ELSEVIER

0308-8146(95)00213-8

Food Chemistry, Vol. 57, No. 2, pp. 245–251, 1996
Copyright © 1996 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0308-8146/96 $15.00+.00

Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra

N. Dupuy, L. Duponchel, J. P. Huvenne,* B. Sombret & P. Legrand

Laboratoire de Spectrochimie Infrarouge et Raman–LASIR–CNRS UPR 2631L, Ecole Universitaire d'Ingénieurs de Lille, Université des Sciences et Technologies de Lille, Bâtiment C5, 59655 Villeneuve d'Ascq Cedex, France

(Received 7 June 1995; accepted 1 November 1995)

Principal component analysis of Fourier transform infrared (FTIR) spectra is performed to classify edible fats and oils with regard to their origin. Convenient sampling methods are proposed to record reproducible spectra: horizontal attenuated total reflection for fat samples and mid-IR optical fibre method for oils. The first principal component extracted from the first derivatized spectra of fats allows separation of butters and vegetable margarines on the basis of the sign of the projection. The interpretation of this first principal component establishes that the difference is due to the concentration of unsaturated fatty acids. For the oil samples, the spectral data must be second derivatized and again the first principal component separates sunflower seed oils from the olive and peanut extracts by the different concentrations of linoleic acid. The major interest of these methods using chemometric analysis of spectral data is in their rapidity, since no chemical treatment of samples is required. Copyright © 1996 Elsevier Science Ltd

INTRODUCTION

The quality of foods may vary greatly as a function of their origin, their method of production or manufacture. Prices may vary and the industry needs precise quality control, especially since the advent of new international and European Community regulations (Soroste, 1994). This quality control (Dupuy *et al.*, 1993) must be rapid, sensitive and cheap and lead to integration with the production process.

Fats and oils (Feinberg *et al.*, 1987) are among the oldest classes of chemical compounds used by humans. They are composed of glyceric esters of fatty acids and are distinguished by their physical state: fats are solid at ambient temperature, whereas oils are liquid. The principal variation in composition is the chain length and degree of unsaturation of the component fatty acids (Sadeghi-Jorabchi *et al.*, 1991). Butter and margarine fats have a high concentration of saturated fatty acids, which accounts for their solid physical state (Hasenhuettl, 1993). Oils such as sunflower seed oil have a high concentration of monounsaturated and polyunsaturated fatty acids. Butter is 100% of milk origin (Gropiron, 1988). Olive oil is 100% of olive origin, which corresponds with a high level of monounsaturated fatty acids.

Since the infrared (IR) spectra contain significant information about all the components of a complex mixture, Fourier transform infrared (FTIR) spectroscopy is a very powerful and general technique for investigating the structure of food components and for monitoring changes in the quality of food. In association with chemometric treatments such as principal component analysis (PCA) (Martens, 1979), vibrational spectroscopy allows classification of foods to be undertaken without any chemical analysis (Dupuy *et al.*, 1995*a*). The main advantage is that no *a priori* information on the mixtures is needed since the significant information is extracted during the statistical treatment. In this study, IR spectra were recorded using two convenient sampling methods: horizontal attenuated total reflection (ATR) for the fats, and optical fibre analysis (Sadeghi-Jorabchi *et al.*, 1994) for the oils. The spectral information constitutes the experimental data which are analysed by PCA to classify fats into butter or margarine groups, and to identify sunflower seed oil from among different oil samples.

MATERIALS AND METHODS

The classification of fats involved 17 commercial samples. Of these, seven were butterfats (F1, F2, F3, F5, F6,

---

*To whom correspondence should be addressed.

F14, F15); of the ten other samples, which were essentially of vegetable origin, four margarines were hardened sunflower oils (F9, F10, F13, F17), three others were unspecified hardened oils (F8, F11, F12), and three were margarine spreads with added butter (F4, F7, F16).

Fifteen commercial oil samples were studied for oil classification. They were sunflower seed oils (O1–O8), olive oils (O9–O11) and peanut oils (O12–O15). All the food samples were obtained from a local supermarket.

