ELSEVIER

Analytica Chimica Acta 348 (1997) 511–532

ANALYTICA CHIMICA ACTA

On mass spectrometer instrument standardization and interlaboratory calibration transfer using neural networks

Royston Goodacre[a,*], Éadaoin M. Timmins[a], Alun Jones[a], Douglas B. Kell[a], John Maddock[b], Margaret L. Heginbothom[c], John T. Magee[c]

[a]*Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, Dyfed, Wales SY23 3DA, UK*
[b]*Horizon Instruments Ltd., Ghyll Industrial Estate, Heathfield, E. Sussex TN21 8AW, UK*
[c]*Department of Medical Microbiology and Public Health Laboratory, Heath Park, Cardiff, CF4 4XN, Wales, UK*

Received 25 June 1996; received in revised form 30 December 1996; accepted 31 December 1996

Abstract

For pyrolysis mass spectrometry (PyMS) to be exploited in areas such as the routine identification of microorganisms, for quantifying determinands in biological and biotechnological systems, and in the production of useful mass spectral libraries, it is paramount that newly acquired spectra be comparable to those previously collected and held in a central reference laboratory. Artificial neural networks (ANNs) and other multivariate calibration models have been used to relate mass spectra to the biological features of interest. However, calibration models developed on one mass spectrometer cannot be used with spectra collected on a second instrument, because of the differences between the instrumental responses of both instruments. We report here that an ANN-based drift correction procedure can be implemented so that newly acquired spectra can be used to challenge models constructed using mass spectra collected on *different* instruments. Calibration samples were run on three different PyMS machines, and ANNs set up in which the inputs were the 150 machine 'a' calibration masses and the outputs were the 150 calibration masses from the machine 'b' spectra. Such associative neural networks could thus be used as signal-processing elements to effect the transformation of data acquired on one machine to those which would have been acquired on a different instrument. Therefore, for the first time PyMS could be used to acquire spectra which could usefully be compared to those previously collected and held in a data-base, irrespective of the mass spectrometer used. The examples reported are for the quantitative assessment of the amount of lysozyme in a binary mixture with glycogen and the rapid identification down to the species level of bacteria belonging to the genus *Eubacterium*. This approach is not limited solely to pyrolysis mass spectrometry but is generally applicable to any analytical tool which is prone to deterioration in calibration transfer, such as IR, ESR, NMR and other vibrational spectroscopies, gas and liquid chromatography, as well as other types of mass spectrometry.

Keywords: Artificial neural networks; Calibration transfer; Chemometrics; Multivariate calibration; Pyrolysis mass spectrometry; Standardization

1. Introduction

PyMS is a rapid and high-resolution method for the analysis of otherwise non-volatile material [1,2], and

*Corresponding author. Tel.: +44 1970 621947; fax: +44 1970 622354; e-mail: rrg@aber.ac.uk.

0003-2670/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0003-2670(97)00062-7

has been widely applied to the discrimination of closely related microbial strains [3–5]. Recent advances in artificial neural networks (ANNs) (see e.g. [6–14]), and other multivariate calibration methods based on supervised learning, which perform linear regression, such as partial least squares regression (PLS) and principal components regression (PCR) (see e.g. [15–20]) have now permitted its exploitation in the *quantitative* analysis of many samples of more (bio)chemical interest (see e.g. [5,21–25]).

Pyrolysis is the thermal degradation of a material in an inert atmosphere, and leads to the production of volatile fragments from non-volatile material. Curie-point pyrolysis is a particularly reproducible and straightforward version of the technique, in which the sample, dried onto an appropriate metal is rapidly heated (<0.5 s is typical) to the Curie point of the metal, which is usually chosen to be 530°C (although other temperatures exist, viz. 358, 480, 510, 610 and 770°C). The volatile fragments or pyrolysate [1] resulting from the Curie-point pyrolysis may then be separated and analysed in a mass spectrometer [2], and the combined technique is then known as pyrolysis mass spectrometry or PyMS.

Almost all biological materials will produce pyrolytic degradation products such as methane, ammonia, water, methanol and H_2S, whose mass:charge (m/z) ratio <50, and fragments with m/z >200 are rarely analytically important [26], the analytically useful data are thus constituted by a set of (150) normalised intensities versus m/z in the range 51–200.

The major problem with PyMS is that long-term reproducibility (>30 days) is poor, such that the mass spectral fingerprints of the same material analysed at two different times are different; this lack of reproducibility is due largely to instrumental drift in the mass spectrometer (and is not confined to PyMS). Therefore, within clinical microbiology PyMS has really been limited to the typing of short-term outbreaks where all micro-organisms are analysed in a single batch [4,27]. For PyMS to be used (a) for the routine identification of micro-organisms and (b) combined with multivariate calibration to quantify biological systems (e.g. metabolites of interest in fermentor broths), new spectra must be able to be compared to those previously collected.

In addition, calibration models developed on a particular mass spectrometer cannot be used with spectra collected on a second instrument, because of the differences between their instrumental responses [28]. To keep those models and effect calibration transfer, it is essential to correct for those differences. This can be achieved by standardisation procedures, which often involve the estimation of the instrumental differences by measuring standardisation samples on both machines, and then correction of those differences. Once developed it would be possible to transfer the calibration model to other instruments, so that old models can be used at the point of measurement rather than in the research laboratory [29].

The issue of calibration transfer has been studied by a number of researchers [30–33]; particular attention has been paid to the standardization of near-infrared spectrometers [34,35], and a recent review describing different standardization methods has been given by de Noord [36]. Amongst those methods proposed for multivariate calibration transfer between infrared instruments, 'piecewise direct standardization' (PDS) has received a great deal of attention [31,34,37–39]; other methods included locally weighted regression [40] while Bouveresse et al. [41] tested Shenk's algorithm to transfer NIR calibration models. Artificial neural networks with optimal associative memory (OAM) have been exploited for background correction of infrared spectra [42], and this approach has now incorporated fuzzy logic and has been extended to fuzzy OAM (FOAM) [43]. The later study demonstrated that multivariate calibrations were improved after analysis using FOAM, relative to the uncorrected spectra, for predicting low levels of glucose from infrared absorbance spectra of glucose in plasma matrices from single-beam data [43].

Smits et al. [44] have implemented a drift correction for pattern recognition using neural networks using *simulated* flow cytometry data. These data sets contained only two variables and the amount of drift was included in neural networks as an *extra* input variable (three input nodes in total). It is, however, often difficult to measure the amount of drift accurately in real systems, especially if the number of input variables is high (typically 150 with pyrolysis mass spectral data); a better method would be to *transform* the spectra collected 'today' to be like those collected previously.

We have found that neural networks can be used successfully to correct for instrumental drift on a

single mass spectrometer so that models created using old, previously collected data can be employed to give accurate estimates of determinand concentration or bacterial identities from newly acquired spectra when calibrated with standards common to the two data sets [45,46]. Calibration samples were run at the two times, and ANNs set up in which the inputs were the 150 'new' calibration masses and the outputs were the 150 calibration masses from the 'old' spectra. Such associative nets could thus be used as signal-processing elements to effect the transformation of data acquired one day to those which would have been acquired on a later date.

The aim of the present study was to assess whether our neural network transformation procedure could be extended from correcting spectra taken on the same instrument at different times to allowing the operator to create calibration models on a 'master' machine and use newly acquired spectra from 'slave' instruments. The examples reported are for the quantitative assessment of the amount of lysozyme in a binary mixture with glycogen and for the rapid identification down to the species level of human bacterial oral isolates belonging to the genus *Eubacterium*.

2. Experimental

2.1. The quantification of lysozyme in glycogen data set

Mixtures were prepared such that 5 μl of a solution contained 0–100 μg (in steps of 5 μg) of lysozyme (from chicken egg white, Sigma), in 100–0 μg glycogen (oyster type II, Sigma); such that the total was always 100 μg; representative of percentage mixtures. The samples were then frozen at −20°C until they were analysed by PyMS.

2.2. The identification of Eubacterium species data set

Five *Eubacterium* groups/taxa were analysed; five *E. timidum*, four *Eubacterium* C_1, five *Eubacterium* C_2, five *Eubacterium* New 1 and five *Eubacterium* isolates from Saint Bartholomew's Hospital (SBH). These isolates have been previously described [47,48] and studied by PyMS [49]; in the later study it was found that the SBH isolates belonged to the *Eubacterium* C_2 taxon. Details of the 19 organisms analysed can be found in [49] and Table 7.

Strains were cultured on Fastidious Anaerobe agar (Lab M Malthus, Bury UK) plus 5% sheep blood and incubated anaerobically in an anaerobic cabinet in an atmosphere of N_2 80% CO_2 10%, H_2 10% for 72 h. The bacteria were harvested with a nichrome wire loop and suspended in phosphate buffered saline to 20 mg ml^{-1}. The samples were then frozen at −20°C until they were analysed by PyMS.

2.3. Pyrolysis mass spectrometry (PyMS)

Three different pyrolysis mass spectrometers were used in this study; two PYMS-200X instruments (one in Aberystwyth and one in Cardiff), and a RAPyD-400 machine (Heathfield). Each of these PyMS instruments were constructed by Horizon Instruments (Heathfield, East Sussex, TN21 8AW, U.K.) and operate on very similar principles; the relevant differences are given in Table 1.

5 μl aliquots of the above samples were evenly applied onto iron–nickel foils to give a thin uniform

Table 1
Instruments used by the three participating laboratories

Laboratory	Machine	Mass spectrometer	Vacuum system
University of Wales, Aberystwyth (UWA)	Machine 1– PYMS-200X	Quadrupole *m/z* Range 11–200	Diaphragm pump + turbo-drag pump
Horizon Instruments (HI)	Machine 2– RAPyD-400	Quadrupole *m/z* Range 11–400	Rotary pump + turbo-molecular pump
Health Park, Cardiff (PHLS)	Machine 3– PYMS-200X	Quadrupole *m/z* Range 11–200	Rotary pump + diffusion pump

Table 2
Details of the two PyMS experiments studied to investigate instrument standardization and calibration transfer

PyMS experiment design	Pyrolysis mass spectra collected		Time difference in days
	Machine A	Machine B	
To quantify lysozyme in glycogen	Machine 1	Machine 2	481
To identify *Eubacterium* human isolates	Machine 1	Machine 3	331

surface coating. Prior to pyrolysis the samples were oven-dried at 50°C for 30 min. Each sample was analysed in triplicate. Three pyrolysis mass spectrometers used for the two experiments and details are given in Table 2. For full operational procedures see [23,25,50]. 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, with a temperature rise time of 0.5 s. These conditions were used for all experiments. The data from PyMS were collected over the *m/z* range 51–200 and may be displayed as quantitative pyrolysis mass spectra (e.g. as in Fig. 1; here normalised to total ion count). The abscissa represents the *m/z* ratio whilst the ordinate contains information on the ion count for any particular *m/z* value ranging from 51–200.

The pyrolysis mass spectra that collected were normalised so that the total ion count was 2^{16} to remove the influence of sample size per se.

Prior to any analysis the mass spectrometer was calibrated using the chemical standard perfluorokerosene (Aldrich), such that *m/z* 181 was one tenth of *m/z* 69.

2.4. Principal components analysis (PCA)

To observe the *natural* relationships between samples the normalised data were then analysed by PCA [17,51–56], according to the NIPALS algorithm [57], using the program Matlab version 4.2c.1 (The MathWorks, Natick, MA, USA), which runs under Microsoft Windows NT on an IBM-compatible PC.

