The Enhanced Definition and Control of Downstream Processing Operations

A thesis submitted to the University of London
for the degree of DOCTOR OF PHILOSOPHY

By

George B. Habib M.Eng (Hons.)

The Advanced Centre for Biochemical Engineering
Department of Biochemical Engineering
University College London
Torrington Place
London, WC1E 7JE

November, 1998

ProQuest Number: 10797768

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10797768

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Monitoring product and contaminants is critically important at all stages of bioprocess operation, development and control. The availability of rapid measurements on product and key contaminants will yield a higher resolution of data points and will allow for more intelligent operation of a process and thereby enhance the definition and characterisation of a bioprocess. The need to control a bioseparation process is due to the variable nature of upstream conditions, process additives and sub-optimal performance of processing equipment which may lead to different requirements for the operating conditions either within batches or on batch to batch basis.

Potential operations for downstream processing of intracellular proteins are the selective flocculation, packed bed and expanded bed chromatographic operations. These processes involve the removal of a large number of contaminants in a single dynamic step and hence are difficult unit operations to characterise and operate in an efficient and reproducible manner.

In order to achieve rapid characterisation and control of these processes some form of rapid monitoring was required. A sampling and monitoring system for analysis of an enzyme produced intracellularly in *S.cerevisiae*, alcohol dehydrogenase (ADH), cell debris, protein and RNA contaminants has been constructed, with a measurement cycle time of 135 s. Both an extended Kalman filter and the Levenberg-Marquardt non-linear least squares model parameter identification technique have been implemented for rapid process characterisation. Estimation of model parameters from at-line data enabled process performance predictions to be represented in an optimum graphical manner and the subsequent determination of ideal operating conditions in a feedback model based control configuration. The application of such a control strategy for the batch flocculation process yielded on average 92% accuracy in achieving optimum operating conditions. A structured and intelligent use of the at-line data would improve process characterisation in terms of speed and stability. It was demonstrated that rapid monitoring of the packed and expanded bed chromatographic operations yielded improved characterisation in terms of higher resolution data points, enabled real time process analysis and control of the load cycle. For the control of the expanded bed operation a predictive technique was applied to compensate for the large dead volume associated with this unit operation. The feedback control resulted in approximately 80% accurate breakthrough setpoint regulation.

Contents

Material and methods 69

Chapter 1

Introduction

1.1 General overview

Section 2:

- Impact of upstream variability on downstream processes.
- Vast number of downstream unit operations makes their choice and sequence difficult.
- The pressure of speed to market.
- Unit interactions should be considered when setting up control configurations.

Section3:

- Benefits of rapid monitoring in terms of accelerating the development cycle- a valuable tool for rapid screening of downstream process protocols, confidence and control.

Section 4:

- Rapid monitoring techniques giving the user an indication of the complexity of product and contaminant measurements.
- Most rapid monitoring systems are of at-line nature and hence yield discrete measurements with some delay.
- Downstream process dynamics make the demand on at-line monitoring systems higher in terms of speed of measurement.

Section 5:

- The basic steps in setting up a control configuration.
- The constraints of bioprocess systems.

Section 6:

- How the control theory (section 5) and monitoring methods (section 4) can be implemented in a possible control configuration for example of downstream processing operations.

Section 7:

- Aim of thesis. Brief statement of what each chapter will discuss.

1.2 Overview of production and recovery of microbial proteins

1.2.1 Introduction

The first part of this section will give an overview of production and recovery of microbial proteins and the impact of upstream operations on downstream processing will be discussed. Several downstream processes are available and the correct sequence of these operations is important to determine in a time-efficient manner due to the increasing pressure of bringing products to market in the quickest time (van Brunt 1985). The objective and performance of downstream operations will be briefly discussed in terms of product yield, purity as well as productivity and operational costs. The operational objectives of downstream processing are important to define both for single unit operations and integrated process sequences as these objectives can be implemented as the setpoints for process control.

1.2.2 Fermentation and its impact on downstream process

The production of biological materials is often split between *upstream* and *downstream* operations. The upstream side includes fermentation and cell culture, whilst the downstream side involves recovery and purification of this product. Interaction between the upstream and downstream operations occurs superficially at the point of transfer. The final product of the upstream operation is the raw material for downstream processes. However the product is not the only material to be delivered downstream, so are all the impurities.

Many differences exist between upstream and downstream operations. For example in terms of cost, upstream operations, tend to be capital intensive, with automated vessels, few operations and relatively lower cost materials. Downstream operations, tend to be labour intensive, many operations, with little automation and higher cost materials. Due to these differences, the running of upstream and downstream operations tend to focus on separate factors, which sometimes clash.

Upstream conditions affect downstream operations, in several ways: *The volume of the materials delivered* to the downstream processes is an important factor to consider. Reduction of the volume or mass of material produced upstream will reduce downstream costs by reducing the volume of vessels needed for processing and the amount of material required. *The amount of impurities* relative to the product affects the efficiency of the purification. Although longer growth times may lead to greater production of for example recombinant proteins, impurities are also created, which may

have a negative impact on the overall product cost. *The nature of impurities*, can be controlled at the upstream stage. By utilising the appropriate media component(s), more suitable impurities will be created, hence facilitating the their removal and reducing purification costs. Variations in the media composition may result in variations in the nature of impurities. Hence if the fermentation includes ingredients with high variability then this must be considered in the purification stages. *The phase present* after the fermentation also influences downstream operations. At present the majority of enzymes produced commercially are obtained extracellularly with few intracellular enzymes having been isolated in large quantities. The main reason for this trend is the great complexity of the recovery process for intracellular materials. In extracellular protein recovery there is no cell disruption step, and there will be fewer other proteins present in the broth. Disadvantages include large liquid volumes and contaminants from the spent liquor. For these reasons combined with the potential offered by the large number of intracellular constituents, intracellular recovery processes have become increasingly commercially important.

All in all the impact of upstream conditions on downstream processing is significant and should be considered during process selection and design. The fluctuations inherent in fermentations will invariably have impact on the performance of the early recovery operations. Due to the fluctuations in feed material to the downstream processes appropriate design in terms of flexibility and robustness is vital. If a downstream process were operated in a constant manner given variable feed material, variable product quality would be the outcome. For more reproducible product quality some form of process monitoring and control is needed.

1.2.3 Downstream processes

Downstream processing in its narrowest definition is the purification of proteins from conditioned media or fermentation broth (Ogez *et al.*1989). The selection of the most appropriate methods for initial stages of downstream processing depends on the production host, location and physical form of the protein in the cell (Naveh 1990). Just as upstream processing is important in increasing productivity, downstream processing is important in reducing production costs. Often the cost of separation and purification of the product is the dominating cost of the whole production process. Wesseling (1994) stated that downstream processing involves at least 50% if not more of the total cost of the total production cost. One way to reduce separation costs and increase product throughput is process automation (Pfund 1987, Ranshoff *et al.* 1990).

Bonnerjea *et al.* (1986) analysed data from 100 papers on protein purification, published during 1984, ten major methods were found. Although this paper identified

no strict sequence of unit operations a distinct trend was obvious. Homogenisation is generally followed by precipitation, then ion-exchange, chromatography, affinity separation and finally gel filtration. The above sequence is a logical one, as precipitation can deal with large quantities of material and is less affected by interfering non-protein materials than the other operations. As the materials used for affinity methods are expensive, it makes sense to use less costly ion-exchange media first to reduce protein loads and remove remaining fouling substances.

Wheelwright (1989) emphasised that even though quite a few downstream sequences are in operation, there is no definite and predictive method or algorithm that one may follow to design a bioseparation protocol for a specific protein or biological product. Currently hands on experience and simplistic guidelines provide the best approach. Characteristic features of the biological product, such as its size, charge, biological affinity and hydrophobic – solubility, may be utilised to help separating it. There are a large number of bioseparation processes available to separate a wide variety of biological products and Scawen and Hammond (1989) emphasised that these processes must be carefully screened which involves a high cost both in terms of man power and time. Great pressure is put on the biochemical engineer to select the correct sequence of downstream processes for a particular product and correct specification for each operation as quick as possible to reduce the time a product gets to market.

Although downstream stages are usually described in terms of separate unit operations they are very interactive. Nucleic acid and cell debris produced on cell rupture can impair fractional precipitation or subsequent chromatography stages. For these reasons an integrated outlook of downstream processing is essential in terms of equipment and process design.

1.2.4 Downstream process performance

Although product yield is the main focus especially in the fermentation stage, it should be emphasised that the effectiveness of a downstream unit operation is a function not only of the ability to achieve a high yield factor, but also very much dependent on the level of purity the purification step can obtain at highest productivity.

The *yield* level of a unit operation can be defined as the ratio of product recovered to total product entering the purification step, i.e. yield is a reflection of product loss. The *purification* level, however is a function of contaminant removal. Both these terms are defined in equations 1.1 and 1.2.

$$Yield = \frac{p}{p_0} \qquad (1.1)$$

$$Purification\ factor = \frac{\left(\frac{p}{c}\right)}{\left(\frac{p_0}{c_0}\right)} \qquad (1.2)$$

Where p & c refer to the amount of product and contaminant recovered, whilst p_0 and c_0 refer to the amount product and contaminant entering the purification operation. It is usually the case for several unit operations that these two objectives are contradictory and inverse. A balance between yield and purity for the overall process and a given purification step is important to pre-determine in the process development stage, as this will influence the choice, sequence and hence control of downstream operations. The equilibrium between yield and purity must be estimated, taking into account the cost of purification, the value of the product and the purity required by market forces. In general the end product quality requirements are largely dependent on the end use of product. Huddleston *et al.* (1991) emphasised that bioseparation processes are defined by the nature of the product and its application. For example, therapeutic proteins must meet stringent standards whilst biological products intended for diagnostic usage do not have to be very pure.

The overall yield of a downstream process is a function of individual unit step yields and the number of operations, and can be expressed mathematically as:

$$Overall\ yield\ (y_t) = \prod_{0}^{i} y_i \qquad (1.3)$$

Where y_i refers to the step yield of unit operation i. Even relatively high step yields will result in a fairly poor overall yield, if a large number of operations are involved. In order to increase the total recovery yield (y_t), not only must the yield be increased for each unit operation but it is also important to reduce the number of purification steps. Chang and Chase (1994) and Smith (1997) used expanded bed (EXB) chromatographic techniques to achieve higher overall product yield levels. EXB techniques allow the direct loading of disrupted cells on a chromatographic column without the need for any prior clarification step, enabling the substitution of a number of traditionally applied unit operations.

Product loss throughout downstream stages are due to three main factors; namely physical damage, irreversible denaturation or physical loss. At every stage of the downstream process protein and enzymes are subjected to physical shear forces.

Much loss of proteins in pipes is due to interfacial denaturation, resulting from shear associated damage (Virkar 1981). Chemical denaturation can be caused by extremes of temperature, pH, ionic strength variations, or oxidation. These can be avoided through prevention of localised concentrations of reagents by good mixing. Physical losses are those where part of the product ends up in a side or waste stream of the process. On scale-up, such losses may increase especially for centrifugation stages, as de-watering becomes more difficult (Fish and Lilly 1984).

Some of the losses of yield experienced in recovering enzymes and proteins after the fermentation are time dependent. They occur due to the attack by proteases and by chemical oxidation and denaturation. Continuous processing can reduce time dependent product losses. However such processes require high equipment reliability, a rigorous process engineering approach and good instrumental control.

Increasing the yield of a unit operation may adversely affect other parts of the process. If the sole objective of the fermentation stage was to maximise product yield, this could directly impact downstream separations resulting in larger volumes of impurities. Hence fermentation optimisation is not equivalent to overall process optimisation. Process integration is also important within downstream operations. For example optimising the recovery of proteins in the homogenisation stage would result in a high level of fragmented cell debris, which will make further downstream operations, such as solid liquid separation more difficult to operate effectively. During the later step an increase in recovery can be achieved at the cost of lower flow rates, hence a lower product output rate. In the flocculation and precipitation stage an increase in yield, is usually followed by a reduction of purification. Zhou *et al.* (1997) illustrated the importance of process interaction during the study of the fermentation-disruption-clarification steps.

Running an unit operation effectively, is not coincident with solely maximising the step yield, but the attempt to maximise the removal of contaminants for a given yield level. A study of the interactions is also important, as this will allow one to determine the ideal inlet composition for operations, hence setting the process engineer a control target for the previous unit's outlet conditions.

1.2.5 Conclusion

An overview was given of upstream and downstream operations as well as the level of impact the former has on the later, in the form of a variable feedstock. Several downstream unit operations are available to the development engineer. Which process step and sequence to adopt is a function of several interacting factors and

currently the most appropriate method of selection is through experience and screening. Considerable pressure is put on the design engineer to identify the correct downstream processing protocol and operating conditions in a time-efficient manner. The variable nature of feed material to downstream processes necessitates the operating conditions to be modified either within batches or on batch to batch basis. Hence some form of monitoring and control is needed.

1.3 Potential benefits of rapid monitoring

1.3.1 Introduction

Monitoring product and contaminant concentrations is critically important at all stages of bioprocess operation, development and control. Biochemical processes are complex systems containing a multitude of different biological components in which their activity and state are extremely sensitive to changes in the physiochemical environment making their rapid monitoring an extremely difficult task. To be able to describe the status of the bioprocess in detail at any stage of its progression would give a more accurate record of the process facilitating and improving process operation, development and control. It should be noted though that process monitoring in biotechnology is extremely primitive (Paliwal *et al.* 1993) constraining the advances especially in bioprocess control. This is particular the case for downstream processing (Mattiason and Haakanson 1991). This section will attempt to discuss the benefits of rapid measurements and conclude on the current state of bioprocess control especially for downstream operations.

1.3.2 Process confidence

Process monitoring is an essential component of biotechnological production and is mandatory to meet the conditions set by the majority of regulatory authorities (Paliwal *et al.* 1993). In view of the fact that process monitoring is required it is paradoxical that quality control measurements are often carried out many hours after processing has completed. The problem is that many measurements have to be carried out afterwards due to the lack of available real time monitoring systems. Traditionally most downstream processing is carried out blind in the process time scale in terms of product and key contaminants giving little confidence in the correct operation, especially as feed conditions are inherently variable. Rapid monitoring would enhance process confidence by allowing the operator to follow the composition of the various process streams in real time and take the required steps in case something unexpected should happen. Rapid monitoring combined with powerful computer systems, can provide real-time data logging and documentation. The real time data

can be presented graphically to facilitate an immediate and comprehensive overview of process conditions and performance (Sonnleitner *et al.* 1991).

1.3.3 Rapid process development

Bioprocess engineers are faced with increasing pressure to extract as much valuable information as possible from each set of experiments performed (Christensen and Marder 1996). Rapid and automated measurements of product and key contaminants can facilitate, improve and accelerate the time costly development cycle both during bench and pilot phase operations.

During bench scale process development rapid monitoring techniques could substitute the laborious time consuming off-line assays eliminating the variability usually associated with manual measurements. Rapid monitoring will also allow more ambitious factorial experiments to be set-up improving experimental design and hence process knowledge. A higher resolution of data points can be demanded improving process characterisation. Further more the rapid data can be implemented for real time process analysis through the use of a supervisory computer. By performing the process analysis in real time the operator can focus on setting up the subsequent operation based on up to date analysis hence accelerating the process development cycle. Rapid monitoring is becoming a necessity in bioprocess development as demand increases for time-efficient ways of extracting information from each experiment (Olsson *et al.* 1998). Takahashi (Takahashi and Taniguchi 1989) demonstrated the at-line monitoring of both product and contaminant levels during ion-exchange and affinity chromatographic operations which resulted in considerable time savings. Rapid monitoring can also accelerate the development of a process in the pilot phase, allowing the engineer to move from one operating window to another in an informed manner (Habib *et al.* 1997).

1.3.4 Process control

The purpose of control is to manipulate the control variables to: (a) maintain the desired outputs at a constant desired value by suppressing the influence of disturbances, (b) stabilise unstable or potentially unstable processes and (c) optimise the performance, such as the yield, productivity or profit. The above objectives are to be achieved under various constraints such as safety, environmental regulations, limited resources and operational constraints. One important operational constraint in bio-systems is the ability to monitor processes in real-time, with sufficient reliability.

Product quality can vary both during biosyntheses and during downstream processing and that quality is a function of a complex set of variables. These variables can

change rapidly as organisms grow and deplete nutrients in a fermenter, as the feedstock to the purification systems varies, or as purification columns in the processing system age. The ability to monitor the processing system rapidly would make it possible to enhance the product quality through the control of system variables (Paliwal *et al.* 1993).

Fermentation monitoring and control has been a field of intense research for many years and a detailed discussion is therefore outside the scope of this section. Several reviews in this area have been published (Royce 1993, Rhelm *et al.* 1991, Thornhill and Royce 1991), where both conventional sensors and more recent biosensors (Scheper *et al.* 1996, Schugerl *et al.* 1996) have been used. Various control strategies have been applied (Vallino and Stephanopolous 1987, Lim and Lee 1991) and promising results have been reported.

The advances in monitoring and control of downstream processes however are less prominent. In Mattiason and Haakanson's (1991) extensive review on *measurement and control of downstream processing* he stated:

"One question to be raised is whether it is the very nature of downstream processing that makes it unsuitable for on-line monitoring and control. This does not seem to be the case. It is merely that there has been no tradition of process control in downstream processing, and hence there are no sensors or other control strategies available".

The lack of interest in rapid monitoring and control in downstream processing is highlighted by the scarce attention given to this area in the revised edition of *biotechnology, "measuring, modelling and control"* (Rehm and Reed 1991). In this comprehensive treatise out of 660 pages only ½ page is given on control of separation processes. Despite this lack of interest the aspects discussed on upstream control and the benefits this brings to the bioprocessing can be projected onto downstream operations.

The availability of measurements of product and contaminant concentrations in real-time would enable the simple open-loop feeding regiments that are currently in use to be replaced by more reliable closed-loop algorithms (Konstantonov *et al.* 1994). Control in downstream processing will increase the reliability concerning reproducibility between repeated batches (Mattiason and Haakanson 1991) and human errors and individual variations can be markedly reduced (Sonnleitner *et al.* 1991). Further more automation permits exploitation of 24 hours a day and 7 days a week (Sonnleitner 1997). However sufficient information about the process state is a pre-requisite for process control. The widespread use of computers in the biochemical industry has provided a powerful tool with which sophisticated control methods can be applied to

improve production efficiency. It is particularly in the areas of optimisation and control, in which the capabilities of the computer can truly be realised. However the success of computer based applications is highly dependent on the level and robustness of on-line monitoring and reliable mathematical models.

Despite the existence of advanced control theories, which are applicable to downstream processing operations in the bio-process industry, the pursuance of such strategies to date has been limited. This is in part attributable to the lack of reliable monitoring techniques particularly for the initial stages of recovery where a complex mixture of product, cell and cell wall debris is present, as well as the limited speed/frequency of monitoring. The need for rapid monitoring in bio-processes has been highlighted by several authors (Ransohoff *et al.* 1990, Paliwal *et al.* 1993, Mattiasson and Haakanson 1991). In Locher's detailed review (Locher *et al.* 1992) of on-line measurements techniques in biotechnology it is clearly stressed that on/at-line monitoring is substantially underdeveloped in biochemical engineering in comparison with other industries. An important difference between upstream and downstream processes is the relative dynamics of the operations. The fast dynamic characteristics of downstream processes puts even more demand on rapid monitoring techniques in terms of speed of measurements.

Today product and the process are well established by the time they become part of the processing chain and most downstream unit operations are monitored and controlled using traditional indirect measurements such as temperature, pH, conductivity, UV absorbance, etc. However, indirect monitoring methods can only pick up general changes in product and contaminant levels making efficient process control virtually impossible.

Advances in at-line measurement methods (see next section) have enabled a few examples of direct control of product and contaminants during downstream processing. Holwill *et al.* (1997) demonstrated the control of a fractional precipitation process using rapid measurements on product levels, enabling disturbances to be detected and the process to be correctly regulated to insure good product recovery in real time. Bracewell *et al.* (1997) and Sonsitza *et al.* (1998) demonstrated the rapid monitoring and control of chromatographic operations applying at-line biosensors, indicating that if traditional UV absorbance readings were applied poor operation of the process would have been the outcome.

1.3.5 Conclusion

Rapid (automated) monitoring has great importance for the biochemical engineering discipline in terms of process confidence, improving and accelerating process development and the potential of process control increasing the reproducibility and hence quality of product. The bottleneck in biotechnological process control is the lack of rapid measurement techniques of controlled biological variables. Several advances in rapid monitoring methods have taken place these will be described in the next section.

1.4 Rapid monitoring methods

1.4.1 Introduction

Process monitoring is an essential component in the biochemical process environment. As discussed in section 1.3 rapid monitoring is vital for process confidence, rapid process development, process control and automation. This section will discuss the various types of monitoring techniques available for the biochemical industry.

In its most ambitious form bioprocess monitoring means effortless access to continuous real-time information about all variables relevant to a given process (Olsson *et al.* 1998). Today very few on-line analysis methods are available and thus at present one has to settle for second best, at-line analysis systems or in some cases rapid off-line measurements as the only means of monitoring a given analyte during a bioprocess. These three monitoring procedures are illustrated in figure 1.1 and defined below.

1) ***Off-line monitoring***. A sample is manually taken, prepared and presented to a detection instrument.
2) ***At-line monitoring***. A sample is automatically taken, prepared and transported to a detector implementing a flow system.
3) ***On-line monitoring***. Measurement can be made directly without the need for a sample to be taken from the bioprocess.

Most rapid monitoring systems are of at-line nature and typically consist of three operating steps: a sampling device, a flow system and a detector. The whole system is usually automated and delivers a discrete signal with a time delay.

This section will briefly discuss the different types of measurement techniques which may be applied for rapid monitoring of bioprocesses be it of on-, at- or off-line nature. Automated sampling systems, three different flow systems applied and different available detection systems will be looked into in terms of their level of specificity and monitoring speed. This section will conclude with an overview of the available systems applicable to downstream monitoring and their potential use for process control.

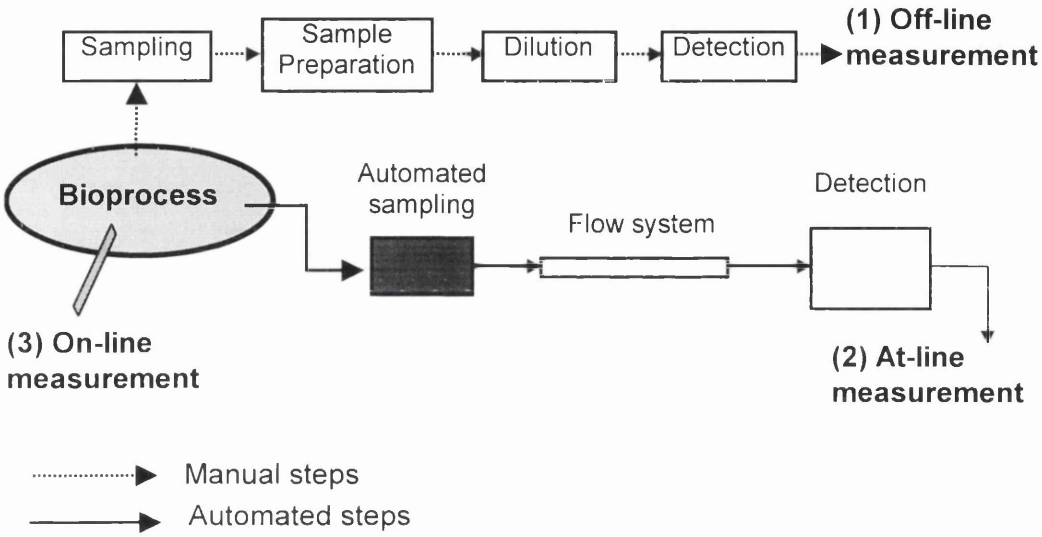


Figure 1.1. Schematic of on-, at-, and off-line measurement techniques.

1.4.2 Automatic sampling

As described above the first part of any at-line bioprocess system is an automated sampling system. Generally two types of samples can be withdrawn from a bioprocess, a solid free sample or a solid-liquid sample, which is usually dependent on what is required to be monitored and whether the detection instrument is sensitive to fouling. The main priority of any sampling system is that it can reproducibly withdraw a representable sample of the bioprocess.

Most of the at-line sampling systems (for fermentation use) use a membrane for removal of interfering material from the sample, dialysis and filtration being most common techniques. In filtration a pressure is applied to force the analyte molecules as well as the solvent through the membrane pores. In dialysis, a concentration gradient is the only driving force for transport through the membrane and molecules of the appropriate size diffuse from the sample. A detailed discussion of membrane based sampling systems was given by Merbel *et al.* (1996). The drawbacks of filtration systems were sample losses and membrane fouling, whilst for dialysis techniques the

constraints were the relatively slow speed of separation and the need for calibration. Membrane sampling systems have been implemented in combination with several at-line monitoring systems (Scheper *et al.* 1996, Merbel *et al.* 1996).

An alternative sampling system based on microcentrifugation has also been developed. Rapid solid-liquid separation using this technique and subsequent HPLC measurements of glucose and acetate from an *Escherichia coli* fermentation was demonstrated by Turner *et al.* (1994). Chard *et al.* (1994) applied a similar solids removal system prior to an enzyme catalysed FIA technique for the rapid monitoring of a fractional precipitation operation. Although automated centrifugation is not widely used due to the relatively high cost, it has the benefits of handling samples fast and reproducibly without fouling.

1.4.3 Flow systems

Flow systems are applied to automatically dilute, prepare and transport a given sample to a detection system for analysis. Three flow systems will be discussed namely, the *flow injection analysis* (FIA), *stopped flow analysis* (SFA) and *sequential injection analysis* (SIA) methods illustrated schematically in figure 1.2.

FIA is based on the injection of a well defined volume of liquid sample into a continuous moving carrier reagent stream (Ruzicka *et al.* 1988). After injection the sample is typically involved in a reaction whilst being propelled towards a detector. The success of FIA rests on (1) the reproducible injection of the sample (2) timing of the sample movement and (3) controllable dispersion of the sample in the carrier stream. The advantages of FIA systems are the low response time, low sample volume and high sampling frequency whilst the drawbacks are the lack of robustness and large quantities of reagent required. Christensen and Marder (1996) lists a number of publications which have demonstrated the successful use of FIA.

During SFA a well defined volume of sample and reagent is mixed and transported to a detector by a carrier stream. The flow is temporarily stopped and the reaction initiated by the sample-reagent mixture is followed usually in terms of a kinetics assay. The advantages of a SFA compared to the FIA is increased sensitivity of the measurement, the ability to evaluate the kinetics of a reaction and the reduction in use of reagent. The drawbacks are usually a slower frequency of measurements.

The development of SIA was to overcome the problems seen with the two above flow systems, such as single analyte analysis, large consumption of reagent and the need for complicated flow manifolds with several flow channels rendering the system more liable to wear of mechanically moving parts. By introducing a single piston pump and multi-position valve connected to a number of reagents reservoirs, sample reservoirs, enzyme reactors, detectors the above problems can be solved (Ruzicka *et al.* 1990). SIA systems are often based on stopped flow techniques, where sample dispersion residence times and signal detection are fully separated phenomena (Christensen and Marder. 1996).

FIA, SFA and SIA have been extensively implemented as the flow system of many at-line analysis methods due to the large number of unit operations and sample preparation steps which can be included as part of the analysis manifold. The use of such techniques for rapid monitoring of fermentation and downstream process have been demonstrated extensively in literature and they differ in the manner in which samples are detected.

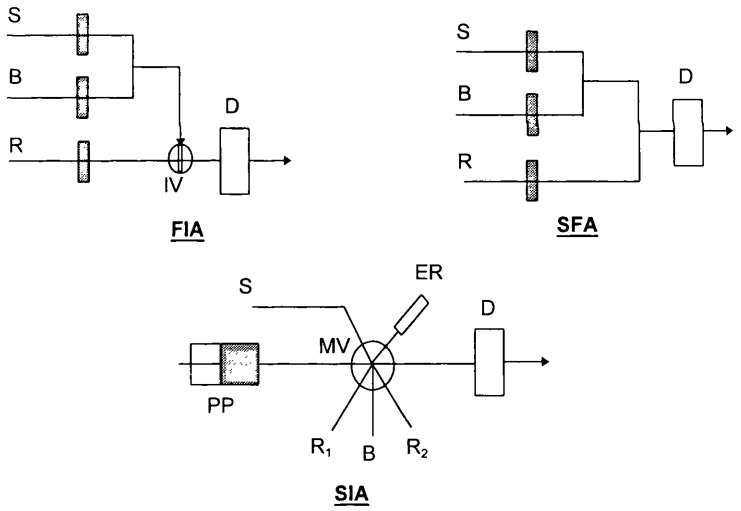


Figure 1.2. Schematic of the FIA, SFA and SIA flow systems. The letters S, B, R, PP, D, ER, MV and IV refer to sample, buffer, reagent, piston pump, detector, enzyme reactor, multi-position valve and injection valve.

1.4.4 Chromatographic techniques

Chromatographic techniques such as LC and GC are well established analytical separation techniques. The sample is separated into its constituents in the flow system and the individual components are subsequently quantified by a non-specific detector (spectrophotometer). Chromatographic techniques allow for multi-component

analysis and therefore have a high information density. In addition the system is very stable and robust since they are based on physical principals. The main drawbacks of such monitoring systems are the relative high cost and especially the low analysis frequency, typically less than 5 h^{-1} for LC (Olsson *et al.* 1998).

Capillary electrophoresis (CE) has been established as a promising chromatographic technique due to its high resolution capacity. Electrophoresis is a process in which charged species are separated based upon different migration rates in an electrical field and can be applied for the determination of peptides, proteins, nucleic acids and many types of biopolymers. CE analysis times are around 10 minutes per cycle. The relatively slow response of chromatographic systems has limited their use to fermentation monitoring.

1.4.5 Enzymatic measurements

Enzyme activity is generally monitored by measuring the enzyme catalysed reaction initiated by the addition of substrate under suitable conditions. By tradition these assays have been performed as off-line analysis. By adopting one of the flow systems discussed above automated enzymatic assays may be performed in a faster and more reproducible manner by eliminating the manual dilution and reagent mixing steps. Though it is rare that the analyte or the product into which it is converted can be detected with sufficient selectivity and sensitivity the analyte can often be quantified by detecting co-substrate or co-products. Two classes of enzyme reactions where co-substrates or co-product are easily detectable are oxidases and dehydogenases. Various detection principals can be used for quantification of these reactions such as spectrophotometry, caliometry, fluorometry and amperometry. Mattiason and Haakanson (1991) and Olsson *et al.* (1998) have compiled examples of FIA techniques which have been used in combination with enzymatic catalytic reactions and their respective detectors. For downstream application Holwill *et al.* (1997) adopted a SFA for the rapid measurement of alcohol dehydrogenase (ADH) activity during a fractional precipitation process. Takashashi *et al.* (1989) demonstrated the use of a FIA for the real time measurement of alkaline phosphate during ion-exchange and affinity column chromatography. As enzymatic off-line assays are traditionally performed post process completion the speed of measurements is not vital, hence enzymatic reactions are usually allowed to run to completion or during kinetic measurements a large number of points are taken prior to rate estimation. During at-line measurements short cuts are taken to reduce assay times, such as only looking at the initial kinetics

information and often not allowing the reaction to go to completion. This is possible since sample handling is extremely reproducible.

1.4.6 Immunoanalysis

Analysis of individual proteins may also be performed through the binding of antibodies to a particular region on the individual protein molecule. If an antibody is available for the protein of interest analysis based upon precipitation, detection of radioactively labelled antibodies (radio-immunoassay (RIA)) or the amount of enzyme activity which can be linked to a particular protein by an antibody (ELISA) may be used to quantify the amount of the individual protein present. Recently some of these immunoassays have been adopted for rapid monitoring of a biochemical process. Especially the flow injection of ELISA technique appears to be a quick, reproducible method that is easy to automate (Nilsson *et al.* 1992a).

Immunoassay procedure have two parts; the actual immunoreaction and the subsequent detection of these immunocomplexes formed. A wide variety of methods have been established; one technique applied by Mattiason and Haakanson (1991) is described below. First the antibody was immobilised to a solid support and packed in a small column. The sample to be analysed for its native antigens was mixed with a known amount of enzyme-labelled antigen. The antigen mixture was passed through the antibody column through one of the flow systems described above. Time for contact was in the range of 10-20 s, i.e. far too short for equilibrium in the binding reaction to be established. However, since flow and sample addition was extremely reproducible, it was possible to operate the system under conditions far from equilibrium (Mattiasson & Borrebaeck 1978). After a short washing step to remove any non-specifically bound enzyme-labelled antigen, the level of the binding reaction was evaluated by feeding a stream of substrate for the marker enzyme. The higher the concentration of the native antigen in the sample, the smaller the amount of labelled antigen that was bound by the affinity column and thus a lower product pulse was registered. Finally regeneration of the antibody column was performed prior to the next assay cycle.

Several different transducers have been used in conjunction with binding reactions, the choice of which varies from case to case. Mattiasson and Haakanson (1991) have compiled a list of examples of analysis performed and detectors used in flow injection binding assays. The times involved vary and are dependent on the time taken for

sample mixing and transportation and reaction kinetics. Nilsson *et al.* (1992b) demonstrated the use of a flow injection immunoassay for the rapid monitoring of chromatographic separation with an assay time of 7 minutes.

1.4.7 Biosensors

A biosensor can be defined as analytical device that combines a biological component with a transducer (Mattiason and Haakanson 1991). The biological component can be an enzyme, an organelle, a micro-organism, a tissue or an antibody and confers specificity on the system. The transducer (optical, thermistor or some type of electrode) is usually placed in close proximity to the biological component and converts the biological signal into an electrical signal, which is proportional to the analyte concentration. Biosensors would be ideal for on-line monitoring as they provide a close to real time continuous and very specific measurement of a given analyte. However the drawbacks of biosensors is that they suffer from poor stability, problems associated with sterility (fermentation use) and need frequent re-calibration. Additionally the biosensor may not work optimally at the conditions prevailing in the bioprocess. Consequently biosensors are currently used for analysis of cell free samples, obtained from either manual or automatic sampling systems. Biosensors have been used as detectors in FIA systems for fermentation (Scheper *et al.* 1996) and affinity chromatographic process (Bracewell *et al.* 1997) where in the later example monitoring frequencies were approximately 120 h^{-1} and therefore applicable for downstream process control.

1.4.8 Traditional measurements software sensors

Traditional measurements such as temperature, pH, conductivity and turbidity readings are traditionally taken during upstream and most downstream unit operations. These are usually regulated to fixed setpoint in order to keep environmental conditions suitable for product stability and to aid good separation. However they give little information on the biological state of the process by themselves.

Software sensors use data available on-line to give estimates of some indicators of the biological state (such as biomass, product and substrate concentration) using process models (Glassey *et al.* 1997). An integral component of software sensors is the model, which relates the available measurements to the estimated process variables. Model building can be performed from first principles through a series of mass balances or through statistical techniques such as neural networks and multivariate analysis

methods. Montague *et al.* (1992) demonstrated the use of a software sensor using an artificial neural network based model to predict biomass concentration on-line by applying information on the feed rate, carbon dioxide evolution rate and fermentation age. All the publications found apply software sensors to fermentation systems, due the vast number of already present on-line measurements on chemical and physical properties. The application of such techniques for downstream process monitoring was not found. However traditional measurements such as, pH, temperature, conductivity, shear rate, process time and turbidity readings may contain information to predict the biological state of for example a precipitation or flocculation process.

1.4.9 Spectral analysis

Spectroscopic techniques such as ultra violet (UV), visible (VIS) and infrared (IR) as well as fluorescence, electron spin resonance and nuclear magnetic resonance have broad application in biochemistry. They can yield a large amount of information about biological processes. However for rapid analysis of bioprocesses at present only a few of these techniques have been applied.

Single wavelength measurements are traditionally applied in the UV and VIS region for an indication of protein (UV280), nucleic acid (UV260) and turbidity readings (absorbance at 600-670 nm) which give an estimate of biomass levels. Their level of specificity is low due to the interference of the multi-component mixture, require clear samples or the need for dilution prior to spectroscopic analysis.

The benefits of spectral scans have recently been demonstrated both using UV-VIS and NIR spectroscopy. By implementing the information contained in spectral scans through multivariate analysis techniques more specific biological information can be gained. Rapid monitoring of cell debris, protein and RNA in yeast homogenate was demonstrated during a flocculation process using analysis of spectral scans using both NIR (Yeung *et al.* 1998) and UV-VIS (Nuoi *et al.* 1998). With recent advances in instrumentation (diode-array spectrophotometers, fibre optic probes) and multivariate analysis spectroscopic instruments are finding their way into various applications for monitoring and control purposes, especially NIR techniques (Yu and Phillips 1992). NIR absorption is generally much weaker in intensity than those found in the UV-VIS and mid infrared absorption bands. This is seen as useful as sample analysis can be performed without the need for dilution or requirement of short optical pathlengths (Plugge and Vlies 1992). UV-VIS traditional spectroscopic techniques are limited in

their on-line monitoring application for fermentation and early recovery processes due to the turbid nature of the process streams resulting in scattering (Brimmer and Hall 1993). Dilution is required, however this can result in a more controlled and stable sample for analysis. UV-VIS spectroscopic instruments are also cheaper and diode-array technology for this spectral range more advanced. Several publications have demonstrated the benefits of NIR spectroscopy mostly for on-line monitoring of product, substrate, nutrient and biomass using diffusive reflectance modes employing fibre-optic probes of during *E. Coli* (Macaloney *et al.* 1996), yeast (Cavinato *et al.* 1990) fermentation. If such techniques could be implemented during such turbid mixtures, their application in downstream monitoring should be possible.

1.4.10 Conclusion

Today only very few on-line analysis methods are available that in a non-invasive manner deliver a continuous real time signal of the bioprocess performance. Hence at present at-line monitoring techniques are the closest to real-time data and in many cases rapid labour intensive off-line assays are employed to follow a bioprocess. Several at-line measurement techniques have been discussed in terms of the analyte specificity, monitoring times and manner of automation. In most cases some form of automated sample preparation, flow system and detection instrument was employed. The detection instrument creates the specificity of the at-line system.

When applying at-line measurements for real time process characterisation and control the time delays associated with the discrete measurements have to be accounted for in terms of the process dynamics. The relatively fast dynamics of downstream processes combined with the discrete nature of at-line measurements results in a low resolution of data points for process characterisation and control. By combing the at-line information with earlier experiences of the process relatively good predictions of the process development can be made. As the time delays of at-line data are reduced, a better connection between measurement and response can be obtained. It should be stressed that traditional parameters should also be registered. This means that a data basis for proper action will be available, based on several factors monitored. Although the whole area of monitoring and control in downstream processing is currently under resourced new techniques based on spectroscopic and biosensor technology are advancing enabling on-line measurements and hence the possibility of robust closed loop process control.

1.5.Control Theory

1.5.1 Introduction

The purpose of this section is to review some of the techniques available to the process engineer for setting up control configurations and to identify the main limitations, especially when applied to downstream bioprocesses. Process control terminology and basic control considerations will be discussed.

The "steady-state" operation of processes is one of the fundamental concepts of chemical and biochemical engineering. It has proved highly successful in analysing the behaviour of existing processes or plants, and in designing new ones. It must be appreciated, however, that the steady-state is a highly idealised situation. Although a process is designed to run at steady state, it is likely that the conditions in it will be changing continuously with time. Such behaviour may be caused by, *short-term disturbances,* such as changes in the flowrate, pH, temperature or composition of the feed stream, due to changes upstream. *Long-term disturbances*, such as, the loss of column efficiency during chromatography or flocculant effectiveness with time due to ageing, and *deliberate control actions*, due to changes in set-point values by the process control engineer. The steady-state operating conditions for a given process, are normally determined at the design stage so as to optimise a certain objective criterion, such as maximum productivity or specific purity. The objective of control is to manipulate the available control variables of a given process or unit operation in order to:

- ◊ Maintain the desired outputs at a constant value or to follow a certain profile, by suppressing the influence of external disturbances.
- ◊ Stabilise processes in terms of robustness.
- ◊ Optimise yield, purity productivity or operating costs.

The above objectives have to be achieved under various constraints, such as safety, environmental regulations and operational constraints. It must also be recognised that some processes are inherently unsteady, i.e. batch operations. Downstream processes are seen as a sequence of batch processes although specific unit operations can be run in a continuous manner. The complexity of batch processes is their inherent unsteady nature and require a start-up and shut-down control strategy.

1.5.2 Variable classification

The control engineer usually classifies the process variables under two main groups; *input variables*, which denote the effect of the surroundings on the process, and *output*

variables, which convey the effect of the process on the surroundings, see figure 1.3. Both the input and output variables can be further classified into sub-categories. Input variables can either be *manipulated variables, design variables* or *disturbances.* The former occurs if the input variables can be adjusted freely by the human operator or by a control mechanism, whilst the latter is the case if their values are not the result of an adjustment by an operator or a control system. According to their direct measurability, the disturbances can be further classified into *measured* and *unmeasured* disturbances.

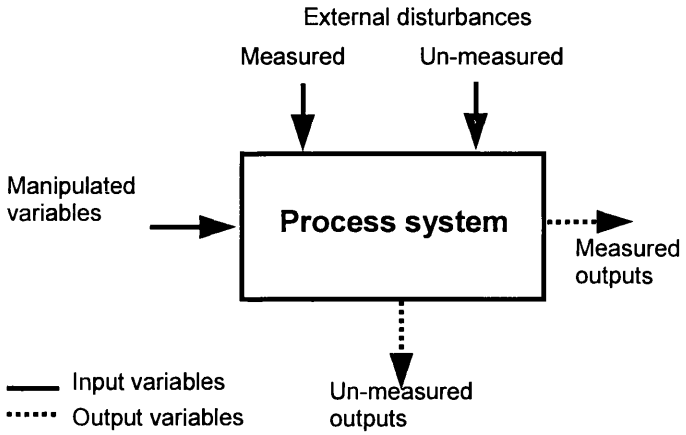


Figure 1.3. Types of process variables.

Output variables are classified into either measured or unmeasured output variables, depending if their values are known directly. The measured output variables may then be used as *controlled variables*, if the process engineer chooses to utilise them for control purposes.

1.5.3 Definition of control objectives

Precise definition of control objectives is vital for the design of an effective control strategy. Initially control objectives are defined qualitatively; subsequently they are quantified, usually in terms of the output variables. Control objectives and setpoint values, are usually determined in the process design stages. For example, a predetermined flocculant dose may be specified as the control setpoint. However in many cases the feed material for early bioprocess recovery systems is variable due to fluctuations in upstream conditions. A more appropriate control setpoint would be to specify the fraction of contaminant removal for example.

Due to the interactive nature of downstream bioprocesses the objectives for a unit operation should take into account the ideal feed requirements of the subsequent

operation. For example; the objectives of a homogeniser, as a unit operation is solely to disrupt cells and achieve maximum product release. However the overall objectives of the homogeniser-centrifugation process is not only to achieve good product release, but also to attain satisfactory removal of cell debris. A trade off exists as for more product release a higher number of passes is required resulting in more fragmented cell debris, which is difficult to separate during centrifugation. The control engineer will have to specify certain minimum and maximum thresholds for product yield and contaminant removal and use these as setpoints, i.e. a setpoint of at least 95% product release.

1.5.4 Selection of measurements

Whatever our control objectives are, we need to monitor the performance of the process, and hence one needs to determine which variables should be measured in order to monitor the process performance. Variables in bioprocesses are very difficult if not impossible to monitor on-line. Three types of measurements can be defined: *Primary measurements*, is when the control objective can be directly measured, *secondary measurements* when the control objective is estimated through other measurement values and *tertiary measurements*, is when the disturbances are monitored. The tertiary measurements are relevant for feedforward control systems, which will be described later.

1.5.5 Selection of manipulated variables

Usually in a process there are a number of available input variables which can be adjusted freely in order to compensate for any disturbances. Which ones we select to use as manipulated variables is vital, as the choice will affect the quality of the control actions. Some manipulated variables, have a direct fast and strong effect on the controlled output variables, others do not. Some manipulated variables are easy to manipulate, i.e. flow rates others are more difficult (ionic strength, pH). Most downstream operations tend to be single input - multiple output (SIMO) systems, hence the choice of manipulated variables is limited to the one. The systems tend to be multiple input-multiple output (MIMO) systems, which, for practical reasons, have been reduced to SIMO systems. For example, the homogenisation step, has two main input variables which can be manipulated directly, namely, the number of passes and the operating pressure. However the operating pressure is always maintained constant during a run, hence the MIMO system gets reduced to a SIMO system.

1.5.6 Determination of control configuration

After determining the control objectives, the possible measurements, and the available manipulated variables, one needs to define the best control configuration for the given process, i.e. which measurements (controlled variables) should be interconnected with which manipulated variables. Two general types of control configurations can be defined, feedback and feedforward control loops, illustrated in figure 1.4.

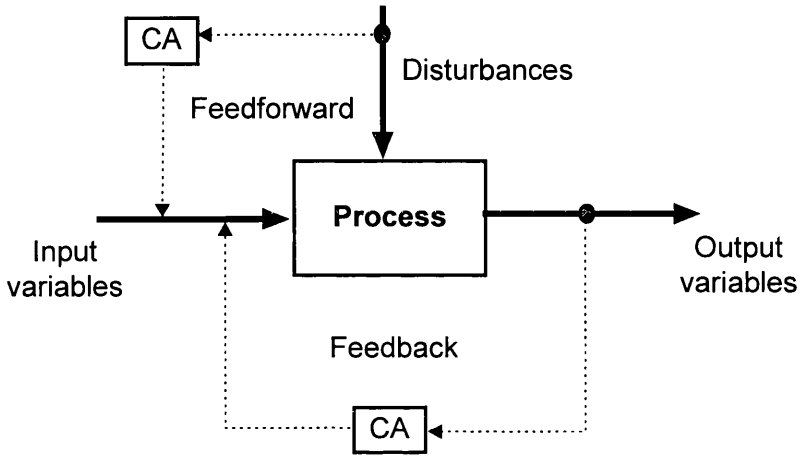


Figure 1.4 Overview of feedback and feedforward control loops. CA refers to the control action performed on the input variable(s).

Feedback (FB) control configuration uses the direct measurement of the controlled variables to adjust the values of manipulative variables. The objective is to keep the controlled variable at desired level. For example the release of protein during homogenisation can be used to regulate the number of passes through the homogeniser.

Feedforward (FF) control configuration uses direct measurement of the disturbances to adjust the values of the manipulated variables. Feedforward control can be a very effective means of improving the dynamic response of a control system when simple feedback is not satisfactory. However FF control assumes that it is possible to measure some of the major disturbances that drive the process away from the desired set point. If one knows how a process will respond to a disturbance, one can, in principle, generate a control signal that will compensate for the predicted response to a disturbance before it occurs, thereby holding the process at the desired state. In order to keep the value of this output at the desired level, we need to change the value of the manipulated variable by such an amount as to eliminate the impact that the disturbance would have on the output. By how much though? The control engineer must know the following relationships :

$$Process\ output = f(manipulated\ variables)$$

$$Process\ output = g(disturbances)$$

This example demonstrates how important mathematical modelling is for design of a feedforward control system. Without good and accurate mathematical modelling we cannot design efficient feedforward control systems. By adopting a combination of FF and FB control an effective control strategy can be created. The FF controller, would increase the speed of recovery, whilst the FB controller would act as a fine tuner. A control configuration for a centrifugal operation can for example adopt a FF control system to monitor any changes (disturbances) in feed conditions and regulate the discharge rate or flowrate accordingly.

