J Sci Food Agric 1993, **63**, 297–307

Rapid Assessment of the Adulteration of Virgin Olive Oils by Other Seed Oils Using Pyrolysis Mass Spectrometry and Artificial Neural Networks

Royston Goodacre, Douglas B Kell*

Department of Biological Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, UK

and Giorgio Bianchi

Istituto Sperimentale per la Elaiotecnica, Contrada 'Fonte Umano' n 37, 65013 Città S Angelo, Pescara, Italy

(Received 23 December 1992; revised version received 1 April 1993; accepted 7 July 1993)

Abstract: Curie-point pyrolysis mass spectra were obtained from a variety of extra-virgin olive oils, prepared from various cultivars using several mechanical treatments. Some of the oils were adulterated (according to a double-blind protocol) with different amounts of seed oils (50–500 ml of soya, sunflower, peanut, corn or rectified olive oils per litre of mixed oil). Canonical variates analysis indicated that the major source of variation between the pyrolysis mass spectra was due to differences between the cultivars, rather than whether the oils had been adulterated. However, artificial neural networks could be trained (using the back-propagation algorithm) successfully to distinguish virgin oils from those which had been adulterated.

Key words: Curie-point pyrolysis mass spectrometry, artificial neural networks, chemometrics, adulteration, virgin olive oil.

INTRODUCTION

Virgin olive oil is the oil extracted by purely mechanical means from sound, ripe fruits of the olive tree (*Olea europaea* L). Such oils with a free fatty acid content (in terms of oleic acid) below 10 g kg^{-1} are known as 'extra-virgin', whilst oils with good flavour but greater acidity are graded as 'fine' or 'semi-fine' (Kiritsakis and Marakakis 1987, 1991; Kiritsakis 1991; Wessels 1992). Yet lower grades, including those that have been subjected to refining, may be known as 'lampante' or 'pure'. Olive oil is considered to contribute significantly to the health and nutritional benefits of Mediterranean-type diets (Masana *et al* 1991) and, unusually among the vegetable oils, the flavour of virgin olive oils is best enjoyed without refining (Kiritsakis and Min 1989). Olive oil therefore commands a higher price than do other vegetable oils, and these and other properties mean that there is a great temptation to adulterate olive oils with other seed oils (Kiritsakis and Marakakis 1987, 1991; Kiritsakis 1991). Although a number of methods has been proposed for the characterisation of olive oils, and thus for the detection of adulterants (eg Pallotta 1976; Kapoulas and Passaloglou-Emmanoulidou 1981; Eddib and Nickless 1987; Kiritsakis and Marakakis 1987, 1991; Aparicio 1988; Armanino *et al* 1989; Passaloglou-Emmanoulidou 1990; Albi and Guttiérez 1991; Aparicio *et al* 1991a,b; CEC 1991; Coors 1991; Guth and Grosch 1991; Kiritsakis 1991; Norman 1991; Sato *et al* 1991; Aparicio *et al* 1992; Grob 1992; Grob *et al* 1992; Santinelli *et al* 1992), none seems to have found universal acceptance. In the present work, we

* To whom correspondence should be addressed.

J Sci Food Agric 0022-5142/93/$06.00

TABLE 1
Properties of the extra-virgin and adulterated extra-virgin olive oils used in the training set

Codename	*Cultivar*	*Adulterant (ml litre^{-1} mixed oil)*	*Hammer mill*	*Stone mill*	*Percolation*	*Centrifuge*
Lucia	Leccino 1		+		+	
Gabriella	Leccino 2		+		+	
Anna	Leccino 1	100 sansa oil	+		+	
Ezilde	Leccino 1		+			+
Vanda	Leccino 2		+			+
Pietro	Leccino 1	50 sansa oil	+			+
Claudia	Leccino 2	200 sunflower oil	+			+
Patrizia	Leccino 1	50 peanut oil	+			+
Alfonso	Coratina 1		+	+	+	
Mario	Coratina 2		+	+	+	
Rosa	Coratina 1	100 sunflower oil	+	+	+	
Sandra	Coratina 1	300 peanut oil	+	+	+	
Catia	Coratina 1		+	+		+
Walter	Coratina 2		+	+		+
Giulia	Coratina 1	100 soya oil	+	+		+
Paola	Coratina 2	150 sansa oil	+	+		+
Leonardo	Dritta 1		+		+	
Mara	Dritta 2		+		+	
Giuseppe	Dritta 1	200 corn oil	+		+	
Mira	Dritta 2	50 soya oil	+		+	
Ugo	Dritta 1		+			+
Giorgio	Dritta 2		+			+
Augusto	Dritta 1	50 sunflower oil	+			+
Luca	Dritta 2	150 peanut oil	+			+

show that a combination of Curie-point pyrolysis mass spectrometry (PyMS) (Irwin 1982; Meuzelaar *et al* 1982; Aries *et al* 1986; Goodacre and Berkeley 1990) with multivariate data analysis using artificial neural networks (ANNs) (Rumelhart *et al* 1986; Eberhart and Dobbins 1990; Simpson 1990; Hertz *et al* 1991) has permitted us to effect a rapid assessment of the adulteration of extra-virgin olive oils with various seed oils. A preliminary report of this work has appeared (Goodacre *et al* 1992).

METHODS

General protocol

Two sets of samples were prepared at the Istituto Sperimentale per la Elaiotecnica, each consisting of 12 samples of various extra-virgin olive oils plus 12 samples variously adulterated with 5–50% of soya, sunflower, peanut, corn or rectified (sansa) olive oils. The first set was the training set whilst the second set was the test set. To ensure generalisability, the two sets were wholly different (ie no oil in the test set had also been in the training set), and were designed to encompass a variety of representative cultivars, crushing protocols and storage regimes, as illustrated in Tables 1–3. The experiment was performed double-blind, such that the identities of the second set were not known to any of the authors. PyMS was carried out and an ANN trained on the first set as described below, prior to assessing the status of the oils in the (then unknown) test set.

Growth of olives

Olives were obtained from trees aged 10 to > 100 years old from the orchard of the Istituto Sperimentale per la Elaiotecnica, and were harvested during November 1991 using a mechanical shaker and nets, thereby avoiding any contact of the fruit with the soil. Any leaves were removed within 2 days of harvest, and sound, ripe olives used for processing. Batches were stored on the floor of a large room to a thickness < 20 cm, and at a temperature of 6–10 °C until required. The actual cultivars used (and indeed the methods of harvest and extraction) are those most representative of present commercial practice in Italy, and are listed in Tables 1–3.