PCA is a well-known technique for reducing the dimensionality of multivariate data whilst preserving most of the variance, and 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. Moreover, PCA can

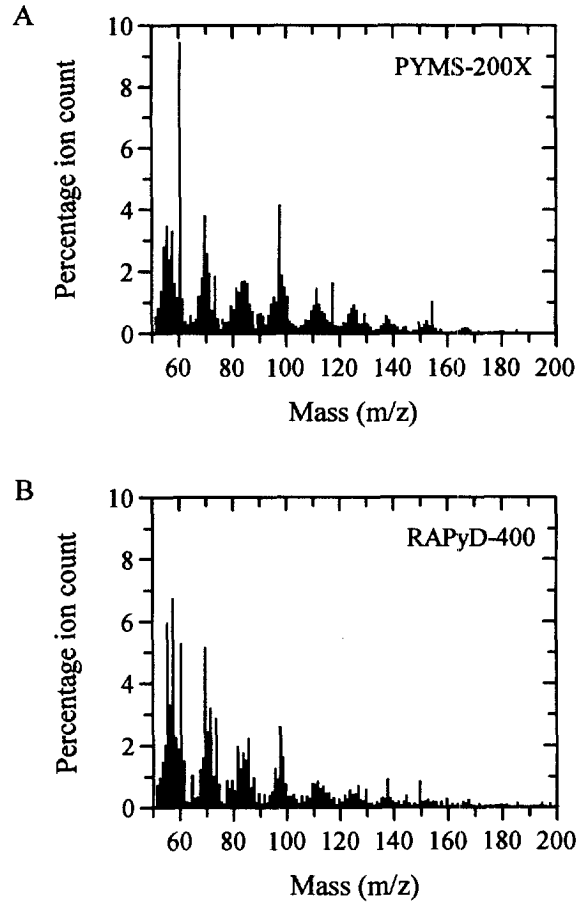

Fig. 1. Normalized pyrolysis mass spectra of 50 μg lysozyme mixed with 50 μg glycogen, analysed on the PYMS-200X instrument (Machine 1) at University of Wales, Aberystwyth (A) and on the RAPyD-400 instrument (Machine 2) at Horizon Instruments, Healthfield (B).

be used to identify *correlations* amongst a set of variables and to transform the original set of variables to a new set of *uncorrelated* variables called principal components (PCs). The objective of PCA is to see if

the first few PCs account for most (>90%) of the variation in the original data. If they do reduce the number of dimensions required to display the observed relationships, then the PCs can more easily be plotted and 'clusters' in the data visualized; moreover, this technique can be used to detect outliers.

2.5. Partial least squares (PLS) for the quantification of lysozyme in glycogen

PLS was used to quantify the levels of lysozyme in mixtures with glycogen. For calibrating PLS models so as to quantify the percentage of lysozyme the mixtures analysed on the PYMS-200X at UWA were divided into two sets; the training set consisted of x % lysozyme and y % glycogen, where $x : y$ were 100 : 0, 90 : 80, 80 : 20, 70 : 30, 60 : 40, 50 : 50, 40 : 60, 30 : 70, 20 : 80, 10 : 90, and 0 : 100. The second, 'unknown' test set consisted of (x % lysozyme: y % glycogen) where $x : y$ were 95 : 5, 85 : 15, 75 : 25, 65 : 35, 55 : 45, 45 : 55, 35 : 65, 25 : 75, 15 : 85, and 5 : 95.

All PLS analyses [15,17,18,58–60] were carried out using an in-house program, developed by Dr Alun Jones, which runs under Microsoft Windows NT on an IBM-compatible PC. Data were also processed prior to analysis using the Microsoft Excel 5.0 spreadsheet.

The first stage was the preparation of the data. This was achieved by presenting the 'training set' as two data matrices to the program; X, which contains the normalised triplicate pyrolysis mass spectra, and Y which represents the percentage of lysozyme (i.e., determinand) in glycogen. The X-data were mean centred and scaled in proportion to the reciprocal of their standard deviations.

The next stage was the generation of the calibration model using the PLS1 algorithm from data obtained on Machine 1. The method of validation used was full cross-validation, via the leave-one-out method [17]. This technique sequentially omits one sample from the calibration; the PLS model is then re-determined on the basis of this reduced sample set. The concentration of lysozyme of the omitted sample is then predicted with the use of the model. This method is required to determine the optimal size of the calibration model, so as to obtain good estimates of the precision of the multivariate calibration method (i.e., neither to under- nor over-fit predictions of unseen data) [16,17,61,62].

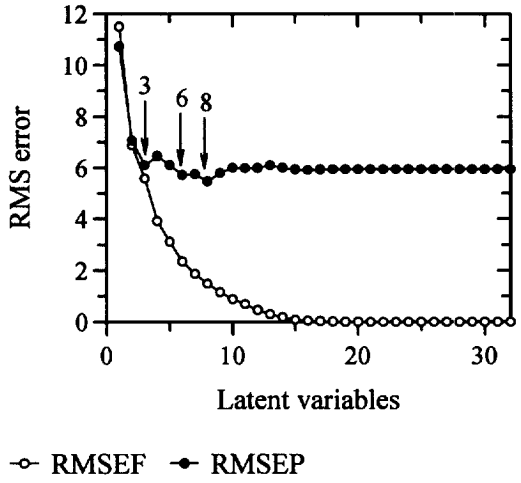

Fig. 2. Effect of the number of PLSl factors on the accuracy of the PLS calibration models used to estimate the percentage lysozyme in glycogen. The open circles represent the RMS error of the data used to create the model (the training set) and the closed circles the RMS from the test set. The arrows indiciate that possible optima were found when 3, 6 or 8 latent variables were used to form the models.

To choose the optimal number of latent variables (PLS factors) to use in predictions after the model was calibrated, the RMS error between the true and desired lysozyme concentrations over the entire calibration model was calculated for the known training set (RMSEF) and the cross-validation set (RMSEP). These RMS errors were then plotted vs. the number of latent variables (factors) used in predictions (Fig. 2). Using this approach, it can be seen that the first minimum is with 3 PLS factors, another minimum was at 6 factors whilst the global was at 8 factors. Therefore, after calibration, the number of PLS factors used in the predictions was 3, 6 and 8; all pyrolysis mass spectra from Machine 2 (before and after correction) were used as the 'unknown' inputs (test data); the model then gave its prediction in terms of the percent lysozyme in glycogen.

2.6. Canonical variates analysis (CVA)

Canonical variates analysis (CVA) is a multivariate statistical technique, here carried out using the GENSTAT package [63], running under MS-DOS 6.2 on an IBM-compatible PC. Before CVA was employed the

following numerical constraint was complied with for a statistically valid analysis [64]:

$$N_v < (N_s - N_g - 1) \quad (1)$$

where N_v is the number of variables (masses) per samples, N_s is the number of samples, and N_g is the number of groups.

Therefore, for 57 samples representing four different groups or taxa this requirement involved reducing the mass spectrum from 150 masses to 51; this was achieved by selecting the 51 most important ones based on their characteristicities [65].

Characteristicity is closely related to the Fisher (F) ratio [66]:

$$F = \frac{\text{Between-group variance}}{\text{Within-group variance}} \quad (2)$$

and has been used to select relevant masses in PyMS spectra for multivariate analysis [2].

Characterisiticity was calculated as described by Eshuis et al. [65]. The first stage is the calculation of the following expressions:

(A) inner variance or reproducibility (r_i) [67]:

$$r_i = \frac{1}{n}\left[\sum_{j=1}^{n} v_{(i,j)}\right] \quad (3)$$

where n is number of duplicate samples and $v_{(i,j)}$ is variance of peak i in sample j.

(B) outer variance or specificity (s_i) [68]:

$$s_i = \frac{1}{n}\left[\sum_{j=1}^{n}(m_{(i,j)} - m_i)^2\right] \quad (4)$$

where n is number of duplicate samples, $m_{(i,j)}$ is mean of peak i in sample j, and m_i is mean of $m_{(i,j)}$ (i.e. mean for all samples of peak i).

The 'characteristicity' (c_i) is then calculated by the following [65]:

$$c_i = \frac{s_i}{r_i}. \quad (5)$$

The mass intensities can then be ranked in order of their characteristicities; large values are more important, smaller ones less so. After this mass selection stage the first 51 most characteristic masses were analysed by CVA.

CVA separated the objects (samples) into groups on the basis of the 51 masses and the a priori knowledge of the appropriate number of groupings [69,70]; this is achieved by minimising the within-group variance and maximising the between-group variance. The *a priori* groups used were the triplicate pyrolysis mass spectra from the five *E. timidum*, four *Eubacterium* C_1, five *Eubacterium* C_2 and five *Eubacterium* New 1 from data collected on Machine 1; the five *Eubacterium* isolates from St Bartholomew's Hospital were used as an external cross-validation set.

The principle of CVA is similar to PCA, but because the objective of CVA is to maximise the ratio of the between-group to within-group variance, a plot of the first two canonical variates (CVs) displays the best 2-D representation of the *group* separation. After CVA was performed on Machine 1 data the unknown test data from Machine 3 were projected into this CVA space.

2.7. Artificial neural networks (ANNs)

All ANN analyses were carried out with a user-friendly, neural network simulation program, NeuFrame version 1,1,0,0 (Neural Computer Sciences, Totton, Southampton, Hants), which runs under Microsoft Windows NT on an IBM-compatible PC. In-depth descriptions of the modus operandi of this type of ANN analysis are given elsewhere [23,25,50].

The algorithm used was standard back-propagation (BP) [6,71] which 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 (f) with a sigmoidal shape, typically the function (f)

$$f = \frac{1}{1 + e^{-x}} \quad (6)$$

where $x = \Sigma$inputs.

For the training of the ANN each input (i.e. normalised pyrolysis mass spectrum from Machine 1) is paired with a desired output (i.e. normalised pyrolysis mass spectrum from Machine 2 or 3); together these are called a training pair (or training pattern). An ANN

is trained over a number of training pairs; this group is collectively called the training set. The input is applied to the network, which is allowed to run until an output is produced at each output node. The differences between the actual and the desired output, taken over the entire training set 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. In the present work, we used a learning rate of 0.1 and a momentum of 0.9. 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.

In addition, the hidden layer (if present) and output layer were connected to the bias (the activation of which is always +1), whose weights will also be altered during training. Before training commenced the values applied to the input and output nodes were normalised between 0 and +1, and the connection weights were set to small random values [8].

2.8. Machine standardization using ANNs

ANN-based methods of machine standardization were employed to transform pyrolysis mass spectra collected on Machine 2 or 3 into those collected previously using Machine 1. This procedure should then allow newly collected mass spectra on Machine 2 or 3 to be directly compared with mass spectra collected using Machine 1 and held in a data base.

Calibration spectra were chosen from the spectra collected on the two machines: (1) for the quantification of lysozyme in glycogen these were the replicate normalised pyrolysis mass spectra containing 0, 25, 50, 75, and 100% lysozyme in glycogen; (2) for the identification of *Eubacterium* species these were the triplicate mass spectra from *E. timidum* ATCC 33093, *Eubacterium* C_1 W1471, *Eubacterium* C_2 SC142, and *Eubacterium* New 1 SC68.

The structure of the ANNs used to correct for drift in this study consisted of an input layer comprising 150 nodes (normalised pyrolysis mass spectra from the calibration samples collected on Machine 2 or 3) connected to the output layer which comprised the mass spectra of the same calibration material analysed on Machine 1; via a single 'hidden' layer containing x nodes; these topologies can be represented as 150-x-150 (and see Fig. 3). Several ANNs architectures were used (details in Section 3) which varied in the number of nodes employed in the hidden layer; 0, 2, 4, 8, 16, 32 and 64 nodes were used.