IR spectra were recorded with a IFS48 Bruker spectrometer. The Specac horizontal ATR accessory equipped with a ZnSe crystal was used for the fat samples. Each sample was exposed to six reflections and

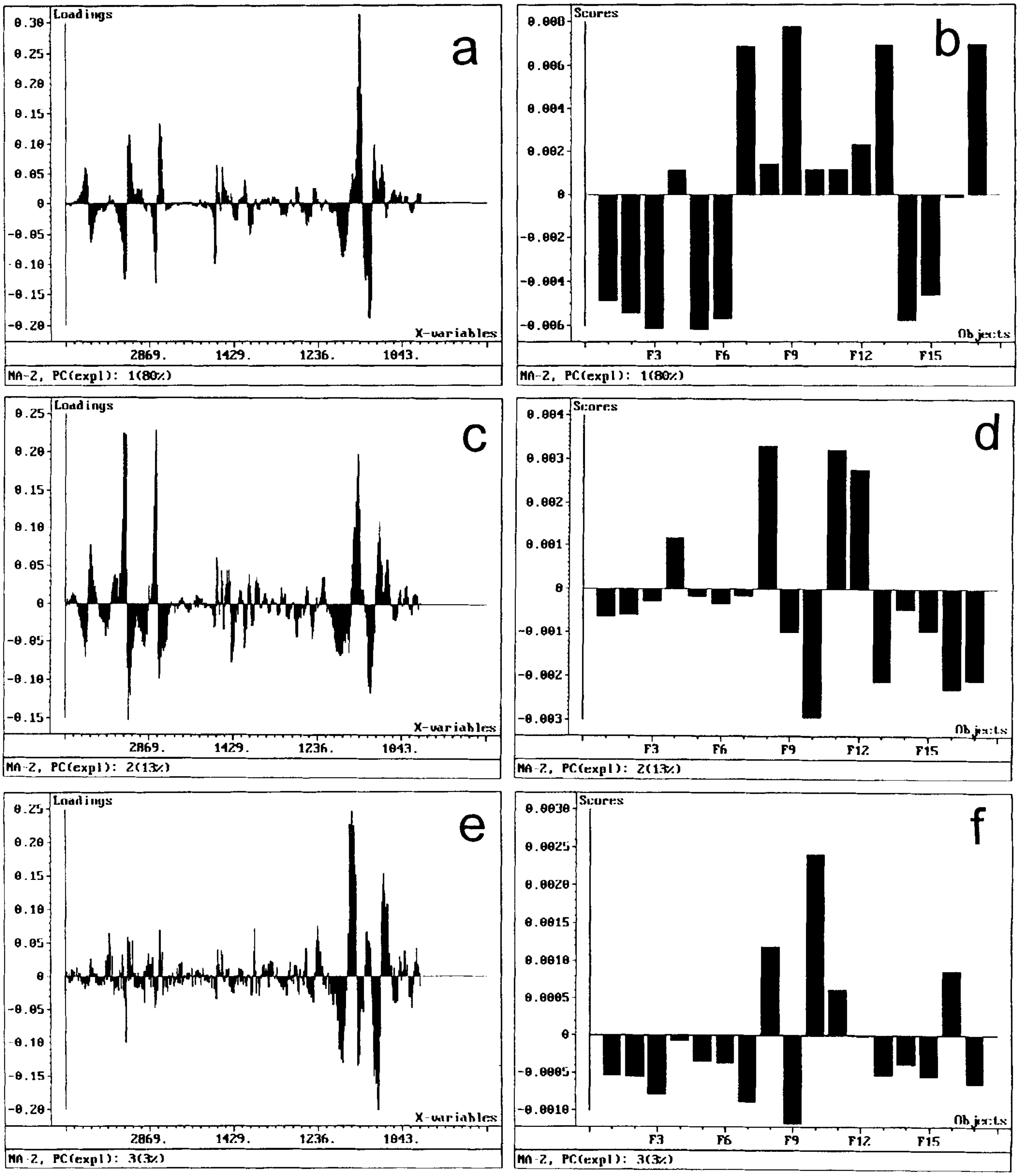

Fig. 1. Principal component analysis of the first derivatized spectra of fats. a, c, e: the three first loadings; b, d, f: the associated scores.

