Rapid Analysis of High-Dimensional Bioprocesses Using Multivariate Spectroscopies and Advanced Chemometrics

A. D. Shaw[1], M. K. Winson, A. M. Woodward, A. C. McGovern, H. M. Davey, N. Kaderbhai, D. Broadhurst, R. J. Gilbert, J. Taylor, É. M. Timmins, R. Goodacre, D. B. Kell[2]

Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion SY23 3DD, UK
[1] *E-mail: ais@aber.ac.uk,* [2] *E-mail: dbk@aber.ac.uk*

B. K. Alsberg, J. J. Rowland
Dept. of Computer Science, University of Wales, Aberystwyth, Ceredigion SY23 3DD, UK

There are an increasing number of instrumental methods for obtaining data from biochemical processes, many of which now provide information on many (indeed many *hundreds*) of variables *simultaneously*. The wealth of data that these methods provide, however, is useless without the means to extract the required information. As instruments advance, and the quantity of data produced increases, the fields of bioinformatics and chemometrics have consequently grown greatly in importance.

The chemometric methods nowadays available are both powerful and dangerous, and there are many issues to be considered when using statistical analyses on data for which there are numerous measurements (which often exceed the number of samples). It is not difficult to carry out statistical analysis on multivariate data in such a way that the results appear much more impressive than they really are.

The authors present some of the methods that we have developed and exploited in Aberystwyth for gathering highly multivariate data from bioprocesses, and some techniques of sound multivariate statistical analyses (and of related methods based on neural and evolutionary computing) which can ensure that the results will stand up to the most rigorous scrutiny.

Keywords. Vibrational spectroscopy, Mass spectrometry, Dielectric spectroscopy, Flow Cytometry, Chemometrics

Advances in Biochemical Engineering/
Biotechnology, Vol. 66
Managing Editor: Th. Scheper

1 General Introduction – Multivariate Analyses in the Post-Genomic Era

"But one thing is certain: to understand the whole you must look at the whole" – Kacser H (1986). On parts and wholes in metabolism. In: Welch GR, Clegg JS (eds) The organisation of cell metabolism, Plenum Press, New York, p 327

As we enter the post-genomic era [1, 2], there is a growing realisation that the search for gene function in complex organisms is likely to require analyses not just of one or two genes or other variables in which an experimenter happens to have an interest but of everything that is going on inside a cell and its surroundings. Such analyses are now occurring at the level of the transcriptome (e.g. [3, 4]), the proteome (e.g. [5–7]) and the metabolome [2], to define, respectively the expressed performance of the genome at the level of transcription, translation and small molecule transactions. However, the present level of analysis of such data is comparatively rudimentary [8].

The bioprocess analyst has long realised that the more (useful) measurements we can make the more likely are we to understand our bioprocesses, and we ourselves have long sought to increase the number of non-invasive, on-line probes available [9, 10]. Classical methods, monitoring factors such as pH, dissolved oxygen tension, and so on, however, are in essence *univariate* methods, and only give information on individual determinands.

The strategy that we have therefore sought to follow is to exploit *multivariate* methods which can measure many variables *simultaneously*. The resulting data floods necessitate the use of robust, multivariate chemometric methods. These too are now available in many flavours, with different strengths and weaknesses.

The purpose of the present review, then, as requested by the Editor, is to review some of the types of method we have developed and exploited in Aberystwyth for the rapid, precise, quantitative, and – where possible – non-invasive measurement of bioprocesses. Our website http://gepasi.dbs.aber.ac.uk may also be consulted. We start with mass spectrometry.

2
Mass Spectrometric Measurements on Bioprocesses

Whilst on-line desorption chemical ionisation mass spectrometry (MS) has been used to analyse fermentation biosuspensions for flavones [11], the majority of MS applications during fermentations have been for the analysis of gases and volatiles produced over the reactor [12–15], or by employing a membrane inlet probe for volatile compounds dissolved in the biosuspensions [16–22]. It is obvious that more worthwhile information would be gained by measuring the non-volatile components of fermentation biosuspensions, particularly when the product itself is non-volatile, which is usually the case.

The introduction of non-volatile components into an MS has typically been via the pyrolysis of whole fermentation liquors. Pyrolysis is the thermal degradation of a material in an inert atmosphere or a vacuum. It causes molecules to cleave at their weakest points to produce smaller, volatile fragments called pyrolysate [23]. An MS can then be used to separate the components of the pyrolysate on the basis of their mass-to-charge ratio (m/z) to produce a pyrolysis mass spectrum, which can then be used as a "chemical profile" or fingerprint of the complex material analysed [24].

Figure 1 gives typical pyrolysis mass spectra of *Penicillium chrysogenum* and of penicillin G, indicating the rich structural and process information that is available from highly multivariate methods of this type.

Pyrolysis MS (PyMS) has been applied to the characterisation and identification of a variety of microbial systems over a number of years (for reviews see: [25–27]) and, because of its high discriminatory ability [28–30], presents a powerful fingerprinting technique applicable to any organic material. Whilst the pyrolysis mass spectra of complex organic mixtures may be expressed in the simplest terms as sub-patterns of spectra describing the pure components of the mixtures and their relative concentrations [24], this may not always be true because during pyrolysis intermolecular reactions can take place in the pyrolysate [31–33]. This leads to a lack of superposition of the spectral components and to a possible dependence of the mass spectrum on sample size [31]. However, suitable numerical methods (or chemometrics) can still be employed to measure the concentrations of biochemical components from pyrolysis mass spectra of complex mixtures.

Heinzle et al. [34] were able to characterise the states of fermentations using off-line PyMS, and this technique was extended to on-line analysis [35].

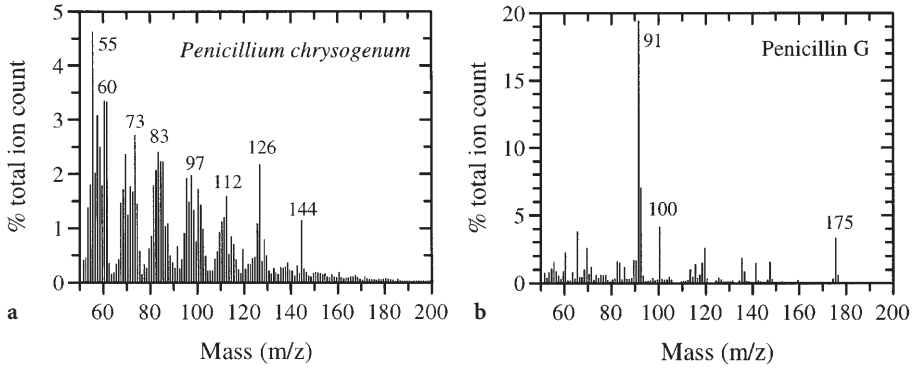

Fig. 1. a Normalised pyrolysis mass spectra of *Penicillium chrysogenum*; this complex 'fingerprint' can be used to type this organism. **b** Normalised pyrolysis mass spectra of 200 µg pure Penicillin G; this somewhat simpler 'biochemical profile' is one of the range of penicillins produced by *Penicillium chrysogenum*

However, they were not very satisfied with their system because there was no suitable data processing for the PyMS spectra. Although Heinzle and colleagues continued to use mass spectrometry for the analysis of volatiles produced during fermentation [13, 36], the analysis of non-volatiles by PyMS has not been investigated further by these authors.

With the advent of user-friendly chemometric software packages, PyMS can now be used for gaining accurate and precise quantitative information about the chemical constituents of microbial (and other) samples [37–39]. Within biotechnology the combination of PyMS with chemometrics has the potential for the screening and analysis of microbial cultures producing recombinant proteins; for instance this technique has permitted the amount of mammalian cytochrome b_5 [40] or α_2-interferon [41] expressed in *E. coli* to be predicted accurately. Chemometrics, and in particular artificial neural networks (ANNs), have also been applied to the quantitative analysis of the pyrolysis mass spectra of whole fermentor biosuspensions [31]. Initially a model system consisting of mixtures of the antibiotic ampicillin with either *Escherichia coli* or *Staphylococcus aureus* (to represent a variable biological background) was studied. It was especially interesting that ANNs trained to predict the amount of ampicillin in *E. coli* (having seen only mixtures of ampicillin and *E. coli*) were able to generalise so as to predict the concentration of ampicillin in an *S. aureus* background to approximately 5%, illustrating the very great robustness of ANNs to rather substantial variations in the biological background. (Genetic algorithms can also be used to simplify analyses of these data [42].) Samples from fermentations of a single organism in a complex production medium were also analysed quantitatively for a drug of commercial interest, and this could be extended to a variety of mutant producing strains cultivated in the same medium, thus effecting a rapid screening for the high-level production of desired substances [31]. In related studies *Penicillium chrysogenum* fermentation biosuspensions were analysed quantitatively for penicillins using PyMS [43] and

this approach has also been used successfully to monitor *Gibberella fujikuroi* fermentations producing gibberellic acid [25, 44], to measure clavulanic acid production by *Streptomyces clavuligerus* [45], and to investigate various differentiation states in *Streptomyces albidoflavus* [46].

In conclusion, PyMS is undoubtedly very useful for the discrimination of micro-organisms at the genus, species and subspecies level, and whilst it has relatively low throughput (2 min per sample), which would make it unsuitable for very-high-throughput screening programmes, it does present itself as a suitable method for the rapid, precise and accurate analysis of the biochemical composition of bioprocesses.

3
Monitoring Bioprocesses by Vibrational Spectroscopies

3.1
Infrared Analysis

The measurement of compounds in bioprocesses, including fermentations, using conventional laboratory techniques such as HPLC, TLC or calorimetric assays is often tedious, invasive, requires sample handling and difficult to do in real time. For a bioprocess where it is important to gain information about the reactor status for feedback control, methods enabling rapid and reliable measurement of components are desirable.

Infrared spectroscopy is a powerful alternative analytical technology for process monitoring which has found wide application as an off-line method in the chemical and food industries. The additional advantage over other methods is that in many circumstances it is possible to quantify a number of components *simultaneously*.

The Near-Infrared (NIR) region extends from 780 nm to 2526 nm (12820 to 3959 cm^{-1}), as defined by the American Society for Testing and Materials. Molecules that contain covalent bonds and have a dipole moment absorb IR radiation. The majority of the bands observed in the NIR are due to overtones or combinations of fundamental vibrations occurring in the Mid-IR (MIR) region that extends from 2.5 to 25 µm (4000 – 400 cm^{-1}) [47]. The light mass of the hydrogen atom and consequently its anharmonic nature means that most of the combination bands in NIR are due to hydrogen-stretching vibrations (3600 – 2400 cm^{-1}). Consequently, the greatest utility of NIR is in the determination of functional groups that contain unique hydrogen atoms [48].