Production of olive oils

Batches of at least 300 kg of olives were used, and processed at the Istituto Sperimentale per la Elaiotecnica

TABLE 2
Properties of the extra-virgin olive oils used in the test set

Codename	*Storage*	*Cultivar(s)*	*Maturity*[a]	*Crushing method(s)*[b]	*Extraction method(s)*[c]
Cagliari	4·5 t underground tank	Several	BB	H	PR,PE,CCC
Milano	4·5 t underground tank	Several	GG,B	S,HHH	PE,CCC
Bolzano	0·45 t container	Coratina	GG,B	S,H	PE,CC
Napoli	1-litre bottle	700 ml Carolea 300 ml Coratina	G,BB	H	C
Perugia	1-litre bottle	Several	G,B	H	C
Firenze	1-litre bottle	Carolea	G,B	H	PR
Lecce	1-litre bottle	Coratina	G,B	H	C
Genova	1-litre bottle	Leccino	G,B	H	C
Urbino	1-litre bottle	Several	G,B	S	PR
Bari	1-litre bottle	Carolea	G,B	S	PR
Trieste	1-litre bottle	300 ml Grossa Cassano 700 ml Coratina	G,B	H	C
Ancona	1-litre bottle	Cipressino	G,B	H	C

[a] Maturity is coded according to the proportion of green (G) and black (B) olives that were present before the oil was extracted. GG,B means that there were significantly more green than black olives, and so on.
[b] Crushing methods: H, hammer mill; S, stone mill; as described in the Methods section.
[c] Extraction system: PR, pressing; PE, percolation; C, centrifugation for more (CCC) or less (C) time.

TABLE 3
Properties of adulterated olive oils used in the test set[a]

Codename	*Adulterant (ml litre^{-1} mixed oil)*
Roma	100 sansa olive oil
Taormina	300 sansa olive oil
Torino	500 sansa olive oil
Siena	100 peanut oil
Venezia	300 peanut oil
Rimini	500 peanut oil
Bologna	100 sunflower oil
Pescara	300 sunflower oil
Palermo	100 corn oil
Messina	300 corn oil
Verona	100 soya oil
Padova	300 soya oil

[a] In all cases, the extra-virgin olive oil was that from the 4·5 t underground tanks of Table 2, codenames *Cagliari* and *Milano* and had been stored as a bottled oil in 1-litre bottles. It may be noted that these oils were a mixture of oils extracted by pressing, percolation and centrifugation, such that the oils are representative of all methods of extraction. Adulterants were added as described in the Methods section.

as follows. Any remaining leaves were removed mechanically. Crushing was effected using either (single crushing) or both (double crushing) of a hammer mill and a stone mill (in that order if both were used), although most samples were crushed only singly (Tables 1 and 2). Hammer mill crushing was for a few minutes, stone mill crushing for 10 min or so. Olive paste so obtained was then mixed in a malaxator, typically for 60 min at 20–30 °C, to allow oil droplets to coalesce. The oil was then separated from the paste by pressing and/or percolation and/or by centrifugation. Pressing was performed by spreading olive paste on filtering diaphragms separated every three to four bags by metal trays and then subjected to a hydraulic pressure of 380–400 kg cm^{-2}. Olive oil and water then flowed to an appropriate tank whence the oil was dewatered by centrifugation in a vertical-axis centrifuge. If separation was via percolation, olive paste from the malaxation was poured into a 'Sinolea' apparatus (Rapanelli, Foligno), which consists of a tank fitted with a stainless-steel grating through which plates pass; motion of these plates causes the oil to adhere to them and thence be removed. Alternatively, rapid separation of the oil/water/solid mixture in the paste was attained via centrifugation alone, using a horizontal-axis centrifuge ('Decanter',

Rapanelli Model Ramef 4000/S) with a mean radius of 30 cm and operating at 4000 *g*.

All olive oils were initially stored either in closed, stainless-steel tanks with a capacity of 400–500 kg or in underground tanks with a capacity of 4·5 t. Small samples of these were sealed in glass ampoules and sent by air mail to Aberystwyth for the analyses, which were carried out within two days of receipt. One of the adulterated oils from the test set (Verona) was broken in the post and was not analysed.

Other oils

Seed oils were purchased in a local Italian supermarket and were produced by the following concerns: peanut oil, Società Alimentare Lucchese Industria Oli (Lucca) and Consilia (Assago, Milano); corn oil, Società Alimentari Lucchese Industria Oli (Lucca); soya oil, Cerol Italia SpA (Ravenna); sunflower oil, ICIC SpA (Ancona); Sansa olive oil, Monini SpA (Spoleto). Adulteration was carried out volumetrically.

Sample preparation for pyrolysis mass spectrometry

Clean iron-nickel foils (Horizon Instruments Ltd, Heathfield, E Sussex, UK) were inserted, using clean forceps, into clean pyrolysis tubes (Horizon Instruments), so that 6 mm was protruding from the mouth of the tube. An appropriate volume of the olive oil sample was smeared on 5 mm of a protruding foil to give a thin uniform surface coating. The samples were oven dried at 50 °C for 30 min, then the foils were pushed into the tube using a stainless-steel depth gauge so as to lie 10 mm from the mouth of the tube. Finally, Viton 'O'-rings (Horizon Instruments) were placed on the tubes. All oils were analysed in quadruplicate.

Pyrolysis mass spectrometry

The pyrolysis mass spectrometer used in this study was the Horizon Instruments PyMS-200X as described by Aries *et al* (1986). The pyrolysate was generated in a vacuum by the heating of a ferro-magnetic foil carrying the sample. Heating was achieved by passing a radio-frequency (0·4 MHz) current for 3 s through a pyrolysis coil which surrounds the sample-coated alloy foil. The foil and sample heated rapidly, within 0·5 s, to the temperature corresponding to the Curie-point of the iron–nickel foil. At this temperature, 530 °C, the alloy ceased to exhibit ferro-magnetic properties and heating finished; on cooling below the Curie-point, inductive heating resumed, so that the foil-pyrolyser system acted as a thermostatic switch maintaining the sample at the Curie-point, until current ceased to flow through the pyrolysis coil. The pyrolysate then entered a gold-plated expansion chamber heated to 150 °C, whence it diffused down a molecular beam tube to the ionisation chamber of the mass spectrometer.

