Authors' final version of paper published in *Biotechnology and Bioengineering*, 63. pp 684-693, 1999

Near Infrared Spectroscopy for Bioprocess Monitoring and Control

Ken S. Y. Yeung[1,2], Mike Hoare[2], Nina F. Thornhill[3*], Tom Williams[4], Jeetendra D. Vaghjiani[2]
[1] Centre For Advanced Instrumentation Systems, [2] The Advanced Centre For Biochemical Engineering, [3] Department of Electronic and Electrical Engineering, University College London, Torrington Place, London WC1E 7JE, [4] Sira Electro-Optics Division, Sira UK Ltd, South Hill, Chislehurst, Kent BR7 5EH
*Author for correspondence

Abstract

This paper describes the calibration of a spectroscopic scanning instrument for the measurement of selected contaminants in a complex biological process stream. Its use is for the monitoring of a process in which contaminants are to be removed selectively by flocculation from yeast cell homogenate. The main contaminants are cell debris, protein and RNA. A low cost instrument has been developed for sensitivity in the region of the NIR spectrum (from 1900 to 2500 nm) where preliminary work found NIR signatures from cell debris, protein and RNA. Calibration models have been derived using a multivariate method for concentrations of these contaminants such as would be found after the flocculation process.

Two strategies were compared for calibrating the NIR instrument. In one case samples were prepared by adding materials representative of the contaminants to clarified yeast homogenate so the contaminant levels were well known but outside the range of interest. In the other, where samples were like those from the process stream after flocculation and floc removal there was uncertainty of analysis of contaminant level but the calibration was in the range of interest.

Calibration using process stream samples gave results close to those derived from traditional assays. When the calibration models were used to predict the contaminant concentrations in previously unseen samples the correlation coefficients between measurements and predictions were above 90% in all cases but one. The prediction errors were similar to the errors in the traditional assays.

1. Introduction

Near infrared (NIR) spectroscopy has become well established in food and agricultural products (Hart et al., 1962; Rosenthal, 1973; Williams and Norris, 1987). Quantitative analysis of protein, moisture and oil can be completed within seconds using portable NIR instruments. With recent advances in instrumentation and multivariate analysis (Lysaght et al., 1991; Mayes and Callis, 1989; Beebe and Kowalski, 1987) NIR spectroscopic instruments are finding their way into various applications to serve important monitoring and control purposes (Yu and Phillips, 1992). Some of these applications include textiles, polymers, biomedicine, petroleum and bioprocesses (Donald and Ciurczak, 1992).

NIR has been demonstrated for use with fermentation broths (Vaccari *et. al*), in characterisation of feed material (Kasprow *et. al*., 1998), and with insect cell culture media and animal cell culture supernatants (Riley et. al., 1996; Yano and Harata, 1994; Harthun *et. al*, 1997).

In bioprocesses, the control task is often difficult due to lack of information relating to the concentration of various key components (Hatch and Hermann, 1990). For example, the traditional spectroscopic techniques in UV are useful in the analysis of protein or nucleic acid in solution (Junker et al., 1989), but are problematic when applied to turbid or multicomponent mixtures. Generally, analysis of biological compounds are performed off-line by wet chemical assays or high performance liquid chromatography (HPLC). These assays require sample preparation and can be time consuming. Moreover, they may involve hazardous or environmentally unacceptable solvents. Thus a rapid and direct instrument will significantly enhance process analysis and control.

This paper reports the calibration of a low cost NIR spectrophotometer and its use for at-line monitoring of complex bioprocesses. The recovery of a yeast intracellular enzyme, alcohol dehydrogenase (ADH), from an unclarified yeast cell homogenate by a selective flocculation process has been chosen for this exploration (Salt *et. al.*, 1995). The ADH is an intracellular protein produced in *Saccharomyces cerevisiae* (Baker's yeast). In the recovery process, *S. cerevisiae* has to be broken to release the ADH. This process also releases unwanted contaminants into the process stream. These include cell debris, protein, RNA, DNA and lipid all of which affect the

performance of subsequent high resolution purification operations. It is the purpose of the recovery process to remove these contaminants (Atkinson and Jack, 1973).

Empirical relationships are used in quantitative analysis of NIR spectra, based on the correlation of the absorption of NIR radiation and the analytical data. Often multivariate analysis such as partial least squares (PLS) is used. Multiple least-squares regression (MLR) is applied in some cases to simple tasks where a distinct spectral response is identified for the constituent (Donald and Ciurczak, 1992). In the system studied in this paper, PLS was adequate to capture the spectral signals and to correlate these with the components of interest.

This paper compares two strategies for the preparation of calibration samples, as explained in Section 2. Section 2 also discusses decisions about which data to use in the PLS calibration. Section 3 gives the materials and methods. Section 4 presents the results, where it becomes clear which calibration strategy was superior. Section 4 also applies the results to the task of process monitoring while section 5 makes conclusions and recommendations.

We have completed an engineering application by optimisation of the calibration strategy and the demonstration of the potential of NIR in the monitoring of a flocculation process. Our experiences with NIR and PLS in the monitoring of biological components match closely those reported by Harthun *et. al*., (1997). For instance, we and they faced similar choices of spectral range, the number of factors and the sizes of calibration sets.

2. Calibration strategy

2.1 Calibration using *Add-Back* and *process stream* samples

In this work, two sets of calibration samples were used for establishing calibration models on cell debris, protein and RNA. In the first set of calibration samples, controlled level of contaminants were added back into the yeast homogenate so that contaminants in the calibration samples were defined. The concept is similar to the 'spiking' used by Riley *et. al*. (1997). These samples are termed *add-back* samples.

Other than for debris it is often difficult to achieve selective and accurate removal of any one contaminant. In *add-back* calibration contaminants were added back to the clarified yeast homogenate from which cell debris (but not protein or RNA) had been removed. The amounts of added protein and RNA were designed to increase the contaminants present by a similar value as those present in the clarified homogenate. Hence while these contaminant levels were well defined, their concentrations were outside the range of interest for the process. A difficulty with these samples is that the contaminants added back were not identical in chemical or physical conformation to the contaminants occurring naturally in the homogenate, a point that is elaborated later.

The second set of calibration samples were similar to those obtained during the flocculation process and were termed *process stream* samples. Assays on protein and RNA of these calibration samples are necessary in order for calibration modelling.

