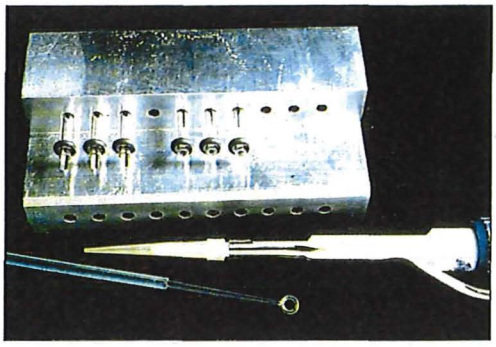
# Food adulteration exposed by neural networks



*Sample block with sample tubes in place.*

New methods to detect food adulteration are urgently needed. **Royston Goodacre, Douglas Kell** and **Giorgio Bianchi** describe a rapid automated approach


Neue Methoden, um Lebensmittelfälschung nachzuweisen, werden dringend benötigt. **Royston Goodacre, Douglas Kell** und **Giorgio Bianchi** beschreiben einen rasch ausführbaren apparativen Ansatz.

De nouvelles méthodes pour détecter l'altération des aliments sont nécessaires d'urgence. **Royston Goodacre, Douglas Kell** et **Giorgio Bianchi** décrivent une approche rapide automatisée.


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 one per cent are known as 'extra virgin', whilst oils with good flavour but greater acidity are graded as 'fine' or 'semi-fine'. Lower grades may be known as 'refined olive oil' or 'olive oil'. Olive oil is thought to contribute significantly to the nutritional and health benefits of Mediterranean-type diets and, unusually for a vegetable oil, the flavour of unrefined oil is considered to be best. Extra virgin olive oil therefore attracts a premium price, and these and other properties mean that there is a temptation to adulterate it with other seed oils.

Pyrolysis is the thermal degradation of complex molecules in a vacuum, causing their reproducible cleavage to smaller, volatile fragments called pyrolysate. Curie-point pyrolysis is a particularly reproducible and straightforward version of the technique in which the sample, placed onto an appropriate metal, is rapidly heated (half a second is typical) to the Curie point of the metal, which may itself be chosen (358, 480, 510, 530, 610 and 770°C are common temperatures). A mass spectrometer can then be used to separate the components of the pyrolysate on the basis of their mass-to-charge ratio ($m/z$) so as to produce a pyrolysis mass spectrum. This can then be used as a 'chemical profile' or fingerprint of the complex material analysed; the combined technique is known as pyrolysis mass spectrometry.

The last three years have seen some very exciting advances in the field of pyrolysis mass spectrometry, several of which have stemmed from advances in chemometrics; that is, the application of novel statistical methods to multivariate analytical chemical data, and in particular the use of supervised learning methods such as those based on (artificial) neural networks.

## Learning curve

Neural networks are a well-known means of uncovering complex, non-linear relationships in multivariate data, and can be considered as collections of very simple computational units. They take numerical inputs (the pyrolysis mass spectral training data) and transform them into desired outputs (whether the oil is of virgin quality or adulterated), here encoded as 0 or 1. The input and output nodes may be connected to the external world and to other nodes within the