All ANNs employed the back-propagation algorithm. The input and output layers were scaled to lie between 0 and +0.9 across the 51–200 mass range; +0.9 was chosen as the maximum to allow masses of higher intensity than those used in the training set to be applied to the input layer. ANNs were trained until the average RMS error of 0.1% was reached; the ANNs were interrogated at various points along this training period to check for over-training.

2.9. Machine standardization using linear methods

To compare the performance of these ANN-based corrections with corrections based on linear corrections alone for the lysozyme in glycogen data set, two methods relying on mass-by-mass transformations were also studied. Linear subtractions were used where the amount of drift in each mass was calculated by first subtracting the normalised mass spectrum collected on Machine 1 from the mass spectrum collected on Machine 2, this was done for the calibration samples and the average drift in each mass computed. These drift correction values were then subtracted from each of the masses newly acquired on Machine 2 mass spectra:

$$\text{Linear method 1} = (\text{new mass}) - (\text{average of } (\text{Machine 2 calibration mass} - \text{Machine 1 calibration mass})). \quad (7)$$

The second linear transformation involved calculating the average mass-by-mass ratio between the mass spectra of the calibration samples collected on Machine 1 and Machine 2 (new). These ratios were then used to scale each of the masses in newly acquired mass spectra collected on Machine 2.

$$\text{Linear method 2} = (\text{new mass})^{*}(\text{average ratio of } \text{Machine 1 calibration mass} : \text{Machine 2 calibration mass}). \quad (8)$$

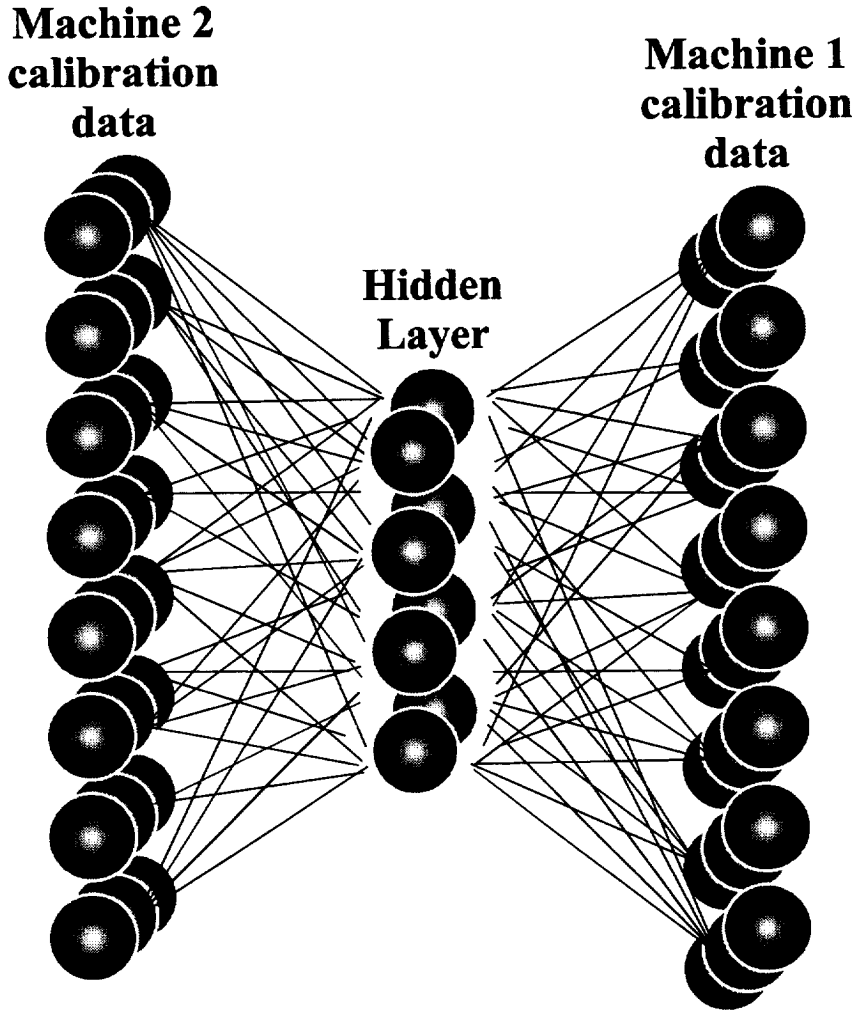

Fig. 3. Architecture of an associative neural network consisting of three layers, trained to transform PyMS spectra from Machine 2 to Machine 1 spectra. In the architecture shown, adjacent layers of the network are fully interconnected. The input and output layer are presented with PyMS data from calibration standards analysed on the two instruments (in this figure there are 24 nodes in these layers; in the present work the number of nodes was actually 150 inputs/masses).

3. Results and discussion

3.1. *Quantification of lysozyme in glycogen; instrument standardization*

Pyrolysis mass spectral fingerprints of 50 μg lysozyme mixed with 50 μg glycogen analysed on Machine 1 from University of Wales, Aberystwyth and the same material analysed at Horizon Instruments (Machine 2) are shown in Fig. 1. Although these mass spectra are complex, as judged by eye, there are some very obvious difference between them; for example the spectrum from the PYMS-200X machine (Fig. 1(a)) has a very intense peak at *m/z* 60, there are other noticeable differences at *m/z* 97, 118 and 145. One way of highlighting any smaller differences between these spectra is simply to subtract one from the other; the resulting subtraction spectrum of the normalised average of three pyrolysis mass spectra of 50 μg lysozyme mixed with 50 μg glycogen

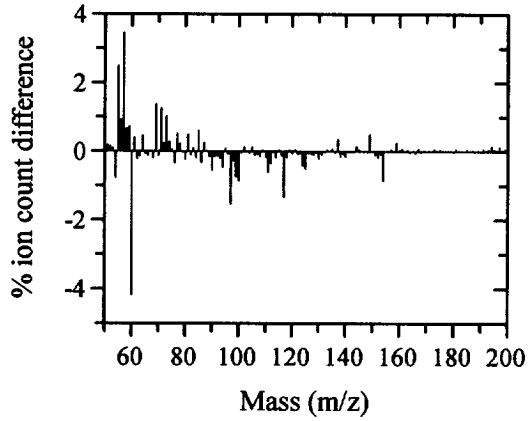

Fig. 4. The subtraction spectrum of the normalised average of three pyrolysis mass spectra of 50% lysozyme mixed with glycogen analysed on Machine 1 (PYMS-200X) (Fig. 1(a)) from the equivalent normalised average spectra from the same material analysed on Machine 2 (RAPyD-400) (Fig. 1(b)). (i.e., subtraction spectrum=Fig. 1(a) and (b)).

analysed on Machine 1 from the equivalent normalised average spectra from the same material analysed on Machine 2 is shown in Fig. 4. The positive half of the graph indicates the peaks that are more intense in the pyrolysis mass spectra collected on the RAPyD-400 instrument (Machine 2), likewise the negative half of the graph indicates masses that are more intense when collected on Machine 1. Fig. 4 indeed shows the same masses indicated above as being characteristic of Machine 1 and also highlights that other masses are different. It is significant that the subtraction spectrum is not monotonic, that is to say there is no obvious trend in the mass spectra of the same material analysed on different mass spectrometers. It is of interest that in our previous studies which investigated instrument drift on a single machine [46] there was an observable trend which resulted in a tip down in the lower *m/z* values and a tip up in high molecular weight fragments. The next stage is therefore to ascertain if these differences due to the different instruments are large enough to be problematic in using PLS models calibrated with data from Machine 1 to give accurate estimates of the percentage lysozyme from pyrolysis mass spectra collected on Machine 2.

Data collected from Machine 1 from mixing lysozyme in glycogen were split into two sets. The calibration set contained the normalized triplicate ion intensities from the pyrolysis mass spectra from 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100% lysozyme in glycogen, whilst the cross validation set contained the 10 'unknown' replicate pyrolysis mass spectra (5, 15, 25, 35, 45, 55, 65, 75, 85 and 95% of the determinand lysozyme in glycogen). We then calibrated PLS models as detailed above with the normalized PyMS data from the calibration sets as the inputs (X-matrix) and the amount of determinand (% lysozyme) mixed in glycogen as the output (Y-matrix). The RMS error in the calibration and cross-validation data are plotted vs. the number of latent variables used in predictions in Fig. 2 where it can be seen that the first minimum is with 3 PLS factors, another minimum was at 6 factors whilst the global was at 8 factors. When the observed vs. known % lysozyme in glycogen were plotted for these three minima (data not shown) it was found that with 3 factors there was a sigmoidal trend and this straightened when 6 and 8 factors were used; that the drop in RMS error between 6 and 8 was small (5.71 to 5.46) 6 was taken to be the best number of latent variables to be used in predictions. PLS models were calibrated with 6 PLS factors and interrogated with the calibration and cross validation sets and a plot of the PLS estimates versus the true % lysozyme in glycogen (Fig. 5) gave a linear fit which was indistinguishable from the expected proportional fit (i.e. $y=x$); the RMS error for the calibration and cross-validation sets were 2.35 and 5.71, respectively (Table 3). It was therefore evident that the network's estimate of the quantity of lysozyme in the mixtures was very similar to the true quantity, both for spectra that were used as the calibration set (open circles) and, most importantly, for the 'unknown' pyrolysis mass spectra (closed circles).

The next stage was to interrogate the PLS model with all the normalised pyrolysis mass spectra of % lysozyme (in 5% steps) in glycogen collected 481 days later on Machine 2. The estimates for these samples are also shown in Fig. 5 (partially shaded squares), where it can be seen that the model's estimate versus the true % lysozyme in glycogen was exceedingly inaccurate. The RMS error in these estimates (Table 3) was 24.1 compared to 4.3 for the same samples analysed on Machine 1. These results clearly show that the pyrolysis mass spectra of the same material had changed significantly when analysed on the two different instruments, thus resulting in an inability to use a PLS model calibrated with data

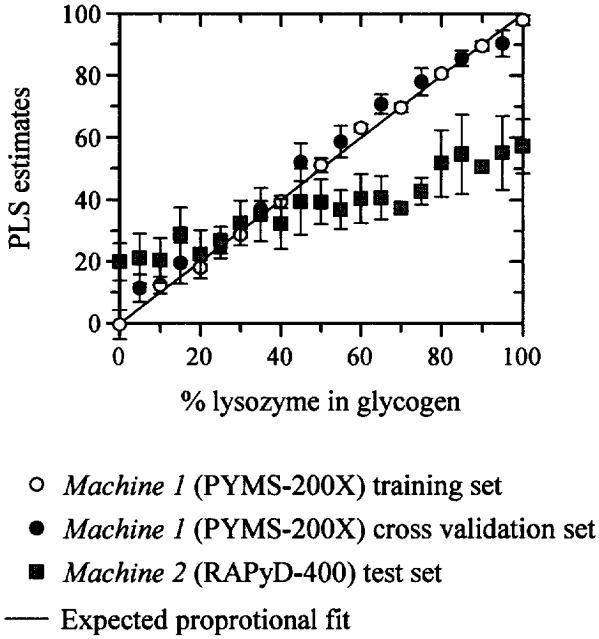

Fig. 5. The estimates from PLS models vs. the true percent lysozyme in glycogen for data collected on Machine 1 (PYMS-200X; UWA) and Machine 2 (RAPyD-400; HI). PLS models were calibrated with PyMS data from Machine 1 using 6 latent variables. The open data points are the averages of the triplicate pyrolysis mass spectra from Machine 1; open circles represent spectra that were used to train the network and closed circles indicate the cross validation set. Partially shaded squares are the averages of the triplicate PyMS spectra collected on Machine 2. Error bars show standard deviation. The expected proportional fit is shown.

collected on Machine 1 to give accurate predictions for data from the same material subsequently collected on Machine 2.