The pyrolysate was bombarded with low-energy electrons (25 eV) producing both molecular and fragment ions (because low energy was used the majority carried only a single positive charge). Non-ionised molecules were deposited on a cold trap, cooled by liquid nitrogen. The ionised fragments were focused by the electrostatic lens of a set of source electrodes, accelerated and directed into a quadrupole mass filter. The ions were separated by the quadrupole, on the basis of their mass-to-charge ratios, and detected and amplified with an electron multiplier. The mass spectrometer scans the ionised pyrolysate 160 times at 0·2 s intervals following pyrolysis. Data were collected over the m/z range 51–200, in one-tenth of a mass-unit interval. These were then integrated to give unit mass. Given that the charge of the fragment was unity the mass-to-charge ratio can be accepted as a measure of the mass of pyrolysate fragments. The IBM-compatible PC used to control the PyMS-200X, was also programmed (using software provided by the manufacturers) to record spectral information on ion count for the individual masses scanned and the total ion count for each sample analysed.

The size of the sample used was typically 5 μl; preliminary experiments indicated how much sample was appropriate to obtain a total ion count in the range $0{\cdot}3–1{\cdot}10^6$; some oils required (and received) only 1 μl to yield such total ion counts. These preparatory experiments also showed that if the sample carrier was heated at 100 °C for 5 s prior to pyrolysis, a procedure which inhibits pyrolysate condensation on the walls of the sample tube and which is an option in the PyMS-200X, evaporated (non-pyrolysed) oil persisted in the mass spectrometer and caused memory effects in the analyses of subsequent samples. Thus, pre-heating of this type was not performed.

Data analysis

The data from PyMS may be displayed as quantitative pyrolysis mass spectra (Fig 1), in which the abscissa represents the m/z ratio whilst the ordinate contains information on the ion count for any particular m/z value ranging from 51 to 200. Data were normalised to a total ion count of 2^{16} to remove the effect of sample size differences.

The normalised data were then processed using the GENSTAT package (Nelder 1979) which runs under Microsoft DOS 5·0 on an IBM-compatible PC. This method has been previously described by MacFie and Gutteridge (1982) and Gutteridge *et al* (1985). In essence, the first stage was the reduction of the data by principal components analysis (Chatfield and Collins 1980; Causton 1987; Gutteridge 1987; Flury and Riedwyl 1988; Martens and Naes 1989), which is a well-

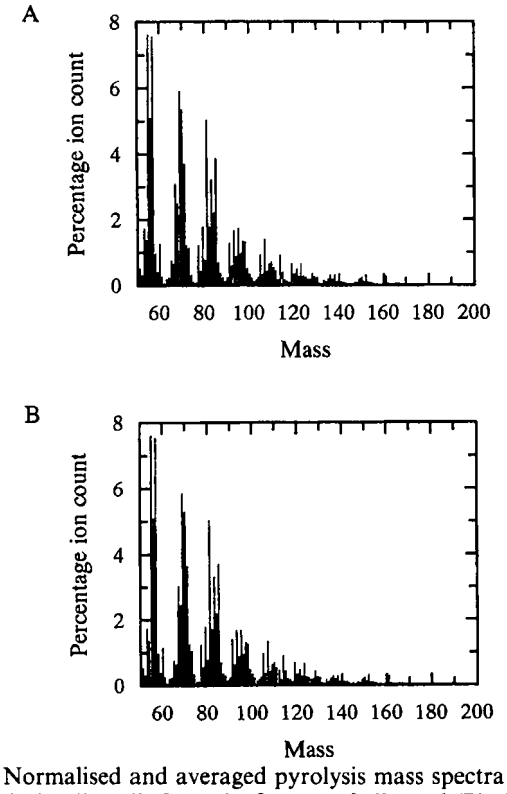

Fig 1. Normalised and averaged pyrolysis mass spectra of (A) all 12 virgin olive oils from the first set of oils, and (B) all 12 of those olive oils that have been adulterated with lower grade oils, also from the first set of oils.

known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance. Data were reduced by keeping only those principal components whose eigenvalues accounted for more than 0·1% of the total variance. Canonical variates analysis (CVA) then separated the samples into groups on the basis of the retained PC and some *a priori* knowledge of the appropriate number of groupings (MacFie *et al* 1978; Windig *et al* 1983). The next stage was the construction of a percentage similarity matrix by transforming the Mahalanobis' distance between *a priori* groups in CVA using the Gower similarity coefficient S_G (Gower 1971). Finally, cluster analysis was employed to produce a dendrogram, using average linkage clustering (Gutteridge *et al* 1985).

All ANN analyses were carried out using a user-friendly, neural network simulation program, NeuralDesk (version 1.2) (Neural Computer Sciences, Southampton, UK), which runs under Microsoft Windows 3.1 on an IBM-compatible PC. To ensure maximum speed, an accelerator board for the PC (NeuSprint) based on the AT&T DSP32C chip, which effects a speed enhancement of some 100-fold, permitting the analysis (and updating) of some 400000 weights s^{-1}, was used. Data were also processed prior to analysis using the Microsoft Excel 4.0 spreadsheet.

For training the ANN, the inputs were the averages of the four normalised replicate pyrolysis mass spectra derived from the olive oil samples, and were further normalised to lie in the range 0–1. For the outputs, virgin olive oils were coded as 1, non-virgins (adulterated oils) as 0.

The algorithm used was standard back-propagation (Rumelhart *et al* 1986), running on the accelerator board. This algorithm employs processing nodes (neurons or units), connected using abstract interconnections (connections or synapses). Connections each have an associated real value, termed the weight, that scale signals passing through them. Nodes sum the signals feeding to them and output this sum to each driven connection scaled by a 'squashing' function with a sigmoidal shape. This 'squashing' function $f = 1/(1+e^{-x})$, where $x = \Sigma$inputs.