3.1.1
Advantages of NIR Application to Bioprocess Monitoring

Peaks in the NIR region are not nearly as distinct as those observed in the fingerprint region of the MIR. As the intensity of first overtones are generally an order of magnitude less than the fundamentals, pathlengths are usually much longer in the NIR. The advantages of these lower intensities include the fact that nonlinearities due to strong absorptions are less likely [49]. NIR analysis can be

employed as a non-destructive process requiring little or no sample preparation and the sample may be re-introduced into the bioreactor. This is advantageous in a process environment where time is an important factor in the analysis [50].

3.1.2 *Instrumentation and Standardisation*

Modern NIR equipment is generally robust and precise and can be operated easily by unskilled personnel [51]. Commercial instruments which have been used for bioprocess analyses include the Nicolet 740 Fourier transform infrared spectrometer [52, 53] and NIRSystems, Inc. Biotech System [54, 55]. Off-line bioprocess analysis most often involves manually placing the sample in a cuvette with optical pathlengths of 0.5 mm to 2.0 mm, although automatic sampling and transport to the spectrometer by means of tubing pump has been used (Yano and Harata, 1994). A number of different spectral acquisition methods have been successfully applied, including reflectance [55], absorbance [56], and diffuse transmittance [51].

At-line sampling may involve a flow-through cell in the NIR spectrometer; in one process a glass-lined steel reaction vessel was used in combination with a fibre optic probe for measurements in a full scale chemical plant reactor [57]. Fibre optic bundles can be used to transmit NIR radiation to the reaction matrix and take signal back to the spectrometer. NIR is notoriously sensitive to changes in temperature and methods for keeping the temperature constant must be incorporated into the instrumentation.

3.1.3 *Interpreting Spectra in Quantitative Terms*

Broad superimposed bands are observed in NIR spectroscopic measurements and in most instances the peaks are not directly proportional to sample concentration. Statistical approaches are therefore required for modelling the behaviour of spectra for quantification. In the application of NIR to real world bioprocess samples, which are highly turbid scattering matrices, quantification of a constituent of interest can be particularly difficult. Vibrations are often observed that are common both to the determinand and the medium and cells in fermentations. Qualitative interpretation, and selection of unique spectral windows for calibration is therefore not always possible. One approach in the determination of wavelengths that can be used to quantify the constituent levels in bioprocess samples is to collect the spectra of raw materials alone and in combination, and then overlay spectra for isolation of unique bands. Second derivative pre-processing of spectral data can enhance spectral features and in addition baseline differences are often eliminated by this calculation; as cell density increases, the effective pathlength traversing through the sample increases because of light scattering by the cells, producing baseline offset [58]. Brimmer and Hall, [55] derived a Multiple Least squares Regression (MLR) equation that compensated for scattering differences attributable to changes in the biomass of the fermentation process. This was accomplished by using

a reference wavelength at which the spectral data varies with penetration depth in a reproducible manner. Background information such as that attributable to water or the sample holder may be subtracted or used as a ratio [53, 59], however, in some instances this correction does not appear to affect the modelling ability of the algorithms [56]. NIR can be applied to whole cells, supernatant and aqueous mixtures of constituent samples, which may be also used to form calibration models [60].

Multivariate calibration methods. These are capable of extracting meaningful information from seemingly uninterpretable NIR spectra of bioprocess samples; however for these methods measurements made using other techniques must be available for training. It may be necessary to form a model for different times in the bioprocess e. g. for the start-up period and for later stages when inhibitors are accumulating and substrates are depleting in the fermentation.

Transferability of spectral data and models in NIR spectroscopy. This subject is an issue that is pertinent to the future use of NIR for bioprocess monitoring. Pre-processing to remove baseline shifts and noise in spectra from individual machines or direct standardisation by data transformation with a representative subset can be used to calibrate across instruments [61].

3.1.4 *Applications*

NIR spectroscopy continues to be applied to on-line fermentation and biotransformation monitoring, for example, of ethanol and biomass in rich medium in a yeast fermentation [62, 63], lactic acid production [64, 65], bioconversion of glycerol to 1,3-dihydroxyacetone [66] and nutrient and product concentrations in commercial antibiotic fermentations [67, 68]. Hall, Macaloney and colleagues [51, 58] reported NIR spectroscopic monitoring of industrial fed-batch *E. coli* fermentation of varying levels of acetate, ammonium, glycerol and biomass which they had previously studied in shake flasks [54], while Yano and colleagues [56] used NIR spectroscopy to determine with good precision the concentrations of ethanol and acetate in rice vinegar fermentations. The spectral signature of biomass with respect to wavelength regions was found to be essentially identical when groups of industrially-important microorganisms [69] were analysed. The concentration of many species may be determined from one spectroscopic measurement, as long as their concentration is 1 mM or greater [59].

New methods of variable selection include evolutionary methods based on Darwinian principles including Genetic Algorithms and Genetic Programming [70] and as such help to deconvolute whole spectral models in terms of which variables are important in the modelling procedure. When applied to a NIR glucose sensor, fewer than 25 variables were selected to produce errors statistically equivalent to those yielded by the full set containing 500 wavelengths and the algorithm correctly chose the glucose absorption peak areas as the in-

formation-carrying spectral regions [71], and these approaches, coupled to digital filtering, appear to be the methods of choice [72, 73].

It is important that calibration models are rigorously validated and in the first instance that all variations are accounted for in the model using diverse samples that are expected to be observed in future bioprocess runs. Some investigators attempt to keep process conditions very reproducible but such conditions are uncommon in an industrial environment. In addition, multivariate calibration models will work well if identical media (composition) and process conditions are used on each successive run. Simple modifications such as use of a different media supplier can affect the spectral background. The predictive ability of the models will then be affected as they will be challenged with samples which they have not been trained to recognise [74].

3.2 MIR Analysis

The higher level of spectral resolution in the MIR range often allows peaks to be assigned to specific medium components or chemical entities. Although analysis of bioprocesses in the MIR range would be especially useful for monitoring products of interest because of the feature rich spectra between 4000 – 200 cm^{-1}, application to on-line aqueous systems at an industrial level is hindered by the broad water absorption across most of the so-called 'fingerprint' spectral range. For off-line analysis this can be overcome simply by drying samples; however, for on-line analysis success with mid-IR monitoring of bioprocesses has been limited to use of transmission cells with extremely short pathlengths or Attenuated Total Reflectance (ATR) spectroscopy. ATR utilises the phenomenon of total internal reflection. ATR can be used essentially as an 'in-line' method, where the sample interface is located in the process line itself, thus eliminating the requirement for an independent sampling system. The sample to be analysed is placed in direct contact with a crystal made from zinc selenide, germanium, thallium/iodide, sapphire, diamond or zirconium. Quantitative monitoring by FT-IR spectroscopy of the enzymatic hydrolysis of penicillin V to 6-aminopenicillanic acid and phenoxyacetic acid using a 25 µl flow through cell with a zinc selenide crystal demonstrated that the IR method allowed better prediction of the process termination time than the standard method based on monitoring the addition of sodium hydroxide [75].

On-line MIR ZnSe ATR analysis of microbial cultures has been used primarily for non-invasive monitoring of alcoholic or lactic fermentations. Alberti et al. [76] reported the use of a ZnSe cylindrical ATR crystal to monitor accurately substrate and product concentrations from a fed-batch fermentation of *Saccharomyces cerevisiae*. Picque et al. [77] also used a ZnSe ATR cell for monitoring fermentations and found that whereas NIR spectra obtained from alcoholic or lactic fermentation samples contained no peaks or zones whose absorbance varied significantly, both transmission and ATR MIR could be used successfully to measure products. Fayolle et al. [78] have employed MIR for on-line analysis of substrate, major metabolites and lactic acid bacteria in a fermentation process (using a germanium window flow-through cell), and

studied the effects of temperature on the ability to quantify the substrates (glucose and fructose) and metabolites (glycerol and ethanol) in an alcoholic fermentation using a ZnSe ATR crystal. Hayakawa et al. [79] described the use of a remote ZnSe ATR probe for determining glucose, lactic acid and pH simultaneously in a lactic acid fermentation process using *Lactobacillus casei*. The benefits of the ATR method of analysis are generally those that would be considered advantageous for any on-line system, being non-destructive, requiring no sample preparation or reagents and only a short analysis time, with minimal expertise necessary in the industrial environment. Practical drawbacks for the technique, particularly for microbial fermentations, centre on the need to purge the flow cell or clean the ATR probe to prevent surface contamination through biofilm formation. Some ATR crystal materials are toxic, limiting certain applications to the use of sapphire, diamond or zirconia. Sapphire crystals are non-transmitting below 2000 cm^{-1} which means that the MIR fingerprint region cannot be investigated with this device [80]. Developments in optical fibre design and coupling to spectrometers makes IR analysis a practical consideration for industrial reactors, as the IR spectrometer can be kept remote from the sampling probe, although at present chalcogenide fibres can only be used over short distances.

Off-line analysis of bioprocesses is clearly less desirable for a rapid response. However, MIR analysis of fermentation samples off-line does offer certain advantages over other techniques. A method we have introduced and called DRASTIC (Diffuse Reflectance-Absorbance IR Spectroscopy Taking in Chemometrics) [81] for MIR analysis of bioprocess samples has been successfully applied to the estimation of drug concentrations in biological samples, including fermentations from a microbial strain development programme [82, 83]. In this technique fermentation samples (5 µl) were applied to wells in an aluminium plate or aluminium-coated plastic 384-well microtitre plate, dried, mounted on a motorised mapping stage and analysed by the diffuse reflectance-absorbance method using a Bruker IFS28 FT-IR spectrometer. This allows rapid non-destructive analysis of samples (typically 1 per second) at a high signal to noise ratio. We were thus able to predict concentrations of ampicillin in a biological background of *E. coli* (see Fig. 2 for example spectra) and *Staphylococcus aureus* cells, and we used spectral data obtained from analysis of fermentations of *Streptomyces citricolor* to predict the concentrations of the carbocyclic nucleosides aristeromycin and neplanocin A. PLS routine was used to create a training set using the MIR spectral data and information provided from HPLC analysis of samples. This method can be fully automated and allows for a particularly high sample throughput rate.

The use of multivariate spectral information is particularly advantageous where quantification of a particular metabolite in a complex biological background is being attempted and application of the technique necessitates the use of chemometric processing techniques for quantification of components.