1.5.7 Design of the controller

In every control configuration the controller is the active element that receives the information from the measurements and takes appropriate control actions to adjust the manipulated variables, in order to compensate for disturbances. The controller's function is therefore to compare the measured output value to a given setpoint, and based on this deviation between desired and actual values perform some control action by applying a certain algorithm. The controllers are named after the particular algorithm they use on the error signal. The most common controller is called a *PID* controller and uses proportional (*P*), integral (*I*) and derivative (*D*) actions. The simplest type of controller is ON/OFF control. Here the control action is switched on when the controlled variable exceeds a specified setpoint and turned off when the controlled variable falls below another specified threshold.

1.5.8 Other control configurations

Selective control systems

These are control systems that involve one manipulated variable and several controlled outputs. Since with one manipulated variable we can control only one output, the selective control systems transfers control action from one controlled output to another according to need. A selective control system selects among several similar measurements the one with highest value or priority and feeds it to the controller. For example, in order to maintain a certain yield level, whilst maximising the removal of contaminants a selective control system can be adopted. The yield factor is the first priority whilst the contaminant removal is the secondary objective.

Sequence control

Since batch and fed-batch systems vary with time, the control system must deal with time- and event- based process conditions and transition phenomena. In a batch precipitation operation sequences of operations need to be followed. First a batch tank is filled with feed material and brought to the right conditions (temperature, pH, salt concentration), a fixed volume of precipitant is added, aggregation of precipitants is allowed to age at specific shear conditions and the termination of the precipitation step is followed by centrifugal separation. The starting of pumps and the opening of valves are all events in time, requiring different control actions. An important question to ask is what concentration of precipitant should be added? For a variable feed stream a different precipitant dose is required to be determined. A FF control configuration could be implemented requiring some form of process model. Alternatively real time optimisation of the process can be performed which determines the ideal precipitant dose setpoint for each batch.

1.5.9 Adaptive optimisation and high level control

The objective of adaptive optimisation and control is to determine optimal operating conditions for a bioprocess that may be unknown or may change with time and subsequently control the process to these conditions using the techniques described above. The optimisation involves determining the best operating variables, which leads to optimum performance as measured by a given performance index. For example, during a precipitation and flocculation process due to the variable nature of the feed material a fixed predetermined setpoint for pH, salt concentration and precipitant or flocculant dose is not possible. One needs to determine in real time optimal operating conditions. During batch mode this may be done by observing the output levels (product and contaminant level if possible) given the specific input conditions (dose level, pH, conductivity) during the gradual dosing of precipitant or flocculant. By applying such an input-output data with estimator process model parameters can be determined. These up to date estimations of model parameters for the given batch run can be used to fully characterise the process and fed to an optimisation procedure to determine the optimum conditions and apply these as control setpoints.

1.5.10 Conclusion

An overview of process control terminology and basic control considerations for downstream processes have been described. The determination of control objectives and subsequent setpoints are dependent on not only the specific objectives of a single unit operations but the behaviour of the subsequent downstream process. The correct

selection of the manipulative and controlled variable as well as the control configuration will determine the effectiveness of the control system to regulate a given process to its pre-specified setpoints.

In downstream processes most input variables can be classified as design variables. Material from upstream operations can be defined as possible disturbances due to the fluctuations observed in fermentation conditions. This leaves the operator with just a few manipulative variables to achieve the control objectives. In some situations the process setpoints may be unknown and change between batches. An adaptive optimisation procedure, adopting a process model, can determine optimum operating conditions in real time.

1.6 Early recovery operations and approaches to their control using direct measurements

1.6.1 Introduction

The purpose of this section is to analyse possible control configurations for a few downstream unit operations. The following sections will give a brief introduction to each purification step, followed by a more detailed description of some operations which will be studied later in this work. An analysis of the control problem using direct measurements on product and contaminants will be given based on existing publications. The unit operations will be discussed in roughly the order they occur in a typical downstream process.

1.6.2 Cell disruption - High pressure homogeniser

Background

Protein purification schemes diverge initially depending on whether the desired product is retained inside or secreted outside the cell. For intracellular products some form of cell disruption is required. Cell disruption methods can be categorised into mechanical and non-mechanical methods, which again can be divided into several types of cell release techniques. Non-mechanical methods include; desiccation, high pH, osmotic shock, chaotropic agents, detergents, freezing, antibiotics or cell wall hydolytic enzymes. In each case the cell membrane may be either totally disrupted or made partially permeable to allow the product to escape. Non-mechanical disruption methods, are usually inflexible, often time consuming batch operations, expensive and can cause further purification problems downstream.

Mechanical disruption methods include; milling, high pressure homogenisation and ultrasonic techniques. High pressure homogenisers are the most widely used methods for large-scale cell disruption processes (Scawen *et al.* 1980) and will therefore be used in the following control example.

High-pressure homogenisers

The most widely studied high-pressure homogeniser is the Manton-Gaulin APV type, which uses a spring-loaded valve originally developed for milk homogenisation. Previous studies on cell disruption using yeast have shown that the disruption to be a first-order process (Hetherington *et al.* 1971) of the input-output form:

$$R_f = \frac{R}{R_m} = 1 - \exp(-KNP^a) \qquad (1.4)$$

Where R_f is the fraction of cells disrupted, R_m the maximum amount protein available for release and R the amount of released protein, K the dimensionless rate constant, P the operating pressure, N the number of passes and a is the exponent to the operating pressure. The rate constant is a function of the suspension temperature, whilst a is dependent on the cell type, and was experimentally found to be constant and approximately 2.9 for yeast and 2.1 for *E.coli.* Such a relationship between pressure and cell release clearly demonstrates the desirability of higher pressure units. The operation conditions of a homogeniser also has a strong impact on the particle size distribution of subsequent cell debris and viscosity of the homogenate, which was shown to follow the response of protein release (Clarkson *et al.* 1993, Siddiqi *et al.* 1991, Mosqueira 1981). As the number of passes through the homogeniser increases the viscosity of the homogenate will rise and the average cell debris size decrease.

Control considerations

The objective of the homogenisation step is to maximise the release of intracellular product whilst maintaining a low carry over of micronised debris that could cause problems for further downstream processes. The smaller debris size range produced by an extra pass may preclude any advantage in extra product release since the debris would be more difficult to recover.

Given equation 1.4 one can classify the various homogenisation input variables. Possible manipulative variables are the pressure (P) and number of passes (N), however the former is traditionally kept constants and specified during the design stages. Process disturbances may arise from variation in upstream conditions resulting in fluctuations in cell wall strength or can arise form changes in homogeniser conditions such as equipment failure.

Whatever the control objectives are, one needs to monitor the performance of the process. Rapid monitoring during homogenisation will allow the release and possible denaturation of the product level to be followed in real time. Such rapid data may be utilised for process confidence, fault detection or control. Currently homogenisation processes are well defined in terms of the required number of passes, operating pressure, temperature, etc. Hence any form of control will focus on keeping input variables to pre-specified operating conditions, i.e. maintaining pressure at 500 bar, temperature at 4°C and number of passes to 5. Such control may be termed traditional chemical engineering control and assumes that no or little variation will be seen in feed conditions.

Holwill *et al.* (1999) demonstrated how at-line monitoring of product during the homogenisation of Baker's yeast could be implemented for fault detection due to equipment failure. Product measurements were performed at-line through a stopped flow analyser giving an indication of the process performance. The release model described in equation 1.4 was implemented to predict the total product release fraction. This was compared to past runs and if a large deviation was observed fault detection was triggered.

Control of product release has not been shown in literature. This may be attributed partly to the difficulty of acquiring rapid process information and the design of the homogenisation unit. By altering for example the manner in which the homogenisation process is operated more scope for direct control is possible. During batch homogenisation the number of passes is discrete making control difficult as the manipulative variable (number of passes) will be an integer and usually well defined i.e. 5 passes. However during continuous homogenisation more flexibility is introduced as in this case the processing time can be used as the manipulative variable.

1.6.3 Flocculation and precipitation

Flocculation and precipitation apply the use of insolubilisation or aggregation by adjustment of the chemical environment (salt concentration, pH, organic solvent, polymer, etc.) to carry out purification. Both operations are usually followed by a solid liquid separation step such as centrifugation. As the selective flocculation process will be studied later in this work, a more detailed theoretical discussion will be given on this unit operation. However possible monitoring and control configurations will be given on published examples on the precipitation process.

Flocculation background

Flocculation provides a highly promising solution to the problem of processing cell homogenate prior to high resolution purification. The flocculation of cell debris using polyelectrolytes (e.g. Polyethyleneimine (PEI)) can be used to enhance the efficiency of separation processes such as centrifugation by causing the debris to aggregate into larger flocs. Simultaneously it is possible to remove contaminants by polymer induced precipitation of, for example, nucleic acids and colloidal proteins. The removal of such contaminants is an important precursor to the use of high resolution separation operations such as packed bed chromatography.

The mechanism of flocculation by polyelectrolytes can involve two processes; s*urface charge neutralisation* and *bridging*. Certain long-chain polymers may absorb in such a way that different segments of the same polymer chain are absorbed on different particles, thus *bridging* the particles despite the electrical repulsion. With polyelectrolytes of opposite charge to the particles, the latter may be partly or completely neutralised by the polymer. This eliminates the electrical repulsion, destabilises the particles and allows them to aggregate. Both of these effects can be important and sometimes both operate simultaneously.

This work will be investigating the flocculation with polyethyleneimine a cationic polyelectrolyte with a wide molecular weight distribution. The key parameter is the charge difference between the polymer and the material in suspension, leading to electrostatic patch flocculation. PEI has been extensively characterised with respect to its physico-chemical properties (Lindquist and Stratton 1975, Horn 1980). It is a low-cost reagent, and used in low concentrations, hence no recycle is required. It is extensively used in the packaging industries and its use for biological products is reviewed elsewhere (Milburn *et al.* 1990).

The flocculation process is highly sensitive to operating conditions such as pH, salt concentration, feed stream composition (Jendrisak 1987) as well as flocculant make-up (age, brand) (Habib *et al.* 1997). This is made worse by the tendency for reversal of flocculation leading to stabilisation of for example cells or debris in suspension. Hence the problem is more difficult than for development of say a precipitation process.

For charge neutralisation driven polymers, such as PEI an increase in ionic strength through an increase in salt concentration, would reduce the effectiveness of

flocculation due the interaction between the polymer and ions and the screening of the particles (Bulmer *et al.* 1994). Bulmer *et al.* further stated that optimum salt concentrations is component specific. The level of pH has a crucial affect on the effectiveness of PEI to flocculate contaminants and product. At low pH values the polymer's charge density is high and therefore a lower concentration of flocculant is required to achieve optimum flocculation. However a more accurate manner of dosing is needed to prevent the flocculation and hence removal of the target product. No practical models are available in literature that describe the effects of different input variables on the flocculation mechanism and hence flocculation process output variables.

The flocculation and precipitation process have to be viewed as an integration of the flocculation/precipitation mechanism and centrifugation operation, with the measure of what is adequately flocculated/precipitated being determined by the centrifuge performance, i.e. both as a consequence of feed zone breakage and of clarification. Hence poor performance of a flocculation or precipitation process could either be resolved by change in flocculant/precipitant concentration, pH, salt concentration, etc. or by altering centrifugation operating conditions.

Precipitation background

A range of protein solubility curves were given by Cohn (1925) for the precipitation process as a function of salt concentration, described in equation 1.5:

$$log\frac{E}{E_0} = kI + \beta \qquad (1.5)$$

where E and E_0 referred to the enzyme activity or protein concentration remaining soluble after precipitation and in the feed. Model parameter k was defined as the salting-out constant which depends on protein and salt properties, β is a function of pH and temperature and I refers to the salt concentration.

Control considerations

Examples of rapid monitoring of a flocculation process was demonstrated by Gregory and Nelson (1986) and Huang and Chen (1996), implementing a novel instrument called a photometric dispersion analyser (PDA). The PDA instrument was able to detect a floc index based on fluctuations in turbidity readings. However these examples were studied for the flocculation of kaolin, i.e. a non-biological material.

Due to the lack of published material on rapid monitoring and control of the flocculation process this section will concentrate on the precipitation process. The objective of a precipitation operation is to obtain a setpoint product yield by maintaining a specific position on the process solubility curve of a target enzyme. A possible process disturbance is the feed material composition which may result in a shift in the solubility curve and thus a change in the amount of product remaining in solution and hence removed/collected in the subsequent solid/liquid separation step. Possible manipulative variables are the pH, temperature and salt concentration levels, however given the limits on product stability the two former variables are usually specified in the design stages.

An example of how direct monitoring of product and subsequent control of a continuous ammonium sulphate precipitation process for the purification of a specific enzyme in *S.cerevisiae* was given by Holwill *et al.* (1997). The adopted control configuration was of feedback nature, where the overall all feed saturation level was manipulated to maintain product levels to a specific position on the solubility curve to ensure a setpoint product yield. The rapid product measurements were carried out by a stopped flow analyser. A microcentrifuge (Richardson *et al.* 1996) was implemented to remove biological particles such as protein precipitate or cell debris whilst maintaining the soluble components. The feedback control configuration implemented a model to establish the position of newly acquired product points on the solubility curve and if these were unsatisfactory the control algorithm would drive the process to meet the solubility setpoint. Holwill *et al.* (1997) demonstrated that the control configuration was capable of maintaining a required product yield by manipulating the saturation level given deliberate disturbances in pH and feed concentration.

It should be noted that in the above work only product was monitored and controlled, hence the issue of purity was not tackled. Although the specified yield would result in appropriate process performance, an optimisation setup would allow the determination of an optimum setpoint.

1.6.4 Chromatography

The term chromatography refers to a group of separation techniques, which are characterised by distribution of the molecules to be separated between two phases, one stationary and the other mobile. In the below text a detailed description on

hydrophobic interaction chromatography will be given as this process is studied in detail later in this work.

This work will focus on hydrophobic interaction chromatography (HIC) which will be operated in both packed bed and expanded bed mode and hence a detailed description of this process is given below. Monitoring and control configurations will be given on published examples on other types of chromatography operations.

Hydrophobic interaction chromatography background

HIC is a technique by which components are separated due to their different strengths of hydrophobic interaction with a generally uncharged absorbent matrix possessing hydrophobic groups. HIC can separate components having similar size and charge on the basis of their hydrophobicity. The technique is particularly useful in the purification of proteins, as most proteins posses hydrophobic regions on their surface. A large number of theories have been proposed for the explanation of hydrophobic interaction chromatography. However most of the theories are essentially based upon the interaction of hydrophobic solutes and water (Tanford 1997, Creighton 1984).

The main factors to consider when selecting HIC media for the optimisation of product selectivity are: ligand type, type of salt, concentration of salt, pH and temperature. The next step is to select the operating conditions in terms of sample load and flowrate for optimum dynamic column performance in terms of productivity. In this work HIC is used as an initial product capture step where the major concern is to remove critical contaminants and reduce volume, hence selectivity during desorption (elution step) is not a prime issue. The entire bed volume is utilised for sample binding and the prime consideration when optimising for highest possible productivity is to find the highest possible sample load over the shortest possible sample application time with acceptable loss in yield. The dynamic binding capacity for the protein of interest should be determined for different sample loads and flowrates. If the flowrate is too high a decrease in dynamic binding capacity will be observed, however this may still be advantageous from a productivity point of view. A high sample load would result in better utilisation of the column capacity, however a larger loss of product would be observed in the breakthrough profile. A low sample volume although increasing the product yield, would result in poor utilisation of the column capacity and hence result in lower productivity. Hence the correct selection of flowrate and sample volume are critical for optimum process performance.

Hydrophobic interaction chromatography is traditionally operated in packed bed mode and can be divided into three operating cycles subsequent to the equilibration of the column. The first operating step is the *load cycle* where a clarified feed stream is loaded on to the column continuously. Components that have the least affinity for the stationary phase break through first. Loading is terminated when the capacity of the column for the product is saturated and breakthrough of product is observed in the eluate. The column is then washed, in the *wash cycle*, to remove unbound impurities. The last operating step is the *elution cycle* where the product is eluted off the column.

In section 1.3 it was emphasised that there is a need for a reduction in the number of unit operations during downstream processing to increase the overall all product yield. Expanded bed (EXB) chromatography allows for this by enabling disrupted cells to be loaded directly onto a chromatographic column without the need for any form of clarification step, thereby substituting several traditionally used unit operations. An extensive review on EXB systems and their application is given by Smith (1997).

When liquid is pumped upwards through a bed of adsorbent not constrained by an upper adapter, the bed can expand and the voidage between the adsorbent beads increase allowing particulate material to travel through the column without being trapped. The design of the adsorbent material for EXB systems is crucial, as the expansion of the bed is a function not only of the liquid composition and viscosity but also of the density and size distribution of the matrix.

In general EXB processes are operated in a similar fashion to packed bed operations. The below text will briefly highlight the main differences. Prior to loading, the settled bed is expanded through the application of equilibration buffer in an upwards direction. The load cycle is then initiated in the same direction. In general, the load sample will have quite different physical properties to the equilibration buffer and often having a higher viscosity density. The influence of feed stock properties on the EXB performance has been discussed by Frej *et al.* (1993). The Washing of EXB systems is usually performed in the upwards direction to remove both colloidal cell debris or cells and unbound contaminants. Once particulate material has been removed the direction of the wash may be reversed to pack the matrix and the elution cycle is initiated. Similar scouting experiments to those applied during the packed bed chromatography are required to specify optimum pH, salt type and concentration, ligand type, sample volume and flowrate for the optimisation of the EXB operations.

Control considerations

The objectives of a chromatographic operation is to separate a specific product from contaminants giving the highest yield at the required purity as quickly, cheaply and easily as possible. Many of the input variables discussed above are usually determined in the design stages leaving a few manipulative variables applicable for control. The control objectives can be split into three parts one for each chromatographic operating cycle. In the load cycle the objective is to load to a setpoint product breakthrough. In the wash cycle the aim is to remove un-bound contaminants from the column and in the elution cycle the objective is to collect the fraction of eluate with the highest product concentration.

Two publications were found on the rapid monitoring and the subsequent control of product directly during a chromatographic operation (Sosnitza *et al.* 1998, Bracewell *et al.* 1998).

Sosnitza *et al.* (1998) demonstrated the ability to rapidly monitor an ion exchange chromatography process for the purification of sugar molasses into high value components. A flow injection analysis biosensor system was adopted to measure product levels (servine and sucrose) every 2-5 minutes. The at-line monitoring system applied was slow compared to the dynamics of the process, where product fractions were eluted within 5-10 minutes and therefore could not be adopted for real time control during the elution cycle for correct fraction collection. Elution fractions were therefore collected in 1 minute intervals and analysed rapidly one after the other. Based on the obtained elution profile the subsequent chromatographic operations could be controlled. In this example the need for fraction collection control was driven by the disturbances found in the feed material resulting in fluctuations in elution profiles.

Bracewell *et al.* (1998) adopted an optical biosensor for immunorecognition of protein products during affinity chromatography to obtain a rapid description of the loading and subsequent breakthrough. The biosensor was operated within a flow injection analysis regime with a total assay time of 30 s. This provided sufficiently rapid data for process control. In combination with the at-line data a model to forward predict the outcome of the breakthrough profile was adopted to enable the correct termination of the load cycle.

1.6.5 Conclusion

It has been indicated that very few examples are available in literature on the direct control of downstream operations in terms of product and/or contaminants. This is partly due to the lack of available rapid monitoring techniques. Even when rapid data on product is available it is often of at-line nature hence associated with time delays, is discrete and noisy. Therefore the application of such at-line information will have to incorporate some form of data filtering. Most of the published work on bioseparation process control has been of open loop nature, i.e. an operator was employed to regulate the manipulative variables. The design of some bioprocesses makes the implementation of control configurations difficult. Bioseparation processes are usually batch operations with just a few manipulative variables giving the control engineer little scope for control. The added constraint of strict regulatory specifications on bioprocess operating conditions means that any manipulation of input conditions in order to compensate for possible disturbances will have to fall within boundaries specified by the regulatory authorities.

1.7 Aim of thesis

This chapter has attempted to give the reader an overview of the current state of rapid monitoring and control of biosepration processes. The benefits of rapid measurements have been put forward in terms of facilitating process development and making process control feasible. Several rapid monitoring techniques have been described however their application for downstream process characterisation and control have been limited due to the relatively fast dynamics of downstream operations. Hence the bottleneck in biotechnological process control is the lack of rapid measurement techniques of controlled biological variables.

The continuous pressure of speed to market of a new drug imposes considerable pressure on the development engineer to gain as much information as possible from each process development run, especially of costly pilot plant operations. Additionally the need for rapid screening of several downstream process protocols advocates the use of rapid monitoring techniques to accelerate the selection of the ideal bioseparation sequence. Real time process information can give the development engineer a powerful tool to facilitate the running of a process to gain optimum information around critical operating areas to enhance the knowledge of a unit operation. Rapid measurements are therefore fundamental for the improvement of the process development cycle in terms of speed and efficiency.

The nature of bioseparation processes makes them inherently variable due to fluctuations in upstream conditions, process additives and sub-optimal performances of processing equipment. Such fluctuations lead to different requirements for operating conditions either within batches or on a batch to batch basis. Rapid monitoring allows for a process to be followed in real time and operating conditions altered in order to compensate for any disturbances, i.e. process control is possible. Process control will reduce the variability of process performance yielding better and more reproducible product quality.

The aim of this thesis was therefore to investigate the use of rapid monitoring methods for rapid process characterisation and control both of product and key contaminant levels during the recovery of a target enzyme, alcohol dehydrogenase (ADH) from S. cerevisiae. The recovery operations examined were a selective flocculation (batch and continuous), a packed bed and expanded bed chromatographic operations.

The following briefly describes the objectives addressed in each of the following chapters. Each chapter has its own summary, introduction and conclusion during which the major points are highlighted.

Chapter 2 describes some theoretical considerations, which were applied during this work. Multivariate statistical techniques will be briefly described especially the partial least squares (PLS) method and its application for the calibration of spectral data for the prediction of contaminant levels in Baker's yeast homogenate. Three flocculation models will be presented and discussed in terms of their main features and ease of application for real time process characterisation. Lastly this chapter describes two model parameter identification techniques namely the Levenberg-Marquradt non-linear least squares and extended Kalman filter which will be implemented for the estimation of model parameters during a process run to obtain an up to date mathematical description of the performance of the flocculation process.

Chapter 3 presents the material and methods applied during this work. The description of the off-line and at-line monitoring techniques as well as the process and sampling methods will be given.

Chapter 4 demonstrates the applicability of the at-line monitoring techniques for the rapid identification of product (alcohol dehydrogenase) and key contaminant (RNA,

protein and cell debris) concentrations. The calibration of the UV-VIS spectral data through the use of a factorial experiment will be presented and the application of the PLS technique to enable calibration model building and subsequent use for contaminant level prediction. The overall analysis of the at-line monitoring setup consisting of an automated microcentrifuge, stopped flow analyser and detection of both a kinetics reaction and rapid spectral analysis is presented.

Chapter 5 describes the use of the at-line monitoring setup for the rapid characterisation of the batch flocculation process. The three flocculation models described in chapter 2 are implemented for process characterisation and the most effective model selected.

Chapter 6 demonstrates the possibility of process optimisation and control of the batch flocculation process through the implementation of the at-line information on product and contaminants, a process model and the use of a model parameter estimation technique. The control of 8 batch operations will be presented given their variable feed conditions. The advantages and disadvantages of two model parameter identification techniques as well as two different control configurations will be discussed.

Chapter 7 illustrates the rapid and efficient characterisation of the continuous flocculation process through an intelligent and informed manner of operation.

Chapter 8 demonstrates the use of the at-line data on product and contaminants for the rapid characterisation of the packed and expanded bed chromatographic operations. Real time process analysis and process control will be addressed.

Chapter 9 concludes the thesis by drawing out the major findings of the work and by presenting some suggestions for future work arising from the research.

Chapter 2

Theoretical considerations

2.1 Introduction

This section will discuss some theoretical aspects implemented in this work. The first part of this section describes multivariate statistical techniques such as principal component analysis (PCA), principal component regression (PCR) and partial least squares (PLS) for the prediction of contaminant levels from spectral data. The second part of this section describes three proposed flocculation models, which mathematically attempt to describe the characteristics of the flocculation operation in terms of the amount of component remaining in solution. Their complexity and relevance for at-line process characterisation will be discussed. The last part of this section describes two model parameter identification techniques, namely an extended Kalman filter (EKF) and the Levenberg-Marquardt non-linear least squares (NLLS) techniques.

2.2 Multivariate statistics

2.2.1 Introduction

Multivariate statistics is a collection of powerful mathematical tools that can be applied to (bio)chemical analysis when more than one measurement is acquired for each sample (Beebe and Kowalski 1987). An example will be given to facilitate the explanation of such techniques. This example will illustrate how protein, RNA and debris contaminant levels in yeast samples may be predicted from ultraviolet-visible (UV-VIS) spectral data.

Most analytical methods consist of two phases: *calibration* and *prediction*. The first step is to construct a data matrix ($\mathbf{X}$) from the instruments responses (UV-VIS spectral scans) at the selected wavelengths (220 – 500 nm) for a given set of calibration samples. A concentration matrix ($\mathbf{Y}$) of protein, RNA and debris values is then formed using independent chemical assays or through the use of standards. The goal of the calibration phase is to produce a model that relates the UV-VIS spectra to the values obtained by the off-line analysis or standards.

The next step of the calibration phase is the foundation of the entire analysis. The analyst must choose an appropriate mathematical method that will best produce **Y** given the **X** matrix. Different methods are available and three such techniques will be discussed: multiple linear regression (MLR), principal component regression (PCR) and partial least squares (PLS). When the calibration model has been determined predictions of unknown samples can be made from their spectral scans. Figure 2.1 illustrates matrices **X** and **Y** for this example, the calibration and subsequent prediction procedure.

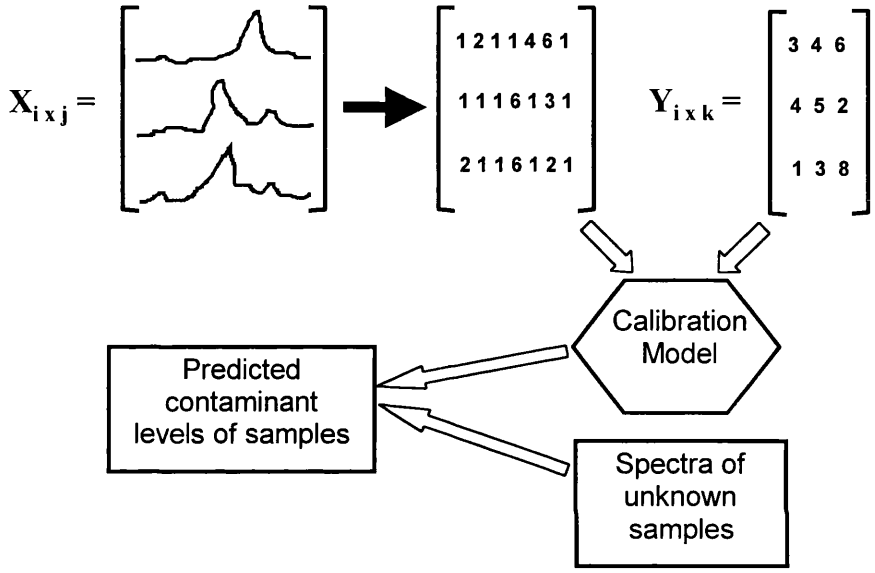


Figure 2.1 *Configuration of the spectra data matrix* ***X****, the measurement matrix* ***Y****, calibration and prediction procedures. The* ***X*** *matrix is a 3 x 7 matrix with three rows of UV-VIS spectra from the analysis of three samples and 7 columns corresponding to the chosen wavelengths. The* ***Y*** *matrix (3x3) consists of three samples (rows) and the three contaminants, protein, RNA and debris (columns).*

2.2.2 Calibration model building

A technique available for calibration model building is multiple linear regression (MLR), which assumes that the best approach to estimate **Y** and **X** matrices is to find a linear combination of the variables **X** that minimises the error in reproducing **Y**. A major drawback of using such an approach is that a model chosen solely according to the MLR criterion attempts to use all of the variance in the **X** matrix, including any irrelevant information or noise to model **Y**. Hence prediction errors might arise through the MLR model building.

Other modelling techniques include the PCR and PLS methods. To understand these statistical procedures the technique of principal component analysis (PCA) should be

explained. In real samples, there are usually many different variations that make up a spectrum: the contaminants in the sample mixture, inter-contaminant interactions, instrument variations, such as detector noise, changing environmental conditions that affect absorbance, and differences in sample handling. However the largest variations in the calibration spectra set should be the changes in the spectrum due to the different concentrations of the constituents of the mixtures. By being able to calculate a set of "variation spectra" that represent the changes in the absorbance at all the wavelengths in the spectra, than this data should be used for model building instead of the raw spectral data.

Principal component analysis (PCA) is a statistical technique for multivariate data compression and information extraction (Davis 1986, Wise and Gallagher 1996). Principal components are the eigenvectors of a variance co-variance matrix. They provide an insight into the structure of this matrix and are the basic tools of many factors analyses. PCA allows for the **X** matrix to be compressed into a set of *scores* (**T**) and *loadings* (**P**) matrices respectively, as illustrated in figure 2.2. Hence by using the scores and loading matrices a reconstruction of the **X** matrix is possible and the spectral data which is not explained by the principal components is described by the residual matrix ($\mathbf{E_x}$). The scores matrix (**T**) is a measure of how much a particular principal component is present in a particular calibration sample. Just as a spectrum is represented by a collection of absorbances at a series of wavelengths, it can also be a series of scores for a given set of factors. The loadings (**P**) express relationships between the principal components and each wavelength absorbance for each scan, informing us which wavelengths are dominant in influencing the model and how they are correlated with each other. The advantage of this decomposition on the **X** data is that the most important variations in the data are abstracted to the first few principal components, which are easier for interpretation of the spectral data.

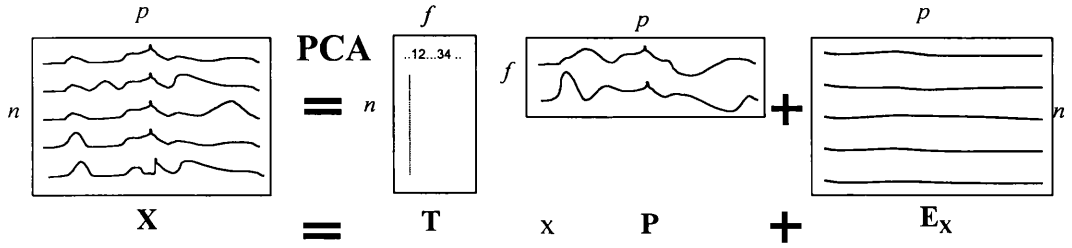


Figure 2.2. *Schematic of PCA decomposition of the data matrix **X**. The letters **T**, **P** and E_X refer to the scores, the loading and residual matrices, whilst f refers to the number of principal components, n is the number of spectra and p is the number of data points per spectra.*

The principal component regression (PCR) method combines the PCA spectral decomposition with an inverse least square regression method to create a quantitative model which may be used to predict the concentration matrix (**Y**). As mentioned above the scores matrix (**T**) is unique to each calibration spectrum. It is therefore possible to regress the calibration data set (**Y**) against the scores (**T**) matrix. The calibration model finds a correlation between **X** spectral and **Y** analytical data using the score **T** matrix and establishing a separate loading **Q** matrix to describe the correlation. This is illustrated in equation 2.2. Equation 2.1 below is taken from figure 2.2. Equation 2.3 describes the calibration model, which can be derived by re-arranging equation 2.1 and substituting into 2.2. (It should be noted that **P** is an orthonormal matrix, i.e. $\mathbf{PP}^T = \mathbf{I}$).

$$\mathbf{X} = \mathbf{TP} + \mathbf{E}_x \qquad (2.1)$$

$$\mathbf{Y} = \mathbf{QT} + \mathbf{E}_y \qquad (2.2)$$

$$\mathbf{Y} = \mathbf{QXP}^T + \mathbf{E}_y \qquad (2.3)$$

Where all of the above matrices have been defined earlier apart from $\mathbf{E}_y$ which is the residual to **Y**. It is important to note that PCR is a two-step process; the PCA loading and scores are calculated and then the scores are regressed against the concentration matrix (**Y**). The PCR method can build accurate calibration models, provided that the selected variables (scores/loadings) are physically related to the properties (concentration matrix **Y**) they are regressed against. However the PCA scores and loadings are calculated independently of any knowledge of the sample concentrations. They merely represent the largest common variations among all the spectra in the training set. There is no guarantee that the PCA scores and loading matrices directly correspond to the contaminants of interest, hence its predictive ability will suffer.

Partial least squares (PLS) is a quantitative spectral decomposition technique that is closely related to PCR. However, in PLS the decomposition is performed in a slightly different fashion. Instead of first decomposing the spectral matrix (**X**) into a set of scores and loadings, and regressing them against the concentration matrix (**Y**) as a separate step, PLS actually uses the concentration information during the decomposition process. This results in spectra containing higher contaminant concentrations to be weighted more heavily than those with low concentrations. Hence the loading and scores matrices calculated using PLS are quite different from those of PCR. The main idea of PLS is to get as much concentration information as possible into the first few loading vectors.

PLS is taking advantage of the correlation relationship that already exists between the spectral data and the contaminant concentrations. Since the spectral data (**X**) can be decomposed into its most common variations, so can the concentration data (**Y**). This generates two scores and loading matrices, one for the spectral data and the other for the concentration data. The two sets of scores are related to each other through some form of regression and a calibration model is constructed. During the PLS model building the decomposition on both the spectral and concentration data is performed simultaneously.

The approach taken by PLS is very similar to that of PCA, except the factors (loadings and scores) are chosen to describe the variables in **Y** and **X.** This is accompanied by using the columns of the **Y** matrix to estimate the factors for **X**. At the same time the column of **X** are used to estimate the factors of **Y**. The resulting models are:

$$\mathbf{X} = \mathbf{T}_{PLS}\mathbf{P}_{PLS} + \mathbf{E}_{XPLS} \qquad (2.4)$$

$$\mathbf{Y} = \mathbf{U}_{PLS}\mathbf{Q}_{PLS} + \mathbf{E}_{YPLS} \qquad (2.5)$$

Where the elements of $\mathbf{T}_{PLS}$ and $\mathbf{U}_{PLS}$ are the scores of **X** and **Y**, respectively and the elements $\mathbf{P}_{PLS}$ and $\mathbf{Q}_{PLS}$ are the loadings determined through PCA driven by the PLS technique. The matrices $\mathbf{E}_{XPLS}$ and $\mathbf{E}_{YPLS}$ are the errors associated with the modelling **X** and **Y** with the PLS model. The $\mathbf{T}_{PLS}$ factors are not optimal for estimating the columns of **X** as was the case with PCA, but are rotated so as to simultaneously describe the **Y** matrix. As $\mathbf{U}_{PLS}$ and $\mathbf{T}_{PLS}$ are linked and have a linear relationship combining equations 2.4 and 2.5 can create the calibration model. The optimum number of principal components applied during the PLS calibration model can be derived from the PRESS (prediction residual error sun of squares) method (Charm Work manual 1996). For more information on the mechanics of calibration and prediction using PLS algorithm refer to Geladi and Kowalski (1986).

2.2.3 Conclusion

This section has briefly discussed three different methods of creating a calibration model for the subsequent prediction of contaminant concentrations from their spectral data. The three methods looked into were the MLR, PCR and PLS techniques. MLR models the concentration matrix (**Y**) from all of the available spectra data (**X**) using a least squares criteria. Hence the MLR calibration model will include the spectra variation due to for example instrument noise and sampling errors, resulting in noisy predictions. PCR first decomposes the spectra data through the use of PCA and

regresses the decomposed set of information to the concentration matrix (**Y**). This reduces the dimensionality making model building a less difficult computational task. By applying a PCA prior to the regression step means that only the most important variation in the spectra are used for model building reducing the risk of including noise into the calibration model. Finally PLS was discussed which decomposes the spectra and the concentration matrix simultaneously. The advantages of PLS over PCR is that it incorporates more information in the model building phase. PLS will be implemented in this work for the analysis of UV-VIS spectra for the prediction of contaminants levels from unknown samples.

2.3 Flocculation model for process characterisation

2.3.1 Introduction

The purpose of this section is to discuss methods of representing the flocculation curve mathematically. The mathematical description should take into account the main features of the flocculation mechanism and allow for real time model parameter estimation for rapid process characterisation and control. A over simplistic mathematical representation will lead to inaccuracy in process prediction, whilst too complex a model will result in difficulties in real time model parameter identification.

The main features of the flocculation mechanism were discussed in chapter 1, section 1.6.3; a summary is given below.

(1) With increasing flocculant concentration higher levels of components will be flocculated and thereby be removed in the solid-liquid separation step.

(2) For certain components a maximum flocculation level is only possible. (For most flocculation systems there is an optimal flocculation concentration, however this does not necessarily result in floc sizes which can be totally removed in the solid-liquid separation step.)

(3) Re-stabilisation of flocs may occur at high flocculant levels.

(4) Ionic strength and pH conditions affect the efficiency of flocculation.

Most mathematical flocculation models discussed in literature (Moudgil and Behl 1991) have been based on the binary collision of particles in a manner consistent with orthokinetic flocculation. These mathematical models put forward, are usually based on

several constraints, such as by Moudgil and Behl (1991) assuming that flocculation is not reversible, and their practical use have been limited.

Two different views on mathematical modelling may be applied. The simplest technique considers external representation of the system i.e. a black box or empirical model. An alternative view of process modelling is from first principles where total mass, component, energy and momentum conservation equations are used. This section will concentrate on the former method in describing the flocculation characteristics.

Three flocculation solubility models will be put forward and their ability to meet the criteria listed above will be discussed. How their model parameters affect the shape and position of the flocculation profile will be investigated through the use of analytical solutions of the model derivatives when possible and/or through graphical examples.

2.3.2 The two-parameter model

Richardson *et al.* (1990) adopted a two-parameter model to describe the solubility curves for a precipitation process:

$$y_2 = \frac{E_i}{E_0} = \frac{1}{1+\left[\frac{x}{a}\right]^n} \qquad (2.6)$$

where E_i and E_o are the component (product and contaminants) concentrations before and after flocculation. In both cases the suspension has been treated by a solid-liquid separation step. The variable x is a measure of the precipitant dose and model parameters a (same units as x) and n (dimensionless) are related to system conditions such as pH, ionic strength, etc. Although the above model has been successfully used to describe the precipitation solubility curve by several authors (Richardson *et al.* 1990, Holwill *et al.* 1997), it is not appropriate in describing the flocculation system. As the above equation will always tend to zero as x becomes larger than parameter a it fails to describe two of the main flocculation modelling criteria, namely the maximum removal level and the re-suspension phenomena.

2.3.3 The 3(1)-parameter model

The 3(1)-Parameter model is an adaptation to the precipitation model using an additional model parameter:

$$y_{3(1)} = \frac{E_i}{E_0} = \frac{1}{1+\left(\frac{x}{a}\right)^n} + cx \qquad (2.7)$$

where c (reciprocal units to x) may also be related to system conditions such as pH, ionic strength, etc. Equation 2.8, shows that the point inflection for the 3(1)-parameter model occurs at a flocculant dose approximately equal to parameter a (See appendix C for more detail) and the slope at this point is described by equation 2.9, i.e. a function of all three parameters.

$$\frac{d^2 y_{3(1)}}{dx^2} = 0 \quad \Rightarrow x \approx a \quad \text{, when } n > 4.5 \qquad (2.8)$$

$$\frac{dy_{3(1)}(a)}{dx} = -\frac{n}{4a} + c \qquad (2.9)$$

Solving the derivative of $y_{3(1)}$ for zero with respect to x did not render any apparent analytical solution and hence a quantitative analysis of equation 2.7 was not possible. The behaviour of the 3(1)-model for different sets of model parameters is illustrated in figure 2.6. The plot shows how model parameter a influences the position of the flocculation profile (curve (1) to (2)) as described by equation 2.8 and 2.9. Parameter n can manipulate the gradient of the flocculation curve (curve (1) to (3)), with a steeper profile gradient for larger n values. The minimum of the 3(1) model seems to be dependent on all three model parameters, however mainly on parameter c as indicated by curves (2) and (4). As x becomes large the first term of equation 2.7 will tend to zero and the flocculation yield $y_{3(1)}$ will approach cx. An important feature of the 3(1) model is its non-linearity, and the behaviour of the $dy_{3(1)}/dx$ term at different levels of flocculant. From figure 2.6 it can be seen that the effect of flocculant concentration on the flocculation yield term is greatest close to the point of inflection.

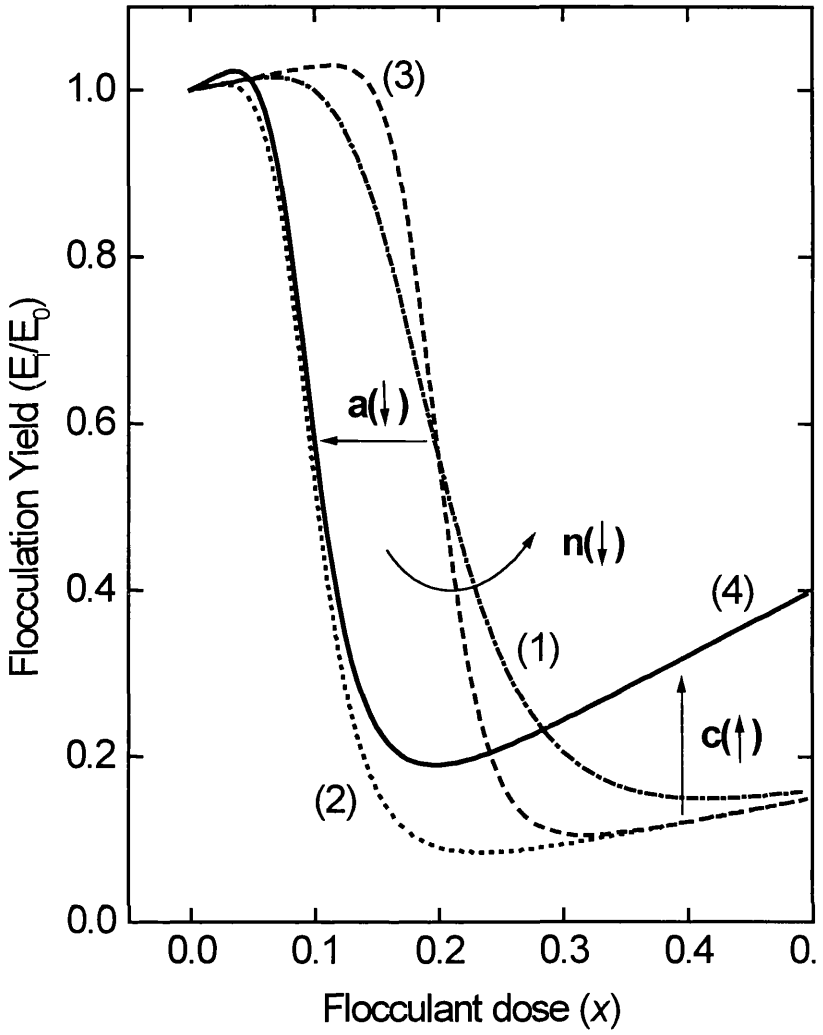


Figure 2.3. *Behaviour of the 3(1)-Parameter model for different sets of model parameters. The arrows on the plot show the shift in the flocculation profiles due to model parameter changes. The arrows in the brackets indicate an increase or decrease in the given model parameter. The sets [a, n, c] of model parameters applied are; (1)..... [0.2, 5, 0.3],* (2) ••••••• [0.1, 5, 0.3], (3) - - - - - - [0.2, 10, 0.3], (4) ——— [0.1, 5, 0.8].

2.3.4 The 3(2)-parameter model

The 3(2)-Parameter model (equation 2.10) is empirically derived. This model describes how with higher flocculant concentration levels greater proportions of a particular component will flocculate and thereby be removed in the solid-liquid separation step. The model also describes a certain minimum level of removal of a specific component. However the re-stabilisation effect is not represented.

$$y_{3(2)} = \frac{E_i}{E_0} = 1 - A\left(1 - Exp\left(-\left(\frac{x}{B}\right)^C\right)\right) \qquad (2.10)$$

where the model parameters A (dimensionless), B (same units as x) and C (dimensionless) are dependent on system conditions. Differentials of equation 2.10 provide insight on how the three model parameters influence the shape and position of the 3(2)-model.

$$\frac{dy_{3(2)}}{dx} = AB^{-C}Cx^{(C-1)}EXP\left(-\left(\frac{x}{B}\right)^C\right) \qquad (2.11)$$

$$\frac{d^2y_{3(2)}}{dx^2} = 0 \;\Rightarrow\; x = B\left(\frac{C-1}{C}\right)^{\frac{1}{C}} \qquad (2.12)$$

For C > 3.5 the $\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$ term in equation 2.12 will approach unity (see appendix C for more detail), hence it can be argued that parameter B determines the point of inflection on the 3(2)-parameter model profile. The minimum level of removal for a specific component is determined by parameter A, as when x >> B, $y_{3(2)}$ will approach $1\text{-}A$. This leaves parameter C to influence the slope of the curve. Although from equation 2.11, it can be seen that all parameters influence the slope to some degree, parameter C is a power term and hence will have the greatest impact. Figure 2.4 illustrates the above arguments graphically. The upward shift in curve (1) to (2) was due to a decrease in parameter A, which resulted in a decrease in the maximum removal level. The shift in curve (2) to (3) was due to a decrease in parameter B, which yielded a shift in inflection point. Curves (1) and (4) illustrate the decrease in flocculation profile slope due to a decrease in parameter C.