the number fixed by the 45° angle between the beam and the crystal surface. To increase the repeatability, the extraction process was improved. One gram of fat was dissolved in 6 ml of ethyl ether. The organic phase was recovered and dried by Na_2SO_4. A small quantity of the liquid (0.3 ml) was put on the flat crystal and evaporated to leave a thin film of the extracted fat. The spectrum of the oil sample was recorded with a Graseby Specac fibre probe (Dupuy *et al.*, 1995*b*). The 1.5-m vitreous chalcogenide fibre was equipped with a two-reflection ZnSe ATR crystal which allows direct measurements by immersing the probe in the sample. The FTIR spectrometer is allowed to purge to minimize spectral contributions due to atmospheric carbon dioxide and water vapour. We co-added symmetrical interferograms on 100 scans for ATR measurements and 500 scans for optical fibre measurements. A reference spectrum was recorded before each sample and the absorbance spectrum was computed between 4000 cm^{-1} and 700 cm^{-1} for ATR measurements (900 cm^{-1} for optical fibre) at 4 cm^{-1} resolution with the triangular apodization function of the standard Bruker software. In order to resolve complex peaks due to additive spectral overlaps and to eliminate possible shapes of the baseline, the spectra were derivatized. Differentiation and smoothing were performed simultaneously according to the Savitzky & Golay (1964) algorithm.

PCA is a method for extraction of the systematic variations in one data set (Martens & Naes, 1989). The method can be used for classification as well as for description and interpretation. PCA is oriented towards modelling the variance/covariance structure of the data matrix into a model which considers the noise as an error. The components are found during the calibration one by one. Each principal component represents the main systematic variation in the data set which can be modelled after the extraction of the previous one. For instance, in Fig. 1 the principal components called 'loadings' are presented in the first column (Fig. 1(a), 1(c), 1(e)) as a distribution of the variance with respect to the wavenumbers. For each sample, the initial data may be reconstructed by the sum of the loading multiplied by coefficients called 'scores'. The second column of Fig. 1 presents for each principal component the scores with respect to the sample numbers (F1–F17). The common characteristics are modelled in one or several principal components for which the scores are not significantly different according to the species. On the other hand, the information which differentiates the species contributes to principal components whose scores were significant. The classification of the samples was done on the basis of the scores since the characteristics of each species were established by the interpretation of the specific loading.

The PCA of the derivatized spectral data is performed by the UNSCRAMBLER software Version 3.1 from CAMO (Computer Aided MOdelling, Trondheim, Norway).

RESULTS AND DISCUSSION

Reproducibility of the spectral data

Since the fats are soft materials, a good optical contact is easily obtained with the flat ATR crystal (Fahrenfort, 1961). Nevertheless, the repeatability measured on the intensities at five wavenumbers on the derivatized spectra was not excellent. An evolution was observed during the time required to record ten spectra of the same sample. This observation has been interpreted as a physical change of the fat due to the temperature at which the sample was exposed. The decantation of the aqueous part of the sample makes the contact bad. For this reason, separation of the fat by ethyl ether extraction was performed before each spectral recording. This method leads to very good reproducibility on ten spectra of the same extract (A, Table 1) and on the spectra of ten extracts of the same fat (B, Table 1).

The repeatability of the spectra recorded on ten samples of the same oil using the fibre probe has been established. Because of relatively weak transmission of this mid-IR fibre, about 5% for a 1.5-m length, the variability is important when the number of accumulated scans is less than 500. For this number of scans, which involves a recording time of about 4 or 5 min, the calculated coefficient of variation is 6 or 7% (Table 2). This result obtained with the optical fibre may be compared to the one measured by horizontal ATR of the same oil on a hollow ZnSe crystal.

Classification of fats

In order to reduce the number of data points, we worked on two spectral ranges, 3060–2750 cm^{-1} and 1500–1000 cm^{-1}, of the first derivatized spectra. In Fig. 1 it can be seen that the two first extracted loadings explain 93% of the data matrix variance. The third component and the further ones contribute less than 3% of the residual variance and we may assume that these components model non-significant variations such as noise or sampling variations.

The first principal component explains 80% of the data set variance and represents the most dominant information since the scores associated with this component reveal that the extracted spectral features are highly correlated with the fat origin (Fig. 1(b)). The

Table 1. Reproducibility of ten first derivatized spectra

Wavenumber (cm^{-1})	Reproducibility[a] (%)	
	A	B
2929	1.45	0.94
2856	0.80	1.48
1751	0.68	0.81
1176	2.00	2.90
725	1.05	2.20
Mean	1.20	1.70

[a] A: same extract of one fat. B: ten extracts of the same fat.