The training of the network thus consisted of the preparation of a set of pairs of patterns where one-half of the pair is input to the network and the other represents the known or expected response. The stimulus pattern is applied to the network, which is allowed to run until an output is produced at each output node. The differences between the actual output and that expected, taken over the entire set of patterns, are fed back through the network in the reverse direction to signal flow (hence back-propagation) modifying the weights as they go. This process is repeated until a suitable level of error is achieved (Rumelhart *et al* 1986; Wasserman 1989; Simpson 1990). In the present work, we used a learning rate of 0·1 and a momentum of 0·9.

The structure of the ANN used in this study (and see also Goodacre and Kell (1993) and Goodacre *et al* (1993)) consisted of three layers containing 159 nodes made up of the 150 input nodes (normalised pyrolysis mass spectra), one output node (status as a virgin or non-virgin oil), and one 'hidden' layer containing eight nodes (150-8-1). Eight units were chosen for the hidden layer since it is widely recognised that good generalisation often comes from using a number of nodes in the hidden layer that approximates the natural logarithm of the number of nodes in the input layer. Each of the 150 input nodes was connected to the eight nodes of the hidden layer which in turn were connected to the output node. In addition, the hidden layer and output node were connected to the bias, making a total of 1217 connections, whose weights would be altered during training. Before training commenced the input and output nodes were normalised between 0 and +1, and the connection weights were set to small random values, except the bias which was always set to +1 (Wasserman 1989). Each epoch (one complete calculation in the network) represented 1217 connection weight updatings and a recalculation of the root mean squared (RMS) error between the true and desired outputs over the entire training set. A plot of the RMS error versus the number of epochs represents the 'learning curve', and was used to estimate the extent of training. Training can be said to have finished when the network has found the lowest RMS error, ie the global minimum on the error surface.

RESULTS AND DISCUSSION

After the collection of pyrolysis mass spectra (Fig 1), the first stage was to perform multivariate statistical methods using the GENSTAT package to establish the relationships between the members of the first set of oils. Each of the oils, represented by four replicate spectra, was coded to give 24 individual groups, the resulting dendrogram being shown in Fig 2 and CVA plots displayed on Fig 3. In the dendrogram (Fig 2) it can be seen that the 24 oils cluster into five groups. This can also be observed in the CVA plots; when the first two canonical variates (Fig 3(A)) are plotted, displaying 68·28% of the total variance, only four clusters can be seen, though cluster 5 can be seen to split away from clusters 2 and 3 when the third canonical variate, which accounts for 13·05% of the total variance, is viewed (Fig 3(B)). The most important thing to note is that in Fig 2 the virgin (upper case) and adulterated olive oils (lower case) are not differentiated into two distinct groups. Rather than clustering according to the oil's virginity or otherwise, the olive oils have split mainly on the basis of their cultivars: cluster 1 comprises four *Leccino* and one *Dritta* cultivars (*Giorgio*); cluster 2 contains only oils of *Coratina* cultivar; cluster 3 of four *Leccino* and one *Coratina* cultivars (*Sandra*); cluster 4 members are all cultivar *Dritta*; finally, the single oil (*Walter*) in cluster 5 belongs to the *Coratina* cultivar. Furthermore, in six cases the virgin olive oil and the same olive oil that has been adulterated grouped together; thus, in cluster 1 *Ezilde* groups with *Pietro* (*Pietro* is *Ezilde* (950 ml) contaminated with sansa olive oil (50 ml)); in cluster 2 *Alfonso* groups with *Rosa* (100 ml sunflower oil l^{-1} mixed oil) and *Catia* with *Giulia* (100 ml soya oil l^{-1}); and finally in cluster 3 *Leonardo* groups with *Giuseppe* (200 ml corn oil l^{-1}), *Mara* with *Mira* (50 ml soya oil l^{-1}), and *Ugo* clusters with *Augusto* (which is *Ugo* adulterated with 50 ml sunflower oil l^{-1}).

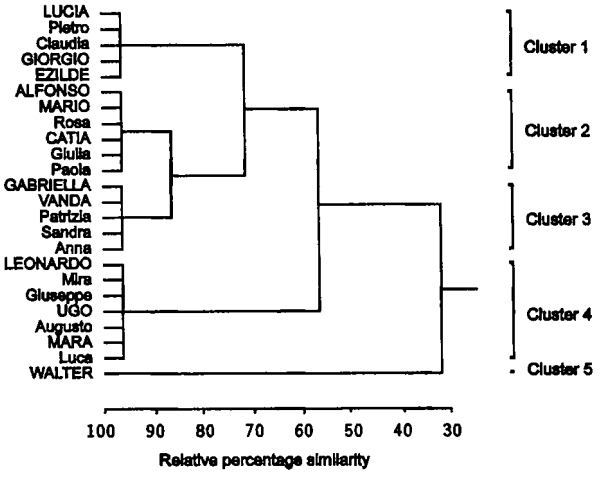

Fig 2. Dendrogram representing the relationships between the first set of oils (the training set) based on PyMS data analysed by GENSTAT. The coded virgin olive oils are in upper case, whilst the adulterated oils are in lower-case letters.

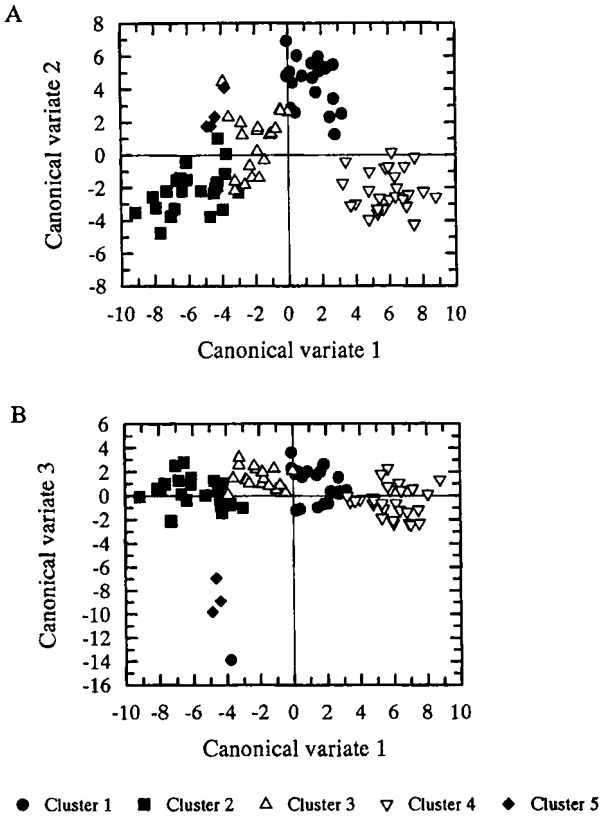

Fig 3. Ordination plot based on PyMS data analysed by GENSTAT showing the relationship between the first set of olive oils. (A) The first and second canonical variates accounted for 68·28% of the total variance, and (B) the third accounted for 13·05% (81·33% total). The five clusters are synonymous with those represented in the dendrogram in Fig 2.