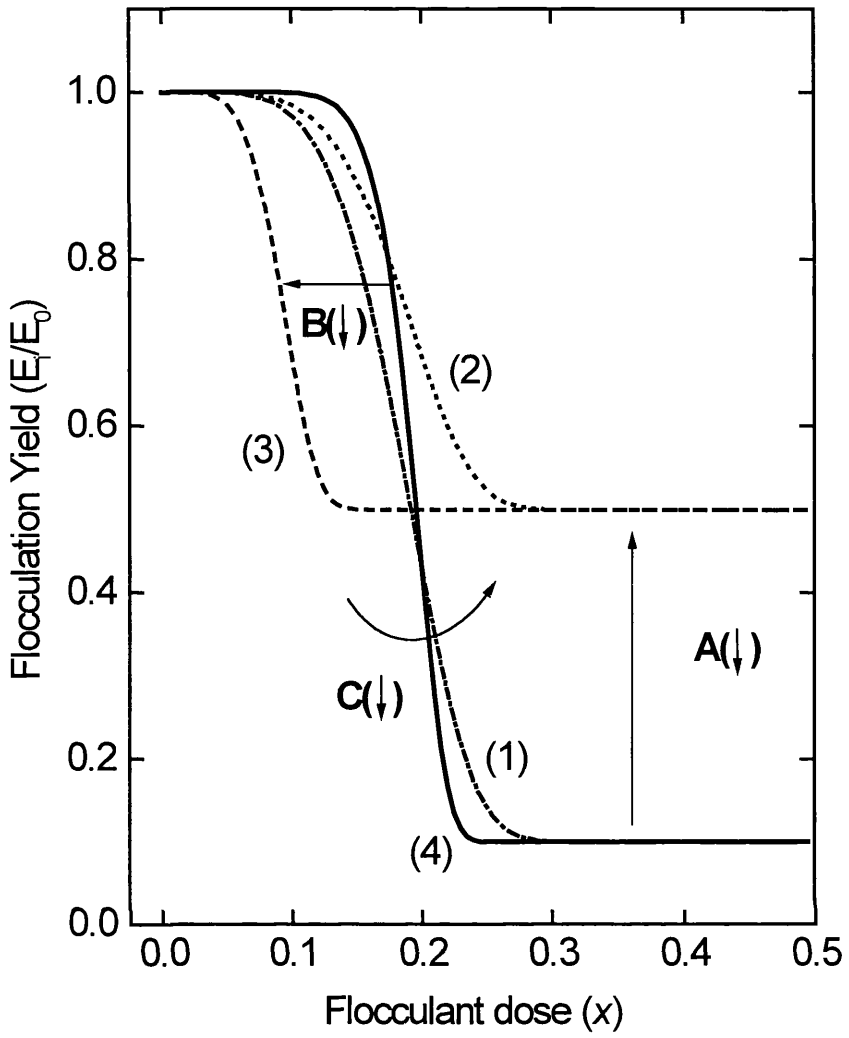


Figure2.4. *Behaviour of the 3(2) Parameter model for different sets of model parameters. The arrows on the plot show the shift in the flocculation profiles due to model parameter changes. The arrows in the brackets indicate an increase or decrease in the given model parameter. The sets [A, B, C] of model parameters applied are; (1) -.-.-.-.- [0.9, 0.2, 5], (2) ····· [0.6, 0.2, 5], (3) - - - - - - [0.6, 0.1, 5], (4) ——— [0.9, 0.2, 10].*

2.3.5 Four parameter model

The four-parameter model is an extension of the two-parameter model and includes a driving mechanism to promote re-stabilisation at high flocculant concentrations. The mathematical representation can be seen below, equation 2.13.

$$y_4 = \frac{E_i}{E_0} = \frac{\left[1+\left[\frac{x}{b}\right]^m\right]}{\left[1+\left[\frac{x}{a}\right]^n\right]} \qquad (2.13)$$

where model parameters a (same units as x) and n (dimensionless) have the properties given in equation 2.6 and parameters b (same units as x) and m (dimensionless) are related to system conditions to drive the re-stabilisation mechanism. Due to its analytical complexity the impact of the parameters on the model is best described graphically figure 2.5.

Figure 2.5 illustrates the shift in the flocculation profiles due to changes in the four model parameters. It can be seen from curves (1) and (2) that a decrease in parameter a resulted in an increase in the flocculating mechanism (denominator of equation 2.13) which led to a shift to the left and a reduction in the gradient of the re-suspension part of the profile. The shift in curve (2) to (3) was a result of an increase in parameter m, making the re-suspension driving force (numerator of equation 2.13) larger. The change in parameter m did not seem to effect the maximum removal level or the flocculant dose where this minimum occurs. Curves (3) and (4) illustrate the profile change due to a decrease in parameter n, which decreases the flocculation driving mechanism. The maximum removal level also decreased as a result, however the flocculant level where this occurred did not change. Finally the shift from curve (4) to (5) was caused by a decrease in parameter b, which resulted in an increase in the re-suspension mechanism. Both a shift and a decrease in the maximum removal level were observed. All in all it can be concluded that the behaviour of the 4-Parameter model is more complex to understand than that of the three parameter models. The parameters a and b seem to influence the position of flocculation inflection, yield minimum and re-suspension points in terms of flocculant concentration. Parameters n and m appear to influence the gradient of the flocculation and re-suspension parts of the flocculation profile, as well as the minima point to some extent.

It can be seen from figure 2.5 that the 4-parameter model describes the same flocculation characteristics as the 3(1)-parameter model. The benefits of an additional model parameter is model flexibility however, when attempting to perform rapid model parameter identification the computational cost is larger.

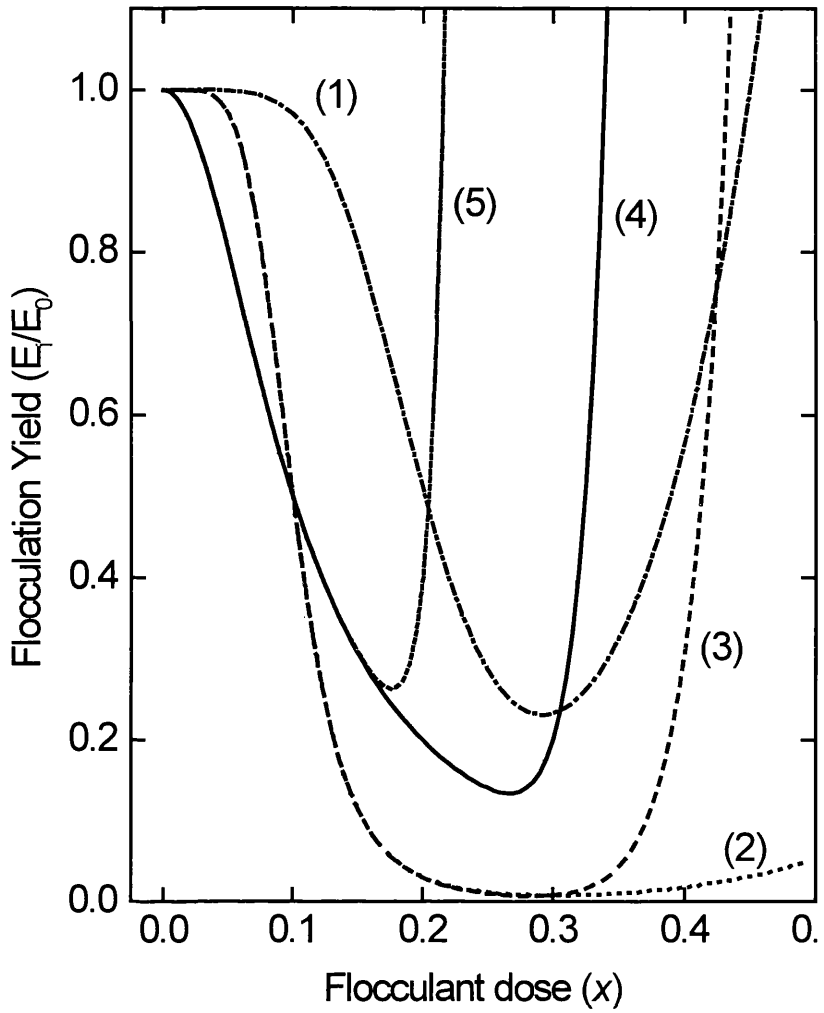


Figure 2.5. *Behaviour of the 4-Parameter model for different sets of model parameters. The model parameters sets [a, b, n, m] applied are;*
(1)-·-·-·-·- [0.2, 0.3, 5, 10], (2) ····· [0.1, 0.3, 5, 10], (3) - - - - - - [0.1, 0.3, 5, 20],
(4) ——— [0.1, 0.3, 3, 20], (5) ..... [0.1, 0.2, 3, 20].

2.3.6 Conclusion

Three mathematical representations of the flocculation mechanism have been put forward and their ability to describe the flocculation process studied. The models differed in their complexity, number of model parameters and manner in which they described the flocculation solubility curve.

The 3(1)-parameter and 4-parameter model put forward described all of the flocculation features discussed in the introduction, although the analysis of these models were relatively complex. The 3(2) parameter model described the majority of the flocculation

characteristics apart from the re-stabilisation affect at high flocculant concentration. For process control purposes this feature is less important as the flocculation process will not be operated in this region. The benefits of the 3(2)-parameter model is that its model parameters clearly describe the performance of the flocculation profile, allowing the operator to rapidly and effectively evaluate the performance of the process for different environmental conditions.

The final application of the characterisation model is an important issue to consider. For non-time constraint situations where a large number of data points are available such as for post process analysis a more complex and accurate model can be implemented. However, for real time applications during a process run with initially a scarce number of data points, a less complex model may be more suitable. A further analysis of these three flocculation solubility (particulate remaining in solution) models will be investigated in terms of experimental data in chapter 5.

2.4 Model parameter estimation techniques

2.4.1 Introduction

Two model parameter identification techniques will be discussed in this section, namely the non-linear least squares Levenberg-Marquardt and extended Kalman filter methods and a simple linear example will be given, to give the reader an insight into their functionality.

The objective of process identification is to obtain an accurate process model. The accuracy of a model can be defined in several ways, each resulting in different problem formulation. Most parameter estimation techniques require three basic factors; input-output data from the process which requires to be modelled, a class of model and a criterion. Parameter estimation can then be formulated as an optimisation problem, where the best model is the one that best fits the data according to the given criteria. There are several ways of combining, experimental data, model classes and criteria. There are also many different ways to organise the computations. Consequently, there is a large number of different identification methods available.

Two basic/classic identification techniques are available to the engineer. These are based on the "least squares" and the "maximum likelihood" methods, which have different criteria. The least squares (LS) method involves fitting a curve to the available

data in such a way that the sum square of the residuals to the data points is minimised. The maximum likelihood (ML) method introduces a weighting factor into the estimation process, in order to make allowances for the accuracy of the measuring instrument as well as the accuracy of the process. Hence the ML method estimates parameters that make the collected data most plausible, whilst the least squares approach sets the model parameters to find the best fit.

Several identification methods have been developed based on the LS method, for linear and non-linear systems (Levenberg-Marquardt), for on-line application (recursive least squares) and to account for time varying model parameters (least squares with a forgetting factor). Probably the most important development using the ML method for parameter estimation is the Kalman Filter (KF) for linear systems and extended Kalman filter (EKF) for non-linear models.

The Non-linear Least Squares (NLLS) analysis method applied here is the Levenberg-Marquardt technique. This method is a combination of the general, quadratic-base, linear regression technique, and the gradient-descent method for function optimisation, by which the gradient of a multi-dimensional function is followed backwards toward a local or global minimum of the sum of squares error function. In this method, a starting point for the model parameters are input by the user, and the iterated technique repeatedly finds better combinations of the parameter values which fit the input data with less overall error.

As mentioned above the first fundamental concept of the Kalman filter is its use of the maximum likelihood method, the second key concept is its recursive or iterative nature, brought forth by Kalman (1960). Du Plessis (1967), reviews the equations used in the Kalman filter, and states that its recursive nature cannot be over emphasised. No past data need ever be stored, as each estimation is identical in procedure to all those which took place before it, but each has a new weighting factor (Kalman gain) computed to take into account the sum total effect of all the previous estimates. To start the cyclic process of the Kalman filter, three quantities are needed: 1) The expected value of the model parameters prior to taking of any readings. 2) The mean square deviation about nominal value of the model parameters. 3) The mean square error in the measurements. In practice one or more of these conditions are only approximately known, and are often, at best, only estimates or educated guesses. Hence in some situations for poorly defined processes, the Kalman filter will not be tuned optimally and the filter is less able to follow certain system disturbances.

The choice of model identification technique will depend on several factors, such as the application of the estimation technique, the type and accuracy of the model, the measurement noise and frequency, and the extent of model parameter variations with changing conditions and with time?

A LS method has been applied previously (Holwill *et al.* 1997) for identifying model parameters for the precipitation method, and Holwill states it has several shortcomings, notably the need for several measurements before a stable estimate is reached, particularly with very noisy data, in comparison to the Kalman filter.

In general the use of the EKF does not guarantee convergence of parameters to their true values (Yoshimura *et al.* 1980). Heijden's *et al.* (1989) review on state estimators, states that trial and error is often necessary to make the EKF work satisfactorily, even though the mathematical approach may suggest otherwise. Ramamurthi *et al.* (1993), compared the performance of a successively linearised horizon based estimator, with the EKF and a non-linear programming (NLP) approach and conclude that the main drawbacks of the EKF is that it requires prior knowledge, and that the tuning of the filter can be a difficult task. The main disadvantage of the NLP approach is the computation time required to obtain convergence of the model parameters. It should be noted that several attempts to compare different identification methods have been made and it seems to be inconclusive, in the sense that there is no method that is universally best (Åström 1984).

2.4.1 Extended Kalman filter

This section will briefly discuss the concept behind Kalman filtering. For linear systems the Kalman filter (KF) may be applied whilst for non-linear systems the extended Kalman filter (EKF) is more appropriate. The main difference between these two approaches is that when the KF is implemented the system is fully linearised for all data points, whilst during EKF application, the model is only linearised around the current operating point, and hence derivatives of the system (model) are introduced. During the Kalman filter model parameter estimation, a probability based weighting factor (Kalman gain), determined from the covariance matrices of the system and observation disturbances is used to apply adjustments to the model parameters using the latest measurements. Such a technique estimates the most likely model parameter values with the given information assuming that the observation and system noise are Gaussian random processes. Equations 2.14 and 2.15 are the process model.

$$y(i) = f(X(i)) + v(i) \quad (2.14)$$

$$X(i) = \Phi X(i\text{-}1) + w(i) \quad (2.15)$$

where y is the observation (measurement), the X matrix contains model parameters and describes the state of the process, v describes the measurement noise, w refers to the system noise and Φ is the transition matrix. The Kalman filter estimate of up to date model parameters based on newly acquired measurements is given by equation 2.16.

$$\hat{X}(i) = \hat{X}(i-1) + K(i) \cdot [y(i) - f(\Phi \hat{X}(i), i)] \quad (2.16)$$

where K is the Kalman gain vector and the $\hat{X}$ matrix contains the estimated model parameters. The first term on the right hand side of equation 2.16 is the estimate from the previous iteration step. The term in the squared brackets is called the *innovation* and is the difference between the measurement and the model prediction. If the measurement and the model prediction value are similar the innovation term will be small and little adjustment is made to the model parameters. However, if the difference between the measurement and the model prediction is large, the amount of adjustment to model parameters will increase. The level of adjustment is dictated by the Kalman gain, which acts as a weighting factor. The Kalman gain (K) can be determined through equations 2.17 to 2.20, and is a function of several factors.

$$P_1(i) = \Phi P(i-1) \Phi^T + Q \quad (2.17)$$

where;

$$P(i) = [I - K(i)H(i)]P_1(i) \quad (2.18)$$

$$K(i) = P_1(i)H^T(i) \cdot [H(i)P_1(i)H^T(i) + Z]^{-1} \quad (2.19)$$

$$H(i) = \left[\frac{\partial f}{\partial X}\right]_i \quad (2.20)$$

where P matrix is the combined system covariance, H vector contains the model derivative terms, Z is the covariance of measurement noise and Q is the operator determined system covariance term. Equation 2.17 shows that the Kalman weighting is dependent upon the relative magnitudes of the measurement noise (Z), the combined parameter co-variances (P) and the model derivative vector (H). The combination over time of this balance will give the optimum estimates for the parameter matrix, given the statistical assumptions are correct. The above equations describing the Kalman filter are based on probabilistic theorem and are accounted for in more detail in many texts (Gelb 1974, Åström and Wittenmark 1984, Catlin 1981).

If we use the flocculation model expressed in equation 2.10 (also shown in equation 2.21) as the system example, i.e. a static three parameter model, and the objective is to estimate the model parameters for each new measurement step (*flocculant concentration (x_i), flocculation performance (y_i)*) the above matrix notation would be:

$$y = f(X,i) + v(i) = 1 - A\left(1 - Exp\left(-\left[\frac{x_i}{B}\right]^C\right)\right) + v(i) \qquad (2.21)$$

$$X(i) = \begin{bmatrix} A \\ B \\ C \end{bmatrix}_i ; \qquad H(i) = \begin{bmatrix} \frac{\partial y}{\partial A} & \frac{\partial y}{\partial B} & \frac{\partial y}{\partial C} \end{bmatrix}_i$$

$$Q(i) = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}_i ; \qquad P(i) = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}_i \qquad (2.22a, b, c)$$

$$K(i) = \begin{bmatrix} K_A \\ K_B \\ K_C \end{bmatrix}_i ; \qquad \Phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} ; \qquad Z = \sigma_m^{\,2}$$

where the model parameters are A, B and C, and the system and combined system covariance terms for parameters A, B and C, are q_{11}, q_{22}, and q_{33} and p_{11}, p_{22}, and p_{33}. The Kalman gain vector consists of K_A, K_B and K_C, which are the Kalman weighting values specific to the model parameters. The measurement covariance term is σ_m^2 based on prior analysis of the measuring instrument, where σ_m is the standard deviation of the observed values. The P and Q matrices are initialised with information based on past experience and, P is subsequently recalculated at every step by the algorithm. Vector H consists of the model derivative terms (shown in Appendix C) and therefore is position specific.

2.4.2 Least squares method

The least squares method involves fitting a curve to the available data in such a way that the sum square of the residuals to the data points is minimised. The sum of squares can be expressed as:

$$S = \sum_{i=1}^{n} [f(x_i, A, B, C) - y_i]^2 \qquad (2.23)$$

where A, B and C are model parameters and the n values of (x_i, y_i) are known measurements. When these known measurements are inserted in the right hand side of equation 2.23 S becomes an expression function of only three unknowns A, B and C. The objective is to choose the set of model parameters that minimises S. There are several methods available to minimise the sum square residual term (S). For linear models the partial derivatives of S may be solved for zero to yield the optimum set of model parameters, shown in equation 2.24.

$$\begin{bmatrix} \frac{\partial S}{\partial A} \\ \frac{\partial S}{\partial B} \\ \frac{\partial S}{\partial C} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad (2.24)$$

For the flocculation model expressed in equation 2.21, the minimisation of S is a more difficult task due to the model's non-linearity. Alternative minimisation approaches can be implemented. The simplest to apply are based on a searching technique such as the simplex method, which only uses information from the values of the function itself (no derivatives), and as a result, although being a robust minimisation method, it sometimes takes a long time to find a minimum.

Alternative methods use information contained in the partial derivatives of the sum of squared residual function (S), in order to follow the function downhill until S reaches a minimum. These alternative approaches fall into two broad categories: the directional set and quasi-Newton methods. The most widely used method in non-linear least squares model parameter identification is the Levenberg-Marquardt (LM) algorithm. This method is based on the insight that (a) near the actual minimum, the function S should be very nearly parabolic in the parameters, and thus quadratic methods will be most efficient, and (b) far from the actual minimum, quadratic methods take too tiny a step, and steepest descent methods employing the gradient are more efficient. In this method, a starting point for the model parameters is input by the user, and the LM iterated technique repeatedly finds better combinations of the parameter values which fit the input data with less overall error. For more detailed information on this technique refer to Seber and Wild (1989).

2.4.3 Linear example

This section will give a simple example of how the least squares and Kalman filter estimation techniques behave in terms of predicting a single model parameter of a static linear function. Suppose we are attempting to estimate the model parameter a for the linear model in equation 2.25.

$$y = f(c_i) = ac_i + v \qquad (2.25)$$

where y_i is a measurement, c_i is the input variable, a is the model parameter and v is the noise associated with the measurements due to the inaccuracy of the measuring instrument. We have available an initial guess as to what the model parameter might be, and also know that our system fluctuates, i.e. model parameter a will change from run to run and even within a run. The expected variation of model parameter a is based on operator experience. Given we have n measurements (c_i, y_i), the following text will discuss the different approaches in estimating the model parameter.

Least squares approach

As discussed earlier the objective of the least squares method is to minimise the sum of squared residuals between the observed/measured and expected values. The sum of squared residuals can be written as:

$$S = \sum_{i=1}^{n} [f(c_i, a) - y_i]^2 = \sum_{i=1}^{n} [ac_i - y_i]^2 \qquad (2.26)$$

To find model parameter a in order to minimise S we need to differentiate S with respect to a and solve for zero.

$$\frac{dS}{da} = 0 \Rightarrow a\sum_{i=1}^{n} c_i^2 - \sum_{i=1}^{n} c_i y_i = 0 \qquad (2.27)$$

$$a = \frac{\sum_{i=1}^{n} c_i y_i}{\sum_{i=1}^{n} c_i^2} \qquad (2.28)$$

Hence for every measurement (c_i, y_i), a new model parameter estimation can be calculated by inserting the values into equation 2.28.

Kalman filter approach

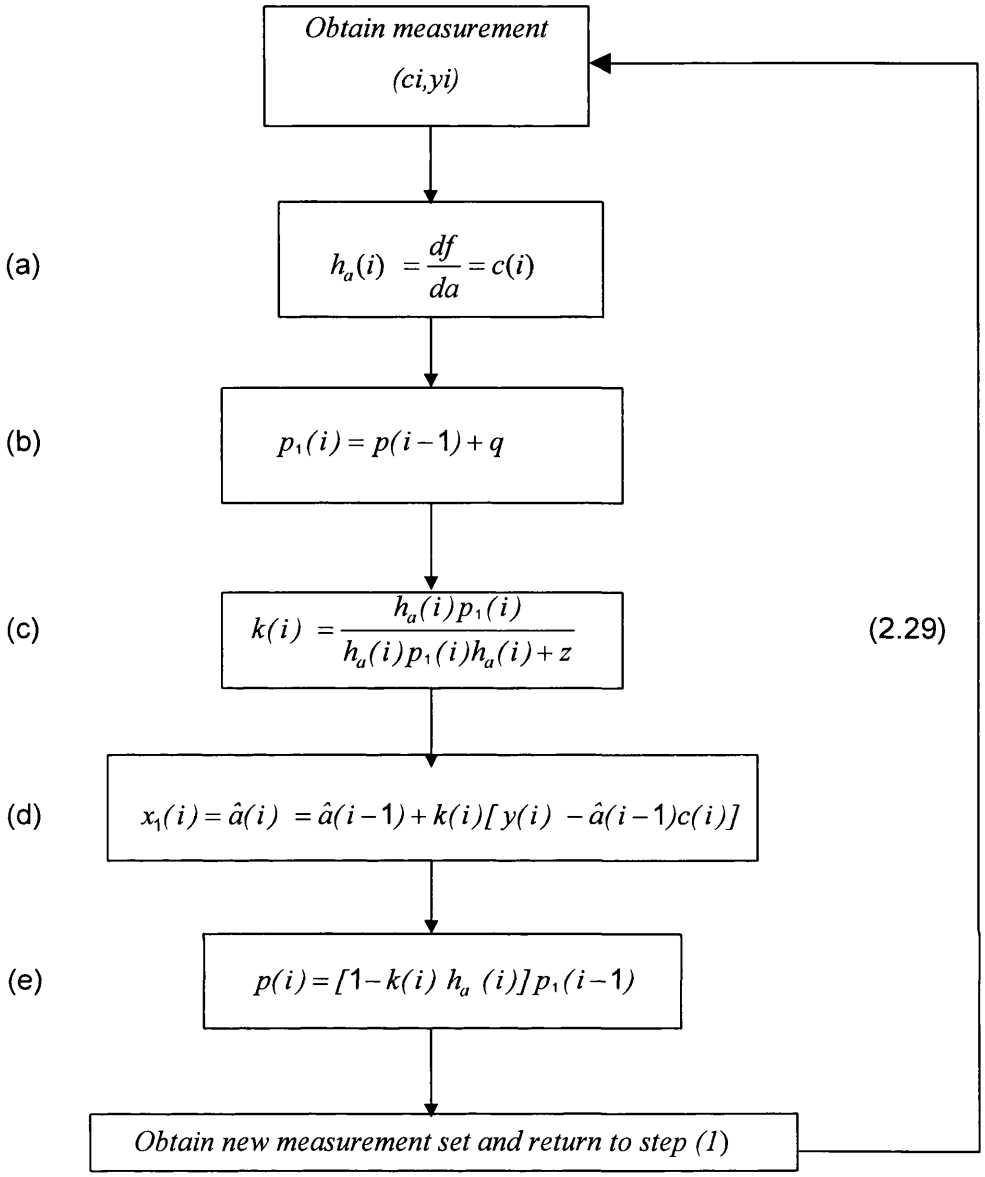
As mentioned above the Kalman filter is based on the maximum likelihood theorem. Hence model parameter estimation are weighted with a factor (Kalman gain) which takes into account the relative quality of the measurement instrument and the approximate (known) variation of the model parameter from run to run to obtain the optimum model parameter estimates.

Several additional factors need to be specified before the Kalman filter can be applied for parameter estimation of the above linear model. Factors such as the system and measurement noise values, and combined system co-variance need to be defined. Given these are chosen the Kalman filter would follow the re-cursive iterative procedure described in equations 2.29 (steps (a) to (e)) for each new measurement step (c_i, y_i) in order to estimate the new model parameters. Due to the static linear system with one model parameter the Kalman filter matrix notation has been reduced to a scalar system, where h_a refers to the function derivative with respect to the model parameter a. The combined system co-variance is $p\ (=\sigma_a^2)$, the system and measurement noise terms are referred to as q and $z(=\sigma_y^2)$, the Kalman gain weighting factor is k and the dynamics matrix is $\Phi = 1$.

For the Kalman filter sequence to start a data point (c_i, y_i) is acquired.

Step (a) then determines the model derivative (h_a). For the linear example $h_a(i) = c(i)$, i.e. function of model structure and position specific. This is used during EKF for the linerisation of the model around the sample point.

Step (b) determines the combined system covariance term $p(i)$. This is merely based on the previous combined system noise term $p(i-1)$ and the operator chosen system noise term (q). The system covariance term (q) is applied so that the combined system covariance always has a minimum value to allow for variation in model parameter estimation. The larger the chosen system covariance term the larger the allowable variation in model parameter estimation between iterations.



Step (c) involves the determination of the Kalman gain (k). As seen from this equation the Kalman gain is a function of the combined covariance (*step (b)*), the measurement covariance term (z) as well as the model derivative h_a (*step (a)*). If the measurement instrument was very accurate relative to the variance in the model parameter estimates ($p >> z$), the Kalman gain ($k(i)$) would approach $1/h_a(i) = 1/c(i)$. In other words one simply would use the raw data to evaluate new model parameter values (inserting $k(i)$ into *step (d)* gives $a(i) = y(i)/c(i)$). On the other hand if the measurement noise was very large relative to the combined system covariance ($z >> p$), the gain term would approach zero. In this instance, one would reject the measurement of $y(i)$ and simply use the nominal value of the model parameter as our estimate (from *step (d)* $a(i) = a(i-1)$). The Kalman gain is also a function of the combined system covariance and the model derivative terms. By choosing an initial high combined system covariance term,

the Kalman gain will initially be high and with time decrease (as the objective of the Kalman filter is to minimise the system covariance). The magnitude of the Kalman gain also depends on the position of the data point on the process model i.e. function $c(i)$. For the linear example the lower the $h_a(i)$ term the higher the Kalman gain.

Step (d) performs the model parameter estimation based on the newly determined Kalman gain. The Kalman gain acts as a weighting factor to modify the previous estimated model parameter according to the size of the innovation terms (term in squared brackets).

Step (e) adjusts the combined system covariance term (p) to account for the new optimum estimate of the model parameter, which reduces the covariance term.

Figure 2.6 presents three examples where different levels of measurement noise (z) values were implemented. The examples adopted equation 2.28 and 2.29 (steps (a) to (e)) for the prediction of model parameter a for the least squares and Kalman filter techniques. A worked example in tabular format is also given in table 2.1, which goes through a step by step approach. The tuning constants and the initial parameter guess are listed in the figure and table legends.

Steps(i)	c(i)	y(i)	$a_{LS}(i)$	$h_a(i)$	$p^-(i)$	k(i)	$a_{KF}(i)$	p(i)
0	7.509	43.735	5.824	7.509	8.02	0.126	5.781	0.420
1	7.208	55.084	6.696	7.208	0.440	0.066	6.670	0.230
2	2.453	15.795	6.683	2.453	0.250	0.023	6.657	0.236
3	7.747	53.978	6.781	7.747	0.256	0.049	6.775	0.158
4	8.105	64.381	7.099	8.105	0.178	0.039	7.148	0.121
5	5.668	39.515	7.084	5.668	0.141	0.027	7.121	0.120

Table 2.1. *A simple parameter identification example applying a least squares and Kalman filter estimation approach. The prediction of model parameter a for the least squares and Kalman filter are termed a_{LS} and a_{KF} and were determined through equations 2.28 and 2.29. The actual model parameter was set to 7 and a 10% standard deviation was added to each observation. The system covariance (q) was set to 0.02 (i.e. a 2.8 % standard deviation), the initial combined covariance term p was 8 and the measurement noise (z) was 25. The initial parameter guess for a was 5.*

Figure 2.6 plot (a) shows the estimates of the model parameter for a relatively well tuned Kalman filter and the least squares approach. Both estimation techniques target the actual model parameter after a few data points. However when the tuning of the Kalman filter is poor the estimation of the model parameter is less effective as seen from plots (b) and (c). Plot (b) illustrates how the Kalman filter behaved with a measurement covariance value (z) of zero i.e. the operator assumed that the measuring instrument was 100% accurate. The estimates rely solely on the raw data

resulting in a noisy outcome due to the actual 10% standard deviation in the observed measurements. On the other hand when the measurement covariance term was set too high z ($=\sigma_m^2$) = 300, plot (c), the estimation of the model parameter relied mainly on previously estimated values, due to the little confidence in observed values. The outcome was a slow approach to the actual model parameter.

With more complex non-linear models involving multiple parameters, the least squares approach becomes a more difficult task with the need for more sophisticated minimisation algorithms as discussed earlier. For the simple example given here the two approaches seem to be equally efficient when the Kalman filter is tuned correctly. All in all the Kalman gain is a function of several factors many of which are operator chosen. This enables high flexibility in model parameter prediction especially for a well known system, however for a badly defined process poor Kalman filter tuning will result in poor estimation results.

2.4.4 Conclusion

Two model parameter identification techniques based on the least squares and maximum likelihood (Kalman filter) criteria have been discussed. Although the Kalman filter is applied in an unconventional manner (for model parameter identification rather that state estimation), this approach has been included due to its ability to take into account the inaccuracy in the model and measurements. A simple example was given to illustrate the main differences and functionality of the two estimation techniques. Which method is most efficient is a function of multitude of factors, such as complexity of model, frequency, noise and position of observations, variation of model parameter from run to run and knowledge of the system one is attempting to characterise.

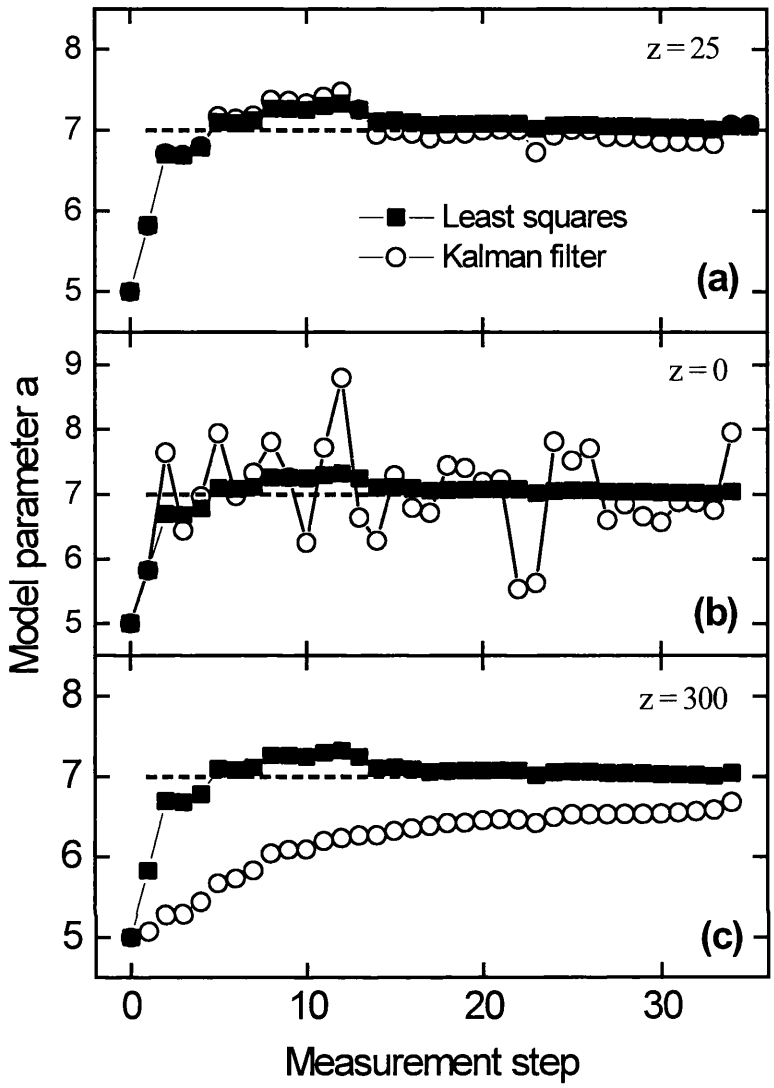


Figure 2.6. *The least squares and extended Kalman filter estimation of model parameter a for different measurement noise values. The prediction of the model parameter a for the least squares and Kalman filter approaches were determined trhough equations 2.28 and 2.29. The actual model parameter was set to be 7 and a 10% standard deviation was added to each observation. The system covariance (q) was set to 0.02 (i.e. a 2.8 % standard deviation) whilst the initial combined covariance term (p) was set to 8. Three different measurement noise terms were adopted, $z \in [0, 25,300]$. The (c_i,y_i) measurements were randomly chosen, where $c_i \in [0 ; 10]$, $y_i \in [0; 70]$. The initial guess for parameter a was 5.*

Chapter 3

Material and Methods

3.1 Feed material and flocculant

3.1.1 Polyethyleneimine (PEI)

PEI (Fluka, 50% w/v, Mr 600,000 – 1,000,000) was dissolved in de-ionised water or phosphate (KH_2PO_4) buffer and the pH was adjusted with 4M HCL. PEI solutions were prepared over the range 5 – 20 mg mL^{-1}, with pH values between 5 – 8 and ionic strengths (KH_2PO_4) 0 - 100mM.

3.1.2 Yeast Homogenate

Blocks of Baker's yeast *Saccharomyces cerevisiae* (The Distillers Company, Surrey, UK) were suspended in phosphate buffer and was adjusted with 4M HCL or 4M NaOH. Suspended Baker's yeast volumes between 3-4 L (125 – 280 g (wet weight) cells L^{-1}) were disrupted, in a model 15M 8BA Manton Gaulin high-pressure homogenizer (APV, Crawley, Sussex, UK) at 500 bar and 5 discrete passes controlled to a temperature of 4-5°C. Suspended Baker's yeast volumes larger than 10 L were disrupted in a pilot scale high pressure homogeniser (Model K3, APV, Crawley, UK), run in continuous mode at 500 bar at a throughput of 280 L h^{-1}, equivalent to 5 discrete passes. The temperature was kept below 5°C.

3.1.3 Standards for spectral calibration

Standards used for spectral calibration were yeast ribonucleic acid (RNA, highly polymerised), bovine serum albumin (BSA, fraction V) which were both supplied by Sigma Chemical Ltd (Poole, Dorset, UK). Both RNA and BSA standards were suspended in phosphate buffer (pH 6.5, 50mM KH_2PO_4). Washed cell debris was obtained from yeast homogenate (280 g (wet weight) L^{-1}, pH 6.5, 50mM KH_2PO_4), 3 L) which was centrifuged in a Beckman (Model J2-M1, Spinco, Beckman Instruments, California, USA) with a fixed angle rotor (type JA-17) at 10,000 rpm for 10 minutes. The sediment was washed with phosphate buffer and re-spun (10 minutes at 10,000 rpm) in the Beckman centrifuge. The wash and spin cycle was repeated three times.

3.2 Off-line assays

3.2.1 Total protein assay

The Bradford (Bradford 1976) assay was used to determine total protein levels. The assay is based on the shift in absorbance from 465 nm to 595 nm which occurs when Coomassie Blue G-250 dye binds to proteins in acidic solution. A commercially available dye was used (Bio-rad Protein Assay reagent, Bio-rad, Hemel Hempstead, UK).

3.2.2 ADH assay

Alcohol dehydrohgenase (ADH) activity was assayed according to the method of Bergmeyer (1933), by following the absorption change at 340 nm due to the conversion of NAD^+ to NADH accompanying the conversion of ethanol to acetaldehyde. ADH catalyses the below reaction:

$$Ethanol + \beta NAD^+ \xleftrightarrow{ADH} Acetaldehyde + \beta NADH + H^+ \quad (3.1)$$

The formation of acetaldehyde from ethanol may be monitored by measuring the increase in absorbance at 340 nm due to the formation of βNADH. The reaction mixture consisted of 600mM ethanol (Fluka), 1.0mM glutathione (Sigma), 0.62mM semicarbazide HCl (sigma), 1.8mM NAD^+ (Sigma) in 50mM Tris HCl (Sigma) buffer at pH 8.8. Semicarbazide inhibits the reverse reaction. One unit of enzyme activity is defined as the amount of ADH necessary to catalyse the conversion of 1μmol of ethanol to acetaldehyde per minute at 25°C. All off-line assays were performed in 10mm path length cuvettes and the reaction was started by the addition of the enzyme. Potassium dihydrogen phosphate buffer (molarity and pH of sample) was used to dilute samples to produce a linear change in absorbance. Assays were performed in triplicates. For at-line assays a 1mm path length flowcell was implemented and a reaction time of 20-30 seconds adopted.

3.2.3 Cell debris measurements

Dry weights as well as optical measurements were used to measure cell debris levels.
Dry cell weights were determined by drying an Ependorf of 1mL homogenate sample suspension at 100°C for 24 & 48 h. The Ependorfs were pre-weighed and dried. The dry cell weights were determined gravimetrically.
Absorbance values at a wavelength of 650nm (OD_{650}) were recorded of samples using an Ultraspec 2000 (Pharmacia Biotech, Upsala, Sweden) against a buffer blank.

Samples were diluted with phosphate buffer to yield absorbance values in the linear region of the spectrophotometer (0 – 1 Au).

3.2.4 Ribonucleic Acid (RNA) assay

RNA was assayed using a method based on the orcinol assay (Bulmer 1992). The method was adapted for yeast as follows. Orcinol reagent was prepared by dissolving orcinol 3 g L^{-1} in concentrated hydrochloric acid and adding ferric chloride (10% w/v, 1mL). The orcinol reagent was stored in the dark and used fresh. Samples (100 μL) were precipitated with 60% perchloric acid (100 μL) in an Eppendorf tube and stored ar 4°C for 24 h and then centrifuged (13,500 rpm, 10 min). The supernatant (100 μL) was mixed with sodium hydroxide (NaOH, 2M, 100 μL) and incubated 92h, 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.

3.3 Sample preparation and flow system

3.3.1 Micro-centrifuge for solid-liquid separation

The microcentrifuge was implemented for automatic solid-liquid separation of flocculation samples. The micro-centrifuge is purpose built for the solid-liquid separation of small volumes of samples of biological material, constructed by John T, Bailey (Engineering) Ltd. (Smallfields, Surrey). It consists of an enclosed miniature bowl operated by compressed air with a friction brake, a solenoid valve, and three peristaltic pumps. The apparatus was housed by a 340 x 290 x 190 mm [L, W, D] stainless steel box, figure 3.1. A similar micro-centrifuge was adopted and described in more detail by Turner (Turner 1993) for fermentation sampling and Chard (Chard 1997) for precipitant separation.

The peristaltic pumps (303 D/A, Watson-Marlow Ltd.) were used to pump feed samples and wash buffer to the micro-centrifuge bowl. A fourth peristaltic pump situated on the stopped flow analyser (see section 3.3.2) was used to pump off the supernatant liquid. The compressed air to drive the micro-centrifuge was supplied by a Bambi compressor (Model 225/1000, Bambi air compressors Ltd, Springhill, Birmingham). The solenoid valve controlled the supply of a vacuum to the micro-centrifuge bowl to suck off excess feed and wash to waste. A water pump generated the vacuum. The micro-centrifuge bowl shown in figure 3.2, can accommodate 0.54 mL of sample, has a tulip shape with a top internal diameter of approximately 20 mm. Two stainless steel tubing are used to

supply feed and wash buffer to the micro-centrifuge bowl, and another two tubes situated closer to the bowl bottom are used to pump supernatant out of the bowl and to suck off waste.

The peristaltic pumps, solenoid valve, friction brake and the supply of compressed air to the micro-centrifuge were controlled by the supervisory computer through 7 digital signals (0-5V). A National Instruments (AT-DIO-32F, USA) digital I/O interface was adopted to execute and convey the relevant signals to the micro-centrifuge.

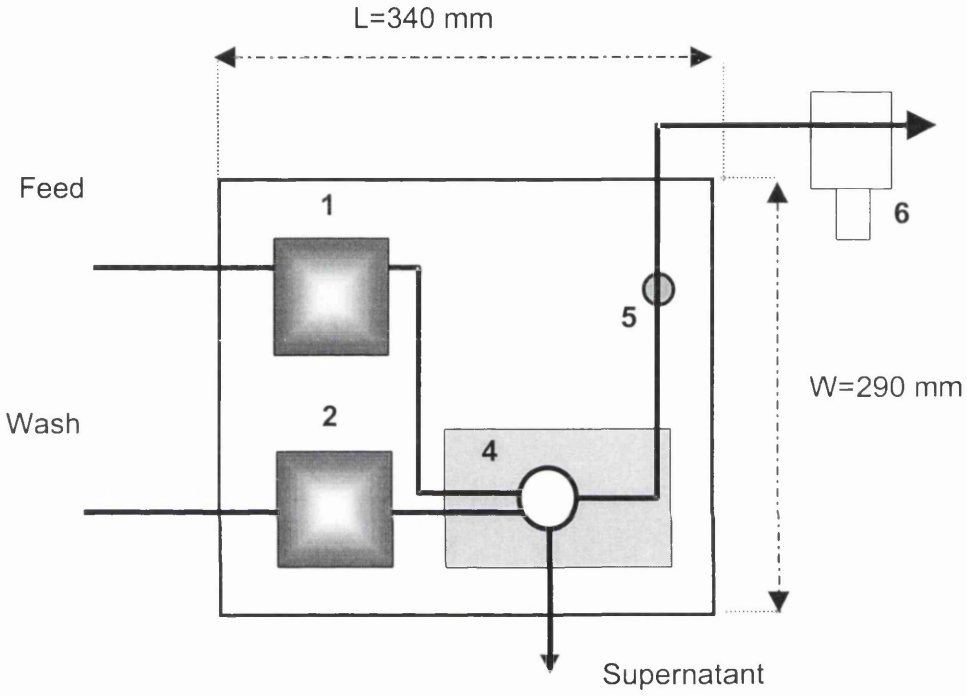


Figure 3.1. *Schematic of the micro-centrifuge. Three peristaltic pumps (1 & 2) are used to feed and wash the micro-centrifuge bowl (4). A solenoid valve (5) controls the supply of a vacuum to the micro-centrifuge bowl to allow excess feed and wash to be sucked to waste. The vacuum is generated by a vacuum pump (6) driven by tap water.*

The general operating procedure of the micro-centrifuge is summarised below

- Approximately 5-6 mL of sample was pumped into the micro-centrifuge bowl, whilst the vacuum was switched on resulting in the filling and subsequent removal of sample. This was done to pre-wash the bowl with sample. The vacuum was switched off and an extra 1 mL of sample was pumped into the bowl for centrifugation.
- The compressed air valves were switched on and the brake switched off for 10-15 seconds, which resulted in the bowl spinning at a high speed (compressed air at 4 bar(g) resulted in the bowl spinning at 55,000 rpm

(Chard 1997)). The vacuum was switched on for 0.8 seconds, 1.2 seconds before the compressed air was turned off to remove drops on the feed inlet tube tip and the surplus feed material in the vacuum tip. This helped avoid contamination of the supernatant liquid (Turner 1993, Chard 1997). The brake was switched on and the centrifuge came to a stop.

- The supernatant liquid was pumped out of the bowl and the wash sequence initiated. This involved switching the wash pump, vacuum and spinning the bowl in discrete pulses, i.e. pump on for 1 second, spin for 1 second, vacuum on for 1-2 sec. This sequence was repeated 5-6 times.

The total run time to sample and perform solid-liquid separation step was approximately 60 s. The wash cycle was performed during the sample dilution, reagent mixing and product and contaminant detection step.

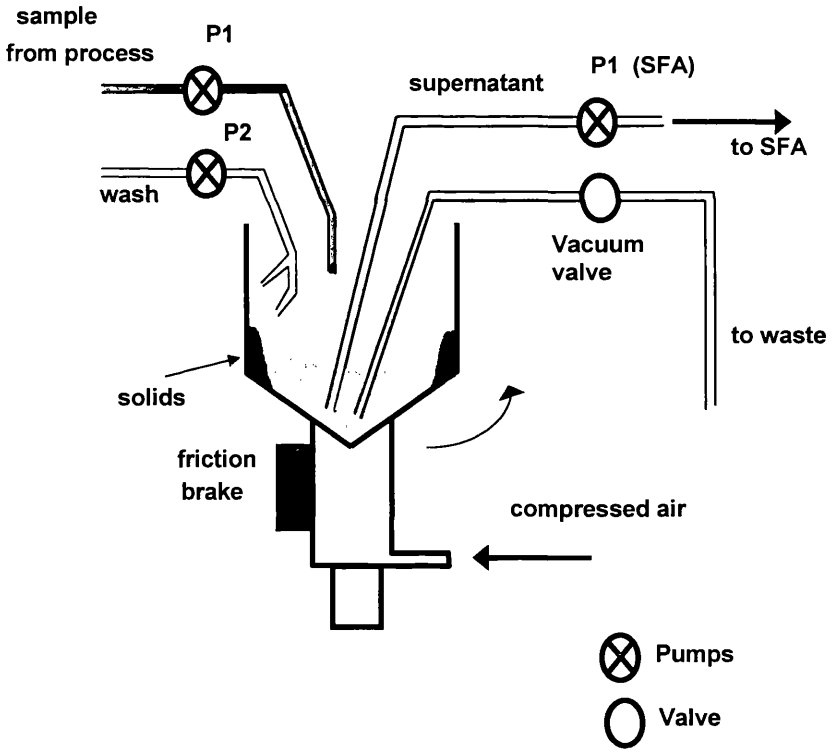


Figure 3.2. *Schematic of the micro-centrifuge bowl. P1 and P2 refer to peristaltic pumps 1 and 2 and P1 (SFA) refers to the peristaltic pump situated on the stopped flow analyser for drawing off supernatant liquid.*

3.3.2 Stopped flow analyser

The stopped flow analyser (SFA) was built in-house. It consisted of two 8 roller peristaltic pumps (308 D/A, Watson-Marlow Ltd), one 3 roller peristaltic pump, a 2-way solenoid valve (Bio-Chem Valve, model 100P3MP12, UK) and a 3-way solenoid valve (Bio-Chem Valve, model 080T312, UK), housed in a steel case with dimensions 400 x 310 x 180mm [L, W, D], illustrated in figure 3.3. The design of the SFA was based on the one used by Chard (Chard 1997).

