Zbl. Bakt. 284, 516–539 (1996)

Quantitative Analysis of Multivariate Data Using Artificial Neural Networks: A Tutorial Review and Applications to the Deconvolution of Pyrolysis Mass Spectra

ROYSTON GOODACRE, MARK J. NEAL, and DOUGLAS B. KELL

Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed, UK

Received April 29, 1994 · Accepted January 11, 1996

Summary

The implementation of artificial neural networks (ANNs) to the analysis of multivariate data is reviewed, with particular reference to the analysis of pyrolysis mass spectra. The need for and benefits of multivariate data analysis are explained followed by a discussion of ANNs and their optimisation. Finally, an example of the use of ANNs for the quantitative deconvolution of the pyrolysis mass spectra of *Staphylococcus aureus* mixed with *Escherichia coli* is demonstrated.

Introduction

Multivariate data consist of the results of observations of many different characters (variables) for a number of individuals (objects) (73, 74). Each variable may be regarded as constituting a different dimension, such that if there are n variables each object may be said to reside at a unique position in an abstract entity referred to as n-dimensional hyperspace. This hyperspace is necessarily difficult to visualise, and the underlying theme of multivariate analysis (MVA) is thus *simplification* (22) or dimensionality reduction, which usually means that we want to summarise a large body of data by means of *relatively* few parameters, preferably the two or three which lend themselves to graphical display, with minimal loss of information.

In the case of spectroscopy, variables are usually represented by properties such as the absorbance at particular wavelengths (e.g. ref. 73). A spectral technique which seems ideally suited to analysis by multivariate methods is pyrolysis mass spectrometry (PyMS). Pyrolysis is the thermal degradation of complex molecules in a vacuum which causes their cleavage to smaller, volatile fragments separable by a mass spectrometer (80) on the basis of their mass-to-charge ratio (m/z). Almost all biological materials will produce pyrolytic degradation products such as methane, ammonia, water, methanol and H_2S, whose $m/z < 50$, and fragments with $m/z > 200$ are rarely analytically important for bacterial discrimination (8) unless very special conditions are em-

ployed (98); the analytically useful multivariate data are then constituted by a set of 150 normalised intensities versus m/z in the range 51 to 200.

Conventionally, at least within microbiology and biotechnology, because PyMS has been used as a taxonomic aid (8, 43, 55, 64, 71, 80), the reduction of the multivariate data generated by the PyMS system (and indeed of those generated by other arrays of sensors; e.g. gas chromatography (70), spectroscopic methods (74), and nuclear magnetic resonance (67) is normally carried out using principal components analysis (PCA; 21, 22, 30, 34, 56, 74). This is a well-known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance, and so is an excellent technique for observing the *natural* relationships between multivariate samples. Whilst it does not take account of any groupings in the data, neither does it require that the populations be normally distributed, i.e. it is a non-parametric method (in addition, it permits the loadings of each of the m/z ratios on the principal components to be determined, and thus the extraction of at least some chemically significant information). The closely-related canonical variates analysis (CVA) technique then separates the samples into groups on the basis of the principal components and some *a priori* knowledge of the appropriate number of groupings (70, 109). Provided that the data set contains "standards" (i.e. type or centro-strains) it is evident that one can establish the closeness of any unknown samples to a known organism, and thus effect the identification of the former, a technique termed 'operational fingerprinting' by *Meuzelaar* et al. (80). An excellent example of the discriminatory power of the approach is the demonstration (44) that one can use it to distinguish *E. coli* strains which differ only in the presence or absence of single antibiotic-resistance plasmids. However, only rarely has the chemical basis for any such differences either been sought or found.

Analyses of the above type fall into the category of "unsupervised learning", in which the relevant multivariate algorithms seek "clusters" in the data (30). This allows the investigator to group objects together on the basis of their perceived closeness in the *n*-dimensional hyperspace referred to above. Such methods, then, although in some sense quantitative, are better seen as qualitative since their chief purpose is merely to *distinguish* objects or populations. More recently, a variety of related but much more powerful methods, most often referred to within the framework of chemometrics, have been applied to the "supervised" analysis of multivariate data. In these methods, of which multiple linear regression (MLR), partial least squares regression (PLS) and principal components regression (PCR) are the most widely used, one seeks to relate the multivariate spectral inputs to the concentrations of target determinands, i.e. to generate a quantitative analysis, essentially *via* suitable types of multidimensional curve fitting or regression analysis (14, 15, 17, 74–76, 79). Although non-linear versions of these techniques are increasingly available (e.g. 35, 63, 67, 100, 110, 111, 112), the usual implementations of these methods are linear in scope. A related approach to chemometrics, which is inherently nonlinear, however, is the use of (artificial) neural networks (ANNs) (see below).

For a given analytical system there are some patterns (e.g. mass spectra) which have desired responses which are known (i.e. the concentration of target determinands). These two types of data (the representation of the objects and their responses in the system) form pairs which for the present purpose are called inputs and targets. The goal of supervised learning is to find a *model* or *mapping* that will correctly associate the inputs with the targets (Fig. 1).

Thus the basic idea in these supervised learning techniques is that there are minimally 4 data sets to be studied, as follows. The "training data" consist of (i) a matrix of *s*

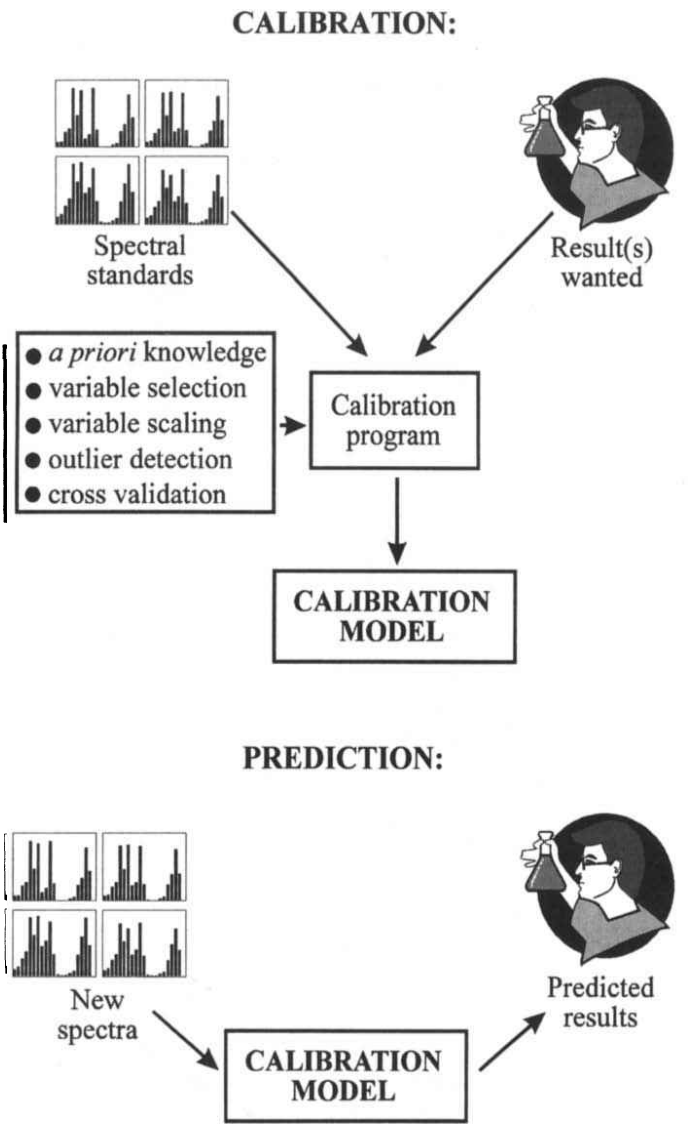


Fig. 1. The process of multivariate calibration *via* supervised learning consists of two stages. Calibration: establishing a CALIBRATION MODEL for later prediction of "results" (e. g., amount of determinand) from "spectra" by matching spectral standards and the known results wanted from a set of calibration samples (the training set) in some sort of supervised learning calibration program; this may be by using a neural network simulation, or with a program that performs multiple linear regression, partial least squares or principal component regression. Some background knowledge may be used in the formation of the model such as *a priori* knowledge, variable selection and scaling, outlier detection and cross validation.
Prediction: converting instrumental data for new samples into predictions of wanted results (e. g., in terms of determinand concentration) using the above previously established CALIBRATION MODEL in the computer program.

rows and *n* columns in which *s* is the number of objects and *n* the number of variables (these may be the absorbance at particular wavelengths, or in the present case the normalised ion intensities at a particular m/z; Fig. 1), and (ii) a second matrix, again consisting of *s* rows and typically 1 or two columns, in which the columns represent the variable(s) whose value(s) it is desired to know (these are the result(s) wanted; Fig. 1) and which for the training set have actually been determined by some existing, "benchmark" method. This variable may be the concentration of a target determinand, and is always paired with the patterns in the same row in (i). The "test data" also consist of two matrices, (iii) and (iv), corresponding to those in (i) and (ii) above, but the test set contains different objects. As the name suggests, this second pair is used to test the accuracy of the system; alternatively they may be used to cross-validate the model. That is to say, after construction of the model using the training set (i, ii) the test data (iii) (these may be new spectra; Fig. 1) are then "passed" through the calibration model so as to obtained the model's prediction of results. These may then be compared with the known, expected responses (iv).

