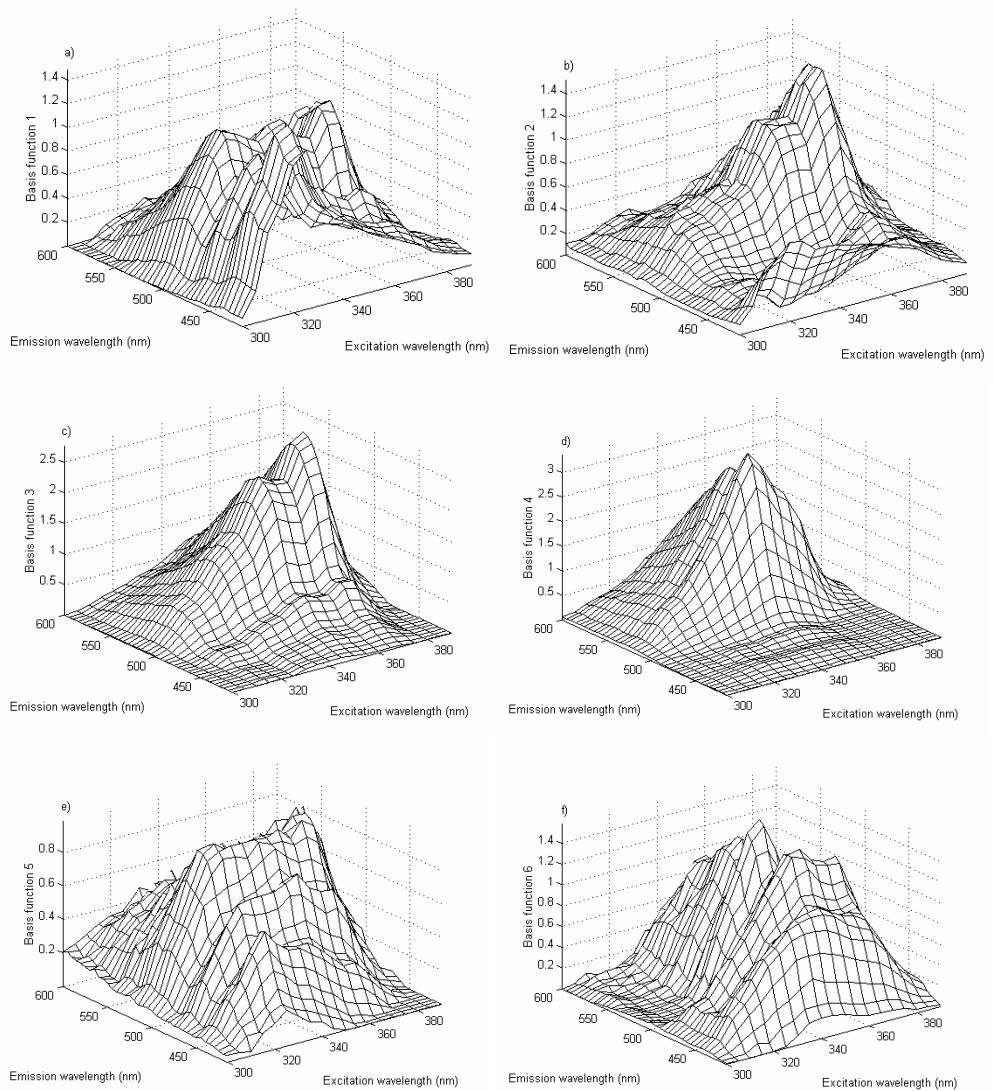
Figure 7. NMF basis functions obtained for the PDO "Siurana" EV olive oils (data set 2).

Table 2. Results of Fisher's LDA applied to the oils of the two "Siurana" regions

Class	Correct classification (%)	
	Recognition ability	Prediction ability
1 (SC)	100	100
2 (SM)	100	100
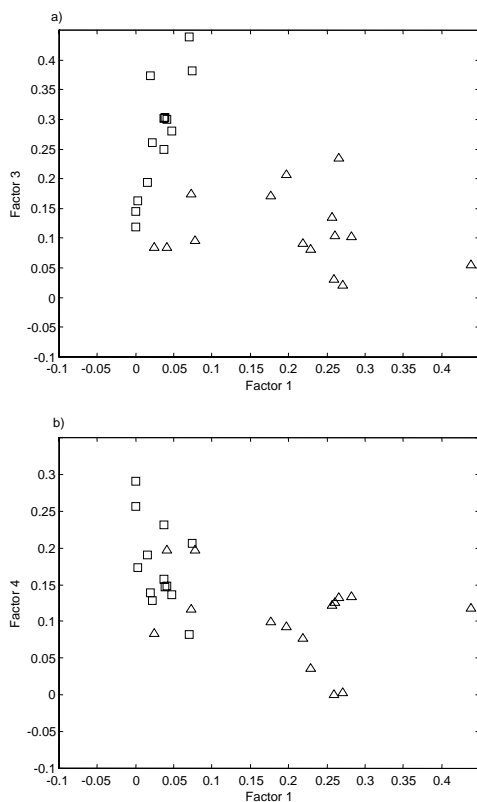
Figure 8. NMF encodings for data set 2. (a) factor 1 vs. factor 3, (b) factor 1 vs. factor 4. SC (△), SM (□) oils.