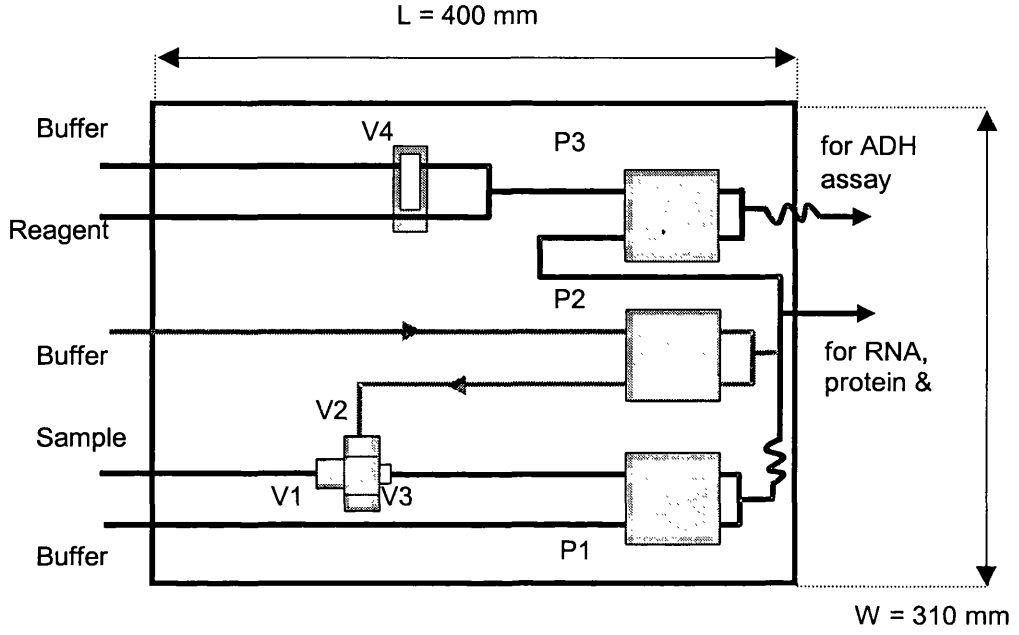


Figure 3.3. *Schematic of the stopped flow analyser (SFA). Two 8 roller peristaltic pumps (P1, P3) for sample dilution and reagent mixing. The two-way valve (V4) allowed the switching from reagent to buffer when the wash sequence was initiated. The three-way valve (V1, V2, V3) was implemented to switch from supernatant sample to buffer to enable the washing of the tubing and flow-cells.*

The two 8-roller peristaltic pumps were used for diluting and mixing sample with buffer and reagent to appropriate levels before transporting the prepared samples to the appropriate destinations (ADH-spectrophotometer and diode-array spectrophotometer). The 3-roller pump was used during the wash cycle. The 2-way solenoid valve enabled the user to switch from reagent mix to wash buffer during the cleaning step. The 3-way solenoid valve was also used during the wash cycle. Figure 3.4 and 3.5, illustrate the sample preparation sequence and wash cycle for both flocculation and chromatography monitoring. The total run time from sampling through to at-line product and contaminant measurement was 135 s for the flocculation process and approximately 60 s for the chromatography process. The later did not require a solid-liquid separation step.

The dilution levels were controlled by choosing different tube diameters for sample and buffer. The sequencing of pump and valve actions were determined by knowing the flow-rate, the length and diameter of the tubes. The peristaltic pumps and solenoid valves were controlled by a supervisory computer through 7 digital signals (0-5V) (National Instruments, AT-DIO-32F, 32-bit digital I/O interface).

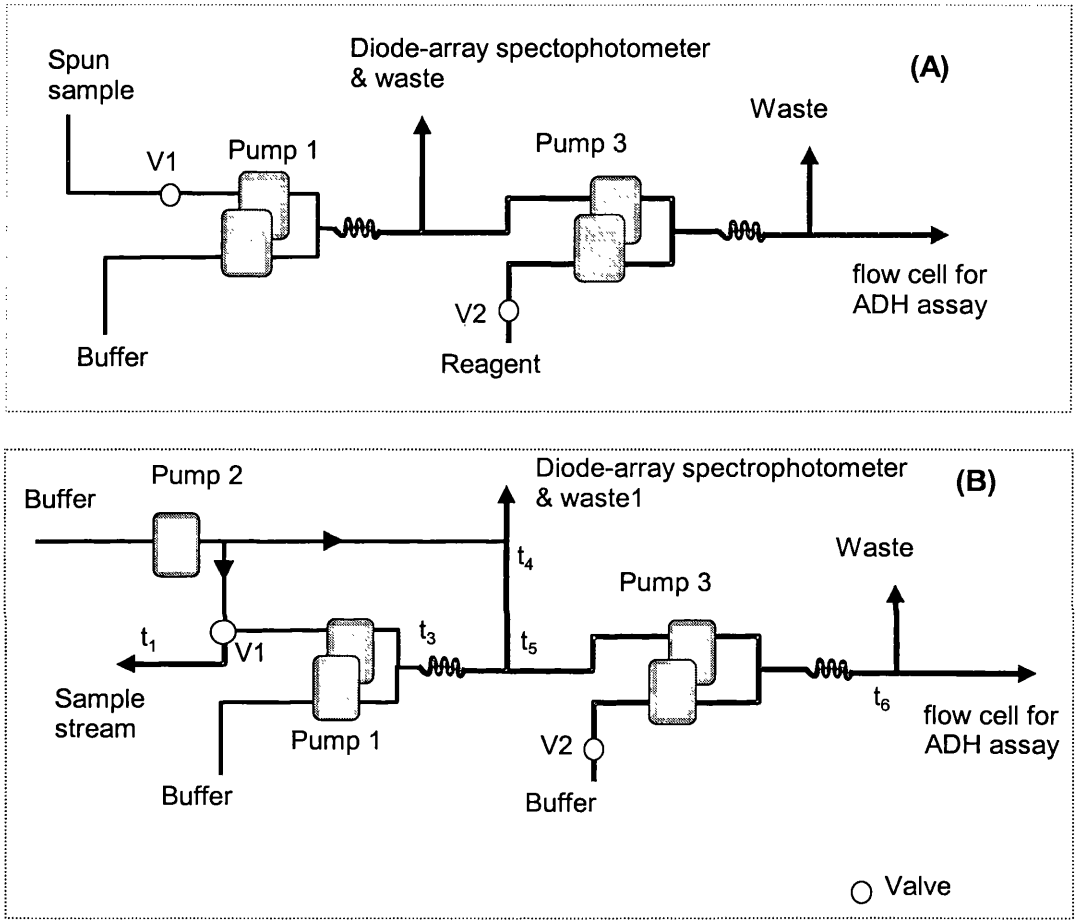


Figure 3.4*. Stopped flow analyser setup for the flocculation process. The top diagram (A) illustrates the sample dilution and reagent mixing sequence. The sample was diluted with buffer after pump 1 and a T-piece splits the stream. Diluted sample is sent to the diode-array spectrophotometer for spectral scanning. Pump 3 was then switched on and the diluted sample was mixed with reagent and transported to the spectrophotometer for ADH assaying.*

The bottom diagram (B) illustrates the wash cycle. Buffer was pumped (pump 2) initially to the diode-array spectrophotometer to wash the flow-cell by keeping valve V1 shut. Pump 1,2 & 3 and valve V1 & V2 were then switched to wash the ADH flow-cell as well as tubing t_1, t_3, t_4, t_5 and t_6. The above wash cycle is altered slightly during the monitoring of the chromatography work, whereby no buffer is washed back through the sample tube t_1.

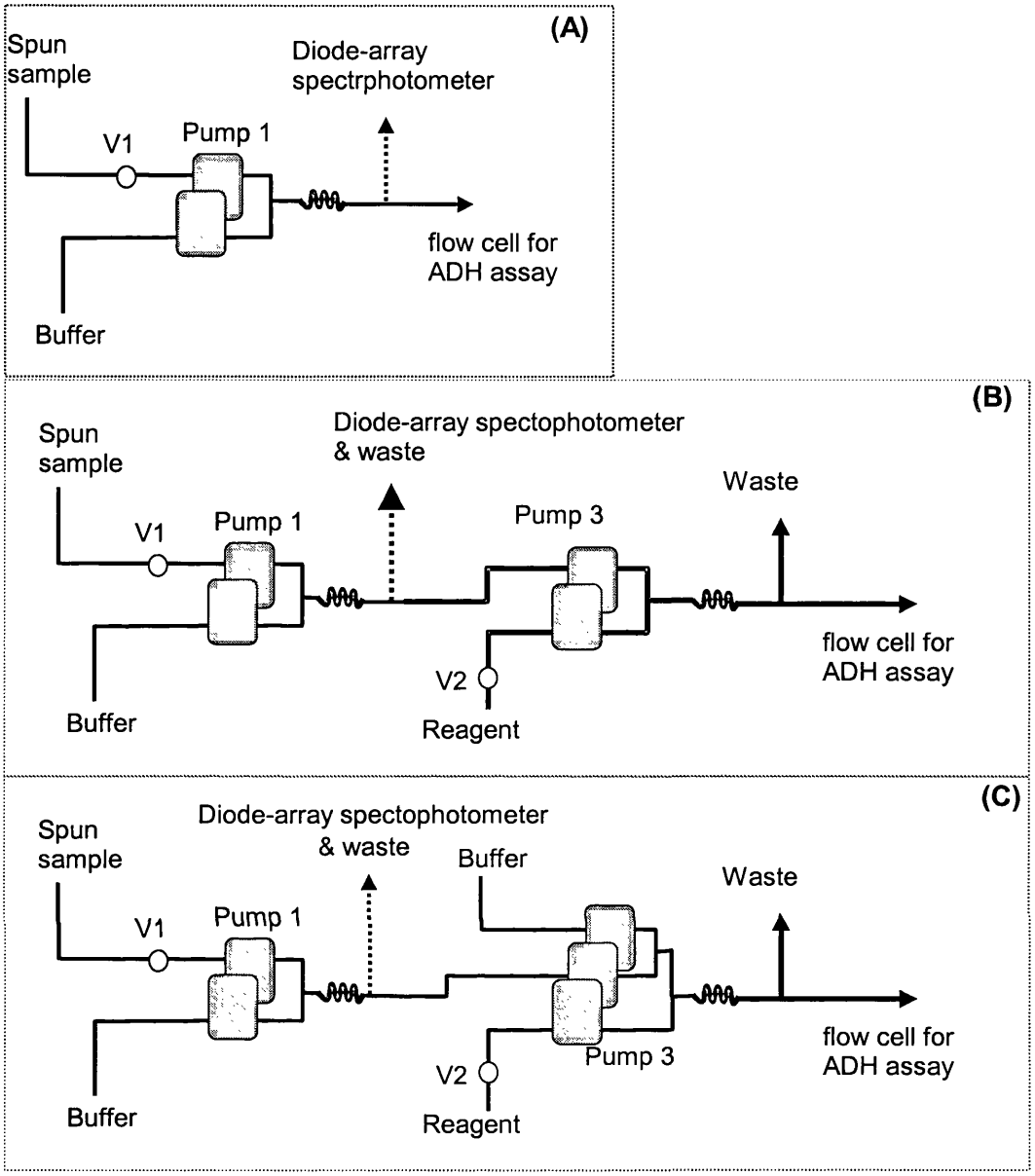


***Figure 3.5**. Stopped flow analyser setup for the chromatography process. Diagrams (A), (B) and (C) illustrate the three SFA configurations to cope with different sample dilution levels during the chromatographic process. The dilution levels for configuration (A) was 1:17, (B) 1:289, and (C) 1:2312 (expanded bed process) or 1:4913 (packed bed process). The wash cycle during chromatographic monitoring was similar to that during of the flocculation setup, however no buffer is washed back through the sample tube (see figure 3.4 (B)).*

3.4 At/on-line monitoring Instruments

3.4.1 At-line spectrophotometer for ADH measurements

A spectrophotometer (Ultrospec 2000, Pharmacia-Biotech, Upsala, Sweden) was used in conjunction with a flow-cell (1mm pathlength, Hellma, Essex). The

spectrophotometer was interfaced with the supervisory computer through a RS232 serial link. Prepared sample from the SFA was transported to and from the spectrophotometer. The flow-cell was washed with buffer between samples during a run whilst 0.5 M NaOH was used between batch runs.

3.4.2 Photometric dispersion analyser

A photometric dispersion analyser (PDA) (PDA 2000, Rank Brothers Ltd.) was used to monitor floc levels in a flowing suspension. In the PDA the flowing suspension was illuminated by a narrow beam of light, from a high intensity light emitting diode (wavelength of 820nm), so that a fairly small sample volume was examined. Since the suspension was flowing, the number of particles in the light beam was continuously changing because of local variations in composition, and these variations caused fluctuations in the intensity of transmitted light. The light intensity was monitored by a sensitive photodiode, the output of which was converted to a voltage proportional to the intensity. The output voltage had a large direct current (dc) component related to the turbidity of the suspension, and a smaller fluctuating component alternating current (ac), due to the random variations in particle number. Both the dc and the ac values were sent to the supervisory computer through an analogue board (ComputerBoards, CIO-DAS16/Jr). The sample suspension was fed through a transparent plastic tubing, which fits into a perspex block that houses two precisely aligned fibre-optic probes, on the PDA. The PDA was only used in the dc mode as the concentration of solids in the monitored homogenate did not allow for a stable ac signal.

3.4.3 At-line diode-array spectrophotometer for contaminant measurements.

The diode-array spectrophotometer was built in-house. It consisted of a diode–array spectrometer and fiber optic probe (Carl Zeiss, MMS polychromator, Jena, Germany), a deuterium light source, the appropriate lenses to focus the light on to the optical fiber input and a fan to prevent over heating, housed in a steel case with dimensions of 400, 310, 180mm, [L, W, D], shown in figure 3.6. The photodiode array was made by Hamamatsu, with a wavelength range between 215 - 740 nm, and a pixel number of 256. The wavelength resolution was 2.2nm/pixel and wavelength accuracy was below 0.5 nm. The sample to the diode-array spectrophotometer was supplied to a 1mm pathlength flow-cell (Hellma). Twenty scans from 220 to 500nm were taken to give an average spectral scan of the sample every 1.2 s. The spectrophotometer was interfaced with the supervisory computer via an analogue board (ComputerBoards, CIO-DAS16(Jr.), UK) for data logging. An equivalent wash cycle of the flowcell to that

described in section 3.4.1 was adopted. The spectral scans were used to predict total protein, RNA and cell debris levels, through a multivariate statistical technique, discussed in section 2.1 and section 4.3.

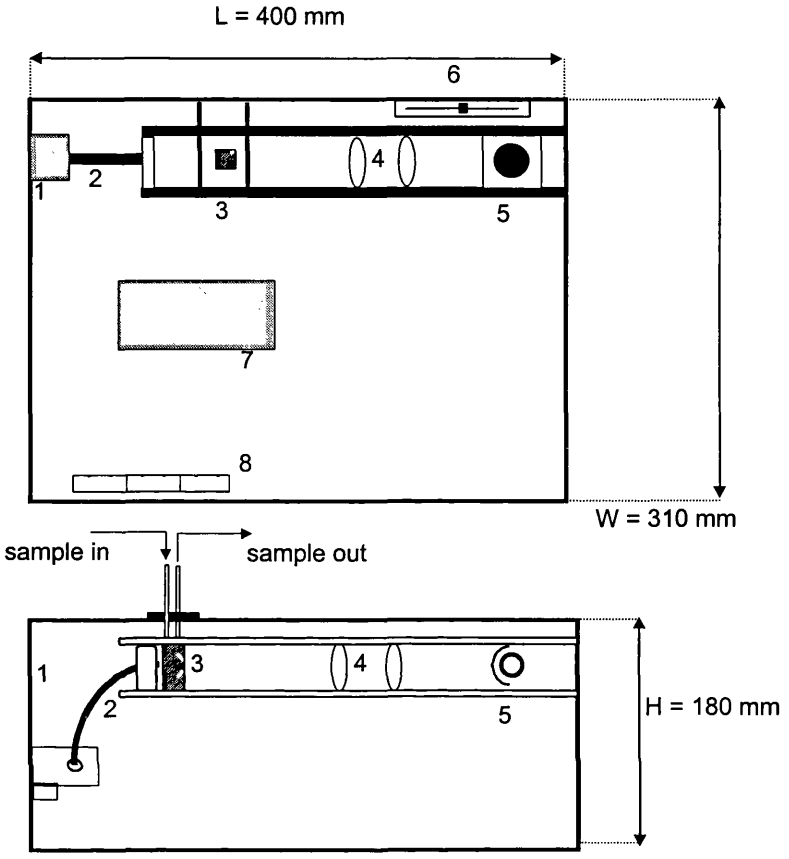


***Figure 3.6**. Schematic of the diode-array spectrophotometer. Top and bottom diagrams illustrate the top and side views of the custom built diode-array spectrophotometer, consisting of a Carl Zeiss MMS polychromator spectrometer (1). Spectral scans where performed on samples pumped into the flowcell (3). Lenses (4) were used to focus the light through the flowcell on to the tip of the optical fibre input (2). Fan (6) and air extractor (8) were used to prevent overheating. A power source (7) was applied to run the lamp.*

3.5 Batch flocculation process

3.5.1 Preparation of feed and PEI solution

Suspended Baker's yeast (125 g L^{-1} (wet weight) 3-4 L) was disrupted according to the procedure discussed in section 3.1.2. PEI flocculant solutions were made at 5 and 10 mg mL^{-1} concentrations. The yeast homogenate and PEI solutions were adjusted to the desired pH (titration with 10M NaOH or HCl) and ionic strength (KH_2PO_4).

3.5.2 Experimental apparatus

The batch flocculation process with the at-line monitoring setup is illustrated in figure 3.7, below. A baffled 1000 mL (750 mL working volume) vessel was filled with feed yeast homogenate and PEI flocculant was dosed into the tank using a variable gear pump (MV-ZP-15, Ismatec, Zurich, Switzerland). The dose flowrate was monitored by an inductive magnetic flowmeter (model DM4, Turbo, Kohn, Germany) and controlled by a Turnbull Control System (TCS) (model 6366, process controller, Worthing, UK), which applied a standard PID algorithm. The TCS was interfaced to a supervisory computer (IBM compatible, Pentium, 133 MHz) via a RS232 serial link. Changes to pump setpoints could be executed either by an operator through the supervisory computer for open loop control or automatically by a control algorithm for closed loop control. The homogenate-PEI mixture was well mixed by a variable speed drill unit (Heidolph, Gemany, 90 - 1800 RPM), operating at 1000 RPM with a six-flat bladed-disk turbine. A more detailed illustration of the bath flocculation tank is shown in figure 3.8, indicating the relevant dimensions.

3.5.3 Sampling and at-line monitoring

A sample was continuously drawn off (25 mL min^{-1}) by a 10-roller peristaltic pump (Minipuls 3, Gilson M312, UK) and recycled back into the batch tank. PDA readings were automatically monitored and logged by the supervisory computer, which would also trigger the at-line sampling and monitoring sequence. A sample (5.6 mL) was drawn off the sampling loop and assayed for product and key contaminants according to the method discussed in section 3.3 and 3.4. At-line measurements were recorded approximately every 135 s and were associated with a 5 s time delay, due to the time taken for the sample to travel from the batch tank to the sampling point.

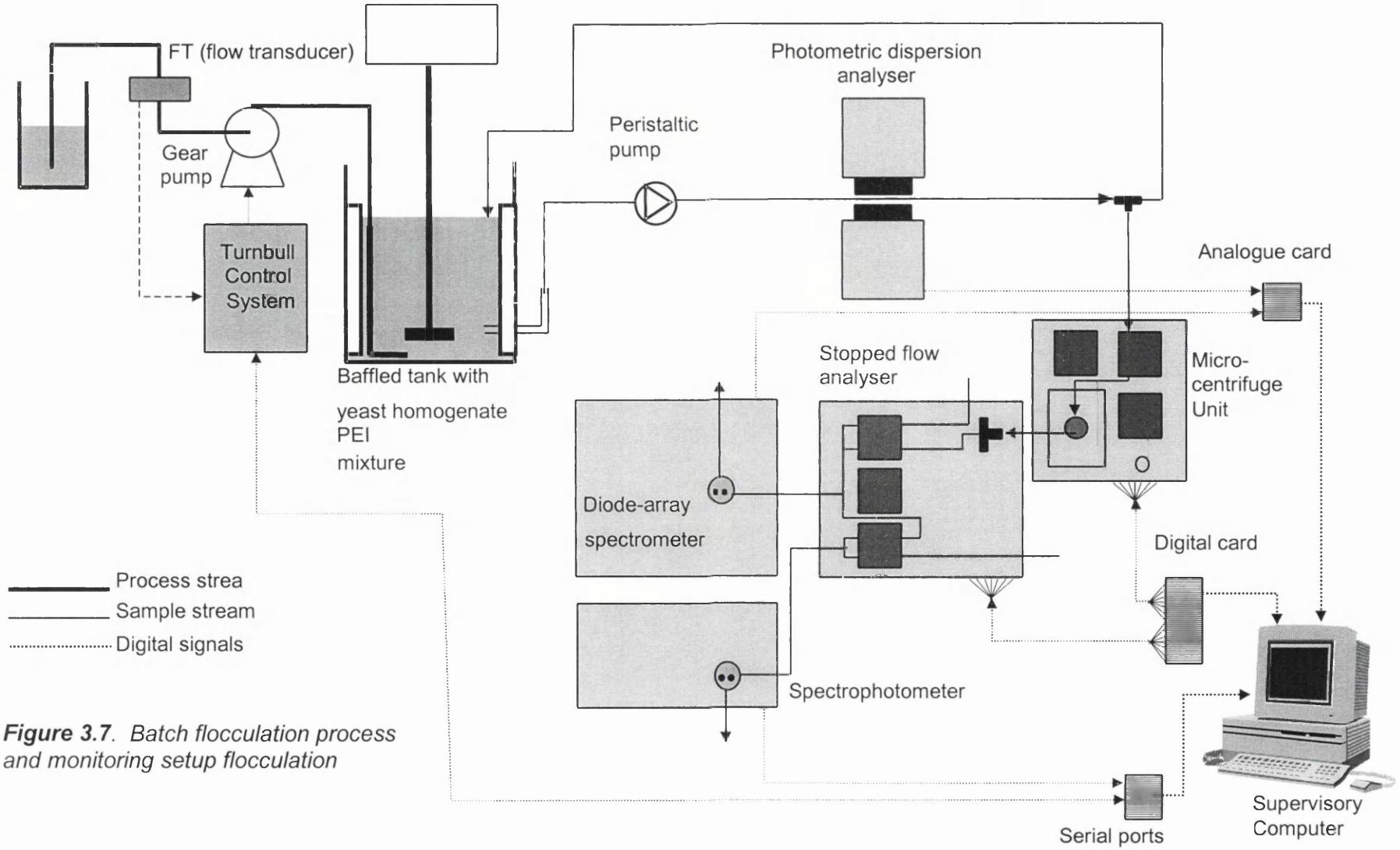


***Figure 3.7**. Batch flocculation process and monitoring setup flocculation*

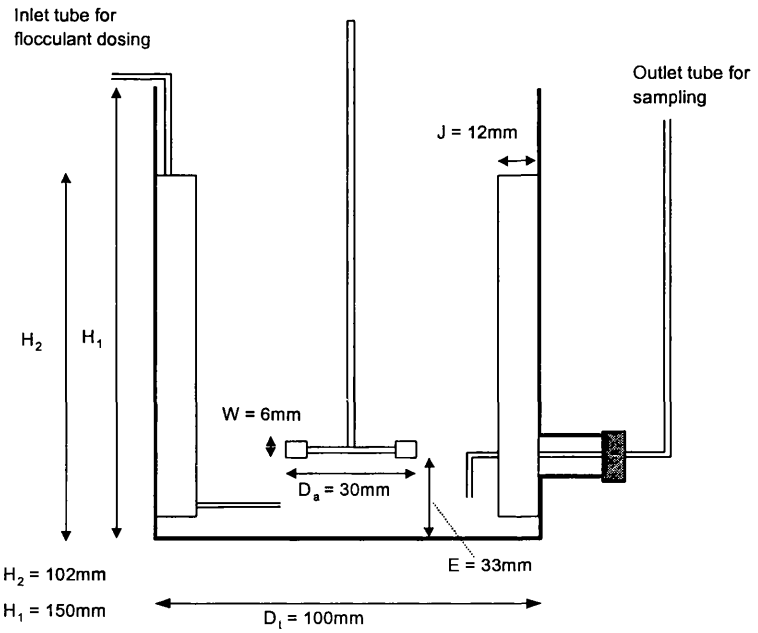


Figure 3.8. *Batch flocculation tank dimensions. The inlet flocculant tube was placed such that PEI would be dosed below the impeller.*

3.6 Continuous flocculation process

3.6.1 Preparation of homogenate and flocculant solutions

A similar procedure to that applied during the batch flocculation process was employed.

3.6.2 Experimental apparatus

The flocculation equipment consisted of two variable speed gear pumps (MV-ZP-15, Ismatec, Zurich, Switzerland), feeding the homogenate and PEI streams into a Y-piece followed by a needle valve (model 11562, Hoke, Harrow, UK) used to ensure good mixing was adopted. Both pumps were controlled by feedback loops with flow rates monitored by inductive magnetic flowmeters (model DM4, Turbo, Kohn, Germany). The flow rate signals were fed back to a Turnbull Control System (TCS) (model 6366, process controller, Worthing, UK), which applied a standard PID algorithm to produce the relevant output signal to control pump speeds at the required setpoints. The TCS was interfaced to a supervisory computer (IBM compatible, Pentium, 133 MHz) via a RS232 serial link. Changes to pump setpoints could be executed either by an operator through the supervisory computer for open loop control or automatically by a control algorithm for closed loop control. A specific PEI volume fraction was achieved by controlling the PEI to homogenate flowrate ratio. The total flowrate was kept constant at a flowrate of 22.6 mL min^{-1}

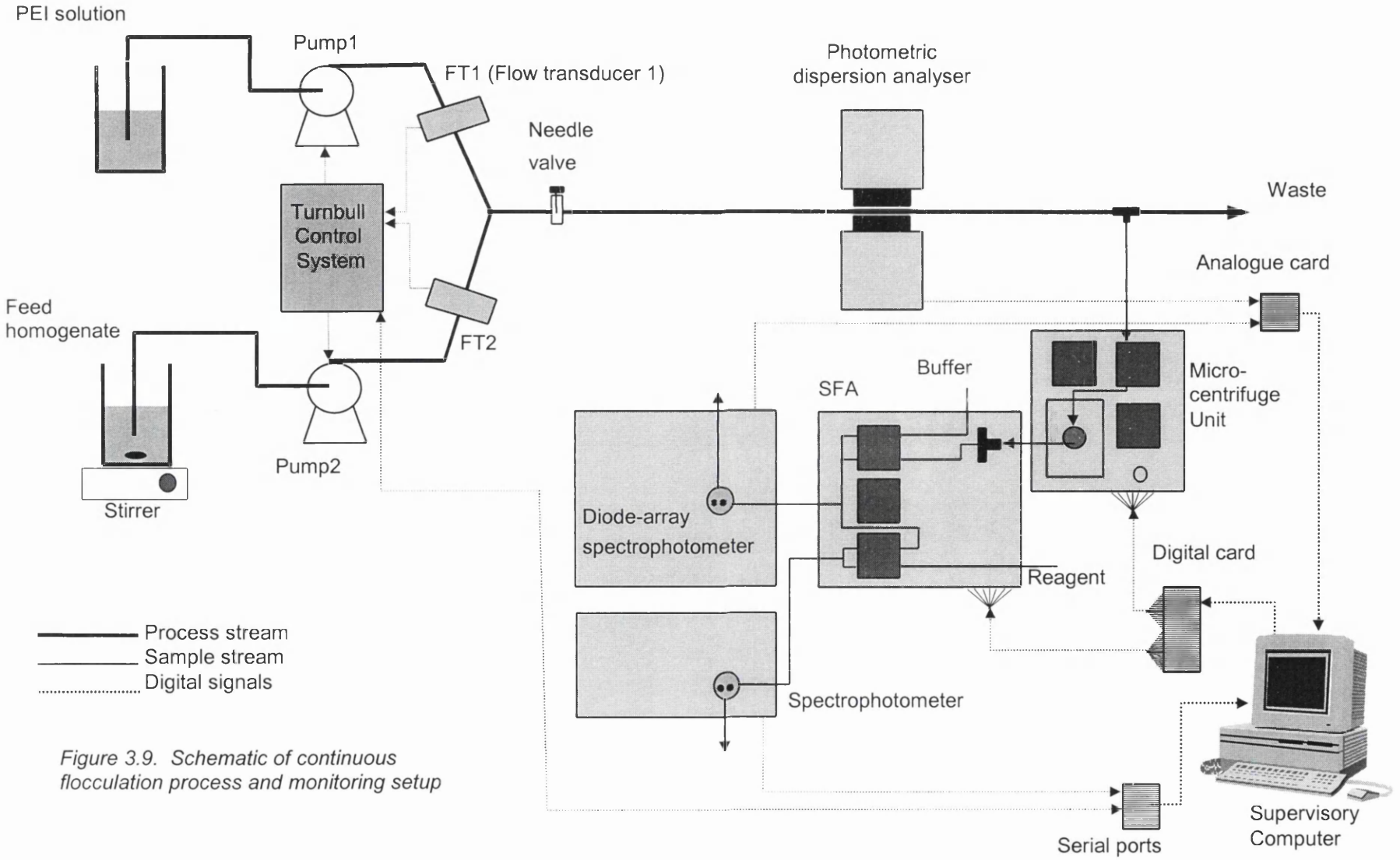


Figure 3.9. Schematic of continuous flocculation process and monitoring setup

3.6.3 Sampling and at-line monitoring

The PEI-homogenate mixture would be pumped to the photometric dispersion analyser (PDA) where continuous PDA dc readings were acquired automatically and logged by the supervisory computer. The at-line sampling set-up (discussed in section 3.3 and 3.4) would be actuated by the supervisory computer. The sampling sequence would automatically draw off approximately 5.6 mL of sample and feed it to the micro-centrifuge followed by the SFA and finally to both spectrophotometers for product and contaminant measurements. A 10 s time delay was associated with the at-line data, due to the time taken for the PEI-homogenate mixture to flow from the Y-piece to the sampling point.

3.7 Packed bed HIC chromatography

3.7.1 Experimental apparatus

Bench scale

The experimental and at-line monitoring set-up for the packed bed column is shown in figure 3.10. The bench scale system under study was a 0.05 m diameter XK50/40 packed bed (Pharmacia Biotech AB, Upsalla, Sweden). A peristaltic pump (Model 505DU Watson Marlow, Cornwall, UK) was used to pump different liquids onto the column during load, wash and elution stages. A valve configuration (Model 4port-2way and 4port-4way, Pharmacia Biotech) was chosen to enable the reversal of flow through the column.

The outlet of the column was connected to a UV monitor (Model UV1, Pharmacia Biotech AB, Sweden) set to 280 nm. A fraction collector (Model SuperFrac, Pharmacia Biotech AB, Sweden) was used to pool eluate from the column. An at-line monitoring system described in section 3.4 and 3.5 was adopted to draw off approximately 2 mL min^{-1} of sample from the outlet stream of the column for rapid measurements.

Pilot scale

The experimental setup for the large scale packed bed column is shown in figure 3.10. A 0.2 m diameter Bioprocess Glass Column BPG200/500 (Pharmacia Biotech, AB, Sweden) was implemented. Two stainless steel tanks with top mounted agitators were connected to a T-piece upstream to a 4 port-2 way valve (V1) (Pharmacia Biotech AB, Sweden). A 120 L stainless steel mobile buffer tank was also connected to the T-piece used for holding clarified supernatant during the loading onto the column. A peristaltic

pump (Model 505DI, Watson Marlow, UK) was used to pump liquid on to the column. The valve configuration adopted was similar to that applied for the bench scale setup.

A similar UV-monitor and at-line sampling setup to that adopted for the bench scale system was applied.

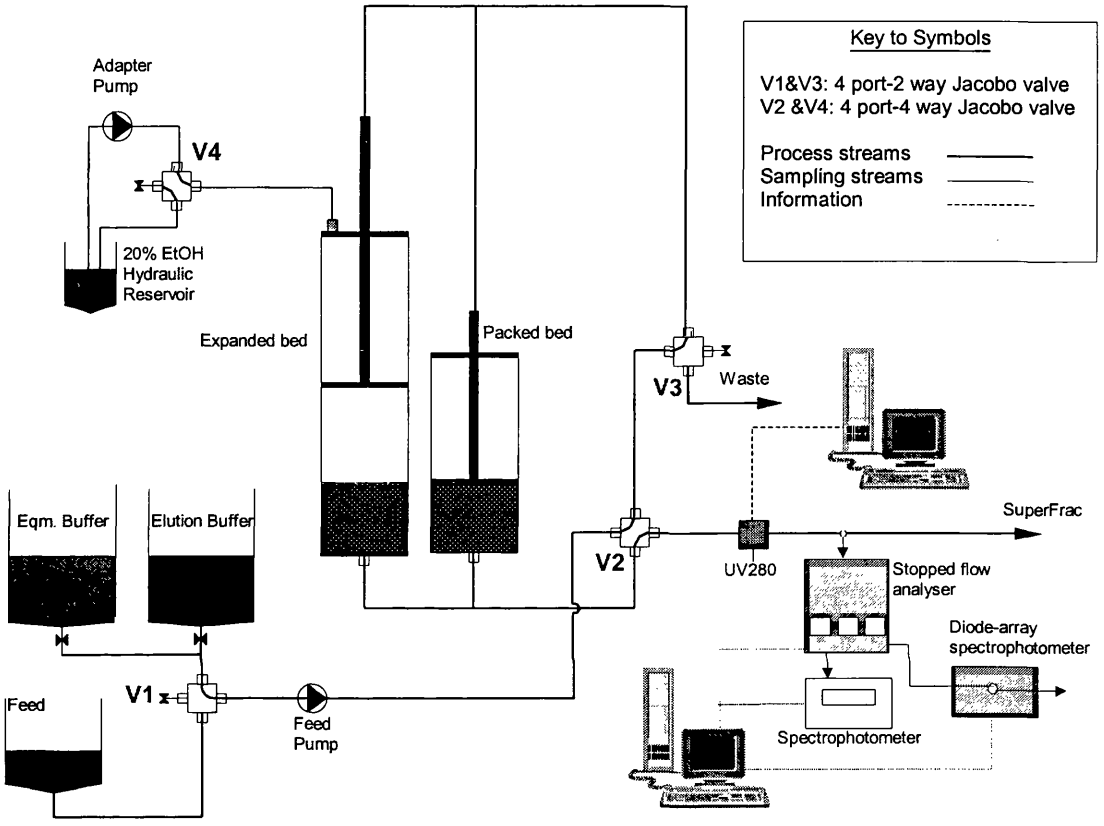


Figure 3.10. *Schematic layout of the packed bed and expanded bed equipment with ancillary equipment and at-line monitoring system. The packed bed column represents both the bench and large scale systems. The above schematic is not to scale.*

3.7.2 Preparation of cell homogenate

Bench scale

Baker's yeast (280 g (wet weight) L^{-1}, 4L) in buffer (0.1M KH_2PO_4, pH 6.5) was disrupted in a high pressure homogeniser (Model Lab 60, APV Crawley, UK) for 5 discrete passes at 500 bar(g) with the temperature maintained at 4°C.

Pilot scale

Baker's yeast (280 g (wet weight) L^{-1}, 60L) in buffer (0.1M KH_2PO_4, pH 6.5) was disrupted in a high pressure homogeniser (Model K3, APV Crawley, UK) for 5 discrete passes at 500 bar(g) with the temperature maintained at 4°C

3.7.3 Bench scale batch PEI flocculation and centrifugal clarification.

PEI (2% (w/v), pH 6.5, 0.75L) dissolved in buffer (0.1M KH_2PO_4, pH 6.5) was added to the 3L of yeast homogenate (PEI volume fraction of 0.2 (v/v)) and stirred by a magnetic flea for 5 minutes in a 4L beaker. The PEI-homogenate mixture (3L) was centrifuged in a Beckman centrifuge (Model J2-M1, Spinco, Beckman Instruments, Pan Alto, California, USA) with a fixed angle rotor (type JA-17) at 10,000 rpm for 10 minutes. The supernatant was diluted with ammonium sulphate (100% saturated solution in 0.1 M KH_2PO_4, pH 6.5) to adjust the salt concentration to 0.78M and further diluted to approximately 10 mg (protein) mL^{-1}, for subsequent packed bed chromatographic purification.

3.7.4 Large scale continuous PEI flocculation and centrifugal separation

Two lobepumps (Monopumps, UK) were used to continuously pump a 1% (w/v) PEI solution (pH 6.5, 0.1M KH_2PO_4) and the yeast homogenate to a T-piece junction to insure good mixing. A PEI volume fraction of 0.3 (v/v) was chosen after a lab-scale test indicated that this PEI dose insured good clarification. The PEI-homogenate mixture was continuously pumped (30 Lh^{-1}) to an intermittent discharge disk-stack centrifuge (Model CSA-1, Westfalia Separator AG, Oelde, Germany) equipped with a hydro-hermetic (soft shear) feed zone. Full discharge was performed every 150 s. The supernatant from the CSA-1 was further clarified by a tubular bowl centrifuge (Model 1P, Sharples, Camberley, Surrey, UK) using peristaltic pump (Model 502S, Watson Marlow) at a flowrate of 30 L h^{-1}. The centrifuge was operated at a bowl speed of 45,000 rpm. Ammonium sulphate (100% saturated solution in 0.1 M KH_2PO_4, pH 6.5) was added to the 1P supernatant pool to adjust the salt concentration to 0.78M and no further dilution of the pool was required.

3.7.5 Column chromatography

Bench Scale

An XK50/40 was packed with Phenyl Sepharose FF matrix (low sub) to a final height of 0.125m. The column was equilibrated with 10 column volumes of *buffer A* (0.78 M $(NH_4)_2SO_4$ in 0.02M KH_2PO_4, pH 7) at a superficial velocity of 2 m h^{-1} (65.3 mL min^{-1}). Prepared supernatant was loaded on to the column at the same flowrate and in the same direction as the equilibration buffer. In the wash cycle where unbound material was washed from the column, flow to the column was switched from supernatant to *buffer A* still in the same direction. The wash cycle was terminated after approximately 7 column volumes of buffer. Elution of the bound material was performed at 0.75 m h^{-1}

in the same direction of the load and wash, with a step decrease in salt concentration using *buffer B* (0.02 M, KH_2PO_4, pH7).

An automated clean in process (CIP) was employed between chromatographic cycles in order to regenerate the matrix. An FPLC liquid chromatography controller (Model LCC500 Plus, Pharmacia Biotech AB) was used to control two motorised valves (Model IMV) and a peristaltic pump (Model P-6000). The CIP protocol was the following:

- 1M NaOH (2.4 L, 0.3 m h^{-1})
- DI-H_2O (0.9L, 0.1 m h^{-1})
- 30% (v/v) Isopropyl alcohol (0.9 L, 0.1 m h^{-1})
- 25% (v/v) Acetic acid (0.9L, 0.1 m h^{-1})
- DI-H_2O (1.5, 0.1 m h^{-1})
- 20 % (v/v) ethanol (1.5L, 0.1 m h^{-1})

Pilot Scale

The BPG200/500 0.2 m diameter column was packed with approximately 393 mL of settled Phenyl Sepharose FF (low sub) matrix to a final packed bed height of 0.129 m. After equilibration of the column with *buffer A*, clarified supernatant was loaded at a liquid velocity of 2 m h^{-1} (1047 mL min^{-1}) until the breakthrough of ADH reached approximately 5%. Unbound material was washed from the column in the reverse direction to loading with approximately 7.5 column volumes of buffer A. Elution of the bound material was performed at 0.75 m h^{-1} (393 mL min^{-1}) in the same direction of the load and wash, with a step decrease in salt concentration using *buffer B* (0.02 M, KH_2PO_4, pH 7). The column was regenerated with 10 column volumes of 1M NaOH, which was washed from the column with DI-water before storage in ethanol (20% (v/v)).

3.7.6 Monitoring of the chromatographic process

During the load, wash and elution cycle fractions were collected every 120 s and the concentration of ADH of each fraction assayed immediately on collection. Protein levels were assayed subsequent to the final chromatographic cycle. At-line ADH levels were also monitored adopting the SFA system described in section 3.3.2. At-line measurements of ADH were taken approximately every 60 s. Protein levels of the pooled fractions were also assayed adopting the SFA and rapid scanning system (section 3.4.3).

3.8 Expanded bed chromatography

3.8.1 Experimental apparatus

The experimental and at-line monitoring setup for the 0.05 m diameter STREAMLINE (ST-50) expanded bed (Pharmacia Biotech) was similar to that of the 0.05 m diameter packed bed system and is shown in figure 3.10. An additional peristaltic pump (Model 505DU Watson Marlow) was used to transfer ethanol (20% (v/v)) from a separate reservoir to the head space above the top adapter, enabling the position of the upper adapter to be adjusted during chromatographic operation. The valve configuration to the bench scale packed bed system enabled the reversal of flow through the expanded bed.

3.8.2 Preparation of column feed

Baker's yeast (280 g (wet weight), L^{-1}, 4L) was disrupted in a high pressure homogeniser as described in section 3.1.1). The homogenate was diluted with ammonium sulphate (100% saturated solution in 0.1 M KH_2PO_4, pH 6.5) to adjust the salt concentration to 0.78M and further diluted to approximately 10 mg (protein) mL^{-1}, for subsequent chromatographic purification.

3.8.3 Column chromatography

The settled bed height of STREAMLINE-Phenyl (low sub) matrix was 0.148 m. The bed was expanded from its initial settled configuration by increasing the superficial liquid velocity of DI-water upwards through the column, until a maximum liquid velocity of 3 m h^{-1} was achieved. The bed was then equilibrated with *buffer A* at a liquid velocity of 2 m h^{-1} (65.3 mL min^{-1}) until the bed height had stabilised (approximately 60 min). The prepared homogenate feed was then loaded on to the bed until an outlet ADH concentration equivalent to a pre-specified breakthrough was observed on the at-line monitoring equipment. In the wash cycle unbound material was washed from the column with *buffer A* in the same direction and superficial velocity of the load i.e. in expanded mode. The wash cycle was terminated when the on-line UV trace returned to its initial base line. The bed was allowed to settle and the upper adapter was positioned on the top of the settled bed at a height of 0.148 m for elution.

Elution was performed using a step decrease in salt concentration by applying *buffer B* at a liquid superficial velocity of 0.75 m h^{-1} (24.5 mL min^{-1}) in the opposite direction to the load and wash. The elution cycle was terminated after the at-line ADH profile had

peaked and returned to its baseline. The system was cleaned using an automatic CIP cycle equivalent to that adopted for the bench scale packed bed system.

3.8.4 Monitoring of the chromatographic process

During the load, wash and elution cycles off-line and on-line assays of ADH and total protein levels were performed according to the method described in section 3.7.6. RNA and cell debris levels of pooled fractions were also measured using the SFA and the UV-VIS diode-array spectrophotometric system described in section 3.4.2.

3.9 Computer software and hardware

3.9.1 Data acquisition

The software implemented for data acquisition was LabVIEW (National Instruments, USA) and was run on an IBM compatible PC (Hewlett Packard, Pentium 133 MHz, USA). Table 1 illustrates the different communication links between the supervisory computer and the various instruments.

Table 3.1: Communication links between computer and instruments

Micro-centrifuge	Digital signals out
SFA	Digital signals out
Diode-array spectrophotometer	Analogue signals in
Pharmacia Biotech Spectrophotometer	Serial link (RS232), two way communication
PDA	Analogue signal in
TCS	Serial link (RS232), two way communication

3.9.2 Multivariate analysis

The multivariate analysis of spectral data was carried out using a partial least squares statistical technique available in a commercial software called *Charm Works V 1.1* (Process Analysis & Automation Ltd., UK).

Chapter 4

Demonstration of at-line techniques

4.1 Summary

The aim of this chapter is to demonstrate the feasibility of at-line monitoring of alcohol dehydrogenase (ADH) product and key contaminants such residual cell debris, protein and RNA for a flocculation process. The at-line monitoring setup consisted of an automated microcentrifuge for solid-liquid separation, a stopped flow analyser (SFA) for automatic sample preparation and transportation and two spectrophotometers (diode-array and fixed wavelength) for product and contaminant detection. Product concentrations were measured through an enzyme catalysed reaction whilst the contaminant levels were predicted from UV-VIS spectral data.

4.2 Introduction

Cell debris dry weight levels will be correlated to optical density readings to enable the rapid monitoring of cell debris residue. This chapter shows results from performance of the at-line monitoring of ADH described in section 3.2. UV-VIS spectral analysis implementing the partial least squares (PLS) multivariate analysis technique was examined for the rapid at-line prediction of key contaminants such protein, RNA and cell debris levels. The performance of a flocculation process in terms of photometric dispersion analyser (PDA) readings (described in chapter 3) was also investigated. An overall error analysis was performed on the at-line monitoring setup consisting of automated microcentrifugation, stopped flow analysis and product and contaminant detection to establish the measurement errors involved with the at-line system.

4.3 Results and discussion

4.3.1 At-line monitoring of cell debris residue using optical density at 650nm

Figure 4.1 on page 99 illustrates a linear response with high coefficients of determination between wet weight and dry weight for Baker's yeast homogenate against optical density values at a wavelength of 650nm (OD_{650}). The linear correlations are shown in figure 4.1 and may be implemented to convert OD_{650} values to wet / dry weight levels. If yield values are used no conversion is required due to this linear response. **Figure 4.2**, shows dry weight against wet weight values derived from figure 4.1. As expected a linear response is achieved with a correlation factor of 0.095 from wet to dry weight. These correlations are used to estimate the solids concentration of process streams although it must be recognised that deviations may be expected, for example where the ratio of dissolved coloured compounds to solids is changed significantly. In the text below cell debris concentrations will be expressed in OD_{650} units, i.e. absorption units (Au).

4.3.2 At-line monitoring of ADH

Figure 4.3 Plot A, compares at-line with off-line measurements of ADH activity for pure ADH samples. **Plot B** shows at-line ADH measurements for homogenised Baker's yeast against wet weight concentrations. A linear response between off-line and at-line measurements (Plot A) and at-line ADH measurements and yeast homogenate wet weight concentrations (Plot B) was found with coefficients of determination (R^2) for both relationships above 0.99. Prior to at-line assaying sample dilution and reagent mixing was performed by a stopped flow analyser. Figure 4.3 illustrates that at-line ADH measurements can be taken of both clear and turbid samples automatically, reproducibly with a reduced assay time of 10 seconds without the loss of accuracy. This at-line technique is used to rapidly monitor ADH activity of process streams.