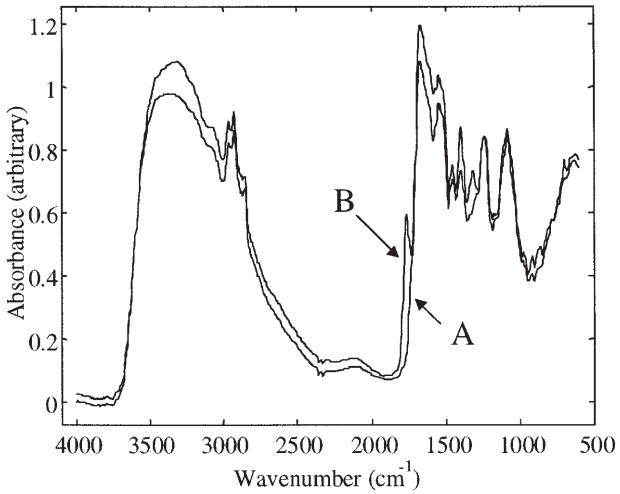

Fig. 2. MIR diffuse reflectance spectra of *Escherichia coli* cells without (*A*) and with (*B*) 20 mM ampicillin

3.3 Monitoring Bioprocesses Using Raman Vibrational Spectroscopy

Recent exploitation of biotechnological processes for pharmaceutical and food industries has necessitated rapid screening and quantitative analysis of the specific components. Therefore, there is continuing need for developing on-line methods for monitoring such biological processes [84–86]. The ideal method [87] would be rapid, non-invasive, reagentless, precise and cheap, although to date, with the possible exception of near-IR spectroscopy almost no such single method has been found. Generally these bioprocesses progress from translucent to increasingly opaque matrices as the microbial cells multiply and become highly light scattering and rich in molecular vibrational information. The use of specific molecular vibrations allowing specific fingerprinting of singular or multi-components for identification and quantification using the vibrational FT-IR and Raman spectroscopies for monitoring these bioprocesses can provide suitable alternatives to the present day process monitoring.

Raman spectroscopy relies on vibrational signals generated by focusing a laser beam onto the sample to be analysed, where most of the incident photons are either transmitted through the sample, absorbed by it, or scattered (elastic scattering). In a very few cases, approximately 1 in 10^9, the vibrations and rotations of the scattering molecules cause energy quanta to be transferred between molecules and photons in the collision process (inelastic scattering). A monochromator and a detector are then used to measure these inelastically scattered photons to give a Raman spectrum.

Raman spectroscopy can be used to analyse aqueous biological and bioorganic samples e.g., bacteria, spores, diseased tissues, neurotransmitters,

protein structures, membrane lipids, biochemical assays, drug-nucleotide interactions, constituents of oils, water for toxic analytes and bioprocesses.

During the last few years there has been a renaissance in Raman instrumentation suitable for the analysis of biological systems, initially with the development of Fourier Transform (FT)-Raman instruments in which the wavelength of the exciting laser is in the near-infrared laser (usually a Nd:YAG (neodymium doped yttrium-aluminium garnet) at 1064 nm) rather than in the visible region, an arrangement which therefore avoids the background fluorescence typical of biological samples illuminated in the visible [47, 88–104]. In addition, and at least as importantly, exceptional Rayleigh light rejection has come from the development of holographic notch filters [105–108], and a recent innovation is the use of Hadamard-transform-based spectrometers [109, 110].

Although the FT approach to both infrared and Raman spectroscopy possesses well-known advantages of optical throughput [47, 111], there are still problems for FT-Raman with many aqueous biological samples as water may absorb both the exciting laser radiation at 1064 nm and the Raman scattered light. In addition, it is often necessary to co-add many hundreds of spectra to produce high-quality data from biological systems, and acquisition times are frequently 15–60 min. More recently, therefore, it has been recognised that charge coupled device (CCD) array detectors are ideal elements for use in *dispersive* (non-FT) Raman spectroscopy. However, they normally have very low quantum efficiency at 1064 nm photons. Thus holographic notch filters and CCD array detectors have been combined with a dispersive instrument, using diode laser excitation at 780 nm (a wavelength which suppresses fluorescence from most samples but which penetrates water well). The cooled CCD is a multi-channel device which has exceptional sensitivity and very low intrinsic noise (dark current), so that the signal:noise ratio is improved by at least 2 orders of magnitude (compared with an uncooled CCD) and data acquisition is correspondingly fast [89]. These and other major technical advances [112, 113] now make Raman a very promising tool for the rapid, non-invasive and multiparameter analysis of aqueous biological systems, including the estimation of metabolite concentrations in ocular tissue [114, 115].

In 1987, Shope and colleagues [116] used attenuated total reflectance (ATR) Raman spectroscopy for the on-line monitoring of the fermentation by yeast of sucrose to ethanol, using the argon ion laser line at 514.5 nm. Gomy et al. [117, 118] monitored their alcoholic fermentation using the same laser with a fiber optic probe attached to a Raman spectrometer but analysed the ethanol levels only at higher wavenumber (2600–3800 cm^{-1}) . This was because the Raman monitoring of these processes using 514.5 nm excitation gave significant fluorescence in the lower wavenumber region, as can be observed in the spectra shown in these papers.

Although fluorescence has been a major hindrance for the use of Raman spectroscopy in biology, Shope and colleagues [116] clearly showed that the narrow Raman peaks were distinct from the broad features of fluorescence and proposed the use of full widths at half-height of the peaks for chemical quantitation from Raman spectra. Shope et al. [116] used a least squares fit to analyse the Raman spectra for quantification of the production of ethanol during the

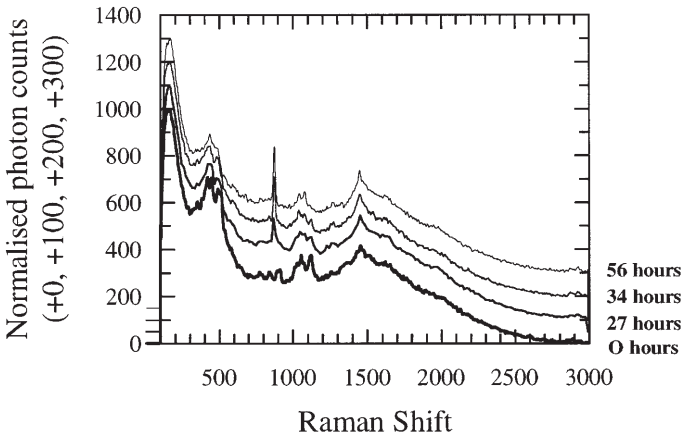

Fig. 3. Comparison of smoothed, normalised spectra from a biotransformation of glucose to ethanol, taken at intervals through the experiment, showing the change in the spectrum over time. Spectra are artificially displaced by 100 photon counts for clarity

yeast fermentation process. Finally, Spiegelman and colleagues [119] have recently shown that the amount of glucose in aqueous solution can be measured using Raman spectroscopy.

4 Measurement of Biomass

This laboratory long ago devised [120] the use of radio-frequency dielectric spectroscopy [121, 122] for the on-line and real-time estimation of microbial and other cellular biomass during laboratory and industrial fermentations. The principle of operation is that only intact cells (see [123] for what is meant in this context by the word 'viable'), and nothing else likely to be in a fermentor, have intact plasma membranes and that the measurement of the electrical properties of these membranes allows the direct estimation of cellular biomass (Fig. 4).

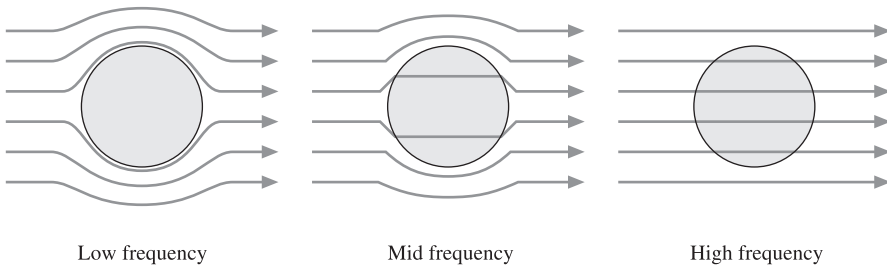

Fig. 4. Fields and cell membranes. At low frequency the field cannot penetrate the cell wall and is dropped almost entirely across the outer membrane such that the membrane amplifies the field across itself by a factor of up to 1000 From *left* to *right* – Low frequency, Mid frequency, High frequency

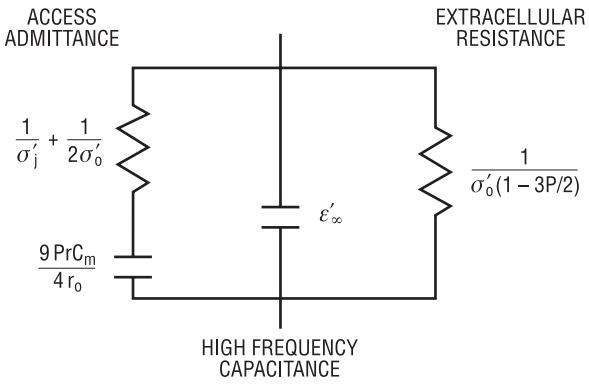

Fig. 5. Standard linear equivalent circuit of an assumed linear dielectric cell membrane can be modelled with simple standard components. This assumption breaks down if the field is amplified across the membrane as in Fig. 1 to a degree sufficient to produce nonlinearity

This situation is modelled as the equivalent circuit of Fig. 5, where all the components are assumed linear.

The probe has been long and successfully commercialised (see http://www.aber-instruments.co.uk) and since we have reviewed this approach on a number of occasions (e.g. Kell et al. 1990, Davey 1993a, b, Davey et al. 1993a, b) we will not do so here, save to point out (in the spirit of this review) the trend to the exploitation of *multi-frequency excitation* for acquiring more (and more robust) information on the underlying spectra. [124, 125]. Most recently, we have also devised a number of novel routines for correcting for the electrode polarisation that can occur under certain circumstances [126, 127], and have turned our attention to the *nonlinear* dielectric spectra of biological systems.

4.1
Dielectrics of Biological Samples – Linear or Nonlinear?

The dielectric response of biological tissue has long been assumed linear. Thus an enzyme is treated as a hard sphere which relaxes linearly in an a.c. field at all but high field strengths [128]. In a suspension of cells, the electric field cannot penetrate to the interior of the cell at the low frequencies currently of interest in nonlinear dielectric spectroscopy [129], and is dropped almost entirely across the outer membrane of the cell which is predominantly capacitive at these frequencies, as was shown in Fig. 4.

However, an enzyme which has different dipole moments in different conformations during its operation (Fig. 6) may affect and be affected by electromagnetic fields [130]. Change between states is unlikely to be smoothly or linearly related to the field due to the constraints imposed on the enzyme by its environment in the membrane, so the dielectric response of the material is nonlinear even at low applied fields [131].