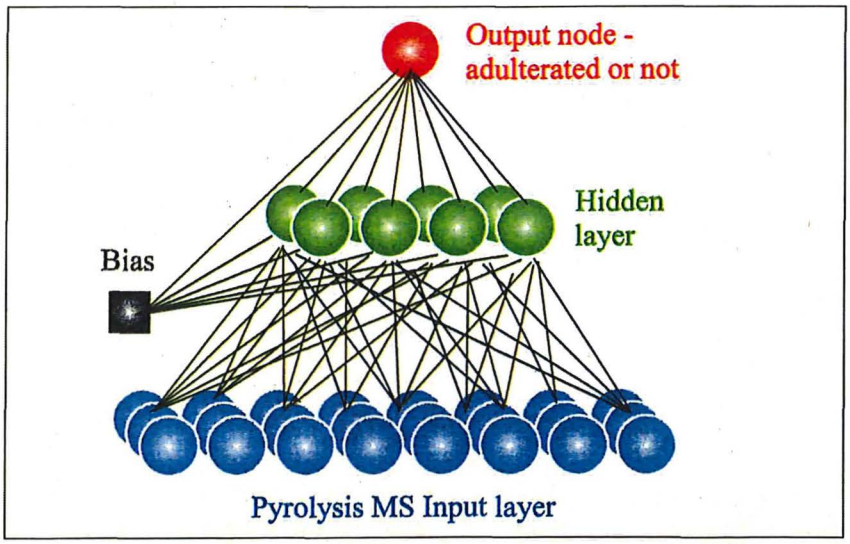



**Figure 1** *A fully-interconnected feedforward neural network consisting of 24 input, eight hidden, and one output node(s). Each node in the hidden layer is connected to all the nodes in the input and output layers. The actual number of nodes in the pyrolysis mass spectrometry input layer was 150 (one for each mass from 51 to 200).*

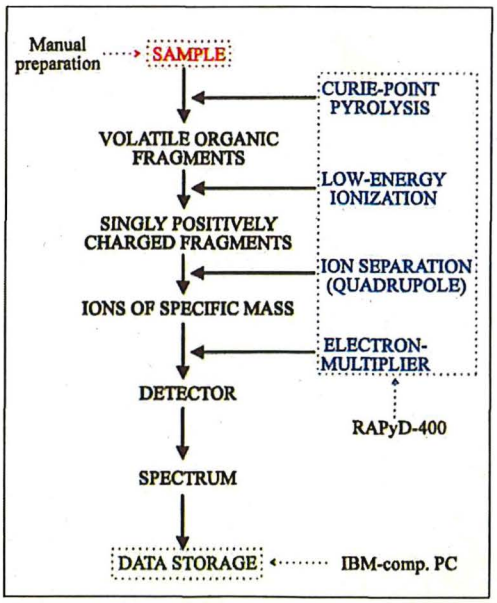



**Figure 2** *Flow diagram of the main steps in pyrolysis mass spectrometry.*

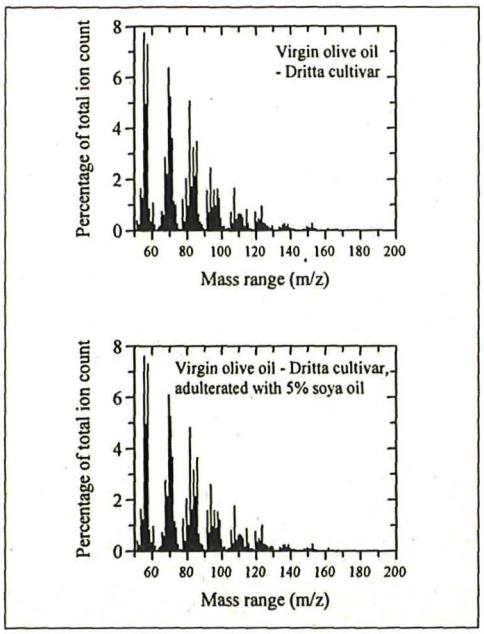



**Figure 3** *Normalised pyrolysis mass spectra of an extra virgin olive oil, Dritta cultivar, and the same oil adulterated with five per cent soya oil.*

network (for a diagrammatic representation see figure 1). The way in which each node transforms its input depends on the connection weights (or connection strength) and bias of the node, which can be modified. The output of each node to another node or the external world then depends on both its weight strength and bias and on the weighted sum of all its inputs, which are then transformed by a (normally non-linear) weighting function referred to as its activation or squashing function. The great power of neural networks stems from the fact that it is possible to train them. One can acquire sets of multivariate data (such as pyrolysis mass spectra) from standard materials of known identities and train neural networks using these identities as the desired outputs. Training is effected by continually presenting the networks with the known inputs and outputs and modifying the connection weights between the individual nodes and the biases, typically according to some back-propagation algorithm, until the output nodes of the network match the desired outputs to a stated degree of accuracy. The trained networks may then be exposed to unknown inputs (such as spectra); they will immediately provide a global optimum best fit to the outputs.

