Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytica Chimica Acta 515 (2004) 75–85
ANALYTICA CHIMICA ACTA
www.elsevier.com/locate/aca

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

Received 29 August 2003; received in revised form 17 December 2003; accepted 12 January 2004

Available online 21 February 2004

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 multiway 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 $\lambda_{ex} = 300$–400 nm, $\lambda_{em} = 400$–695 nm and $\lambda_{ex} = 300$–400 nm, $\lambda_{em} = 400$–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 europaea* 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 olive 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 (FTIR) [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 400–550 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

* Corresponding author. Tel.: +34-977-558155; fax: +34-977-559563.
E-mail address: pgv@quimica.urv.es (F. Guimet).

0003-2670/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2004.01.008

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 to 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 factor 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 luminiscence 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 and λ_{em} = 400–700 nm. The step size and bandpass 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 pre-processing

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 nm, λ_{em} = 400–600 nm mainly shows peaks due to oxidation products and Vitamin E [8] (Fig. 2).

Studies were carried out first in 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-centring was used in PCA and U-PCA. No pre-processing 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. 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 from 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 pre-processing methods were tested. Column mean-centring led to PCA being dominated by the chlorophylls peak, which is much more intense than the peaks of

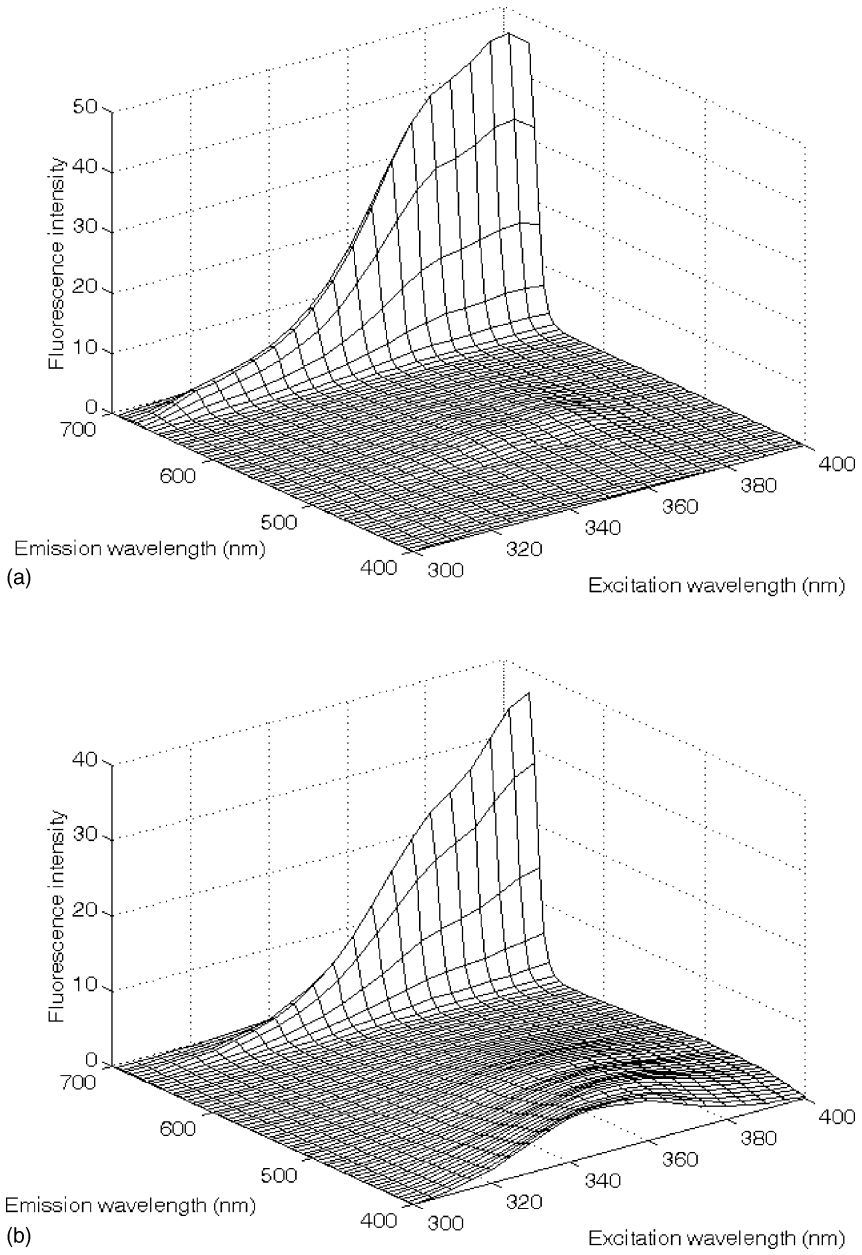

Fig. 1. EEMs of a virgin (a) and a pure (b) olive oil between $\lambda_{ex} = 300$ and 400 nm, $\lambda_{em} = 400$ and 695 nm.

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).

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

λ_{ex} (nm)	λ_{em} (nm)	PCA ($\lambda_{ex} = 365$ 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.

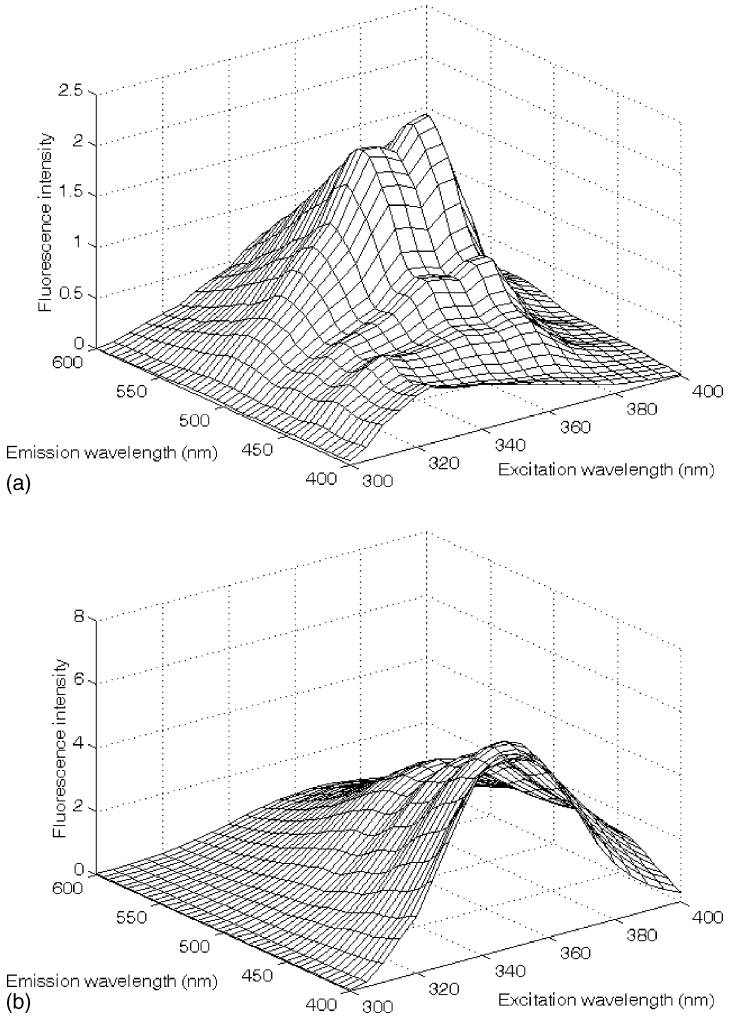

Fig. 2. EEMs of a virgin (a) and a pure (b) olive oil between $\lambda_{ex} = 300$ and 400 nm, $\lambda_{em} = 400$ and 600 nm.