Table 2. Reproducibility of ten first derivatized spectra of one oil

Wavenumber (cm^{-1})	Reproducibility[a] (%)	
	A	B
2931	3.6	7.5
2858	3.3	6.0
1749	3.4	4.5
1173	3.4	3.1
1105	4.0	6.6
Mean	3.5	5.5

[a]A: measured by horizontal ATR; B: measured with the fibre probe.

butters are negatively projected while the scores of the margarines are all positive. The first principal component intensity is maximum in the spectra' ranges 3060–2800 cm^{-1} and 1200–1050 cm^{-1}. A more precise analysis of this first loading shows that the differentiation is done at three significant spectral points noted at 3008, 1140 and 1108 cm^{-1} on the first derivatized spectra. At 3008 cm^{-1} the contribution is connected to the C–H vibration band of the ethylenic group (Bellamy, 1975), which is characteristic of oleic or linoleic structures in comparison with the saturated esters. We can assign it to the margarine species since these two unsaturated components are more concentrated in margarine than in butter, which has a fairly high concentration of short or medium chain saturated fatty acids (Table 3).

In the same way, the second significant absorption at 1108 cm^{-1} can be assigned to oleic acid (as will be seen in the following section on vegetable oils). Thus, it constitutes another characteristic of the margarine group associated with the unsaturated fatty acid level in the fat. But the more intense contribution to the first loading appears at 1140 cm^{-1}. Its interpretation requires precise comparison with the first derivatized spectra (Fig. 2) of the two classes: butter (sample F3) and margarine (sample F9). We can observe that this important contribution is due to a slight shift of the band at 1135–1140 cm^{-1} combined with the presence of lateral bands with different intensities in the derivatives. So, in contrast with the previous observations which have been clearly interpreted, this important contribution is more difficult to assign precisely. We can only assume that it corresponds to very small differences in the profile of a very characteristic vibration band observed for all the species of fatty acids. The chemometric treatment is able to extract this very subtle but significant information.

Therefore, we can propose this projection on the first component as a classification method of fats between butter or margarine clusters. According to the chemical composition of the various fats, the differentiation is partially done on the concentration of oleic and linoleic acids.

Classification of oils

The spectral data were selected in the regions 3050–2700 cm^{-1} and 1900–940 cm^{-1}, which are more characteristic of the chemical functions of the triglycerides. This choice allows elimination of the fibre contribution which appears between 2150 and 2000 cm^{-1}. Since the classification of the oil samples, whose origin is always vegetable, is more difficult than the fat classification, the data matrix has to be constructed with the second derivatized spectra (Whitbeck, 1981). In this case, the PCA needs the extraction of six components to explain 70% of the data variance. Moreover, we may consider that only the first three are significant since each following component contributes less than 10% of the variance (Fig. 3). These weak contributions are more distributed than those of the previous classification. They correspond to the aleatory variations of spectra which appear in this study to be at least as important as the significant data.

Nevertheless, the scores (Fig. 3(b)) associated with the first loading (Fig. 3(a)), which interprets 25% of the data variance, are all positive for the sunflower seed oils (O1–O8) since they are negative for the seven other oils extracted from olives or peanuts. So we may consider that the projection of the experimental data with regard to the first component allows the identification of the origin as 'sunflower' in comparison with other origins.

Three regions appear to contain significant information on the first loading. At the higher wavenumbers, one peak at 3003 cm^{-1}, already discussed for the fat classification, is readily associated with the concentration of unsaturated fatty acid esters in the different oils (Table 3). At 2852 cm^{-1}, a difference is established between the intensities of the CH_2 or CH_3 vibrations. Between 1200 and 1100 cm^{-1}, the 1102 cm^{-1} contribution is assigned to oleic acid, which is highly concentrated in the olive and peanut extracts. At 1147 cm^{-1}, a significant signal is assigned to a shoulder observed in the original spectra of the different kinds of oils.