Flocculated samples were prepared on laboratory scale. These samples were centrifuged to remove the flocculated contaminants and the remaining supernatants were used for calibration and validation. Here the uncertainty in the chemical analysis impacts on the calibration process. In particular, the measurement error of the RNA assay has implications for the structure of the PLS calibration model.

DNA is also a contaminant of the process stream, although at a lower level than RNA. A comparison of synthetic samples having similar concentrations of RNA and DNA showed the NIR response to DNA was much less than the response to RNA. It was therefore decided to focus on the detection of RNA as a measure of the nucleic acid contamination.

2.2 Multivariate calibration

A partial least square (PLS) approach was used for all calibrations and predictions. PLS is a regression between the spectra of calibration samples (the spectral information or **X** matrix) and the analytical data of these calibration samples (the concentration or **Y** matrix). The PLS procedure has been widely reported elsewhere (Geladi and Kowalski, 1986; Höskuldsson, 1988). The key steps are data selection, application of the PLS algorithm and the determination of the number of factors to use in the model.

Selection of data is an important step in a regression procedure such as PLS. The aim is to use the available data in such a way as to meet the objectives of the work. Prediction errors in regression arise from three sources. They are (a) the errors in the **Y** (concentration) data (b) errors in the **X** (spectral) data and (c) error in the choice of the structure of the model, for instance a failure to capture non-linearity. Here, the errors in the concentration data dominate. A full multivariate calibration would combine all the spectra with all the concentration data to make simultaneous predictions of all three contaminants. The concern is that the large measurement error in the RNA assay would propagate to and degrade predictions of the other concentrations. Therefore a comparison has been made of the simultaneous approach and an approach in which three separate calibration models were derived, one for each contaminant. Data selection in the latter case involves selection of the appropriate column of the concentration matrix and, in the case of RNA, the exclusion of any samples with gross errors.

Selection of the spectral data involves a choice of the wavelength range. The range selected was 1900nm to 2500nm because preliminary work showed spectral responses across in this range to the contaminants. The response can be highlighted through an inspection of the loading profiles for the first few principal components in the PLS analysis. Kasprow *et.al.* (1998) also selected their wavelength range through such considerations.

3. Materials and Methods

3.1 Process application

The composition of contaminants in the process stream is listed in Table 1 (Bulmer, 1992). The process objective was to remove the contaminants with minimum product loss. The process involved various stages of purification as illustrated in Figure 1. A polymer polyethyleneimine (PEI) was fed into the yeast homogenate to selectively flocculate the contaminants (Milburn *et al*., 1990). This allows an effective removal of the bulk of the contaminants by centrifugation. This clarified and partly purified yeast homogenate goes into one or two precipitation stages for further removal of protein before the final purification by HPLC.

3.2 Materials

Standards used were yeast ribonucleic acid (RNA, highly polymerised), bovine serum albumin (BSA, fraction V), both supplied by Sigma Chemical Ltd (Poole, Dorset, UK). The assay chemicals were orcinol, ferric chloride, perchloric acid. These were all supplied by BDH Chemicals Ltd (Poole, Dorset, UK). Polyethyleneimine was supplied by Fluka Chemicals (Dorset, UK).

3.3 Methods

3.3.1 Yeast homogenate

The yeast homogenate was prepared from packed Baker's yeast, *Saccharomyces cerevisiae*, supplied by Distillers Company Ltd. (Sutton, Surrey, UK). The Baker's yeast was re-suspended in 100 mM phosphate buffer, pH 6.5, to a final cell concentration of 500 g wet packed weight L^{-1}. The yeast suspension was disrupted using a high pressure homogeniser (Model Lab 40; APV Gaulin, APV, Crawley, Sussex, UK) for two discrete passes at 1200 bar and maintained at approximately 4°C by cooling. Following homogenisation the homogenate was clarified using a centrifuge (Beckman, J2-MI) at 16,000 rpm for 0.3 h at 4°C. Finally, a volume of 40 mL of clarified yeast homogenate was pipetted out from the cell debris and the less dense lipid layer and stored at 4°C prior to usage. The cell debris pellet was washed by re-suspending the pellet in phosphate buffer by vortexing and re-centrifuging at 16,000 rpm for 0.3 h at 4°C. The supernatant and lipid layer were discarded and the pellet of cell debris was mixed to homogeneity. This was then re-suspended in phosphate buffer to give a concentration of 150 g L^{-1} (wet weight) and it was used as the cell debris contaminant in the *add-back* calibration experiment.

3.3.2 *Add-back* calibration samples

Stock solutions of protein and RNA were prepared from standard materials. BSA was made up to a final concentration of 156 g L^{-1} with phosphate buffer (100mM, pH 6.5). RNA was made up to a final concentration of 69 g L^{-1} with the same phosphate buffer (100mM, pH 6.5), and cell debris stock was as previously described. Using the clarified yeast homogenate and the three stocks of suspensions, 36 different calibration samples were prepared.

3.3.2 *Process stream* samples

PEI was diluted into phosphate buffer (100mM, pH 6.5) at 2% w/v. The un-clarified yeast homogenate described above was diluted 1:1 with phosphate buffer to give homogenate equivalent to 250 g wet packed weight yeast L^{-1}. The PEI stock was added to this homogenate to cause flocculation. Four sample sets were prepared independently, each set containing ten samples with various extents of flocculation. The PEI stock concentrations used for each set of samples were 0, 2, 6, 10, 15, 20, 25, 30, 35 and 40 percents (volume/volume). Two sets of these *process stream* samples were used for calibration as before. The remaining two sets were used for calibration model validation purposes and for process monitoring experiments. The reason for using two sets for calibration was to give duplicate samples at each level.

3.3.3 Assays for biological contaminants

Cell debris (turbidity) was ascertained by measuring the supernatant absorption of an appropriate dilution of the samples at 650nm against a buffer blank. Protein was measured using the dye-binding method of Bradford (Bradford, 1976).