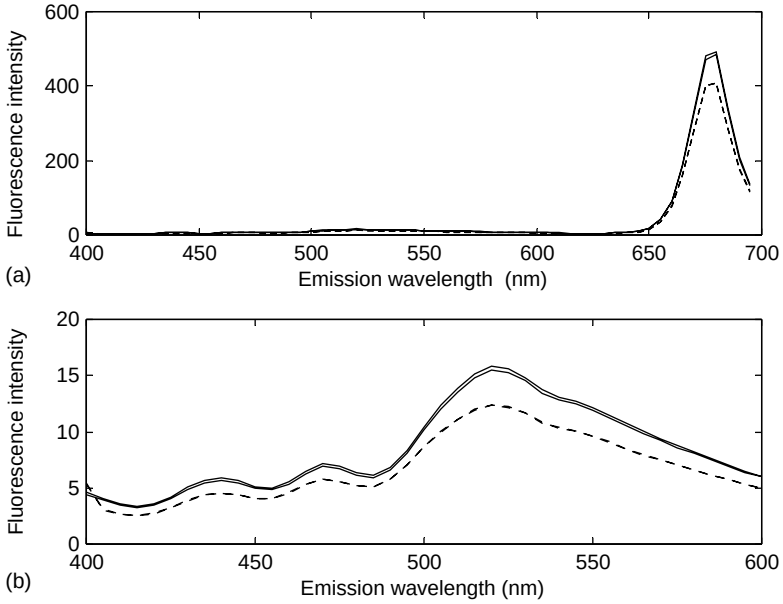

Fig. 3. Total component spectra of a virgin olive oil. (a) Fluorescence emission spectra between 400 and 695 nm; (b) fluorescence emission spectra between 400 and 600 nm. Continuous line: before oxidation; dotted line: after 3 h heating. Two replicates were measured.

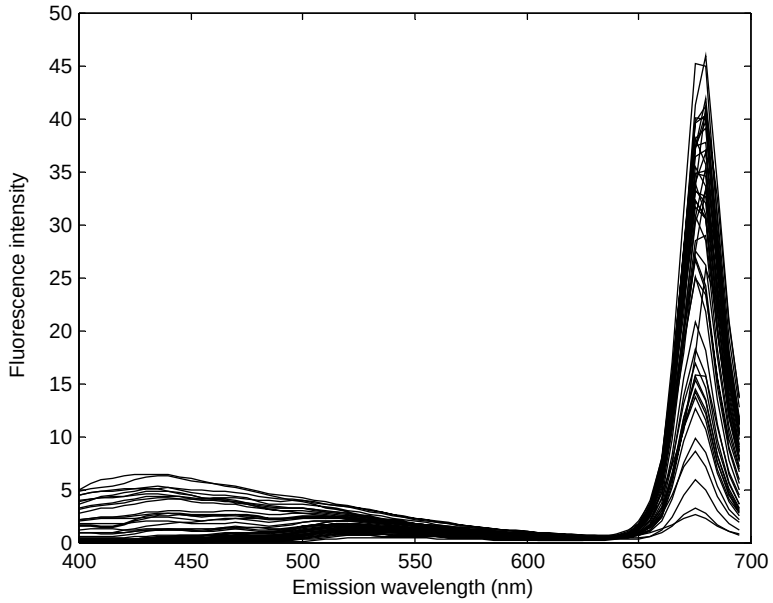

Fig. 4. Fluorescence emission spectra (λ_{em} = 400–695 nm) of the 49 olive oils, measured at λ_{ex} = 365 nm.

This PC combines the contribution of the chlorophyll (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.