In other experiments GENSTAT was used to analyse the first set of oils where all the virgin olive oils were coded as a single group and all the adulterated oils as another group. The first canonical variate displayed 100% of the total variation, because two groups can be separated in one dimension. It is possible to depict this as a histogram (data not shown) where the abscissa represents the canonical variate distance and the ordinate contains information on the number of samples that appear in that area. For two groups of pyrolysis mass spectra to be statistically discrete (CVA) the group means must be separated by more than 3·84 canonical variate units which represents the 95% confidence limit constructed round each mean by the χ^2 distribution on one degree of freedom (Krzanowski 1988; Ward *et al* 1993). For the separation of virgins from adulterated olive oils this distance was 0·982; using this method of analysis it is not therefore statistically possible to separate these two classes (virgin/adulterated) on the basis of their pyrolysis mass spectra.

Given that the results shown in Figs 2 and 3 indicated that the individual cultivars of the virgin olive oil were statistically distinguishable, it is to be expected that this should be reflected in the distances obtained using the

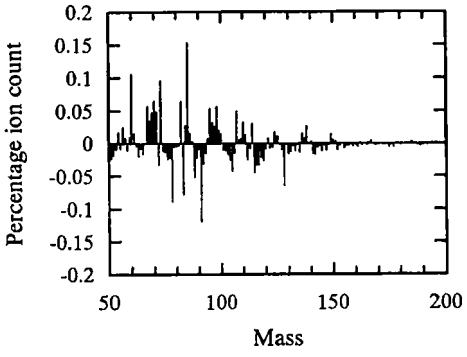

Fig 4. A subtraction spectrum of the normalised average of all the adulterated olive oils (Fig 1(A)) from the average spectra of all the virgin olive oils (Fig 1(B)).

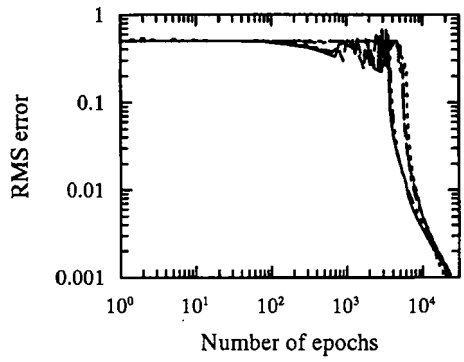

Fig 5. The learning curve(s) for neural networks employing the standard-back propagation algorithm with one hidden layer consisting of eight nodes, and trained to assess virginity using data from the first set of oils.

above technique of observing the CVA group means. Thus, cultivars *Leccino* and *Coratina*, represented by *Lucia* and *Alfonso* respectively, were separated by 14·064 units; *Leccino* and *Dritta*, exemplified by *Lucia* and *Mara* respectively, were 11·66 units apart; finally *Alfonso* (illustrating *Coratina*) and *Mara* (portraying *Dritta*) were separated by 10·696 canonical variate units.

The question therefore arises as to whether it is possible accurately to assess the difference between virgin olive oils and those olive oils that have been adulterated with lower grade oils. Figure 1 shows the normalised averaged pyrolysis mass spectra of all 12 virgin olive oils (Fig 1(A)) and all 12 of the adulterated oils (Fig 1(B)) from the first set of oils. It is clear that the distinction between these two spectra by eye is difficult and any differences have to be visualised by other means. Figure 4 shows a simple subtraction of the normalised average of all the adulterated olive oils (Fig 1(A)) from the average spectra of all the virgin olive oils (Fig 1(B)). The negative half of the graph therefore indicates the peaks that have arisen from the adulterant oils, whilst any peaks in the positive half may be presumed to reflect chemical characteristics that are common to virgin olive oils. It is particularly noteworthy that the (normalised) intensity range of the subtraction spectra in Fig 4 was very small, from only −0·12% to +0·15%. When subtraction of the pyrolysis mass spectra of the different cultivars of virgin olive oils was made (data not shown) the range in intensities was from −0·46% to +1·80% when subtracting *Alfonso* from *Lucia*, −1·78% to +2·42% for *Mara* from *Lucia*, and when *Mara* was subtracted from *Alfonsa* the range was from −1·32% to +2·10%. This strongly indicates that the univariate differences between cultivars of virgin olive oil were greater (by an order of magnitude) than the univariate differences between the virgin olive oils and the same oils that had been adulterated with between 50 and 300 ml of a lower grade oil per litre of mixed oil.

We have seen that the above methods of multivariate analysis, which separate spectra linearly (in terms of their Mahalanobis distance (Gutteridge *et al* 1985)), were unable to yield any reliable information on the virginity or otherwise of the oils. The next stage was therefore to examine the ability of artificial neural networks, which should uncover non-linear relationships between the two types of oils, to assess the virginity of the test set (unknowns) using the first set (knowns) as the training data.

We therefore trained ANN with normalised averaged ion intensities from the pyrolysis mass spectra from the training set (the first set of oils); the 12 virgin olive oils were coded 1 at the output node, and the 12 adulterated oils were coded 0. The 150-8-1 ANN were trained using the standard back-propagation algorithm, and the effectiveness of training was expressed in terms of the RMS error between the actual and the desired outputs; examples of these 'learning curves' are shown in Fig 5. Training was stopped after the RMS error had reached 0·001 and we interrogated using all oils from both sets. Training was effected five times, using randomised, small initial values for the starting weights; even though the learning curves do not superimpose (Fig 5), it is clear, because they follow the same trend, that despite the randomised starting connection weights, training was executed (ie the error surface in weight space was negotiated) in a rather reproducible manner.

