ELSEVIER

Bioelectrochemistry and Bioenergetics 39 (1996) 185–193

BIOELECTROCHEMISTRY AND BIOENERGETICS

Deconvolution of the dielectric spectra of microbial cell suspensions using multivariate calibration and artificial neural networks

David J. Nicholson *, Douglas B. Kell, Christopher L. Davey

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

Received 28 August 1995; in revised form 25 September 1995

Abstract

Dielectric spectroscopy at radiofrequencies has been widely used for the on-line and real-time estimation of cellular biomass. However, the presence of substantial amounts of non-biomass insoluble solids, such as wheatgerm, may interfere with these measurements in certain industrial media. Dielectric spectroscopy was combined with artificial neural networks (ANNs) to provide an estimation of the cellular biomass present in suspensions of yeast that had been contaminated in some cases with much higher concentrations of wheatgerm, so as to deconvolute the dielectric properties of the mixtures. It was found that an ANN, trained by backpropagation on the dielectric spectra produced by suspensions of varying amounts of yeast and wheatgerm, was able successfully to predict both yeast and wheatgerm content from unseen mixture data. Multivariate statistical methods, such as partial least squares (PLS) and principal component regression (PCR), could also be used successfully to deconvolute such dielectric spectra. It is concluded that such methods provide a powerful adjunct to the conventional quantitative analyses of dielectric data.

Keywords: Dielectric spectroscopy; Artificial neural networks; Biomass estimation; Chemometrics; Process control

1. Dielectric spectroscopy and biomass estimation

There is a continuing need for measurement methods which can enable the on-line and real-time estimation of microbial biomass [1–4]. Dielectric spectroscopy measures the passive or non-faradaic electrical properties of living cells [5–12], and within the β-dispersion frequency range, mainly encompassing radiofrequencies in the region 0.1–10 MHz, it is nowadays widely used to provide an on-line estimate of living biomass in cell suspensions and fermentations of industrial and laboratory interest [13–23]. Whilst every effort is made to stabilize the physical and chemical properties of such systems in order to obtain spectra of the highest possible quality, it is inevitable that fluctuations or noise in the data, produced by aeration and the presence of necromass and other insoluble solids, will tend to distort the spectra obtained. Some industrial media containing particular solid substrates, such as wheatgerm, have been known to cause interferences [24]. Despite this, and a small degree of artifactual distortion which may occasionally be caused by lead inductances at high frequencies and electrode polarization impedance at the low end of the frequency range, the method is normally sufficiently robust to be able to produce reliable and accurate results.

Fig. 1 shows a typical β-dispersion such as that which may be obtained from a suspension of biological cells. The important point to note is that the fall in capacitance (ΔC) measured across the β-dispersion frequency range is directly attributable to the area of electrically polarizable cell membrane present in the suspension and hence the amount of cellular biomass. The capacitance (C in farads) and conductance (G in siemens) of a cell suspension, as measured by a particular dielectric spectrometer, are dependent on the electrode geometry [14], and it is therefore appropriate to normalize and convert capacitance to relative permittivity (ϵ' which is dimensionless) (Eq. (1)) and conductance to conductivity (σ' in S m^{-1}) (Eq. (2)). Thus ΔC, when normalized, becomes $\Delta\epsilon'$ and is known as the dielectric increment

$$\epsilon' = CK/\epsilon_0 \quad (1)$$

$$\sigma' = GK \quad (2)$$

$$K = \sigma'/G \quad (3)$$

The cell constant K (in m^{-1}) (Eq. (3)) is dictated by the electrode geometry and ϵ_0, a constant, is the permittivity of free space (8.854×10^{-12} F m^{-1}).

* Corresponding author. Tel.: +44 1970 623111 ext. 4188; fax: +44 1970 622354; e-mail: djn93@aber.ac.uk.

SSDI 0302-4598(95)01890-5

As explained above, it is the dielectric increment ($\Delta\epsilon'$) of the β-dispersion that is proportional to the concentration of biomass present in a cell suspension. However, for technical or other reasons, it may not be possible to measure the permittivity at frequencies that are very low or very high with respect to the characteristic frequency of the β-dispersion. One solution to this problem is to assume that the dielectric data of interest may be described accurately by the Cole–Cole equations [25], which model changes in the permittivity (Eq. (4)) and conductivity (Eq. (5)) as a function of frequency

$$\epsilon'_\omega = \left[\frac{\Delta\epsilon'\left(1+(f/f_c)^{1-\alpha}\sin(90\alpha)\right)}{1+2(f/f_c)^{1-\alpha}\sin(90\alpha)+(f/f_c)^{2-2\alpha}}\right]+\epsilon'_\infty \quad (4)$$

$$\sigma'_\omega = \left[\frac{-\Delta\sigma'\left(1+(f/f_c)^{1-\alpha}\sin(90\alpha)\right)}{1+2(f/f_c)^{1-\alpha}\sin(90\alpha)+(f/f_c)^{2-2\alpha}}\right] + (\Delta\sigma'+\sigma'_L) \quad (5)$$

The terms used in Eqs. (4) and (5) are as defined above, together with the conductivity terms $\Delta\sigma'$ and σ'_L which are analogous to the permittivity terms $\Delta\epsilon'$ and ϵ'_∞. To obtain the parameters of the Cole–Cole equation from a measured set of dielectric data, it is usual to fit a spectrum using an iterative process known as "curve fitting" [5,16,26–29];. The curve-fitting process is usually carried out by computer programs, typically using the Levenberg/Marquardt algorithm for non-linear, least-squares fitting [5,26,27,30,31]. Such curve fitting is initialized by providing estimates of the equation parameters $\Delta\epsilon'$, f_c, ϵ'_∞ and the Cole–Cole α and a set of permittivity or conductivity data, after which the algorithm is iterated until the curve plotted fits the data with minimum error. So-called "robust weighting" [32,33] is used to minimize the contribution of outlying cases.