Table 3. Mean composition of triglycerides in the different classes of fats and oils studied

Fatty acid (g kg^{-1})	Butter	Margarine	Peanut oil	Olive oil	Sunflower oil
Myristic acid	83.9	33.1	0	0	0
Palmitic acid	205.5	158.1	100	102.2	60
Stearic acid	88.6	65.5	33	35	45
Oleic acid	195.3	274.5	400	727.7	189
Linoleic acid	14	111.6	300	65	640
Saturated acids	526	311.5	200	145	120
Monounsaturated acids	235	329.3	450	740	190
Polyunsaturated acids	20.4	127.4	301.2	71.4	650

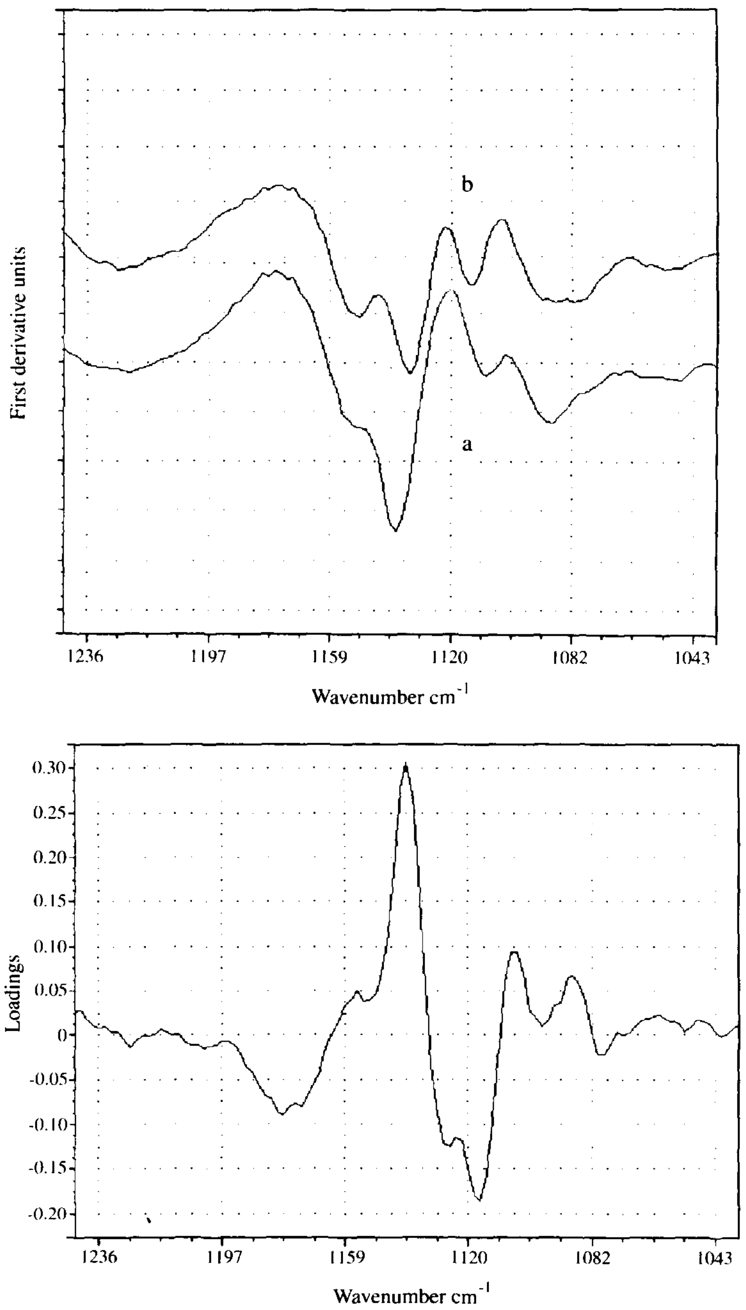

Fig. 2. Comparison of the first derivatized spectra of one butter sample and one margarine sample and the first principal component (Fig. 1(a)) in the 1250–1050 cm^{-1} region.