As in all other data analysis techniques, these supervised learning methods are not immune from sensitivity to badly chosen initial data (113). Therefore the exemplars for the training set *must* be carefully chosen; the golden rule is "garbage in – garbage out". An excellent example of an unrepresentative training set was discussed some time ago on the BBC television programme *Horizon;* a neural network was trained to attempt to distinguish tanks from trees. Pictures were taken of forest scenes lacking military hardware and of similar but perhaps less bucolic landscapes which also contained more-or-less camouflaged battle tanks. A neural network was trained with these input data and found to differentiate most successfully between tanks and trees. However, when a new set of pictures was analysed by the network, it failed to distinguish the tanks from the trees. After further investigation, it was found that the first set of pictures containing tanks had been taken on a sunny day whilst those containing no tanks were obtained when it was overcast. The neural network had therefore thus learned simply to recognise the weather! We can conclude from this that the training and tests sets should be carefully selected to contain representative exemplars encompassing the appropriate variance over all relevant properties for the problem at hand.

It is also imperative that the objects fill the sample space. If a neural net is trained with samples in the concentration range from 0 to 50% it is unlikely to give a accurate estimates for samples whose concentrations are greater than 50%, that is to say the network is unable to *extrapolate.* Furthermore for the network to provide good *interpolation* it needs to be trained with a number of samples covering the desired concentration range (47).

Artificial neural networks

ANNs are an increasingly well-known means of uncovering complex, non-linear relationships in multivariate data, whilst still being able to map the linearities. ANNs can be considered as collections of very simple "computational units" which can take a numerical input and transform it, usually *via* summation, into an output (see 1, 2, 6, 20, 24, 28, 40, 59, 60, 66, 78, 87, 88, 93, 97, 102 and 103 for excellent introductions; and 4, 7, 9–12, 18, 23, 25, 29, 37, 42, 45, 47, 52, 65, 69, 77, 85, 89, 92, 96, 99, 105 and 114 for applications in analytical chemistry and microbiology).

The relevant principle of "supervised" learning in ANNs is that, as with the multivariate calibration described above, they take numerical inputs (the training data, which are usually multivariate) and transform them into "desired" (known, predetermined) outputs. The input and output nodes may be connected to the "external world" and to other nodes within the network (for a diagrammatic representation see Fig. 2). The way in which each node transforms its input depends on the so-called "connection weights" (or "connection strength") and "bias" of the node, which are modifiable. The output of each node to another node or the external world then depends on both its

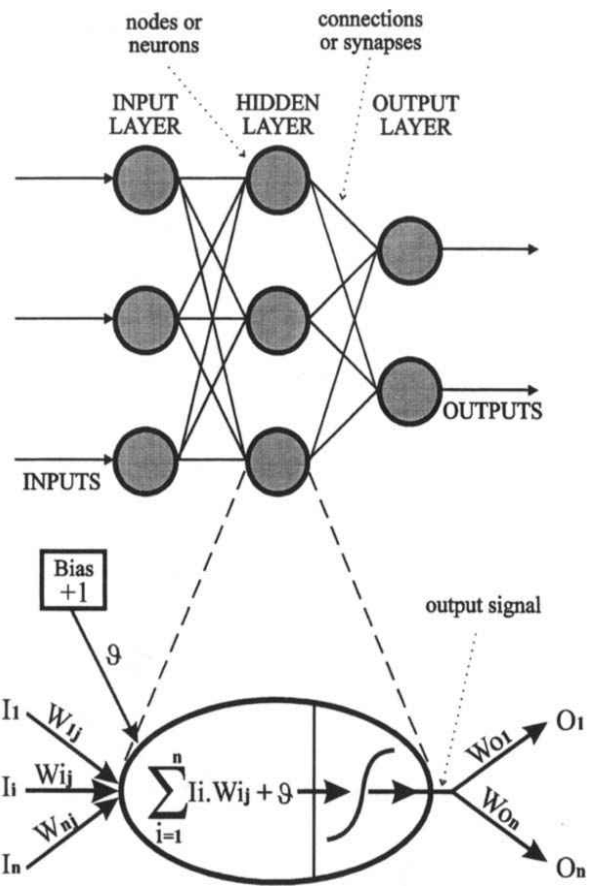


Fig. 2. A neural network consisting of 3 inputs (data for PyMS actually consisted of 150 inputs/masses) and 2 outputs (for the PyMS study this was a single node which represented the %*S. aureus*) connected to each other by 1 hidden layer consisting of 3 nodes (for PyMS this was actually 8). In the architecture shown, adjacent layers of the network are fully interconnected although other architectures are possible. One of the nodes in the hidden layer is given in more detail showing the information processing by node. An individual node sums its input (the Σ function) from nodes in the previous layer, including the bias (ϑ), transforms them *via* a "sigmoidal" squashing function, and outputs them to the next node to which it is linked *via* a connection weight.

weight strength, 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, threshold or squashing function. As with other supervised learning methods, the great power of neural networks stems from the fact that it is possible to "train" them. One can acquire sets of multivariate data (which may be pyrolysis mass spectra) from standard materials of known identities and train ANNs 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 kind of back-propagation algorithm (93), until the output nodes of the network match the desired outputs to a stated degree of accuracy. The trained ANNs may then be exposed to unknown inputs (i. e. spectra) and they and will immediately provide the globally optimal best fit to the outputs.

Provided the ANN gives the correct results for "unknown" data (i. e. data unseen by the calibration system, but known by the operator) it may be said to have "generalised". The operator can now be confident that when *genuine* unknown spectra are passed through the neural network, the predictions will be accurate and precise, provided of course that these spectra bear some relationship to the training set.

The so-called *false-sample problem* often poses a major stumbling block for the routine application of these numerical techniques for the analysis of multivariate data (27). This occurs, for example, when an operator makes a calibration model to quantify substance *a* in *b*, but then tests the model on (say) *c* mixed with *b*. It should be obvious that the neural network calibration model is now being asked to extrapolate beyond the knowledge domain on which it was trained, and will thus fail to give an accurate prediction of the determinand. As well as exploiting appropriate software methods (e. g. statistical ones) to detect whether a test sample falls within the domain of validity of the training set, it is also possible to include suitable "false" samples and corresponding "dummy" output variables in the training set; this practice should allow the neural network to detect widely different samples (53). In this sense, then, analysis with ANNs of this type can fairly be regarded within the framework of the multivariate calibration approach as outlined in Figure 1.

The following summarises the fundamental nomenclature and the rudimentary mathematical concepts that are used to describe and analyse ANN processing.

The processing units

ANNs employ processing nodes (neurons or units), connected using abstract interconnections (connections or synapses). Connections each have an associated real value, termed the weight (w_i), that scales the input (i_i) passing through them (Fig. 2); this also includes the bias (ϑ), which also has a modifiable weight. Nodes sum the signals feeding to them *(Net)*:

$$Net = i_1w_1 + i_2w_2 + i_3w_3 + \ldots + i_iw_i + \ldots + i_nw_n =$$

$$= \sum_{i=1}^{n} i_iw_i + \vartheta$$

Activation functions

The sum of the scaled inputs and the node's bias, are then scaled to lie between 0 and +1 (or sometimes between −1 and +1) by an activation function to give the nodes

output *(Out);* this scaling is typically achieved using a logistic "squashing" (or sigmoidal) function:

$$Out = \frac{1}{(1 + \exp^{-Net})}$$

It is widely thought that a continuously differentiable sigmoidal squashing function (93, 97, 102) is most appropriate, although some advantage may accrue to the use of a linear activation function on the output nodes.

Neural network topology

ANN topologies, or architectures, are formed by organising nodes into layers (also termed fields or slabs) and linking these layers of neurons with modifiable weighted interconnections. A diagrammatic representation of a neural network consisting of 3 inputs and 2 outputs connected to each other by 1 hidden layer consisting of 3 nodes is shown in Figure 2. In the fully connected topology shown each of the 3 nodes in the input layer is connected to the 3 in the hidden layer, by 9 connection weights, which in turn are connected to the 2 output nodes, by a further 6 connection weights. In addition, there is also a bias (extra node), which always has an activation level of +1, which is connected to nodes in the hidden and output layers (but not the input layer) *via* modifiable weighted connections (5 in the example shown in Fig. 2). Such an architecture can be written as a 3-3-2 ANN, and is commonly referred to as a fully interconnected feedforward multilayer perceptron.

Other architectures are possible such as direct linear feed through, where in addition to the above the nodes in the input layer is also connected directly to the output layer. There are many other topologies where not all the nodes between layers are connected, the connections may be chosen or random.

It is known from the statistical literature that better predictions can often be obtained when only the most relevant input variables are considered (81, 90, 95). Therefore neural networks that prune larger networks are an active area of study (e. g., 33, 57, 18, 68, 84, 91, 104). It is also possible to grow neural networks from small ones (16, 32, 36, 82, 83).

The most widely used neural network topology is one that is fully connected (Fig. 2) and where the input and output nodes are connected *via* a single hidden layer. One reason that this architecture is so attractive for the quantitative analysis of multivariate spectral data is that it has been shown mathematically (26, 39, 61, 62, 108) that a neural network consisting of only one hidden layer, with an arbitrarily large number of nodes, can learn any, arbitrary (and hence non-linear) mapping of a continuous function to an arbitrary degree of accuracy. It is the presence of this hidden layer which permits the nonlinear mapping, since similar networks lacking a hidden layer can only effect a multivariate *linear* mapping (93). In addition, such ANNs are widely considered to be relatively robust to noisy data, such as those which may be generated by mass spectrometry or gas chromatography.