RNA was assayed using a method based on the orcinol assay (Bulmer, 1992; Munro and Fleck, 1966). The method was adapted for yeast as follows. Orcinol reagent was prepared by dissolving orcinol (3 gL^{-1}) in concentrated hydrochloric acid to which was added ferric chloride (10%w/v, 1 mL). Samples (100 μL) were precipitated with 60% perchloric acid (100 μL) in an Eppendorf tube and stored at 4°C for 24 h and then centrifuged (13,500 g, 0.12 h). The supernatant (100 μL) was mixed with sodium hydroxide (NaOH, 2M, 100 μL) and incubated (2 h, 37°C). Orcinol reagent (800 μL) was added to the samples which were then placed in boiling water for 0.3 h, then cooled and centrifuged (13,000 rpm, 0.6 h). The samples were read against a reagent (orcinol) blank at 665nm.

The standard deviations of these measurements were assessed by conducting several repeats. Some of them have also been reported in the literature (Bradford, 1976; Dehghani *et. al*., 1995). They were: 4 % for protein, 6% for optical density and 9.5% for RNA. The implications of the large error in the RNA assay are explored later.

3.3.4 NIR spectra

The NIR spectra were obtained using an in-house built spectrophotometer costing less than £5000 (Sira UK Ltd, Chislehurst, UK). The instrument was based on a 120

grooves/mm holographic grating covering 1100 to 2500 nm CP140-2-21 (Instrument S.A. Ltd, Middlesex, UK) and a single element lead sulfide (PbS) detector with two stages of thermal electric cooler controller, IRI 2700 and TC-328, respectively (Graseby Infrared, Orlando, USA). This scanning spectrophotometer was set to collect NIR radiation between 1900-2500 nm. The spectral data are collected every 4 nm with a total scan time of 210 s. The pre-amplified analogue output from the PbS detector is collected by a PC via a data acquisition card AT-MIO-16XE-50 (National Instruments, Berkshire, UK).

3.3.5 Multivariate Analysis

An IBM 486 compatible PC was used for all analyses. All spectra were baseline corrected using a pre-scanned reference spectrum (distilled water). The transmission spectra were then smoothed by a spline routine (Thornhill *et al.*, 1994). Since the spectral data are smoothed and baseline corrected their reproducibility is high; errors in the spectral data were therefore considered negligible. All multivariate calibrations and predictions were carried out by a PLS routine that is available in a commercial software package, Unscrambler version 5.0 (CAMO, Trondheim, Norway).

The number of factors was assessed by the cross validation procedure (Martens and Naes, 1989) that was supplied by the Unscrambler package. Randomly selected data sets are omitted from the calibration and used instead for validation, a procedure that is repeated several times in order to generate a statistical sample from which to calculate an average prediction error. In *add-back* calibration, for instance, five at a time of the 36 *add-back* spectra were omitted. Unscrambler recommends the number of factors giving the smallest average prediction error.

Standard errors of prediction (SEP) are shown with all PLS predictions. The SEP indicates the standard deviation of the errors between the predicted and measured values. In the case where predictions were made on unseen validation samples the offset is also given, where the offset is the mean value of the error.

4. Results and Discussion

4.1 Calibration using *add-back* samples

The transmission spectra of all 36 *add-back* calibration samples are shown in Figure 2. The spectra of samples with low, medium and high concentration of cell debris are grouped together and shown in the three windows on the right hand side. Within

each window, the effect of protein on the NIR spectra can also be visually identified for the samples with low cell debris in the top window where different protein levels cause three distinct bands. It is not possible to see the influence of RNA by eye, however, nor the influence of protein at other levels of cell debris. The conclusion from the visual inspection is that the spectra are sensitive to the concentrations of contaminants but that PLS analysis will be needed to capture the more subtle effects.

Table 2 presents results for the three *add-back* calibration models for optical density, protein and RNA. The correlation coefficients are all above 89%, and the number of factors range from 1 to 5. The fact that the PLS procedure reduced 36 variables to a maximum of four or five shows that the NIR spectra were capturing real influences from the contaminants. In their challenging application, Harthun *et.al*. (1997) found similar reductions for analysis of $mg \cdot L^{-1}$ concentrations of protein in animal cell culture supernatant, typically needing 4-9 factors with sets of 25 calibration samples.

Figures 3(a), 3(c) and 3(e) show the back-predicted calibration samples versus the known values. For protein and RNA it is better to use expected concentrations than the measured ones because the sample preparation error is smaller than the variability in the Bradford assay. Such a use of the data is justified because the variance of the prediction error in a regression procedure depends upon the variance of errors in the **Y**-data and one should therefore use the best information available. The OD measurement is used in Figure 3(a) because it has a smaller error than the expected value. An OD measurement would be needed on the stock material in order to calculate the expected OD but the stock material would need $\times 100$ dilution to bring it in range. That dilution introduces additional error.

There are fewer RNA points shown on Figure 3(e) than the number of samples because some obvious outliers were removed from the data set. Their exclusion is justified because they had assignable causes known to arise during sample preparation from RNA stock; the RNA was difficult to dilute.

4.2 Calibration using *process stream* samples

The two sets of process stream calibration samples were combined to give a total of 20 samples. They were scanned at 650 nm for optical densities as a measure of cell debris concentration, and measured for protein and RNA. These measurements and the NIR spectra of the samples were used to establish three separate PLS calibration

models for cell debris, protein and RNA. Table 2 presents the correlation coefficients and numbers of factors used, while figures 3(b), 3(d)) and 3(f) show back-predicted results compared with the measurements. Here it can be seen that the SEP values are similar to but a little larger than the measurement error. Note that SEP cannot generally be smaller than the measurement error because it is the standard deviation of the distances between the measurements and the line with a gradient of +1. Even a perfect calibration which captured all relevant influences could not predict random measurement errors. The larger SEP for protein reflects a single poorly predicted point at a measured value of 26 $g \cdot L^{-1}$.

A comparison was also made with a PLS model in which all components were predicted simultaneously (Table 3). Overall, the correlations coefficients and SEPs for the simultaneous model were worse than for the three separate models. This finding can be explained by considering the influence of propagation of errors in the concentration measurements. For instance, the use of only the protein measurements in a protein calibration model eliminates the influences of the errors in the OD and RNA measurements. On the other hand, it also eliminates any useful information that those measurements may contain so that OD and RNA become unknown background influences. We judged that the three separate PLS models were better since they gave overall higher correlations and smaller SEP values than the simultaneous PLS model. Thus in this case it seems that the propagation of measurement error is the more dominant effect. We recommend exploring both approaches since it is hard to determine *a-priori* which effect will dominate.