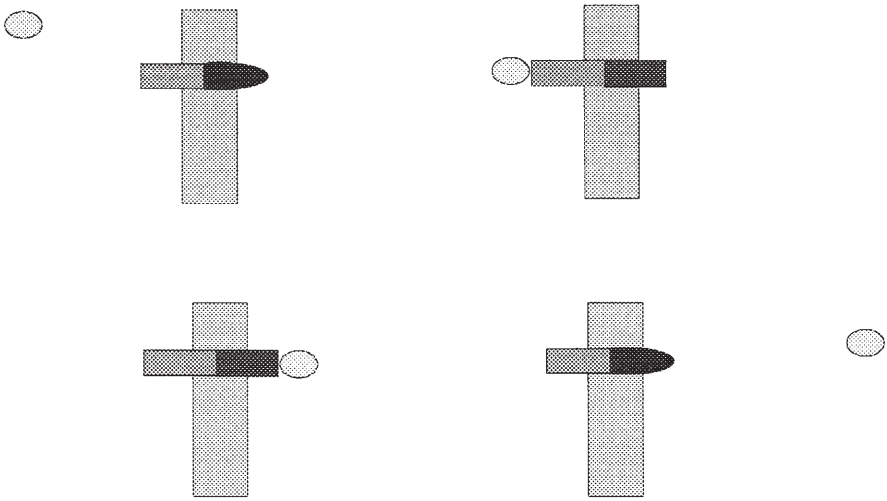

Fig. 6. Enzyme transporting ion across membrane via conformational change. If the different conformations have different dipole moments, the enzyme will be sensitive to electric fields and will be detectable by its effect on these fields

The equivalent circuit of Fig. 5 is no longer very useful since its individual components are no longer linear. This behaviour shows up as the generation by the tissue of harmonics of the applied frequency [129].

A nonlinear dielectric spectrometer has been designed around a standard IBM PC; and realised almost completely in software, with a minimum of extraneous hardware [129].

4.1.1
The Nonlinear Dielectric Spectrometer

A sinusoidal (or otherwise) signal is generated by the PC and applied to the outer terminals of a 4-terminal electrode system. The resulting signal across the inner electrodes is fed back differentially to the PC. This signal is then transformed into its power spectrum and the harmonics studied (Fig. 7).

Of course things are never quite this simple. At the low frequencies (a few Hz to a few kHz) studied so far, there is a strong polarisation layer around the driver electrodes. The i/V relation of this layer is both strongly nonlinear and highly variable with time, and its effects must be removed from the (weak) harmonics generated by the biology, if direct visualisation of the harmonic spectra is needed.

A reference spectrum (dB power spectrum) is taken using the supernatant of the suspension under test. This is the polarisation signature. This is then subtracted from the equivalent spectrum from the whole suspension. This procedure deconvolves the polarisation harmonics from those produced by the tissue nonlinearity (Fig. 8).

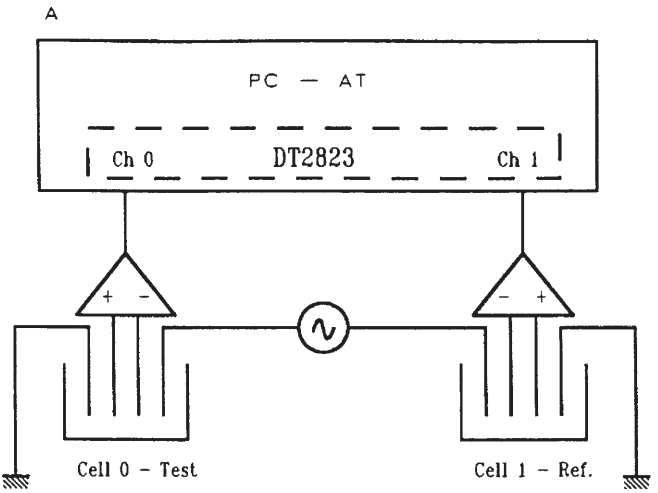

Fig. 7. Dielectric spectrometer schematic: Two standard four-terminal electrode chambers are connected to A/D converters and on into a PC. Fourier analysis is done by the PC to produce the nonlinear dielectric spectra

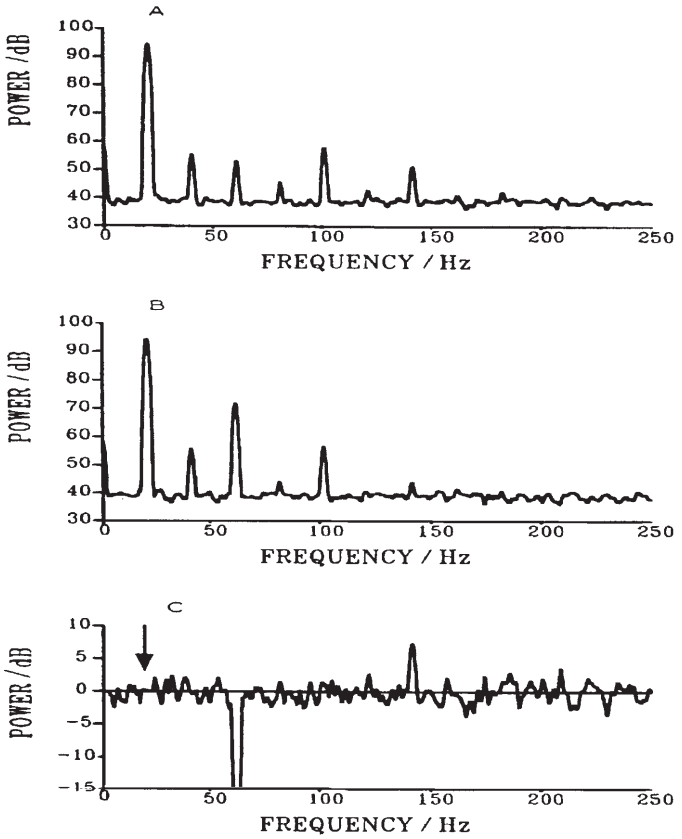

Fig. 8. Reference, suspension, and difference spectra of resting yeast. Predominently odd harmonics only are produced in this metabolic state signifying a symmetric system in equilibrium

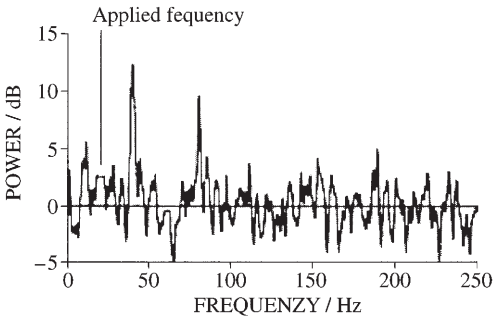

Fig. 9. Difference spectrum of metabolising yeast cells Even harmonics appear under these conditions showing an activation of the ATPase signifying the disturbance of the equilibrium of Fig. 5

4.1.2
Nonlinear Dielectrics of Yeast Suspensions

In a suspension of *Saccharomyces cerevisiae*, an inhibitor study along with use of mutant strains showed that the predominant source of the nonlinear signature in this organism is the membrane-located H+ ATPase. The harmonics are highly voltage- and frequency-windowed, with the peak of the frequency window for the resting enzyme coinciding neatly with its k_{cat} value. In a resting state, at equilibrium, the suspension generates almost entirely odd-numbered harmonics, as in Fig. 8, suggesting symmetry about the equilibrium of the ATPase. If glucose is addded to the suspension to fuel proton transport by the ATPase, then the shift away from equilibrium breaks the symmetry and even-numbered harmonics appear, giving a measure of the activity or inactivity of this enzyme and the consequent metabolic state of the yeast cells as shown in Fig. 9.

Analysing the behaviour of the harmonics over a range of frequencies/voltages allows the rapid collection of a very large amount of metabolism-dependent information.

4.1.3
Multivariate Analysis

Recently, work has focused on the use of multivariate methods to form models capable of predicting the factors causing responses. Much of this work has centred on the prediction of glucose levels in yeast fermentations from the cellular responses. A major practical advantage of multivariate methods is that there is no requirement for a reference sample to be taken.

Initial experiments used principal component analysis (PCA) to investigate the multivariate response. PCA is a non-parametric method which outputs linear combinations of the input values (the "principal components"), such that the majority of variation is concentrated in the first few components.

PCA does not attempt to relate cause and effect; it merely serves to highlight the larger variations in the data. Nevertheless, the results obtained from PCA

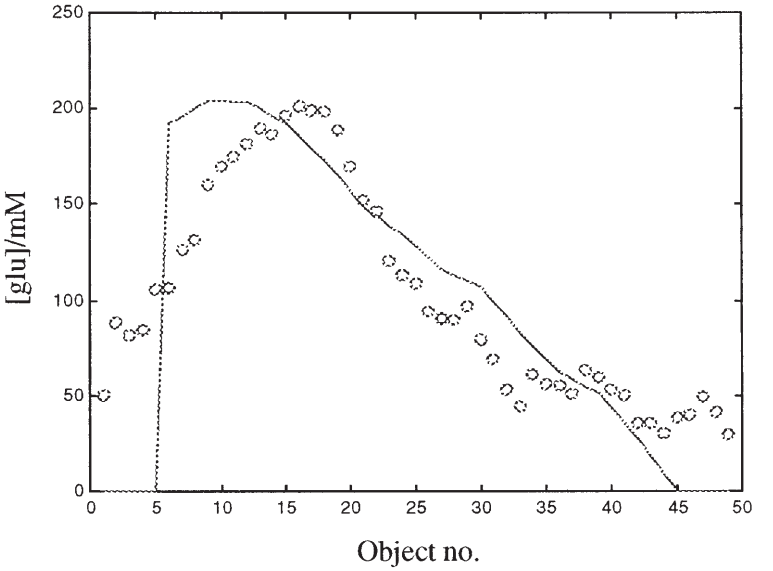

Fig. 10. PLS based prediction of glucose levels in one yeast batch fermentation by a model formed and validated on glucose levels in two other independent fermentations. The rmsep is 41% of the mean value of the data

proved promising, showing large variations which could be due to the cells' activity in response to glucose.

Subsequent work has used partial least squares regression (PLS) to form predictive models of glucose concentration during batch fermentations as shown in Fig. 10 (where object number = sample number and gives a measure of the progress of the fermentation). PLS produces models by projecting the large number of response X-variables (the harmonics in the NLDS spectra) into a smaller number of 'Latent' variables, while retaining as much relevant variability as possible. The variables in this space are then used to form a regression onto the predicted Y-variables (the actual glucose levels measured by a reference method). This "two-way'" modeling tends to form much more accurate models than other simple linear multivariate methods (e.g. principal component regression and multiple linear regression) as it automatically detects relevant X-variables and preferentially forms the model on these. The precision of the prediction is assessed by the commonly used Root Mean Square Error of Prediction (rmsep) [132]. Three independent datasets are required; one to form the model, one to validate the model, and one which the modelling process has not seen to test the model against 'unknown' data

Examination of the "residual" unmodelled variation in these experiments indicates that there is a nonlinearity in the relationship between the X and Y variables. This detracts from the models' accuracy. To this end the inherently nonlinear capabilities of ANNs have been employed with an improved predictive capability resulting in the prediction of Fig. 11.