The above 3-3-2 architecture is rather simple and the question arises "How does one choose the number of nodes in the hidden layer?" For pyrolysis mass spectra there are 150 m/z values and thus 150 nodes in the input layer. It is important not to have too many nodes in the hidden layer(s) because this may allow the neural network to learn by example only and not to generalise (5). We have found that a suitable rule of thumb is that good generalisation often comes from using a number of nodes in the hidden

layer that approximates to the natural logarithm of the number of nodes in the input layer. Thus with 150 input nodes this equates to 8. For a single determinand (output node) this architecture would be represented as a 150-8-1 ANN. However, it is noteworthy that during our work we have found that fewer nodes in the hidden layer (indeed often even zero) can be used to quantify PyMS data successfully.

Preparation of data

As mentioned previously it is of paramount importance to have the correct exemplars in the training and test sets. It is necessary also that these samples fill the sample space. Neural networks are very bad at extrapolating and to ensure good interpolation they need to be trained with samples equally-spaced over the desired concentration range; for a range of binary mixtures, it has been found, using PyMS, that 11 samples spaced every 10% will allow the network to generalise well (47).

Once the data used to train the ANN are collected, and the concentrations of the determinand ascertained using "wet chemistry", they are split into two data sets in which one half of the pair are the inputs (stimulus) of the network (for present purposes this would be the pyrolysis mass spectra normalised to the total ion count) and the other are the known or expected responses (i. e. the concentrations of the determinand[s]).

At this stage it is important to determine whether the training and/or test sets contain any outlying samples; these are usually detected using principal components analysis. It is known that if outliers are included in the construction of a calibration model then inaccuracies in the predictions from new multivariate data using the model are likely to occur (74).

Some of the m/z values may be omitted from the training data, a practice termed "pruning". It is unlikely however that the operator will know *a priori* which masses to remove so at least for typical back-propagation neural networks it is best to start with the intensities from all 150 masses.

The input and output nodes are next normalised between 0 and +1. It has been found (47, 52) that network generalisation is improved if the output layer is scaled to exploit less than the full range of the normalised scale, and the optimum for PyMS appears typically to be between +0.1 and +0.9. We have also found on occasion that scaling input nodes *individually*, i. e. over their own range rather than over the whole range encompassed in the entire input space, has improved learning rates dramatically (>100 fold) (52, 86).

Training the neural network

The first step is to choose the algorithm to be used for training the ANN. There are numerous algorithms available, and indeed the list of new ones expands continually. The most commonly employed is the standard back-propagation (BP) algorithm (93, 106, 107). Other algorithms which we have exploited include stochastic back-propagation, also termed learning by pattern (78), quick propagation (31), and Weigend weight elimination (104). The following describes training ANNs with the standard back-propagation algorithm.

Before training commences the connection weights are set to small random values, including the weights connecting the bias to the hidden and output layers (102). Next 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 (93, 97, 102).

For any given ANN, set of connection weight values, and training set there exists an overall RMS error of prediction. An error surface can be constructed by using one dimension in a multidimensional space to represent each connection weight, and an additional one for the RMS error. The BP algorithm performs gradient descent on this error surface by modifying each weight in proportion to the gradient of the surface at its location. Two constants, *learning rate* and *momentum*, control this process; for standard BP a learning rate of 0.1 and a momentum of 0.9 often give the best results. Learning rate scales the magnitude of the step down the error surface taken after each complete calculation in the network (epoch), and momentum acts like a low pass filter, smoothing out progress over small bumps in the error surface by remembering the previous weight change.

It is known that gradient descent can sometimes cause networks to get "stuck" in a depression in the error surface should such a depression exist. These are termed "local minima" (97, 102). However, it has been found empirically that local minima are seldom problematic for larger networks dealing with problems such as those presently under discussion, since the chance of encountering a multidimensional depression that is bounded in every dimension is relatively small.

One complete calculation in the network is called an epoch. This is equivalent to one complete pass through all the training data, calculating for *each member* of the training set. For a 150-8-1 ANN topology, trained with the standard back-propagation algorithm (93, 106, 107), where the weights are updated after all the training data are seen, one epoch represents 1217 connection weight updatings (1200 weights between the input and hidden layer (150×8), 8 weights between the hidden layer and the output node (8×1), and 9 weights from the bias to the 8 nodes in the hidden layer and the single output node) and a recalculation of the root mean squared (RMS) error between the true and desired outputs over the entire training set. In contrast, one epoch for an ANN trained using the stochastic back-propagation (78) would also incluce the weight updatings after *each* of the training pairs is passed through the neural network.

Stability and convergence

During training a plot of the RMS error versus the number of epochs represents the "learning curve", and may be used to estimate the extent of training. Training may be said to have finished when the network has found the lowest RMS error. Provided the network has not become stuck in a local minimum, this point is refered to as the global minimum on the error surface.

It is known (45, 52, 93, 102) that neural networks can become over-trained. An over-trained neural network has usually learnt perfectly the stimulus pattern it has seen but can not give accurate predictions for unseen stimuli, and it is no longer able to generalise. For ANNs accurately to learn the concentrations of determinands in biological systems networks must obviously be trained to the correct point. It is therefore imperative that ANNs should be trained several (perhaps many) times to ascertain whether they converge reproducibly. In addition, a superior method to reveal when the neural network will best generalise is to use the (or a) test set to cross-validate the model. During training the network may be interrogated with new stimulus patterns so as to

generate outputs at the output node. The stimuli used may be pyrolysis mass spectra, whose determinand concentration is known to the operator but not to the neural network. The error between the output seen and that expected may then be calculated thus allowing a second learning curve for the test set to be drawn. Training is stopped when the RMS error on the test or cross-validation data is lowest.

Table 1. Some parameters which one may vary during the production of a feedforward back-propagation neural network calibration model to improve learning/convergence and generalisation

1.	Number of hidden layers:
	One is thought sufficient for most problems. More give a big increase in computational load.
2.	Number of nodes in hidden layer:
	Rule of thumb says *ln* (number of inputs).
3.	Architecture:
	Fully interconnected feedforward net is most common. Many others exist such as adaptive resonance theory, Boltzmann machine, direct linear feedthrough, Hopfield networks, Kohonen networks.
4.	Number of exemplars in training set:
	Need enough to fill parameter space and to allow generalisation. When fewer are used then the network can "store" all the knowledge.
5.	Number of input variables:
	Those that do not contribute positively to discrimination may impair generalisation and are best removed by pruning the input data.
6.	Scaling of input and output variables:
	Individual scaling on inputs improves learning speed dramatically. There is a need to leave headroom, especially on the output layer.
7.	Updating algorithm:
	There are many variants on the original "vanilla-flavoured" back-propagation (BP) – most of which give small but worthwhile improvements. Others include radial basis functions, quick-prop, stochastic BP, Weigend weight eliminator. Standard back-propagation (93) is still the most popular.
8.	Learning rate and momentum:
	Need to be carefully chosen so that the net does not get stuck in local minima nor "shoot" off in the wrong direction when encountering small bumps on the error surface. For standard BP a learning rate of 0.1 and a momentum of 0.9 are best.
9.	Stability:
	Best to reserve some of the training data for cross-validation.

Once trained to the best generalisation point, the neural network may then be challenged with stimuli whose determinand concentrations are unknown, thereby allowing the operator accurately to predict the concentration of the determinand, without recourse to labour-intensive and often difficult wet chemistry.

It is evident that there are a large number of parameters which one may vary during the production of a neural network calibration model. For convenience, the most relevant for simple back-propagation networks are summarised in Table 1.

Outside the rather inaccessible mathematical literature, there has been relatively little work on the statistical validation of neural network predictions. So although they can be trained to the optimal point, when challenged with a new stimulus the network will give its answer, but as yet, it is not possible to put accurate confidence limits on the prediction (59). However, the link between statistics and neural networks is now becoming increasingly realised (13, 19, 41, 52, 85), and it is arguably, therefore, only a matter of time before true statistical confidence limits (beyond simple mean ± standard deviation on replicates) will be applied to neural network outputs.

The quantification of a binary mixture of bacterial cells by neural network analysis of their pyrolysis mass spectra

Our own aims have been to expand the application of the PyMS technique from (bacterial) taxonomy to the rapid and *quantitative* analysis of the chemical constituents of microbial and other samples, and we have therefore sought to apply ANNs to the deconvolution and interpretation of pyrolysis mass spectra. Thus, we have been able to follow the production of indole in a number of strains of *E. coli* grown on media incorporating various amounts of tryptophan (45), to estimate the amount of casamino acids in mixtures with glycogen (47), to quantify the (bio)chemical constituents of complex biochemical binary mixtures of proteins and nucleic acids in glycogen, and to measure the concentrations of tertiary mixtures of bacterial cells (52). More recently, within biotechnology, we have used PyMS and ANNs for the quantitative analysis of recombinant cytochrome b_5 expression in *E. coli* (49), and for effecting the rapid screening of the high-level production of desired substances in fermentor broths (54).

With regard to classifications and discriminations, we have also exploited the combination of PyMS and ANNs for the rapid and accurate assessment of the presence of lower-grade seed oils as adulterants in extra virgin olive oils (50, 51), for the identification of strains of *Mycobacterium* (37) and *Propionibacterium* spp. (48, 53). In the later studies we also exploited Kohonen's self-organising feature map (66) succesfully to carry out unsupervised learning, and hence the classification of the *P. acnes* strains (48, 53).