4.3 Process monitoring

The ability of NIR to identify and determine the amount of residual contaminants in a process stream gives two advantages. Firstly, NIR can be an indicator for the next or final stage of a recovery process. For example, if the levels of the contaminants are high then the stream should be stopped from going onto high resolution chromatography, since the columns used can be easily damaged by contaminants. Secondly, NIR gives important information for control of the process to maximise removal of contaminants.

NIR monitoring of two new previously unseen sets of flocculated samples was carried out and compared to the assayed measurements. These new flocculated samples

provided the opportunity for validation of the PLS models. The focus towards the process application means that the *add-back* calibration model was validated with samples like those from the process stream, not with unseen *add-back* samples.

The results are listed in Table 4 and plotted in Figure 4 where the profiles are as expected for the flocculation process. They show that as PEI feed concentration increases more cell debris, protein and RNA are removed from the yeast homogenate.

In all cases the *process stream* calibrated models gave better predictions than the *add-back* calibrated models, as shown by the smaller offsets and generally smaller SEP values. The optical density and protein concentrations from the *process stream* models match the measured values accurately and the RNA predictions from the *process stream* models are close when one considers the standard deviations of the measured values. A reason for poor performance of the *add-back* models is that the lower OD and protein concentrations presented in the process streams require back-extrapolation outside the ranges in which they were calibrated. In addition (section 4.5) physical differences between the *add-back* and *process stream* calibration samples also contribute to errors of prediction using *add-back* calibration. The conclusion is that the calibration model established from *process stream* samples should be used for process monitoring.

4.4 Establishing the control set point

The calibrated NIR system can monitor all three prime contaminants frequently during operation of the flocculation process and aid the selection of the appropriate PEI feed control set point. . The measured values in Figure 4 show that the removal of the cell debris, protein and RNA tends to occur simultaneously. They each show an 'elbow' in the concentration profile with a steep gradient for PEI additions below about 20% v/v. Above the elbow, however, it is known that over-flocculation by excessive PEI feed can reverse the efficiency of contaminant removal, as well as leaving excessive PEI in the process stream. Therefore the target should be to maintain operation at the elbow. Process disturbances such as the ionic strength of the yeast homogenate mean that the elbow will not always be achieved at exactly 20% PEI, however. Thus the process might be controlled by adjusting PEI feed so that protein levels monitored by NIR are always maintained at the level of the elbow.

Alternatively, the process might be controlled in an adaptive manner in which the position of the elbow is mapped out once per day or at the start of processing of a new

batch of material. Rapid NIR monitoring of protein or other contaminants for a range of PEI addition rates would determine the current position of the elbow and thus a suitable operating point for the next few hours or days of running.

A bootstrapping approach to process development can be achieved through the use of NIR. The *process stream* calibration has shown that it is feasible to use just 20 calibration samples even though 20 is a small number; 50 to 100 independent samples would be preferred. The result demonstrates that calibration using a small number of samples enables the NIR instrument to give immediate benefits to the process even while further refinements are under way. Additional calibration samples can be taken while the process is operating. For instance, measurements on process stream samples during an adaptive mapping of the elbow would provide useful additions to the calibration set. It is also worth noting that the SEPs are expected to reduce as the number of samples increases.

4.5 Comparison of *add-back* and *process stream* PLS models

Prediction of contaminants using the *add-back* calibration was not satisfactory (Figure 4). Apart for the issue of back-extrapolation, there is evidence that particle size is a cause of differences between the add-back and process-stream samples.

The calibration model for optical density provides insight. For OD, only one principal component is needed for the reconstruction. The optical density model is of the form:

$$N_i(\lambda) - \overline{N}(\lambda) = t_i p(\lambda) + e_i(\lambda)$$

$$OD_i - \overline{OD} = t_i q + f_i$$

where $N_i(\lambda)$ is the i'th spectrum, $\overline{N}(\lambda)$ is the mean of all the spectra, $p(\lambda)$ is the loading profile for the principal component and t_i is the score for the i'th spectrum. OD_i is the measured optical density of the *i*'th sample, t_i is the score, q is the loading for the OD data, and e_i and f_i are the residual errors. This one-PC model shows that the deviation of each NIR spectrum from the mean is proportional to $p(\lambda)$, apart from the error $e_i(\lambda)$.

It is instructive to compare the values of these quantities for the *add-back* and *process-stream* calibration sets. For instance, plots of $p(\lambda)$ (not shown) for the *add-back* and *process stream* models are similar to one another and reflect the underlying

shapes of the NIR spectra. Plots of $\overline{N}(\lambda)$ show an expected difference in amplitude and the parameters $\overline{OD}$ of the *add-back* and *process stream* models are different because the samples were prepared in different OD ranges..

The parameters q for the two cases were 0.127 and 0.257; any difference in q is unexpected and is significant because it indicates a different relationship between the intensity of the NIR spectral response and the OD values between the *add-back* and *process stream* cases.

The explanation for the difference is thought to be that the flocculating effect of residual PEI in the process stream leads to a difference in particle size between *add-back* and *process stream* samples.. Evidence for this suggestion is that both the NIR transmission and the OD have been observed to change when trace amounts of PEI were added to a sample (PEI causes flocculation but by itself does not present a significant NIR signature).

Transmission is a function of particle size as well as of the concentration of material in the sample (Kerker, 1969). Therefore the relationship between the optical density of a sample and the concentration of particles is constant only as long as the particle size does not change. The conclusion is that the PLS model developed for one particle size cannot be validated with samples having a different particle size.

4.6 Cost and performance of a factory NIR instrument

Cost would be a critical factor in any decision about factory-wide deployment of an NIR instrument. More grooves/mm improve the spectral resolution but they also increase the cost. Likewise reducing the scan step size improves the spectral resolution but increases the cost of mechanical components. Sensitivity is determined by the quality of the PbS detector and the detector electronics. A commercial NIR spectrometer typically costs £50000, has a spectral resolution of 2nm, a scan time of 40s and signal-to-noise ratio of at least 1000:1 (Yeung, 1998). By contrast, the UCL/SIRA instrument cost £5000, had spectral resolution of 4nm, a scan time of 210s and s.n.r. of 450:1.