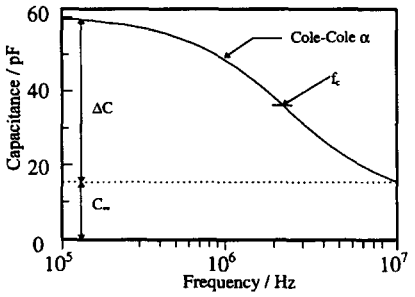

Fig. 1. A typical β-dispersion such as that which may be obtained from a suspension of yeast cells. The high-capacitance, low-frequency plateau, equal to $\Delta C + C_\infty$, is the point at which cell plasma membranes retain the greatest electrical charge. The high-frequency, low-capacitance plateau is the point at which cell polarization and capacitance reach their lowest value C_∞, and any residual capacitance measured here is produced mainly by dipoles of water and other small molecules. This decrease in cell polarization between $\Delta C + C_\infty$ and C_∞ produces an inverted sigmoidal capacitance curve which is characteristic of the β-dispersion. The Cole–Cole α is a measure of how steeply the capacitance or conductance rises with respect to frequency. The characteristic frequency (f_c) of a particular dispersion occurs at the midpoint between ΔC and C_∞, i.e. the frequency at which the capacitance is equal to $(\Delta C/2) + C_\infty$. It is the size of ΔC (pF) which, when normalized and converted to relative permittivity, becomes $\Delta\epsilon'$, known as the dielectric increment, that is directly proportional to the area of charged cell membrane and hence to the amount of living biomass present.

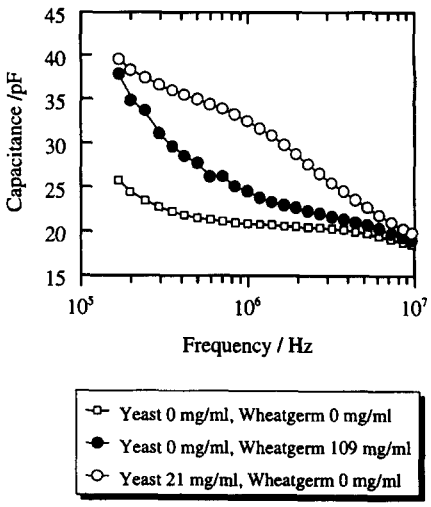

Fig. 2. The dielectric properties of yeast and wheatgerm suspensions. Capacitance measurements were made using a dielectric BM at the frequencies and with the concentrations of yeast and wheatgerm indicated. The cell constant was 0.6 cm^{-1}. When no yeast cells or wheatgerm are present, the apparent increase in capacitance measured at the lower frequencies over that at the higher frequencies is an artifact of the system produced by electrode polarization. Wheatgerm alone exhibits a dielectric dispersion in this frequency range, even though it has no intact cellular membranes and is incapable of producing a classical β-dispersion, so it is apparent that accurate estimations of the yeast cell biomass in this frequency range could be substantially interfered with by the presence of wheatgerm.

Dielectric spectra obtained under normal conditions are generally noise free, and curve fitting is able to produce reliable results (see, for example, Ref. [31]). However, we have noticed special conditions in which interference in the estimation of the radiofrequency dielectric increment of a suspension may occur as a result of the presence of substances which have a substantial α-dispersion that merges with the β-dispersion of interest. Electrode polarization may also contribute problems of this type. Figs. 2 and 3 show some dielectric data that exhibit this general problem. Under these conditions, where there is no known governing equation, non-linear, least-squares fitting can evidently not be applied.

The major virtue of the Cole–Cole equations is that an entire dielectric spectrum may be predicted from data collected over a limited frequency range, although this calculation may be computationally rather intensive [22]. Artificial neural networks (ANNs) can be trained to learn the more significant parameters of the Cole–Cole equation from small sets of simulated dielectric data, and thereby can successfully predict dielectric spectra from sets of unseen data [29]. In the work presented here, ANNs were

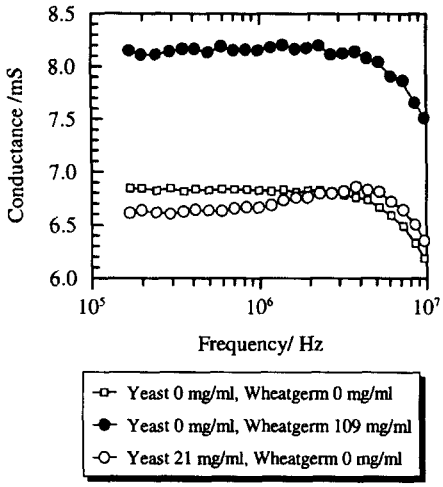

Fig. 3. The dielectric properties of yeast and wheatgerm suspensions. Conductance measurements were made using a dielectric BM at the frequencies and with the concentrations of yeast and wheatgerm indicated. The cell constant was 0.6 cm^{-1}. With wheatgerm only and when no yeast cells or wheatgerm are present, there is no significant dielectric dispersion at the lower frequencies (which if present would be apparent as a decrease in conductance over that at the higher frequencies), whereas yeast cell biomass does exhibit a dispersion in conductivity over this frequency range, especially noticeable between 1 and 4 MHz. The tip down in conductance at the high end of the frequency range is an artifact caused by inductance.

trained with experimental dielectric data. We show that ANNs, and other multivariate statistical methods, are able to deconvolute yeast spectra from interferences produced by the presence of an insoluble solid, wheatgerm, so as to provide accurate dry weight estimates of both yeast and wheatgerm content.