## Double blind

Two sets of samples were prepared in Italy by Giorgio Bianchi, each consisting of 12 samples of various extra virgin olive oils plus 12 samples variously adulterated with 5-50 per cent of soya, sunflower, peanut, corn or rectified olive oils. The experiment was performed double-blind, such that the identities of the second set were not known to any of the experimenters.

The pyrolysis mass spectrometry analysis method is very simple (figure 2); sample preparation is quick and easy and involves the application of a few microlitres of the oils to iron-nickel foils. Curie-point pyrolysis MS was performed in Douglas Kell's laboratory at 530°C; two typical spectra are shown in figure 3, where it is obvious that it is not possible to distinguish them by eye.

After the collection of these pyrolysis mass spectra, conventional multivariate statistical methods were used to establish the relationships between the members of the first set of oils. Canonical variates analysis and cluster analysis (dendrograms) were used to reduce the dimensionality of these data from the 150 masses to only two or three latent variables whilst preserving most of the variance. They are thus excellent methods for highlighting the differences between the oils (figure 4). These plots clearly indicate that the major source of variation was due to differences between the cultivars, rather than to adulteration. The salient point here is that these classical unsupervised methods, routinely used by pyrolysis mass spectrometry experts, could not be used to assess the virginity or even purity of these olive oils.

## Better separation

Neural network analyses were then carried out using a user-friendly, neural network simulation program. Neural networks consisting of an input layer of the 150 normalised averaged ion intensities from the pyrolysis mass spectra of the first set of oils (see above) with mass range 51-200, and one hidden layer of eight nodes, were trained using the standard back-propagation algorithm, coding virgin oils as 1, non-virgins as 0 (figure 1). The effectiveness of training was expressed in terms of the root mean square (RMS) error between the actual and the desired outputs over the entire training set. When training ended (an RMS error of 0.1 per cent was attained), the trained network was interrogated with the normalised averaged ion intensities of the pyrolysis mass spectra from the second unknown set of 23 oils (one was damaged in the

*The olive: pure and unrefined.*

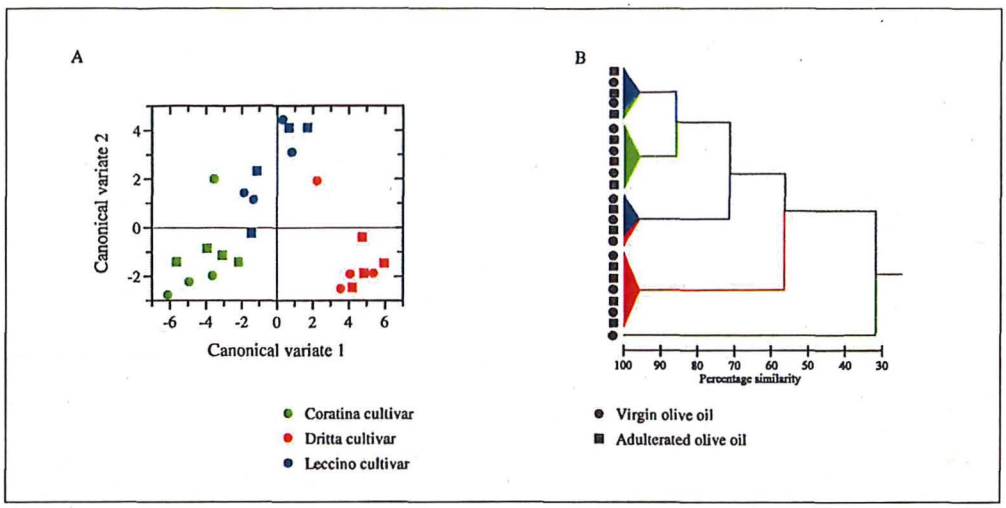



**Figure 4** *Canonical variates analysis plot (A) and dendrogram (B) representing the relationships between the first set of 24 oils (the training set) based on pyrolysis mass spectrometry data analysed by GENSTAT.*

post). When the code was broken, it was found that the neural network had correctly assessed all the oils (figure 5). 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).