Figure 9 shows the score plot of Fisher's LDA for the PDO "Siurana" oils and Table 2 the percentage of correct classification. As it can be seen, a 100% of correct classification was achieved for the training and the test sets.

These results were compared to those obtained from PCA combined with Fisher's LDA. The six-factor PCA model was the most suitable (99.77% expl. var.). The prediction ability for the SC group was 86%. For the rest of sets, classifications of 100% were obtained. Thus, NMF was superior to PCA for discriminating between "Siurana" oils.
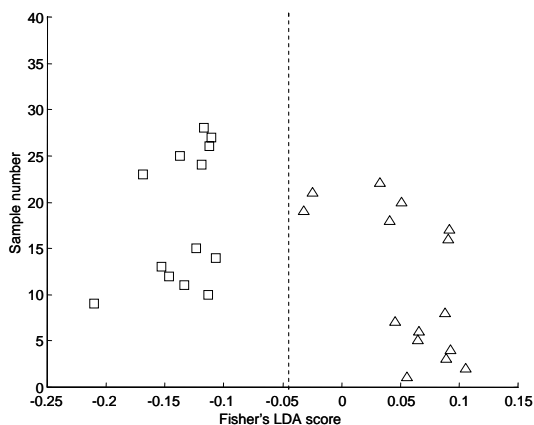
Figure 9. Score plot from Fisher's LDA applied to data set 2. SC (△), SM (□) oils. Boundary between classes (– – –).

4.2.3. Discrimination between EV oils and adulterations with OP oils

NMF was applied to the same 15 EV olive oils as in the "Siurana" regions case. Then, the rest of samples were projected onto the model (the 13 EV oils of the test set and the ADs of both the training and the test set). The NMF encodings of the 15 EV oils of the training set constituted class 1 for Fisher's LDA, and those of the 28 ADs prepared from these samples formed class 2. The remaining 13 EV oils and their 65 ADs were used as test set for Fisher's LDA.

As in the case of discrimination between the "Siurana" regions, the NMF model using 6 factors (99.97% expl. var.) was the most suitable. Thus, the same basis functions (Fig. 7) were considered. NMF factors 2, 5 and 6 captured the features that best distinguished between the two classes of oils. This can be seen from the encodings of these factors (Fig. 10). Thus, the general trend is that the ADs have the highest encodings on factor 2 and 6, whereas the non-adulterated samples have the highest encodings on factor 5. However, there are a few ADs that do not follow this trend.

Figure 11 shows the score plot of Fisher's LDA for the set of non-adulterated and adulterated EV oils, and Table 3 the percentage of correct classification. For non-adulterated samples, a 93% of correct classification was obtained for the training set, and a 92% for the test set, with only one misclassified sample on both sets. In

the case of the ADs, all the samples of the training set were correctly classified, whereas the prediction ability was 98%, with one sample misclassified.
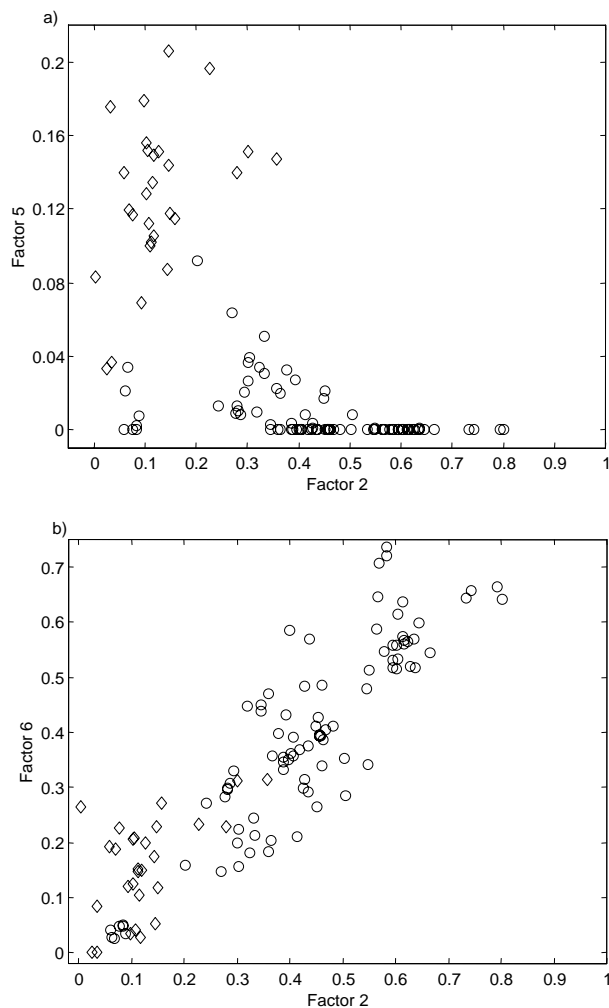