2. Artificial neural networks

Fig. 4 illustrates the architecture of a typical feed-forward ANN or multilayer Perceptron (MLP) [34]. The MLP architecture was inspired by that of biological neurons and consists of a number of neurons or processing units arranged in discrete layers. Each layer of units is connected to the subsequent layer by synaptic connections or weights and it is the adaptation or training of these weights, in accordance with an appropriate training algorithm, which enables the network to "learn". Learning is defined here as the ability to recognize or correctly to classify unseen patterns after a period of training. Activation spreads across the network from the input layer, via the hidden layer, to the output layer, hence the term "feed-forward" network. The hidden layer is so called because its inputs and outputs do not communicate with sources outside the network. Each neuron calculates the weighted sum of its inputs and passes the result through a threshold or activation function as described in Fig. 4. The selection of the number of units in each layer, the nature of the activation function and the type of learning algorithm best suited to the task in hand must be determined empirically. There is a substantial literature which suggests that an MLP together with a supervised learning algorithm is likely to yield good results in this type of quantitative analysis task, since it is known that MLPs of this type can effect the non-linear mapping of arbitrary inputs to arbitrary outputs. In chemical analyses, ANNs have been applied to a variety of problems [35–41], including the deconvolution of pyrolysis mass spectra [42–45]. An overview of the use of ANNs for quantitative chemical analysis, together with the use of multivariate statistical methods may be found in Ref. [46], whilst neural and statistical methods are compared more generally in Refs. [47–50]. Finally, Refs. [51] and [52] provide excellent entrées to the literature on statistical, neural and other machine learning methods.

The biological inspiration which lay behind the early ANN research led to a belief that they should be implemented in a parallel rather than a serial fashion [53], and the term parallel distributed processing (PDP) was coined to describe this. In practice most ANNs are implemented on conventional serial machines and the speed of modern microprocessors is such that their efficiency is not greatly impaired for small problems. The major strength of the PDP programming style, albeit implemented in a serial environment, is its computational simplicity. Curve fitting is a lengthy process which can require much computational overhead, whereas an ANN, if given good training examples, may learn rapidly and once trained yields results "instantaneously". To ensure good generalization (i.e. the

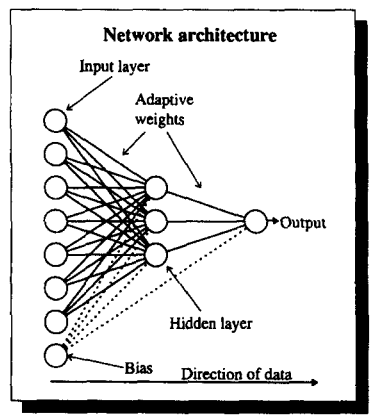

Fig. 4. The architecture of a classical feed-forward ANN. Each of the open circles represents a neuron or processing unit. The units are arranged in three layers: input, hidden and output. Each layer is linked to the next by synapses or weights W and each weight is modified during training until the desired output is achieved. When the network is interrogated, input is presented at the input layer and the activation of each of the units in the hidden and output layers is calculated. The activation x of a unit is defined as the weighted sum or dot-product of its inputs X_i and weights W_i (Eq. (6)) passed through an activation function, $f(x) = 1/[1 + \exp(-(x/g))]$, which is usually sigmoidal and in which g, a gain term, is usually unity (as herein). The function of the bias unit is akin to that of a threshold in that it provides a continuous positive input which allows the hyperplane which the network's weights are representing to move away from the origin.

ability to give correct results for data different from those used to train the network), the data used to train a network are separated into two groups known as the training and the test set. The learning algorithm used here is backpropagation (BACKPROP) which was developed independently by several research groups and individuals; a good description of the algorithm may be found in Refs. [53] and [54] and the overview of Ref. [55]. BACKPROP is a supervised algorithm and this means that the training set comprises "training pairs", i.e. a training example and its known output, the training signal. Each training example is the vector of values representing the complete dielectric spectrum at the frequencies used, and each training signal is the dry weight of yeast or wheatgerm present when that spectrum was obtained. An effective training set should comprise a representative sample of the whole input space so that the network is able correctly to classify all the unseen examples in the test set without the need for extrapolation.

The training set, in the form of numerical data scaled between 0 and 1, is presented to the input layer and fed forward to the next so-called "hidden" layer via a set of connections. Each connection has an associated multiplication factor or "weight" (see Fig. 4). Input to each of the neurons in the hidden and output layer is determined by calculating the weighted sum or dot-product of the input vector multiplied by the weight vector (Eq. (6))

$$x = \sum_{i=1}^{n} W_i X_i \tag{6}$$

W_i is the weight on the ith connection and X_i is the input. Having calculated its input, each neuron then applies a transfer or threshold function, usually sigmoidal (Eq. (7)), to that input, the result of which is passed on to the next layer

$$o_j = \frac{1}{1 + e^{-(x/g)}} \tag{7}$$

o_j is the output of the jth node, x is the input to that node and g is a gain term (generally unity). An additional input and associated weight known as a "bias" is also applied to the hidden and output layers. In order to be able to classify unseen data correctly, it is necessary for the network to form a multidimensional internal representation of the training data. This representation may be thought of as a hyperplane and the addition of the bias promotes the network's ability to move that hyperplane away from the origin.

The network is initialized by setting the weights to small random numbers between 0 and 1, after which the training set is presented seriatim as training pairs, i.e. known input together with known output. Training is achieved by adjusting the weights in the network until it is able correctly to classify the training examples and generalize that classification to an unseen test set. One pass of the complete training set is known as an epoch. The backpropagation training algorithm works in two passes: the forward pass, in which an error signal, which is proportional to the difference between the actual and the desired output, is calculated at the output layer, and the backward pass in which the weights in both layers are adjusted to minimize that error. The error signal generated by the network may be envisaged as an "error surface" containing peaks and troughs, and the object of the training process is to find the global minimum on that surface, without becoming trapped in any of the local minima; this process is known as gradient descent. In an attempt to overcome the problem of becoming trapped in local minima, an extra gain term or momentum is included in the learning algorithm. The magnitude of the momentum term ensures that rapid changes in weights, which tend to occur in the early stages of training, are carried over to subsequent epochs. The way in which the training data are presented to the network can also affect how easily the error reaches a true global minimum, and for this reason it is always better to present the training set members in random order at each epoch; this procedure is known as training by pattern. Training by pattern also helps to prevent oscillations in the error signal.