An example of the exploitation of ANNs for the deconvolution of the pyrolysis mass spectra of *Staphylococcus aureus* mixed in *Escherichia coli* is given below.

Preparation of the bacterial binary mixtures

The bacteria used were *E. coli* W3110 and *S. aureus* NCTC6571. Both strains were grown in 2 L liquid media (glucose (BDH), 10.0 g; peptone (LabM), 5.0 g; beef extract (LabM), 3.0 g; H_2O, 1 L) for 16 h at 37 °C in a shaker. After growth the cultures were harvested by centrifugation and washed in phosphate buffered saline (PBS). The dry weight of the cells were estimated gravimetrically and used to adjust the weight of the

final slurries using PBS to approximately 40 mg/mL. Two sets of mixtures were then prepared. The training set consisted of x% *E. coli* and y% *S. aureus,* where x : y were 100 : 0, 75 : 25, 50 : 50, 25 : 75, and 0 : 100. The second, "unknown" test set consisted of (x% *E. coli* : y% *S. aureus*) where x:y were 90 : 10, 80 : 20, 70 : 30, 60 : 40, 40 : 60, 30 : 70, 20 : 80, and 10 : 90.

Pyrolysis mass spectrometry

Aliquots (5 µL) of the bacterial suspensions were evenly applied onto iron-nickel foils. Prior to pyrolysis the samples were oven dried at 50 °C for 30 min. Samples were run in triplicate. The pyrolysis mass spectrometer used in this study was the Horizon Instruments PYMS-200X, as initially described by *Aries* et al. (3); for full operational procedures see *Freeman* et al. (38), *Goodacre* and *Kell* (45), *Goodacre* et al. (46, 53) and *Magee* et al. (72). The sample tube carrying the foil was heated, prior to pyrolysis, at 100 °C for 5 s. Curie-point pyrolysis was at 530 °C for 3 s. Data were collected over the m/z range 51 to 200.

Data Analysis

The data from PyMS may be displayed as quantitative pyrolysis mass spectra (e. g. see Fig. 3). The abscissa represents the m/z ratio (mass) whilst the ordinate contains information on the ion count for any particular m/z value ranging from 51 to 200. Data were normalised as a percentage of total ion count to minimise the influence of sample size *per se*.

The data (normalised as above but not weighted by their standard deviations) were analysed by principal components analysis (PCA) using the program Unscrambler II Version 4.0 (CAMO A/S, Olav Tryggvasonsgt. 24, N-7011 Trondheim, Norway; and see 74) which runs under Microsoft MS-DOS 6.2 on an IBM-compatible PC. PCA was employed as a method to detect outliers. If any outlying samples were discovered these would be removed from the construction of the neural network calibration model to prevent (gross) inaccuracies in future predictions.

Neural network simulations

All ANN analyses were carried out using a user-friendly, neural network simulation program, NeuralDesk (Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton, Hants, U.K.), which runs under Microsoft Windows 3.1 (or Windows NT) 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 over a 386 processor, permitting the analysis (and updating) of some 400,000 weights per second, was used. Data were also processed prior to analysis using the Microsoft Excel 4.0 spreadsheet.

The structure of the ANN (see above) consisted of 3 layers made up of the 150 input nodes (normalised pyrolysis mass spectra), 1 output node (the determinand; i. e., the percentage of *S. aureus* mixed with *E. coli*), and one "hidden" layer containing 8 nodes (150-8-1). The algorithm used was standard back-propagation (BP) (93). The input layer was scaled between 0 and 1 over the whole dataset (i. e. not on individual nodes) whilst the output layer was scaled to exploit less than the full range of the normalised scale between 0 and 1 (47, 52), that is, from –10 to 110% of *S. aureus* (thus the normalised scaling range was from 0.1 to 0.9).

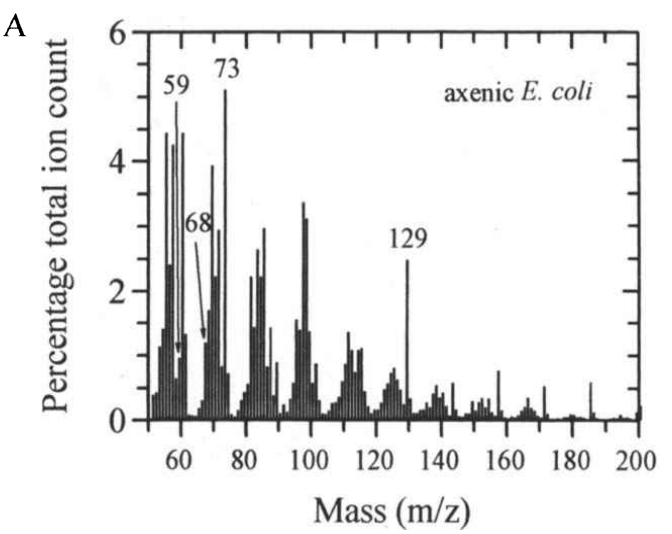
For training the ANNs, the training set consisted of the inputs which were the normalised triplicate pyrolysis mass spectra derived from 0,25, 50, 75 and 100% *S. aureus* in *E. coli,* and the desired outputs which were the actual (true) percentage of *S. aureus.* The network was trained and the RMS error of the training and test sets were calculated in order to ascertain when the network would best generalise. Finally after training, all pyrolysis mass spectra of the *S. aureus/E. coli* mixture were used as the "unknown" inputs (test data); the network then output its estimate in terms of the percentage of *S. aureus* mixed with *E. coli.*

Results and Discussion

Pyrolysis mass spectral fingerprints of 200 μg *E. coli,* pure *S. aureus* and of 100 μg *S. aureus* mixed with 100 μg *E. coli* (50%: 50% mix) are shown in Fig. 3. These pyrolysis mass spectra are very complex, such that their visual distinction is very difficult. At first there appears to be relatively little difference between them, with the particular exception of m/z 129 which is quite intense in the spectra from axenic *E. coli* (Fig. 3a) and the 50 : 50 mixture of *S. aureus* and *E. coli* (Fig. 3c), but much reduced in the spectra of pure *S. aureus* (Fig. 3b).

Figure 4 shows a simple subtraction of the normalised averages of three spectra of *E. coli* from three of *S. aureus.* The positive half of the graph indicates the peaks that are more intense in the spectra of *S. aureus* and shows some similarities to the pyrolysis mass spectrum of axenic *S. aureus* (Fig. 3b); these were notably m/z 59 and 68. Similarly the negative half of the subtraction spectra (Fig. 4) shows some analogies to the spectrum of pure *E. coli* (Fig. 3a); the most distinct peaks in the difference spectrum being m/z 73 and 129.

If these masses can be considered characteristic for *S. aureus* and *E. coli,* the intensities of m/z 59 and m/z 68 for *S. aureus,* and m/z 73 and m/z 129 for *E. coli* should be proportional to the relative proportion of these two bacteria in the mixtures. Plots of the average intensities of these four masses against the percentage of *S. aureus* in the binary mixtures, with standard error bars and the best linear fits, are shown in Figure



5. It can be seen that m/z 59 and m/z 68 do indeed alter in a fashion that is approximately linear with the % *S. aureus*, and that the intensities of m/z 73 and m/z 129 do decrease in a roughly linear manner.

One might hazard that it is possible simply to use the intensities of either m/z 59 or m/z 68 to estimate the percentage of *S. aureus* in these bacterial mixtures. However,

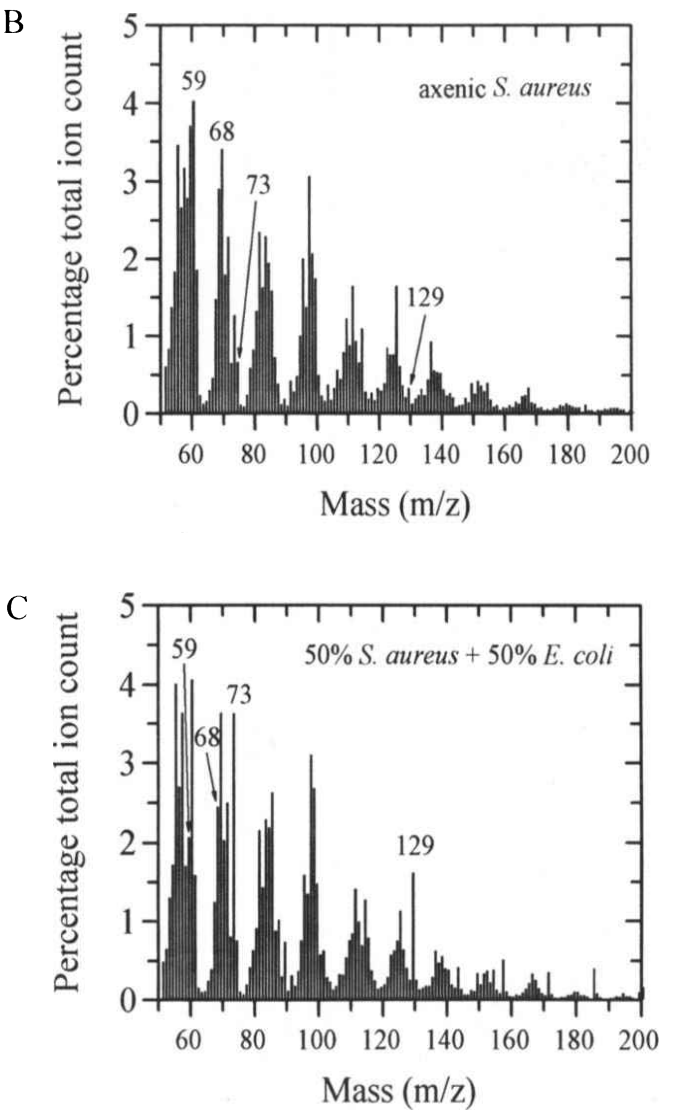


Fig. 3. Normalised pyrolysis mass spectra of 200 μg *E. coli* (A), 200 μg *S. aureus* (B), and of 100 μg *S. aureus* mixed with 100 μg *E. coli* (C). These spectra were obtained as described using a Curie-point temperature of 530 °C for 3 s with the Horizon Instruments PYMS-200X pyrolysis mass spectrometer.