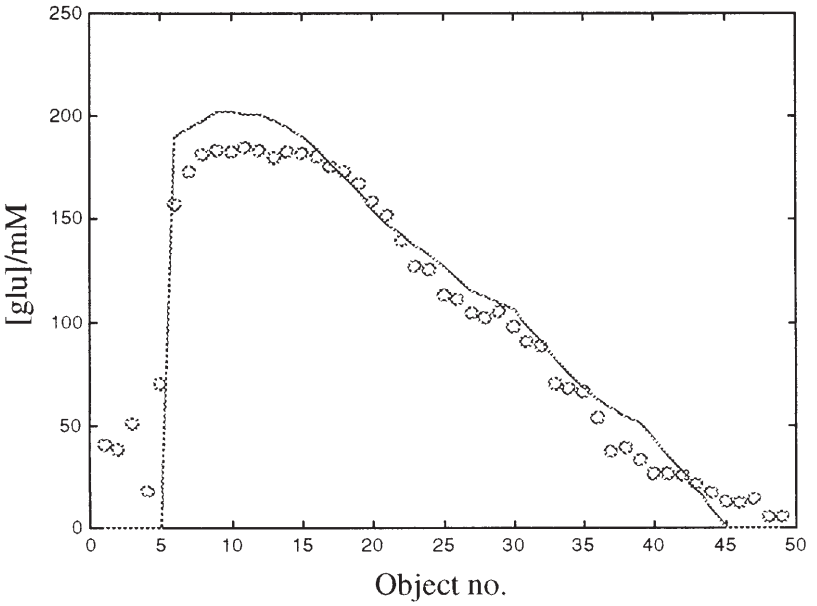

Fig. 11. Neural net prediction of glucose levels in one yeast batch fermentation by a model formed and validated on glucose levels in two other independent fermentations. This experiment uses uncoated gold electrodes. The rmsep is 19%

The current area of interest is in trying to reduce the electrode instabilities which are responsible for large baseline offsets when a model based on one fermentation is used to predict results from another. This can be done in either hardware, by coating the electrode to stabilise the interface [133], or in software, by using more powerful modelling methods such as Genetic Programming (GP) to automatically remove the effect of these instabilities from the model [134].

4.1.4
Electrode Polarisation and Fouling

In biological NLDS work, electrode polarisation is a serious problem at the low frequencies (up to a few tens of Hz) where the biology typically reacts most strongly to the electric field; and its fluctuations can be similar in size to, or bigger than, the small changes due to biological activity (e.g. upon glucose metabolism). It is therefore vital to control electrode polarisation insofar as is possible. To obtain nonlinear electrochemical reproducibility, electrode surfaces must be scrupulously clean, and this is very difficult to achieve. If any contamination is present, the biologically relevant signal may be unstable, distorted or concealed completely [135].

Electrode cleaning to ensure repeatable nonlinear dielectric spectra is a complex and empirical task, due to the lack of knowledge of the exact form of the

causative mechanisms operating in the electrode/electrolyte interface. No repeatable and certain ways of obtaining a quiet and repeatable reference signal from an individual electrode surface have been found but simple abrasion works best. Once clean, electrodes may stay stable for days, or become unstable within a few minutes. Continual control readings, performed as indicated above, are vital during any series of experiments to be sure the electrode surface behaviour has not substantially altered during the experiments, in which case the results must be abandoned and the experiments repeated. This Byzantine process can make the process of obtaining a lengthy series of results with continually clean electrodes a nightmare.

4.1.5 *Electrode Coating*

To prevent a protein from adhering to a metal surface, the surface can be coated with a sheet of poloxamers. These are a triblock copolymer consisting of PEO-PPO-PEO, in which two polyethylene oxide (PEO) chains are attached to a hydrophobic polyproylene oxide (PPO) anchor. This prevents the protein binding by steric repulsion overpowering the attraction between the protein and the coating layer [136]. This coating layer stabilises the electrode interface slightly and prevents protein fouling, allowing the electrodes to be used after a simple cleaning and coating procedure. They then stay useable for a month.

The coating allows three independent datasets leading to the prediction of PLS prediction of Fig. 12 (to be compared with that of Fig. 10) to be obtained

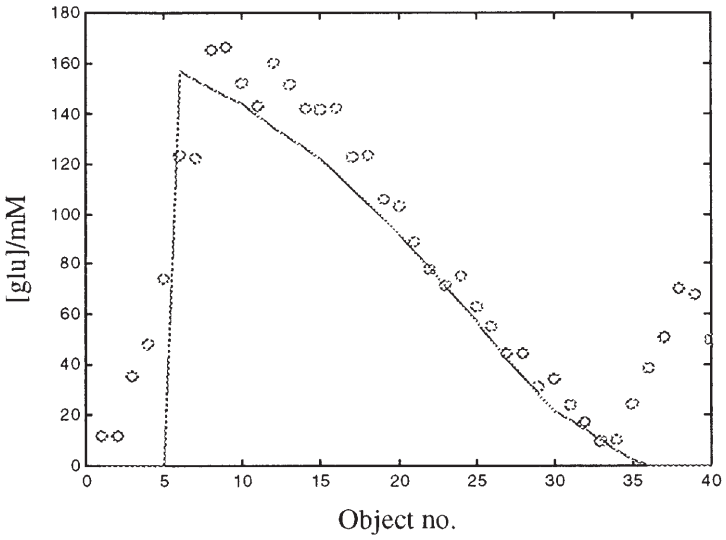

Fig. 12. PLS prediction of glucose levels in one yeast batch fermentation by a model formed and validated on glucose levels in two other independent fermentations. This experiment uses polymer coated electrodes. The rmsep is 35%

rapidly and conveniently, without the prohibitive electrode problems discussed above. It is also found that the coating linearises the data and allows PLS to perform better in relation to nonlinear modelling methods.

4.1.6
Genetic Programming

Genetic programming [137] is an evolutionary technique which uses the concepts of Darwinian selection to generate and optimise a desired computational function or mathematical expression. It has been comprehensively studied theoretically over the past few years, but applications to real laboratory data as a practical modelling tool are still rather rare. Unlike many simpler modelling methods, GP model variations that require the interaction of several measured nonlinear variables, rather than requiring that these variables be orthogonal.

An initial population of individuals, each encoding a potential solution to the optimisation problem, is generated randomly and their ability to reproduce the desired output is assessed. New individuals are generated either by mutation (the introduction of one or more random changes to a single parent individual) or by crossover (randomly re-arranging functional components between two or more parent individuals). The fitness of the new individuals is then assessed, and the fitter individuals from the total population are more likely to become the parents of the next generation. This process is repeated until either the desired result is achieved or the rate of improvement in the population becomes zero. It has been shown [137] that if the parent individuals are chosen according

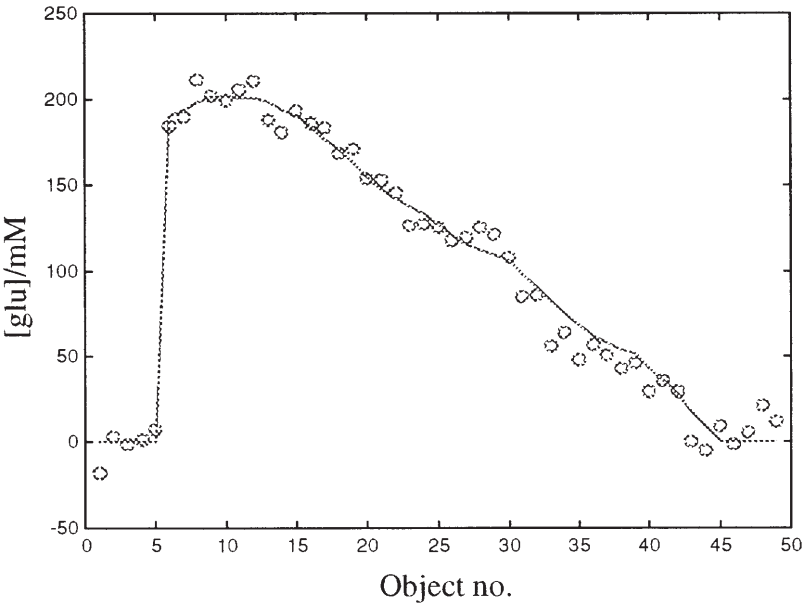

Fig. 13. Genetic Program prediction of the data of Figures 7 and 8. The rmsep is 9%

to their fitness values, the genetic method can approach the theoretical optimum efficiency for a search algorithm.

This technique allows the prediction of Fig. 10 and 11 to be improved to produce Fig. 13.

Given the very heavy computational load of GP, it would not be the method of choice for problems which yield to simpler approaches. However the above data show that it can be very beneficial on problems that have defeated other methods.

4.1.7
Other Microbial Systems

NLDS has also been successfully applied in this laboratory to measurements of photosynthesis in *Rhodobacter capsulatus* [138]; of glucose levels in erythrocytes, both invasively and non-invasively [135]. It has also been used successfully to detect the subtle interaction of weak low-frequency magnetic fields with membrane proteins of aggregating amoebal cells of *Dictyostelium discoideum*. Using PCA, a significant distinction was shown between cells previously exposed to pulsed magnetic fields (PMF) of 0.4 mT and 6 mT and their respective controls. Significant distinction was also shown between cells exposed to 50 Hz sinusoidal magnetic fields of 9 μT and 90 μT and their respective controls. NLDS was able to demonstrate a dose response with respect to both duration of exposure and field strength. In all cases significant changes in intracellular biochemistry had also been shown. There is some evidence to support a hypothesis that voltage gated calcium channels are involved in the response of *Dictyostelium* to PMFs [139, 140].

5
Flow Cytometry

Flow cytometry [141, 142] is a technique that allows the measurement of multiple parameters on individual cells. Cells are introduced in a fluid stream to the measuring point in the apparatus. Here, the cell stream intersects a beam of light (usually from a laser). Light scattered from the beam and/or cell-associated fluorescence are collected for each cell that is analysed. Unlike the majority of spectroscopic or bulk biochemical methods it thus allows quantification of the heterogeneity of the cell sample being studied. This approach offers tremendous advantages for the study of cells in industrial processes, since it not only enables the visualisation of the distribution of a property within the population, but also can be used to determine the relationship *between* properties. As an example, flow cytometry has been used to determine the size, DNA content, and number of bud scars of individual cells in batch and continuous cultures of yeast [143, 144]. This approach can thus provide information on the effect of the cell cycle on observed differences between cells that cannot be readily obtained by any other technique.