It is therefore necessary to apply a mathematical correction procedure to compare two sets of data of the same material directly. As described in the methods section, standards (calibration spectra) were chosen from the two different data sets and these were the triplicate normalised pyrolysis mass spectra containing 0, 25, 50, 75, and 100% lysozyme in glycogen. These standards were used by the two linear transformations and the various 150-*x*-150 neural networks which varied in the numbers of nodes used in their hidden layers.

Data from Machine 2 were first transformed by the linear methods as detailed above and used to challenge PLS models calibrated with Machine 1 data, using six latent variables, to quantify the % lysozyme in mixtures with glycogen. Fig. 6 shows the estimates of PLS models vs. the true % lysozyme in glycogen for data collected on Machine 2 after correction for instrumental drift by either (a) linear subtraction (open triangles) or (b) a linear mass-by-mass ratio correction (closed triangles). It can be seen clearly that both methods failed to compensate for the differences between the spectra collected on Machine 2 compared with the same collected on Machine 1. Table 3 gives details of the RMS errors between the expected and observed % lysozyme and these were 20.5 and 17.6 for the subtraction and ratio methods respectively, compared with 24.1 for when no correction was applied. That the linear mass-by-mass ratio method gave slightly lower RMS errors than using subtractions of drift and so was better for compensating is because the latter will introduce some negative *m/z* intensities in the transformed spectra; this phenomenon is not

Table 3
Comparison of the average RMS error for quantifying lysozyme in glycogen using PLS models calibrated with 3, 6 or 8 latent variables from data collected on Machine 1 (PYMS-200X). The test sets were data from Machine 1 (PYMS-200X), Machine 2 (RAPyD-400), and two linear methods to transform data collected on Machine 2 to resemble that collected on Machine 1 and hence attempt to compensate for instrument differences

Number of latent variables used for PLS	RMS error[a]		
	3	6	8
Machine 1 – training set	5.57	2.35	1.49
Machine 1 – cross validation set	6.11	5.71	5.46
Machine 1 – all data	5.83	4.29	3.92
Machine 2 – all data	31.88	24.08	21.13
Linear subtraction correction – all data	14.43	20.53	19.60
Linear ratio correction – all data	15.83	17.57	16.93

[a] RMS error is the error between the expected and observed percentages of lysozyme mixed with glycogen.

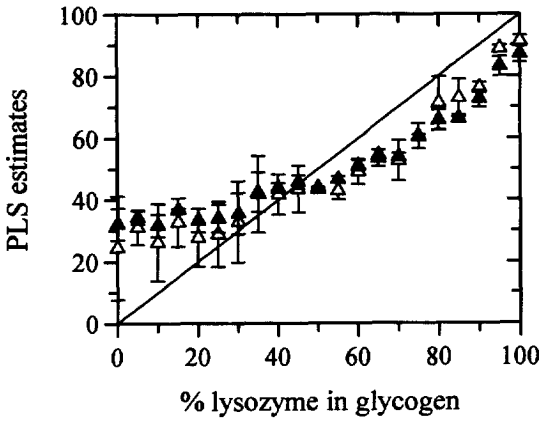

Fig. 6. The estimates from PLS models vs. the true percent lysozyme in glycogen for data collected on Machine 2 (RAPyD-400; HI) after instrument standardization using linear methods. PLIS models were calibrated with PyMS data from Machine 1 using six latent variables. The open triangles are the averages of the triplicate pyrolysis mass spectra after the linear subtraction correction; closed triangles represent spectra after correction using the mass-by-mass ratio correction method. Error bars show standard deviation. The expected proportional fit is shown.

possible with real data and is in this instance a consequence of having to normalise to total ion count.

The next stage was to compare the performance of various neural networks trained to transform mass spectra collected on Machine 2 to those previously collected on Machine 1. Seven different 150-x-150 ANNs, which differed in the number of nodes (x) used in their hidden layers, were employed which were trained using the standard back-propagation algorithm with the normalised triplicate pyrolysis mass spectra containing 0, 25, 50, 75, and 100% lysozyme in glycogen as the inputs from Machine 2 and the outputs from the same calibration standards collected on Machine 1 as the outputs. The following hidden layer sizes were employed which differed in the number of total number of weights in the ANN and are ranked in order of increasing number of weights or complexicity; 2, 4, 8, 16, 32, 64, and one ANN containing no hidden layer (i.e. a 150-150 architecture). Table 4 gives details of the training times, in terms of epochs, to train these ANNs to various RMS error points; this RMS error is defined as the average RMS error for the training set between the observed and expected 150 outputs. It can be seen that ANNs employing only 2 or 4 nodes in the hidden layer failed to reach 0.005 and 0.002 RMS error respectively and this implies that these ANNs which contained 752 and 1354 weights were not complex enough to compensate for the instrumental differences. Indeed, this result might be taken to indicate that there were more than four features (or hyperplanes through 150-dimensional input space) which could describe the effects of the differences between the spectra from the two instruments. This is not surprising given that the subtraction spectra (Fig. 4) displayed no obvious monotonic trends.

ANNs with 8, 16, 32 and 64 nodes were trained to an RMSEF of 0.002 which took longer to train as a function of model complexicity and typically 1.8×10^6, 1.9×10^6, 2.3×10^6, and 3.2×10^6 epochs, respectively (Table 4). The 150-150 ANN which contained 22 650 weights trained very slowly and training was stopped when the RMS error was 0.005; this took 5.0×10^6 epochs. Table 5 gives details of the RMS errors between the expected and observed quantities of % lysozyme after challenging PLS models calibrated with Machine 1 data, using six latent variables, to quantify the % lysozyme in mixtures with glycogen. It can be seen that when 8, 16, 32 and 64 nodes were used in the hidden layer of these 150-x-150 ANNs all gave similar RMS errors; 6.08, 6.44, 6.20 and 7.34, and it was observed that the error increased with model complexicity. The parsimony principle, as described by Seasholtz and Kowalski [62], states that the fewer the number of parameters (or weights) in a calibration model the better, provided there is not a deterioration in predictive accuracy. To satisfy this principle 150-8-150 ANNs were deemed to give the best model with the fewest weights (2448). The 150-8-150 ANN were therefore trained further to 0.001 RMS error which took 5.8×10^6 epochs, and it was observed that the error in PLS predicting % lysozyme increased slightly from 6.08 (at 0.002 RMS error for the mass spectra) to 6.09. It was therefore taken that optimal correction for the difference between instruments was for the 150-8-150 network trained to 0.002 RMS error, this took 1.8×10^6 epochs and the actual time taken

Table 4
Number of epochs taken to reach the indicated average RMS errors in the training set over the 150 masses in the output layer of 150-*x*-150 ANNs trained to transform data collected on Machine 2 to resemble those collected on Machine 1 and hence compensate for instrument differences. Seven ANNs were assessed which differed in the number of nodes in the hidden layer

RMS error[a]	Number of epochs						
No. nodes in hidden layer	0	2	4	8	16	32	64
No. weights in ANN	22650	752	1354	2558	4966	9782	19414
0.05	399	9	9	14	16	19	23
0.025	2.0×10^{3}	251	128	179	213	161	160
0.01	2.3×10^{5}	2.3×10^{4}	2.9×10^{4}	3.1×10^{4}	4.2×10^{4}	5.6×10^{4}	6.6×10^{4}
0.009	4.0×10^{5}	4.2×10^{4}	4.1×10^{4}	4.6×10^{4}	6.3×10^{4}	8.0×10^{4}	9.6×10^{4}
0.008	6.9×10^{5}	8.7×10^{4}	5.9×10^{4}	6.5×10^{4}	9.2×10^{4}	1.2×10^{5}	1.4×10^{5}
0.007	1.2×10^{6}	1.7×10^{5}	8.4×10^{4}	9.5×10^{4}	1.3×10^{5}	1.7×10^{5}	2.1×10^{5}
0.006	2.3×10^{6}	5.7×10^{5}	1.3×10^{5}	1.5×10^{5}	1.9×10^{5}	2.5×10^{5}	3.3×10^{5}
0.005	5.0×10^{6}	3.1×10^{6}[b]	2.5×10^{5}	2.6×10^{5}	2.9×10^{5}	4.1×10^{5}	5.5×10^{5}
0.004	nd	nd	5.2×10^{5}	5.1×10^{5}	6.1×10^{5}	7.0×10^{5}	9.7×10^{5}
0.003	nd	nd	1.1×10^{6}	8.6×10^{5}	1.1×10^{6}	1.2×10^{6}	1.7×10^{6}
0.002	nd	nd	4.6×10^{6}[c]	1.8×10^{6}	1.9×10^{6}	2.3×10^{6}	3.2×10^{6}
0.001	nd	nd	nd	5.8×10^{6}	nd	nd	nd

[a] Average RMS error between the expected and observed 150 outputs of 150-*x*-150 ANNs trained to correct for differences in the mass spectra between Machine 1 and Machine 2.
[b] The average RMS error reached 0.005743 and would not decrease further.
[c] The average RMS error reached 0.002201 and would not decrease further. nd, not determined.

Table 5
Comparison of the average RMS error between that expected and that seen for quantifying lysozyme in glycogen using PLS models calibrated with six latent variables from data collected on Machine 1 (PYMS-200X). The test data were data from Machine 2 (RAPyD-400) after transformation using various ANNs, differing in the number of nodes used in their hidden layers, trained to various stages as indicated in Table 4

RMS error[a]	Average RMS error for prediction of percentage lysozyme						
No. nodes in hidden layer	0	2	4	8	16	32	64
0.05	774.76	470.33	1349.39	396.27	689.61	764.45	328.24
0.025	322.76	86.47	41.19	32.93	180.64	236.65	102.90
0.01	15.98	22.03	20.55	19.19	19.20	15.57	15.08
0.009	12.09	22.18	19.38	17.40	17.16	14.23	13.30
0.008	9.87	18.30	16.99	14.91	15.04	12.98	12.07
0.007	8.60	18.73	14.54	12.68	12.31	11.78	11.24
0.006	7.89	18.86	12.93	11.04	9.64	10.19	10.42
0.005	7.72	18.86	12.13	10.04	8.00	8.72	9.60
0.004	nd	nd	11.38	9.21	7.10	7.80	9.05
0.003	nd	nd	9.79	8.07	6.75	6.99	8.31
0.002	nd	nd	7.44	6.08	6.44	6.20	7.34
0.001	nd	nd	nd	6.09	nd	nd	nd

[a] Average RMS error between the expected and seen 150 outputs of 150-*x*-150 ANNs trained to correct for differences in the mass spectra between Machine 1 and Machine 2. nd, not determined.

was approximately eight days on a 486 DX4, with 24 MB of memory.