4.3.3 Rapid contaminant monitoring using UV-VIS spectral scans and PLS

Partial least squares (PLS) a multivariate analysis technique was adopted to calibrate the Zeiss diode array spectrophotometer for RNA, protein and cell debris (OD_{650}) prediction using spectral scans in the ultraviolet and visible (UV-VIS) region (230 nm to 500 nm). The theoretical aspects of multivariate analysis in particular principal component analysis (PCA) and PLS were discussed in section 2.1.

Figure 4.4 illustrates the effects of varying protein (BSA), RNA and cell debris levels (OD_{650}) in phosphate buffer (100 mM, pH 6.5) on the UV-VIS spectra between 230-500 nm. The spectral scans were taken after diluting samples by a factor of 16 with phosphate buffer (100mM, pH 6.5) and introduced to the diode array spectrophotometer via a 1 mm pathlength flowcell manually. **Plot A** illustrates the spectral scan due to an increase in cell debris resulting in OD_{650} levels of 0 Au, 1.77 Au to 5.5 Au whilst keeping BSA and RNA levels at zero. An increase in cell debris levels resulted in an apparent linear upward shift in the spectral scans (in terms of absorbance), i.e. a twofold increase in OD_{650} levels resulted in an apparent twofold upward shift in the spectra. It should be noted that the spectral scans fall in an absorbance range between 0 – 0.125 Au for the increase in OD_{650} values from 0 to 5.5 Au.

Plots B and **C** demonstrate the response of the UV-VIS spectra due to increases in RNA (0 mg mL^{-1}, 2.5 mg mL^{-1} to 5 mg mL^{-1}) and BSA (0 mg mL^{-1}, 4 mg mL^{-1}, 12 mg mL^{-1} to 20 mg mL^{-1}) concentrations in phosphate buffer. Increases in both RNA and BSA levels resulted in apparent linear upward shifts in the spectral scans in the UV region 230 – 310 nm. The absorbance range associated with the RNA change was between 0 – 0.8 Au with a global and local maximum around 260 nm and 280 nm, whilst beyond a wavelength of 310 nm the absorbance values dropped to zero. Similar to the RNA response two peaks around 260 nm and 280 nm were observed for BSA variations, however the global maximum in this case was at 280 nm. The absorbance range lies between 0 – 0.05 Au, and beyond 310 nm the spectral absorbance falls to zero. The relatively high absorbance values at 260 nm and 280 nm for protein and RNA is due to primarily the presence of tyrosine and tryptophen. Many authors have applied this phenomenon for spectroscopic measurements of protein and nucleic acids (Warburg and Christian 1941, Ehresmann *et al.* 1973, Kalb and Bernlohr 1977). However such techniques up to data have not had much success due to their implementation of only two wavelength absorbance values and hence poor predictive ability. By implementing the information contained in spectral scans through multivariate analysis techniques more specific biological information can be gained.

Table 4.1 on page 103 lists the 36 samples and their compositions applied in the factorial experiment for the model calibration. Each sample composition had a volume of 3 mL consisting of 1 mL per component. The calibration range was set to enable the prediction of protein, RNA and OD_{650} levels equivalent to that found in centrifuged Baker's yeast homogenate at a wet weight concentration of 150 g L^{-1}. Maximum levels of the contaminants were therefore set at 20 mg mL^{-1} of protein, 5 mg mL^{-1} of RNA and

5.5 Au of cell debris (these are conservative estimates, i.e. above expected levels). The calibration samples were made up from standards of yeast RNA, BSA protein and washed cell debris. A 3x3x4 (OD_{650}, RNA, BSA) factorial experiment set-up was decided, resulting in 36 calibration samples. A larger number of protein samples were chosen due to the weak absorbance response compared to that of RNA in the UV region (figure 4.4). A 1:16 factor dilution was performed prior to feeding the sample to a 1mm pathlength flowcell situated in the diode-array spectrophotometer

Figure 4.5 illustrates the scores plot of the calibration scans in terms of principal components 1 and 2, derived from the first step of PLS namely principal component analysis (PCA). The numbers on the plot refer to the spectral scan taken of a particular sample number listed in table 4.1. Sample clusters (samples which are geometrically close in a scores plot are termed a cluster) have been circled (by eye), indicating that there are no obvious outliers and the calibration modelling may proceed without the need for repeats. If for example sample number 1 was situated close to the [33,34,35,36]-cluster on the scores plot one would classify sample number 1 as an outlier as it has nothing in common with the above cluster in terms of cell debris, RNA and BSA concentrations. An outlier may occur due to errors in sample make-up or spectral scanning errors. The scores plot clearly shows that there is a trend due to RNA, protein and cell debris, indicated by the arrows on the plot, i.e. just from two principal components a vast amount of information about the calibration samples can be observed, qualitatively. RNA and cell debris concentration changes result in shifts in clusters, whilst BSA variations can be seen as a progression within a cluster. The cluster formation and shifts due to contaminant variation is very dependent on the co-ordinates one chooses to plot the calibration samples in (two co-ordinates can only express cluster shifts due to two components). A scores plot of principal components 1 and 3 would illustrate a shift in clusters due to changes BSA concentration (Appendix A, Figure A.1) and not a progression within clusters. Principal components 1 and 2 express most of the information from the original spectral scan data, hence shifts in clusters due to RNA and cell debris concentration changes may be because these components absorb more readily in the UV-VIS region than BSA variations for the given calibration samples.

Figure 4.6, illustrates prediction residual error sum of squares (PRESS) levels against the number of principal components. This plot demonstrates that only 5 or 6 principal components are needed in building the calibration model to predict effectively the calibration contaminant levels through the UV-VIS spectra between 230 and 500 nm. Six principal components were chosen to build the calibration model.

Figures 4.7 and **4.8** show the quantitative accuracy of the calibration model in predicting contaminant concentrations in the standards themselves. As discussed in section 2.2, there are two main stages in using PLS for contaminant prediction. The first step is building the prediction model from the calibration set, **X** and **Y** blocks. The **X** block contains the spectral scans from each sample. The **Y** block contains the calibration standards to be predicted. PCA enables the original data (**X** block) to be decomposed and expressed in terms of a few principal components (figure 4.5). The calibration model finds a correlation between **X** (spectral scans) and **Y** (calibration samples) through the use of the first few principal components. The second stage of PLS is the prediction through the use of the calibration model.

Figures 4.7 and 4.8 illustrate the prediction technique applying the calibration model derived adopting PLS, 6 principal components and the spectral scans between 230 – 500 nm of each calibration sample. The linear response between actual and predicted values are shown in figure 4.7 for OD_{650}, RNA and Protein (BSA) and a linear fit (least squares) resulted in coefficients of determination of $R^2_{protein} = 0.986$, $R^2_{RNA} = 0.997$ and $R^2_{OD650} = 0.999$. Figure 4.8 illustrates the same data as figure 4.7 however here both the predicted and actual values are plotted against the calibration sample number (table 4.1). This shows the spread of predicted data around the actual values, indicating the larger scatter of the protein predictions, which was probably due to the weaker adsorbance signals seen in figure 4.4, plot C.

Figure 4.9 shows the quantitative ability of the prediction technique to estimate protein and cell debris levels in Baker's yeast homogenate samples from their UV-VIS spectral scans (230nm to 500nm). RNA predictions will be discussed later, figure 4.10. Figure 4.9, **plots A** and **B** illustrate measured and predicted protein and OD_{650} levels against the homogenate sample number. Although similar trends in measured and predicted data for both protein and OD_{650} profiles were observed, the predicted protein profile was shifted upwards compared to that of the measured protein profile resulting in systematic error throughout the homogenate samples. Figure 4.9, **plots C** and **D** show the predicted and measured protein and cell debris data plotted against each other resulting in a highly linear response with coefficients of determination (R^2) of 0.995 and 0.991 for protein and OD_{650}. The linear equations for the contaminants are shown in the figure legend, both of which have an intercept at (0,0) and slopes of 0.71 (measured/predicted) and 1.1 (measured/predicted) for protein and debris. These linear correlations will be implemented to correct the predicted data. This linear correction can be argued to describe the transition from prediction of contaminants in

buffer to predictions of contaminants in homogenate solution. A prediction error was determined by taking the average residual values between the measured and corrected predicted data, described in equation 4.1.

$$Prediction\ error = \frac{\sum_{0}^{k} \left| c_m - c_p \times \alpha \right|}{n} \qquad (4.1)$$

where c_m is the measured contaminant concentration, c_p is the predicted contaminant concentration, α is the linear correction factor and n is the number of k samples. If contaminant yield values are used the above corrections to the predicted data are not required. The prediction error for protein and debris was 0.41 mg mL^{-1} and 0.136 Au.

Figure 4.10 illustrates predicted and measured RNA data for a flocculation run using Baker's yeast homogenate at a concentration of 125 g L^{-1} and a stock polyethyleneimine (PEI) solution of 0.01 g L^{-1}. Homogenate-PEI mixtures were spun in a laboratory centrifuge prior to contaminant estimation. Predicted and measured data are plotted against PEI volume fraction figure 4.10, **plot A**. Figure 4.10, **plot B** shows predicted versus measured RNA data for the same flocculation run. The predicted data were a good match to that of the measured values. The linear relationship (shown in the figure legend) had a coefficient of determination of 0.991, intercept at (0,0) and a slope of 0.961 (measured/predicted). The prediction error based on equation 4.1 was 0.11 mg mL^{-1}.

The ability of the calibration model to predict RNA, protein and cell debris concentration from their UV-VIS spectral scans was demonstrated to be relatively good. The accuracy (based on figure 4.7) was highest for RNA, cell debris and then protein which is coincident with the level of absorbance for each of the contaminants in the UV-VIS spectra. Predicted protein and RNA profiles were shifted upwards, predicting higher values compared to their chemical measurements. Noui *et al.* (1998) argued that this phenomena was due to the interference of other components found in Baker's yeast homogenate such as lipids and DNA which increased the absorbance readings resulting in higher predictions. As there exists a highly linear relationship between the predicted and measured contaminant levels with the intercept at (0,0) if relative contaminant values (for example yield terms) were estimated the systematic prediction error would be eliminated, and there would be no need to correct the predicted data. The prediction technique is used for at-line contaminant estimation of process streams.

4.3.4 Photometric dispersion analyser (PDA)

Figure 4.11 illustrates the behaviour of a batch flocculation run in terms of its cell debris yield levels (debris (OD_{650}) remaining in solution after centrifugation as a fraction of debris in spun feed) and photometric dispersion analyser light intensity (PDA dc) readings. The mechanism of the PDA instrument was described in section 3.3. PDA measurements were taken of the PEI-homogenate mixture according to the continuous sampling system described in section 3.5. Figure 4.11 illustrates that the PDA profile follows a similar trend to that of the cell debris yield data. The addition of flocculant (PEI) to the feed homogenate promotes flocculation (aggregation) of cell debris particulate as described in section 1.6.3 and hence facilitates their removal during centrifugation. The formation of the flocs also results in an increase in the PEI-homogenate mixture turbidity thereby decreasing the PDA light intensity (PDA dc) measurements. Although similar trends were found between the PDA measurements and debris yield data a direct correlation between these two profiles does not seem to exist. For example the minimum PDA value does not coincide with that of the debris yield profile. Further analysis of the PDA data's ability to characterise the flocculation behaviour will be examined in chapter 5.

4.3.5 Error analysis of the at-line monitoring set-up and flocculation process

To determine the overall error of the at-line monitoring system an investigation into the variations in the performance of the micro-centrifuge, stopped flow analyser and both the diode-array spectrophotometer and single wavelength spectrophotometer was examined in terms of coefficient of variation.

The first step in the sampling sequence is the micro-centrifuge which was implemented for solid-liquid separation (section 3.3). **Figure 4.12** illustrates OD_{650} yield values (y_{OD650} = $OD_{650}(i)$ / $OD_{650}(0)$, where i refers to sample i and 0 to the feed sample) against PEI volume fraction for a flocculation run using both the micro-centrifuge and a lab scale centrifuge for solid-liquid separation. Therefore a lower yield value means better removal of debris. The batch flocculation run used Baker's yeast homogenate at a concentration of 125 g L^{-1} and a stock flocculant solution of 0.01 g L^{-1} polyehyleneimine (PEI). Optical density readings were taken in the same spectrophotometer (Pharmacia Ultro spec 2000). The Micro-centrifuge was operated under 4 Bar (approximately 55,000rpm (Chard 1997)) for 15 seconds, whilst the lab centrifuge was operated at 13,000rpm for 2 minutes. Dilutions prior to optical density readings were performed manually for the lab centrifuge and automatically for the

micro-centrifuge by the stopped flow analyser (1:17 factor dilution). Above PEI volume fractions of 0.025 (v/v) no off-line dilutions were required. Figure 4.12, shows that the micro-centrifuge generally under performed in terms of OD_{650} yield, compared to the lab scale centrifuge. The optimum PEI dose in terms of cell debris removal (OD_{650} reduction) fell in the region of 0.05-0.1 (v/v) PEI volume fraction for both the lab-centrifuge and the micro-centrifuge with OD_{650} yield values of 0.02 and 0.1. The under performance of the micro-centrifuge was studied by Turner (1993) and Chard (1997) and was argued to be due to the re-suspension of solids in the supernatant stream.

The coefficient of variation (CV = standard deviation/mean) for both lab and micro-centrifuged flocculation samples were highest close to the optimum cell debris separation area, with a maximum of 22% at low OD_{650} yield values and lowest for PEI volume fraction of zero, i.e. homogenate feed. The average CV for the micro-centrifuge and lab-scale centrifuge were 8.7 % and 8.9 %. It should be noted that the CV values also account for the variations due to dilutions and sample transportation.

Figure 4.13, shows the percent residual errors of at-line measurements of OD_{650}, RNA, protein and ADH of Baker's yeast (125 g L^{-1}) homogenate samples. The homogenate samples were first centrifuged in the micro-centrifuge, followed by sample dilution and reagent mixing by the stopped flow analyser before the at-line assay for ADH and at-line spectral scan predictions were taken. The percentage residual error and average percentage residual error are defined as:

$$\%Residual\ error = 100\left(\frac{\mu - x_i}{\mu}\right) \tag{4.2}$$

$$\%Average\ residual\ error = 100\frac{1}{n}\sum_{0}^{k}\frac{\left|\mu - x_i\right|}{\mu} \tag{4.3}$$

Where μ is the mean value, x_i is the actual value (ADH, RNA, debris, or protein) at sample i, and n is the total number of k samples. Figure 4.13 illustrates that the at-line OD_{650} readings are associated with the percentage largest residual errors, then protein, RNA and ADH values follow. The average percentage residual errors for ADH, RNA, protein and OD650 were 2.2%, 1.9%, 3.4% and 4.5%, whilst the average coefficient of variation were estimated to be ±2.7%, ±3.5%, ±4.8% and ±6.2%. Note protein, RNA and ADH levels may be assumed soluble in feed homogenate, hence any variations in the performance of the micro-centrifuge should not affect these components. The variations in at-line measurements of RNA, protein and ADH may be attributed directly

to stopped flow analyser dilution and reagent mixing, contaminant prediction and ADH assaying inconsistencies. Cell debris at-line measurements however are highly dependent on the efficiency of the solid-liquid separation step and hence associated with the higher coefficient of variation (±6.2%).

Figures 4.14 and **4.15** illustrate results derived from error analysis on the Turnbull control system (TCS) and two gear pumps implemented in both the continuous and batch flocculation process. To determined the errors associated with each of the pumps a flowrate setpoint was set and measured flowrates were compared in terms of coefficient of variation. Figure 4.14 shows the measured against setpoint flowrate (in TCS units) for the feed and flocculant pump at relevant flowrates (low flocculant and higher feed flowrates). The correlation coefficient as expected is very high with R^2 values for both pumps of above 0.999 and slopes of 1. Although the standard deviation for all flowrates were relatively constant for each setpoint, the coefficient of variation (CV) was seen to be highest for lower flowrates especially for the flocculant pump, with the highest variation around the mean values of ± 6%. **Figure 4.15**, illustrates setpoint versus measured volume fractions for the variable gear pumps on the bench scale flocculation rig using water. The volume fraction X_{P1} was defined as $X_{P1} = P_1/(P_1+P_2)$, where P_1 was the flocculant pump flowrate and P_2 the homogenate feed pump flowrate. A linear fit between the actual and setpoint volume fraction data resulted in an R^2 value of 0.999 and the average coefficient of variation was estimated to be ±2.5%. From figure 4.15 it is can be observed that the lower volume fraction setpoint incur the larger errors, which was due to the limitations on the minimum flowrate of flocculant pump.

Table 4.2, lists the individual apparatus errors and the overall at-line measurement variations in terms of RNA, protein OD_{650} and ADH. Two different overall estimates have been given. The first overall at-line monitoring error estimate (table 4.2, row 9) is an average value based on homogenate-PEI samples. This was achieved by adding the errors from the individual steps (row 4,5,6,7,8), according to the principal of serial error addition described in equation 4.4 (below). The micro-centrifuge average error was obtained from figure 4.12, whilst errors involved with sample dilution and reagent mixing was obtained through dilution experiments involving homogenate yeast and buffer. It was assumed that the errors due to the stopped flow analyser (dilution and reagent mixing) was constant throughout a flocculation run. The second overall at-line estimate error estimate (table 4.2, row 10) was based on repeat measurements of homogenised Baker's yeast samples at a concentration of 125 g L^{-1} (figure 4.13). As

seen from table 4.2 the overall coefficient of variation of at-line measurements based on yeast homogenate is considerably lower than that estimated from the addition of individual errors based on an average error throughout the flocculation profile. This was mainly due to the large variation associated with the micro-centrifuge (±8.4%). In a flocculation process a fraction of ADH, RNA and protein are rendered insoluble and hence effected by the variations in the solid-liquid separation step.

$$(\Delta AB)^2 = (\Delta A)^2 + (\Delta B)^2 \qquad (4.4)$$

Where Δ refers to error or coefficient of variation.

4.4 Conclusion

A linear relationship between cell debris levels and optical density readings at 650 nm were found enabling the estimation of debris contaminant through rapid spectrophotometric measurements. An existing off-line kinetics ADH assay has been successfully converted for at-line purposes and demonstrated to be reproducible for a shorter assay time through the use of a stopped flow analyser. Rapid UV-VIS spectral analysis through a multivariate analysis technique has allowed the prediction of RNA, cell debris and protein levels in yeast homogenate. Automatic sampling, solid-liquid separation, and sample preparation has been demonstrated enabling the at-line measurements of ADH, RNA, protein and cell debris. An overall error analysis of the at-line monitoring setup was estimated indicating that the largest errors were associated with the solid-liquid separation step through a micro-centrifuge (±8.4%). The overall at-line measurement errors for the contaminants and product were ±8.6% and ±8.9%.

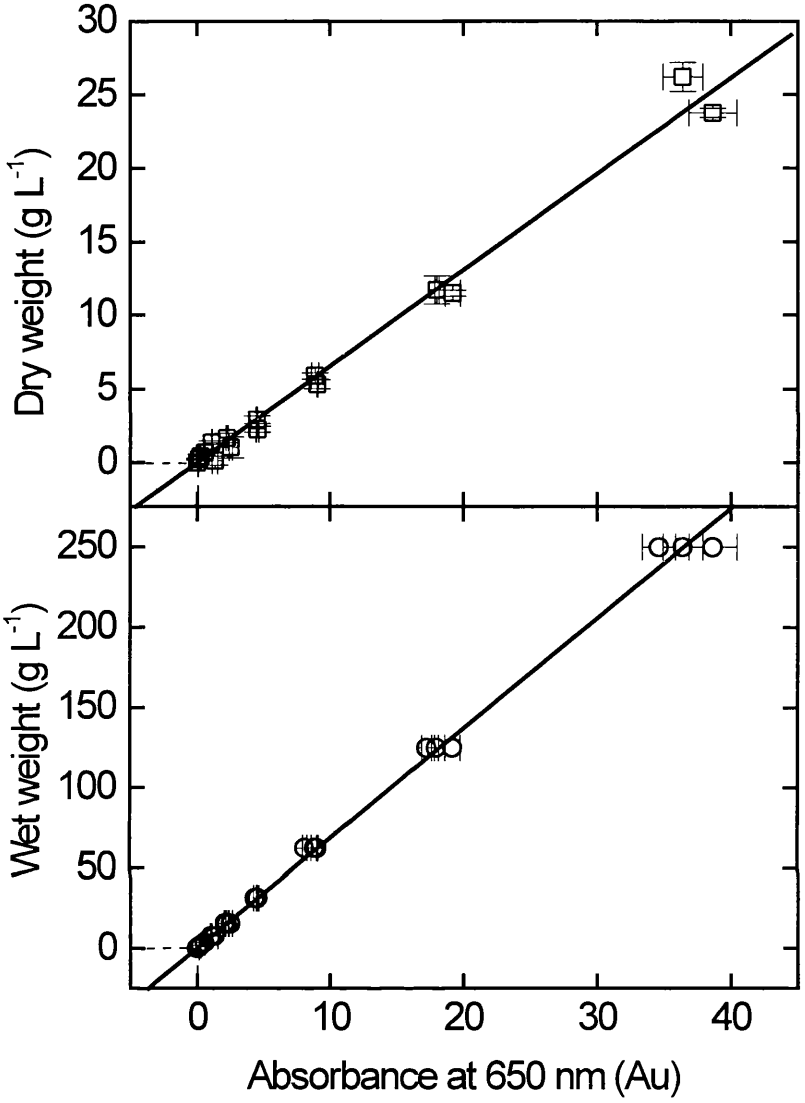


Figure 4.1. *Correlation of wet weight and dry weight with OD_{650} for homogenised Baker's yeast. Homogenised Baker's yeast at a concentration of 250 g (wet weight) L^{-1} was diluted serially to create samples at 250 g L^{-1}, 125 g L^{-1}, 62.5 g L^{-1}, 13.25 g L^{-1}, 15.63 g L^{-1}, 7.82 g L^{-1}, and 3.91 g L^{-1}. Dry weight measurements were taken of 1 mL homogenate samples and dried in an oven at 100°C for 48 h. Linear fits (least squares fit) are shown forced through (0,0) and weighted for variance.*

(□) [Wet weight] = 6.85 x OD_{650} R^2 = 0.99

(○) [Dry weight] = 0.653 x OD_{650} R^2 = 0.988

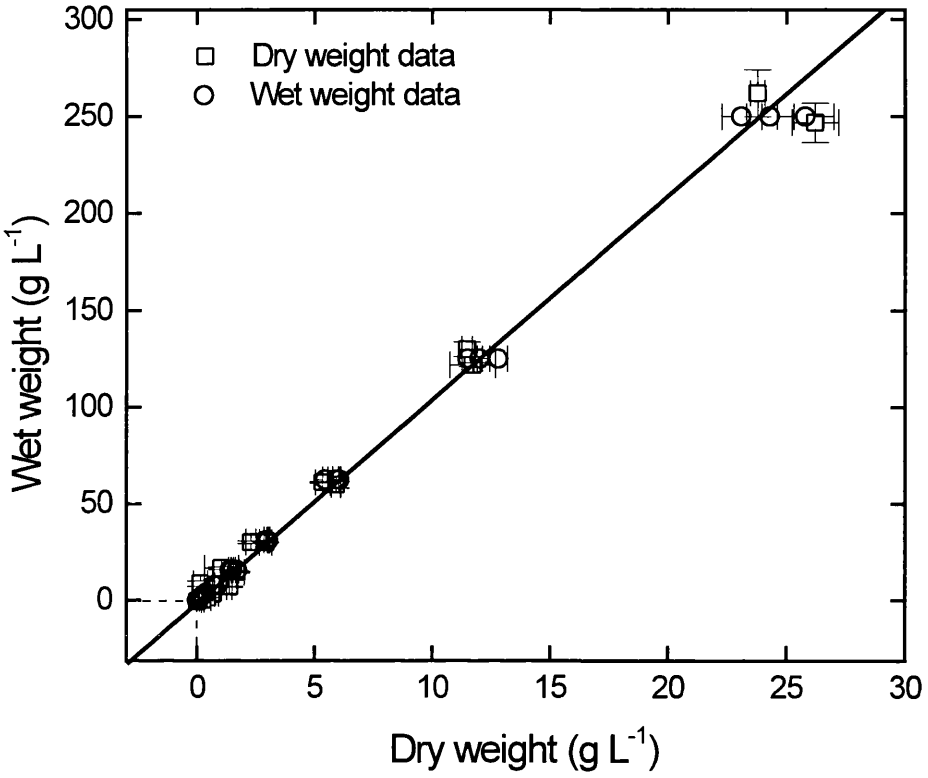


Figure 4.2. *Wet weight dry weight correlation for homogenised Baker's yeast. . Homogenised Baker's yeast at a concentration of 250 g (wet weight) L^{-1} was diluted serially to create samples at 250 g L^{-1}, 125 g L^{-1}, 62.5 g L^{-1}, 13.25 g L^{-1}, 15.63 g L^{-1}, 7.82 g L^{-1}, and 3.91 g L^{-1}. Dry weight measurements were taken of 1 mL homogenate samples and dried in an oven at 100°C for 48 h. The data points are derived from combining the measurements from figure 4.1. Linear fits (weighted least squares) of both the wet weight and the dry weight data against OD650 have been plotted, and the average correlation relating these two variables to each other is:*

[Dry weight] = 0.0953 x [Wet weight]

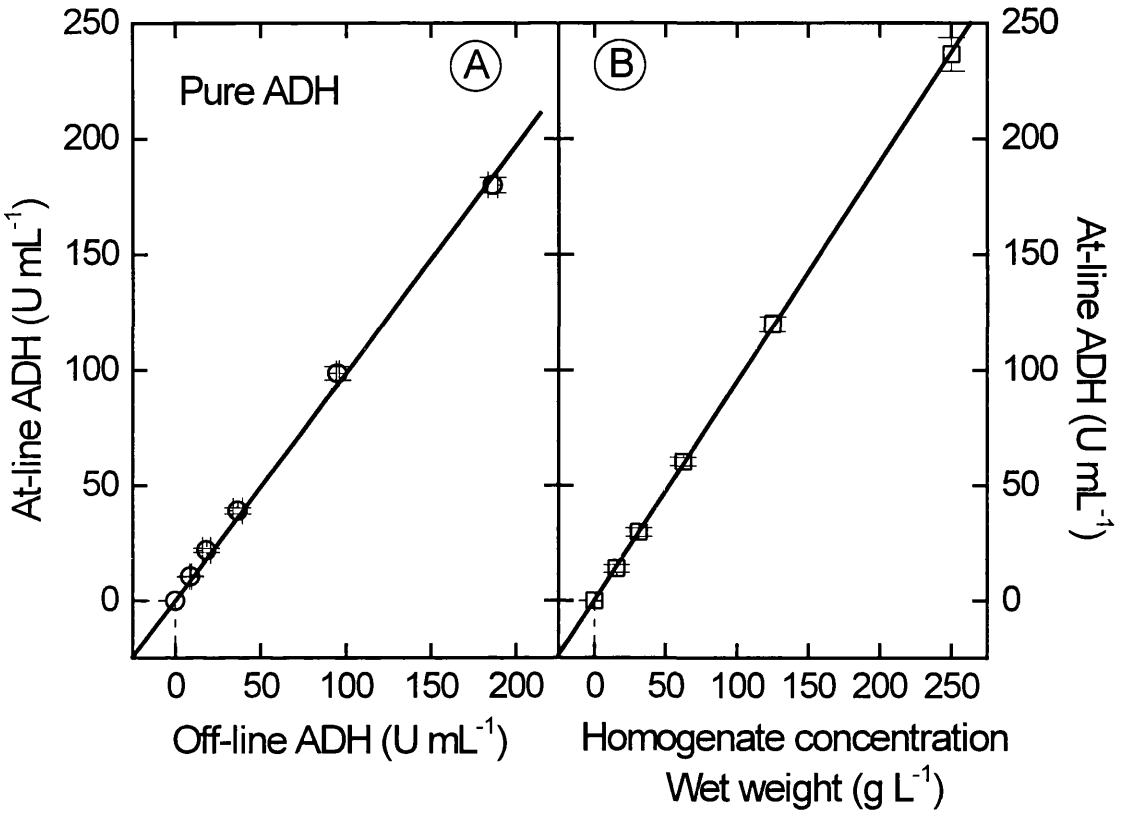


Figure 4.3. *At-line and off-line ADH measurements of pure ADH and Baker's yeast homogenate samples.*

Plot A. At-line measurements versus off-line ADH assays for pure Sigma-ADH samples. Approximately 3300 U mL^{-1} of Sigma ADH was diluted [1] (pH 6.5, ionic strength 100mM (KH_2PO_4)) 10, 20, 50, 100 and 200-fold to make up five samples. Both off-line and at-line assays were performed in the same spectrophotometer (Pharmacia, Ultro spec 2000). The linear fit (weighted least squares fit) between at-line and off-line measurements (in triplicate) resulted in a coefficient of correlation of 0.999.

Plot B. At-line ADH measurements of homogenised Baker's yeast samples. Homogenised Baker's yeast at a concentration of 250 g (wet weight) L^{-1} (pH 6.5, ionic strength 100 mM (KH_2PO_4)) was diluted serially (1, 2, 4, 8, 16-fold dilutions) to form samples at 250 g L^{-1}, 125 g L^{-1}, 62.5 g L^{-1}, 31.25 g L^{-1}, 15.63 g L^{-1}. The linear fit (weighted least squares fit) resulted in an R^2 value of 0.999.

The at-line monitoring set-up adopted a 1 mm pathlength flowcell, a 10s assay time and automatic sample dilution, reagent mixing and transportation to spectrophotometer by a stopped flow analyser (SFA).

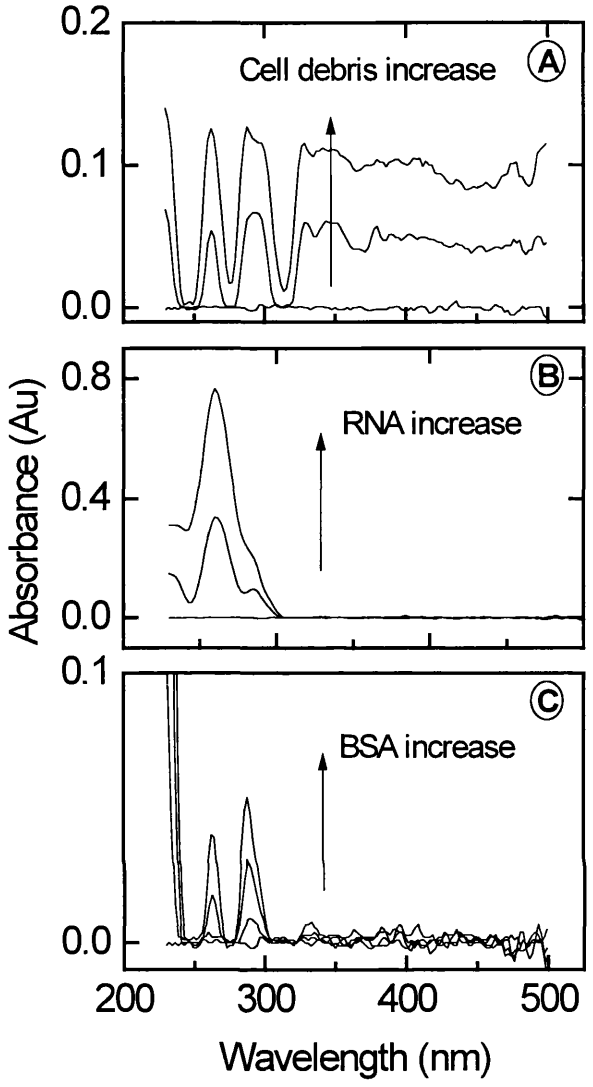


Figure 4.4. *UV-VIS spectral scans due to cell debris, RNA and BSA concentration changes. Washed cell debris, RNA and BSA were added to phosphate buffer (100mM (KH_2PO_4), pH 6.5). Plot* ***A*** *illustrates the spectral scan (indicating by arrow) due to an increase in cell debris resulting in OD_{650} levels of 0 Au, 2.77 Au to 5.5 Au whilst keeping BSA and RNA levels at zero. Plot* ***B****, shows the spectral scan due to an increase in RNA levels from 0 mg mL^{-1}, 2.5 mg mL^{-1} to 5 mg mL^{-1} whilst keeping cell debris and BSA levels at zero. Plot* ***C****, shows the spectral scan due to an increase in BSA levels 0 mg mL^{-1}, 4 mg mL^{-1}, 12 mg mL^{-1} to 20 mg mL^{-1} whilst keeping cell debris and RNA levels at zero. The above spectral scans were taken after diluting samples by a factor of 16 with phosphate buffer and introduced to the diode array spectrophotometer via a 1mm pathlength flowcell.*

Sample Number	Sample composition [1mL,1mL, 1mL]	Sample number	Sample composition [1mL,1mL, 1mL]	Sample number	Sample composition [1mL,1mL, 1mL]
1	AXM	13	BXM	25	CXM
2	AXN	14	BXN	26	CXN
3	AXO	15	BXO	27	CXO
4	AXP	16	BXP	28	CXP
5	AYM	17	BYM	29	CYM
6	AYN	18	BYN	30	CYN
7	AYO	19	BYO	31	CYO
8	AYP	20	BYP	32	CYP
9	AZM	21	BZM	33	CZM
10	AZN	22	BZN	34	CZN
11	AZO	23	BZO	35	CZO
12	AZP	24	BZP	36	CZP

Washed debris	RNA	BSA
C = 1.03 Au at 650 nm & 0.054 mg mL^{-1} of protein	**Z** = 0.938 mg mL^{-1}	**P** = 3.75 mg mL^{-1}
B = 0.52 Au at 650 nm & 0.032 mg mL^{-1} of protein	**Y** = 0.469 mg mL^{-1}	**O** = 2.25 mg mL^{-1}
A = Buffer	**X** = Buffer	**N** = 0.75 mg mL^{-1}
		M = Buffer

Table 4.1. *Factorial setup for UV-VIS spectral scan calibration using contaminant standards. Buffer refers to 100 mM phosphate buffer at 6.5 pH and BSA refers to bovine serum albumin protein. The sample compositions had a total volume of 3 mL (1mL of each contaminant). Washed cell debris and RNA were varied three way, whilst BSA protein was varied 4 way to form the full factorial experiment with 36 samples. Washed cell debris had 0.054 mg mL^{-1} of protein whilst it was assumed that the RNA levels were negligible.*

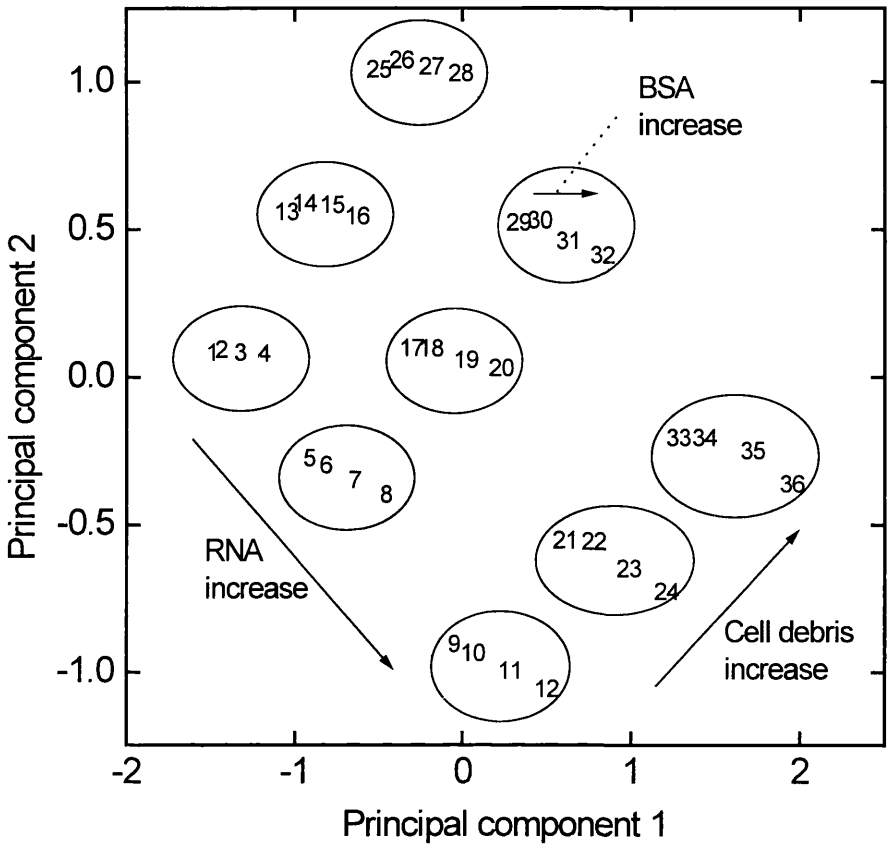


Figure 4.5. *Scores plot of UV-VIS spectral scan calibration data in terms of principle components 1 and 2. The numbers on the plot refer to the spectral scan taken of the particular sample number in the factorial experiment (table 4.1). Sample clusters (samples which have a geometrically close proximity in a scores plot) have been circled. Trends due to RNA, BSA and cell debris variations are indicated by the arrows on the plot.*

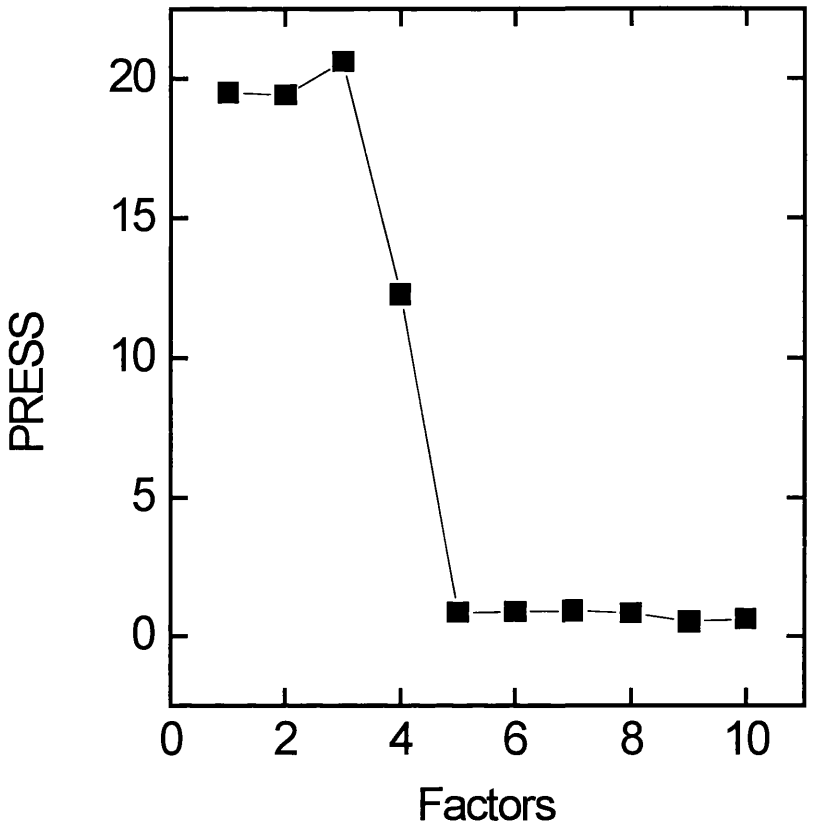


Figure 4.6. *PRESS (prediction residual error sum of squares) versus the number of principal components (factors) for the calibration data. The above plot shows that the optimum number of components is 5-6.*

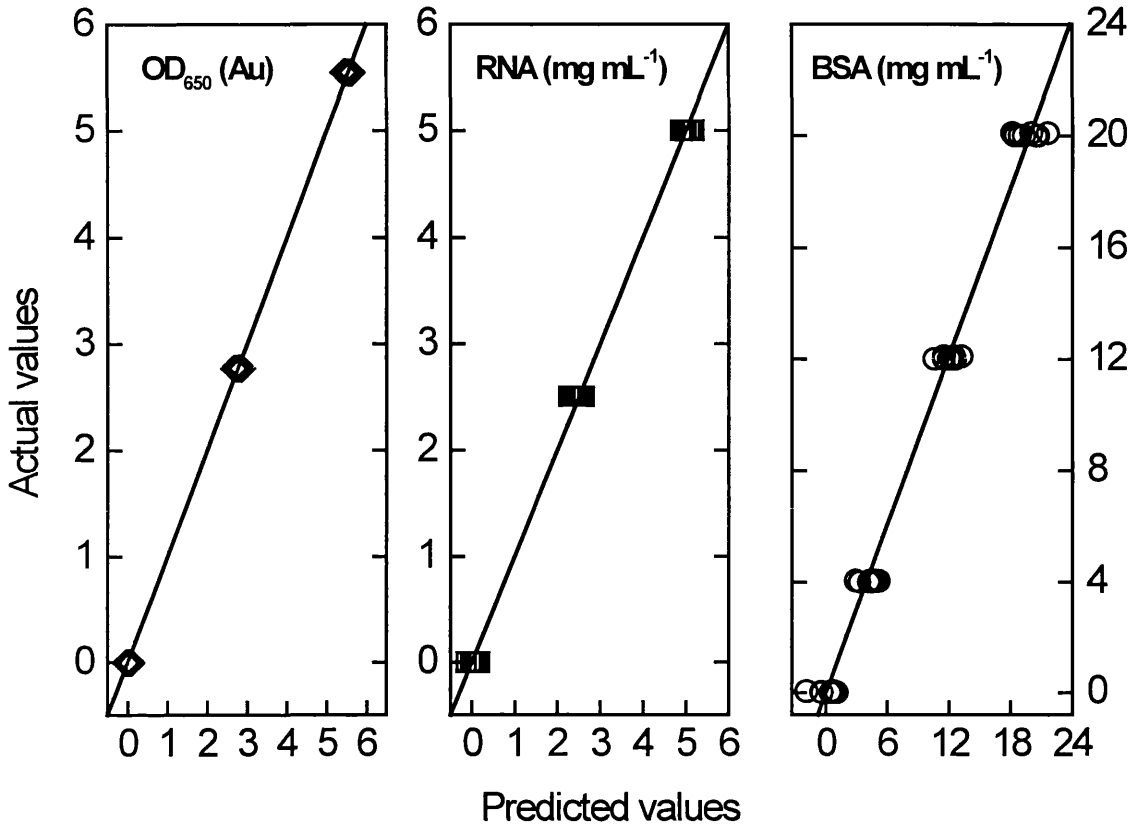


Figure 4.7. *Predicted OD_{650}, RNA and BSA data against actual values for the calibration samples. Each of the above plots consists of 36 data points where the actual readings are listed in table 4.1. The results have coefficients of determination of 0.999, 0.997 and 0.986 for OD_{650}, RNA and BSA. The PLS correlation is based on 6 principal components.*

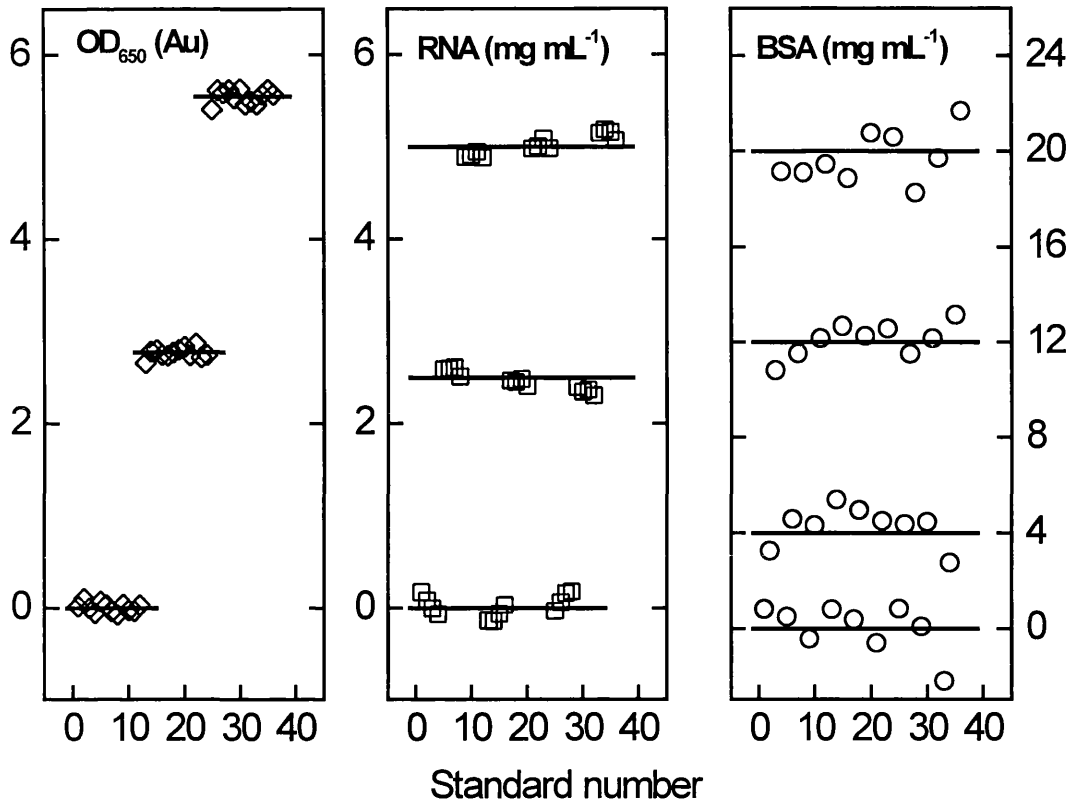


Figure 4.8. *Predicted and actual OD_{650}, RNA and BSA data plotted in terms of calibration sample number. The solid lines refer to the actual OD_{650}, RNA and BSA levels. The results have coefficients of determination of 0.999, 0.997 and 0.986 for OD_{650}, RNA and BSA. The PLS correlation is based on 6 principle components.*

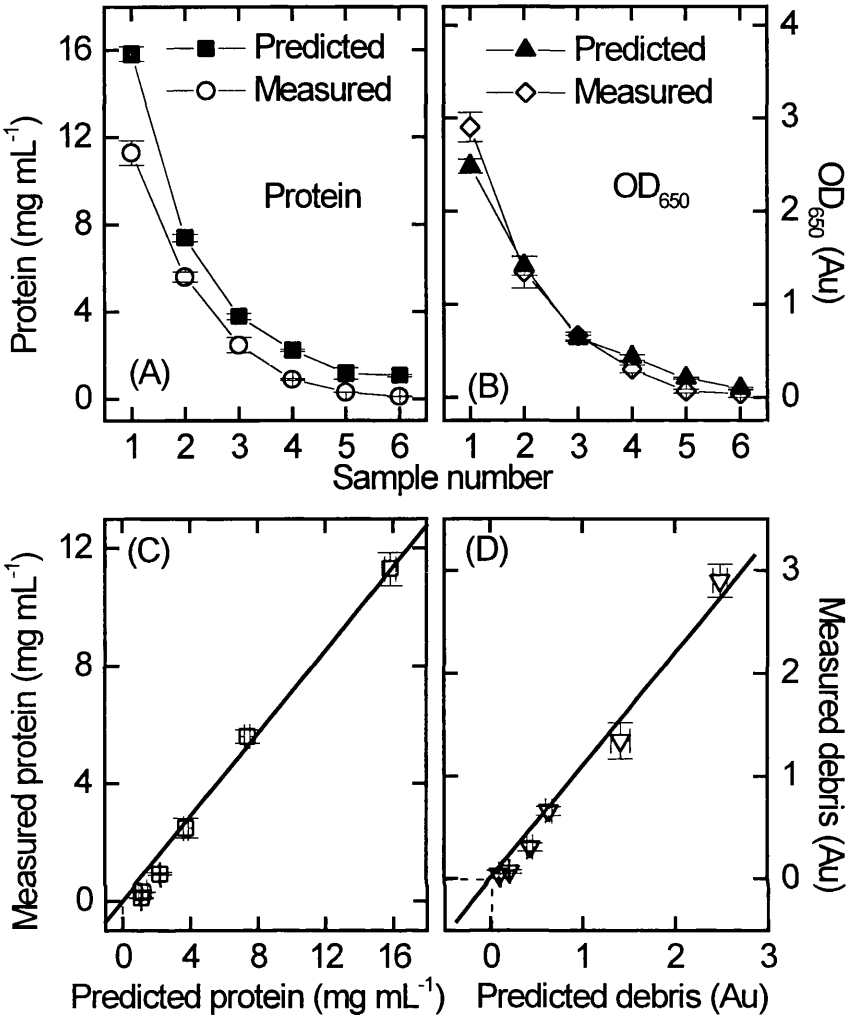


Figure 4.9. *Measured and predicted protein and OD_{650} data for spun Baker's yeast homogenate samples. Homogenate Baker's yeast at a concentration of 125 g (wet weight) L^{-1} was diluted with phosphate buffer to create samples at 62.5 g L^{-1}, 31.25 g L^{-1}, 15.63 g L^{-1}, 7.82 g L^{-1}, and 3.91 g L^{-1}. The samples were spun at 13,000 rpm for 5 minutes in a laboratory centrifuge prior to off-line and prediction analysis. The data points and error bars are derived from the mean and standard deviation of triplicate measurements for both the measured and predicted data. . The predicted contaminant levels were based on the PLS calibration model using 6 principal components.*

Plot (A) and (B). Predicted and measured data in terms of sample number.