The results of this study were first reported in *Nature*, and it was the first application of artificial neural networks for the successful identification of biological samples from their pyrolysis mass spectra. This investigation permitted the rapid and exquisitely sensitive assessment of the adulteration of extra virgin olive oils with various seed oils, a task previously both labour intensive and difficult.

Previously, workers have relied on multivariate methods (such as CVA and dendrograms) to attempt to cluster pyrolysis mass spectra, but, since they rely on unsupervised learning and linear (orthogonal) transformations of the raw multivariate data, they cannot provide the truly best analytical discriminations, and thus have achieved only limited success. Neural nets, however, have proved very advantageous in the analysis of pyrolysis mass spectrometry data (as this study clearly shows), and these 'intelligent systems' have now been adopted by other users of pyrolysis mass spectrometry.

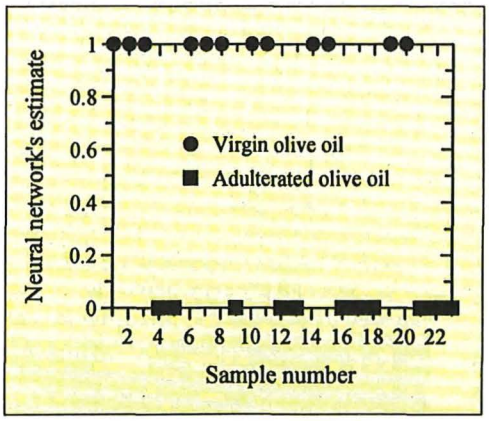



**Figure 5** *Results of the estimates of neural networks trained on the 24 oils in the training set and applied to the 23 oils in the unseen test set. Neural networks were trained using the standard back propagation algorithm, to 0.1 per cent RMS error. Filled circles represent spectra from virgin olive oils and filled squares those from oils which had been adulterated. Each oil was assessed correctly.*

## Rapid and automatic

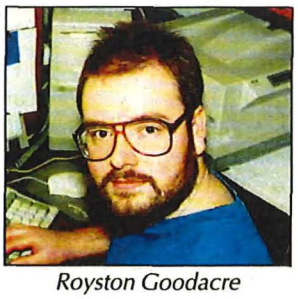
Pyrolysis mass spectrometry has the major advantages of speed (the typical sample time is less than two minutes) and automation, which allows approximately 300 samples to be analysed daily. When more olive oils are analysed by pyrolysis mass spectrometry (currently under joint investigation in the UK by Douglas Kell and in Italy by Giogio Bianchi), covering the numerous olive producing regions, the approach should allow identification of the geographical origin of the olives used in the production of an oil (the initial results look very promising). In microbial biotechnology we have been shown that pyrolysis mass spectrometry and neural networks can be used to quantify fermenter broths directly (without sample preprocessing) for the metabolites of interest. This approach could also be applied to foodstuffs not only to give a qualitative yes/no answer for adulteration but also the level of contamination if present (initial results look promising).

Since any biological material can be pyrolysed in this way, the combination of pyrolysis mass spectrometry and neural networks constitutes a rapid, powerful and novel approach to the assessment of food adulteration.

• More information regarding pyrolysis MS and neural nets can be found on the **World Wide Web** at http://gepasi.dbs.aber. ac.uk/roy/rgcv.htm.

We would like to thank Tony Pugh for photography. The work reported here was initially funded by the Chemicals and Pharmaceuticals Directorate of the UK BBSRC, under the terms of the LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, Neural Computer Sciences and Zeneca Bioproducts plc.

***'Pyrolysis mass spectrometry has the advantages of speed and automation'***



*Royston Goodacre*

**Royston Goodacre** is funded by the Wellcome Trust to research chemometric methods for pyrolysis MS in biological systems at the University of Wales, where **Douglas Kell** holds a personal chair. **Giorgio Bianchi** is the director of Istituto Sperimentale per la Elaiotecnica, Italy