Finally, below 1000 cm^{-1}, very important contributions are extracted with the first principal component. Taking into account the high disturbance of the profile, there are suspicions about the interpretation of these data. However, when we cut out the spectra above 1000 cm^{-1}, it is not possible to classify the samples. Thus, in spite of its noisy profile, there is a need to concede confidence to this region. In fact, the reference spectra of the oleic and linoleic methyl esters and the three oils justify this third contribution since the profile of the band at 960 cm^{-1} is very different in each case. It can be seen in the second principal component (Fig. 3(c)) that the doubt about the low wavenumber region was real since another contribution is extracted there. However, no classification is possible on the basis of this second point (Fig. 3(d)), leaving the aleatory behaviour out. Thus it can be assumed that the signal below 1000 cm^{-1} is simultaneously due to significant absorption of oil samples and noise resulting from the low transmission of the optical fibre. The statistical treatment selectively extracts these two types of information.

CONCLUSION

A convenient method for identifying the origin of edible fats or oils without any chemical separation and any *a priori* knowledge about their composition is presented.

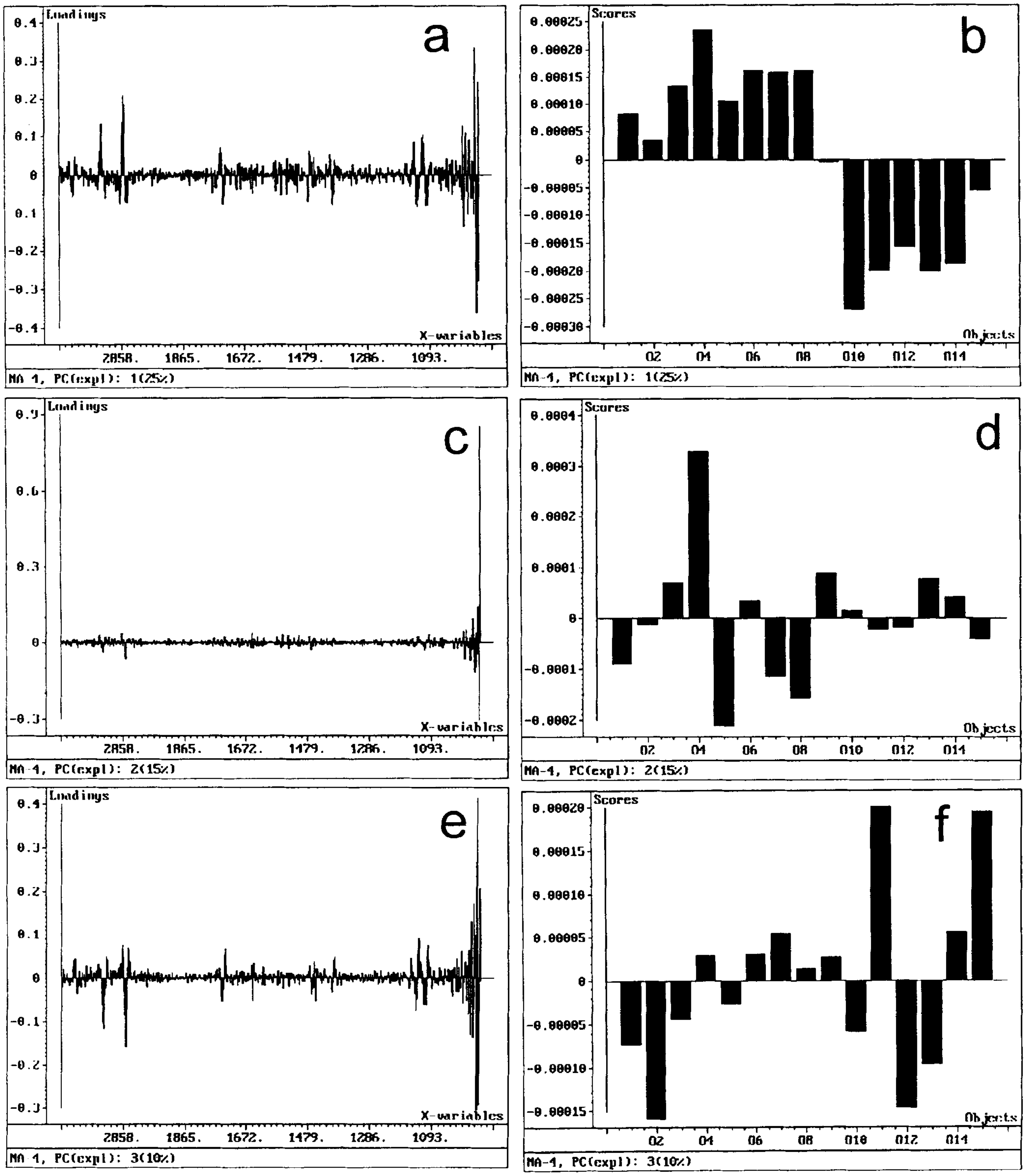

Fig. 3. Principal component analysis of the second derivatized spectra of oils. a, c, e: the three first loadings; b, d, f: the associated scores.