Having chosen a suitable training and test set and the appropriate learning algorithm, it is necessary to know when the network is trained. A plot of the error vs. the epoch number is known as a learning curve. Initially, during training, a rapid fall in error is recorded for both the training and test set examples until both error signals converge; after convergence the two signals remain at approximately the same level until the training set error decreases and the test set error increases. When this divergence takes place the network is said to be overtrained, and it is at an empirically determined point during the convergence period that most networks are able to make their most accurate predictions. A network is said to be trained when it is able to generalize successfully from the training set to the test set, i.e. from seen to unseen data. Fig. 5 shows the root mean square (RMS) error learning curves obtained from an ANN being trained to predict the concentration of yeast from dielectric data generated from the mixtures described in Table 1. The network used 50 input units, 3 hidden units and 1 output unit (50–3–1 architecture) and was trained with the standard BACKPROP algorithm. This process of checking a network's performance on unseen test set data during training is a form of cross-validation.

3. Materials and methods

Stock suspensions contained Baker's yeast *Saccharomyces cerevisiae* as the cell suspension, obtained locally as a paste, and wheatgerm ("natural" wheatgerm, W. Jordan Cereals, Biggleswade, UK) as the interfering non-cellular insoluble solid. The yeast was suspended in 10

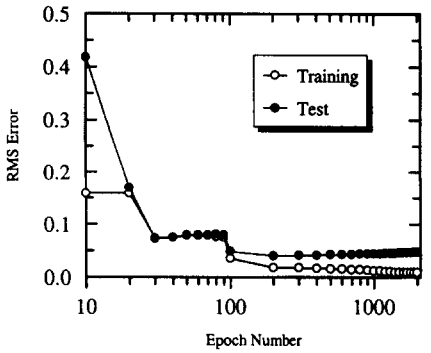

Fig. 5. Learning curves from ANN training to predict yeast cell concentrations from the dielectric data obtained on the mixtures described in Table 1. The network used a 50–3–1 architecture and was trained using standard BACKPROP. The inputs were the capacitance and conductance data at the 25 frequencies shown in Fig. 2. The recorded RMS error for both the training set (seen) and test set (unseen) data drops rapidly during the first 30 or so epochs of training until both error rates converge and remain superimposed one upon the other. The two error signals stay approximately equal for another 50 epochs whereupon they begin to diverge as the test set error increases and the training set error decreases. A network is said to be trained when it is able to generalize successfully from the training set to the test set, i.e. from seen to unseen data, and the process of checking a network's performance on unseen data is a form of cross-validation. When the error plots begin to diverge and it is apparent that the network will only classify the training set correctly, the network is said to be overtrained. The software used was a commercial package (NeuDesk, Neural Computer Sciences, Unit 3, Lulworth Business Centre, Nutwood Way, Totton, Southampton, SO4 3WW, UK) running on a PC.

mM KH_2PO_4 (pH 6.5) buffer, washed twice by centrifugation to remove any starch/packing material and then resuspended in the same buffer; the wheatgerm was also suspended in the same medium. Suspensions of different conductivities were obtained by the addition of a fixed quantity of concentrated KH_2PO_4, determined beforehand in a control experiment, to provide an increase of 2 mS cm^{-1} (per addition) when measured at 1 MHz. Increasing the conductivity of the suspending medium causes the characteristic frequency (f_c, Fig. 1) of the β-dispersion to rise and a decrease causes it to fall, as described in Ref. [21]. It should be noted that this movement of f_c caused by changes in the conductivity of the suspending medium does not affect the magnitude of $\Delta\epsilon'$, the dielectric increment, and hence the viability of dielectric biomass determination.

Table 1
The concentrations of yeast and wheatgerm used in the training and test sets

Training set		Test set	
Yeast (mg ml^{-1})	Wheatgerm (mg ml^{-1})	Yeast (mg ml^{-1})	Wheatgerm (mg ml^{-1})
0.00	0.00	5.3	13.65
0.00	27.30	5.3	40.95
0.00	54.60	5.3	68.25
0.00	81.90	5.3	95.55
0.00	109.20	5.3	122.85
0.00	136.50	15.9	13.65
10.60	0.00	15.9	40.95
10.60	27.30	15.9	68.25
10.60	54.60	15.9	95.55
10.60	81.90	26.5	13.65
10.60	109.20	26.5	40.95
21.20	0.00	26.5	68.25
21.20	27.30	37.1	13.65
21.20	54.60	37.1	40.95
21.20	81.90	47.7	13.65
31.80	0.00		
31.80	27.30		
31.80	54.60		
42.40	0.00		
42.40	27.30		
53.00	0.00		

As described above, the ANN used here requires the construction of a training and test set (from dielectric data) such that the training set used completely covers the range to be quantified. Table 1 shows the dry weight (in mg ml^{-1}) of yeast and wheatgerm present for each member of the training and test sets. The dry weights were obtained by uniform sampling and drying overnight at 100 °C. The training set was obtained from dielectric measurements of 21 different mixtures of the yeast and wheatgerm stock solutions and the test set from 15. Prior to mixing, the two stock solutions were kept agitated at a constant speed and temperature, it being particularly important to keep the wheatgerm in suspension and not to allow it to stick to the bottom or sides of the beaker.

The dielectric spectrometer used here was a model 214A biomass monitor (BM) manufactured by Aber Instruments Ltd., Science Park, Aberystwyth, SY23 3AH, UK, the data from which were logged by a PC running MINISCAN [56] software. An individual scan consisted of measurements of the capacitance (pF) and conductance (mS) made at each of 25 frequencies over the β-dispersion frequency range of 0.2–10 MHz, whose order was randomly chosen. Ten replicate measurements were made at each of the 25 frequencies and the mean was calculated and recorded by the computer; replicates were made in order to compensate for any momentary fluctuations in capacitance/conductance which may have been produced by bubbles or the physical impact of particulate matter on the electrodes. The cell constant of the BM/electrode combination was 0.6 cm^{-1}.