Figure 10. NMF encodings for EV oils and ADs (data set 3). (a) factor 2 vs. factor 5, (b) factor 2 vs. factor 6. EV oils (◊), ADs (o).
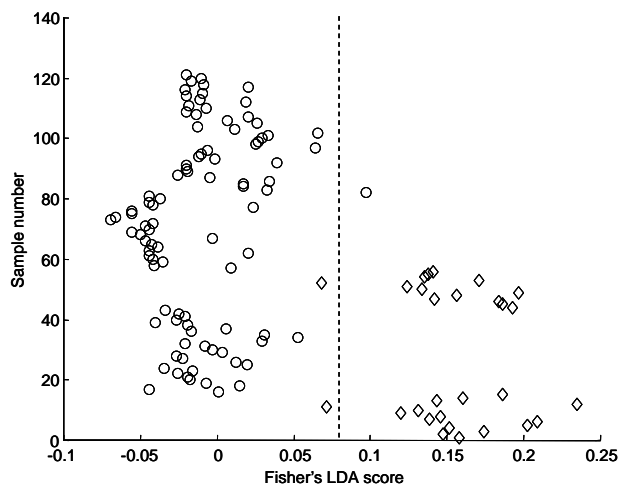
Figure 11. Score plot from Fisher's LDA applied to data set 3. EV oils (◊), ADs (o). Boundary between classes (– – –).

Table 3. Results of Fisher's LDA applied to the EV olive oils and the ADs

Class	Correct classification (%)	
	Recognition ability	Prediction ability
1 (EV)	93	92
2 (AD)	100	98

Fisher's LDA was also applied to the PCA scores. The five-factor PCA model was the most discriminating (99.64% expl. var.). In this case, the recognition ability was 100% for both classes, but the prediction ability was 85% and 98% for EV oils and ADs, respectively.

5. CONCLUSIONS

We have shown that NMF is a powerful technique for learning a meaningful parts-based representation of spectroscopic data. The algorithm was applied to the fluorescence EEMs of three sets of olive oils. The basis functions represented specific features of the whole fluorescence landscape of oils, which could be related to some of their fluorophors. The non-negativity constraints of the algorithm

enabled to obtain realistic solutions, because only positive basis functions and encodings are obtained.

The capabilities of NMF used together with Fisher's LDA for discriminating between the different types of oils were also studied. In all cases, classifications above 90% were achieved. In some cases, NMF yielded better classifications than PCA.

The results look very promising with regard to applicability of NMF to second-order fluorescence data for discrimination purposes. However, further research on the combination of this method with other discrimination methods, such as quadratic discriminant analysis (QDA), discriminant partial least squares (PLS) regression or density methods is needed in order to improve classifications.

ACKNOWLEDGEMENTS

REFERENCES

[1] R. Aparicio, R. Aparicio-Ruíz, J. Chromatogr. A, 881 (2000) 93-104.

[2] A.K. Kiritsakis, Olive Oil: From the Tree to the Table. 2nd ed., Food and Nutrition Press Inc., Trumbull, 1998, p.237.

[3] I. Marcos, J.L. Pérez, M.E. Fernández, C. García, B. Moreno, J. Chromatogr. A, 945 (2002) 221-230.

[4] I.J. Wesley, R.J. Barnes, A.E.J. McGill, JAOCS, 72 (1995) 289-292.

[5] V. Baeten, M. Meurens, M.T. Morales, R. Aparicio, J. Agric. Food Chem., 44 (1996) 2225-2230.

[6] G. Downey, P. McIntyre, A.N. Davies, J. Agric. Food Chem., 50 (2002) 5520-5525.

[7] H. Yang, J. Irudayaraj, JAOCS, 78 (2001) 889-895.

[8] Y.W. Lai, E.K. Kemsley, R.H. Wilson, Food Chem., 53 (1995) 95-98.

[9] M.J. Dennis, Analyst, 123 (1998) 151R-156R.

[10] L. Küpper, H.M. Heise, P. Lampen, A.N. Davies, P. McIntyre, Appl. Spectrosc., 55 (2001) 563-570.

[11] E.C. López-Díez, G. Bianchi, R. Goodacre, J. Agric. Food Chem., 51 (2003) 6145-6150.

[12] R. Sacchi, L. Mannina, P. Fiordiponti, P. Barone, L. Paolillo, M. Patumi, A. Segre, J. Agric. Food Chem., 46 (1998) 3947-3951.

[13] L. Mannina, A.P. Sobolev, A. Segre, Spectrosc. Eur., 15 (2003) 6-14.

[14] G. Vigli, A. Philippidis, A. Spyros, P. Dais, J. Agric. Food Chem., 51 (2003) 5715-5722.

[15] D.L. García-González, L. Mannina, M. D'Imperio, A.L. Segre, R. Aparicio, Eur. Food Res. Technol., 219 (2004) 545-548.

[16] K. Papadopoulos, T. Triantis, C.H. Tzikis, A. Nikokavoura, D. Dimotikali, Anal. Chim. Acta, 464 (2002) 135-140.

[17] A.G. Mignani, P.R. Smith, L. Ciaccheri, A. Cimato, G. Sani, Sensor Actuat. B-Chem., 90 (2003) 157-162.

[18] N.B. Kyriakidis, P. Skarkalis, J. AOAC Int., 83 (2000) 1435-1439.

[19] S.M. Scott, D. James, Z. Ali, W.T. O'Hare, F.J. Rowell, Analyst, 128 (2003) 966- 973.

[20] A. Sayago, M.T. Morales, R. Aparicio, Eur. Food Res. Technol., 218 (2004) 480-483.