These learning curves (Fig 5) display some very interesting neurodynamics. It can be seen that initially training was very slow; the RMS error fluctuated around 0·5 until after some 2000–3000 epochs, and the network did not appear to be learning. Between some 2500 and 4500 epochs the RMS error rather 'suddenly' and reproducibly decreased and after approximately 20000 epochs the RMS error reached 0·001.

When training had ceased (ie as determined by the attainment of an RMS error of 0·001 averaged over the training set) the network was interrogated with the normalised averaged ion intensities of the pyrolysis mass spectra from both sets of oils. Not surprisingly, the network's estimate of the virginity of the training set was the same as the known virginities in all five trainings. The results of the network's final analysis of the unknown test set is shown in Table 4. The network gave the same

TABLE 4
Identity of the double-blind olive oils used in the test set as judged by the ANN

Codename	*Network's answer*[a]	*Virgin or adulterated olive oil*
Perugia	1	Virgin
Lecce	1	Virgin
Urbino	1	Virgin
Rimini	0	Adulterated
Taormina	0	Adulterated
Napoli	1	Virgin
Milano	1	Virgin
Trieste	1	Virgin
Torino	0	Adulterated
Cagliari	0·8[b]	Virgin
Bolzamo	1	Virgin
Venezia	0	Adulterated
Roma	0	Adulterated
Genova	1	Virgin
Bari	1	Virgin
Pescara	0	Adulterated
Padova	0	Adulterated
Palermo	0	Adulterated
Firenze	1	Virgin
Ancona	1	Virgin
Siena	0	Adulterated
Messina	0	Adulterated
Bologna	0	Adulterated

[a] The network was trained and interrogated five times. The scores given are the average of the five runs (to ±0·001), where virgin is coded 1 and adulterated oil coded 0.
[b] The network indicated that oil *Cagliari* was of virgin quality (1) on four of the five trainings.

results on each of the five training runs, with the exception of oil *Cagliari* which was coded virgin on four out of the five instances. When the code was broken it transpired that the network had correctly assessed each oil. In a typical run, the virgins were assessed with a code of 0·99976±0·000146 (range 0·99954–1·00016) and the non-virgins with a code of 0·001079±0·002838 (range 0·00026–0·01009).

A useful way of discerning the manner in which the network trained differentially on the individual exemplars is to show its assessment of the virginity of the members of the training and test sets at different values of the overall RMS error. These data are displayed in Fig 6, where it may be seen both for the training (Fig 6(A)) and test sets (Fig 6(B)) that even at an RMS error of as much as 0·1 the network was able adequately to differentiate the virgins from the adulterated oils.

It is likely that the reason the network was able to generalise well (ie correctly to assess oils which it had not seen) was because the training set contained a wide assortment of virgin olive oils and a variety of varying amounts (50–300 ml per litre of mixed oil) of different adulterants. In this study a variety of cultivars was used,

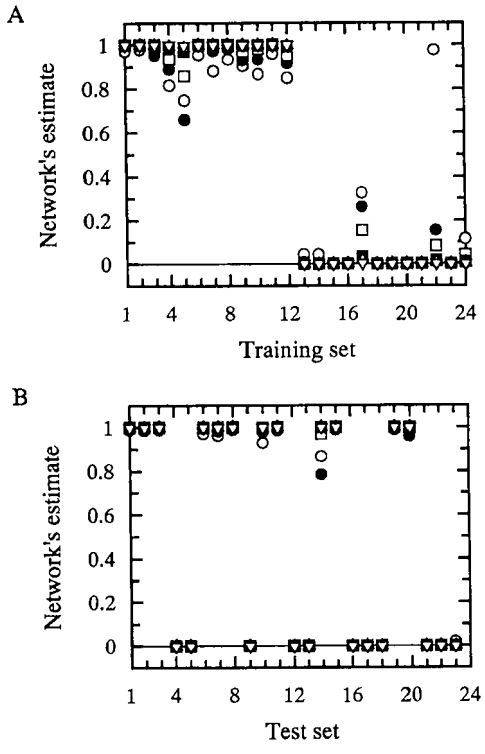

Fig 6. Results of the estimates of trained 150-8-1 ANN for both (A) the training set, and (B) the test set (unknowns) of oils. The ANN were interrogated after the RMS error values indicated (0·25–0·001) were reached.

the olives were of differing maturity, and the olives were processed in various ways (Table 1). Clearly, one might also vary the harvest time and have the appropriate exemplars in the training set. In any multivariate calibration method (Martens and Næs 1989), one should try and create a calibration model which encompasses all the variations in individual samples likely to be met. Thus, if all potentially variable factors are represented in the training set, then the resultant network should be more robust to any slight variations in the pyrolysis mass spectra of oils of a given type.

In other studies, similar feedforward ANN were set up using the standard back-propagation algorithm but containing no hidden layers. In this case, training was stopped after a final RMS error of 0·002 was attained; although there were obviously fewer nodes, the overall speed of learning was significantly impaired relative to those when a hidden layer was present. It was interesting to observe (Fig 7) that, although generalisation was also slower than when using one hidden layer, the networks were still able correctly to assess the virginity of the training set (Fig 7(A)). However, the status of the members of the test set (Fig 7(B)) was correctly assessed with only 82·6% accuracy compared to the 100% accuracy shown previously in Table 4, and the virgin olive oils *Urbino* (3), *Trieste* (8), *Cagliari* (10) and *Ancona* (20) were incorrectly estimated as being adulterated. This

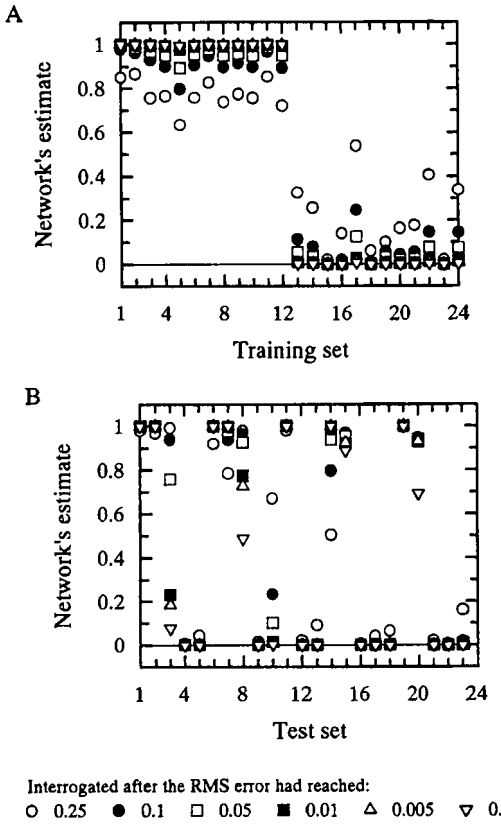

Fig 7. Results of the estimates of trained 150-0-1 ANN for both (A) the training set, and (B) the test set (unknowns) of oils. The ANN were interrogated after the RMS error values indicated (0·25–0·001) were reached.

suggests that the differences in the pyrolysis mass spectra of virgin olive oils and the adulterated olive oils could not adequately be fitted to a linear model in 150-dimensional space, consistent with the failure of CVA to effect such a discrimination (see above).