Was the performance of the instrument compromised by its low cost or was it adequate for the application? The fact that the errors in prediction of contaminant concentrations were similar to those of the laboratory assays suggests that its performance is adequate. Likewise, the close match during validation (Figure 4) leads

to the conclusion that it is fit for the purpose of monitoring of the flocculation process.

5. Conclusions and recommendations

The paper has described the use of near infra red (NIR) analysis for monitoring of a flocculation process in which alcohol dehydrogenase (ADH) is recovered from a bioprocess stream containing yeast homogenate. The low cost NIR instrument showed responses to cell debris, protein and RNA. It could measure concentrations of these contaminants in the process stream with accuracy similar to that of the laboratory assays, and was fast enough to act as an at-line instrument

The NIR spectra were interpreted through the multivariate statistical technique of partial least squares (PLS). The PLS calibration model used the NIR spectra (1900-2500nm) together with either the known or assayed compositions of the samples that generated the spectra. The work compared two strategies for preparation of calibration materials. One strategy synthesised calibration samples (the *add-back* samples) by adding known amounts of the contaminants to previously clarified yeast homogenate. The second strategy used *process stream* samples which had been independently assayed for the level of each contaminant.

An advantage of the *add-back* approach was in the precise knowledge of the relative amounts of the contaminants. The *add-back* approach had the disadvantage, however, that the contaminants were present at higher than normal levels because there was a baseline level of contaminant in clarified yeast homogenate. *Process stream* samples, by contrast, contained the contaminants at their normal levels but the levels had to be measured by laboratory assay.

The calibration models were validated using previously unseen samples representative of the flocculation process. The calibration produced from the *process stream* samples was found to be more accurate than the *add-back* model because the *add-back* model had to back-extrapolate outside the range in which it was calibrated. It is also believed that residual amounts of PEI in the process stream led to a difference in particle size that was not modelled by the *add-back* calibration experiments. We therefore recommend use of *process stream* calibration for monitoring of cell debris, protein and RNA.

Acknowledgements

The authors thank Mark Bulmer and George Habib of the Advanced Centre for Biochemical Engineering, University College London, UK, for suggestions about chemical assays and flocculation process. Support from the EPSRC and DTI through the Sira/UCL Postgraduate Training Partnership, and from the BBSRC through the Advanced Centre for Biochemical Engineering is gratefully acknowledged.

6. References

Atkinson A. and Jack G.W. 1973. Precipitation of nucleic acids with polyethyleneimine and the chromatography of nucleic acids and proteins on immobilised polyethyleneimine. *Biochimica et Biophysica Acta.* **308**:41-52.

Beebe K.R. and Kowalski B.R. 1987. An introduction to multivariate calibration and analysis. *Anal. Chem.*. **59**:1007A-1017A.

Bradford M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.", *Analytical Biochemistry*. **72**:248-254.

Bulmer M. 1992. Ph.D. Thesis. *University of London*.

CAMO, Trondheim, Norway, 1993, *Unscrambler User Manual, Version 5.0.*

Dehghani M., Bulmer M., Gregory G.E. and Thornhill N.F. 1995. Measurement variability analysis for fermentation process assays. *Bioprocess Engineering.* **13**:239-243.

Donald A.B. and Ciurczak E.W. 1992. Handbook of near infrared analysis. Marcel Dekker.

Geladi P. and Kowalski B.R. 1986. Partial least squares regression: A tutorial. *Analytica Chimica Acta*. **185**:1-17.

Hart J.R., Norris K.H. and Golumbic C. 1962. Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extract. *Cereal Chemistry*. **39**:94-99.

Hatch R.T. and Hermann T. 1990. Monitoring and control of fermentations with an in-situ steam sterilizable optical density probe. presented at *Annual AIChE Mtg.*, Chicago, IL, November.

Harthun S., Matischak K., and Friedl P. 1997. Determination of recombinant protein in animal cell culture supernatant by near-infrared spectroscopy. *Analytical Biochemistry*. **251**: 73-78.

Höskuldsson A. 1988. PLS regression methods. *Journal of Chemometrics*. 2:211-228

Junker, B.H., Wang D.I.C. and Hatton T.A. 1989. Fluorescence sensing of fermentation parameters using fibre optics. *Biotech. Bioeng.* **32**:55-63.

Kasprow R.P., Lange A.J., and Kirwan D.J. 1998. Correlation of fermentation yield with yeast extract composition as characterised by near-infrared spectroscopy. *Biotechnology Progress*. **14**:318-325.

Kerker, M. 1969. The scattering of light and other electromagnetic radiation, Academic Press. London.

Lysaght M.J., Zee, J.A. and Callis, J.B. 1991. Laptop chemistry: A fiber-optic, field-portable, near-infrared spectrometer. *Rev. Sci. Instrume*. **62**:507-515.

Martens H. and Naes T. 1989. Multivariate calibrations. *J. Wiley & Sons*, New York.

Mayes D.M. and Callis J.B. 1989. A photodiode-array-based near-infrared spectrophotometer for 600-1100 nm wavelength region. *Applied Spectroscopy*. **43**:27-32.

Milburn P., Bonnerjea J., Hoare M., and Dunnill P. 1990. Selective flocculation of nucleic acids, lipids, and colloidal particles from a yeast cell homogenate by polyethyleneimine, and its scale-up. *Enzyme Microb. Technol*. **12**:527-532.

Munro H.N. and Fleck A. 1966, The determination of nucleic acids., *from Methods of Biochemical Analysis.* **65**:113-176.

Riley M.R., Rhiel M., Zhou X., Arnold M.A., and Murhammer D.W. 1997. Simultaneous measurement of glucose and glutamine in insect cell culture media by near infrared spectroscopy. *Biotechnology and Bioengineering*. **55**: 11-15.

Rosenthal R.D. 1973. The grain quality analyzer; A rapid and accurate means of determining the percent moisture, oil and protein in grain and grain products. *American Association of Cereal Chemists Meeting 12-13 April 1973*, Denver, Co, USA.

Salt D.E. Hay S. and Thomas O.R.T, Hoare M., and Dunnill P. 1995. Selective flocculation of cellular contaminants from soluble proteins using polyethyleneimine: A study of several organisms and polymer molecular weights. *Enzyme and Microbial Technology*. **17**:107-113.

Thornhill N.F., Manela M., Campbell J.A. and Stone K.M. 1994. Two methods of selecting smoothing splines applied to fermentation process data. *AIChE.J.* **40**:716-725.