[21] E. Sikorska, A. Romaniuk, I.V. Khmelinskii, R. Herance, J.L. Bourdelande, M. Sikorski, J. Koziol, J. Fluorescence, 14 (2004) 25-35.

[22] F. Guimet, J. Ferré, R. Boqué, F.X. Rius, Anal. Chim. Acta, 515 (2004) 75-85.

[23] F. Guimet, R. Boqué, J. Ferré, J. Agric. Food Chem., 52 (2004) 6673-6679.

[24] F. Guimet, J. Ferré, R. Boqué, Anal. Chim. Acta, 544 (2005) 143-152.

[25] F. Marini, F. Balestrieri, R. Bucci, A.D. Magrì, A.L. Magrì, D. Marini, Chemom. Intell. Lab. Syst., 73 (2004) 85-93.

[26] D.D. Lee, H.S. Seung, Nature, 401 (1999) 788-791.

[27] C.W. Lee, H. Kang, K. Jung, H.J. Kim, Lect. Notes Comput. Sc., 2756 (2003) 470-477.

[28] D. Guillamet, J. Vitrià, Proceedings of the 11th International Conference on Image Analysis and Processing, Palermo, 2001, pp. 256-261.

[29] J.P. Brunet, P. Tamayo, T.R. Golub, J.P. Mesirov, PNAS, 101 (2004) 4164-4169.

[30] G. Buchsbaum, O. Bloch, Vision Res., 42 (2002) 559-563.

[31] H. Asari, unpublished Technical Note (2004). http://zadorlab.cshl.edu/asari/nmf.html, last access April 19, 2005.

[32] D.D. Lee, H.S. Seung, Adv. Neural Info. Proc. Syst., 13 (2001) 556-562.

[33] J. Tous, A. Romero, J. Plana, L. Guerrero, I. Díaz, J.F. Hermoso, Grasas y Aceites, 48 (1997) 415-424.

[34] SLM AMINCO, Technical Note No 101, Urbana.

[35] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, 2nd ed., Kluwer Academic/Plenum Publishers, New York, 1999, p. 25.

[36] Matlab, The Mathworks, South Natick, MA, USA, 2002. http://www.mathworks.com, last access: April 18, 2005.

[37] Broad Institute, Massachussets Institute of Technology and Harvard University. http://www.broad.mit.edu/, last access: April 18, 2005.

[38] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Vol. 20, Part B, Elsevier, Amsterdam, 1998, p. 207.

Chapter 6

**Conclusions and
Suggestions for Future Research**

6.1 CONCLUSIONS

The first part of this chapter summarizes the conclusions of the work presented in this thesis.

1. EEFS can be applied directly to olive oils for obtaining information about their characteristics. Using EEMs (i.e. second-order data) is more advantageous than recording a single fluorescence spectrum at one excitation wavelength (multivariate data). The reason is that more information about the fluorescent species is considered. As a result, when EEMs are recorded differentiation between oil types improves compared to when a single fluorescence spectrum is considered.

2. Unsupervised pattern recognition methods such as unfold-PCA, PARAFAC and HCA are very useful for the exploratory analysis of olive oils from bi-dimensional fluorescence data. They enabled to distinguish between different types of commercial olive oils (virgin, pure and olive-pomace) on the basis of their EEMs. In addition, unfold-PCA and PARAFAC allowed relating the fluorescence landscapes of oils to some of the fluorescence species present in oils (vitamin E, oxidation products and chlorophylls). It has to be taken into account that the samples used at this stage were not reference samples, since they were purchased in shopping centres. Thus, the conclusions stated here are referred to this type of samples. It is also important to note that the number of samples used in this study was reduced, especially OPOs. Hence, if the number of available samples had been higher, the results would have probably been different. In addition, all samples came from Spanish regions and other origins were not considered. Thus, the results of the exploratory analysis have to be seen as a trend considering the mentioned limitations.

3. Application of exploratory methods to olive oil differentiation was performed considering two spectral ranges. The first one (λ_{ex} = 300-400 nm; λ_{em} = 400-695 nm) contained the chlorophyll peak, whereas the second one (λ_{ex} = 300-400 nm; λ_{em} = 400-600 nm) did not include it. Chlorophylls had strong influence on all the models because of their high fluorescence intensity in virgin and pure olive oils. To overcome this problem data was scaled when the fluorescence region of the chlorophylls was included in the models. Nevertheless, even scaling data,

differentiation between oil types was worse when the fluorescent region of the chlorophylls was considered compared to when it was no used.

4. Considering the spectral range without chlorophylls (λ_{ex} = 300-400 nm; λ_{em} = 400-600 nm) differentiation between oil types is manly due to the oxidation products in oils (conjugated hydroperoxides, and carbonyl compounds). These compounds exhibit a wide peak below λ_{em} = 500 nm in the fluorescence EEMs. On the contrary, vitamin E, which emits mainly above λ_{em} = 500 nm has low discrimination power.

5. PARAFAC improves discrimination between oil types compared to unfold-PCA. In addition, PARAFAC has the advantage that extracts the underlying spectra of the main families of fluorophores. This method can also be used as a fingerprint of the oil types because the PARAFAC profiles are good approximations of the original EEMs.