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-centred. 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

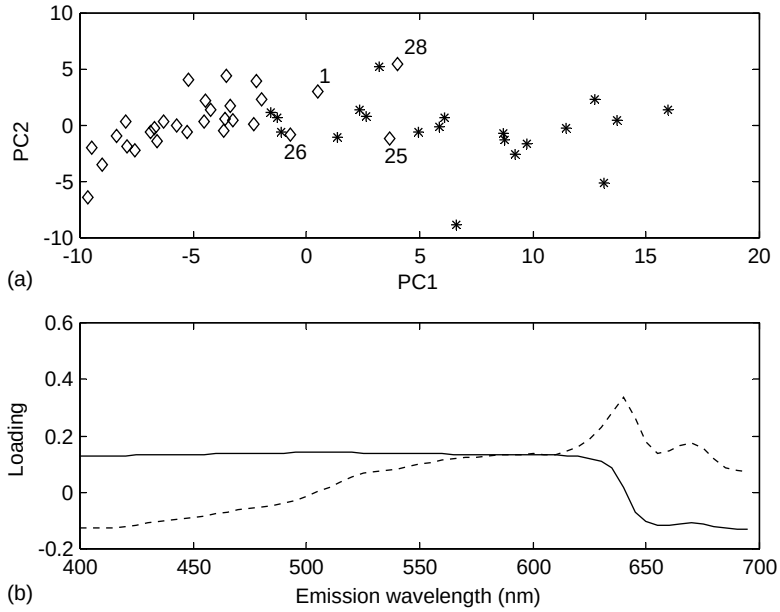

Fig. 5. (a) Score plot from PCA calculated of the emission spectra of the oils (λ_{em} = 400–695 nm) at λ_{ex} = 365 nm. (◇) Virgin olive oils; (✳) pure olive oils. (b) Loadings plot. Continuous line: PC1; dotted line: PC2. Column autoscaled data.

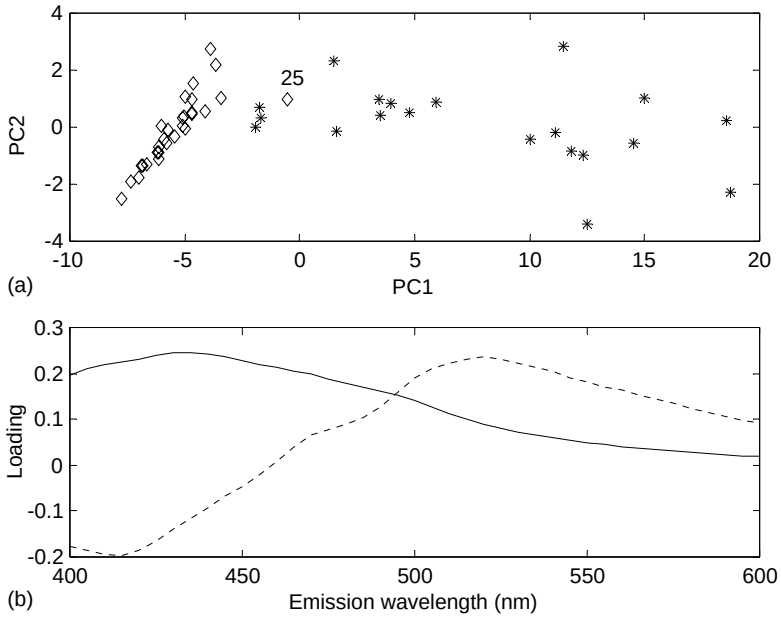

Fig. 6. (a) Score plot from PCA calculated of the emission spectra of the oils (λ_{em} = 400–600 nm) at λ_{ex} = 365 nm. (◇) Virgin olive oils; (✳) pure olive oils. (b) Loading plot. Continuous line: PC1; dotted line: PC2.

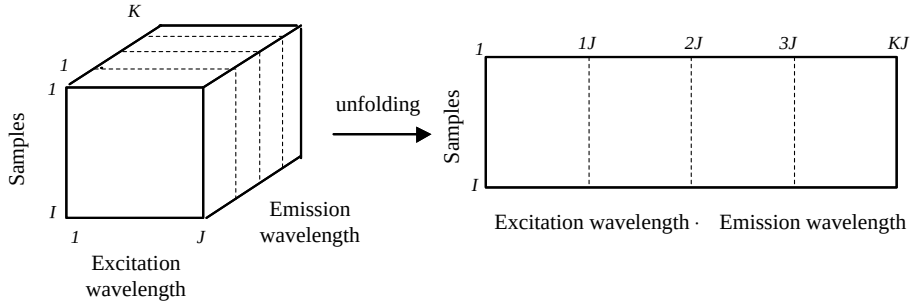

Fig. 7. Arrangement of the EEMs in a cube and unfolding by combining the spectral modes.