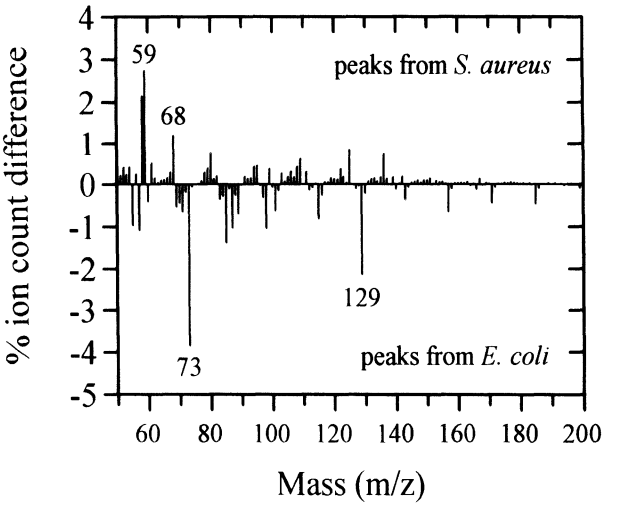


Fig. 4. A subtraction spectrum of the normalised average of three pyrolysis mass spectra of *E. coli* (Fig. 3a) from the average spectra of three from *S. aureus* (Fig. 3b).

there are two main problems with this: the first is that the variation in the intensity of these two masses is quite large, particularly above 70% *S. aureus* where the standard deviation error bars are very big. The other, albeit minor, problem is that although the relationships between the % ion counts of these masses and the % *S. aureus* are *linear*, they are not *proportional* (i. e., the line does not pass through the origin), which means that the source of m/z 59 and m/z 68 is not purely from *S. aureus*, and there is some contribution from *E. coli;* this can be clearly seen in the spectra of pure *E. coli* (Fig. 3a) which contribute 1.0% and 1.7% to m/z 59 and m/z 68 respectively (Fig. 5). Likewise, the masses derived from the pyrolysis of cells of *E. coli* (i. e., m/z 73 and m/z 129) do not alter in a proportional manner, and there is some contribution from *S. aureus* (1.3% and 0.3% respectively [Fig. 5]); again this can be observed in the spectra of axenic *S. aureus* (Fig. 3b). Thus, changes in these single ions can not be used accurately to estimate the percentage of cells of *S. aureus* mixed with *E. coli.* In other, less favourable cases, there may be interactions between constituents of the pyrolysate, which would change the mass spectra in a *non-linear* fashion (54, 94, 101).

The next stage was to look at the relationship between the pyrolysis mass spectra of the binary bacterial mixtures using principal components analysis (PCA). PCA is the best method for reducing the dimensionality of multivariate data whilst preserving most of the variance; in our pyrolysis mass spectral data this reduction was from the 150 m/z values to the first 2 or 3 principal components (PCs). Plots of the first two PCs of the variance in the PyMS for *S. aureus* in *E. coli* which account for 89.6% and 7.2% of the total variation respectively, are shown in Figure 6; this graph show that most of the variation was preserved in the first PC. It was also evident that the first PC served roughly to account for (or describe) the difference in the amount of *S. aureus* mixed with *E. coli.* The second PC did, however, contain some information in some of the samples (especially in mixtures containing 25, 75 and 80% *S. aureus*). That the second PC only accounted for 7.2% of the total variation indicates that the 25, 75 and

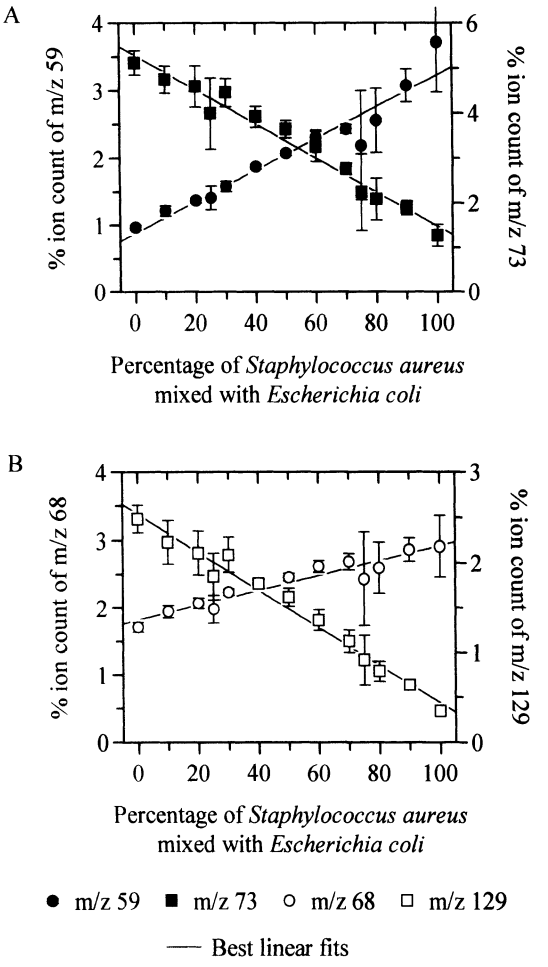


Fig. 5. Plots of the percentage intensity of m/z 59 and 73 (A) and of m/z 68 and 129 (B) against the percentage of *S. aureus* mixed with *E. coli*. Error bars show standard deviation based on triplicates. The best linear fits are shown.

80% *S. aureus* mixtures were not *significantly* outlying; thus the inclusion of the pyrolysis mass spectra of the 25 and 75% mixtures in the training set should not adversely affect the calibration model produced by the neural network analysis.

We therefore trained ANNs, using the standard back-propagation algorithm, with the normalised ion intensities from the triplicate pyrolysis mass spectra from the training sets as the inputs and the amount of *S. aureus* (0, 25, 50, 75 and 100%) mixed with *E. coli* as the output, the latter being scaled to lie between 0.1 and 0.9. The effectiveness of training was expressed in terms of the RMS error between the actual and desired network outputs; this "learning curve" is shown in Figure 7 (continuous line). The "learning curve" of the test data (broken line) is also shown in Figure 7; it can be

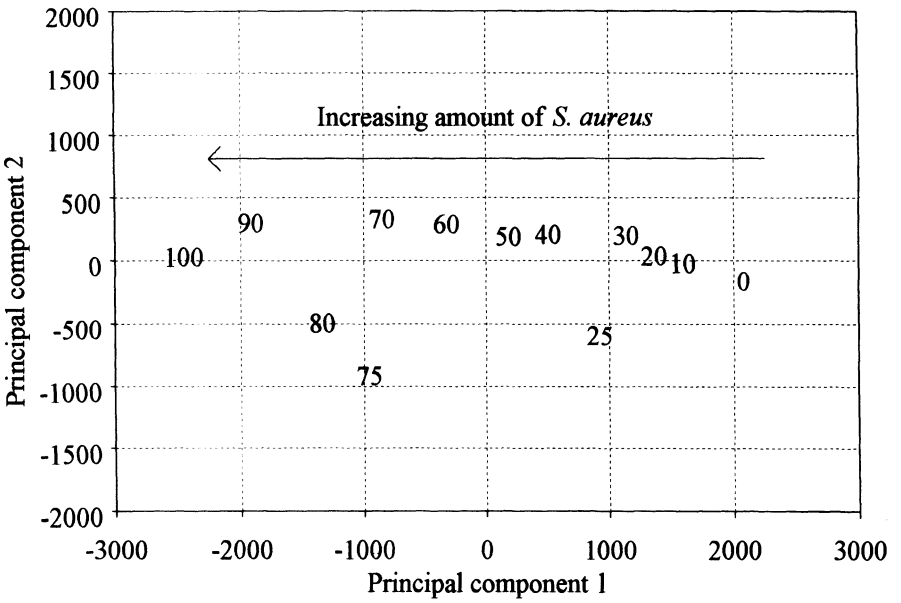


Fig. 6. Principal components biplots based on PyMS data showing the relationship between the PyMS of various amounts of *S. aureus* mixed with *E. coli*. The first two principal components are displayed and they account for 89.6% and 7.2% of the total variation respectively. The numbers represent the percentage of *S. aureus* in the binary mixture.