6. Inspection of unfold-PCA and PARAFAC score plots is an easy and rapid way of detecting outliers. In addition, as the loadings obtained contain information about the spectra, information about the cause of the different characteristics of such samples (e.g. a major degradation) can be extracted. Anyway, a deeper study about the properties of these samples should be done by reference method of analysis.

7. The suitability of HCA to discriminate between a set of commercial Spanish olive oils (virgin, pure and olive-pomace) was studied. Concretely, HAC using the Euclidean distance as a similarity measure and the average linkage was applied. In this study, various preprocessing methods were compared. The best results were obtained when the unfolded EEMs in the range between (λ_{ex} = 300-400 nm; λ_{em} = 400-600 nm) were row normalized to length one and column autoscaled. In these conditions, all the samples were clustered in the correct group. Again, the results obtained here must to be considered bearing in mind that the samples were purchased in shopping centres and that the number of samples was reduced.

8. The relationship between the fluorescence EEMs of oils and some quality parameters (PV and K_{270}) was studied. Two methods were compared to study the correlation between the fluorescence spectra and these parameters: MLR computed on the PARAFAC scores and N-PLS regression. The latter provided better fits and

lower prediction errors. With N-PLS, the correlation between EEMs and PV was r = 0.94 and that between EEMs and K_{270} was r = 0.99, which are acceptable values. It has also been shown that EEFS is able to detect samples highly degraded at early stages (with high PV), because they exhibit strong fluorescence between λ_{ex} = 315-370 nm; λ_{em} = 415-460 nm. Thus, fluorescence in this region can be related to high amount of primary oxidation products (conjugated hydroperoxides). However, low correlation was observed between fluorescence and other quality parameters of olive oils (e.g. K_{232}). Thus, EEFS is not a definitive technique for oil analysis, but can be used as a complementary technique for obtaining additional information to that obtained from the official methods of analysis.

9. It has been shown that OPO adulteration in EVOOs can be detected by fluorescence measurements. Such adulteration increases the fluorescence intensity at emissions below 500 nm. This is due to the high content of oxidation products in OPO oils. The capability of this technique to detect other adulterants has not been tested yet.

10. Three-way classification methods applied to second-order fluorescence data are able to discriminate between EVOOs from the PDO "Siurana" these oils adulterated with OPOs at low levels (5% w/w). The Hotelling T^2 and Q statistics can be applied to unfold-PCA as a fast screening method for detecting such adulteration. However, in order to improve discrimination between non-adulterated and adulterated samples, some supervised pattern recognition method must be applied. Fisher's LDA computed on the unfold-PCA scores provides quite good classifications, but DN-PLSR is superior. In the study presented in this thesis, DN-PLSR yielded to a 100% of correct classifications for both the training and the test sets. The adulteration level can also be quantified by using N-PLSR. The ability of EEFS and the chemometric methods presented here to detect OPO adulteration below the 5% level has not been tested yet.

11. PDO "Siurana" EVOOs from "Camp de Tarragona" and "Montsant" regions have different composition, which gives them different stability. These differences are captured by their fluorescence EEMs. It has been shown that oils from "Camp de Tarragona" in general emit more fluorescence at emission below 500 nm, due to a larger content of oxidation products, which gives them less stability compared to oils from "Montsant" region. Two three-way classification methods have been

applied to discriminate between these two types of oils (DU-PLSR and DN-PLSR). Both methods yielded good classifications. Nevertheless those obtained from DU-PLSR were slightly better (94% of correct classification for "Siurana-Camp" oils and 100% for "Siurana-Montsant" oils). It has to be taken into account that this method has only been used to discriminate between oils from two very close geographical areas. No studies about discrimination between oils from other origins have been made so far.

12. NMF has been applied to fluorescence spectroscopy for the first time. It has been shown that this method can decompose the EEMs of olive oils into parts that can be related to the fluorescence species contained in oils. The main advantage of NMF respect to PCA is that it provides a more realistic interpretation of the data because only positive solutions are allowed. In addition, a new classification method based on NMF combined with Fisher's LDA has also been proposed. The capabilities of the method have been checked by applying it to classify oils in three different situations. In all cases, good classifications were obtained (90-100%).

6.2 SUGGESTIONS FOR FUTURE RESEARCH

This thesis has been focused on studying the possibility of applying three-way methods of analysis to the fluorescence EEMs of olive oils for discrimination. It has been shown that three-way classification methods can be applied for discriminating between olive oils in several situations. However there are some aspects that still need to be studied in more detail. They are summarized below.

1. This thesis has shown that using EEFS and three-way methods enables to detect OPO adulteration in EVOOs. However, more research should be done so as to build a robust methodology based on EEFS and three-way methods of analysis for detecting other typical adulterants (e.g. sunflower, soy, maize, rapessed and nut oils) in EVOOs.

2. This thesis has shown that three-way classification methods are able to discriminate between oils from two geographical regions within the PDO "Siurana" production area on the basis of their EEMs. This study could be extended to the authentication of the origin of oils from other regions. This

includes distinguishing between oils from various PDOs and detecting misclassified samples.

3. Another interesting issue to be explored is the suitability of EEFS and three-way methods for discriminating between olive oil varieties and between oils produced in different harvesting years.

4. Some research on olive oil degradation using EEFS and multi-way methods could be done. For instance, studying oil degradation over time or the effect of frying.