4. Results and discussion

We have demonstrated that it is possible to train an ANN with dielectric data collected on varying concentrations of yeast and an insoluble solid, wheatgerm. The dielectric data were the capacitance and conductance values recorded at a number of frequencies within the β-dispersion frequency range of 0.2–10 MHz. Once trained, the ANN was able to predict the concentration of yeast and wheatgerm when interrogated with a novel set of dielectric data. The learning curves plotted for an ANN being trained

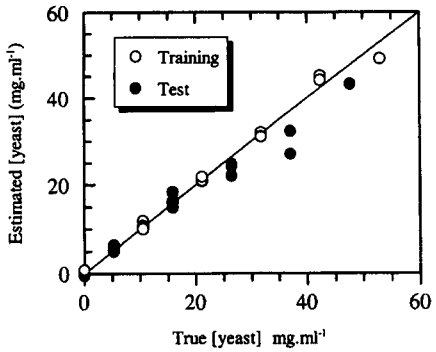

Fig. 6. Estimated yeast cell concentrations from the dielectric data of mixtures of yeast and wheatgerm as in Table 1. The training run was that shown in Fig. 5, after 1000 epochs. The line is the line of identity and the correlation coefficients of the estimates for the training set and the test set data to this line are 0.99 and 0.98 respectively.

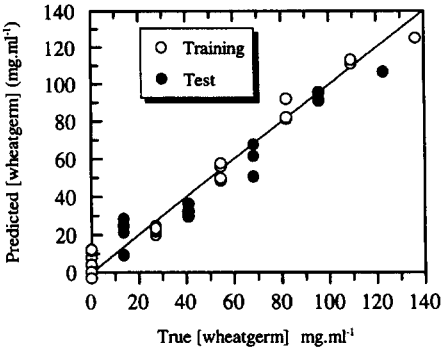

Fig. 8. Estimated wheatgerm concentrations from the dielectric data of mixtures of yeast and wheatgerm as in Table 1. The training run was that shown in Fig. 7, after 1000 epochs. The line is the line of identity and the correlation coefficients of the estimates for the training set and the test set data are 0.95 and 0.82 respectively.

to predict yeast concentration from dielectric data measured on the mixtures in Table 1 are shown in Fig. 5. The network architecture comprised 50 input units, 3 hidden units and 1 output unit. The number of network inputs was dictated by the number of frequencies (25) at which experimental measurements of the capacitance and conductance were made. A single output unit provided the quantity of yeast present. The appropriate number of hidden units was determined experimentally in the range 0–5, three proving to be optimal as judged by the RMS error on the test set and the network's ability to generalize. The network performed well (see Fig. 6) and clearly demonstrated an ability to deconvolute the yeast signal from that of wheatgerm. An ANN with the same architecture, but trained for wheatgerm content, also performed well, although it learned somewhat more slowly, giving the learning curves shown in Fig. 7 and the results shown in Fig. 8.

Figs. 9–12 show the learning curves and results for an ANN with 25–3–1 architecture, i.e. networks trained to predict yeast and wheatgerm content with capacitance data only. The difference between the correlation coefficients of the estimated and true values for the test and training sets for networks trained on capacitance and conductance data

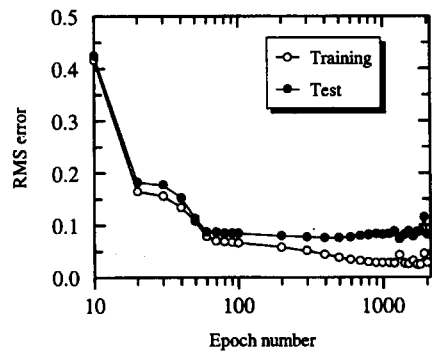

Fig. 7. Learning curves from ANN training to predict wheatgerm concentrations from the dielectric data obtained on the mixtures described in Table 1. The network used a 50–3–1 architecture and was trained using standard BACKPROP.

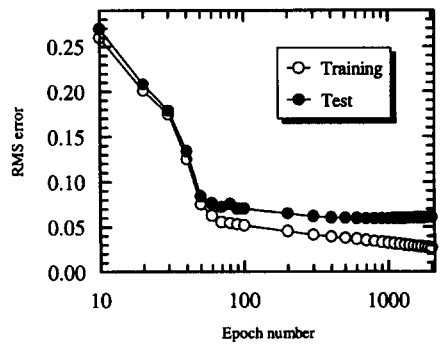

Fig. 9. Learning curves for ANN being trained to predict yeast concentration from the dielectric data obtained on the mixtures described in Table 1. The network used a 25–3–1 architecture and was trained using the standard BACKPROP algorithm. The experiment was the same as that in Fig. 5 except that only the capacitance (no conductance) data were used.

and those trained with capacitance data alone is not very great; this leads to the conclusion that under the current experimental conditions capacitance data alone are sufficient for the quantification task in hand.

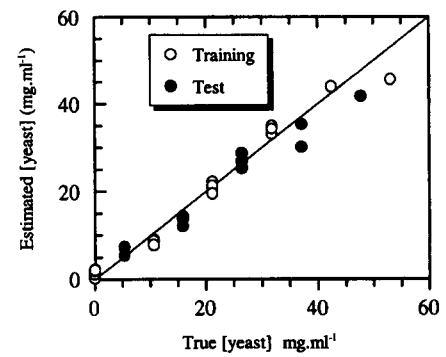

Fig. 10. Estimated yeast cell concentrations from the dielectric data of mixtures of yeast and wheatgerm as in Table 1. The neural net run was that shown in Fig. 9, and the values are those after 500 epochs. The experiment was the same as that in Fig. 5 except that only capacitance (no conductance) data were used. The line is the line of identity and the correlation coefficients of the training set and the test set data to this line are 0.96 and 0.84 respectively.