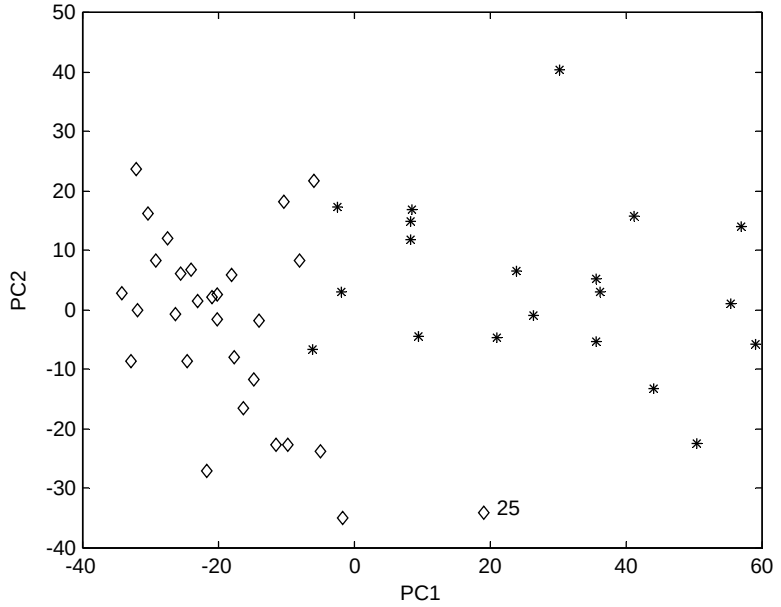

Fig. 8. Score plot from U-PCA of the 49 × 1260 matrix (column autoscaled). (◇) Virgin olive oils; (✳) pure olive oils.

compositional variability is much larger than for virgin oils. Moreover, the virgin olive oils group are 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 can not 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.

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 $49 \times 21 \times 60$ and $49 \times 21 \times 41$ (samples $\times$ λ_{ex} $\times$ 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 chlorophylls peaks) were obtained. Then PCA was calculated on the unfolded matrices [19,20].

As in Section 3.3, when the fluorescence emission zone of chlorophylls was included in the model, the most remark-