5. In this thesis, we have proposed NMF as a method to obtain meaningful parts-based representations of bi-dimensional fluorescence data. This method has also been applied for classification purposes. In further research, NMF could be compared with PARAFAC, which also enables to obtain positive solutions when non-negativity constraints are applied.

6. Further research could still be done to study the viability of applying other classification methods to the fluorescence EEMs of olive oils. Some of these methods may be QDA and neural networks, for instance. Another suggestion is to develop new three-way classification methods and to apply them to oil EEMs.

7. Studies on the viability of the developed three-way methods to other second-order data (i.e. GC-MS and HPLC-DAD) for the analysis of olive oils could be carried out.

Appendix

LIST OF ABBREVIATIONS

AD	Admixture
ANOVA	Analysis of variance
BP	Back propagation
CA	Cluster analysis
COPO	Crude olive-pomace oil
CV	Cross-validation
CVA	Canonical variate analysis
DNA	Deoxyribonucleic acid
DCV	Double cross-validation
DN-PLSR	Discriminant multi-way partial least squares regression
D-PLSR	Discriminant partial least squares regression
DU-PLSR	Discriminant unfold partial least squares regression
EEC	European Economic Community
EEFS	Excitation-emission fluorescence spectroscopy
EEM	Excitation-emission matrix
EU	European Union
EVOO	Extra virgin olive oil
FID	Flame ionization detector
FT-IR	Fourier transform infrared
FT-Raman	Fourier transform Raman
GC	Gas chromatography
HAC	Hierarchical agglomerative clustering
HCA	Hierarchical cluster analysis
HDL	High-density lipoprotein
HPLC	High performance liquid chromatography
HS-MS	Headspace-mass spectrometry
IOOC	International Olive Oil Council
IRMS	Isotope ratio mass spectrometry
LDA	Linear discriminant analysis
LDF	Linear discriminant function
LDL	Low-density lipoprotein
LVOO	Lampante virgin olive oil
MIR	Mid-infrared
MLR	Multiple linear regression

MPCA	Multi-way principal component analysis
MS	Mass spectrometry
NIPALS	Non-linear iterative partial least squares
NIR	Near-infrared
NMF	Non-negative matrix factorization
NMR	Nuclear magnetic resonance
N-PLS	Multi-way partial least squares
OPO	Olive-pomace oil
OVOO	Ordinary virgin olive oil
PARAFAC	Parallel factor analysis
PC	Principal component
PCA	Principal component analysis
PDO	Protected denomination (designation) of origin
PGI	Protected geographical indication
PLS	Partial least squares
PLSR	Partial least squares regression
POO	Pure olive oil
PRESS	Predicted residual error sum of squares
PV	Peroxide value
QDA	Quadratic discriminant analysis
RBF	Radial basis function
ROO	Refined olive oil
RMSECV	Root mean square error of cross-validation
RMSEP	Root mean square error of prediction
ROPO	Refined olive-pomace oil
SC	Siurana-Camp de Tarragona
RSE	Residual square error
SFAM	Simplified fuzzy adaptive resonance theory mapping
SIMCA	Soft independent modelling of class analogy
SM	Siurana-Montsant
SSP	Sum of squares and products
TLS	Total luminescence spectroscopy
U-PCA (unfold-PCA)	Unfold principal component analysis
UV	Ultraviolet
VIS	Visible
VOO	Virgin olive oil

LIST OF PAPERS AND MEETING CONTRIBUTIONS

List of papers presented by the author in this thesis (in chronological order):

1. F. Guimet, J. Ferré, R. Boqué, F.X. Rius.
   Application of unfold principal component analysis and parallel factor analysis to the exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy.
   Analytica Chimica Acta 515 (2004) 75-85.
   (Chapter 3)

2. F. Guimet, R. Boqué, J. Ferré.
   Cluster analysis applied to the exploratory analysis of commercial Spanish olive oils by means of excitation-emission fluorescence spectroscopy.
   Journal of Agricultural and Food Chemistry 52 (2004) 6673-6679.
   (Chapter 3)

3. F. Guimet, J. Ferré, R. Boqué.
   Rapid detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis.
   Analytica Chimica Acta 544 (2005) 143-152.
   (Chapter 5)

4. F. Guimet, R. Boqué, J. Ferré.
   Study of oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis.
   Grasas y Aceites 56 (4) (2005) 292-297.
   (Chapter 5)

5. F. Guimet, J. Ferré, R. Boqué, M. Vidal, J. Garcia.
   Excitation-emission fluorescence spectroscopy combined with three-way methods of analysis as a complementary technique for olive oil characterization.
   Journal of Agricultural and Food Chemistry (accepted for publication).

(Chapter 4)

6. F. Guimet, R. Boqué, J. Ferré

    Application of non-negative matrix factorization combined with Fisher's linear discriminant analysis for classification of olive oil excitation-emission fluorescence spectra.

    Chemometrics and Intelligent Laboratory Systems (accepted for publication).

    (Chapter 5)

List of contributions to international and national meetings (in chronological order):

1. F. Guimet, J. Ferré, R. Boqué, F.X. Rius.

    Exploratory analysis of olive oils by means of excitation-emission matrix fluorescence spectroscopy and three-way pattern recognition methods.