Plot (C) and (D). Predicted against measured data. Linear fits were forced through (0,0) and the equation and coefficients of determination (R^2) are listed below:

[Protein_measured] = [Protein_predicted] x 0.71 $R^2 = 0.995$

[Debris_measured] = [Debris_predicted] x 1.1 $R^2 = 0.991$

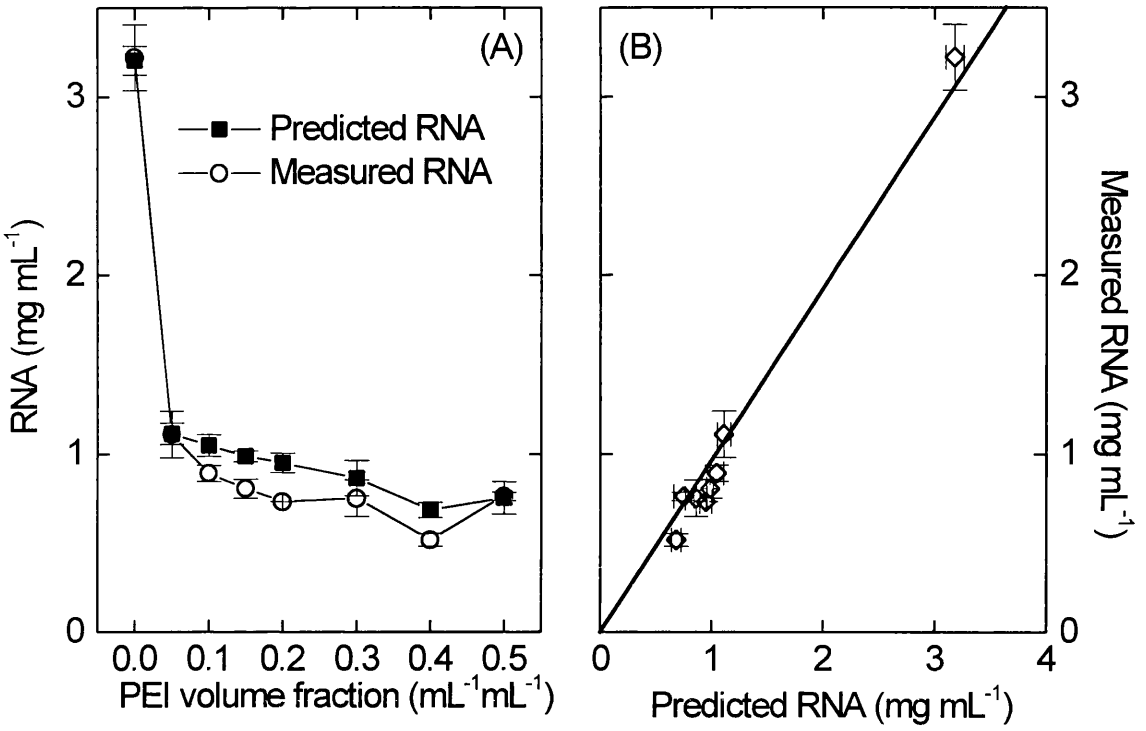


***Figure 4.10**. Measured and predicted RNA data of flocculation samples.*

Plot A. Predicted and measured RNA data for a flocculation run. Baker's yeast homogenate at a concentration of 125 g (wet weight) L^{-1} (pH 6.5 and ionic strength of 50mM (KH_2PO_4)) was flocculated with a stock polyethyleneimine (PEI) solution of 0.01 g L^{-1} (pH 6.5 and ionic strength of 50 mM). Homogenate with PEI volume fractions of 0, 0.05, 0.1, 0.15, 0.2, 0.3, 0.4, and 0.5 (v/v) were spun for 5 minutes at 13,000 rpm in a laboratory centrifuge. The data points and error bars are derived from the mean and standard deviation of triplicate measurements for both the measured and predicted data. The predicted contaminant levels were based on the PLS calibration model using 6 principal components.

Plot B. Predicted against measured RNA data of flocculation samples. The linear fit was forced through (0,0) and the equation and coefficients of determination (R^2) are listed below:

$$[RNA\_measured] = [RNA\_predicted] \times 0.961 \qquad R^2 = 0.991$$

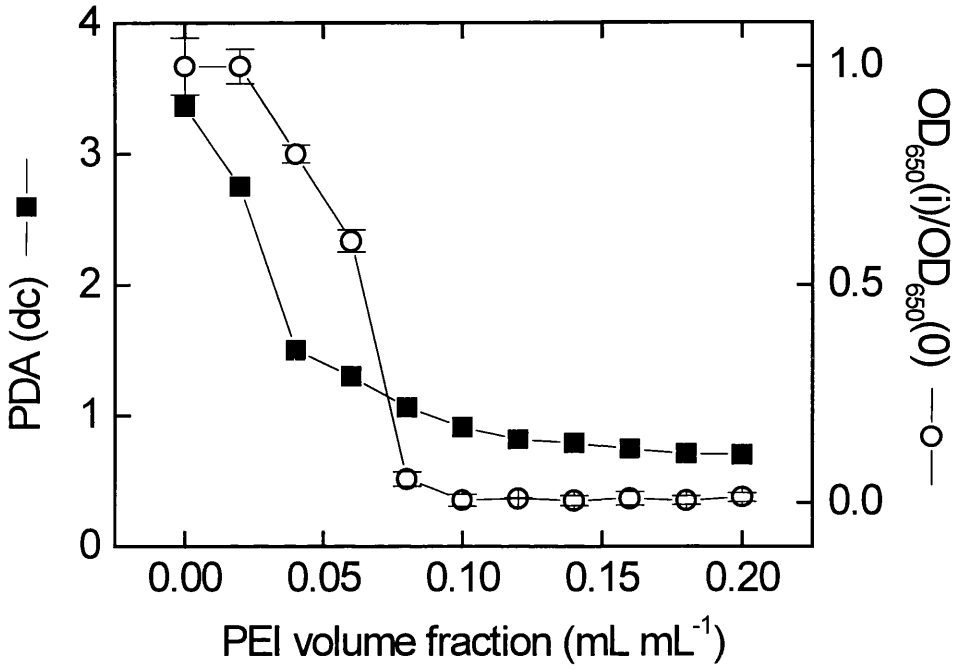


Figure 4.11. *Behaviour of PDA data during a batch flocculation operation.*

Yeast homogenate (125 g (wet weight) L^{-1}) was flocculated with PEI (0.01 g L^{-1}). Both solutions were at pH 6.5 and ionic strength of 50 mM (KH_2PO_4).

PEI was added incrementally to a batch tank containing the yeast homogenate. Optical density readings at λ = 650 nm (OD_{650}) of spun samples (13,000 rpm, 5 min in a lab centrifuge) in triplicates. The OD_{650} data is plotted as a yield value ($OD_{650}(i)/OD_{650}(0)$), where i and 0 refer to sample number and feed.

PDA (dc) data were taken of the homogenate-PEI mixture through a continuous sample loop described in section 3.5.

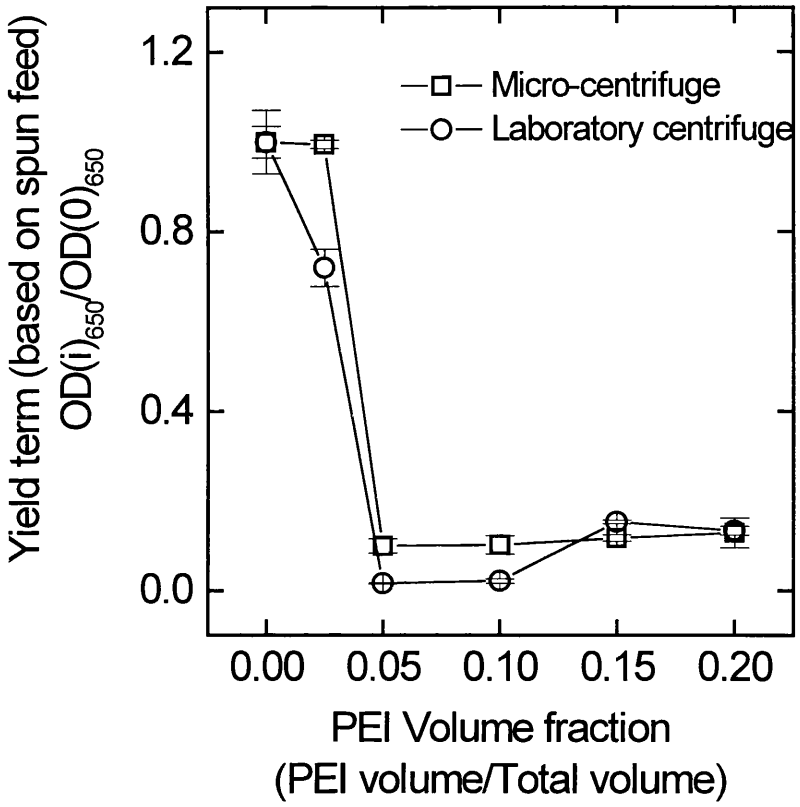


Figure 4.12. *OD650 yield values against PEI volume fraction for a flocculation run using both the micro-centrifuge and a lab scale centrifuge for solid-liquid separation. The batch flocculation run used 125 g (wet weight) L^{-1} Baker's yeast homogenate and a stock flocculant solution of 0.01 g L^{-1} PEI. Both solutions were at pH 6.5 and ionic strength of 50 mM (KH_2PO_4). Optical density readings were taken in the same spectrophotometer (Pharmacia Ultro spec 2000). The Micro-centrifuge was operated at 4 Bar(g) (approximately 55,000 rpm (Chard 1997)) for 15 seconds, whilst the lab centrifuge was operated at 13,000 rpm for 2 minutes. Dilutions prior to optical density readings were performed manually for the lab centrifuge and automatically for the micro-centrifuge by the stopped flow analyser. Above PEI volume fractions of 0.025 (v/v) no off-line dilutions were required. The average coefficient of variation for the lab and micro centrifuge (and dilution) were ±8.9% and ±8.7%.*

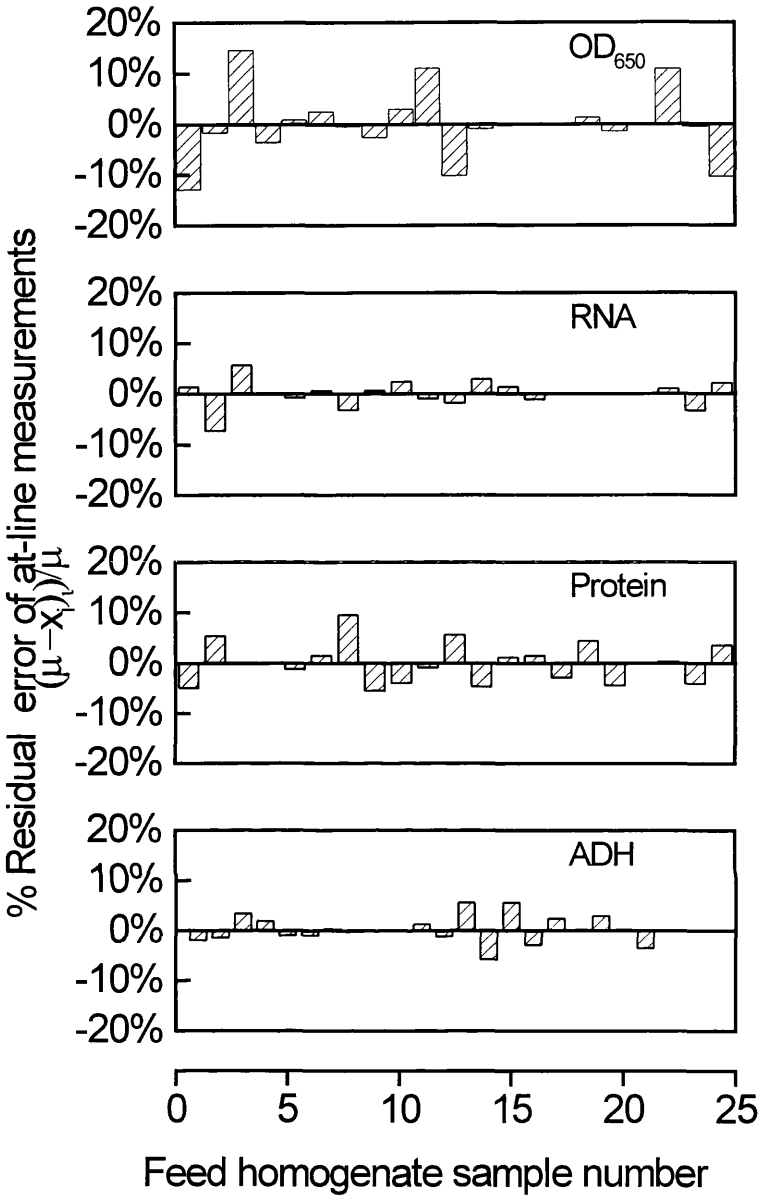


Figure 4.13. *Percent residual errors of at-line measurements of OD_{650}, RNA, protein and ADH of Baker's yeast homogenate (125 g (wet weight) L^{-1}, pH 6.5, 50 mM (KH_2PO_4)). The homogenate samples were first centrifuged in the micro-centrifuge, followed by sample dilution and reagent mixing by the stopped flow analyser before the at-line assay (for ADH) and at-line spectral scan prediction were taken. The μ term on the plot refers to the mean at-line measurement value whilst x_i refers to the at-line measurement of sample i. The average percent residual error of at-line measurement for OD_{650}, RNA, protein and ADH were 4.5%, 1.9%, 3.4% and 2.2% and the average coefficient of variation were ±6.2%, ±3.5%, ±4.8% and ±2.6%.*

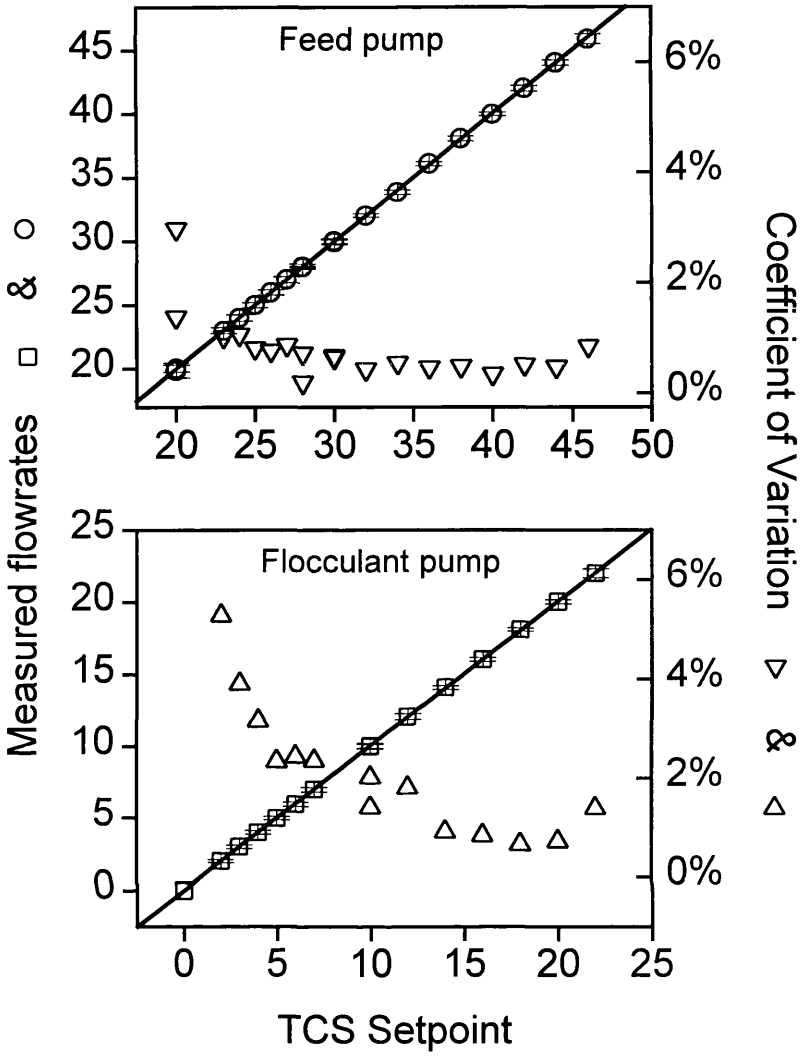


Figure 4.14. *Measured flowrates and coefficient of variation against setpoint flowrates for flocculant (P1) and feed (P2) gear pumps for the bench scale flocculation rig using water. The coefficient of variation (CV) was defined as the standard deviation over the mean flowrate reading, i.e. a percentage variation around the mean value. The flowrates are defined in TCS units. Linear fits resulted in coefficients of determination of 0.999 for both pumps and an average CV was estimated to be ±2% for the flocculant pump and ±0.7% for the feed pump.*

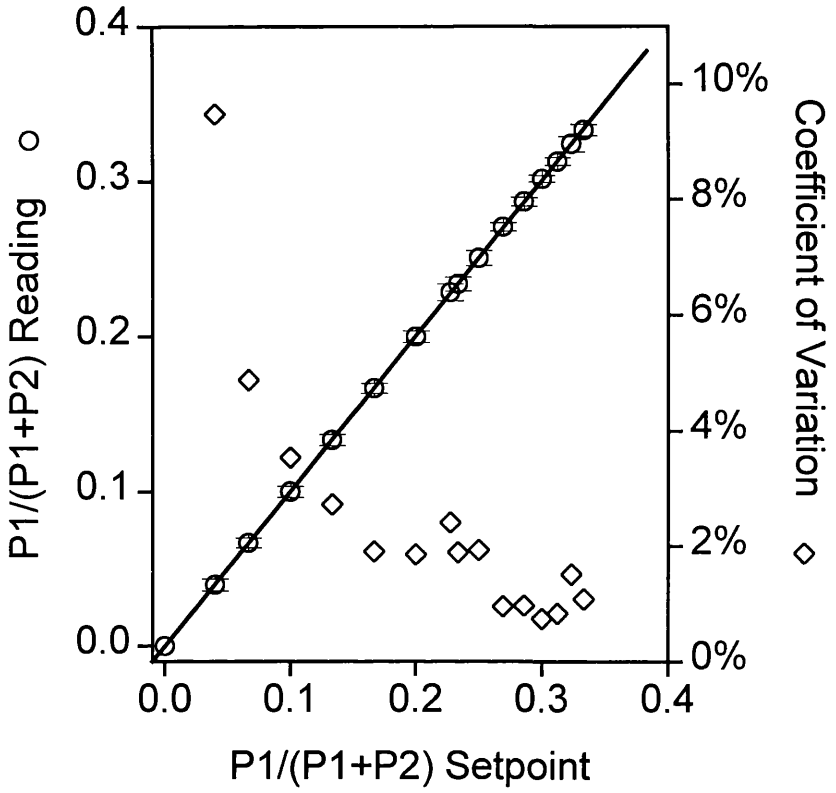


Figure 4.15. *Setpoint versus actual volume fractions for the variable gear pumps on the bench scale flocculation rig using water. The volume fraction X_{P1} was defined as*

$$X_{P1} = \frac{P_1}{P_1 + P_2}$$

Where P_1 and P_2 refer to the flocculant pump flowrate and feed pump flowrate. A linear fit between the actual and setpoint volume fraction data resulted in an R^2 value of 0.999 and the average coefficient of variation was estimated to be ±2.5%.

		Flow rates	ADH	OD650	RNA	Protein	From Figure
1	Flocculant Pump	**±2.5%**					4.14
2	Feed Pump	**±0.7%**					4.14
3	Volume fraction	**±2.6%**					4.15
4	Micro-centrifuge		±8.4%	**±8.4%**	±8.4%	±8.4%	4.12
5	SFA_P_1		±2%	**±2%**	±2%	±2%	
6	SFA_P_3		±2%				
7	ADH-Spec.		±0.5*				
8	Diode-array Spec.			±1%*	±1%*	±1%*	
9	At-line measurements based on average flocculation conditions		±8.9%	±8.7%	±8.7%	±8.7%	
10	At-line measurements based on feed		±2.7%	±6.2%	±3.5%	±4.8%	4.13

Table 4.2. *Overall analysis of at-line monitoring setup in terms of coefficient of variation. The figures in bold refer to measured variations whilst others are derived values from the principle of serial error addition, i.e.* $(\Delta AB)^2 = (\Delta A)^2 + (\Delta B)^2$, *where* Δ *refers to error or coefficient of variation. Results in row 10 are based on yeast homogenate samples at a concentration of 125 g (wet weight)* L^{-1} *(pH 6.5, 50 mM* (KH_2PO_4)*). Results in row 9 are based on the addition of serial errors for average variations during flocculation conditions described in figure 4.14. The last column indicates the figure from which the appropriate data has been taken. SFA_*P_1 *and SFA_*P_2 *refer to the first and second peristaltic pump on the stopped flow analyser for sample dilution and reagent mixing prior to assay analysis. The coefficient of variation for SFA_*P_1 *and SFA_*P_2 *where estimated through dilution tests using yeast homogenate and phosphate buffer. During the dilution for contaminant predictions only SFA-*P_1 *was used (chapter 3).*

- *Errors obtained from instrument manuals.*

Chapter 5

Rapid characterisation of the batch flocculation process

5.1 Summary

This chapter will present real time measurements of product and key contaminants for eight batch flocculation processes. An investigation into the use of three mathematical descriptions to characterise the flocculation process will be performed and how such characterisation can enable rapid definition for process analysis will be demonstrated.

5.2 Introduction

Flocculation behaviour is affected by several environmental conditions such as pH, ionic strength (salt concentration), flocculant dose and the concentration of flocculable components in the feed. This chapter will present at-line data of both product (alcohol dehydrogenate) and key contaminants (RNA, cell debris and protein) for several batch flocculation operations carried out under different feed volumes, feed and flocculant pH and ionic strength, flocculant stock solution concentration and flocculant dose rate. The batch processes were monitored using the at-line setup described in chapter 3.

The flocculation profiles of product and key contaminants will be characterised through three mathematical expressions, discussed in section 2.3. Such process modelling allows the design engineer to analyse the performance of a given flocculation run in terms of a few key factors, namely the model parameters. Furthermore the application of real time process characterisation enables the possibility of robust process control for process optimisation and reproducibility. Which mathematical description to apply for real time analysis will be discussed and the most appropriate model selected.

5.3 Mathematical characterisation of the flocculation process

Three mathematical equations will be implemented in the characterisation of the batch flocculation process. The mathematical descriptions are discussed in detail in chapter 2 and are summarised below in table 5.1. The flocculation models only have one variable namely the flocculant dose level. Input variables affecting floc removal

through centrifugation are not represented as a micro-centrifuge was applied for solid-liquid separation which is assumed ideal due to its high efficiency, i.e. all solids formed will be separated.

4-parameter equation	$\frac{E}{E_0} = \frac{1+\left(\frac{[PEI]}{b}\right)^m}{1+\left(\frac{[PEI]}{a}\right)^n}$
3(1)-parameter equation	$\frac{E}{E_0} = \frac{1}{1+\left(\frac{[PEI]}{a^*}\right)^{n^*}} + c[PEI]$
3(2)-parameter equation	$\frac{E}{E_0} = 1 - A\left(1 - exp\left[-\left(\frac{[PEI]}{B}\right)^C\right]\right)$

Table 5.1. Mathematical representation of flocculation profile. The model parameters a, b, c, n, m, a^*, n^*, A, B and C are a function of pH, ionic strength, feed and PEI stock solution conditions. The flocculant dose level is $[PEI]$. The parameters E and E_0 refer to spun flocculated samples and spun feed, hence the ratio of these two factors (E/E_0) is a performance indicator, i.e. the flocculation yield. All three models are non-linear with respect to $[PEI]$. For more detail on the above mathematical descriptions refer to chapter 2.

Model parameter identification will be performed using the Levenberg-Marquardt non-linear least squares technique on the at-line data for both product and key contaminants. Graphical representation of the model fits will be presented for the batch operations and the "goodness of fit" of the three flocculation models will be determined through a coefficient of determination (R^2);

$$R^2 = 1 - \frac{\sum[f(x_i) - y_i]^2}{\sum[\bar{y} - y_i]^2} \qquad (5.1)$$

where y_i is the actual flocculation yield at a flocculant dose x_i, $f(x_i)$ is the flocculation yield determined through the flocculation model and $\bar{y}$ is the mean flocculation yield for all levels of flocculant concentrations. The second term in equation 5.1 gives us an indication of how the model fit compares to the mean flocculation yield. If the model fit is good the numerator will approach zero due to small variation between model

prediction and actual levels, resulting in an R^2 term tending to unity. However if the model fit is poor the variation between the actual and both the predicted and mean yield values would be similar resulting in R^2 values close to zero. All in all the R^2 term gives an indication of the fraction of data which is explained by the model.

5.4 Mass balancing of the batch flocculation process

A mass balance on flocculant and homogenate concentrations for each sampling step was required to determine flocculant dose levels. The below equations describe the flocculant volume fractions in terms of homogenate, flocculant, sample volume and PEI dose flow-rate. The density of both homogenate and flocculant is assumed to be similar.

$$x_{PEI} = \frac{V_{PEI}(i-1)}{V_{PEI}(i-1)+V_H(i-1)} \tag{5.2}$$

$$V_{HS}(i) = (1 - x_{PEI}(i))V\text{s} \tag{5.3}$$

$$V_{PEIS}(i) = x_{PEI} V_S \tag{5.4}$$

$$V_H(i) = V_H(i-1) - V_{HS}(i) \tag{5.5}$$

$$V_{PEI}(i) = V_{PEI}(i-1) - V_{PEIS}(i) + F_{PEI}[t(i) - t(i-1)] \tag{5.6}$$

where x_{PEI} is the flocculant volume fraction, V_H, V_{PEI}, V_S, V_{HS} and V_{PEIS} are the volume of homogenate in the process vessel, flocculant in the process vessel, volume of sample, homogenate in sample and flocculant in sample, F_{PEI} is the dose flow-rate of flocculant, t is time, and i refers to the sampling step. For start conditions; $V_{H(0)}$ = start homogenate volume and $V_{PEI(0)}$ = 0.

5.5 Results

The results of eight batch runs will be presented in terms of direct measurements of flocculant and homogenate volumes, photometric dispersion analyser light intensity readings (referred to as PDA dc levels for the rest of this chpater) and at-line measurements of product and contaminants. Table 5.2 summarises the operating conditions applied for each batch run. Mathematical characterisation of the batch operations will be performed using the three flocculation models listed in table 5.1. The first part of this section will present the results from Batch Run1 in detail. The second part will present an overview of batch runs 2 to 8, underlining the main points of

interest. A qualitative discussion on how the environmental conditions such as pH, ionic strength and PEI concentration as well as batch operational variables such as feed volume and dose rate effect the flocculation profile will be presented. The efficiency of the three models in fitting the at-line data will be evaluated by comparing their coefficient of determination. Finally the batch runs will be compared to each other quantitatively, through the implementation of model parameters and the influence of start conditions on batch behaviour will be discussed.

Batch Run	1	2	3	4	5	6	7	8
Homogenate start volume (mL)	500	700	700	500	500	500	700	700
Homogenate pH	6.5	7.0	5.5	6.2	6.2	5.5	6.0	5.5
Homogenate ionic strength (mM)	50	50	50	50	50	50	20	20
Start Cell debris (Au)	3.1	2.4	2.3	3.3	3.5	3.7	3.2	3.2
Start Protein (mg mL^{-1})	13.8	11.4	12.1	13.2	14.2	11.2	14.8	13.1
Start RNA (mg mL^{-1})	3.4	2.8	2.2	3.7	3.5	3.0	3.3	3.4
Start ADH (U mL^{-1})	114	111	83	128	124	118	134	137
PEI dose rate (mL min^{-1})	2.82	2.82	2.82	2.82	2.82	2.82	F_{PEI}	F_{PEI}
Stock solution of PEI (mg mL^{-1})	10	10	10	5	5	5	10	10
PEI pH	6.5	7.0	5.5	5.9	5.9	5.0	6.0	5.5
PEI ionic strength (mM)	50	50	50	50	50	50	20	20
Max PEI volume fraction (mL mL^{-1})	0.46	0.21	0.25	0.27	0.23	0.30	0.16	0.15

Table 5.2. Overview of eight batch flocculation runs and their operating conditions. Start protein, RNA, cell debris and ADH refer to the concentration of contaminant and ADH in the spun feed homogenate prior to PEI dosing. F_{PEI} *refers to a variable flocculant dose rate for batch runs 7 & 8.*

5.5.1 Batch run 1

A batch volume of 500 mL of yeast homogenate at a concentration of 125 g L^{-1}, pH 6.5 and ionic strength of 50 mM (KH_2PO_4) was dosed at a flow-rate of 2.82 mL min^{-1}, with a polyethelyneimine (PEI) stock solution of 10 mg mL^{-1}, pH 6.5 and ionic strength of 50mM (KH_2PO_4). A volume of 5.6 mL was sampled from the batch tank every 130-140 seconds. The process was monitored using the set-up described in section 3.5.

Figure 5.1 on page 130, illustrates the PDA dc data (**plot A)** and PEI, homogenate, total volumes, and the PEI volume fraction (x_{PEI}) (**plot B)**. Due to the similar volumes in PEI dosing and sampling small variations in the total volume were observed. The discrete sampling steps taken every 130-140 seconds resulted in the step changes seen in the total and homogenate volume profiles. A similar pattern is seen in the PEI volume but more pronounced during higher volume fraction levels. The PDA dc values decreased with increasing PEI dose, due to the flocculation of particulates, hence increase in turbidity and decrease in light intensity. The PDA dc profile may be divided into three regions. The initial steep flocculation stage (a) where initial addition of PEI had large influence on the level of flocculation. The subsequent levelling off region (b) which occurred after approximately 13 minutes (x_{PEI} = 0.07 mL mL^{-1}) where an increase in the PEI dose had less effect on the level of flocculation and lastly the re-suspension phase (c) where the PDA dc started to increase.

The results obtained from the at-line monitoring setup of product, alcohol-dehydrogenate (ADH), key contaminants, protein, RNA, cell debris and PDA (dc) levels are shown in **Figures 5.2** and **5.3**. Figure 5.2 illustrates the flocculation profiles (53 data points for each profile) in terms of PEI volume fractions (PEI volume/Total volume), whilst figure 5.3 shows the batch profiles with respect to PEI weight ratios (PEI weight/Homogenate weight). As one of the objectives in this chapter is to discuss the rapid monitoring and characterisation of the process, real time information will be applied to represent the flocculation data. Due to the possible fluctuations in homogenate conditions, the only truly known and hence controllable input variable for the batch operations is the volume of flocculant, hence volume fraction terms will be used when characterising the flocculation runs in terms of the models described earlier.

The apparent optimum operating window for good recovery of product and removal of key contaminants is a PEI volume fraction (x_{PEI}) around 0.1 mL mL^{-1} and a weight ratio ($w_{PEI/H}$) of 0.01 mg mg^{-1}. An excess PEI dose may lead to the removal of product, and

eventually the re-suspension of contaminants and product. Over dosing will also have an effect on the operation of the subsequent downstream processes hence it is important to control to the minimum flocculant dose when attempting to run the flocculation process under optimum conditions. The cell debris data increased initially which can be due to the initial flocculation/precipitation of components which result in too small a floc size to be removed in the solid-liquid separation step.

Although the PDA dc data gives us some indication of the flocculation performance, the behaviour of the individual components are not represented in this information. If the flocculant dose was terminated at the minimum PDA dc point (x_{PEI} = 0.2 mL mL^{-1}) over dosing would occur leading to the loss of product. The transition step from the initial steep PDA dc flocculation stage (a) to the less steep region (b) seems to contain information on when to terminate the dosing of flocculant. It should be noted that product and contaminant levels were measured post microcentrifuge separation, whilst PDA dc values were taken prior to this step. Optimum flocculation and subsequent separation conditions need not be similar to those which result in optimum floc size, as flocs will be separated in the microcentrifuge when they have reached a critical size rather than optimum size. Hence we should expect that optimum flocculation conditions after solid-liquid separation to occur before the minimum point in the PDA dc profile (if the PDA dc is giving us a true indication of floc size).

Figures 5.4, 5.5 and **5.6** illustrate the at-line data in terms of yield factors and their respective profile fits using the three models in combination with a non-linear least squares (NLLS) fitting technique. The yield terms refer to spun flocculation samples as a fraction of spun feed. Optimum operating conditions resulted in the removal of approximately 45% of protein, 80% of RNA and 90% of cell debris. It should be noted that higher cell debris removal levels can be expected, as the above at-line yield data are a reflection of the re-suspension and contamination problem associated with the microcentrifuge (see chapter 3 and 4).

The estimated model parameters and coefficient of determination (R^2) for each of the flocculation models (see table 5.1) are listed below the figures. As the empirical 3(2)-parameter model does not allow for the re-suspension phenomena due to excess PEI levels (discussed in chapter 2), only data up to a volume fraction of 0.3 mL mL^{-1} was applied. The "goodness of fit" of the three models can be observed visually or by comparing the R^2 values for each contaminant and product profile. Such a comparison indicated that the 3(2)-parameter model was more appropriate at fitting the real time data. However the re-suspension phenomena was not accounted for. By fitting the at-

line monitored data with a mathematical description one can reduce the amount of information required to characterise a given flocculation run, i.e. the 3(2)-parameter model allowed the specification of the flocculation process with three parameters. By relating the operating conditions to the model parameters it enabled an easy and efficient manner to evaluate the performance of the flocculation process. This analysis technique will be discussed in more detail later.

5.5.2 Batch runs 2 to 8

Figures 5.7 to 5.27 illustrate the at-line data of both product and key contaminants as well as the model fits using the 3(2)-parameter and 4-parameter models for the batch runs 2 to 8. Model fits using the 3(1)-parameter model will not be shown, although a summary of its effectiveness to predict the flocculation data will be given later in table 5.4. This section will discuss the behaviour of the flocculation runs in terms of different environmental (pH, ionic strength) and operating conditions (feed volume, PEI concentration).

Figure 5.7, illustrates the results obtained from the rapid analysis of ADH, protein, RNA, cell debris and PDA dc levels for Batch Run2. It can be observed that the ADH activity increased with increasing flocculant levels. It is well known that the activity of an enzyme is dependent on the environment in which it is in. Factors such as temperature, pH, ionic strength, contaminants/inhibitors, activating agents, etc can influence the activity of an enzyme. Flocculation with PEI removes a large number of contaminants (nucleic acids, protein, cell debris, lipids), alters the balance of ions due to its anionic conformation, might remove anti-activating agents and can flocculate and remove ADH. The activation of ADH due to PEI is therefore a complex mechanism, function of several interactive factors. Activation of ADH has also been observed in the presence of ammonium sulphate (Holwill *et al.* 1997, Smith 1997) and both authors concluded that the mechanism behind the activation was unclear. The activation phenomenon is observed for most of the batch runs.

The start feed (homogenate) volume for batch runs 2 (**figures 5.7 to 5.9**) and 3 (**figures 5.10 to 5.12**) was 700 mL compared to the 500 mL volume used during batch Run1. The larger start volume allowed for a higher resolution of data points during the initial stages of the flocculation profiles, due to a slower increase in the flocculant volume fraction. During the operation of batch runs 2 and 3, approximately 12 measurements were acquired between a PEI volume fraction of 0 – 0.1 mL mL^{-1}, compared to 9 data points during Run1. The resolution of data points in terms of

flocculant volume fraction is a function of the start feed volume, sampling frequency and PEI dose rate. To increase the resolution of the flocculation profiles, one may decide to increase the feed volume, sampling frequency or reduce the flocculant dose rate.

The behaviour of a flocculation profiles is affected by several factors one of which is pH. The lower the pH, the higher the net charge making PEI (a cationic flocculant) more effective in flocculating negatively charged components (Horn 1980, Cordes *et al.* 1990). **Figure 5.10**, illustrates the characterisation of the product and contaminant levels in terms of PEI volume fraction for Batch Run3. The effect of pH change on the behaviour on the flocculation profiles is clearly evident by comparing batch runs 1 and 3 (**figures 5.2 & 5.10** or **figures 5.3 & 5.11**). The apparent optimum operating window for good removal of contaminants for Batch Run3 was a PEI volume fraction (x_{PEI}) of 0.05 mL mL^{-1}, whilst for batch Run1 this was 0.1 mL mL^{-1}. Hence the apparent flocculant dose optimum has halved as a result of a decrease in pH from 6.5 to 5.5. Although a pH increase (pH 6.5 to 7) was seen between batch runs 1 and 2 the apparent optimum flocculant volume fraction still decreased from x_{PEI} = 0.1 mL mL^{-1} to 0.08 mL mL^{-1}. It should be noted that although all other start conditions remained similar (the start volume only effects the mass balancing of the process and not the flocculation mechanism), the composition of the feed in terms of cell debris, RNA and protein concentrations where different for the two batch runs as indicated in table 5.2. The contaminant levels in Batch Run1 were higher than those of Batch Run2 hence a larger amount of PEI dose would have been required.

Figures 5.13 to **5.15** illustrate the at-line data in terms of flocculant volume fraction for Batch Run4. The stock solution of PEI was made to 5 mg mL^{-1} for this run, compared to 10 mg mL^{-1} for batch runs 1, 2, 3. The 2-fold dilution in PEI stock solution has allowed for a higher resolution of data points, not in terms of PEI volume fraction but with respect to the batch run's apparent optimum PEI dose. For batch Run1 (figure 5.2) 8 data points (per profile) were acquired up to the apparent optimum flocculant dose of 0.1 mL mL^{-1}, whilst for Batch Run4 (figure 5.13) 14 data sets were obtained before the apparent flocculant dose optimum (0.16 mL mL^{-1}) was reached. Since PEI flocculates contaminants by directly interacting or essentially titrating anionic charged groups, it was expected that the amount of PEI required to flocculate a given percentage of total contaminants depended upon the ratio of flocculable components to flocculant (Jendrisak 1987). As the PEI stock solution was diluted 2-fold, approximately double as much PEI was required to obtain similar flocculation conditions to that of batch Run1. Note that a pH change (Batch Run4 was run under

pH 6.2) was also introduced making the direct comparison more difficult. An increase in the feed volume (discussed earlier) enables a slower increase in PEI volume fraction due to simple mass balancing and the position of the optimum flocculant dose does not change in terms of flocculant volume fraction. When diluting the PEI stock solution a shift in the flocculation profiles occurs, pushing the flocculation optimum to higher volume fraction levels. Although the resolution of data points in terms of PEI volume fraction remains the same, due to the increase in the optimum flocculant dose more measurements could be acquired in the critical region of the flocculation profiles. Batch runs 5 and 6 (figures 5.16 to 5.21), illustrate flocculation profiles also dosed with a 5 mg mL^{-1} concentration PEI stock solutions. Batch Run5 behaved similarly to batch Run4, however Run6 due to its lower pH conditions (pH = 5) although operating with a diluted PEI stock solution had flocculation profiles similar to Batch Run1. The lower pH conditions made PEI more effective in flocculating the contaminants, shifting the profiles resulting in an apparent optimum PEI volume fraction of 0.1 mL mL^{-1}.

Figures 5.22 to **5.27** illustrate at-line data and model fits for batch runs 7 and 8. During these two operations two additional input variables have been changed, namely the flocculant dose rate and the batch ionic strength. Both batch operations started with a feed volume of 700 mL and an initial linear decreasing dose rate F_{PEI}, where $F_{PEI} = F_0 - \alpha t$, and F_0 = 10 mL min^{-1}, α = 0.01 mL min^{-2} and t was time in minutes. When the dose rate had reached 0 mL min^{-1}, a constant PEI feed rate was introduced at 2.82 mL min^{-1}. The linear decreasing dose rate resulted in a decrease in the number of data points acquired per volume fraction step in the first part of the flocculation profile, i.e. a reduction in profile resolution. During the operation of batch Run8 only 4 measurements were acquired between a PEI volume fraction of 0 - 0.05 mL mL^{-1}, compared to 7 data points during the batch runs 2 and 3.

The effect of ionic strength on the flocculation profiles is best seen by comparing batch runs 3 (figures 5.10 to 5.12) and 8 (figures 5.25 to 5.27). Both batch operations were run under similar pH and PEI stock solution conditions, however the ionic strength during Batch Run8 was 20 mM compared to 50 mM for Run3. From figure 5.25 to 5.27 it can be observed that during batch Run8 ADH was removed form solution at PEI volume fraction above 0.1 mL mL^{-1}. Another effect in the reduction of batch ionic strength is the shift in flocculation profiles. This phenomenon is difficult to see from batch runs 3 and 8 (also runs 5 and 7), however is more clearly observed when comparing the inflection point for each set of batch flocculation profiles. Such plots are illustrated in figures 5.29 to 5.31 and are discussed later. Lowering the ionic strength reduces the interaction of phosphate with PEI and the screening of the contaminants

and product, making PEI more effective as a flocculating agent (Bulmer *et al.* 1994). Bulmer also claims that the total level of cell debris remaining in solution will be reduced with lower ionic strength.

Deliberate pH changes were introduced during batch runs 5, 7 and 8. During the batch operation of Run5, a pH change from pH 6.5 to pH 5.7 was made through the titration of 10M HCL, resulting in a slight decrease in the ADH remaining in solution, see figures 5.16. The deliberate pH changes performed on batch runs 7 (pH 5.9 to pH 6.2, addition of 10 NaOH) and 8 (pH 5.5 to pH 6.6 addition of 10 NaOH) resulted in an increase in pH levels, which was an attempt to re-suspend already flocculated ADH enzyme. The changes in pH in the final stages of these batch runs saw little change in contaminant or product removal, indicating that the complexes formed between PEI and the various compounds are relatively strong compared to the shear forces within the batch tank.

5.5.3 Flocculation process characterisation

Tables 5.3, 5.4 and **5.5** show an overview of the eight batch runs in terms of pH, ionic strength, start contaminant and product concentrations and the estimated model parameters for the three mathematical flocculation descriptions discussed earlier. The coefficient of determination for each batch and flocculation profile is listed indicating the effectiveness of the three models in fitting the at-line data. Several different minimum coefficient of determination values were set and percentage failures of the individual models to achieve this fitting criterion was plotted, **figure 5.28**. It can clearly be seen that the 3(1)-parameter model was the least effective in fitting the at-line data, followed by the 4-parameter and 3(2)-parameter model. Although the 3(2)-parameter model seems to be the most effective in describing the flocculation behaviour, the 4-parameter model has the ability to describe the re-suspension phenomenon which occurs at high floc dosages and therefore should be applied if a total flocculation profile is required. For real time process optimisation and control the 3(2)-parameter model is most appropriate due to its efficient fitting of at-line data in the area of interest and the fewer number of model parameters making real time parameter estimation simpler and faster.

Characterisation of a process through a mathematical description enables an efficient and simple method to analyse the performance of the operation in terms of a few factors (model parameters) quantitatively. **Figures 5.29, 5.30** and **5.31** show how pH and initial contaminant concentrations in the feed influenced the maximum contaminant

removal level (parameter A) and the point of inflection (~ parameter B, when $C > 3.5$) for the contaminant profiles. In order to compare batch runs 4, 5 and 6 with the rest of the batch runs, the inflection point for these operations were divided by a factor 2 to account for the 2-fold dilution in their PEI stock solution.

As illustrated from the figures there seems to be a strong correlation between the pH levels and the point of inflection for all three contaminants. Low pH levels facilitates contaminant flocculation and therefore reduces the point of profile inflection. Batch Run2 however does not seem to follow this concept. Lower feed concentrations of contaminants during this run may have compensated for the higher pH level. The maximum percentage removal of contaminants does not appear to be influenced by pH or feed contaminant levels. It should be noted that batch runs 7 and 8 were carried out at lower ionic strength levels, which appears to result in higher maximum removal levels for all the contaminants. The influence of ionic strength on the final level of contaminant removal was also observed by Bulmer *et al.* (1994). The position of the inflection points for batch runs 7 and 8 occurred at lower PEI volume fraction levels compared to batch runs 5 and 3, which had similar operating conditions apart from higher ionic strength levels.

Figure 5.32, illustrates the maximum fraction of contaminant removed (**plot A**) and the point of profile inflection for the eight batch runs. Cell debris, RNA and protein removal levels lie in the range of 82-91%, 57-82% and 47-66%. **Plot B** illustrates the inflection point of the contaminant profiles, showing that for all the batch runs RNA, protein and then cell debris is flocculated / titrated out of the homogenate solution in that order. The variation in the inflection point is small at low pH values due the increase in the efficiency of PEI to flocculate the contaminants narrowing the optimum operating window. The flocculation order of the contaminants indicate that there may be a specific type/band of RNA and protein that aggregate due to PEI, whilst cell debris has a wider range of sizes and charges.