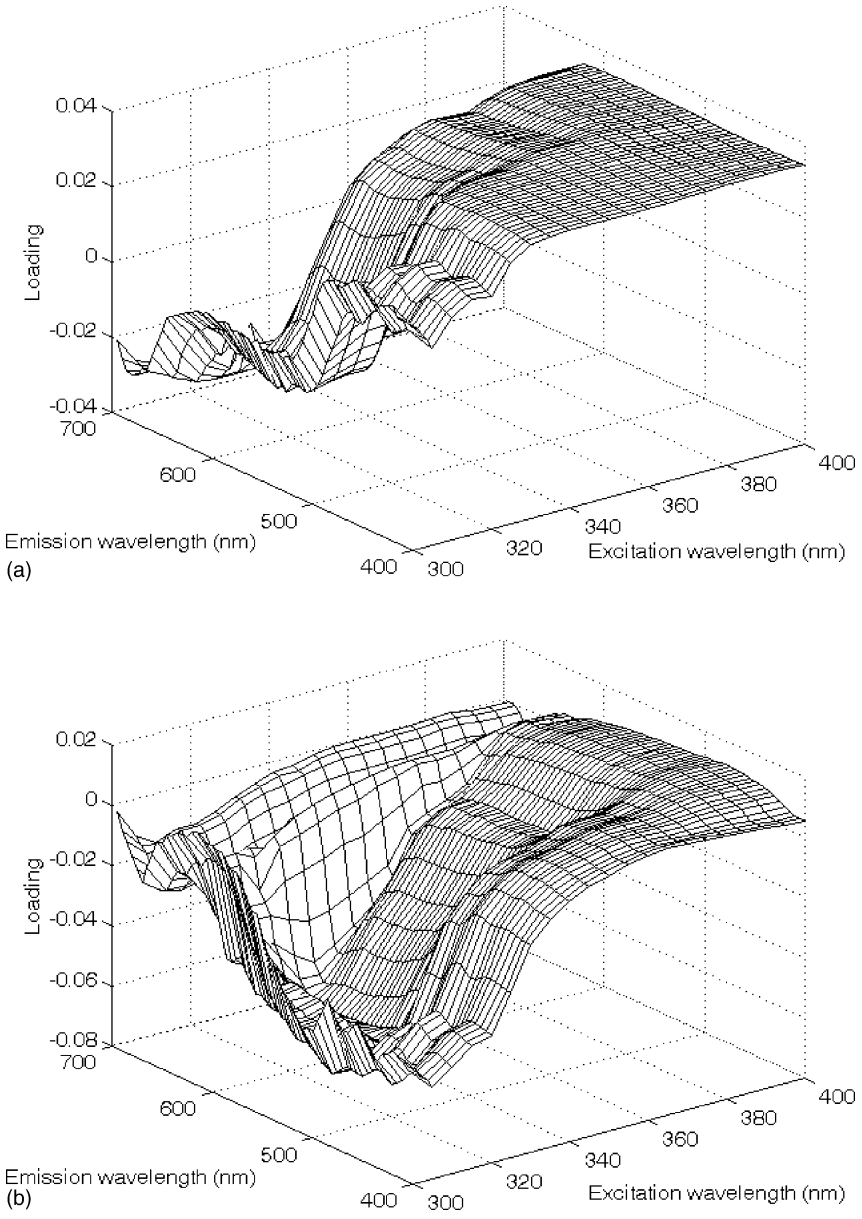

Fig. 9. Refolded loadings from U-PCA of the 49×1260 matrix (column autoscaled). (a) PC1; (b) PC2.

able 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). The loadings of PC1 and PC2 are similar to those found when PCA was applied to the fluorescence spectra at $\lambda_{ex} = 365$ 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 slow peak between $\lambda_{em} = 600$ 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-centred. 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 $\lambda_{ex} = 300$–400 and $\lambda_{em} = 400$–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 the 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.

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 $\lambda_{ex} = 300$–400 nm, $\lambda_{em} = 400$–695 nm. Data were scaled within the emission mode: EEMs were unfolded to a 60 × 1029 matrix ($\lambda_{em} \times$ (sample × 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 $\lambda_{em} = 440$, 475 and 525 nm (Fig. 12d). The second compo-

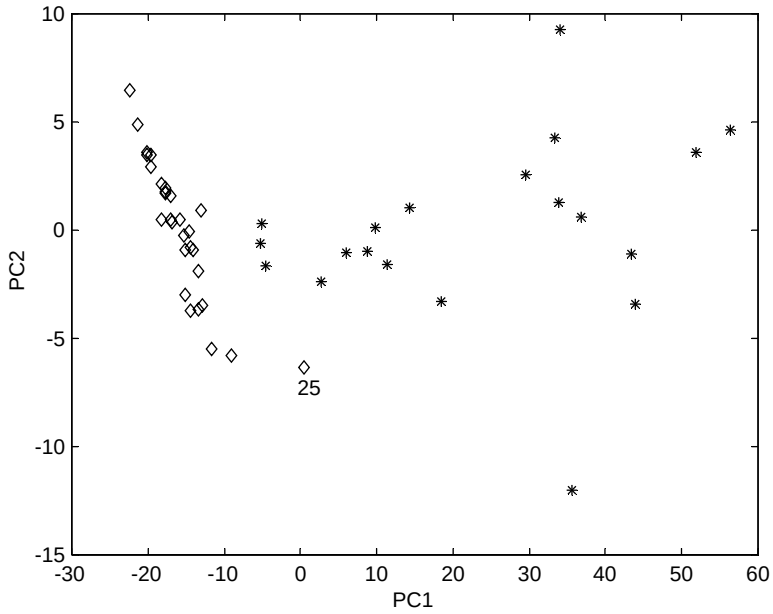

Fig. 10. Score plot from U-PCA of the 49 × 861 matrix (column centred). (◇) Virgin olive oils; (✳) pure olive oils.