After training 150-8-150 neural networks to the above point to transform new mass spectra collected on Machine 2 into those previously collected on Machine 1, the first stage was to observe how similar the transformed mass spectra were to the raw mass spectra of the same material. Fig. 7 shows the

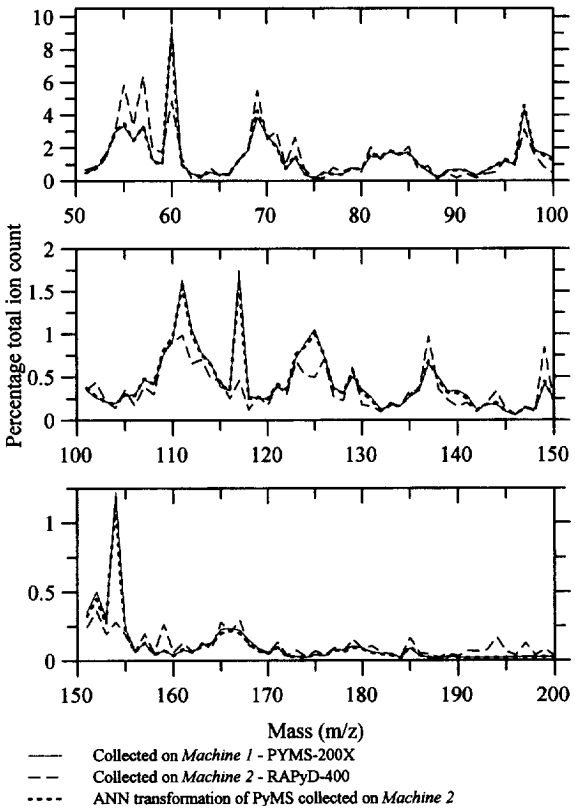

Fig. 7. Normalized pyrolysis mass spectra of 65 : 35 lysozyme : glycogen mixture, analysed on Machine 1 (thin filled line), Machine 2 (broken line), and the spectra from Machine 2 after correction by a 150-8-150 neural network trained to transform data collected on Machine 2 into data collected on Machine 1 (bold broken line). ANNs were trained until the average RMS error of the 150 outputs was 0.002, this took approximately 2×10^6 epochs.

normalized pyrolysis mass spectra of 65% lysozyme mixed with glycogen (chosen because it had not been used to train the neural network) analysed on Machine 1 (thin filled line), Machine 2 (broken line), and the spectra from Machine 2 transformed by a 150-8-150 neural network (bold broken line). It is clear that there is, as expected from Figs. 1 and 4, some difference in the mass spectra collected on the two different instruments, but that the transformed spectrum shows little or no difference compared to the real mass spectra collected on Machine 1; it was therefore evident that this ANN-based correction procedure had indeed compensated for the differences between the mass spectrometers.

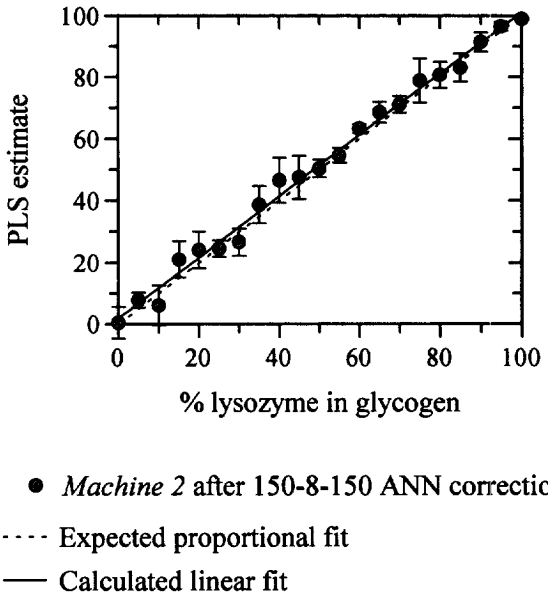

Fig. 8. The estimates from PLS models vs. the true percent lysozyme in glycogen for data collected on Machine 2 (RAPyD-400; HI) after correction using 150-8-150 neural networks. ANNs were trained until the average RMS error of the 150 outputs was 0.002, this took approximately 2×10^6 epochs. Error bars show standard deviation based on replicates. The expected proportional fit and the calculated linear fit are shown. The slope and intercept of the calculated fit are 0.992 and 1.87, respectively.

The next stage was to use these neural network transformed spectra to challenge PLS models calibrated with PyMS data from Machine 1 to quantify the amount of lysozyme in mixtures with glycogen; Fig. 8 shows the estimates of PLS models vs. the true % lysozyme. It can clearly be seen that the neural network-transformed mass spectra gave much better estimates than do the linear transformed spectra (Fig. 6); indeed these gave a linear fit which was indistinguishable from the expected proportional fit (i.e. $y=x$); the error in the estimates was 6.08 and the slope and intercept of the best fit line (dotted line) was 0.992 and 1.87, respectively. It was therefore evident that the PLS estimate of the % lysozyme in glycogen was very similar to the true quantity after correction.

Further to highlight the success of the neural network corrections over the linear ones the transformed mass spectra were analysed with the data collected on Machine 1 and Machine 2 by PCA (Fig. 9); the first plot (Fig. 9(a)) shows the effect on transforming data

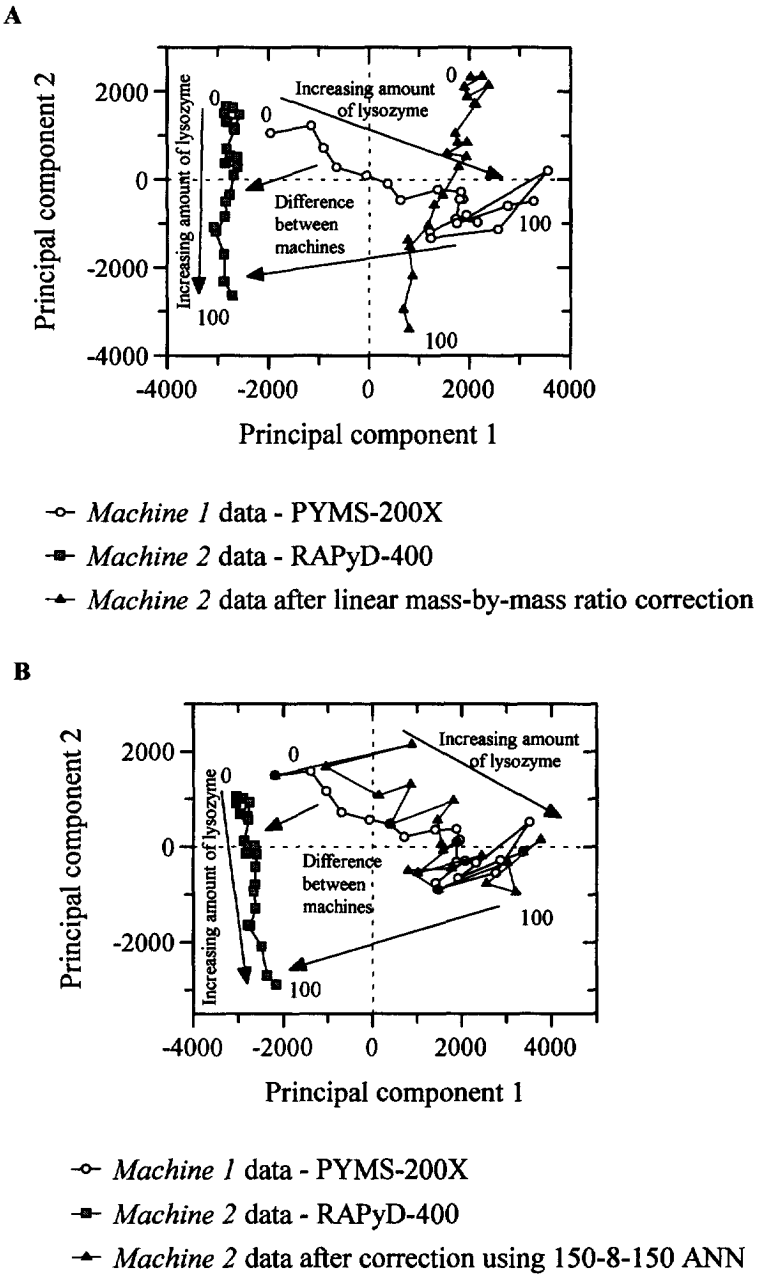

Fig. 9. Principal components analyses plots based on the PyMS data for lysozyme mixed with glycogen for data collected on Machine 1 and Machine 2 compared with either a linear mass-by-mass ratio correction (A) or 150-8-150 neural network (B) correction of the mass spectra collected on Machine 2 transformed to those collected on Macine 1. The first two principal components accounted for 67.8% and 25.3% (93.1% total) of the total variance in plot A and 76.8% and 14.5% (91.3% total) in plot B. In both plots arrows are used to indicate the features in the mass spectra which are extracted by PCA and which account for the increasing amount of lysozyme and for the effect of instrumental differences.

from Machine 2 to Machine 1 using the linear mass-by-mass ratio method whilst the second plot (Fig. 9(b)) shows the effect of a 150-8-150 ANN transformation. In both plots the first PC describes the features in the mass spectra which account for the differences between the two instruments whilst the second PC is controlled by the features in the mass spectra which account for increasing amount of lysozyme. It can be seen that both transformations 'move' the Machine 2 data closer to Machine 1 but that the neural network transformation is more successful because the line from these transformed data (Fig. 9(b)) overlap the line from the data collected on Machine 1 more. It is particularly noteworthy that the linear transformed estimates are parallel with the Machine 2 untransformed data (Fig. 9(a)) whereas the 150-8-150 ANN transformation produces data which map accurately onto the data from Machine 1 (Fig. 9(b)); this may be explained by the ability of the ratio transformation to correct the mass spectra only in a linear manner.

Although these 150-8-150 ANN were very successful for correcting between the two instruments the time taken to train these ANNs was long and typically took eight days; with the current advances in computing technology, particularly in the area of processor power this should not be a problem. However, we have previously observed that *individual* scaling of the input nodes of a 150-8-1 ANN considerably decreased the time taken to train by 100-fold because the range of each input in the population is made equal [72]; this does, however, have the undesired effect that if there is any noise in masses with low intensity then this can affect the ability of ANNs to generalise [73], since models with irrelevant or noisy variables will always tend to perform poorly [74]. Therefore, the effect of scaling the inputs and outputs on the basis of the highest ion counts throughout the entire data set vs. scaling the inputs and outputs of each *m/z* independently over the data set was studied.

In addition to scaling the lower headroom by +0% and the upper headroom by +10% other scaling regimes were set up as detailed in Table 6; all ANNs were trained for 5×10^5 epochs. It can be seen that the various scalings used had little effect on the RMS error between the observed and known values for the concentration of lysozyme, when the input and output layers were scaled across the *whole* mass range. In contrast, if the input and output layers were scaled for *each* input and output node then the RMS errors decreased. Indeed, the lowest RMS error was 5.02 (for PLS models calibrated with six latent variables) and this was when the ANNs were scaled +30% on both the lower and upper headroom, that is to say each *individual* mass was scaled such that the lowest mass was set to 0.3 and the highest mass to 0.7. This suggests that these mass spectra were fairly free of noisy variables and that a reduction in training time was possible; this lowest RMS error (5.02) was achieved in 5×10^5 epochs compared to the best 150-8-150 ANN (Tables 4 and 5), scaled across the mass range, which took 1.8×10^6 epochs (~3.5 times longer) to reach an RMS error of 6.08.