Williams P.C. and Norris K. 1987. Near-infrared technology in agricultural and food industries. *American Association of Cereal Chemists*, St. Paul, Minneapolis.

Vaccari G., Dosi E., Campi A.L., Gonzalez-Vara y R A., Matteuzzi D., and Mantovani G. 1994. A near-infrared spectroscopy technique for the control of fermentation processes: An application to lactic acis fermentation. *Biotechnology and Bioengineering*. **43**:913-917.

Yano T., and Harata M. 1994. Prediction of the concentration of several constituents in a mouse-mouse hybridoma culture by near infra-red spectroscopy. *Journal of Fermentation and Bioengineering*. **77**:659-662.

Yeung K.S.Y. 1998. PhD Thesis. *University of London*.

Yu K. and Phillips J.A. 1992. The use of infrared spectroscopic techniques in monitoring and controlling bioreactors. *IFAC Modeling and Control of Biotechnical Process*, Pergamon Press, Oxford: 7-13.

FIGURE AND TABLE CAPTIONS

Figure 1. The use of NIR monitoring in the early stage of the ADH recovery process. A feed of PEI is introduced to yeast homogenate. The contaminants after flocculation and the first stage of centrifugation are monitored by NIR analysis and information is fed back to the controller to optimise the PEI feed.

Figure 2. The NIR transmission spectra of all *add-back* calibration samples. The spectra of samples with low, medium and high concentration of cell debris are shown in the three windows on the right hand side.

Figure 3. Predictions against measured values for the *add-back* and *process stream* calibration samples. The measurement errors are given by the horizontal bars, the vertical bars are the SEP values.

Figure 4. Process monitoring of two process stream using NIR analysis. The figure compares measured values with values predicted by the *add-back* and *process stream* models.

Table 1. Statistics of contaminants within Baker's yeast.

Table 2. Comparisons of performance of separate PLS models.

Table 3. Comparisons of performance of simultaneous PLS models.

Table 4. Comparison of performance of *add-back* and *process stream* calibration models on the prediction of two sets of *process stream* validation samples.

Bakers yeast	Overall Percent
Polysaccharide	42
Protein	40
DNA	2
RNA	5
Lipid	7
Ash	4

Table 1

	Separate PLS models					
	Add-back			***Process stream***		
	OD	**Protein**	**RNA**	**OD**	**Protein**	**RNA**
Factors	1	4	5	1	3	3
Correlation	0.982	0.891	0.937	0.981	0.915	0.919
S.E.P.	0.075	4.351	0.877	0.057	2.343	1.358

Table 2

	Simultaneous prediction PLS model					
	Add-back			***Process stream***		
	OD	**Protein**	**RNA**	**OD**	**Protein**	**RNA**
Factors		3			3	
Correlation	0.978	0.9003	0.6083	0.984	0.917	0.807
S.E.P.	0.088	4.618	3.514	0.051	2.458	3.978

Table 3

	Experiment 1			**Experiment 2**		
Add-back	**OD**	**Protein**	**RNA**	**OD**	**Protein**	**RNA**
Correlation	0.996	0.950	0.713	0.988	0.952	0.896
S.E.P.	0.015	3.781	3.724	0.030	3.432	2.950
Offset	0.358	-12.94	-12.00	0.343	-21.30	-0.061
Process stream	**Experiment 1**			**Experiment 2**		
Correlation	0.996	0.988	0.981	0.984	0.909	0.838
S.E.P.	0.028	0.858	0.787	0.060	2.222	2.556
Offset	0.006	4.136	-3.090	-0.037	4.711	-2.453

Table 4.

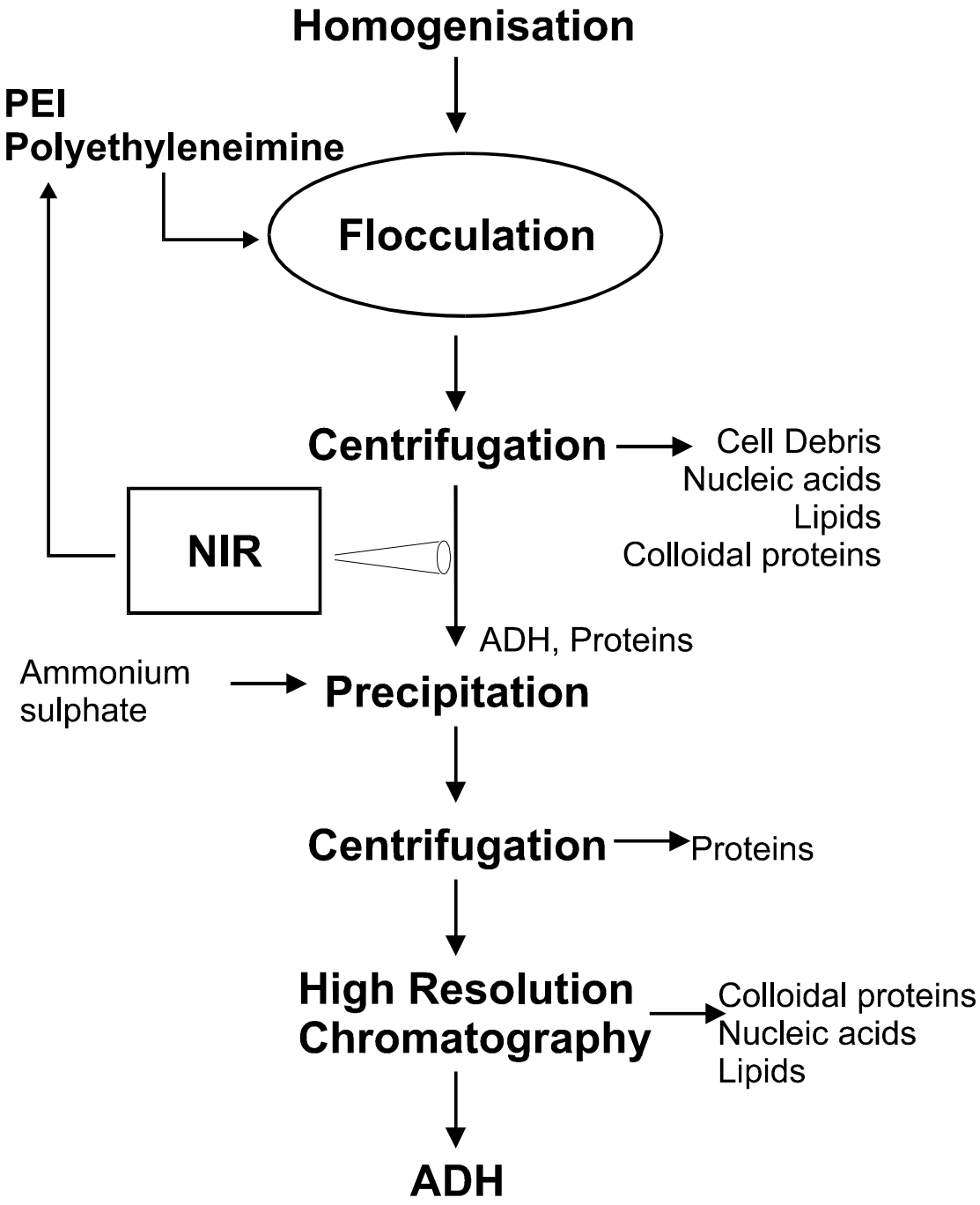
Homogenisation
PEI
Polyethyleneimine
Flocculation
Centrifugation
Cell Debris
Nucleic acids
Lipids
Colloidal proteins
NIR
ADH, Proteins
Ammonium
sulphate
Precipitation
Centrifugation
Proteins
High Resolution
Chromatography
Colloidal proteins
Nucleic acids
Lipids
ADH