A linear correlation between batch pH conditions and the point of contaminant profile inflection is illustrated in **figure 5.33**. The rapid at-line monitoring has allowed for a higher resolution of data points especially in the critical flocculation regions enabling such a correlation to be studied. Batch runs 7 and 8 were not included due to their lower ionic strength in comparison to the other six batch runs. As there seems to be a strong correlation between pH levels and the resultant flocculation behaviour, pH may be applied as a manipulative variable for process control. In simulation studies "what if" scenarios may be run allowing the design engineer to get a good

prediction/understanding of how the flocculation process would behave under specific operating conditions.

5.6 Discussion

Real time monitoring of the batch flocculation process has been demonstrated in terms of product and key contaminants. Rapid monitoring allows the luxury of a higher resolution of data points for process characterisation. The total processing time for batch Run1 was approximately 2 hours. To acquire the same amount of information on product and key contaminant levels as that of at-line data, one can expect several hours/days post run off-line measurements. Hence real-time monitoring results in considerable cost saving in terms of time. During the operation and real-time monitoring of the batch runs several factors can be altered to allow for even higher resolution of data points around critical areas of the flocculation process. The dose rate may be altered to enable a more gradual addition of flocculant and therefore a slower change in flocculation conditions. The start volume can be made larger and/or the stock solution of PEI can be diluted. However the dose rate is constrained by the dose pump minimum and maximum possible flow rates. The start volume is restricted to the working volume of the batch tank.

The batch processes in this chapter were characterised in terms of three mathematical models. The most appropriate mathematical description to apply for flocculation characterisation is dependent on how and for what the model is used for. For post process analysis a complex model can be applied where there are no time constraints and a large number of data is available. However for real time process characterisation with initially a small number of data points a simple model with few model parameters is more suitable. From the three models investigated it can be concluded that the 3(2)-parameter model is the most appropriate for real time application. It has only three model parameters that require identification compared to that of the 4-parameter model and results in a better fit to process data than that of the 3(1)-parameter model. The goodness of fit was determined through a coefficient of determination (R^2), indicating the fraction of data explained by the model. Although the 3(2)-Parameter model is an empirical model the parameters describe important features in the flocculation profiles, such as maximum level of contaminants removal and position of the point of inflection of the profiles. However it should be noted that this model does not have the ability to describe the re-suspension phenomenon which occurs during high PEI dose levels. The 4-parameter model would be a more suitable function to implement for total flocculation profile fitting.

Characterisation of a process in terms of a mathematical description enables an efficient and simple method to analyse process information in terms of a few factors (model parameters) quantitatively. From such analysis we have seen that for the given batch runs pH has a linear response on the point of contaminant profile inflection and hence on the appropriate flocculant dose for process optimisation. However there does not seem to be a correlation between the start conditions chosen or pH levels and the maximum level of contaminant removal. For a full study on the effect of ionic strength levels on flocculation behaviour additional batch operations need to be performed for different salt concentrations.

Due to the linear relationship between pH levels and contaminant profile inflection point, regulation of pH conditions may be used as a manipulative variable for the control of the flocculation process. The pH level does not seem to influence the total level of contaminant removal and does not seem specific to individual components. Therefore the most appropriate use of pH regulation for control should be undertaken before or after the addition of the flocculating agent. For operations where feed and flocculant make up fluctuations are large, high pH conditions may be adopted to allow for a larger operating window facilitating correct flocculant dosing. For processes where batch to batch fluctuations are small, low pH levels may be applied tightening the operating window and reducing the required flocculant volumes. Operational requirements may state that the flocculant concentration should remain within a pre-specified level, hence if flocculant control is not adequate in driving the process to an appropriate operating condition, pH control may be implemented. Note if pH control is used shear levels which result in the reversibility of the flocculant-contaminant complex would have to be applied.

The application of the photometric dispersion analyser for the monitoring of floc levels gave some indication of how the flocculation process was performing. However the PDA data did not contain information on behaviour of individual components during the process run. Further work on correlating the slope of the PDA profile to optimum operating conditions could give the operator a rapid technique to identify appropriate dosing levels. The main problem in applying the photometric dispersion analyser is its sensitivity to too turbid solutions such as the feed homogenate which is used for the batch operations. Due to this the PDA instrument was applied in a different manner to that recommended. For the batch processes light intensity readings at a wavelength of 820 nm were taken and used as an indication of floc levels (described in more detail in chapter 3). The larger the flocs the more light was absorbed and hence a lower light intensity was monitored in terms direct current (dc). The recommended manner of

operation is to monitor the direct current as well as the fluctuating ac signal, which has been shown to be a sensitive indicator of the state of aggregation of the suspension (Gregory and Nelson 1984). However for the level of solids concentration applied in this work the fluctuating signal was very erratic/unstable and hence only PDA light intensity measurement (PDA dc) could be applied.

Most of the batch flocculation runs saw some level of product (alcohol dehydrogenase) activation. ADH activation has been experienced during ammonium sulphate addition (Smith 1997). The mechanism of ADH activation due to PEI is unclear. Holwill *et al* (1997) experienced similar problems during ammonium sulphate precipitation of clarified Baker's yeast homogenate. Their solution was to monitor both spun and un-spun samples and by taking the ratio of these two measurements an account for ADH activation could be made. For the flocculation process this method was not seen to be suitable due to the presence of insoluble particulate/cell debris which could effectively dilute the concentration of the ADH compared with the analysis of the clarified flocculation samples. As the above authors were monitoring clarified homogenate the change in solid volume was relatively small. There is no clarification step prior to the flocculation process hence by comparing spun and un-spun samples account of the volume of solids removed would have had to be made.

5.7 Conclusion

Rapid monitoring of product and key contaminants has been demonstrated for eight batch flocculation runs. Variable start (environmental) flocculation conditions in terms of pH, ionic strength, PEI stock solution, PEI dose rate and start feed volume were carried out and characterised in real time through the at-line data.

Three mathematical descriptions have been investigated in their ability to describe the flocculation behaviour in terms of flocculant volume fractions (dose), where the model parameters are related to the environmental conditions. An empirical model, the 3(2)-parameter model was concluded to be most appropriate for real time characterisation due to its effectiveness to describe the flocculation behaviour with the least number of model parameters.

Finally the performance of the flocculation process was analysed quantitatively in a simple and efficient manner through the use of the model. A linear relationship between pH levels and flocculation behaviour was observed.

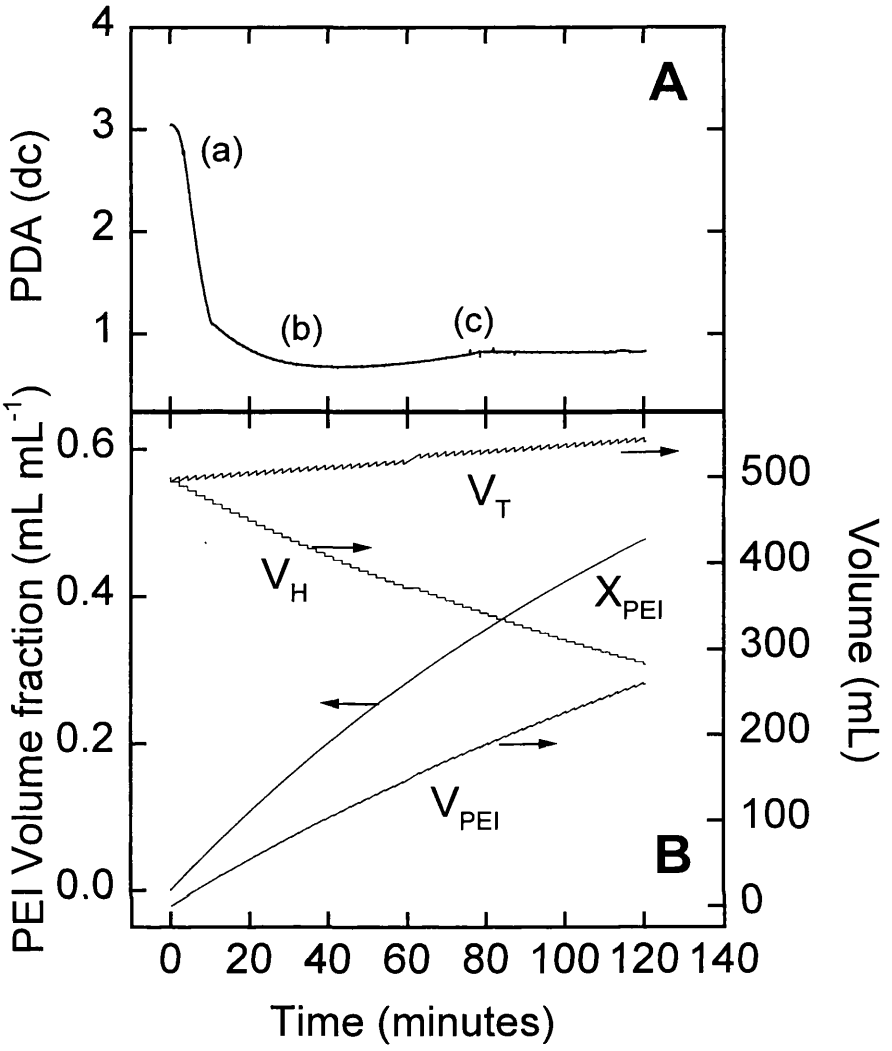


Figure 5.1. *Batch Run1 characterisation in terms of PDA, homogenate and PEI volumes.*

Plot A. PDA direct current (PDA dc) response with respect to time.

Plot B. Characterisation of PEI volume fraction x_{PEI} , homogenate volume V_H, PEI volume V_{PEI} and total batch volume V_T with respect to time (minutes) determined from equations 5.2-5.6. The homogenate solids concentration was 125 g(wet weight) L^{-1} and the PEI solution was 10 mg mL^{-1}. Both solution were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) were automatically taken off every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min.

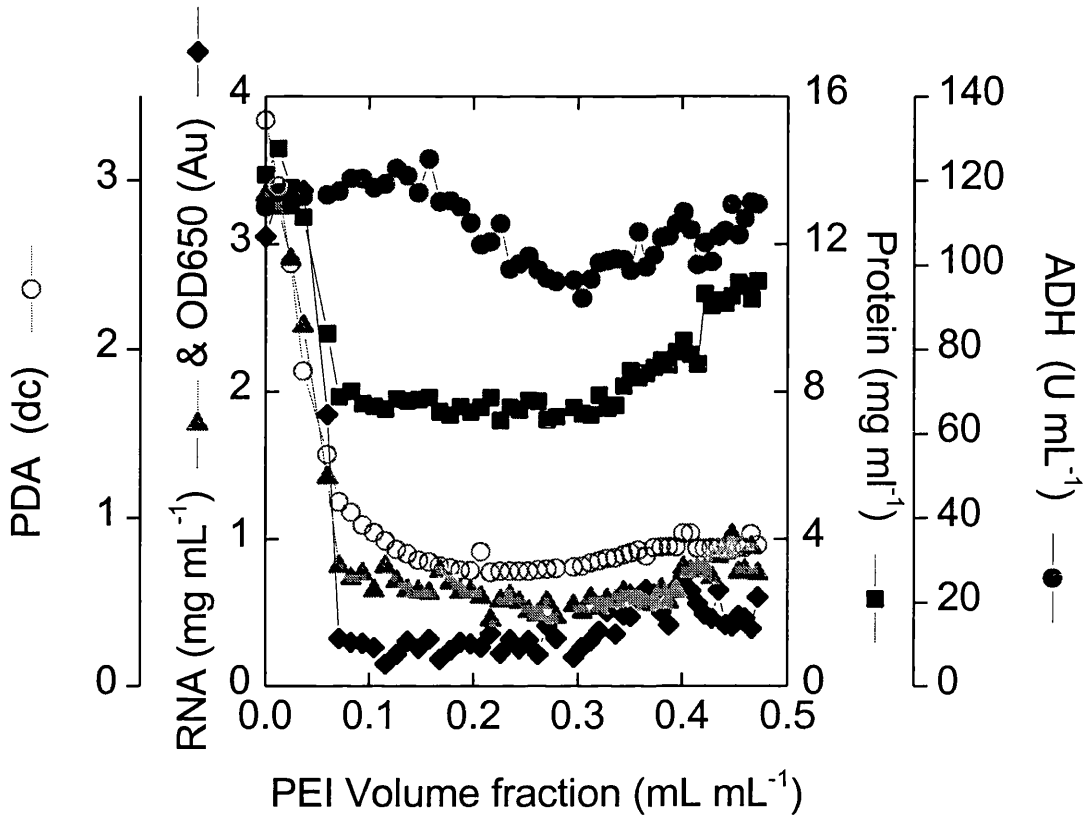


Figure 5.2. *Characterisation of Batch Run1 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

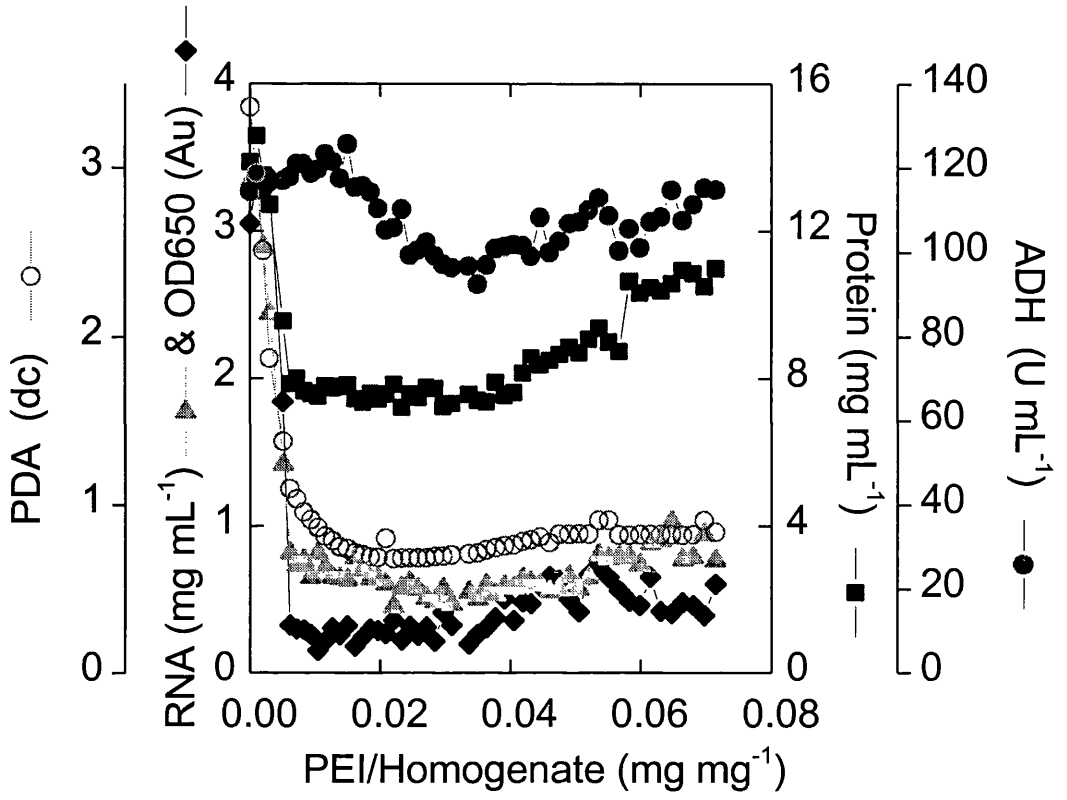


***Figure 5.3**. Characterisation of Batch Run1 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI-homogenate weight ratio (PEI weight / Homogenate weight). Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10*

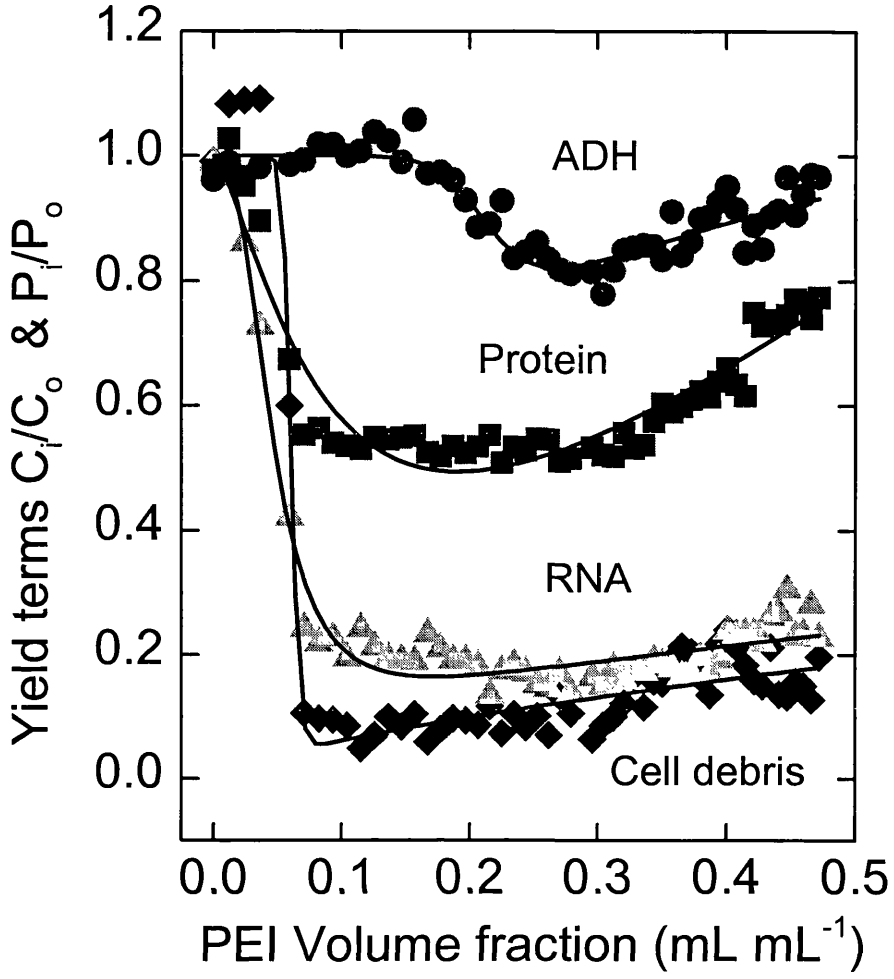


Figure 5.4. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run1. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH	0.219	0.224	10.76	11.04	0.994
Protein	0.102	0.224	1.39	2.44	0.839
RNA	0.047	0.108	2.68	3.18	0.968
Cell Debris	0.061	0.071	19.71	20.41	0.979

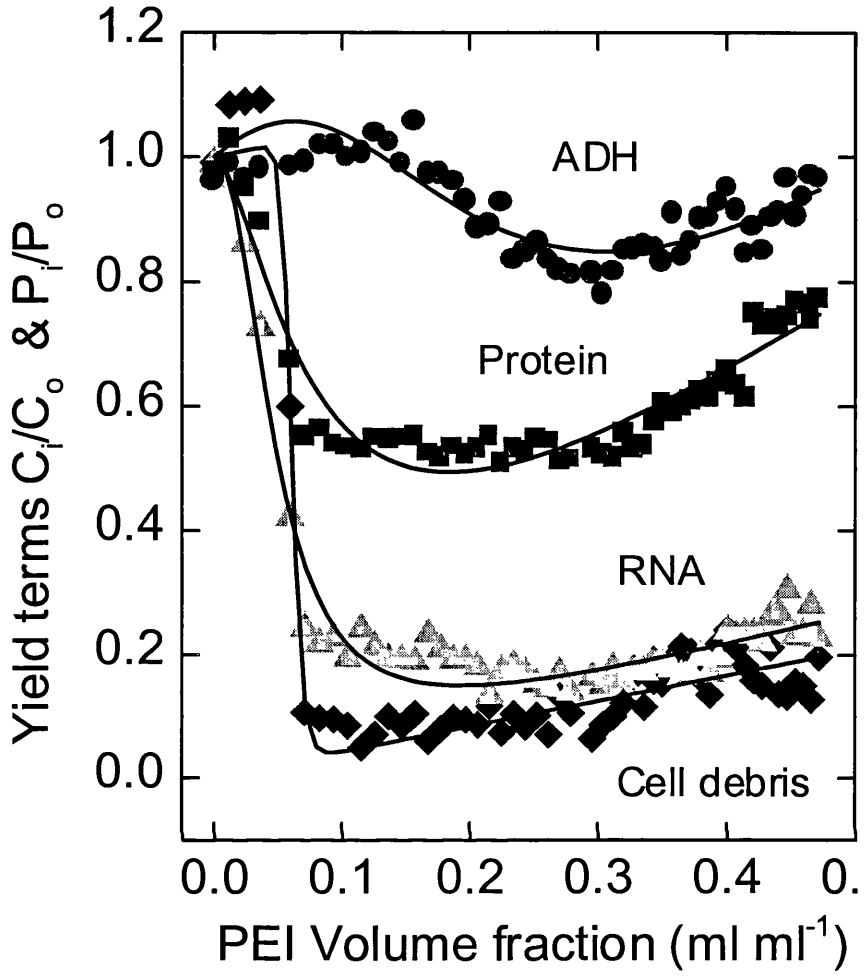


Figure 5.5. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(1)-parameter model for Batch Run1. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	n	c	R^2
ADH	0.228	2.31	1.68	0.989
Protein	0.082	1.51	1.44	0.842
RNA	0.049	2.15	0.52	0.964
Cell Debris	0.061	14.84	0.416	0.979

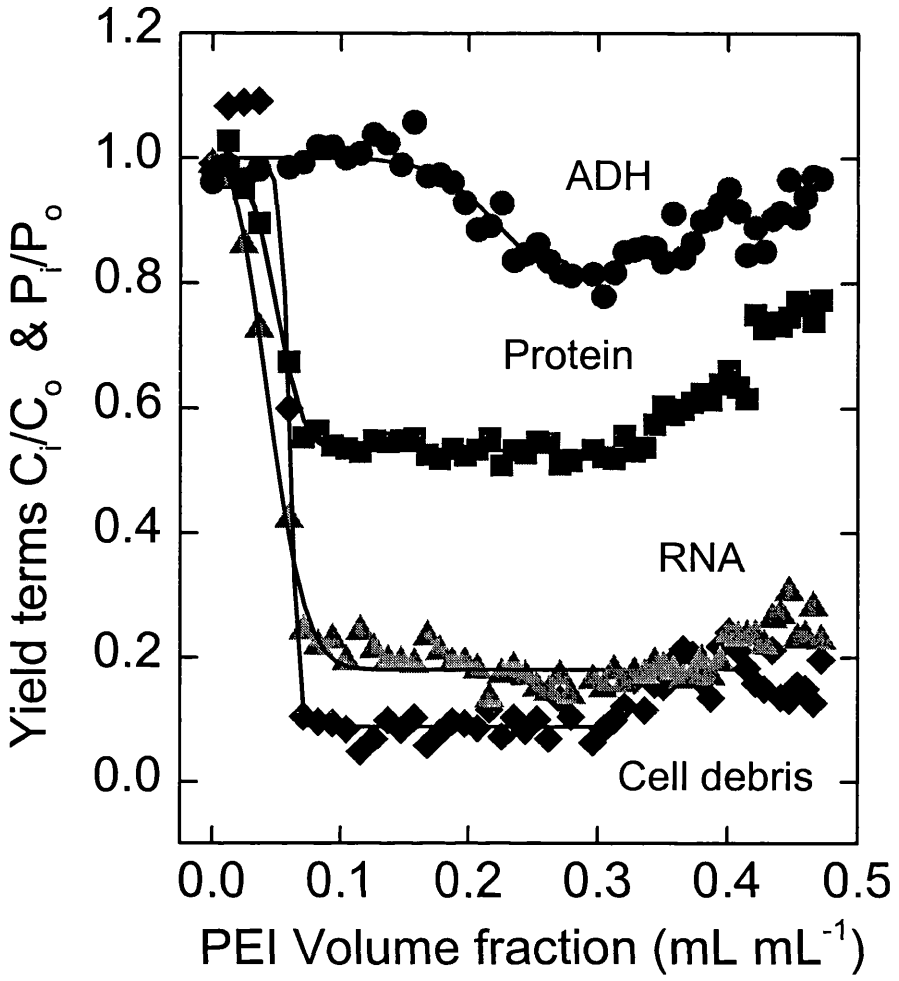


Figure 5.6. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run1. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 6.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 120 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination). The least squares model fitting only implemented data up to a volume fraction of 0.3 (mL mL^{-1})*

	A	B	C	R^2
ADH	0.190	0.228	7.1	0.995
Protein	0.468	0.055	3.24	0.973
RNA	0.82	0.052	2.32	0.977
Cell Debris	0.911	0.062	12.11	0.983

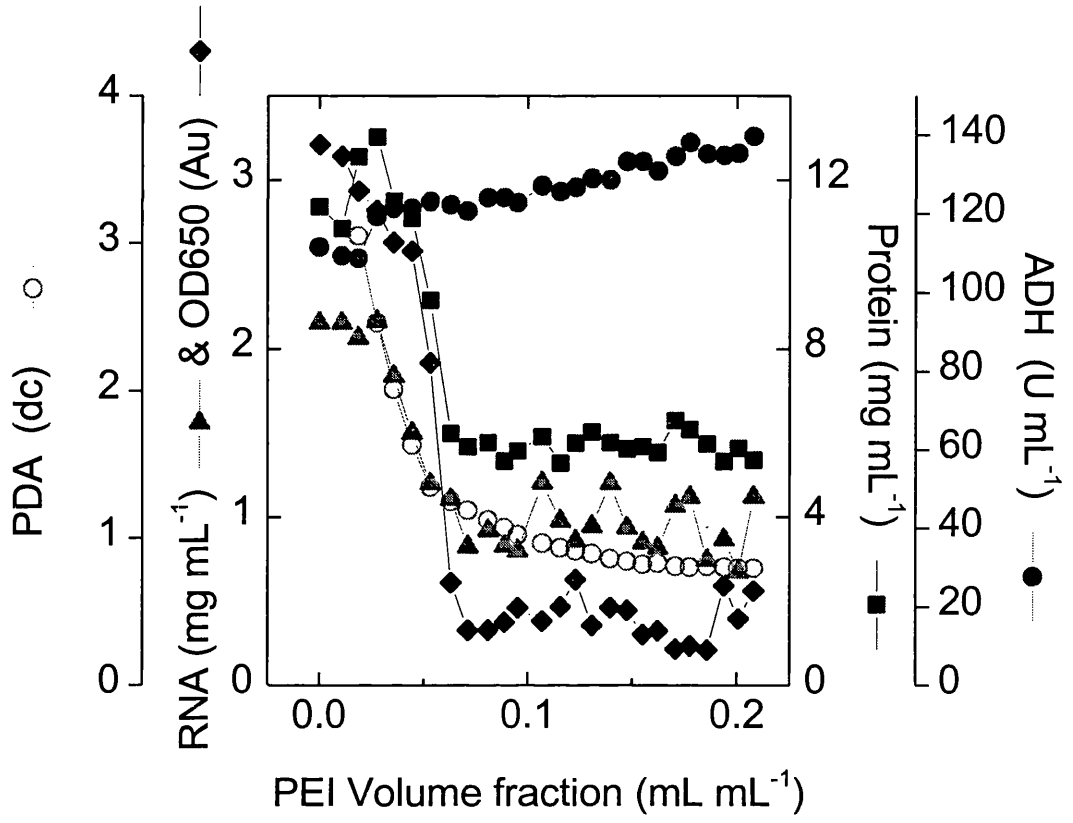


Figure 5.7. *Characterisation of Batch Run2 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 7 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

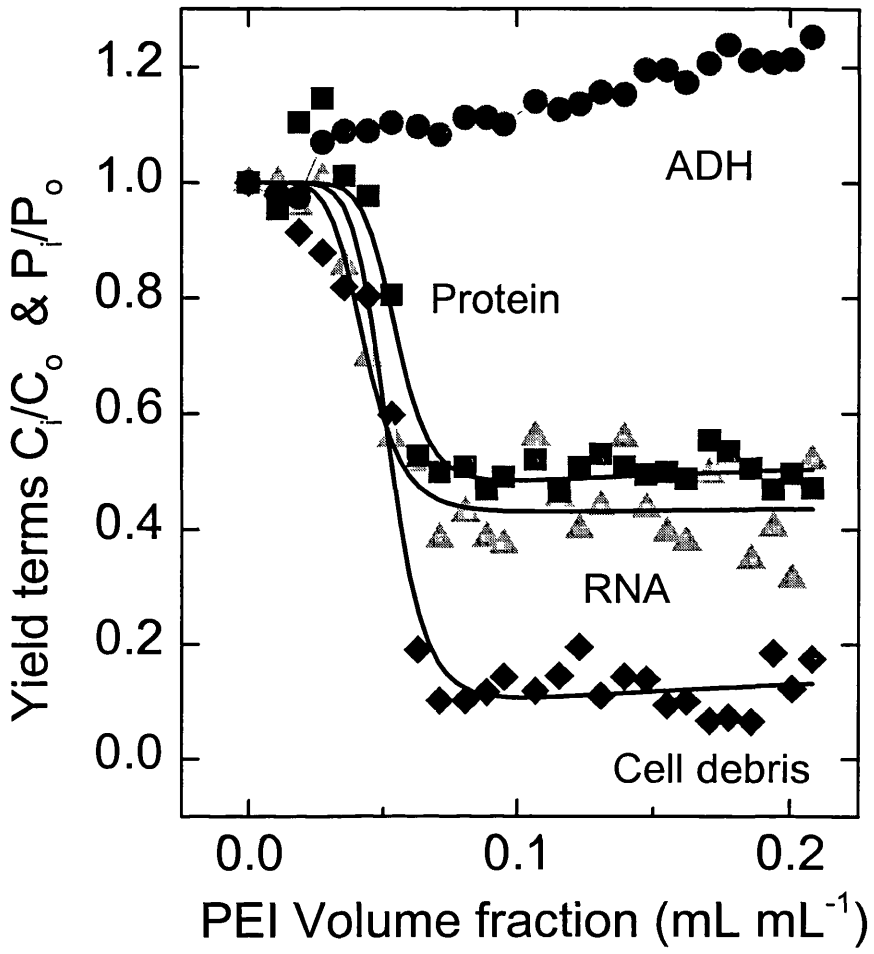


Figure 5.8. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run2. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 7 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH					
Protein	0.055	0.06	8.58	8.65	0.995
RNA	0.044	0.05	6.27	6.29	0.989
Cell Debris	0.052	0.07	7.81	8.20	0.971

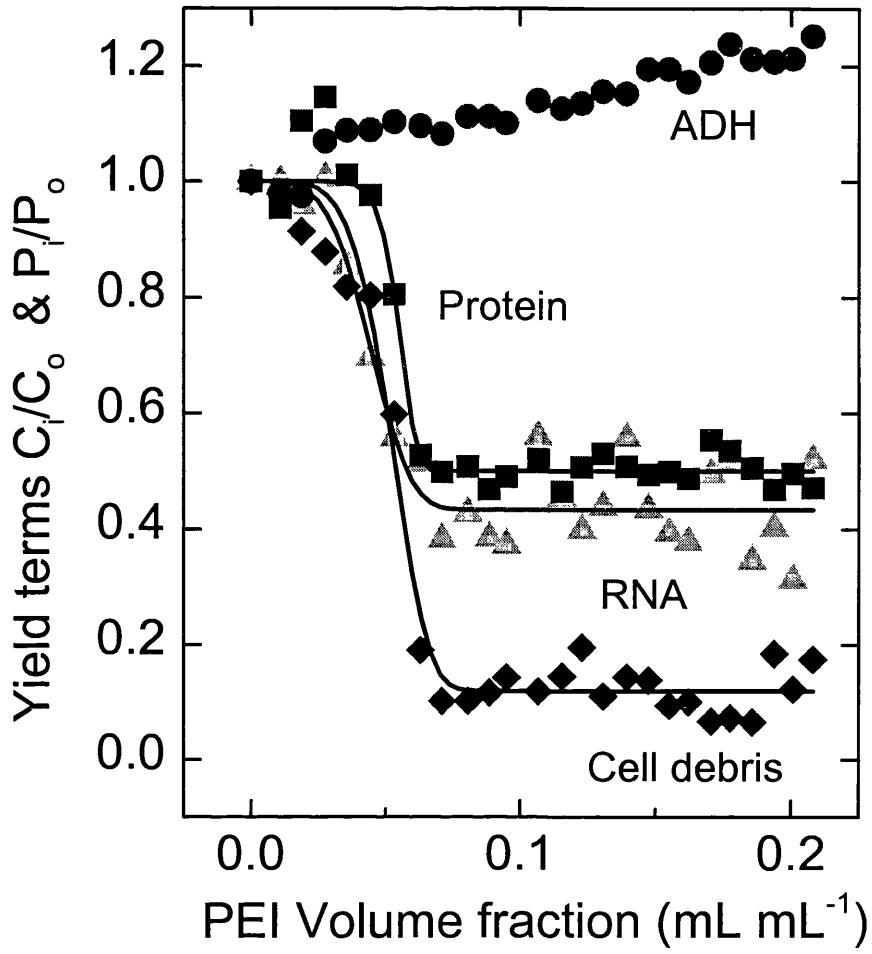


Figure 5.9. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run2. Baker's yeast homogenate (700 mL at 125 (wet weight) g L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 7 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH				
Protein	0.5	0.057	11.2	0.996
RNA	0.57	0.048	4.17	0.999
Cell Debris	0.88	0.056	5.3	0.978

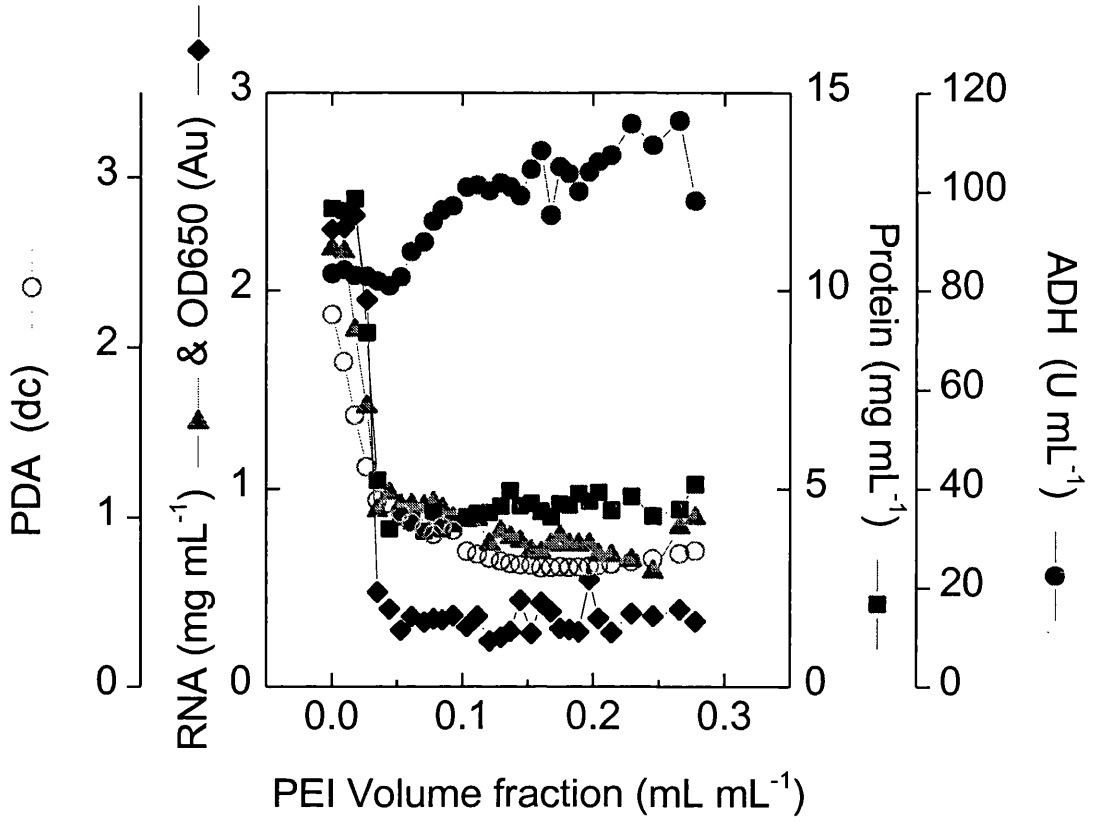


Figure 5.10. *Characterisation of Batch Run3 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction mL mL^{-1}. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 5.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 85 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

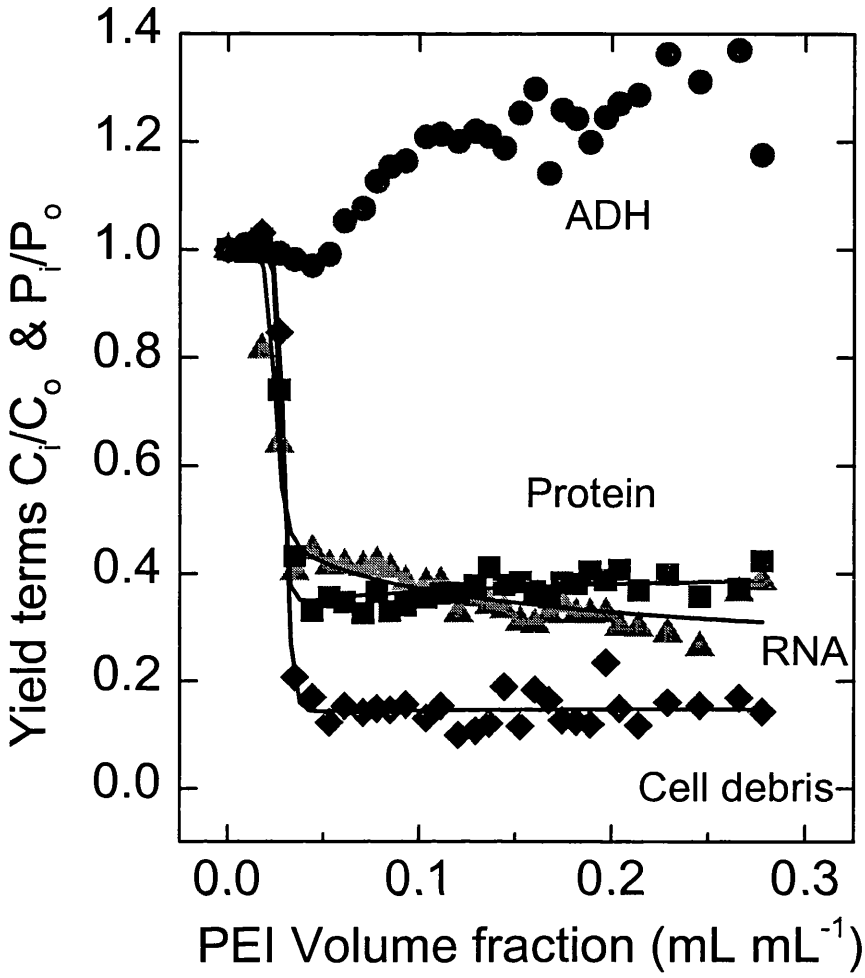


Figure 5.11. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run3. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 5.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 85 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH					
Protein	0.028	0.029	16.79	16.84	0.987
RNA	0.024	0.026	10.26	10.08	0.978
Cell Debris	0.03	0.033	16.2	16.19	0.977

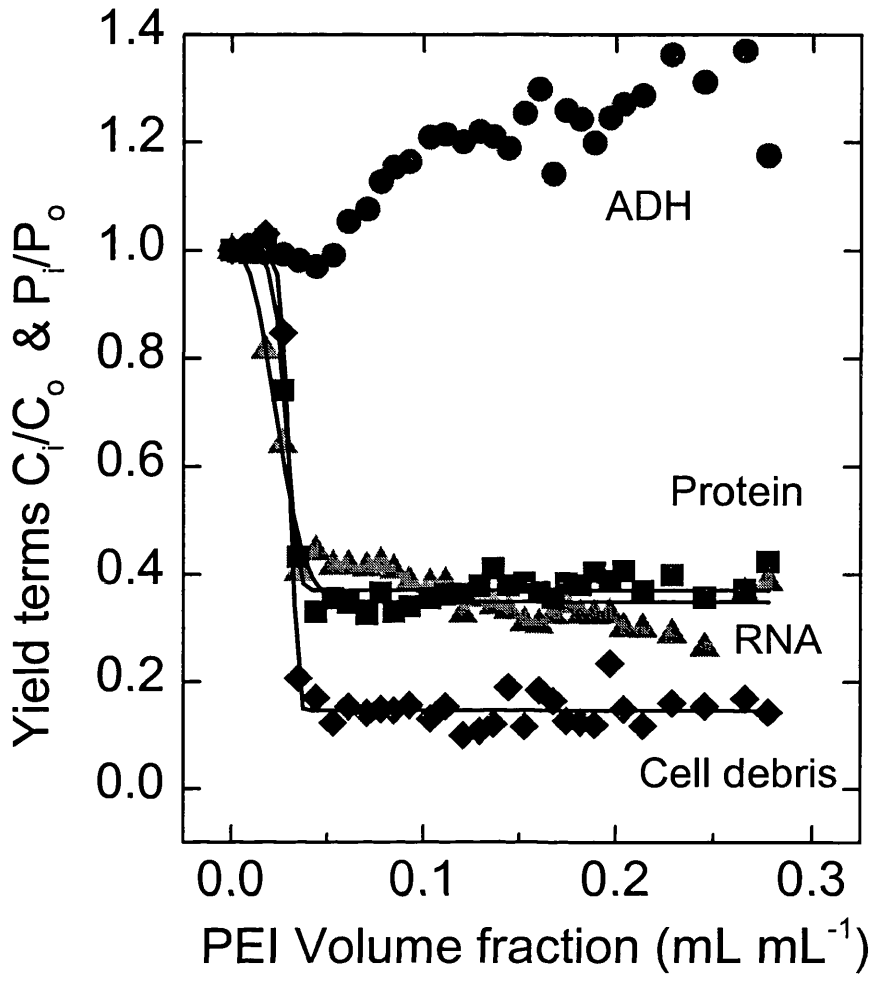


Figure 5.12. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run3. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}) was flocculated using PEI flocculant (10 mg mL^{-1}). Both solutions were at pH 5.5 and at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 85 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH				
Protein	0.63	0.03	6.16	0.984
RNA	0.65	0.028	2.44	0.975
Cell Debris	0.85	0.032	9.84	0.979

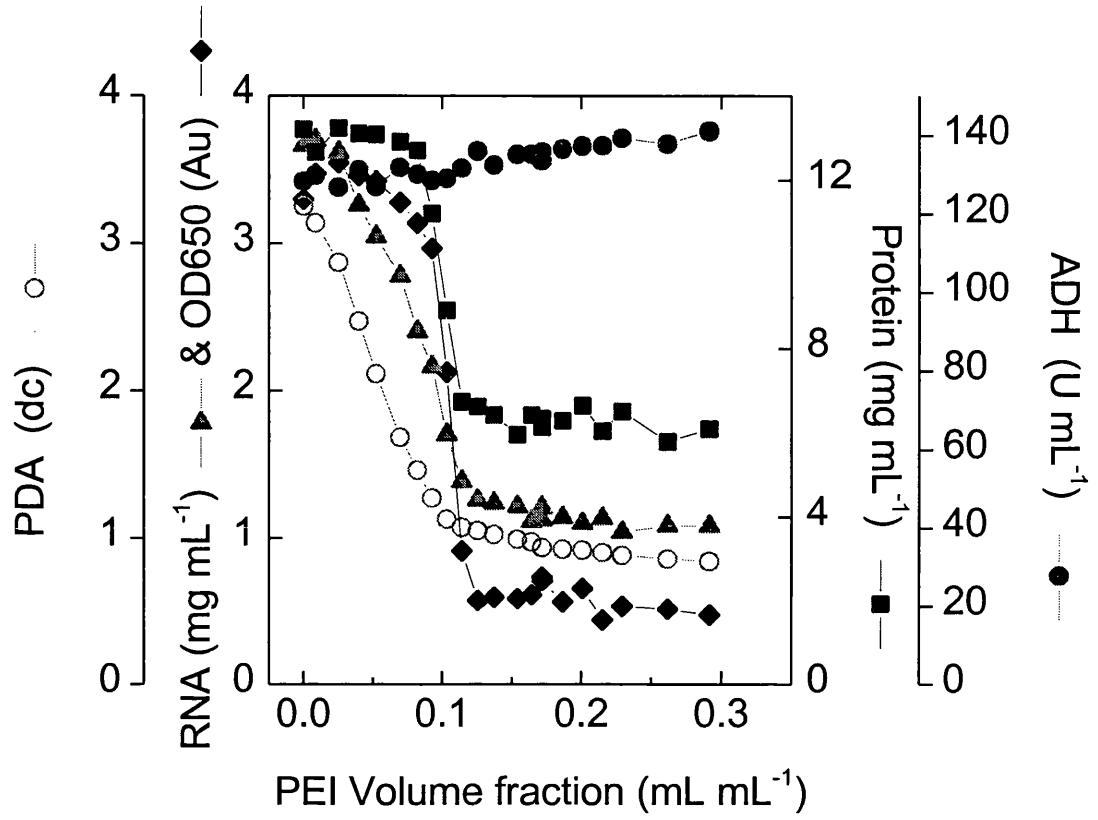


Figure 5.13. *Characterisation of Batch Run4 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

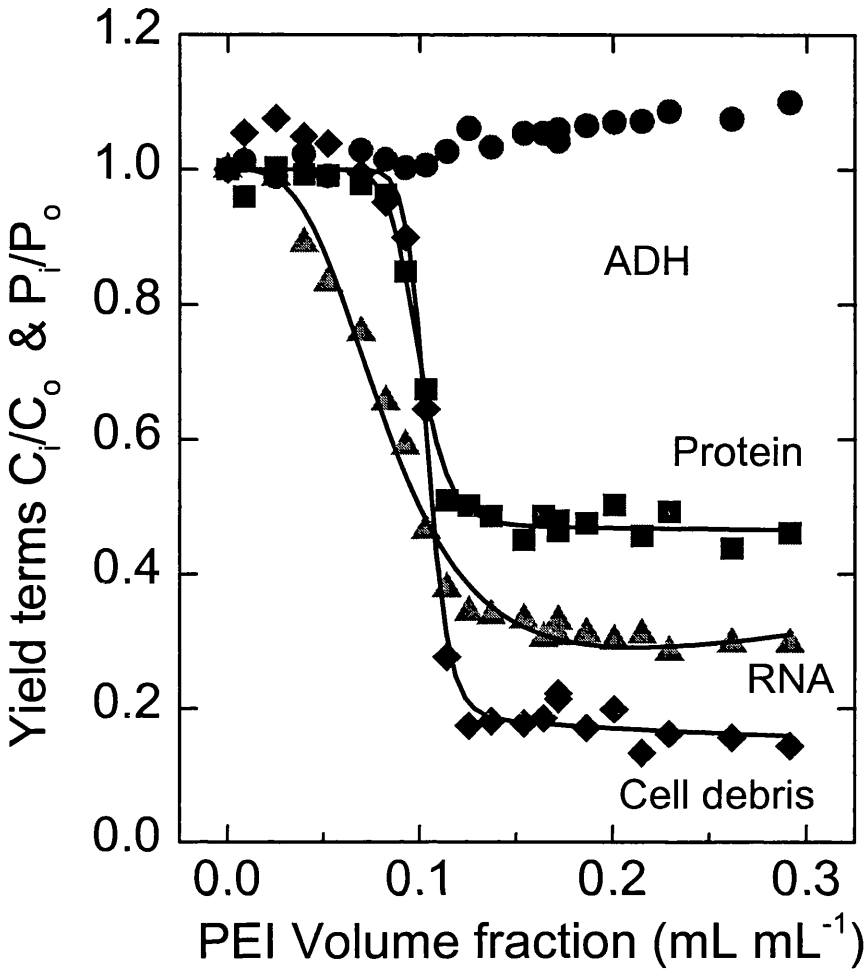


Figure 5.14. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter for model Batch Run4. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	N	m	R^2
ADH					
Protein	0.099	0.105	12.69	12.67	0.994
RNA	0.088	0.143	3.28	3.79	0.991
Cell Debris	0.104	0.113	19.03	18.84	0.993