Table 6
Effect of scaling the input and output layer of 150-8-150 ANNs trained to transform data collected on Machine 2 to resemble that collected on Machine 1 and hence compensate for instrument differences. The comparisons are based on the average RMS error for quantifying lysozyme in glycogen using PLS models calibrated with 3, 6 or 8 latent variables from data collected on Machine 1 (PYMS-200X)

Headroom		RMS error[a]	No. latent variables used		
Lower	Upper		3	6	8
Scaling the input and output lalyers across the total range[b]					
0	10	0.0045	9.72	8.71	9.59
0	20	0.0047	10.04	8.47	9.12
0	30	0.0048	11.75	8.53	9.32
10	10	0.0036	5.83	8.51	11.27
20	20	0.0039	8.79	8.37	11.14
30	30	0.0041	9.14	7.92	8.60
Scaling each input and output node individually[c]					
0	10	0.0251	7.33	5.88	6.45
0	20	0.0261	7.95	8.85	10.34
0	30	0.0248	7.26	12.04	9.90
10	10	0.0261	6.50	7.61	7.30
20	20	0.0222	7.09	7.10	6.89
30	30	0.0169	5.27	5.02	4.87

[a] Average RMS error between the expected and seen 150 outputs of 150-8-150 ANNs trained for 0.5×10^6 epochs to correct for differences in the mass spectra between Machine 1 and Machine 2.
[b] The input and output layers were scaled across the *whole* mass range, with the percentage headroom indicated, such that the lowest mass (+ lower headroom) was set to 0 and the highest mass (+ upper headroom) to 1.
[c] The input and output layers were scaled for *each* input node, with the percentage headroom indicated, such that the lowest mass (+ lower headroom) was set to 0 and the highest mass (+ upper headroom) to 1.

In conclusion, PLS can be calibrated with pyrolysis mass spectral data to quantify lysozyme in glycogen. However, these PLS models cannot be used with mass spectra from identical material collected on a different mass spectrometer. 150-8-150 neural networks can be used to compensate successfully for the difference observed in these pyrolysis mass spectra so that PLS models created using the PYMS-200X instrument from UWA (Machine 1) can be used with spectra from the RAPyD-400 instrument from Horizon Instruments (Machine 2). It is likely that this success was due to the ability of ANNs to map non-linearities as well as linearities since linear transformation methods alone could not be used to correct for instrument drift.

3.2. *The identification of Eubacterium species; instrument standardization*

As detailed above CVA was used to separate the bacteria (objects) into four groups on the basis of the 51 most characteristic masses and the *a priori* knowledge of the appropriate number of groupings; the *a priori* groups used were the triplicate pyrolysis mass spectra from the five *E. timidum*, four *Eubacterium* C_1, five *Eubacterium* C_2 and five *Eubacterium* New 1 from data collected on Machine 1 (PYMS-200X, UWA). The five *Eubacterium* isolates from St Bartholomew's Hospital (SBH) were known to belong to the *Eubacterium* C_2 taxon [49] and were thus used as an external cross-validation set, and were projected into this CVA space. All the SBH isolates were found to cluster with the five *Eubacterium* C_2 isolates (data not shown). Now that the model had been tested, the next stage was to project in the 23 mass spectra analysed on Machine 3 (PYMS-200X, Public Health Laboratory, Heath Park); the *E. timidum* W690 strain was not analysed on this instrument because during freezer storage the sample had been lost.

Fig. 10 shows the pseudo-3D CVA plot based on PyMS data from Machine 1 analysed by GENSTAT

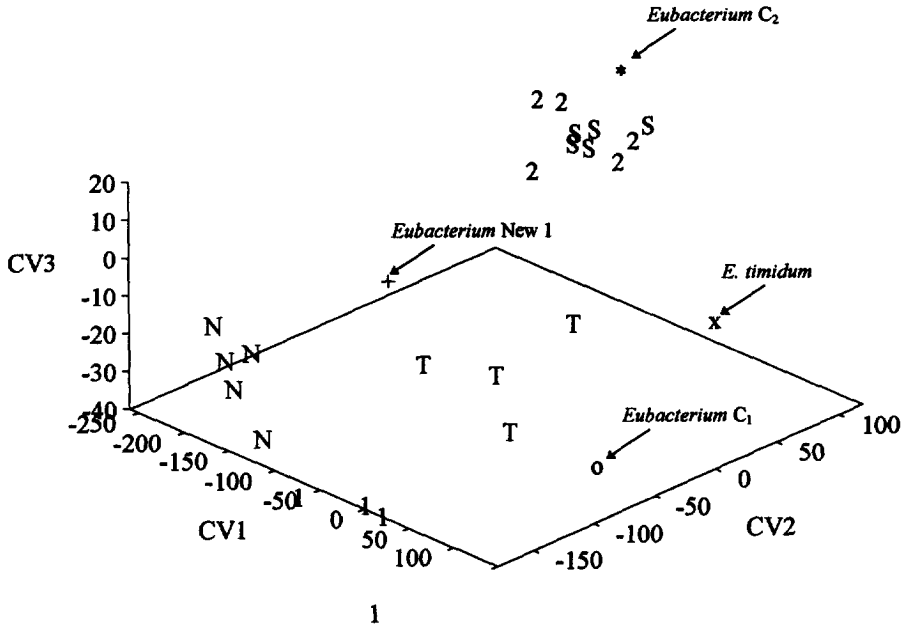

Fig. 10. Pseudo-3D CVA plots based on PyMS data from Machine 1 (PYMS-200X, UWA) analysed by GENSTAT showing the relationship between five *E. timidum* (×), four *Eubacterium* C_1 (○), five *Eubacterium* C_2 (2133*) and five *Eubacterium* New 1 (+). CVA was given the *a priori* information according to the four different *Eubacterium* sp. and calibrated with the 51 most characteristic masses; the symbols refer to the group centroids. The test set was the 23 average pyrolysis mass spectra from Machine 3 (PYMS-200X, PHLS) and were projected into this CV space. The test set are coded as follows; T=*E. timidum*, 1=*Eubacterium* C_1, 2=*Eubacterium* C_2, N=*Eubacterium* New 1, S=*Eubacterium* isolates from Saint Bartholomew's Hospital.

showing the relationship between five *E. timidum* (×), four *Eubacterium* C_1 (○), five *Eubacterium* C_2 (*) and five *Eubacterium* New 1 (+). The test set was the 23 averaged pyrolysis mass spectra from Machine 3 and were projected into the CV space. It can been seen that the only isolates that appeared close to the correct group mean were the five *Eubacterium* C_2 (labelled 2) and the five *Eubacterium* isolates from St Bartholomew's Hospital (labelled S), although if a different angle were used they were in fact quite distant. The *E. timidum* (labelled T), *Eubacterium* C_1 (labelled 1) and the *Eubacterium* New 1 (labelled N) isolates were not projected near their respective group/taxon centres. On closer inspection it may be observed that all isolates from Machine 3 were projected accurately in the first CV but that the second CV was very poorly projected; it may be considered therefore that the failure in CV2 is due to the differences between the two sets of data collected on the two different mass spectrometers.

It is therefore necessary to apply a mathematical correction procedure to compare these two sets of data directly. Calibration spectra from both instruments were chosen as detailed above and in Table 7 and

Table 7
The identities of the bacteria in the test set and training set from Machine 1 (PYMS-200X, UWA) as judged by PLS2 using 4 latent variables calibrated with mass spectral data from Machine 1, compared with new data from Machine 3 (PYMS-200X, PHLS) and that data after transformation using various 150-8-150 ANNs trained for 5×10^5 epochs

Identity	Results from PyMS data from Machine 2				Results from PyMS data from Machine 3				Results after correction using a 150-8-150 ANN			
	T	1	2	N	T	1	2	N	T	1	2	N
E. tim. ATCC3309 3[a]	0.8	0.1	0.2	−0.1	**0.2**	**0.9**	*−0.5*	**0.4**	0.8	0.1	0.2	−0.1
E. tim. W557	1.0	−0.1	0.1	0.0	**0.5**	**0.6**	*−0.3*	**0.2**	**0.5**	*−0.1*	**0.6**	**0.0**
E. tim. W690[b]	1.1	0.0	−0.2	0.1	—	—	—	—	—	—	—	—
E. tim. W693	1.1	0.0	−0.2	0.0	**0.6**	**0.7**	*−0.6*	**0.3**	0.7	0.1	0.2	0.0
E. tim. W2847	1.0	−0.1	0.2	0.0	**0.6**	**−0.1**	**0.2**	**0.3**	**0.2**	**0.2**	**0.4**	**0.1**
Eub. C_1 W1471[a]	0.1	1.0	−0.1	0.0	**−0.3**	**1.5**	**−0.6**	**0.4**	0.1	0.1	−0.1	0.0
Eub. C_1 W687	0.0	0.9	0.0	0.0	**−0.1**	**1.2**	**−0.5**	**0.5**	0.0	0.9	−0.1	0.2
Eub. C_1 W1475	0.0	1.0	0.0	0.0	**−0.2**	**1.1**	**−0.6**	**0.7**	0.0	0.9	−0.1	0.2
Eub. C_1 W1470	0.0	1.0	0.0	0.0	**−0.2**	**1.3**	**−0.6**	**0.5**	0.1	0.9	−0.1	0.1
Eub. C_2 SC142[a]	0.0	0.0	1.0	0.0	0.0	0.2	0.8	0.1	0.0	0.0	1.0	0.0
Eub. C_2 SC108	0.1	0.1	1.0	−0.2	−0.1	0.2	1.0	−0.1	0.0	−0.1	1.1	0.0
Eub. C_2 W1365	0.0	−0.2	1.0	0.2	−0.1	−0.1	0.9	0.2	−0.1	0.0	1.0	0.1
Eub. C_2 W733	0.0	0.0	0.9	0.1	**0.0**	**0.0**	**0.7**	**0.4**	−0.1	0.0	1.0	0.1
Eub. C_2 W2848	0.0	0.0	0.8	0.2	**0.0**	**0.1**	**0.6**	**0.3**	**−0.2**	**−0.1**	**0.9**	**0.4**
Eub. New 1 SC68[a]	0.0	0.0	0.1	0.9	**−0.1**	**0.4**	**−0.4**	**1.0**	0.0	0.0	0.1	0.9
Eub. New 1 SC88P	0.0	0.0	−0.1	1.1	**−0.1**	**0.3**	**−0.5**	**1.2**	0.0	0.0	0.1	0.9
Eub. New 1 SC41B	0.0	0.0	0.0	0.9	**0.0**	**0.4**	**−0.5**	**1.1**	0.0	0.0	0.1	0.9
Eub. New 1 SC37	0.0	0.1	0.0	0.9	**−0.2**	**0.4**	**−0.4**	**1.1**	−0.1	0.1	0.1	0.9
Eub. New 1 SC87K	0.0	0.0	0.1	0.9	0.0	0.2	−0.3	1.1	−0.1	0.0	0.1	0.9
SBH463	0.1	0.2	0.7	0.0	**0.0**	**0.4**	**0.4**	**0.1**	0.0	0.0	0.8	0.2
SBH481	**0.0**	**0.3**	**0.4**	**0.2**	*−0.1*	**0.4**	**0.3**	**0.4**	0.1	0.1	0.7	0.1
SBH462	0.1	0.1	0.7	0.1	−0.1	0.2	0.8	0.1	0.0	−0.1	1.1	0.0
SBH403	0.0	0.1	0.8	0.1	0.0	0.2	0.6	0.2	0.0	0.0	0.8	0.2
SBH477	0.0	0.1	0.7	0.2	**0.0**	**0.1**	**0.6**	**0.3**	−0.1	0.1	0.8	0.2
mis-identified (**bold**)	1/24 or 4.2%				16/23 or 69.6%				3/23 or 13.0%			
mis-identified (*italics*)	0/24 or 0%				5/23 or 21.7%				2/23 or 8.7%			

T=*E. timidum*, 1=*Eubacterium* C_1, 2=*Eubacterium* C_2, N=*Eubacterium* New 1. SBH=*Eubacterium* isolates from Saint Bartholomew's Hospital.