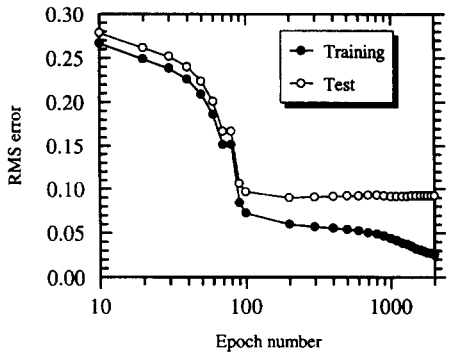

Fig. 11. Learning curves for ANN being trained to predict the concentration of wheatgerm from the dielectric data obtained on the mixtures described in Table 1. The network used a 25–3–1 architecture and was trained using the standard BACKPROP algorithm. The experiment was the same as that in Fig. 5 except that only capacitance (no conductance) data were used.

In order to compare the ANNs performance with that of conventional multivariate statistical regression techniques, we used two methods: partial least squares (PLS) and principal component regression (PCR) [57]. PLS was used to predict the dry weight of yeast (Figs. 13 and 14) and wheatgerm (Figs. 15 and 16). PCR produced very similar results (not shown). Both of these methods utilize a calibration stage to form a model from a representative sample of the data to be described and employ cross-validation as a measure of performance. The calibration and cross-validation processes used with PCR and PLS may be considered to be equivalent to the use, with an ANN, of a supervised learning algorithm, such as BACKPROP, in conjunction with cross-validation by unseen test data as described in Section 2 (see Ref. [57] for statistical procedures and Ref. [46] for ANN techniques). The PLS (and PCR) procedures were also capable of providing a robust model for the deconvolution of dielectric spectra and, at least at these concentrations (cf. Ref. [31]), the contribu-

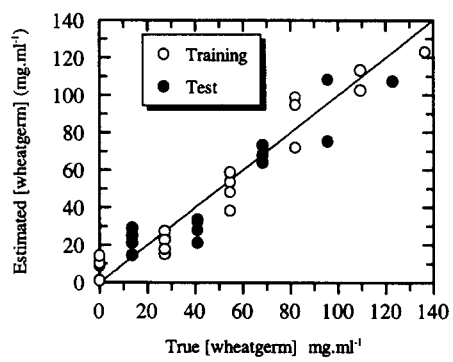

Fig. 12. Estimated wheatgerm concentrations from the dielectric data of mixtures of yeast and wheatgerm as in Table 1. The neural net run was that shown in Fig. 11 and the values are those after 200 epochs. The experiment was the same as that in Fig. 5 except that only capacitance (no conductance) data were used. The line is the line of identity and the correlation coefficients for the training and test sets are 0.97 and 0.94 respectively.

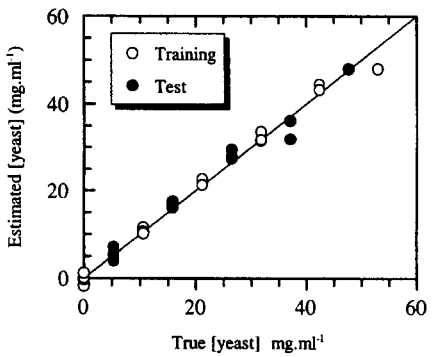

Fig. 13. Estimated yeast cell concentrations based on a PLS model formed using full leave-one-out cross-validation (the optimal model shown contained two PLS factors). Both capacitance and conductance data were employed as in Fig. 5 (dielectric data obtained on the mixtures described in Table 1). The line is the line of identity and the correlation coefficients for the training and test sets are both 0.99. The software used was a commercial package (Unscrambler II, manufactured by CAMO A/S, Olav Tryggvasonsgt. 24, N-7011 Trondheim, Norway) running on a PC.

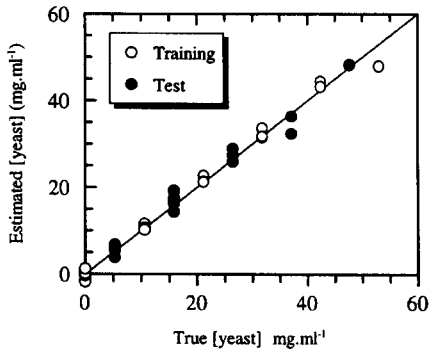

Fig. 14. Estimated yeast cell concentrations based on a PLS model formed using full leave-one-out cross-validation (the optimal model shown contained two PLS factors). Capacitance (not conductance) data were employed as in Fig. 9 (dielectric data obtained on the mixtures described in Table 1). The line is the line of identity and the correlation coefficients for the training and test sets are both 0.99.

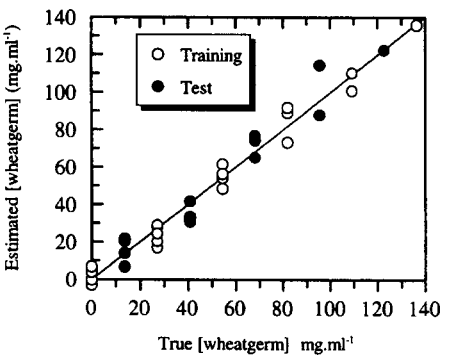

Fig. 15. Estimated wheatgerm concentrations based on a PLS model formed using full leave-one-out cross-validation (the optimal model shown contained two PLS factors). Both capacitance and conductance data were employed as in Fig. 5 (dielectric data obtained on the mixtures described in Table 1). The line is the line of identity and the correlation coefficients for the training and test sets are 0.99 and 0.97 respectively.