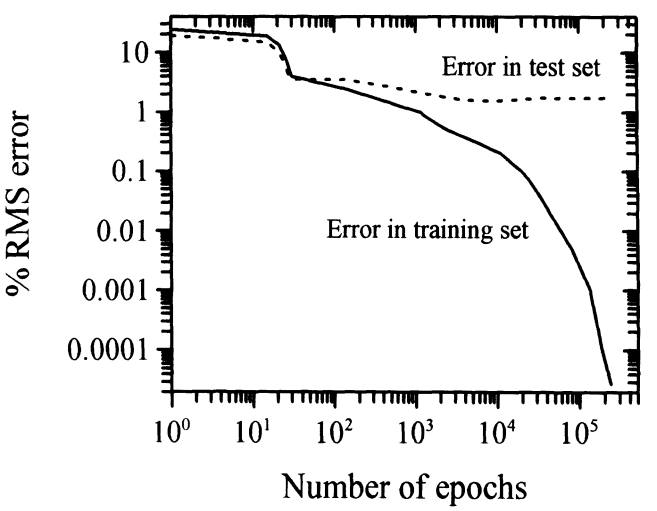


Fig. 7. Typical learning curves for the ANN, using the standard back propagation algorithm and with one hidden layer consisting of eight nodes, trained to estimate the percentage of *S. aureus* mixed with *E. coli*. The continuous line represent the % RMS error of the data used to train the neural network (the training set) and the broken line the data from the test set.

seen that whereas the learning curve of the training set continues to decrease during training the test set's learning curve initially decreases for approximately 5.10^3 epochs and then increases. This indicates that after this point the ANN was being over-trained, and it is important not to over-train ANNs since (by definition) the network will not generalise well (45).

Over-training was also detected when the RMS error of the test set is plotted against the RMS error of the training set (Fig. 8). The minimum RMS error in the test set was reached (1.55%) when the RMS error of the training set was 0.30% and optimal training had occurred. The ANN was then interrogated with the training and test sets and a plot of the network's estimate versus the true amount of *S. aureus* mixed with *E. coli* (Fig. 9) gave a linear fit which was indistinguishable from the expected proportional fit (i. e. y = x). It is therefore evident that the network's estimate of the percentage of *S. aureus* in the mixtures is very similar to the true quantity. This is true both for spectra that were used as the training set and, most importantly, for the "unknown" pyrolysis mass spectra.

In summary, it is evident from these results that ANNs can be trained with the pyrolysis mass spectra of binary bacterial mixtures of *S. aureus* and *E. coli* so as to gain quantitative information of the percentage content of *S. aureus*.

Concluding remarks

The arrival of novel chemometric techniques employing *supervised* learning such as neural networks (and indeed other analysis methods such as multiple linear regression, partial least squares and principal components regression) allow *quantitative* as well as *qualitative* analysis of multivariate data.

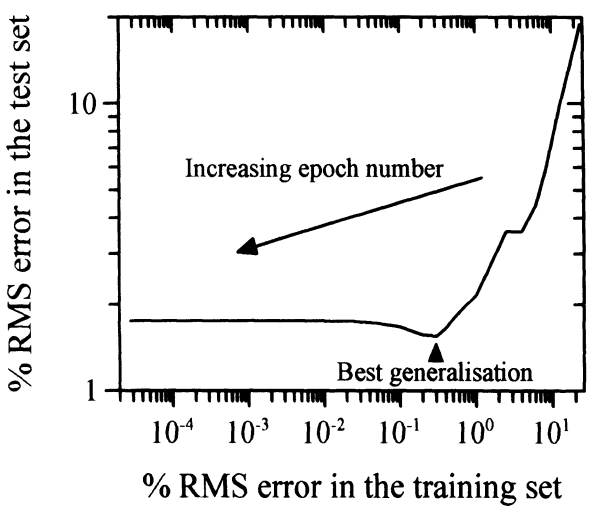


Fig. 8. A plot of the percentage RMS error of the test set versus the percentage RMS error of the training set. This show that optimal training occurred (i. e. at 0.3% RMS error). The number of epochs (and hence extent training) increases from right to left.

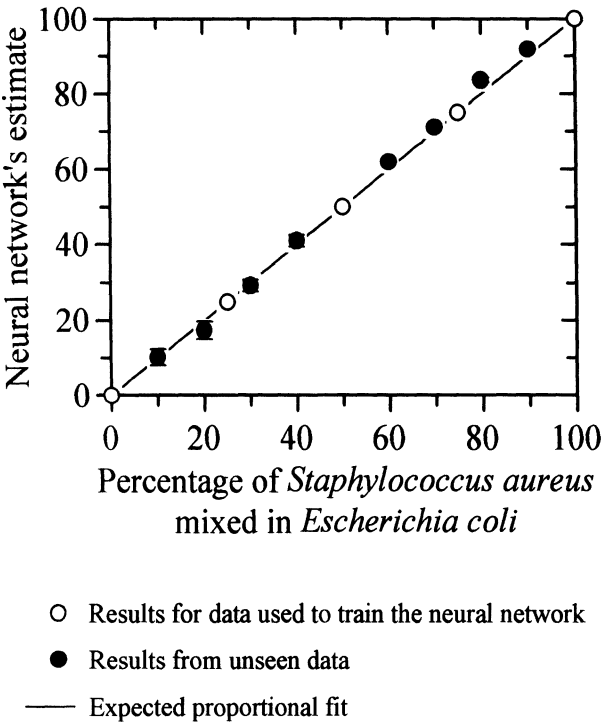


Fig. 9. The estimates of trained 150-8-1 neural networks versus the true percentage of *S. aureus*. ANNs were trained using the standard-back propagation algorithm, to 0.3% RMS error (the point at which Figure 5 had indicated that optimal training took place). Data points are the averages of the triplicate pyrolysis mass spectra. Open circles represent spectra that were used to train the network and closed circles indicate "unknown" spectra which were not in the training set. Error bars show standard deviation. The expected proportional fit is shown.

Pyrolysis mass spectrometry is a rapid, sensitive, relatively low cost, and instrument-based technique which produces a chemical fingerprint of the material to be analysed in less than 2 minutes. The data produced are multivariate and so ideally suited to analysis by neural networks.

The example given above demonstrates the ease with which neural networks can successfully be trained with the pyrolysis mass spectra of samples of a bacterial binary mixture of *S. aureus* and *E. coli* so as to be able accurately to quantify the concentration of *S. aureus*.

In conclusion, the combination of PyMS and ANNs constitutes a novel, powerful and increasingly accessible technology for the precise, accurate and quantitative analysis of the concentrations of appropriate substrates, metabolites and products in (bio)chemical processes generally (54). This offers the microbiologist an alternative to the often labour intensive, slow and difficult methods used, for example, to quantify drugs produced in fermentors.

Acknowledgements. This work was supported by the Biotechnology Directorate of the UK SERC, under the terms of the LINK scheme in Biochemical Engineering, in collaboration with Horizon Instruments, Neural Computer Sciences and Zeneca Bio Products.

References

1. *Aleksander, I.* and *H. Morton:* An Introduction to Neural Computing. Chapman & Hall, London (1990)
2. *Amit, D. J.:* Modeling Brain Function; The World of Attractor Neural Networks. Cambridge University Press, Cambridge (1989)
3. *Aries, R. E., C. S. Gutteridge,* and *T. W. Ottley:* Evaluation of a low-cost, automated pyrolysis-mass spectrometer. J. Anal. Appl. Pyrol. 9 (1986) 81–98
4. *Ball, J. W.* and *P. C. Jurs:* Automated selection of regression-models using neural networks for C-13 NMR spectral predictions. Anal. Chem. 65 (1993) 505–512
5. *Baum, E. B.* and *D. Haussler:* What size net gives valid generalization? Neural Computation 1 (1989) 151–160
6. *Beale, R.* and *T. Jackson:* Neural Computing: An Introduction. Adam Hilger, Bristol (1990)
7. *Beebe, K. R., W. W. Blaser, R. A. Bredeweg, J. P. Chauvel, R. S. Harner, M. LaPack, A. Leugers, D. P. Martin, L. G. Wright,* and *E. D. Yalvac:* Process analytical-chemistry. Anal. Chem. 65 (1993) 199R–216R
8. *Berkeley, R. C. W., R. Goodacre, R. J. Helyer,* and *T. Kelley:* Pyrolysis-MS in the identification of micro-organisms. Lab. Pract. 39 (1990) 81–83
9. *Blank, B. T.* and *S. D. Brown:* Nonlinear multivariate mapping of chemical data using feed-forward neural networks. Anal. Chem. 65 (1993) 3081–3089
10. *Boddy, L.* and *C. W. Morris:* Neural network analysis of flow cytometry data. In: Flow Cytometry in Microbiology, pp. 159–169 (D. Lloyd, Ed.). Springer-Verlag, London (1993)
11. *Borggaard, C.* and *H. H. Thodberg:* Optimal minimal neural interpretation of spectra. Anal. Chem. 64 (1992) 545–551
12. *Bos, A., M. Bos,* and *W. E. van der Linden:* Artificial neural networks as a tool for soft-modeling in quantitative analytical-chemistry – the prediction of the water-content of cheese. Anal. Chim. Acta 256 (1992) 133–144
13. *Bourlard, H. A.* and *N. Morgan:* Connectionist Speech Recognition: A Hybrid Approach. Kluwer Academic Publishers, Boston, MA (1994)
14. *Brereton, R. G.:* Chemometrics: Applications of Mathematics and Statistics to Laboratory Systems. Ellis Horwood, New York (1990)
15. *Brereton, R. G.:* Multivariate Pattern Recognition in Chemometrics. Elsevier, Amsterdam (1992)
16. *Broomhead, D.* and *D. Lowe:* Multivariate function interpolation and adaptive networks. Complex Systems 2 (1988) 321–355
17. *Brown, S. D., R. S. Bear,* and *T. B. Blank:* Chemometrics. Anal. Chem. 64 (1992) 22R–49R
18. *Bruchmann, A., P. Zinn,* and *C. M. Haffer:* Prediction of gas-chromatographic retention index data by neural networks. Anal. Chim. Acta 283 (1993) 869–880
19. *Buntine, W. L.* and *A. S. Weigend:* Bayesian back-propagation. Complex Systems 5 (1991) 603–643
20. *Carpenter, G. A.* and *S. Grossberg:* Pattern Recognition by Self-Organizing Neural Networks. MIT Press, Cambridge, MA (1991)
21. *Causton, D. R.:* A Biologist's Advanced Mathematics, pp. 48–72. Allen & Unwin, London (1987)
22. *Chatfield, C.* and *A. J. Collins:* Introduction to Multivariate Analysis, pp. 57–81. Chapman & Hall, London (1980)