Figure 1

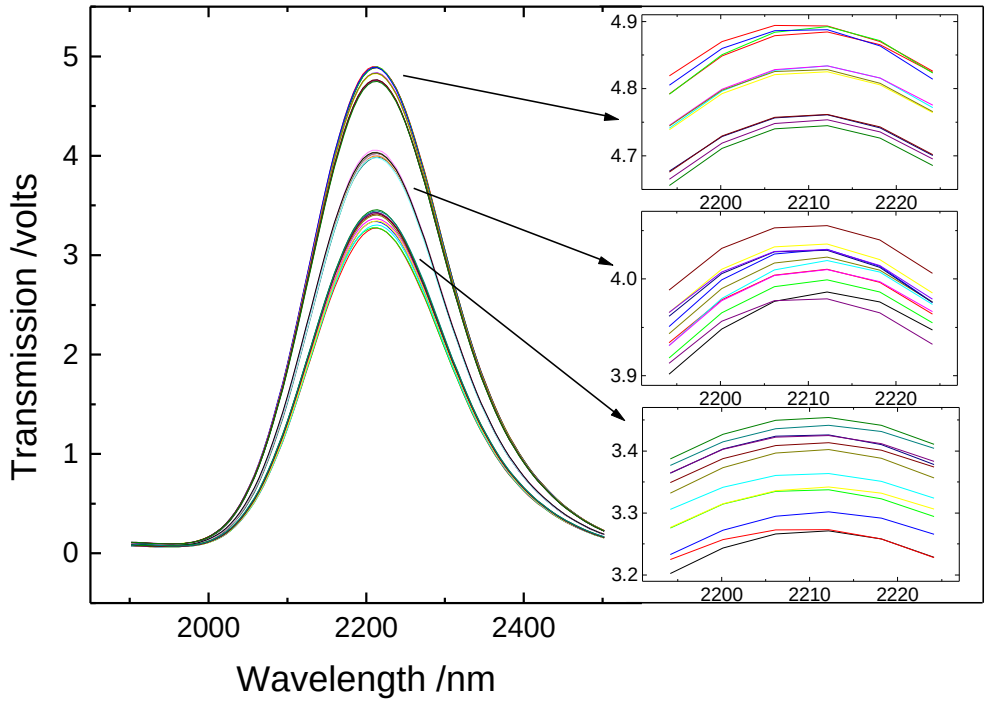
Transmission /volts
Wavelength /nm
0
1
2
3
4
5
2000
2200
2400
4.9
4.8
4.7
4.0
3.9
3.4
3.3
3.2
2200
2210
2220

Figure 2.

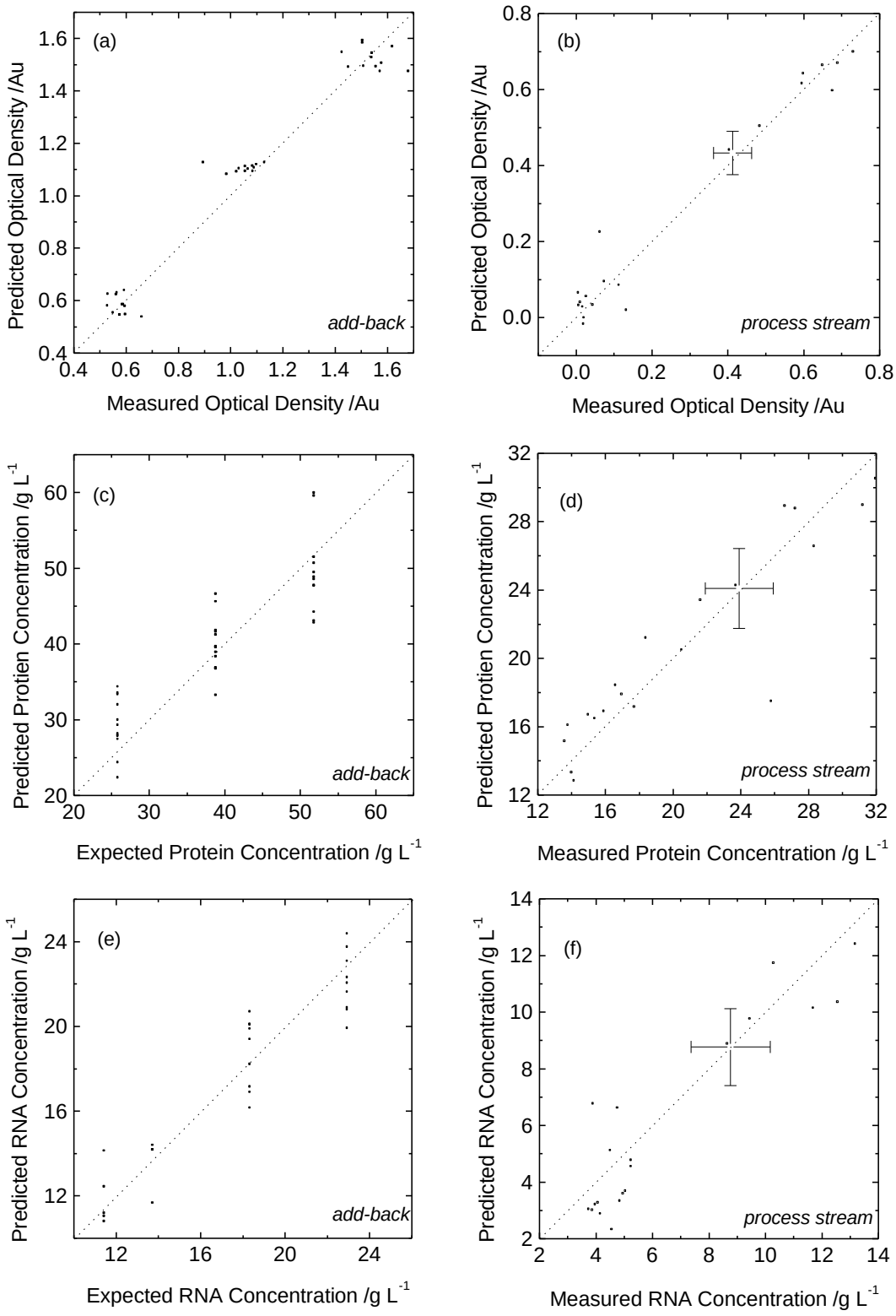
(a)
Predicted Optical Density /Au
Measured Optical Density /Au
add-back
(b)
Predicted Optical Density /Au
Measured Optical Density /Au
process stream
(c)
Predicted Protien Concentration /g L-1
Expected Protein Concentration /g L-1
add-back
(d)
Predicted Protien Concentration /g L-1
Measured Protein Concentration /g L-1
process stream
(e)
Predicted RNA Concentration /g L-1
Expected RNA Concentration /g L-1
add-back
(f)
Predicted RNA Concentration /g L-1
Measured RNA Concentration /g L-1
process stream