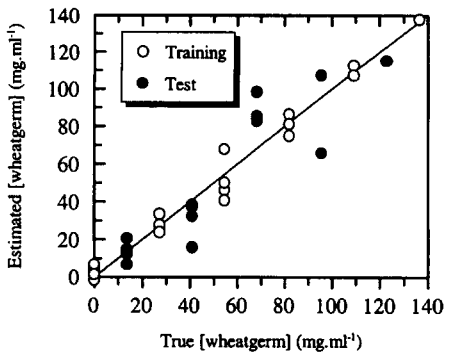

Fig. 16. Estimated wheatgerm concentrations based on a PLS model formed using full leave-one-out cross-validation (the optimal model shown contained two PLS factors). Capacitance (not conductance) data were employed as in Fig. 9 (dielectric data obtained on the mixtures described in Table 1). The line is the line of identity and the correlation coefficients for the training and test sets are 0.99 and 0.91 respectively.

tions of yeast and wheatgerm to the dielectric spectra were linearly separable. Such linear multivariate techniques also have the advantage (relative to the non-linear ANN approach) that they may more easily be interrogated so as to obtain information about the mechanism by which the model serves to fit the observable data. In particular, it is possible to establish from the so-called loadings plots [57] the relative significance of each of the input variables to the model formed. When this was done (data not shown) it transpired, as foreshadowed in the analysis presented earlier, and consistent with it, that the conductance data were indeed of little significance to the multivariate linear model formed in the PLS procedure.

In conclusion, we have shown that the application of multivariate statistical procedures and ANNs allows the successful deconvolution of dielectric spectra for the estimation of microbial biomass under conditions in which substantial interferents are present and where a univariate, single or differential frequency measurement [17] would give highly inaccurate estimates. An ability to determine the concentration of an interferent, in cases where such substances are used as substrates and decrease during the course of a fermentation, could be of great benefit in process control.

Acknowledgements

We are indebted to the Chemicals and Pharmaceuticals Directorate of the UK Biotechnology and Biological Sciences Research Council for financial support of this work, and to Aber Instruments Ltd., Aberystwyth for their contribution to a CASE Studentship for D.J.N. We thank Dr. Royston Goodacre for assistance with the implementation of neural networks.

References

[1] C.M. Harris and D.B. Kell, The estimation of microbial biomass, Biosensors, 1 (1985) 17–84.
[2] D.J. Clarke, B.C. Bake-Coleman, R.J.G. Carr, M.R. Calder and T. Atkinson, Trends Biotechnol., 4 (1986) 173.
[3] B. Sonnleitner, G. Locher and A. Fiechter, Biomass determination, J. Biotechnol., 25 (1992) 5–22.
[4] D.B. Kell and B. Sonnleitner, GMP—Good modelling practice, Trends Biotechnol., submitted for publication.
[5] E.H. Grant, R.J. Sheppard and G.P. South, Dielectric Behaviour of Biological Molecules in Solution, Clarendon Press, Oxford, 1978.
[6] O.F. Schanne and E.R.P. Ceretti, Impedance Measurements in Biological Cells, Wiley, New York, 1978.
[7] R. Pethig, Dielectric and Electronic Properties of Biological Materials, Wiley, Chichester, 1979.
[8] K.R. Foster and H.P. Schwan, Dielectric properties of cells and tissues: a critical review, CRC Crit. Rev. Biomed. Eng., 17 (1989) 25–102.
[9] R. Pethig and D.B. Kell, The passive electrical properties of biological systems: their significance in physiology, biophysics and biotechnology, Phys. Med. Biol., 32 (1987) 933–970.
[10] S. Takashima, Electrical Properties of Biopolymers and Membranes, Adam Hilger, Bristol, 1989.
[11] S. Bone and B. Zaba, Bioelectronics, Wiley, Chichester, 1992.
[12] C.L. Davey and D.B. Kell, The low-frequency dielectric properties of biological cells, in D. Walz, H. Berg and G. Milazzo (Eds.), Bioelectrochemistry: Principles and Practice, Vol. 6, 1995, 159–207.
[13] C.M. Harris, R.W. Todd, S.J. Bungard, R.W. Lovitt, J.G. Morris and D.B. Kell, The dielectric permittivity of microbial suspensions at radio frequencies; a novel method for the real-time estimation of microbial biomass, Enz. Micr. Technol., 9 (1987) 181–186.
[14] D.B. Kell, The principles and potential of electrical admittance spectroscopy. An introduction, in A.P.F. Turner, I. Karube and G.S. Wilson (Eds.), Biosensors; Fundamentals and Applications, Oxford University Press, Oxford, 1987, pp. 427–468.
[15] C.A. Boulton, P.S. Maryan and D. Loveridge, The application of a novel biomass sensor to the control of yeast pitching rate, Proc. European Brewing Convention Congress, Zurich, 1989, pp. 653–661.
[16] (a) D.B. Kell and C.L. Davey, Conductimetric and impedimetric devices, in A.E.G. Cass (Ed.), Biosensors. A Practical Approach, Oxford University Press, Oxford, 1990. (b) D.B. Kell, G.H. Markx, C.L. Davey and R.W. Todd, Real-time monitoring of cellular biomass: methods and applications, Trends Anal. Chem., 9 (1990) 190–194.
[17] G.H. Markx and D.B. Kell, Dielectric spectroscopy as a tool for the measurement of the formation of biofilms and of their removal by electrolytic cleaning pulses and biocides, Biofouling, 2 (1990) 211–227.
[18] (a) G.H. Markx, C.L. Davey and D.B. Kell, To what extent is the magnitude of the Cole–Cole α of the β-dispersion of cells explicable in terms of the cell-size distribution?, Bioelectrochem. Bioenerg., 25 (1991) pp. 195–211. (b) G.H. Markx, C.L. Davey and D.B. Kell, The permittistat: a novel type of turbidostat, J. Gen. Microbiol., 137 (1991) 735–743.
[19] R. Fehrenbach, M. Comberbach and J.O. Pêtre, On-line biomass monitoring by capacitance measurement, J. Biotechnol., 23 (1992) 303–314.
[20] C.L. Davey, The Biomass Monitor Source Book, Aber Instruments Ltd., Aberystwyth, Dyfed, 1993.
[21] C.L. Davey, The Theory of the β-Dielectric Dispersion and its Use in the Estimation of Cellular Biomass, Aber Instruments Ltd., Aberystwyth, Dyfed, 1993.
[22] C.L. Davey, G.H. Markx and D.B. Kell, On the dielectric meth monitoring cellular viability, Pure Appl. Chem., 65 (1993) 1926.