On the basis of their IR spectra, fats are separated into two classes: butter or margarine. Interpretation of the experimental data confirms that the most important fact is the intensity of the spectral bands connected with the relative concentrations of saturated or unsaturated fatty acid triglycerides in relation to their animal or vegetable origin. For the identification of oils, the high concentration of polyunsaturated fatty acid ester allows characterization of the sunflower seed origin compared with the others, i.e. olive or peanut.

A more precise characterization of the different kinds of margarine and the separation of olive and peanut oils were also undertaken. The method can be applied (Dupuy *et al.*, 1995*b*), but as the differences are very subtle the actual results are not completely general and require the study of many further samples and additional classification software.

ACKNOWLEDGEMENTS

We thank the EC concerted actions FLAIR-QUEST (QUality Established by Spectroscopy Techniques) and FAIM (Food Authenticity: Issues and Methodologies) for exchanges and discussions.

REFERENCES

Bellamy, L. J. (1975). *The Infrared Spectra of Complex Molecules*, Vol. I, 3rd edn. Chapman & Hall, London, pp. 48–50.

Dupuy, N., Huvenne, J. P., Sombret, B. & Legrand, P. (1993). Quantitative analysis by mid-infrared spectrometry in food and agro-industrials fields. *J. Mol. Struct.*, **294**, 223–226.

Dupuy, N., Huvenne, J. P., Duponchel, L. & Legrand, P. (1995*a*). Classification of green coffees by FTIR analysis of dry extract. *Appl. Spectrosc.*, **49**(5), 580–585.

Dupuy, N., Huvenne, J. P., Legrand, P. & Le Bourlout, R. (1995*b*). Application du couplage fibre optique/IRTF pour la classification d'huiles végétales par analyse en composantes principales de spectres infrarouges. *Spectra 2000 Analysis*, **185**(5), 36–41.

Fahrenfort, J. (1961). Attenuated total reflection. A new principle for the production of usefull infrared reflection spectra of organic compounds. *Spectrochim. Acta*, **17**, 698–709.

Feinberg, M., Favier, J. C. & Ireland-Ripert, J. (1987). *Répertoire Général des Aliments. Table de Composition des Corps Gras*. Tech. Doc. Lavoisier/Ciqual-Regal, INRA, Paris.

Gropiron, P. (1988). Les industries agricoles et alimentaires. In *Progrès des Sciences et Techniques*. Scriban, Paris.

Hasenhuettl, G. L. (1993). Fats and fatty acids. In *Encyclopedia of Chemical Technologies*, Vol. X, 3rd edn. pp. 252–287.

Martens, H. (1979). Factor analysis of chemical mixture. *Anal. Chim. Acta*, **112**, 423–442.

Martens, H. & Naes, T. (1989). *Multivariate Calibration*. Wiley, New York.

Sadeghi-Jorabchi, H., Wilson, R. H., Belton, P., Edward-Weeb, J. D. & Coxon, D. T. (1991). Quantitative analysis of oils and fats by Fourier transform Raman spectroscopy. *Spectrochim. Acta*, **47A**, 1449–1458.

Sadeghi-Jorabchi, H., Wood, W. M., Jeffery, F., Bruster-Davies, A., Loh, N. & Coombs, D. (1994). Estimation of bio diesel in lubricating oil using Fourier transform infrared spectrometry combined with a mid-infrared fibre optic probe. *Spectrosc. Eur.*, **6**, 16–21.

Savitzky, A. & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, **36**, 1627–1679.

Soroste, A. (1994). In *La Qualité des Produits Alimentaires, Politique, Incitation, Gestion et Contrôle*, ed. J. L. Multon. Dec & Doc, Paris, pp. 320–333.

Whitbeck, M. R. (1981). Second derivative infrared spectroscopy. *Appl. Spectrosc.*, **35**, 93–95.