Figure 3.

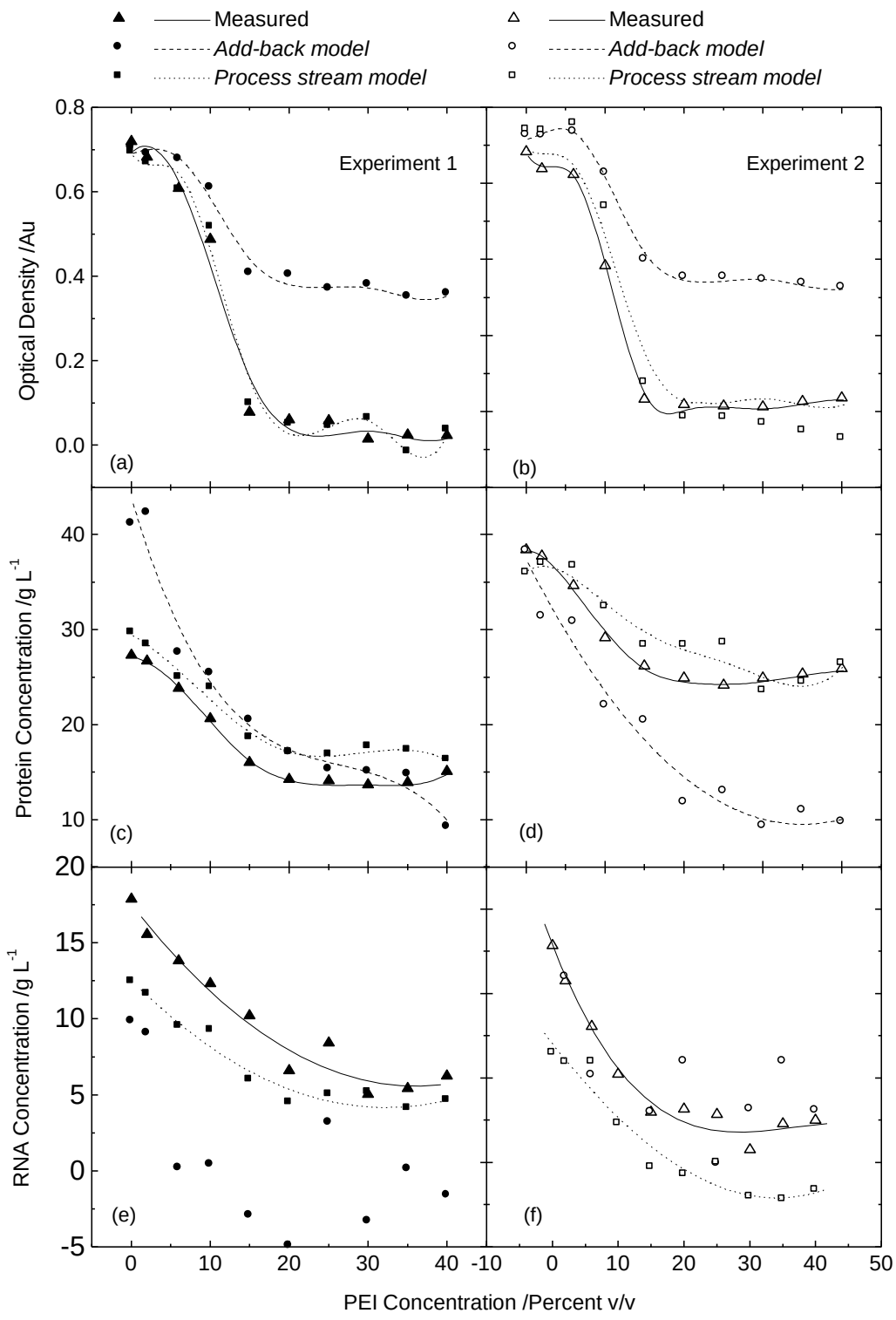
Measured
Add-back model
Process stream model
Measured
Add-back model
Process stream model
Experiment 1
Experiment 2
Optical Density /Au
Protein Concentration /g L-1
RNA Concentration /g L-1
PEI Concentration /Percent v/v
(a)
(b)
(c)
(d)
(e)
(f)

Figure 4.