23. *Chun, J., E. Atalan, A. C. Ward,* and *M. Goodfellow:* Artificial neural network analysis of pyrolysis mass-spectrometric data in the identification of *Streptomyces* strains. FEMS Microbiol. Lett. 107 (1993) 321–325
24. *Cowan, J. D.* and *D. H. Sharp:* Neural nets. Q. Rev. Biophys. 21 (1988) 365–427
25. *Curry, B.* and *D. E. Rumelhart:* MSnet: a neural network which classifies mass spectra Tetra. Comput. Meth. 3 (1990) 213–237
26. *Cybenko, G.:* Approximation by superposition of a sigmoidal function. Mathemat. Control Signals Syst. 2 (1989) 303–314
27. *Drennen, J. K., E. G. Kraemer,* and *R. A. Lodder:* Advances and perspectives in near-infrared spectrophotometry. Critical Rev. Anal. Chem. 22 (1991) 441–475
28. *Eberhart, R. C.* and *R. W. Dobbins:* Neural Network PC Tools. Academic Press, London (1990)
29. *Erwin, E., K. Obermayer,* and *K. Schulten:* Self-organizing maps: Ordering, convergence properties and energy functions. Biolog. Cybernet. 67 (1992) 47–55
30. *Everitt, B. S.:* Cluster Analysis. Edward Arnold, London (1993)
31. *Fahlman, S. E.:* An Empirical Study of Learning Speed in Back Propagation Networks. Carnegie-Mellon University Technical Report CMU-CS-88-162 (1988)
32. *Fahlman, S. E.* and *C. Lebiere:* The Cascade-Correlation Learning Architecture. Report CMU-CS-90-100, Carnegie-Mellon University 1990
33. *Finoff, W., F. Hergert,* and *H. G. Zimmermann:* Improving model selection by nonconvergent methods. Neural Networks 6 (1993) 771–783
34. *Flury, B.* and *H. Riedwyl:* Multivariate Statistics: A Practical Approach, pp. 181–233. Chapman & Hall, London (1988)
35. *Frank, I. E.:* A nonlinear PLS model. Chemom. Intell. Lab. Sys. 8 (1990) 109–119
36. *Frean, M.:* The upstart algorithm: A method for constructing and training feedforward neural networks. Neural Computation 2 (1990) 198–209
37. *Freeman, R., R. Goodacre, P. R. Sisson, J. G. Magee, A. C. Ward,* and *N. F. Lightfoot:* Rapid identification of species within the *Mycobacterium tuberculosis* complex by artificial neural network analysis of pyrolysis mass spectra. J. Med. Microbiol. 40 (1994) 170–173
38. *Freeman, R., F. K. Gould, R. Wilkinson, A. C. Ward, N. F. Lightfoot,* and *P. R. Sisson:* Rapid inter-strain comparison by pyrolysis mass spectrometry of coagulase-negative staphylococci from persistent CAPD peritonitis. Epidem. Infect. 106 (1991) 239–246
39. *Funahashi, K.:* On the approximate realization of continuous-mappings by neural networks. Neural Networks 2 (1989) 183–192
40. *Gallant, S. I.:* Neural Network Learning. MIT Press, Cambridge, MA (1993)
41. *Geman, S., E. Bienenstock,* and *R. Doursat:* Neural networks and the bias/variance dilemma. Neural computation 4 (1992) 1–58
42. *Gemperline, P. J., J. R. Long,* and *V. G. Gregoriou:* Nonlinear multivariate calibration using principal components regression and artificial neural networks. Anal. Chem. 63 (1991) 2313–2323
43. *Goodacre, R.:* Characterisation and quantification of microbial systems using pyrolysis mass spectrometry: Introducing neural networks to analytical pyrolysis. Microbiol. Eur. 2(2) (1994) 16–22
44. *Goodacre, R.* and *R. C. W. Berkeley:* Detection of small genotypic changes in *Escherichia coli* by pyrolysis mass spectrometry. FEMS Microbiol. Lett. 71 (1990) 133–138
45. *Goodacre, R.* and *D. B. Kell:* Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: Application to indole production. Anal. Chim. Acta 279 (1993) 17–26
46. *Goodacre, R., J. E. Beringer,* and *R. C. W. Berkeley:* The use of pyrolysis-mass spectrometry to detect the fimbrial adhesive antigen F41 from *Escherichia coli* HB101 (pSLM204). J. Anal. Appl. Pyrol. 22 (1991) 19–28
47. *Goodacre, R., A. N. Edmonds,* and *D. B. Kell:* Quantitative analysis of the pyrolysis-mass spectra of complex mixtures using artificial neural networks: Application to casamino acids in glycogen. J. Anal. Appl. Pyrol. 26 (1993) 93–114

48. *Goodacre, R., S. A. Howell, W. C. Noble,* and *M. J. Neal:* Sub-species discrimination using pyrolysis mass spectrometry and self-organising neural networks of *Propionibacterium acnes* isolated from normal human skin. Zbl. Bakt., in this volume (1996, pp. 501–515)
49. *Goodacre, R., A. Karim, M. A. Kaderbhai,* and *D. B. Kell:* Rapid and quantitative analysis of recombinant protein expression using pyrolysis mass spectrometry and artificial neural networks: application to mammalian cytochrome b_5 in *Escherichia coli.* J. Biotechnol. 34 (1994) 185–193
50. *Goodacre, R., D. B. Kell,* and *G. Bianchi:* Neural networks and olive oil. Nature 359 (1992) 594
51. *Goodacre, R., D. B. Kell,* and *G. Bianchi:* Rapid assessment of olive oil adulteration using pyrolysis mass spectrometry and artificial neural networks. J. Sci. Food Agric. 63 (1993) 297–307
52. *Goodacre, R., M. J. Neal,* and *D. B. Kell:* Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. Anal. Chem. 66 (1994) 1070–1085
53. *Goodacre, R., M. J. Neal, D. B. Kell, L. W. Greenham, W. C. Noble,* and *R. G. Harvey:* Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. J. Appl. Bact. 76 (1994) 124–134
54. *Goodacre, R., S. Treu, C. Wrigley-Jones, M. J. Neal, J. Maddock, T. W. Ottley, N. Porter,* and *D. B. Kell:* Rapid screening for metabolite overproduction in fermentor broths using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks. Biotechn. Bioengin. 44 (1994) 1205–1216
55. *Gutteridge, C. S.:* Characterization of microorganisms by pyrolysis mass spectrometry. Meth. Microbiol. 19 (1987) 227–272
56. *Gutteridge, C. S., L. Vallis,* and *H. J. H. MacFie:* Numerical methods in the classification of microorganisms by pyrolysis mass spectrometry. In: Computer-assisted Bacterial Systematics, pp. 369–401 (eds., M. Goodfellow, D. Jones, F. G. Priest). Academic Press, London (1985)
57. *Hassibi, B.* and *D. G. Stork:* Second order derivatives for network pruning: Optimal brain surgeon. In: Advances in Neural Information Processing Systems 5, pp. 164–171 (eds., S. J. Hanson, J. D. Cowan and C. L. Giles). Morgan Kaufmann, San Mateo, CA (1993)
58. *Hassibi, B., D. G. Stork,* and *G. J. Wolff:* Optimal Brain Surgeon And General Network Pruning. IEEE Internat. Confer. Neural Networks 1–3 (1993) 293–299
59. *Hecht-Nielsen, R.:* Neurocomputing. Addison-Wesley, Massachusetts (1990)
60. *Hertz, J., A. Krogh,* and *R. G. Palmer:* Introduction to the Theory of Neural Computation. Addison-Wesley, California (1991)
61. *Hornik, K., M. Stinchcombe,* and *H. White:* Multilayer feedforward networks are universal approximators. Neural Networks 2 (1989) 359–368
62. *Hornik, K., M. Stinchcombe,* and *H. White:* Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3 (1990) 551–560
63. *Höskuldsson, A.:* Quadratic PLS regression. J. Chemom. 6 (1992) 307–334
64. *Irwin, W. J.:* Analytical Pyrolysis: A Comprehensive Guide. Marcel Dekker, New York (1982)
65. *Kell, D. B.* and *C. L. Davey:* On fitting dielectric spectra using artificial neural networks. Bioelectrochem. Bioenerg. 28 (1992) 425–434
66. *Kohonen, T.:* Self-Organization and Associative Memory. Springer-Verlag, Berlin (1989)
67. *Kvalheim, O. M., D. W. Aksnes, T. Brekke, M. O. Eide, E. Sletten,* and *N. Telnæs:* Crude oil characterization and correlation by principal component analysis of ^{13}C nuclear magnetic resonance spectra. Anal. Chem. 57 (1985) 2858–2864
68. *LeCun, Y., J. S. Denker,* and *S. A. Solla:* Optimal brain damage. In: Advances in Neural Information Processing Systems 1 (ed., D. S. Touretzky), pp. 598–605. Morgan Kaufmann, New York (1989)