Flow cytometry has been applied to the study of the formation of the biopolymer poly-*b*-hydroxybutyrate (PHB). While the formation of the polymer can be detected by changes in the light scattering behaviour of cells [145], its ac-

cumulation has also been analysed using the hydrophobic fluorescent dye Nile Red [146]. PHB is produced commercially for use in the manufacture of biodegradable plastic materials and this approach has enabled researchers to determine the effect of changes in nutrient limitation conditions on the production and storage of PHB in individual cells [147].

While these examples illustrate the role of flow cytometry in bioprocess monitoring, the analyses have been conducted off-line thus making their use in bioprocess control impractical. Recently, a portable flow cytometer – the Microcyte – [148] has been described, which due to its small size and lower cost (compared to conventional machines) allows flow cytometry to be used as an at-line technique [149]. Rønning showed that this instrument had a role to play in the determination of viability of starter cultures and during fermentation. The physiological status of each individual cell is likely to be an important factor in the overall productivity of the culture and is therefore a key parameter in optimising production conditions.

The problems of converting flow cytometry into an on-line technique are discussed by Degelau and colleagues [150], however, more recently a flow injection flow cytometer for on-line monitoring of bioreactors has been developed by Zhao and colleagues [151]. In the system described a sample is removed from the fermentor under computer control. The sample is degassed prior to passing into a microchamber where it is automatically diluted if necessary prior to the addition of stains or other reagents. Following an appropriate incubation in the microchamber the sample is delivered to the flow cytometer for analysis. This instrument has been used successfully to monitor both the production of green fluorescent protein (Gfp) in *E. coli* and to determine the distribution of DNA content of a *S. cerevisiae* population without the necessity for operator input. With continuing decrease in costs and increase in automation flow cytometry is likely to play an increased role in bioprocess monitoring and control.

6
Data Analysis

Whilst modern instruments may provide much more accurate data than those of years ago, new types of instrument are being developed which provide data of somewhat lesser accuracy, but which have other advantages (e.g. speed, throughput, on-line). Advances in computing methods help in the extraction of meaningful information from such data, which in the past would have been impossible, and so bioinformatics has become an essential part of the experimental procedure.

6.1
Data Pre-processing

Before carrying out any statistical analysis on multivariate data, it is important to ensure that the data are valid, and in a suitable format. This means:

- Ensuring that there are no errors in the data
- Normalising, when necessary

Errors may be caused by data input error (where this is done by hand), or by an incorrectly analysed sample. In the former case, this is typically a wrong number, or a decimal point missed or wrongly placed. Such errors may usually be found by testing the maximum and minimum values of a variable. If one value is found to be significantly different to the others, it is suspect, and should either be corrected (e.g. by referring back to the original experimental results, where available, or moving a decimal point), or the whole object affected deleted. If the measurements for one sample are consistently found to be suspect, normalisation may solve this problem. If it is suspected that the sample was incorrectly analysed, and cannot easily be reanalysed, it should be deleted from the data set.

Many spectroscopic methods will produce results whose magnitude depends upon the amount of sample present during the analysis or prevailing experimental conditions (e.g. Pyrolysis Mass Spectrometry, Raman spectroscopy). In such cases, the samples should be *normalised*, either to an internal standard or variable of consistent value, or, where the totals are expected to be about the same for each object (PyMS), to the total.

For example, to normalise the total of all objects to 1000, each variable x_{ib} before normalisation in object x with n variables becomes after normalisation (x_{ib}):

$$x_{in} = \frac{1000}{\sum_{j=1}^{n} x_i} \times x_{ib}$$

Where the result does *not* depend on such factors, or a normalisation to an internal standard is carried out by the spectrometer or accompanying software automatically (e.g. in Nuclear Magnetic Resonance – NMR), further normalisation *should not be carried out*.

If after normalisation to the total, a variable is found to be suspect and deleted, normalisation must be carried out again. It is possible, when normalising to the total, that such re-normalisation may adversely affect the remaining data. If this is judged to be the case, the whole experiment will need to be repeated.

Most statistical packages will carry out normalisation of the *variables*, typically to $\frac{1}{StDev}$ *for each variable*. The purpose of this is to negate the effect of large variables on the model formed [152]. If the package being used does not provide this facility, or if for some other reason it is believed that a better result will be obtained by using a different normalisation of the variables, this should be carried out at this stage. Such normalisation should always be carried out *after* any normalisation of the objects has been performed.

6.2 Model Simplification

When performing multivariate statistical analysis on a set of data for classification or quantification, it is common practice to use all the variables available.

The belief is that the statistical method used (such as PLS, PCR, MLR, PCA, ANNs) will extract from the data those variables which are most important, and discard irrelevant information. Statistical theory shows that this is incorrect. In particular, the *principle of parsimony* states that a simple model (one with fewer variables or parameters), if it is just as good at predicting a particular set of data as a more complex model, will tend to be better at predicting a new, previously unseen data set [153–155]. Our work has shown that this principle holds.

There have been a number of methods of data reduction proposed, some of which are briefly described here.

One method is to use a *variable ranking* system, in which the best *n* variables (where *n* ranges from 1 to the total number of variables), are tested. The variables used for the value of *n* at which the best model is formed will then be taken to be the optimal. This method has proved very successful, particularly for relatively low noise NMR data from olive oils [156–158], the results clearly showing that the use of all variables in model creation does not yield an optimal result in most cases, and for Raman data [159], where the variables are peak height, width, area and position, the peaks initially chosen being representative of certain bonds within the substance being analysed. It has the advantage of being relatively quick (only *n* models need be formed), and simple to understand. It can also be a great aid to understanding the data being analysed. However, it does not take account of collinearity in the data, nor the possibility that two variables may be additively, but not individually, important

Taking *Fourier transforms* of spectra (e.g. [160, 161]) and selecting a suitable cut-off will eliminate most of the noise whilst retaining most of the information. The precise point of the cut-off is not easy to determine, as there is a trade-off between eliminating noise and losing data. It is also likely that many of the remaining variables will be collinear (essentially saying the same thing), and therefore make the model unnecessarily complicated.

Using the first *n principal components* (where *n* is determined by some metric which attempts to remove components containing only noise) also suffers from the problem of this trade-off, but does have the advantage that no variables remaining will be collinear (therefore they all contribute different information to the model).

Genetic programming, described earlier, picks only certain variables from the model. The rules, which may be in the form of a computer language such as Lisp, or easily interpretable equations, produce a formula from which a result can be calculated (e.g. if *(measurement_1 > 2.37* and *measurement_2 < 0.53)* or *measurement_3 > 4.28* then *sample is adulterated* else *sample is clean*) [162–165]. Rather than being a pre-processing step before statistical analysis, this method combines the variable selection and model formation stages into one.

6.3 Data Partitioning

It is, at this point, important to understand the difference between *unsupervised methods* and *supervised methods*. With the former, there is no indication given to the model creation program (e.g. PCA, self-organising maps) of where any of

the data should lie, or its class or value. With such a technique, therefore, one set of data is sufficient. However, if variable selection is being used to produce the optimum variables for the model, it is better to use two data sets, using one for establishing the best number of variables, and the second for producing the results.

The remainder of this section deals with supervised methods.

6.3.1 *Training and Testing*

In order to create a prediction, the data must be divided up into a *training set* (on which the model is formed) and a *query* or *test set* (using which the model is tested, and the best number of factors or epochs established).

Since most supervised methods of forming a model will use the query set in order to establish the optimal number of factors (or epochs, in the case of an artificial neural network), a completely independent *validation* set is required, to ensure that the model is valid. This data set will not have been seen by the model in any form at any time. The only reason for not using a third data set is where there are insufficient objects to form a meaningful model if the data are divided into three. In such cases, it must be remembered that the results may appear better than they really are, and this fact should be noted in any results. Other methods of forming a model are able to establish the optimal factors or epochs from the training set alone, for example by dividing the training set into two and alternately training the model on one section and testing on the other. In such cases, two data sets are probably sufficient.

Replicates should always be kept in the same data set; not to do so would definitely classify as 'cheating'. If one of two replicates were in the training set, it would be expected that its partner in the validation set would be predicted with accuracy.

6.3.2 *The Extrapolation Problem*

Statistical models are not in general able to extrapolate; that is to say, if for a given variable, the training set data are in the range 3 to 4, there is no way a meaningful prediction can be made if the validation data contains a 5. This means that the training set should encompass the whole of the query and validation sets.

For quantification (e.g. prediction of concentration in a solution), the solution is easy: objects should be placed alternately in the training, query and (where there are sufficient objects) validation sets, ensuring that the objects with the lowest and highest value in the target being predicted are in the training set.

For classification (e.g. identification of country of origin or variety of a sample, or the bacterial strain), it is not quite so straightforward, as it is difficult to know within a class (country of origin, etc.) which data lie at the edge of the spectrum. This may, however, be achieved, by examining the data for each

object within a group, and determining how the variables lie with respect to those of other objects in the same class. With *n* variables, this means looking in *n* dimensional space; clearly not a task that is possible for the mere human. To facilitate this, a program called MultiPlex has been written by Dr. Alun Jones of UWA (an extension of the duplex algorithm described in [166]). Using this program will ensure that the objects are divided between training, query and, if desired, validation, sets appropriately. Provided that any replicates in the samples are correctly identified, it will also ensure that replicates are placed in the same set.

7 Concluding Remarks

"Organisms are not billiard balls, struck in deterministic fashion by the cue of natural selection and rolling to optimal positions on life's table. They influence their own destiny in interesting, complex and comprehensible ways." – S.J. Gould (1993) Evolution of organisms. In: Boyd CAR, Noble D (eds) The logic of life. Oxford University Press, p 5

Biological systems are indeed complex (and this differs from 'complicated' – [167]), but many of their most important features that are of interest to us for specific purposes are in fact of low dimensionality. The key to understanding them then lies in acquiring large amounts of the right kind of data which can act as the inputs to intelligent and sophisticated data processing and machine learning algorithms. These approaches alone – especially those based on induction – will help us unravel their workings [168].

Acknowledgments. We thank the BBSRC, the EPSRC and HEFCW for financial support of our collaborative programme in Analytical Biotechnology, Spectrometry, Chemometrics and Machine Learning.