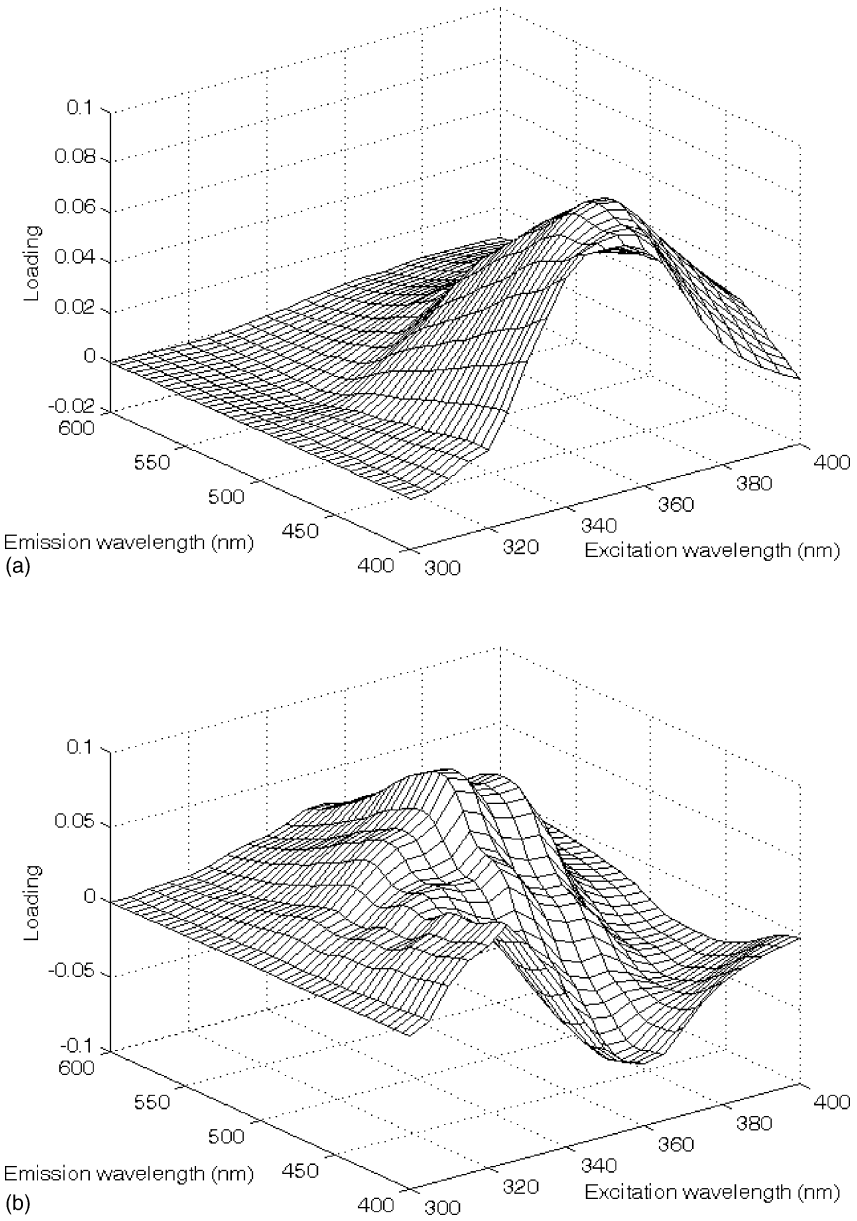

Fig. 11. Refolded loadings from U-PCA of the 49 × 861 matrix (column centred). (a) PC1; (b) PC2.

nent 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.