69. *Long, J. R., H. T. Mayfield, M. V. Henley,* and *P. R. Kromann:* Pattern recognition of jet fuels chromatographic data by artificial neural networks with back-propagation of error. Anal. Chem. 63 (1991) 1256–1261
70. *MacFie, H. J. H., C. S. Gutteridge,* and *J. R. Norris:* Use of canonical variates in differentiation of bacteria by pyrolysis gas-liquid chromatography. J. Gen. Microbiol. 104 (1978) 67–74
71. *Magee, J. T.:* Whole-organism fingerprinting, in Handbook of New Bacterial Systematics, pp. 383–427 (M. Goodfellow and A. G. O'Donnell, Eds.). Academic Press, London (1993)
72. *Magee, J. T., J. M. Hindmarch, B. I. Duerden,* and *L. Goodwin:* A pyrolysis mass spectrometry study of oral pigmented bacteroides. J. Med. Microbiol. 37 (1992) 56–61
73. *Mark, H.:* Principles and practice of spectroscopic calibration. John Wiley & Sons, New York (1991)
74. *Martens, H.* and *T. Næs:* Multivariate Calibration. John Wiley & Sons, New York (1989)
75. *Martin, K. A.:* Recent advances in near-infrared reflectance spectroscopy. Appl. Spect. Revs. 27 (1992) 325–383
76. *Massart, D. L., B. G. M. Vandeginste, S. N. Deming, Y. Michotte,* and *L. Kaufmann:* Chemometrics: A Textbook. Elsevier, Amsterdam (1988)
77. *McAvoy, T. J., H. T. Su, N. S. Wang, M. He, J. Horvath,* and *H. Semerjian:* A comparison of neural networks and partial least-squares for deconvoluting fluorescence-spectra. Biotechn. Bioeng. 40 (1992) 53–62
78. *McClelland, J. L.* and *R. E. Rumelhart:* Explorations in Parallel Distributed Processing: A Handbook of Models, Programs and Exercises. MIT Press, Cambridge, MA (1988)
79. *Meloun, M., J. Militky,* and *M. Forina:* Chemometrics for Analytical Chemistry. Vol 1: PC-aided Statistical Data Analysis. Ellis Horwood, Chichester, UK (1992)
80. *Meuzelaar, H. L. C., J. Haverkamp,* and *F. D. Hileman:* Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials. Elsevier, Amsterdam (1982)
81. *Miller, A. J.:* Subset Selection in Regression. Chapman & Hall, London (1990)
82. *Mjolsness, E., D. H. Sharp,* and *B. K. Alpert:* Scaling, machine learning, and genetic neural nets. Adv. Appl. Math. 10 (1989) 137–163
83. *Moody, J.* and *C. Darken:* Fast learning in networks of locally-tuned processing units. Neural Computation 1 (1989) 281–294
84. *Mozer, M. C.* and *P. Smolensky:* Skeletonization: A technique for trimming fat from a network *via* relevance assessment. In: Advances in Neural Information Processing Systems 1, pp. 107–115 (ed., D. S. Touretzky). Morgan Kaufmann, New York (1989)
85. *Næs, T., K. Kvaal, T. Isaksson,* and *C. Miller:* Artificial neural networks in multivariate calibration. J. Near Infrared Spectrosc. 1 (1993) 1–11
86. *Neal, M. J., R. Goodacre,* and *D. B. Kell:* On the analysis of pyrolysis mass spectra using artificial neural networks. Individual input scaling leads to rapid learning. Proc. World Congress on Neural Networks 1994, San Diego, California, (1994) I–318, I–323
87. *Pao, Y.-H.:* Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, Reading, MA (1989)
88. *Peretto, P.:* An Introduction to the Modelling of Neural Networks. Cambridge University Press, Cambridge (1992)
89. *Rataj, T.* and *J. Schindler:* Identification of bacteria by a multilayer neural network. Binary 3 (1991) 159–164
90. *Rawlings, J. O.:* Applied Regression Analysis. Wadsworth & Brooks, Pacific Grove, CA (1988)
91. *Reed, R.:* Pruning algorithms – a survey. IEEE Trans. Neural Networks 4 (1993) 740–747
92. *Richard, D., C. Cachet, D. Cabrol-Bass,* and *T. P. Forrest:* Neural network approach to structural feature recognition from infrared-spectra. J. Chem. Inf. Comput. Sci. 33 (1993) 202–210
93. *Rumelhart, D. E., J. L. McClelland,* and The PDP Research Group: Parallel Distributed Processing. Experiments in the Microstructure of Cognition. MIT Press, Cambridge, Massachusetts (1986)

94. *H.-R. Schulten* and *R. P. Lattimer:* Applications of mass-spectrometry to polymers. Mass Spec. Rev. 3 (1984) 231–315
95. *Seasholtz, M. B.* and *B. R. Kowalski:* The parsimony principle applied to multivariate calibration. Anal. Chim. Acta 277 (1993) 165–177
96. *Shadmehr, R., D. Angell, P. B. Chou, G. S. Oehrlein,* and *R. S. Jaffe:* Principal component analysis of optical-emission spectroscopy and mass-spectrometry – application to reactive ion etch process parameter-estimation using neural networks. J. Electrochem. Soc. 139 (1992) 907–914
97. *Simpson, P. K.:* Artificial Neural Systems. Pergamon Press, Oxford (1990)
98. *Smith, P. B.* and *A. P. Snyder:* Characterization of bacteria by quartz tube pyrolysis-gas chromatography ion trap mass-spectrometry. J. Anal. Appl. Pyrol. 24 (1992) 23–38
99. *Smits, J. R. M., P. Shoenmaker, A. Stehmann, F. Sijstermans,* and *G. Kateman:* Interpretation of infrared spectra with modular neural-network systems. Chemom. Intell. Lab. Syst. 18 (1993) 27–39
100. *Taavitsainen, V. M.* and *P. Korhonen:* Nonlinear data analysis with latent variable. Chemom. Intell. Lab. Sys. 14 (1992) 185–194
101. *D. Van de Meent, J. W. de Leeuw, P. A. Schenck, W. Windig,* and *J. Haverkamp:* Quantitative analysis of polymer mixtures by pyrolysis mass spectrometry/discriminant analysis. J. Anal. Appl. Pyrol. 4 (1982) 133–142
102. *Wasserman, P. D.:* Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York (1989)
103. *Wasserman, P. D.* and *R. M. Oetzel:* Neural Source: the Bibliographic Guide to Artificial Neural Networks. Van Nostrand Reinhold, New York (1989)
104. *Weigend, A. S., D. E. Rumelhart,* and *B. A. Huberman:* Generalization by weight-elimination with application to forecasting. In: Neural Information Processing Systems 3, pp. 875–882 (eds., R. P. Lippmann, E. Moody and D. S. Touretzky). Morgan Kaufmann, San Mateo, CA (1991)
105. *Weiss, S. H.* and *C. A. Kulikowski:* Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems. Morgan Kaufmann Publishers, California (1991)
106. *Werbos, P. J.:* Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. M. Sc. Thesis, Harvard University, USA (1974)
107. *Werbos, P. J.:* The Roots of Back-Propagation: From Ordered Derivatives to Neural Networks and Political Forecasting. John Wiley & Sons, Chichester (1993)
108. *White, H.:* Connectionist nonparametric regression – multilayer feedforward networks can learn arbitrary mappings. Neural Networks 3 (1990) 535–549
109. *Windig, W., J. Haverkamp,* and *P. G. Kistemaker:* Interpretation of a set of pyrolysis mass spectra by discriminant analysis and graphical rotation. Anal. Chem. 55 (1983) 387–391
110. *Wold, S.:* Nonlinear partial least squares modelling. II. Spline inner relation. Chemom. Intell. Lab. Sys. 14 (1992) 71–84
111. *Wold, S., N. Kettaneh-Wold,* and *B. Skagerberg:* Nonlinear PLS modelling. Chemom. Intell. Lab. Sys. 7 (1989) 53–65
112. *Wythoff, B. J.:* Orthonet: Orthogonal latent variable neural network. Chemom. Intell. Lab. Sys. 20 (1993) 129–148
113. *Zupan, J.* and *J. Gasteiger:* Neural Networks for Chemists: An Introduction. VCH Verlagsgesellschaft, Weinheim (1993)
114. *Zupan, J., M. Novič, X. Li,* and *J. Gasteiger:* Classification of multicomponent analytical data of olive oils using different neural networks. Anal. Chim. Acta 292 (1994) 219–234

Dr. R. *Goodacre*, Institute of Bilogical Sciences, University of Wales, Aberystwyth, Dyfed SY23 3DA, UK, E-mail: rrg@aber.ac.uk
http: //gepasi.dbs.aber.ac.uk/roy/pymshome.htm