References

1. Kell D (1998) Trends in Biotechnology 16:491
2. Oliver SG, Winson MK, Kell DB, Baganz F (1998) Trends in Biotechnology 16:373
3. DeRisi JL, Iyer VR, Brown PO (1997) Science 278:680
4. de Saizieu A, Certa U, Warrington J, Gray C, Keck W, Mous J (1998) Nature Biotechnol 16:45
5. Humphery-Smith I, Cordwell SJ, Blackstock WP (1997) Electrophoresis 18:1217
6. Wilkins MR, Williams KL, Appel RD, Hochstrasser DF (1997) Proteome research: new frontiers in functional genomics. Springer, Berlin Heidelberg New York
7. Blackstock WP, Weir MP (1999) Tibtech 17:121
8. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Proc Natl Acad Sc 95:14863
9. Kell DB (1980) Process Biochemistry 15:18
10. Clarke DJ, Kell DB, Morris JG, Burns A (1982) Ion-Selective Electrode Rev 4:75
11. Lee MS, Hook DJ, Kerns EH, Volk KJ, Rosenberg IE (1993) Biological Mass Spectrometry 22:84
12. Heinzle E, Moes J, Griot M, Kramer H, Dunn IJ, Bourne JR (1984) Analytical Chimica Acta 163:219
13. Heinzle E, Oeggerli A, Dettwiler B (1990) Analytica Chimica Acta 238:101

14. Matz G, Loogk M, Lennemann F (1998) Journal of Chromatography A 819:51
15. Namdev PK, Alroy Y, Singh V (1998) Biotechnology Progress 14:75
16. Bohatka S, langer G, Szilagyi J, Berecz I (1983) International Journal of Mass Spectrometry 48:277
17. Dongre AR, Hayward MJ (1996) Analytica Chimica Acta 327:1
18. Heinzle E, Kramer H, Dunn IJ (1985) Biotechnology and Bioengineering 27
19. Lauritsen FR, Choudhury TK, Dejarme LE, Cooks RG (1992) Analytica Chimica Acta 266:1
20. Lauritsen FR, Nielsen LT, Degn H, Lloyd D, Bohatka S (1991) Biological Mass Spectrometry 20:253
21. Lloyd D, Ellis JE, Hillman K, Williams AG (1992) Journal of Applied Bacteriology 73
22. Weaver JC (1982) Continuous monitoring of volatile metabolites by a mass spectrometer. In: Cohen JS (ed) Noninvasive Probes of Tissue Metabolism. J Wiley, New York
23. Irwin WJ (1982) Analytical Pyrolysis: A Comprehensive Guide. Marcel Dekker, New York
24. Meuzelaar HLC, Haverkamp J, Hileman FD (1982) Pyrolysis Mass Spectrometry of Recent and Fossil Biomaterials. Elsevier, Amsterdam
25. Goodacre R, Kell DB (1996) Current Opinion in Biotechnology 7:20
26. Magee JT (1993) Whole-organism fingerprinting. In: Goodfellow M, O'Donnell AG (eds). Handbook of New Bacterial Systematics. Academic Press, London, p 383
27. Tas AC, Vandergreef J (1994) Mass Spectrometry Reviews 13:155
28. Goodacre R, Berkeley RCW (1990) FEMS Microbiology Letters 71:133
29. Goodacre R, Berkeley RCW, Beringer JE (1991) Journal of Analytical and Applied Pyrolysis 22:19
30. Goodacre R, Rooney PJ, Kell DB (1998) Journal of Antimicrobial Chemotherapy 41:27
31. Goodacre R, Trew S, Wrigley-Jones C, Neal MJ, Maddock J, Ottley TW, Porter N, Kell DB (1994) Biotechnology and Bioengineering 44:1205
32. Schulten H-R, Lattimer RP (1984) Mass Spectrometry Reviews 3:231
33. Van de Meent D, de Leeuw JW, Schenck PA, Windig W, Haverkamp J (1982) Journal of Analytical and Applied Pyrolysis 4:133
34. Heinzle E, Kramer H, Dunn IJ (1985) Analysis of biomass and metabolites using pyrolysis mass spectrometry. In: Johnson A (ed) Modelling and Control of Biotechnological Processes. Pergamon, Oxford
35. Sandmeier EP, Keller J, Heinzle E, Dunn IJ, Bourne JR (1988) Development of an on-line pyrolysis mass spectrometry system for the on-line analysis of fermentations. In: Hienzle E, Reuss M (eds). Mass Spectrometry in Biotechnological Process Analysis and Control. Plenum, New York, p 209
36. Heinzle E (1992) Journal of Biotechnology 25:81
37. Goodacre R, Edmonds AN, Kell DB (1993) Journal of Analytical and Applied Pyrolysis 26:93
38. Goodacre R, Kell DB (1993) Analytica Chimica Acta 279:17
39. Goodacre R, Neal MJ, Kell DB (1994) Analytical Chemistry 66:1070
40. Goodacre R, Karim A, Kaderbhai MA, Kell DB (1994) Journal of Biotechnology 34:185
41. McGovern AC, Ernill R, Kara BV, Kell DB, Goodacre R (1999) Journal of Biotechnology 72:157–167
42. Broadhurst D, Goodacre R, Jones A, Rowland JJ, Kell DB (1997) Anal. Chim. Acta 348:71
43. Goodacre R, Trew S, Wrigley-Jones C, Saunders G, Neal MJ, Porter N, Kell DB (1995) Analytica Chimica Acta 313:25
44. McGovern AC, Broadhurst D, Taylor J, Gilbert RJ, Kaderbhai N, Small DAP, Kell DB, Goodacre R (1999) (in preparation)
45. Kang SG, Lee DH, Ward AC, Lee KJ (1998) Journal of Microbiology and Biotechnology 8:523
46. Kang SG, Kenyon RGW, Ward AC, Lee KJ (1998) Journal of Biotechnology 62:1
47. Schrader B (1995) Infrared and Raman spectroscopy: methods and applications. Verlag Chemie, Weinheim.
48. Ingle Jr JD, Crouch SR (1988) Spectrochemical Analysis, Prentice-Hall, London

49. Martin KA (1992) Applied spectroscopy reviews 27:325
50. Howard WW, Sekulic S, Wheeler MJ, Taber G, Urbanski FJ, Sistare FE, Norris T, Aldridge PK (1998) Applied Spectroscopy 52:17
51. Macaloney G, Draper I, Preston J, Anderson KB, Rollins MJ, Thompson BG, Hall JW, McNeil B (1996) Food and Bioproducts Processing 74:212
52. Yano T, Harata M (1994) Journal of Fermentation and Bioengineering 77:659
53. Marquardt LAA, M.A. Small,G.W. (1993) Anal chemistry 65:3271
54. Macaloney G, Hall JW, Rollins MJ, Draper I, Thompson BG, McNeil B (1994) Biotechnology Techniques 8:281
55. Brimmer PJ, Hall JW (1993) Canadian Journal of Applied Spectroscopy 38:155
56. Yano T, Aimi T, Nakano Y, Tamai M (1997) Journal of fermentation and bioengineering 84:461
57. Norris T, Aldridge PK (1996) Analyst 121:1003
58. Hall JW, McNeill B, Rollins MJ, Draper I, Thompson BG, Macaloney G (1996) Applied Spectroscopy 50:102
59. Riley MR, Rhiel M, Zhou X, Arnold MA (1997) Biotechnology and Bioengineering 55:11
60. McShane MJ, Cote GL (1998) Applied Spectroscopy 52:1073
61. Swierenga H, Haanstra WG, deWeijer AP, Buydens LMC (1998) Applied Spectroscopy 52:7
62. Cavinato AG, Mayes DM, Ge ZH, Callis JB (1990) Analytical Chemistry 62:1977
63. Ge ZC, AG Callis, JB (1994) Analytical Chemistry 66:1354
64. Vaccari G, Dosi E, Campi AL, Gonzalezvara A, Matteuzzi D, Mantovani G (1994) Biotechnology and Bioengineering 43:913
65. Vaccari G, Dosi E, Campi AL, Mantovani G (1993) Zuckerindustrie 118:266
66. Varadi M, Toth A, Rezessy J (1992) Application of NIR in a fermentation process. VCH Publishers, New York
67. Hammond SV (1992) NIR Analysis of Antibiotic Fermentations. In: Murray I, Cowe IA (eds) Making Light Work: Advances in Near-Infrared Spectroscopy. VCH Publishers, New York, p 584
68. Hammond SV (1992) Near-Infrared Spectroscopy – A Powerful Technique for At-Line and Online Analysis of Fermentations. In: Bose A (ed) Harnessing Biotechnology for the 21st Century: Proceedings of the Ninth International Symposium and Exhibition. American Chemical Society, Washington D.C., p 325
69. Validyanathan S, Macaloney G, McNeill B (1999) Analyst 124:157
70. Koza JR (1995) Proceedings of Wescon 95:E2. Neural-Fuzzy Technologies and Its Applications
71. McShane MJ, Cote GL, Spiegelman C (1997) Applied Spectroscopy 51:1559
72. Bangalore AS, Shaffer RE, Small GW, Arnold MA (1996) Analytical Chemistry 68:4200
73. Shaffer RE, Small GW, Arnold MA (1996) Analytical Chemistry 68:2663
74. Hassell DC, Bowman EM (1998) Applied Spectroscopy 52:A18
75. Guzman M, deBang M, Ruzicka J, Christian GD (1992) Process Control and Quality 2:113
76. Alberti JC, Phillips JA, Fink DJ, Wacasz FM (1985) Biotechnology and Bioengineering Symp. 15:689
77. Picque D, Lefier D, Grappin R, Corrieu G (1993) Analytica Chimica Acta 279:67
78. Fayolle P, Picque D, Corrieu G (1997) Vibrational Spectroscopy 14:247
79. Hayakawa K, Harada K, Sansawa H (1997) Abstracts of the 8th European Congress on Biotechnology 275
80. Wilson RH, Holland JK, Potter J (1994) Chemistry in Britain 30:993
81. Winson MK, Goodacre R, Timmins ÉM, Jones A, Alsberg BK, Woodward AM, Rowland JJ, Kell DB (1997) Analytica Chimica Acta 348:273
82. Kell DB, Winson MK, Goodacre R, Woodward AM, Alsberg BK, Jones A, Timmins ÉM, Rowland JJ (1998) DRASTIC (Diffuse Reflectance Absorbance Spectroscopy Taking In Chemometrics). A novel, rapid, hyperspectral, FT-IR-based approach to screening for biocatalytic activity and metabolite overproduction. In: Kieslich K (ed) New Frontiers in Screening for Microbial Biocatalysts. Elsevier Science B.V., The Netherlands, p 61