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

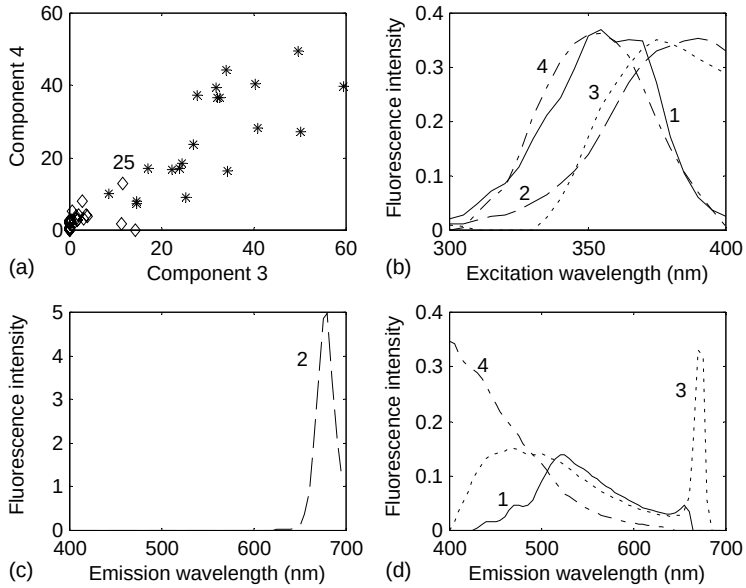

Fig. 12. (a) Score plot on component 3 and 4 of the four-component PARAFAC model calculated from the $49 \times 21 \times 60$ cube of EEM spectra, (b) excitation profiles (λ_{ex} = 300–400 nm), (c) and (d) rescaled emission profiles (λ_{em} = 400–695 nm). Continuous line: component one; broken line: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

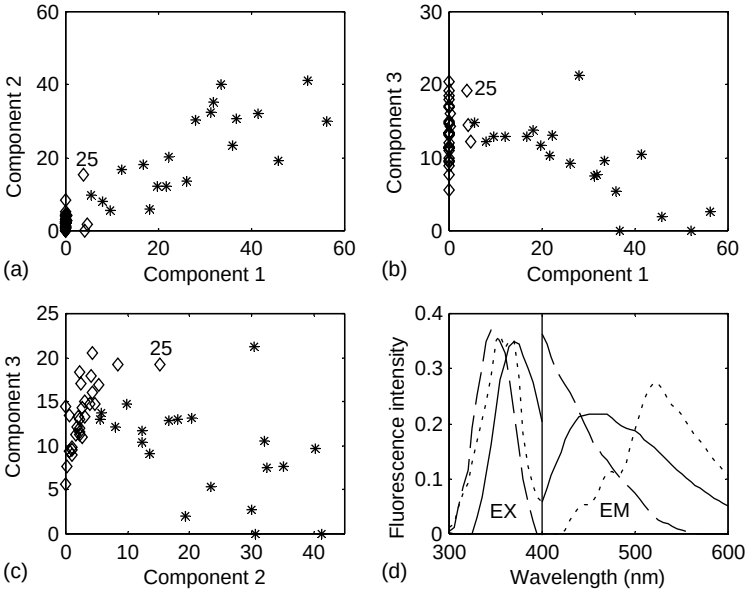

Fig. 13. (a–c) Score plots of the three-component PARAFAC model calculated from the $49 \times 21 \times 41$ cube of EEM spectra, (d) excitation (λ_{ex} = 300–400 nm) and emission (λ_{em} = 400–600 nm) profiles. Continuous line: component one; broken line: component two; dotted line: component three.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

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.

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-centring is enough for U-PCA and no pre-processing 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

We would like to thank the Spanish Ministry of Science and Technology (project no. BQU2003-01142) for financial support and the Rovira i Virgili University for a doctoral fellowship.

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 Mahworks, 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. Chemom. 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.