    V Colloquim Chemometricum Mediterraneum, Ustica (Italy), 2003.

    Poster communication.

2. F. Guimet, J. Ferré, R. Boqué.

    Detection of olive-pomace oil adulteration in extra virgin olive oils from the protected denomination of origin "Siurana" using excitation-emission fluorescence spectroscopy and three-way methods of analysis.

    9th Chemometrics in Analytical Chemistry Conference (CAC'04), Lisbon (Portugal), 2004.

    Poster communication.

3. F. Guimet, J. Ferré, R. Boqué.

    Caracterització d'olis d'oliva mitjançant espectroscòpia de fluorescència d'excitació-emissió i mètodes d'anàlisi de tres vies.

    Workshop de la Xarxa Catalana de Quimiometria, Barcelona (Spain), 2005.

    Poster communication.

4. F. Guimet, R. Boqué, J. Ferré.

    Olive oil characterization by means of excitation-emission fluorescence spectroscopy and three-way PLS methods.

    4th International Symposium on PLS and Related Methods (PLS'05), Barcelona (Spain), 2005.

    Oral communication.

SUMMARY

L'oli d'oliva és un producte de gran importància econòmica i nutricional. Això fa que la seva anàlisi sigui fonamental per tal d'assegurar la seva qualitat i evitar possibles fraus. Les tècniques més emprades en l'anàlisi d'olis d'oliva són les cromatogràfiques, però també se n'han utilitzat d'altres, com l'espectroscòpia ultravioleta-visible, l'espai de cap acoblat a l'espectrometria de masses, l'espectroscòpia d'infraroig proper, infraroig mitjà i Raman amb transformada de Fourier, la ressonància magnètica nuclear i l'olfactometria electrònica.

Degut a què els olis d'oliva contenen espècies fluorescents, com la vitamina E, pigments (clorofil·les i feofitines), productes d'oxidació i compostos fenòlics, l'espectroscòpia de fluorescència també s'ha aplicat en alguns casos a l'anàlisi d'olis, principalment per a detectar adulteracions. Aquesta tècnica té l'avantatge que permet analitzar les mostres en poc temps. A més, les mesures es poden fer directament sobre els olis sense realitzar cap etapa prèvia de dilució o addició de reactius, amb la qual cosa es redueix el temps d'anàlisi i no es generen residus addicionals.

La instrumentació permet obtenir un conjunt d'espectres de fluorescència a diferents longituds d'ona d'excitació en una única mesura. D'aquesta manera, per a cada mostra analitzada s'obté un perfil de fluorescència bidimensional que recull la intensitat de fluorescència a cada longitud d'ona d'excitació i d'emissió. Matemàticament aquest perfil de fluorescència és una matriu de dades i per això s'anomena matriu d'excitació-emissió de fluorescència. Aquest tipus de dades es coneix amb el nom de dades de segon ordre. Tot i que hi ha aplicacions de l'espectroscòpia de fluorescència d'excitació-emissió (EFEE) a l'anàlisi d'olis d'oliva, aquest camp està poc estudiat i l'ús d'aquesta tècnica no està difós en els laboratoris de control de qualitat dels olis.

Els mètodes quimiomètrics que s'apliquen a les dades de segon ordre reben el nom de mètodes de tres vies perquè les dades estan disposades en una estructura tridimensional. En el cas de l'EFEE, quan es mesuren diverses mostres, les matrius de fluorescència es poden disposar en un cub de dimensions (mostres×nombre de longituds d'ona d'emissió×nombre de longituds d'ona d'excitació). Els mètodes

de tres vies han estat àmpliament emprats en anàlisi exploratòria i en anàlisi quantitativa. Tanmateix, els mètodes de classificació de tres vies estan poc desenvolupats.

En aquesta tesi s'estudien les possibilitats de l'EFEE i els mètodes de tres vies per a la caracterització d'olis d'oliva. Els resultats obtinguts estan recollits en articles publicats, els quals són l'eix central dels capítols de la tesi.

El capítol 1 conté la introducció i els objectius de la tesi. En aquest capítol s'expliquen les principals característiques dels olis d'oliva i es fa un recull de les tècniques analítiques emprades habitualment en l'anàlisi dels olis. Finalment, es plantegen els objectius que s'han pretès assolir en aquesta tesi.

El capítol 2 conté les bases teòriques necessàries per entendre la tesi. En aquest capítol s'explica el fonament de la fluorescència, s'introdueixen els conceptes relacionats amb les estructures de tres vies i es presenten els mètodes quimiomètrics utilitzats en els treballs de la tesi. Aquests mètodes es poden dividir en mètodes d'anàlisi exploratòria i de classificació.

En el capítol 3 s'estudien els avantatges de treballar amb dades de fluorescència de segon ordre respecte a utilitzar dades multivariants (un sol espectre de fluorescència). Posteriorment s'apliquen i es comparen diversos mètodes d'anàlisi exploratòria a les matrius de fluorescència d'olis comercials (verges, purs i de sansa d'oliva). Aquests mètodes són l'anàlisi de components principals sobre la matriu desplegada (*unfold-PCA*), anàlisi paral·lela de factors (*PARAFAC*) i anàlisi d'agrupacions jerarquitzada (*HCA*). Per tal d'aplicar aquests mètodes, s'estudien diversos rangs espectrals i mètodes de processament de dades. Es comenten els problemes que comporta incloure les clorofil·les en els models (el pic que apareix entre λ_{em} = 650-695 nm), degut a la gran intensitat d'aquest pic en comparació amb la resta.