83. Winson MK, Todd M, Rudd BAM, Jones A, Alsberg BK, Woodward AM, Goodacre R, Rowland JJ, Kell DB (1998) A DRASTIC (Diffuse Reflectance Absorbance Spectroscopy Taking in Chemometrics) approach for the rapid analysis of microbial fermentation products: quantification of aristeromycin and neplanocin A in *Streptomyces citricolor* broths. In: Kieslich, K (ed) New Frontiers in Screening for Microbial Biocatalysts. Elsevier Science B.V., The Netherlands, p 185
84. Kell DB, Sonnleitner B (1995) Trends Biotechnol. 13:481
85. Montague GA (1997) Monitoring and control of fermenters, Institute of Chemical Engineers London
86. Pons M-N (1991) Bioprocess monitoring and control. Hanser, Munich
87. Kell DB, Markx GH, Davey CL, Todd RW (1990) Trends in Analytical Chemistry 9:190
88. Adar F, Geiger R, Noonan J (1997) Applied Spectroscopy Reviews 32:45
89. Chase B (1994) Appl Spectrosc 48:14 A
90. Gerrard DL (1994) Analytical Chemistry 66:R 547
91. Góral J, Zichy V (1990) Spectrochimica Acta 46 A:253
92. Graselli JG, Bulkin BJ (1991) Analytical Raman spectroscopy, John Wiley, New York
93. Hendra P, Jones C, Warnes G (1991) Fourier Transform Raman Spectroscopy. Ellis Horwood, Chichester
94. Hendra PJ, Wilson HMM, Wallen PJ, Wesley IJ, Bentley PA, Arruebarrena Baez M, Haigh JA, Evans PA, Dyer CD, Lehnert R, Pellow-Jarman MV (1995) Analyst 120:985
95. Hirschfeld T, Chase B (1986) Appl Spectrosc 40:133
96. Keller S, Schrader B, Hoffmann A, Schrader W, Metz K, Rehlaender A, Pahnke J, Ruwe M, Budach W (1994) J Raman Spectrosc 25:663
97. Naumann D, Keller S, Helm D, Schultz C, Schrader B (1995) Journal of Molecular Structure 347:399
98. Parker SF (1994) Specrochim. Acta 50 A:1841
99. Puppels GJ, Colier W, Olminkhof JHF, Otto C, Demul FFM, Greve J (1991) Journal of Raman Spectroscopy 22:217
100. Puppels GJ, Greve J (1993) Adv Spectrosc 20 A:231
101. Puppels GJ, Schut TCB, Sijtsema NM, Grond M, Maraboeuf F, Degrauw CG, Figdor CG, Greve J (1995) Journal of Molecular Structure 347:477
102. Schrader B, Baranovic G, Keller S, Sawatzki J (1994) Fresenius Journal of Analytical Chemistry 349:4
103. Treado PJ, Morris MD (1994) Applied spectroscopy reviews 29:1
104. Twardowski J, Anzenbacher P (1994) Raman and infrared spectroscopy in biology and biochemistry. Ellis Horwood, Chichester
105. Carrabba MM, Spencer KM, Rich C, Rauh D (1990) Appl Spectrosc 44:1558
106. Kim M, Owen H, Carey PR (1993) Applied Spectroscopy 47:1780
107. Puppels GJ, Huizinga A, Krabbe HW, Deboer HA, Gijsbers G, Demul FFM (1990) Review of Scientific Instruments 61:3709
108. Tedesco JM, Owen H, Pallister DM, Morris MD (1993) Analytical Chemistry 65:A 441
109. Treado PJ, Morris MD (1990) Spectrochimica Acta Reviews 13:355
110. Turner JF, Treado PJ (1996) Applied Spectroscopy 50:277
111. Griffiths PR, de Haseth JA (1986) Fourier transform infrared spectrometry. John Wiley, New York
112. Williams KPJ, Pitt GD, Batchelder DN, Kip BJ (1994) Applied spectroscopy 48:232
113. Williams KPJ, Pitt GD, Smith BJE, Whitley A, Batchelder DN, Hayward IP (1994) Journal of Raman Spectroscopy 25:131
114. Erckens RJ, Motamedi M, March WF, Wicksted JP (1997) Journal Of Raman Spectroscopy 28:293
115. Wicksted JP, Erckens RJ, Motamedi M, March WF (1995) Applied Spectroscopy 49:987
116. Shope TB, Vickers TJ, Mann CK (1987) Applied Spectroscopy 41:908
117. Gomy C, Jouan M, Dao NQ (1988) Analytica Chimica Acta 215:211
118. Gomy C, Jouan M, Dao NQ (1988) Comptes Rendus De L Academie Des Sciences Serie Ii-Mecanique Physique Chimie Sciences De L Univers Sciences De La Terre 306:417

119. Spiegelman CH, McShane MJ, Goetz MJ, Motamedi M, Yue QL, Cote GL (1998) Anal Chem 70:35
120. Harris CM, Todd RW, Bungard SJ, Lovitt RW, Morris JG, Kell DB (1987) Enzyme Microbial Technol 9:181
121. Kell DB (1987) The principles and potential of electrical admittance spectroscopy: an introduction. In: Turner APF, Karube I, Wilson GS (eds). Biosensors; fundamentals and applications. Oxford University Press, Oxford, p 427
122. Pethig R, Kell DB (1987) Phys Med Biol 32:933
123. Kell DB, Kaprelyants AS, Weichart DH, Harwood CL, Barer MR (1998) Antonie van Leeuwenhoek 73:169
124. Kell DB, Davey CL (1992) Bioelectrochemistry and Bioenergetics 28:425
125. Nicholson DJ, Kell DB, Davey CL (1996) Bioelectrochemistry and Bioenergetics 39:185
126. Davey CLK, D. B. (1998) Bioelectrochemistry and Bioenergetics 46:91
127. Davey CL, Kell DB (1998) Bioelectrochemistry and Bioenergetics 46:105
128. Debye P (1929) Polar Molecules. Dover Press, New York
129. Woodward AM, Kell DB (1990) Bioelectrochemistry and Bioenergetics 24:83
130. Davey CL, Kell D B (1990) The dielectric properties of cells and tissues what can they tell us about the mechanisms of field/cell interactions. In: O'Connor ME, Bentall RHC, Monahan JC (eds). Emerging Electromagnetic Medicine. Springer, Berlin Heidelberg New York, p 19
131. Kell DB, Astumian RD, Westerhoff HV (1988) Ferroelectrics 86:59
132. Martens H, Næs T (1989) Multivariate calibration. John Wiley, Chichester
133. Woodward AM, Gilbert RJ, Kel DB (1999) Bioelectrochemistry and Bioenergetics (in press)
134. Woodward AM, Davies EA, Denyer S, Olliff C, Kell DB (1999) Submitted for publication in Journal of Electroanalytical Chemistry
135. Woodward AM, Jones A, Zhang X-Z, Rowland JJ, Kell DB (1996) Bioelectrochemistry and Bioenergetics 40:99
136. Jeon SI, Lee JH, Andrade JD, de Gennes PG (1991) Journal of Colloidal and Interface Science 142:149
137. Koza JR (1992) Gentic programming: on the programming of computers by means of natural selection, MIT press Cambridge, MA
138. McShea A, Woodward AM, Kell DB (1992) Bioelectrochemistry and Bioenergetics 29:205
139. Davies EA, Olliff C, Wright I, Woodward AM, Kell DB (1999) Bioelectrochemistry and Bioenergetics (in press)
140. Davies EA, Woodward AM, Kell DB (1999) Bioelectromagnetics (in press)
141. Davey HM, Kell DB (1996) Microbiol Rev 60:641
142. Shapiro HM (1995) Practical flow cytometry, 3rd edn. Alan R. Liss, New York
143. Münch T, Sonnleitner B, Fiechter A (1992) Biotechnology 22:329
144. Münch T, Sonnleitner B, Fiechter A (1992) Journal of Biotechnology 24:299
145. Srienc F, Arnold B, Bailey JE (1984) Biotechnology and Bioengineering 26:982
146. Müller S, Lösche A, Bley T, Scheper T (1995) Applied Microbiology and Biotechnology 43:93
147. Müller S, Lösche A, Bley T (1993) Acta Biotechnol 13:289
148. Gjelsnes O, Tangen R (1994) Norway patent WO 94/29695
149. Rønning Ø (1999) Genetic Engineering News 19:18
150. Degelau A, Freitag R, Linz F, Middendorf C, Scheper T, Bley T, Müller S, Stoll P, Reardon KF (1992) Journal of Biotechnology 25:115
151. Zhao R, Natarjan A, Srienc F (1999) Biotechnology and Bioengineering 62:609
152. Neal MJ, Goodacre R, Kell DB (1994) On the analysis of pyrolysis mass spectra using artificial neural networks. Individual input scaling leads to rapid learning. in Proceedings of the World Congress on Neural Networks. International Neural Network Society San Diego
153. de Noord OE (1994) Chemometrics and Intelligent Laboratory Systems 23:65

154. Flury B, Riedwyl H (1988) Multivariate Statistics: A Practical Approach. Chapman and Hall, London
155. Seasholtz MB, Kowalski B (1993) Analytica Chimica Acta 277:165
156. Shaw AD, di Camillo A, Vlahov G, Jones A, Bianchi G, Rowland J, Kell DB (1996) Discrimination of Different Olive Oils using ^{13}C NMR and Variable Reduction. in Food Authenticity '96. Norwich, UK
157. Shaw AD, di Camillo A, Vlahov G, Jones A, Bianchi G, Rowland J, Kell DB (1997) Analytica Chimica Acta 348:357
158. Vlahov G, Shaw AD, Kell DB (1999) Accepted for publication in Journal of the American Oil Chemists Society
159. Shaw AD, Kaderbhai N, Jones A, Woodward A, Goodacre R, Rowland J, Kell DB (1999) Accepted for publication in Applied Spectrometry
160. Boschelle O, Giomo A, Conte L, Lercker G (1994) La Rivista Italiana delle Sostanze Grasse 71:57
161. Hazen KHA, MA Small, GW (1994) Applied spectroscopy 48:477
162. Gilbert RJ, Goodacre R, Woodward AM, Kell DB (1997) Analytical Chemistry 69:4381
163. Gilbert RJ, Goodacre R, Shann B, Taylor J, Rowland JJ, Kell DB (1998) Genetic Programming based Variable Selection for High Dimensional Data in Proceedings of Genetic Programming 1998. Morgan Kaufmann, Madison, Wisconsin, USA
164. Taylor J, Winson MK, Goodacre R, Gilbert RJ, Rowland JJ, Kell DB (1998) Genetic Programming in the Interpretation of Fourier Transform Infrared Spectra: Quantification of Metabolites of Pharmaceutical Importance in Genetic Programming 1998. Morgan Kaufmann, Madison, Wisconsin, USA
165. Taylor J, Goodacre R, Wade W, Rowland JJ, Kell DB (1998) FEMS Microbiology Letters 160:237
166. Snee RD (1977) Technometrics 19:415
167. Bialy H (1999) Nature Biotechnology, in the press
168. Kell DB, Mendes P (1999) Snapshots of systems: metabolic control analysis and biotechnology in the post-genomic era. In: Cornish-Bowden A, Cardenás ML (eds). Technological and Medical Implications of Metabolic Control Analysis (in press) (and see http://gepasi.dbs.aber.ac.uk/dbk/mca99.htm). Plenum Press, New York