[a] These spectra were also chosen as the calibration samples for drift correction.

[b] Only available for analysis on Machine 1.Values in bold and italics are those *Eubacterium* isolates which were mis-identified: for **bold**—a correct identification was described as the winning position having a value of >0.7 and losing positions <0.3; for *italics*—a correct identification was described as the winning position having the highest value.

these were the triplicate normalised pyrolysis mass spectra from each of the four taxa. These standards (12 spectra) were used to train 150-8-150 ANNs, the input and output nodes were scaled individually such that the lowest mass was set to 0.3 and the highest mass to 0.7 (this was chosen since it was found to be the best for the previous experiment). Training was stopped after 5×10^5 epochs.

The first stage was to carry out PCA to observe the natural relationships between the transformed data using 150-8-150 neural networks with data collected on Machine 1 and Machine 3; the PCA plot is shown in Fig. 11 where it can be observed that different combinations of the first two PCs reflect the effect of instrument differences for each of the taxa (indicated by the arrows). That this drift is not uniform in PC space indicates that the instrument differences are non-linear in nature since PCA carries out only *linear* (orthogonal) transformations of the raw multivariate data. In this PCA plot it can seen that the isolates from *Eubacterium* C_1 (squares), *Eubacterium* C_2 (upward pointing triangles) and *Eubacterium* New 1 (downward pointing triangles), cluster tightly together and that they can be easily separated. In contrast *E. timidum* (circles), although easily separated from the other three taxa do not cluster tightly together. The most important observation from this PCA plot is that the 150-8-150 ANN-transformed mass spectra (partially shaded symbols) overlap with the data collected at Machine 1 (open symbols) and no longer cluster with the data from Machine 3 (closed symbols).

The next stage was therefore to project the neural network-transformed Machine 3 data into the CVA model constructed from mass spectral data collected on Machine 1. Fig. 12 shows the Pseudo-3D CVA plot based on PyMS data from Machine 1 and the transformed test set from Machine 3. It can been seen that the isolates from *Eubacterium* C_1, *Eubacterium* C_2, Saint Bartholomew's Hospital and *Eubacterium* New 1 were all very close to their respective group means from data from Machine 1. The four isolates from *E. timidum* were also moved closer to its group mean but two *E. timidum* W557 and W693 were projected

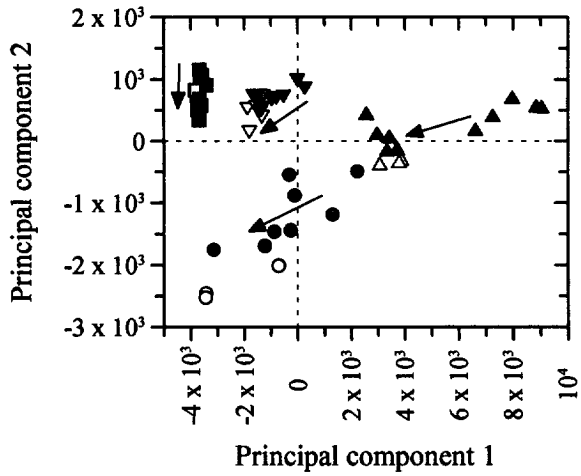

Fig. 11. Principal components analysis plot based on the PyMS data for the *Eubacterium* isolates for data collected on Machine 1 (PYMS-200X, UWA) compared with a 150-8-150 neural network correction of the mass spectra collected on Machine 3 transformed into those collected on Machine 1. The first two principal components accounted for 85.7% and 6.8% (92.5% total) of the total variance. A varying combination of the first and second PCs can be seen to describe the effects of the different instruments (indicated by arrows).

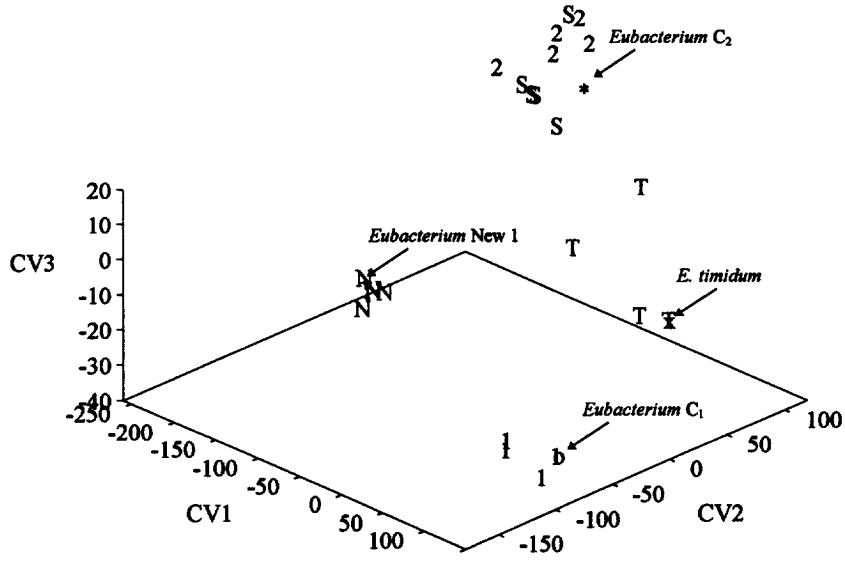

Fig. 12. Pseudo-3D CVA plots based on PyMS data from Machine 1 (PYMS-200X, UWA) analysed by GENSTAT showing the relationship between five *E. timidum* (x), four *Eubacterium* C_1 (○), five *Eubacterium* C_2 (*) and five *Eubacterium* New 1 (+). CVA was conducted as detailed in the text and Figure. 10; the symbols refer to the group centroids. The test set was the 23 average pyrolysis mass spectra from Machine 3 (PYMS-200X, PHLS), after transformation using 150-8-150 neural network corrections, and were projected into this CV space. The test set are coded as follows; T = *E. timidum*, 1 = *Eubacterium* C_1, 2 = *Eubacterium* C_2, N = *Eubacterium* New 1, S = *Eubacterium* isolates from Saint Bartholomew's Hospital.

between the group means for *E. timidum* and *Eubacterium* C_2. Indeed these correspond to the rather loose clustering of the *E. timidum* isolates in the PCA plot (Fig. 11).

Rather than interpret complex 3-D ordination plots a better way to perform identification is to use a supervised learning method that gives a numerical output which can be easily read. Discriminant PLS2 is such an approach and can be exploited to allow easy direct interpretation of the identification of bacteria from their pyrolysis mass spectra. The same 19 replicate mass spectral data from Machine 1 that were used in the above CVA analysis were therefore employed to calibrate PLS models coding the variables in **Y**-matrix as follows; *E. timidum* as 1 0 0 0, *Eubacterium* C_1 as 0 1 0 0, *Eubacterium* C_2 as 0 0 1 0, and *Eubacterium* New 1 as 0 0 0 1.

To choose the optimal number of latent variables to use in predictions, after the PLS model was calibrated the model was challenged with the training set data from Machine 1 and the five replicate spectra from St Bartholomew's Hospital (cross-validation set) using a range of PLS factors; between 1 and 10. It is best to use as few latent variables as possible whilst still obtaining good predictions [17], and it was found that the best model was formed when four PLS factors were used. The results for predictions of the Machine 1 data are shown in Table 7 where it can be seen that the training set were all correctly identified; the results from the five SBH isolates scored highest on the third variable and were thus all identified as belonging to the *Eubacterium* C_2 taxon. This criterion (**A**) was a simple one and was that the identity was related to the winning *Y*-variable; if a second criterion (**B**) was used where a correct identification was taken to be that the winning variable must be >0.7 and <0.3, then the SBH 481 isolated was not classified.

The PLS2 model was then challenged with the mass spectral data from Machine 3 before and after correction using the 150-8-150 ANN transformation method as described above. The results of these two data sets are also shown in Table 7 where it can be seen that for raw untransformed data from Machine 3 using criterion A 21.7% (5 out of 23) were mis-classified and after ANN transformation this dropped to 8.7% and the two *E. timidum* isolates W557 and W693 were mis-iden-

tified as belonging to the *Eubacterium* C_2 taxon. It is noteworthy that these two isolates were also found to be projected away from their group means in CVA space (Fig. 12) and were between *E. timidum* and *Eubacterium* C_2. If the more rigorous criterion **B** for a correct identification was used then before transformation 69.6% (16 out of 23) were mis-identified and after ANN correction this was 13.0% (3 out of 23). The reason this second criterion was used is because it highlights how successful the ANN transformation process has been in causing a more quantized identification; the *Y*-variables for the members belonging to the *Eubacterium* C_2 taxon were very much different from the 0 1 0 0 expected before correction, for example strain W1471 was scored as −0.3 1.5 −0.6 0.4, after ANN correction this was much closer to the real answer and was 0.1 1.0 −0.1 0.0.

In conclusion, projections into CVA space and PLS2 can be calibrated with pyrolysis mass spectral data to identify *Eubacterium* isolated from human oral abscess, and that five isolates from St Bartholomew's Hospital were unequivocally identified as belong to *Eubacterium* C_2. However, these models, calibrated with data from Machine 1, cannot be used to give accurate classifications for mass spectra from the same bacteria collect on a different mass spectrometer (Machine 3) collected 331 days later. 150-8-150 ANNs were used successfully to correct for instrumental differences so that CVA models created using old data from Machine 1 can be used to give accurate isolate identities from newly acquired spectra on Machine 3. It is noteworthy that these isolates were all from oral isolates that had been previously identified as belong to the genus *Eubacterium* and were phenotypically very similar [49]; that these transformation procedures were sensitive enough to corrected for drift was therefore very encouraging.

4. Conclusion

We have shown here and elsewhere that pyrolysis mass spectrometry and multivariate calibration can be used accurately to quantify various mixtures [14,21,23–25] and to identify correctly (micro)biological materials [49,50,75,76]. However, when identical materials were analysed by PyMS on different machines accurate calibration models for (A) the quantitative assessment of the amount of lysozyme in a binary mixture with glycogen and (B) the rapid identification down to the species level of bacteria belonging to the genus *Eubacterium* could not be formed. It was therefore evident that this lack of instrument reproducibility resulted in a lack of calibration transfer.

For PyMS to be used for either the routine identification of microorganisms, or to quantify biological systems, it is paramount that newly acquired spectra from 'slave' machines be compared to those previously collected on a 'master' machine. To correct for the instrumental differences observed calibration samples, or standards, were used which had been analysed on both machines. Two methods relied on linear corrections alone either by subtracting the drift seen in each mass, or by applying a weighting to each new mass based on the average ratio of old calibration mass : new calibration mass. The use of linear correction techniques does, however, assume that the differences are uniform (i.e. linear), which is obviously not the case; therefore, neural network-based transformations were also exploited. ANNs can carry out non-linear as well as linear mappings from the input to the output nodes, and are purported to be robust to the types of noisy data which are often associated with pyrolysis mass spectra [23,73]. In these models the inputs to the ANNs were the new calibration masses from machine 'b' and the outputs were the calibration masses from machine 'a' spectra, several ANNs were employed which contained different numbers of nodes in their hidden layers.