[23] G.D. Austin, R.W.J. Watson and T. D'Amore, Studies of on-line viable yeast biomass with a capacitance biomass sensor, Biotechnol. Bioeng., 43 (1994) 337–341.
[24] C.L. Davey, unpublished observations, 1993.
[25] R.H. Cole and K.S. Cole, Dispersion and absorption in dielectrics. 1. Alternating current characteristics, J. Chem. Phys., 9 (1941) 341–351.
[26] P.R. Bevington, Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1969.
[27] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing (FORTRAN Version), Cambridge University Press, Cambridge, 1990.
[28] J.R. McDonald, Impedance Spectroscopy, Wiley, New York, 1987.
[29] D.B. Kell and C.L. Davey, On fitting dielectric spectra using artificial neural networks, Bioelectrochem. Bioenerg., 28 (1992) 425–434.
[30] D.W. Marquardt, An algorithm for least squares estimation of non-linear parameters, J. Soc. Ind. Appl. Math., 11 (1963) 431–441.
[31] C.L. Davey, H. Davey, D.B. Kell and R.W. Todd, Introduction to the dielectric estimation of cellular biomass in real time, with special emphasis on measurements at high volume fractions, Bioelectrochem. Bioenerg., 28 (1992) 319–340.
[32] F. Mosteller and J.W. Tukey, Data Analysis and Regression, Addison Wesley, Reading, MA, 1977, pp. 353–365.
[33] R.J. Leatherbarrow, GraFit Version 3.0, Erithacus Software, Staines, 1992.
[34] F. Rosenblatt, Principles of Neurodynamics, Spartan, New York, 1962.
[35] T.B. Blank and S.D. Brown, Non-linear multivariate mapping of chemical-data using feed-forward neural networks, Anal. Chem., 65 (21) (1993) 3081–3089.
[36] J.A. Burns and G.M. Whitesides, Feedforward neural networks in chemistry—mathematical systems for classification and pattern-recognition, Chem. Rev., 93 (1993) 2583–2601.
[37] W. Duch and G.H.F. Diercksen, Neural networks as tools to solve problems in physics and chemistry, Computer Phys. Comm., 82 (1994) 91–103.
[38] G. Kateman and J.R.M. Smits, Colored information from a black-box—validation and evaluation of neural networks, Anal. Chim. Acta, 277 (2) (1993) 179–188.
[39] M.L. Thompson and M.A. Kramer, Modeling chemical processes using prior knowledge and neural networks, AIChE J., 40 (8) (1994) 1328–1340.
[40] B. Walczak and W. Wegscheider, Nonlinear modeling of chemical data by combinations of linear and neural-net methods, Anal. Chim. Acta, 283 (1) (1993) 508–517.
[41] J. Zupan and J. Gasteiger, Neural Networks for Chemists, VCH, Cambridge, 1993.
[42] R. Goodacre and D.B. Kell, Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks: application to indole production, Anal. Appl. Pyrolysis, 26 (1992) 93–114.
[43] R. Goodacre, D.B. Kell and G. Bianchi, Rapid assessment using pyrolysis mass spectrometry and artificial neural networks of the adulteration of virgin olive oils by other seed oils, J. Sci. Food Agric., 63 (1993) 297–307.
[44] R. Goodacre, M.J. Neal and D.B. Kell, Rapid and quantitative deconvolution of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks, Anal. Chem., 66 (1994) 1070–1085.
[45] R. Goodacre, S. Trew, 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, Biotechnol. Bioeng., 44 (1994) 1205–1216.
[46] R. Goodacre, M. Neal and D.B. Kell, Quantitative analysis of multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra, Zentralblatt. für Bakteriologie, in press.
[47] B. Cheng and D.M. Titterington, Neural networks—a review from a statistical perspective, Stat. Sci., 9 (1994) 2–30.
[48] B.D. Ripley, Neural networks and related methods for classification, J. R. Stat. Soc., Ser. B-Methodological, 56 (1994) 409–437.
[49] W.S. Sarle, Neural networks and statistical models, Proc. Nineteenth Annual SAS Users Group International Conference, SAS Institute Inc., Cary, NC, USA, 1994, Available by anonymous ftp from: ftp.sas.com/pub/sugi19/neural/neural1.ps.
[50] T. Næs, K. Kvaal, T. Isaksson and C. Miller, Artificial neural networks in multivariate calibration, J. Near Infrared Spectrosc., 1 (1993) 1–11.
[51] S.H. Weiss and C.A. Kulikowski, Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems, Morgan Kaufmann Publishers, CA, 1991.
[52] D. Michie, D.J. Spiegelhalter and C.C. Taylor, Machine Learning, Neural and Statistical Classification, Ellis Horwood, Chichester, 1994.
[53] D.E. Rumelhart and J.L. McClelland, Parallel Distributed Processing, Explorations in the Microstructure of Cognition. Vol. 1: Foundations, MIT Press, Cambridge, MA, 1986.
[54] P.D. Wassermann, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1989.
[55] P.J. Werbos, The Roots of Back-Propagation: From Ordered Derivatives to Neural Networks and Political Forecasting, Wiley, Chichester, 1993.
[56] C.L. Davey, MINISCAN, Institute of Biological Sciences, University of Wales, Aberystwyth, Dyfed, SY23 3DA, 1993.
[57] H. Martens and T. Næs, Multivariate Calibration, Wiley, New York, 1989.