En aquest capítol es mostra que l'EFEE és capaç de detectar diferències de composició entre diversos tipus d'olis. L'ús dels mètodes *unfold-PCA* i *PARAFAC* permet obtenir informació sobre les espècies fluorescents responsables de la diferenciació dels olis, ja que els factors obtinguts a partir dels models es poden relacionar amb les espècies fluorescents contingudes en els olis. L'aplicació de

PARAFAC té l'avantatge addicional que permet extreure els espectres subjacents de les principals famílies de compostos fluorescents.

El capítol 4 inclou un estudi sobre la viabilitat de l'EFEE per a l'avaluació de la qualitat d'olis d'oliva. En aquest estudi es van correlacionar les matrius de fluorescència d'un conjunt d'olis de la mateixa collita amb alguns paràmetres de qualitat mesurats mitjançant mètodes oficials d'anàlisi. La correlació es va estudiar emprant mètodes de tres vies, concretament *PARAFAC* i regressió per mínims quadrats parcials multi-via (*N-PLS*). El mètode *N-PLS* va proporcionar ajustos millors i errors de predicció més petits. Es va veure que els espectres de fluorescència dels olis es poden correlacionar amb l'índex de peròxids i amb l'absorbància a 270 nm. Aquests paràmetres estan directament relacionats amb substàncies que es formen en els olis degut a la seva degradació. També es va mostrar que l'EFEE permet detectar olis altament degradats de manera ràpida, ja que aquests presenten una elevada intensitat de fluorescència per sota dels 500 nm.

El capítol 5 es basa en l'aplicació de mètodes de classificació de tres vies per a discriminar conjunts d'olis a partir de les seves matrius de fluorescència. Els resultats es presenten en tres articles. En el primer d'ells, s'estudia la possibilitat de detectar adulteracions d'oli de sansa en olis d'oliva verge extra a baixes concentracions (5% pes/pes). Per dur a terme aquest estudi, es va utilitzar un conjunt d'olis d'oliva verge extra de la denominació d'origen protegida (DOP) "Siurana" i es van preparar mescles d'aquests olis amb olis sansa d'oliva adquirits en diversos centres comercials. Posteriorment, es va comparar l'habilitat de diversos mètodes quimiomètrics per a discriminar entre les mostres no adulterades i les adulterades. En un primer estadi, es van aplicar els estadístics de Hotelling T^2 i Q sobre les puntuacions (*scores*) d'*unfold-PCA* com a mètode ràpid de cribatge. Després es va aplicar anàlisi discriminant lineal de Fisher (*Fisher's LDA*) per a millorar la discriminació. Aquest mètode va permetre obtenir una habilitat de predicció d'un 85% per a mostres no adulterades i d'un 98% per a mostres adulterades. Aquests resultats es van superar emprant regressió per mínims quadrats parcials multi-via discriminant (*DN-PLSR*), que va permetre obtenir un 100% de classificació correcta en ambdues classes. *N-PLS* també va permetre quantificar el nivell d'adulteració amb relativament poc error per al conjunt de validació (1.2%).

En el segon article del capítol 5 es fa un estudi dels olis d'oliva verge extra de la DOP "Siurana" per fluorescència. La zona de producció DOP "Siurana" es pot dividir en dues regions, anomenades Camp de Tarragona i Montsant, amb diferent orografia, característiques de sòl i condicions climàtiques. Això implica que els olis procedents d'aquestes regions tinguin diferències en la seva composició, la qual cosa influeix en la seva estabilitat enfront a l'oxidació. En aquest treball es mostra que els olis procedents de les dues regions "Siurana" es poden distingir mitjançant l'EFEE i mètodes tres vies. Concretament es van comparar els mètodes regressió per mínims quadrats parcials discriminant sobre la matriu desplegada (*DU-PLSR*) i *DN-PLSR* per a discriminar entre els dos tipus d'olis. Ambdós mètodes van proporcionar bones classificacions, però els resultats van ser lleugerament millors amb *DU-PLSR* (94% per als olis de la zona del Camp de Tarragona i 100% per als de la zona Montsant).

El tercer article del capítol 5 conté un treball sobre l'aplicació de l'algorisme factorització de matrius no negativa (*NMF*) sobre les matrius de fluorescència de tres conjunts d'olis diferents. En aquest estudi es mostra que aquest mètode és capaç de descomposar les MEE dels olis en parts positives que es poden relacionar amb algunes de les espècies fluorescents dels olis d'oliva. Es comenten els principals avantatges d'emprar aquest mètode respecte a altres mètodes de descomposició, com *PCA*, que admeten solucions negatives. En aquest article també es proposa *NMF* combinat amb *Fisher's LDA* com a mètode de classificació. Per tal d'estudiar la viabilitat d'aquest mètode, es van plantejar tres situacions diferents: discriminació entre tipus d'olis (verges, purs i de sansa), detecció d'adulteracions en olis d'oliva verge extra i discriminació entre olis de les dues zones "Siurana". En tots els casos es van obtenir molt bones classificacions (90-100%).