With regard to the quantification of lysozyme in mixtures with glycogen the linear correction methods could not be used to correct for drift (Fig. 6 and Table 3). In contrast the neural network transformation method allowed excellent calibration transfer. This can be determined by observing the lower errors between % lysozyme estimates and known percentages seen in Tables 5 and 6, Fig. 8 and by examining PCA plots of Machine 1, Machine 2 and transformed mass spectra (Fig. 9). Furthermore, the optimal ANN model contained a hidden layer of 8 nodes (150-8-150 architecture) and each input and output node was scaled *individually* to lie between 0.3 and 0.7.

For the identification of human *Eubacterium* isolates 150-8-150 ANNs could also be used to correct

for the drift, so that projections into CVA space from models calibrated with Machine 1 data would give accurate projections for Machine 3 mass spectra. In addition, PLS2 also showed that before correction 16 out of 23 were mis-identified compared to only 3 out of 23 after transformation using neural networks.

In conclusion, neural networks can be used successfully to carry out calibration transfer so that multivariate calibration models created using previously collected data on machine 'a' can be used to give accurate estimates of determinand concentration or bacterial identities (or indeed other materials) from newly acquired spectra on a different instrument when calibrated with standards common to the two data sets.

It should seem obvious that this approach is not limited solely to pyrolysis mass spectrometry but is generally applicable to any analytical tool which is prone to instrumental differences which result in a lack of calibration transfer (which cannot be compensated for by tuning), such as infrared and Raman spectroscopies, nuclear magnetic resonance and gas chromatography, as well as other forms of mass spectrometry.

Acknowledgements

R.G. and E.M.T. thank the Wellcome Trust for financial support (grant number 042615/Z/94/Z). D.B.K. is indebted to the Chemicals and Pharmaceuticals Directorate of the UK BBSRC for financial support.

References

[1] W.J. Irwin, Analytical Pyrolysis: A Comprehensive Guide, Marcel Dekker, New York, 1982.
[2] H.L.C. Meuzelaar, J. Haverkamp and F.D. Hileman, Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials, Elsevier, Amsterdam, 1982.
[3] C.S. Gutteridge, Meth. Microbiol., 19 (1987) 227–272.
[4] J.T. Magee, in M. Goodfellow and A.G. O'Donnell (Eds.), Handbook of New Bacterial Systematics, Academic Press, London, 1993, pp. 383–427.
[5] R. Goodacre and D.B. Kell, Cur. Opin. Biotechnol., 7 (1996) 20–28.
[6] D.E. Rumelhart, J.L. McClelland and The PDP Research Group, Parallel Distributed Processing, Experiments in the Microstructure of Cognition, MIT Press, Cambridge, 1986.
[7] T. Kohonen, Self-Organization and Associative Memory, Springer-Verlag, Berlin, 1989.
[8] P.D. Wasserman, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.
[9] R. Hecht-Nielsen, Neurocomputing, Addison-Wesley, Massachusetts, 1990.
[10] J. Hertz, A. Krogh and R.G. Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley, California, 1991.
[11] S.S. Haykin, Neural networks: a comprehensive foundation, Macmillan, New York, 1994.
[12] C.M. Bishop, Neural networks for pattern recognition, Clarendon Press, Oxford, 1995.
[13] B.D. Ripley, Pattern recognition and neural networks, Cambridge University Press, Cambridge, 1996.
[14] R. Goodacre, M.J. Neal and D.B. Kell, Zbl. Bakt., 284 (1996) 516–539.
[15] H. Martens and T. Næs, in P. Williams and K. Norris (Eds.), Near-Infrared Technology in the Agriculture and Food Industries, Ammerican Association of Cereal Chemists, Inc., St. Paul, Minnesota, 1987, pp. 57–87.
[16] D.M. Haaland and E.V. Thomas, Anal. Chem., 60 (1988) 1193–1202.
[17] H. Martens and T. Næs, Multivariate Calibration, John Wiley, Chichester, 1989.
[18] R.G. Brereton, Multivariate Pattern Recognition in Chemometrics, Elsevier, Amsterdam, 1992.
[19] T.J. McAvoy, H.T. Su, N.S. Wang, M. He, J. Horvath and H. Semerjian, Biotechnol. Bioeng., 40 (1992) 53–62.
[20] S.P. Jacobsson and A. Hagman, Anal. Chim. Acta, 284 (1993) 135–147.
[21] R. Goodacre and D.B. Kell, Anal. Chim. Acta, 279 (1993) 17–26.
[22] R. Goodacre, A.N. Edmonds and D.B. Kell, J. Anal. Appl. Pyrol., 26 (1993) 93–114.
[23] R. Goodacre, M.J. Neal and D.B. Kell, Anal. Chem., 66 (1994) 1070–1085.
[24] R. Goodacre, S. Trew, C. Wrigley-Jones, M.J. Neal, J. Maddock, T.W. Ottley, N. Porter and D.B. Kell, Biotechnol. Bioeng., 44 (1994) 1205–1216.
[25] R. Goodacre, S. Trew, C. Wrigley-Jones, G. Saunders, M.J. Neal, N. Porter and D.B. Kell, Anal. Chim. Acta, 313 (1995) 25–43.
[26] R.C.W. Berkeley, R. Goodacre, R.J. Helyer and T. Kelley, Lab. Prac., 39 (1990) 81–83.
[27] M. Goodfellow, Binary – Comp. Microbiol., 7 (1995) 54–60.
[28] M.J. Whitehouse, J.J. Boon, J.M. Bracewell, C.S. Gutteridge, A.J. Pidduck and D.J. Puckey, J. Anal. Appl. Pyrol., 8 (1985) 515–532.
[29] S.D. Brown, S.T. Sum, F. Despagne and B.K. Lavine, Anal. Chem., 68 (1996) R21–R61.
[30] E. Bouveresse, C. Hartmann, D.L. Massart, I.R. Last and K.A. Prebble, Anal. Chem., 68 (1996) 982–990.
[31] Y. Wang, D.J. Veltkamp and B.R. Kowalski, Anal. Chem., 63 (1991) 2750–2756.
[32] Y.D. Wang and B.R. Kowalski, Anal. Chem., 65 (1993) 1174–1180.

[33] Y.D. Wang, M.J. Lysaght and B.R. Kowalski, Anal. Chem., 64 (1992) 562–564.
[34] Y.D. Wang and B.R. Kowalski, Appl. Spectrosc., 46 (1992) 764–771.
[35] E. Bouveresse and D.L. Massart, Vibr. Spectrosc., 11 (1996) 3–15.
[36] O.E. de Noord, Chem. Intell. Lab. Sys., 25 (1994) 85–97.
[37] Y.D. Wang and B.R. Kowalski, Anal. Chem., 65 (1993) 1301–1303.
[38] Z.Y. Wang, T. Dean and B.R. Kowalski, Anal. Chem., 67 (1995) 2379–2385.
[39] E. Bouveresse and D.L. Massart, Chem. Intell. Lab. Sys., 32 (1996) 201–213.
[40] Z.Y. Wang, T. Isaksson and B.R. Kowalski, Anal. Chem., 66 (1994) 249–260.
[41] E. Bouveresse, D.L. Massart and P. Dardenne, Anal. Chim. Acta, 297 (1994) 405–416.
[42] B.W. Wabuyele and P.de.B. Harrington, Anal. Chem., 66 (1994) 2047–2051.
[43] B.W. Wabuyele and P.de B. Harrington, Appl. Spectrosc., 50 (1996) 35–42.
[44] J.R.M. Smits, W.J. Melssen, M.W.J. Derksen and G. Kateman, Anal. Chim. Acta, 284 (1993) 91–105.
[45] R. Goodacre and D.B. Kell, Patent No. 9607339.0, June 8th 1995.
[46] R. Goodacre and D.B. Kell, Anal. Chem., 68 (1996) 271–280.
[47] W.G. Wade, M.A. Slayne and M.J. Aldred, J. Med. Microbiol., 33 (1990) 239–242.
[48] W.G. Wade, M.A.O. Lewis, S.L. Cheeseman, E.G. Absi and P.A. Bishop, J. Med. Microbiol., 40 (1993) 115–117.
[49] R. Goodacre, S.J. Hiom, S.L. Cheeseman, D. Murdoch, A.J. Weightman and W.G. Wade, Cur. Microbiol., 32 (1996) 77–84.
[50] R. Goodacre, M.J. Neal, D.B. Kell, L.W. Greenham, W.C. Noble and R.G. Harvey, J. Appl. Bacteriol., 76 (1994) 124–134.
[51] C. Chatfield and A.J. Collins, Introduction to Multivariate Analysis, Chapman and Hall, London, 1980.
[52] C.S. Gutteridge, L. Vallis and H.J.H. MacFie, in M. Goodfellow, D. Jones and F. Priest (Eds.), Computer-assisted Bacterial Systematics, Academic Press, London, 1985, pp. 369–401.
[53] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.
[54] D.R. Causton, A Biologist's Advanced Mathematics, Allen and Unwin, London, 1987.
[55] B. Flury and H. Riedwyl, Multivariate Statistics: A Practical Approach, Chapman and Hall, London, 1988.
[56] B.S. Everitt, Cluster Analysis, Edward Arnold, London, 1993.
[57] H. Wold, in K.R. Krishnaiah (Eds.), Multivariate Analysis, Academic Press, New York, 1966, pp. 391–420.
[58] K.A. Martin, Appl. Spectrosc. Revs., 27 (1992) 325–383.
[59] Y.Z. Liang, O.M. Kvalheim and R. Manne, Chem. Intell. Lab. Sys., 18 (1993) 235–250.
[60] Y.Z. Liang and O.M. Kvalheim, Chem. Intell. Lab. Sys., 32 (1996) 1–10.
[61] P.J. Brown, J. Chemom., 6 (1992) 151–161.
[62] M.B. Seasholtz and B. Kowalski, Anal. Chim. Acta, 277 (1993) 165–177.
[63] J.A. Nelder, Genstat Reference Manual, Scientific and Social Service Program Library, University of Edinburgh, 1979.
[64] W.J. Dixon, Biomedical Computer Programs, University of California Press, Los Angeles, 1975.
[65] W. Eshuis, P.G. Kistemaker and H.L.C. Meuzelaar, in C.E.R. Jones and C.A. Cramers (Eds.), Analytical Pyrolysis, Elsevier, Amsterdam, 1977, pp. 151–156.
[66] B.F.J. Manly, Multivariate Statistical Methods: A Primer, Chapman and Hall, London, 1994.
[67] J.B. Kruskal, Psychometrika, 29 (1964) 1–27.
[68] J.B. Kruskal, Psychometrika, 29 (1964) 115–129.
[69] H.J.H. MacFie, C.S. Gutteridge and J.R. Norris, J. Gen. Microbiol., 104 (1978) 67–74.
[70] W. Windig, J. Haverkamp and P.G. Kistemaker, Anal. Chem., 55 (1983) 81–88.
[71] P.J. Werbos, The roots of back-propagation: from ordered derivatives to neural networks and political forecasting., John Wiley, Chichester, 1994.
[72] M.J. Neal, R. Goodacre and D.B. Kell, Proceedings of the World Congress on Neural Networks, San Diego, International Neural Network Society, I, 1994, I318–I323.
[73] R. Goodacre, Appl. Spectrosc., (1996) in press..
[74] D.B. Kell and B. Sonnleitner, Trends Biotechnol., 13 (1995) 481–492.
[75] R. Goodacre, D.B. Kell and G. Bianchi, J. Sci. Food Agric., 63 (1993) 297–307.
[76] R. Goodacre, D.B. Kell and G. Bianchi, Nature, 359 (1992) 594–594.