Further to investigate this point, other linear multivariate statistical models, namely principal component regression (PCR) and partial least squares (PLS) were implemented using the UNSCRAMBLER package (Martens and Næs 1989), as outlined by Martens and Martens (1986). These methods have been shown to estimate the concentration of components in binary and tertiary mixtures from their pyrolysis mass spectra to the same accuracy as could ANN (Goodacre R and Kell DB, in preparation). However, in the present case (coding virgins as 1, non-virgins as 0 as before) neither PCR nor PLS was able to form an adequate calibration model on the training set, and therefore could not sensibly be used to assay the virginity of unknown olive oils (data not displayed).

CONCLUDING REMARKS

The study reported here clearly shows that the combination of pyrolysis mass spectrometry and artificial neural networks was able accurately to assess the contamination of virgin olive oils adulterated with 50–500 ml corn, peanut, soya, sunflower oil, or rectified (sansa) olive oil per litre of mixed oil. PyMS is rapid (the typical sample time is less than 2 min) and automated, which allows approximately 300 samples to be analysed daily. Furthermore, after the initial outlay of some £50000 on the machinery, running costs are relatively cheap, typically about £1 per sample.

When more olive oils are analysed by PyMS, covering the numerous regions that produce olives, then the present approach should also allow one to identify the geographical origin of the olives that were used in the production of an oil. Further studies using pure components should also allow one to obtain chemical information (John 1992) on which compounds present in the adulterants were responsible for features in the pyrolysis mass spectra and thus for the elucidation via ANN of the contamination of virgin olive oils.

Finally, we would mention that the present work has concentrated on the exploitation of supervised learning via fully interconnected feedforward ANN of mass spectra that were subjected to no preprocessing nor feature extraction beyond normalisation to the total ion count. As well as exploiting other neural architectures, further work must also consider the many other well-developed approaches to pattern recognition that exist (e.g. Duda and Hart 1973; Breiman *et al* 1984; Jain and Dubes 1988; Massart *et al* 1988; Massart and Kaufman 1989; Fukunaga 1990; Schalkoff 1992).

In conclusion, we believe the reported combination of PyMS and ANN to be a powerful and novel approach to the assessment of food adulteration generally.

ACKNOWLEDGEMENTS

This work is supported by the Biotechnology Directorate of the UK SERC LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, ICI Biological Products and Fine Chemicals, and Neural Computer Sciences. The Italian Ministero del'Agricoltura e delle Foreste is thanked for financial support.

REFERENCES

Albi M A, Guttiérez F 1991 Study of the precision of an analytical taste panel for sensory evaluation of virgin olive oil. Establishment of criteria for the elimination of abnormal results. *J Sci Food Agric* **54** 255–267.

Aparicio R 1988 Characterisation of foods by inexact rules: the SEXIA expert system. *J Chemometrics* **3** 175–192.

Aparicio R, Ferreiro L, Leardi R, Forina M 1991a Building decision rules by chemometric analysis: application to olive oil. *Chemometrics Intell Lab Syst* **10** 349–358.

Aparicio R, Ferreiro L, Rodriguez J L 1991b Caracterización de alimentos combinado reglas de decisión relacionales y lineales. Una aplicación al aceite de oliva virgen de Málaga [Characterization of foods merging relational and lineal

decision rules. An application to virgin olive oil of Malaga]. *Grasas y Aceites* **42** 132–142.

Aparicio R, Graciani E, Ferreiro L 1992 Chemometric study of the Hilditch theory applied to virgin olive oil. *Anal Chim Acta* **259** 115–122.

Aries R E, Gutteridge C S, Ottley TW 1986 Evaluation of a low-cost, automated pyrolysis-mass spectrometer. *J Anal Appl Pyrol* **9** 81–98.

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

Breiman L, Friedman J H, Olshen R A, Stone C J 1984 *Classification and Regression Trees*. Wadworth International Group, Belmont, CA, USA.

Causton D R 1987 *A Biologist's Advanced Mathematics*. Allen & Unwin, London, UK, pp 48–72.

CEC (Commission of the European Communities) 1991 Commission Regulation No2568/91 of 11 July 1991 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. *Official J Eur Commun L248* **34** 1–83.

Chatfield C, Collins A J 1980 *Introduction to Multivariate Analysis*. Chapman & Hall, London, UK, pp 57–81.

Coors U 1991 Anwendung des Tocopherolmusteers zur Erkennung von Fett-und Ölvermischungen [Utilisation of the tocopherol pattern for recognition of fat and oil adulterations]. *Fat Sci Technol* **93** 519–526.

Duda R O, Hart P E 1973 *Pattern Classification and Scene Analysis*. John Wiley, New York, USA.

Eberhart R C, Dobbins R W 1990 *Neural Network PC Tools*. Academic Press, London, UK.

Eddib O, Nickless G 1987 Elucidation of olive oil classification by chemometrics. *Analyst* **112** 391–395.

Flury B, Riedwyl H 1988 *Multivariate Statistics: A Practical Approach*. Chapman & Hall, London, UK, pp 181–233.

Fukunaga K 1990 *Introduction to Statistical Pattern Recognition* (2nd edn). Academic Press, New York, USA.

Goodacre R, Berkeley R C W 1990 Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. *FEMS Microbiol Lett* **71** 133–138.

Goodacre R, Kell D B 1993 Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production. *Anal Chim Acta* **279** 17–26.

Goodacre R, Kell D B, Bianchi G 1992 Neural networks and olive oil. *Nature* **359** 594.

Goodacre R, Edmonds A N, Kell D B 1993 Quantitative analysis of the pyrolysis mass spectra of complex mixtures using artificial neural networks: application to amino acids in glycogen. *J Anal Appl Pyrol* **26** 93–114.

Gower J C 1971 A general coefficient of similarity and some of its properties. *Biometrics* **27** 857–874.

Grob K 1992 Hyphenated high-performance liquid chromatography capillary gas chromatography. *J Chromatogr* **626** 25–32.

Grob K, Artho A, Mariani C 1992 Determination of raffination of edible oils and fats by olefinic degradation products of sterols and squalene, using coupled LC–GC. *Fat Sci Technol* **94** 394–400.

Guth H, Grosch W 1991 A comparative study of the potent odorants of different virgin olive oils. *Fat Sci Technol* **93** 335–339.

Gutteridge C S 1987 Characterization of microorganisms by pyrolysis mass spectrometry. *Meth Microbiol* **19** 227–272.

Gutteridge C S, Vallis L, MacFie H J H 1985 Numerical methods in the classification of microorganisms by pyrolysis mass spectrometry. In: *Computer-Assisted Bacterial Systematics*, ed Goodfellow M, Jones D & Priest F. Academic Press, London, UK, pp 369–401.

Hertz J, Krogh A, Palmer R G 1991 *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, CA, USA.

Irwin W J 1982 *Analytical Pyrolysis: A Comprehensive Guide*. Marcel Dekker, New York, USA.

Jain A K, Dubes R 1988 *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, USA.

John P 1992 *Biosynthesis of the Major Crop Products*. John Wiley, Chichester, UK.

Kapoulas V M, Passaloglou-Emmanoulidou S 1981 Detection of adulteration of olive oil with seed oils by a combination of column and gas liquid chromatography. *J Amer Oil Chem Soc* **58** 694–697.

Kiritsakis A 1991 *Olive Oil*. American Oil Chemists Society, Champagin, IL, USA.

Kiritsakis A, Marakakis P 1987 Olive oil: a review. *Adv Food Res* **31** 453–482.

Kiritsakis A, Marakakis P 1991 Olive oil analysis. In: *Essential Oils and Waxes*, ed Linskens H F & Jackson J F. Spinger, Heidelberg, Germany, pp 1–20.

Kiritsakis A, Min D B 1989 Flavour Chemistry of Olive Oil. In: *Flavour Chemistry of Lipid Foods*, ed Min D B & Smouse T H. American Oil Chemists Society, Champaign, IL, USA, pp 196–221.

Krzanowski W J 1988 *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press, Oxford, UK, pp 369–400.

MacFie H J H, Gutteridge C S 1982 Comparative studies on some methods for handling quantitative data generated by pyrolysis. *J Anal Appl Pyrol* **4** 175–204.

MacFie H J H, Gutteridge C S, Norris J R 1978 Use of canonical variates in differentiation of bacteria by pyrolysis gas–liquid chromatography. *J Gen Microbiol* **104** 67–74.

Martens H, Næs T 1989 *Multivariate Calibration*. John Wiley and Sons, New York, USA.

Martens M, Martens H 1986 Partial least squares regression. In: *Statistical Procedure in Food Research*, ed Piggot J R. Elsevier Applied Science, New York, USA, pp 293–359.

Masana L, Camprubi M, Sarda P, Sola R, Joven J, Turner P R 1991 The Mediterranean-type diet: is there a need for further modification? *Am J Clin Nutr* **53** 886–889.

Massart D L, Kaufman L 1989 *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*. Krieger, Malabar, FL, USA.

Massart D L, Vandeginste B G M, Deming S N, Michotte Y, Kaufman L 1988 *Chemometrics: A Textbook*. Elsevier, Amsterdam, The Netherlands.

Meuzelaar H L C, Haverkamp J, Hileman F D 1982 *Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials*. Elsevier, Amsterdam, The Netherlands.

Nelder J A 1979 *Genstat Reference Manual*. Scientific and Social Service Program Library, University of Edinburgh, Edinburgh, UK.

Norman K N T 1991 Determination of tetrachloroethylene in olive oil by automated headspace gas chromatography. *Food Addit Contamin* **8** 51–516.

Pallotta U 1976 Analytic problems in the ascertainment of olive oil genuineness. In: *Lipids, Vol 2, Technology*, ed Paoletti R, Jacini G & Porcellati R. Raven Press, New York, USA, pp 387–393.

Passaloglou-Emmanoulidou S 1990 A comparative study of UV spectrophotometric methods for detection of olive oil adulteration by refined oils. *Lebensm Unters Forsch* **191** 132–134.

Rumelhart D E, McClelland J L, The PDP Research Group

1986 *Parallel Distributed Processing. Experiments in the Microstructure of Cognition.* MIT Press, Cambridge, MA, USA.

Santinelli F, Damiani P, Christie W W 1992 The triacylglycerol structure of olive oil determined by silver ion high-performance liquid chromatography in combination with stereospecific analysis. *J Am Oil Chem Soc* **69** 552–556.

Sato T, Kawano S, Iwamoto M 1991 Near infrared spectral patterns of fatty acid analysis from fats and oils. *J Am Oil Chem Soc* **68** 827–833.

Schalkoff R 1992 *Pattern Recognition: Statistical, Structural and Neural Approaches.* Wiley, New York, USA.

Simpson P K 1990 *Artificial Neural Systems.* Pergamon Press, Oxford, UK.

Ward A C, Freeman R, Sisson P R, Barer M, Lightfoot N F 1993 A highly discriminatory method for the direct comparison of two closely related bacteria populations by pyrolysis mass spectrometry. *Zbl Bakt Hyg (J Med Micro Hyg)* (in press).

Wasserman P D 1989 *Neural Computing: Theory and Practice.* Van Nostrand Reinhold, New York, USA.

Wessels H 1992 Olivenölkategorien und ihre Merkmale—Regelungen der Europäischen Gemeinschaft [Olive oil types and their characteristics--regulations of the European Community]. *Fat Sci Technol* **94** 130–137.

Windig W, Haverkamp J, Kistemaker P G 1983 Interpretation of a set of pyrolysis mass spectra by discriminant analysis and graphical rotation. *Anal Chem* **55** 387–391.