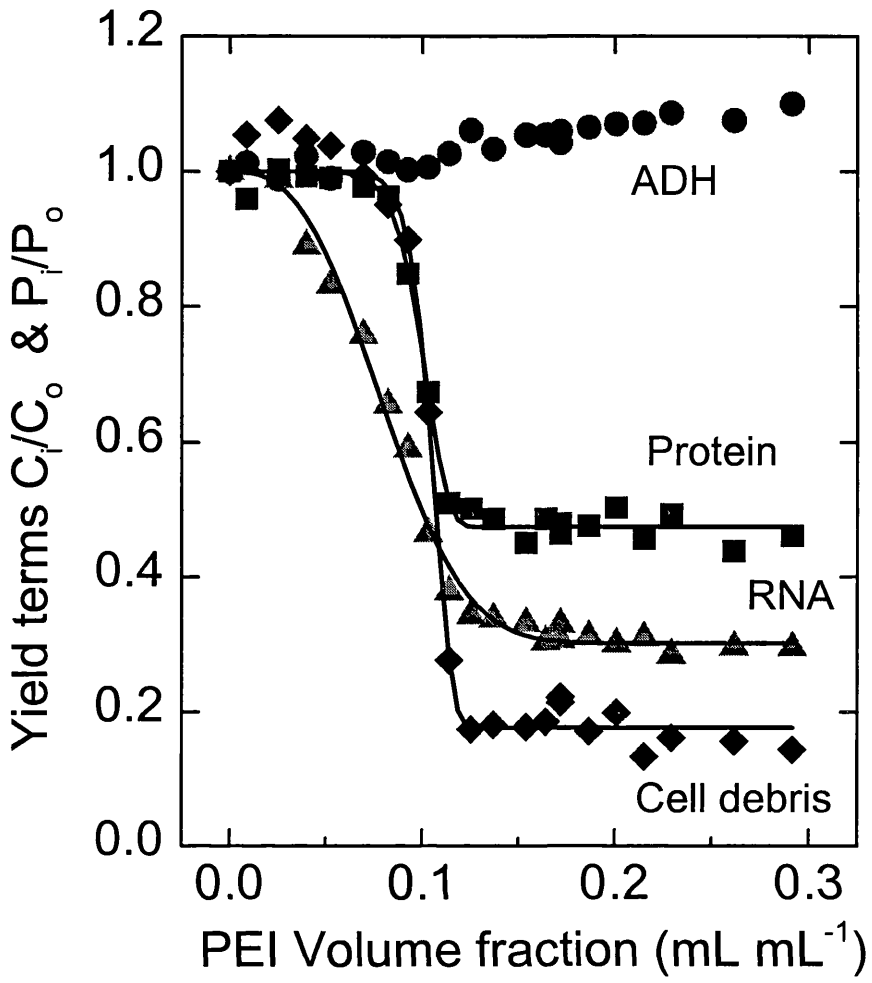


Figure 5.15. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run4. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 60 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH				
Protein	0.52	0.104	10.05	0.994
RNA	0.7	0.092	2.72	0.994
Cell Debris	0.82	0.108	12.97	0.994

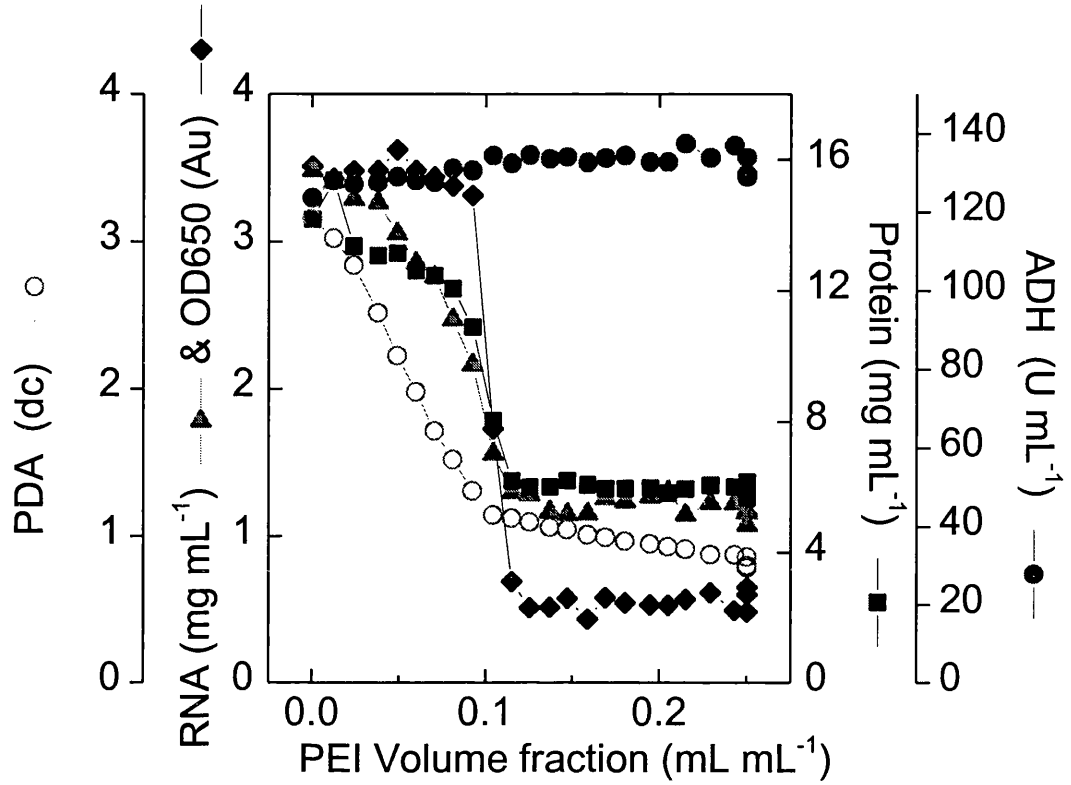


Figure 5.16. *Characterisation of Batch Run5 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 55 min. A pH change (6.5 to 5.7, addition of 10M HCL) was made during the last two data points. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

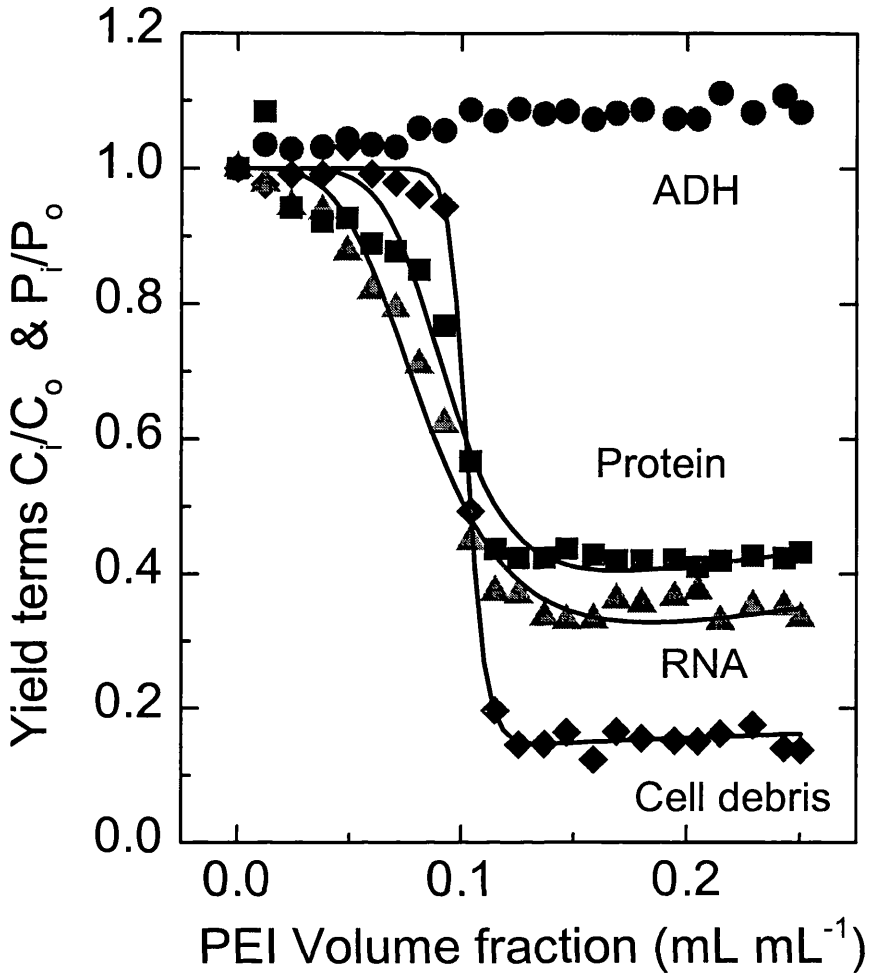


Figure 5.17. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run5. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 55 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH					
Protein	0.095	0.112	6.27	6.53	0.971
RNA	0.088	0.125	4.02	4.5	0.982
Cell Debris	0.103	0.111	24.2	24.3	0.998

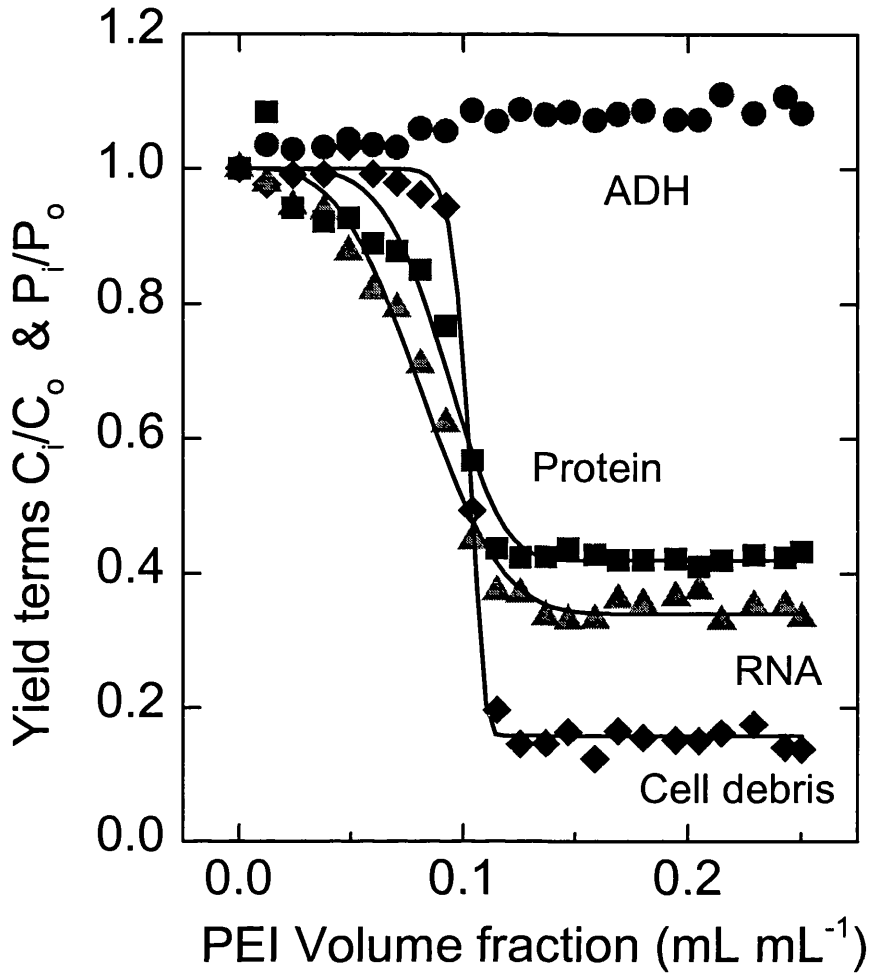


Figure 5.18. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run5. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 6.2) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5.9). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 55 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH				
Protein	0.58	0.099	5.01	0.977
RNA	0.66	0.091	3.31	0.989
Cell Debris	0.84	0.105	20.05	0.998

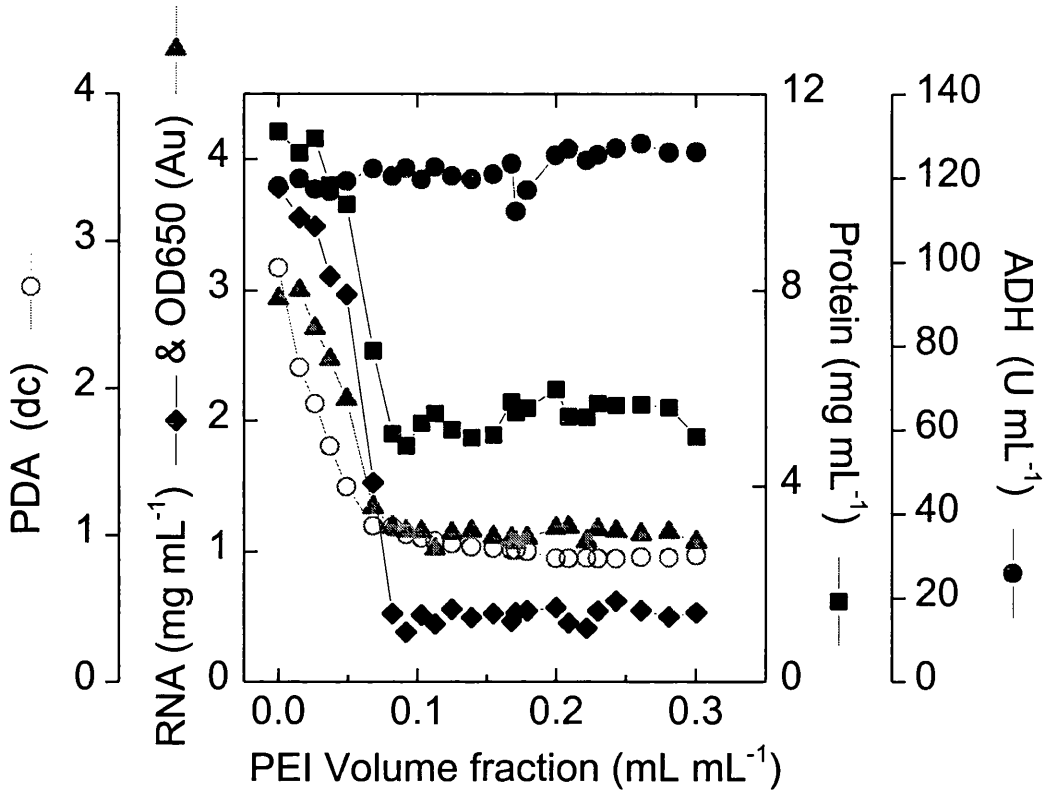


Figure 5.19. *Characterisation of Batch Run6 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 70 min. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

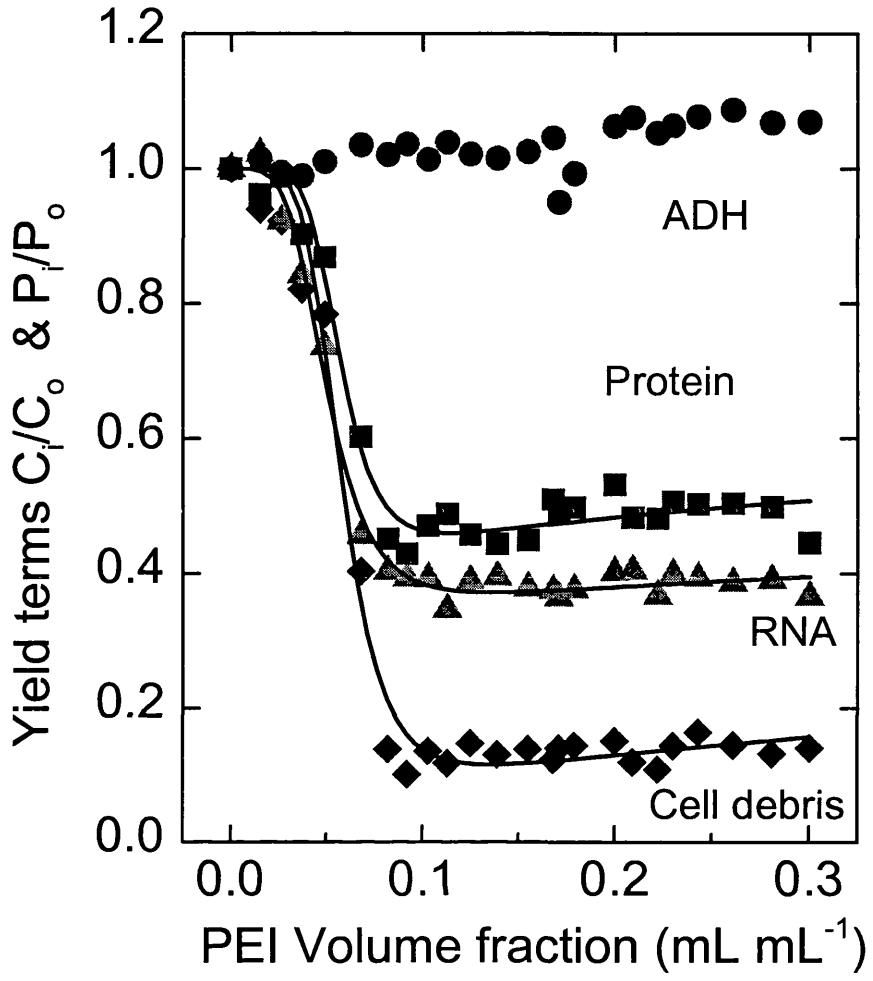


Figure 5.20. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run6. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 75 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH					
Protein	0.058	0.067	5.97	6.09	0.977
RNA	0.051	0.066	4.32	4.42	0.991
Cell Debris	0.058	0.092	537	5.86	0.984

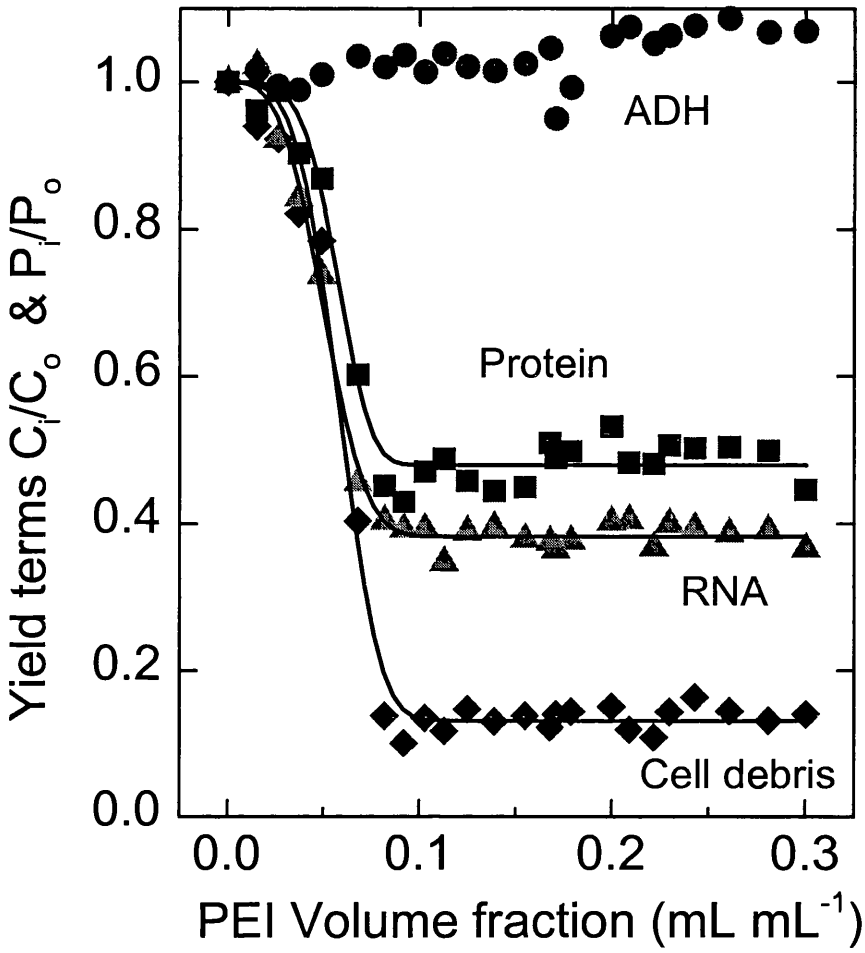


Figure 5.21. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run6. Baker's yeast homogenate (500 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (5 mg mL^{-1}, pH 5). Both solutions were at an ionic strength of 50mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 2.82 mL min^{-1}. The total run time was approximately 75 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH				
Protein	0.52	0.062	4.54	0.978
RNA	0.62	0.056	3.25	0.993
Cell Debris	0.87	0.064	3.85	0.990

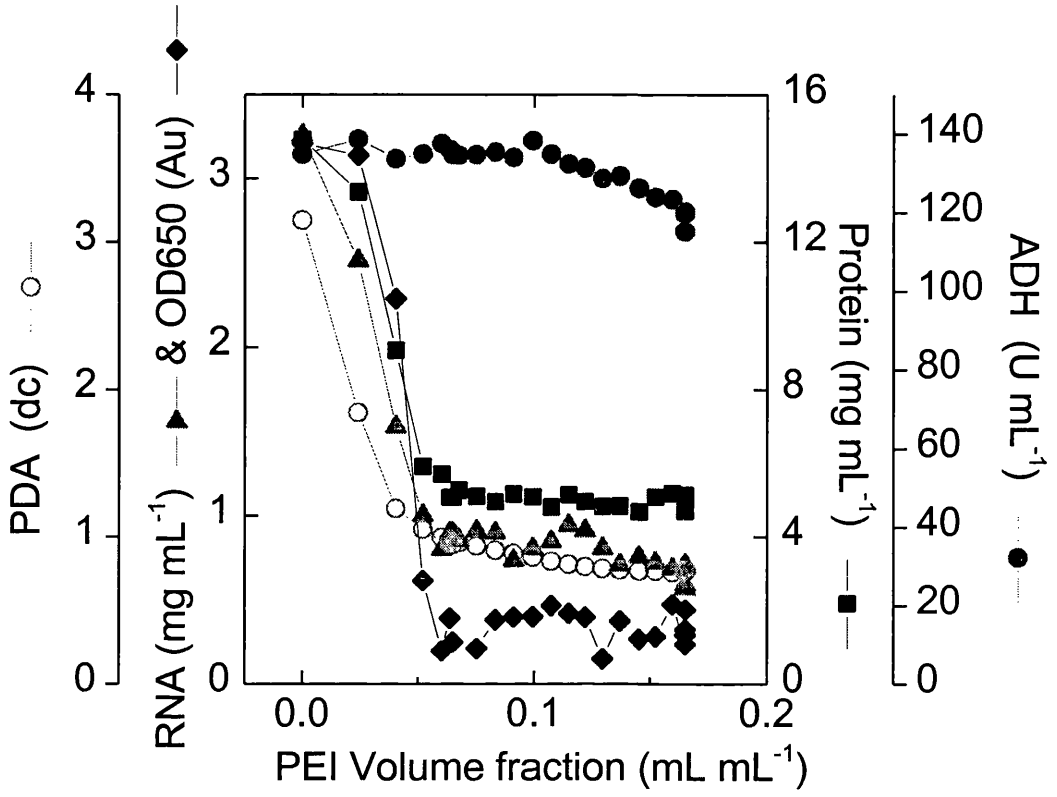


Figure 5.22. *Characterisation of Batch Run7 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 6) was flocculated using PEI flocculant (10 mg mL^{-1}, pH 6). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 55 min. A pH change (5.9 to 6.2, addition of 10M NaOH) was made during the last two data points. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

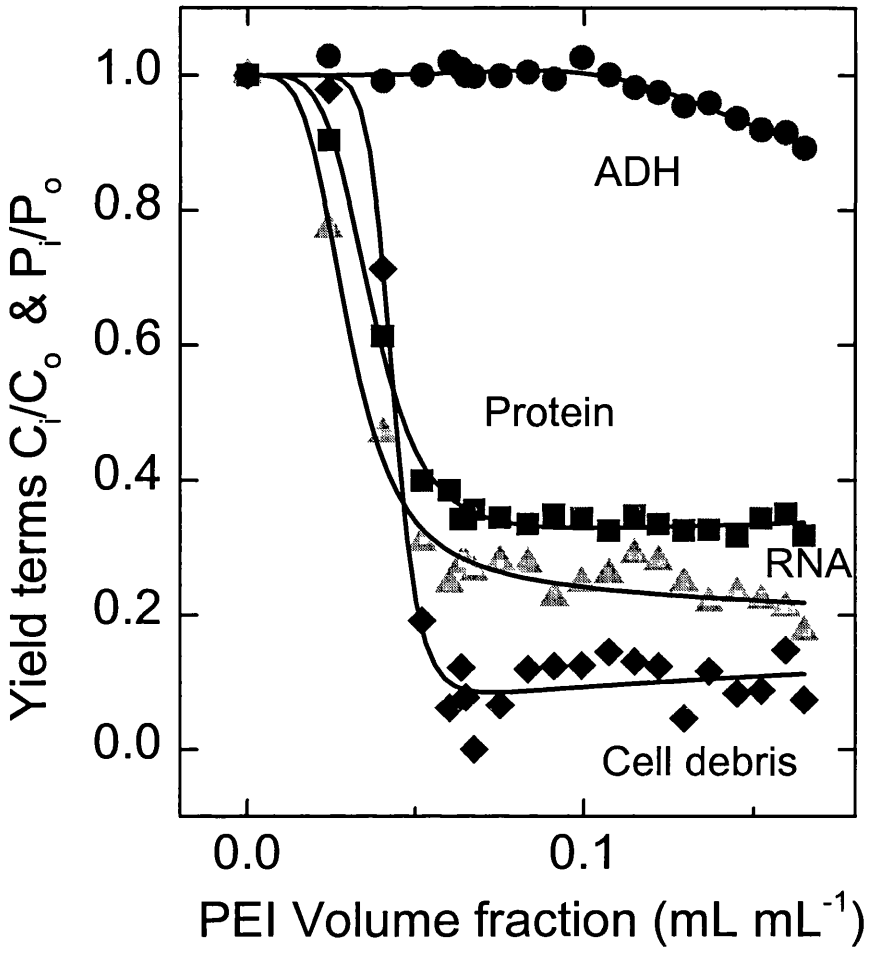


Figure 5.23. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run7. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 6) was flocculated using PEI flocculant (10 mg mL^{-1}, pH 6). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 55 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH	0.124	0.125	5.45	5.15	0.995
Protein	0.037	0.048	4.8	4.88	0.994
RNA	0.029	0.04	4.09	3.93	0.981
Cell Debris	0.043	0.055	11.26	11.64	0.985

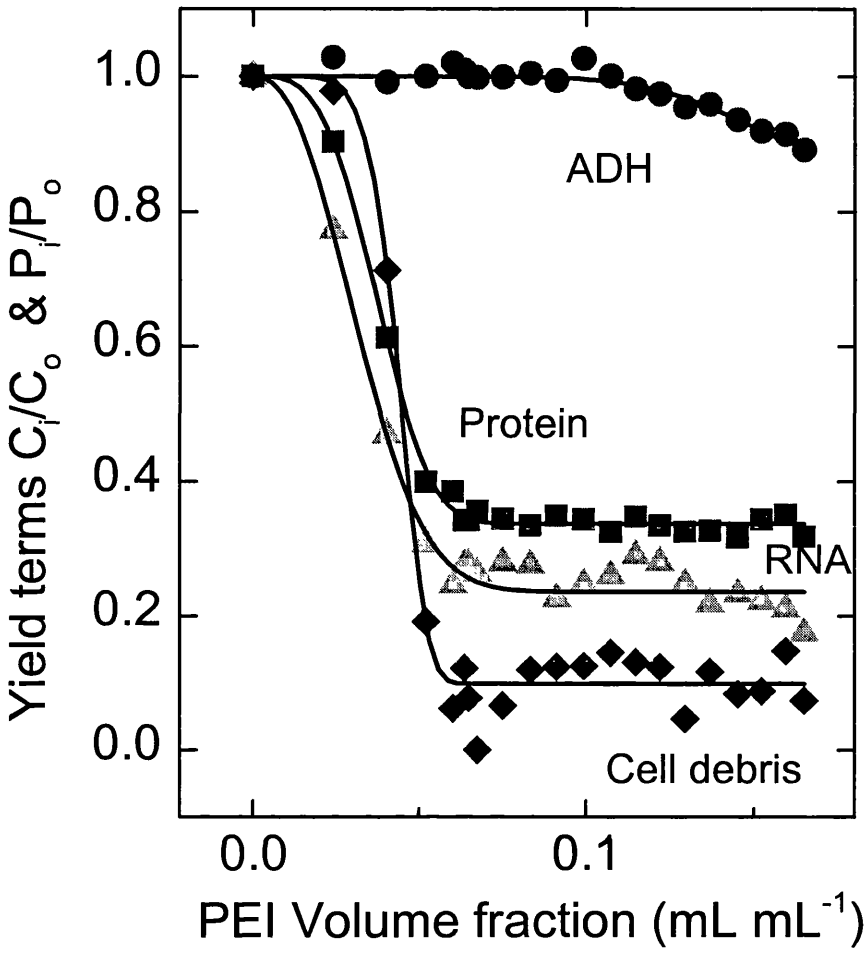


Figure 5.24. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run7. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 6) was flocculated using PEI flocculant (10 mg mL^{-1}, pH 6). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 55 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH	0.12	0.149	8.05	0.945
Protein	0.66	0.043	3.4	0.994
RNA	0.76	0.037	2.32	0.977
Cell Debris	0.9	0.046	6.98	0.985

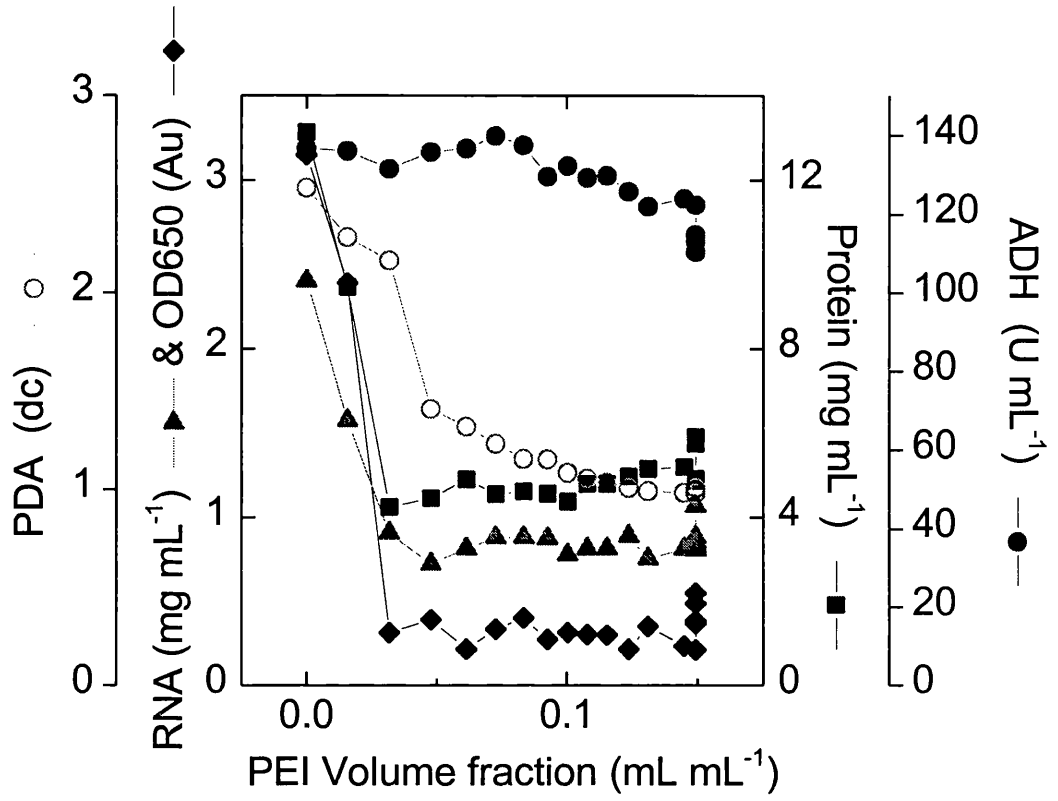


Figure 5.25. *Characterisation of Batch Run8 in terms of at-line total product (ADH) contaminants (RNA, cell debris (OD_{650}) and protein) and PDA dc levels with respect to PEI volume fraction. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (10 mg mL^{-1}, pH 5.5). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 40 min. A pH change (5.5 to 6.6, addition of 10M NaOH) was made during the last two data points. The at-line predicted contaminant data were corrected with the linear correlation in figures 4.9 and 4.10.*

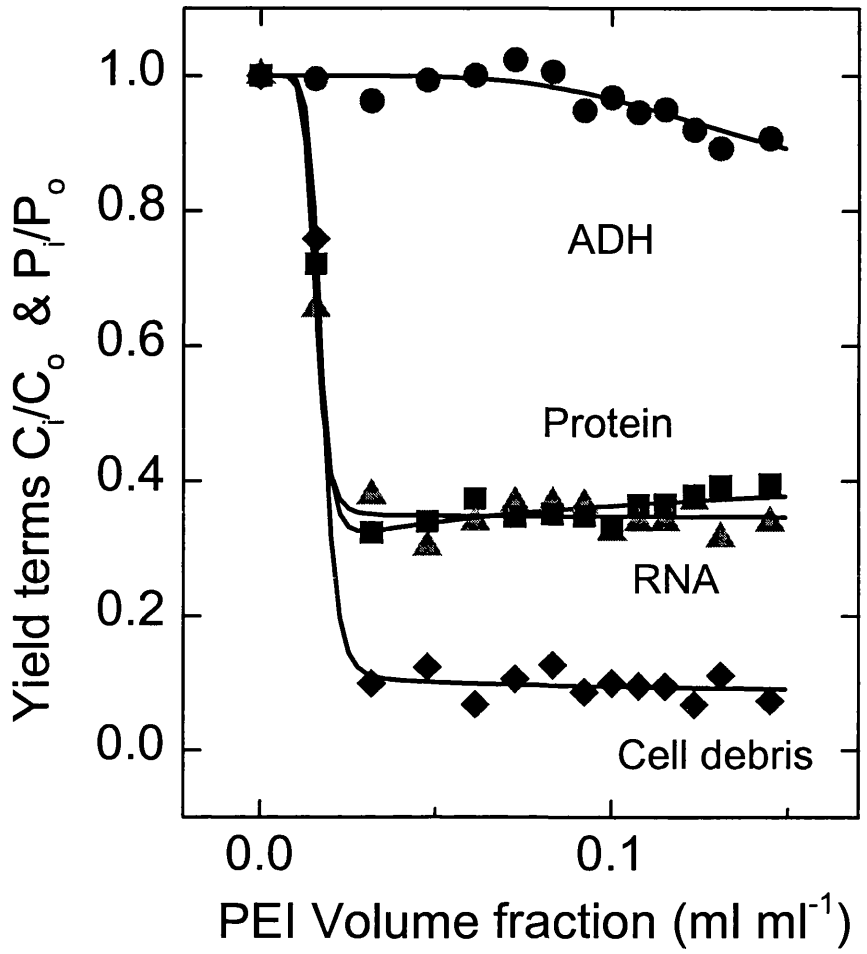


Figure 5.26. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 4-parameter model for Batch Run8. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (10mg mL^{-1}, pH 5.5). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds. The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 40 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	a	b	n	m	R^2
ADH	0.129	0.134	4.2	4.11	0.995
Protein	0.017	0.019	8.79	8.89	0.959
RNA	0.016	0.018	8.61	8.6	0.981
Cell Debris	0.018	0.023	8.64	8.54	0.986

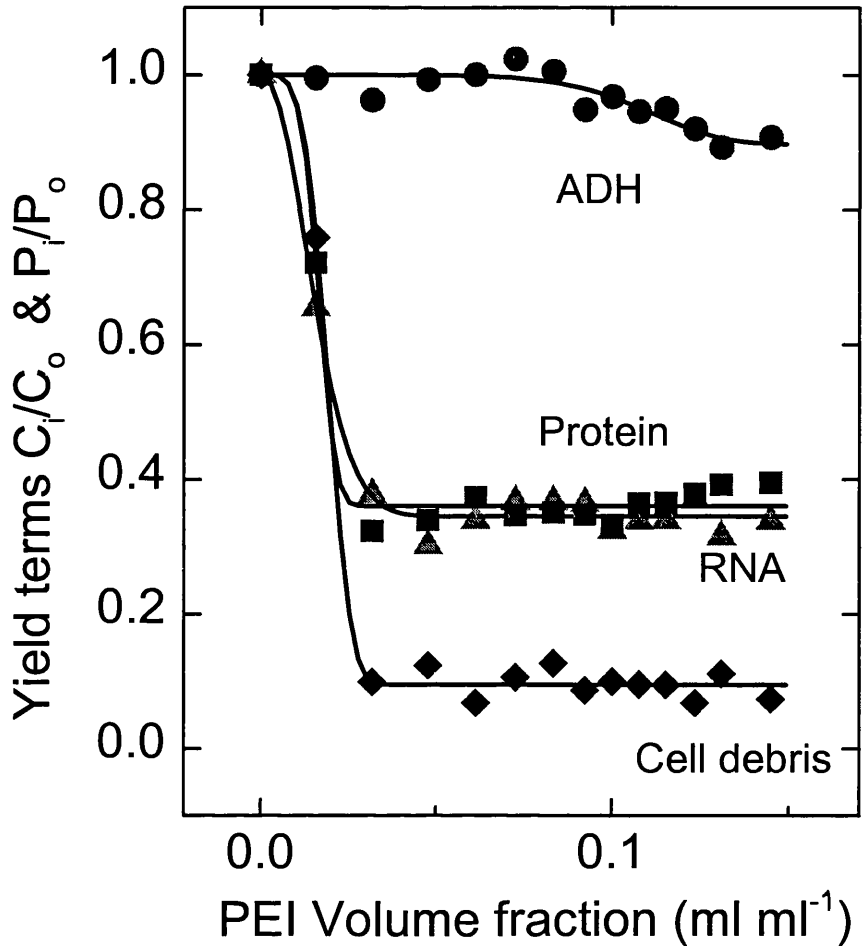


Figure 5.27. *Non-linear least squares (NLLS) fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-parameter model for Batch Run8. Baker's yeast homogenate (700 mL at 125 g (wet weight) L^{-1}, pH 5.5) was flocculated using PEI flocculant (10 mg mL^{-1}, pH 5.5). Both solutions were at an ionic strength of 20mM (KH_2PO_4). Samples (5.6 mL) and at-line measurements were automatically taken every 130-140 seconds The PEI dose flow-rate was controlled to 10-0.01 t (mL min^{-1}), where t is time in minutes. The total run time was approximately 40 min. The NLLS resulted in the below parameter estimations and R^2 term (coefficient of determination).*

	A	B	C	R^2
ADH	0.1	0.115	6.79	0.996
Protein	0.64	0.018	4.95	0.900
RNA	0.66	0.018	2.03	0.985
Cell Debris	0.91	0.021	4.08	0.986

	Batch	1	2	3	4	5	6	7	8
	Homogenate pH	6.5	7	5.5	6.2	6.2	5.5	6	5.5
	PEI pH	6.5	7	5.5	5.9	5	5	6	5.5
	Average pH at the average poin of inflection for all the contaminants	6.5	7.0	5.5	6.2	6.1	5.5	6.0	5.5
	Ionic strength (mM)	50	50	50	50	50	50	20	20
	PEI stock	0.01	0.01	0.01	0.005	0.005	0.005	0.01	0.01
OD650	Start conditions (mg mL^{-1})	3.08	2.42	2.31	3.30	3.51	3.74	3.19	3.19
	A	0.91	0.88	0.85	0.82	0.84	0.87	0.90	0.91
	B	0.062	0.056	0.032	0.108	0.105	0.064	0.046	0.021
	C	12.11	5.30	9.84	12.97	20.05	3.85	6.98	4.08
	Inflection point (mL mL^{-1})	0.062	0.054	0.031	0.107	0.104	0.059	0.045	0.020
RNA	Start conditions (mg mL^{-1})	3.36	2.79	2.21	3.65	3.46	2.98	3.27	3.36
	A	0.82	0.57	0.65	0.70	0.66	0.62	0.76	0.66
	B	0.052	0.048	0.028	0.092	0.091	0.056	0.037	0.018
	C	2.32	4.17	2.44	2.72	3.31	3.25	2.32	2.03
	Inflection point (mL mL^{-1})	0.041	0.045	0.023	0.077	0.082	0.050	0.029	0.013
Protein	Start conditions (mg mL^{-1})	13.85	11.43	12.07	13.21	14.20	11.22	14.77	13.14
	A	0.47	0.50	0.63	0.52	0.58	0.52	0.66	0.64
	B	0.055	0.057	0.030	0.104	0.099	0.062	0.043	0.018
	C	3.24	11.20	6.16	10.05	5.01	4.54	3.40	4.95
	Inflection point (mL mL^{-1})	0.049	0.056	0.029	0.102	0.095	0.058	0.038	0.017
ADH	Start conditions (U mL^{-1})	113.80	111.50	83.43	128.00	124.00	118.00	134.00	136.60
	A	0.19						0.12	0.10
	B	0.228						0.149	0.115
	C	7.10						8.05	6.79
	Inflection point (mL mL^{-1})	0.223						0.147	0.112

R^2 ADH	0.995						0.945	0.996
R^2 Cell Debris	0.983	0.978	0.979	0.994	0.998	0.990	0.985	0.986
R^2 RNA	0.977	0.999	0.975	0.994	0.989	0.993	0.977	0.985
R^2 Protein	0.973	0.996	0.984	0.994	0.977	0.978	0.994	0.900

Table 5.3. *Overview of batch flocculation runs through the 3(2)-parameter model in terms of pH, ionic strength, start contaminant and ADH concentrations. Model parameters A, B and C were estimated through the Levenberg-Marquardt non-linear least squares method (100 iterations). The inflection points was determined through* $B\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$. *The R^2 term is the coefficient of determination. The 3(2)-parameter model is discussed in more detail in section 2.2.*

Batch		1	2	3	4	5	6	7	8
Homogenate pH		6.5	7	5.5	6.2	6.2	5.5	6	5.5
PEI pH		6.5	7	5.5	5.9	5	5	6	5.5
Ionic strength (mM)		50	50	50	50	50	50	20	20
PEI stock (mg mL^{-1})		10	10	10	5	5	5	10	10
Start conditions (Au)		3.08	2.42	2.31	3.30	3.51	3.74	3.19	3.19
OD650	a	0.061	0.053	0.031	0.106	0.111	0.06	0.044	0.0205
	n	14.84	5.43	9.26	11.27	3.16	3.97	8.60	4.00
	c	0.42	4.00	3.48	3.23	6.09	2.31	5.00	5.64
Start conditions (mg mL^{-1})		3.36	2.79	2.21	3.65	3.46	2.98	3.27	3.36
RNA	a	0.049	0.069	0.047	0.097	0.098	0.077	0.037	0.024
	n	2.15	1.80	1.00	2.50	2.85	1.65	1.75	0.827
	c	0.52	8.50	2.92	3.28	4.77	4.11	5.97	8.50
Start conditions (mg mL^{-1})		13.85	11.43	12.07	13.21	14.20	11.22	14.77	13.14
Protein	a	0.082	0.083	0.047	0.127	0.111	0.098	0.047	0.026
	n	1.51	2.45	1.17	3.20	15.33	1.52	1.99	0.98
	c	1.44	11.20	4.96	5.64	3.25	4.61	10.60	11.49
Start conditions (U mL^{-1})		113.80	111.50	83.43	128.00	124.00	118.00	134.00	136.60
ADH	a	0.23						0.24	0.17
	n	2.31						3.82	2.96
	c	1.68						3.36	6.79
R^2 ADH		0.989						0.952	0.995
R^2 Cell Debris		0.979	0.963	0.801	0.971	0.989	0.973	0.970	0.916
R^2 RNA		0.964	0.977	0.961	0.987	0.975	0.928	0.941	0.937
R^2 Protein		0.842	0.981	0.793	0.922	0.956	0.860	0.940	0.346

Table 5.4. *Overview of batch flocculation runs through the 3(1)-parameter model in terms of pH, ionic strength, start contaminant and ADH concentrations. Model parameters a, n and c were estimated through the Levenberg-Marquardt non-linear least squares method (100 iterations). The R^2 term is the coefficient of determination. The 3(1)-parameter model is discussed in more detail in section 2.2.*

Batch		1	2	3	4	5	6	7	8
Homogenate pH		6.5	7	5.5	6.2	6.2	5.5	6	5.5
PEI pH		6.5	7	5.5	5.9	5	5	6	5.5
Ionic strength (mM)		50	50	50	50	50	50	20	20
PEI stock (mg mL^{-1})		10	10	10	5	5	5	10	10
Start conditions (Au)		3.08	2.42	2.31	3.30	3.51	3.74	3.19	3.19
OD650	a	0.061	0.052	0.0296	0.104	0.103	0.058	0.043	0.018
	n	19.71	7.81	16.20	19.03	24.20	5.37	11.26	8.64
	b	0.071	0.070	0.033	0.113	0.111	0.092	0.055	0.023
	m	20.41	8.20	16.19	18.84	24.30	5.86	11.64	8.54
Start conditions (mg mL^{-1})		3.36	2.79	2.21	3.65	3.46	2.98	3.27	3.36
RNA	a	0.047	0.044	0.024	0.088	0.088	0.051	0.029	0.016
	n	2.68	6.27	10.26	3.28	4.02	4.32	4.09	8.61
	b	0.108	0.050	0.026	0.143	0.125	0.066	0.040	0.018
	m	3.18	6.29	10.08	3.79	4.50	4.42	3.93	8.60
Start conditions (mg mL^{-1})		13.85	11.43	12.07	13.21	14.20	11.22	14.77	13.14
Protein	a	0.102	0.055	0.0276	0.099	0.095	0.058	0.0374	0.0165
	n	1.39	8.58	16.79	12.69	6.27	5.97	4.80	8.79
	b	0.224	0.060	0.029	0.105	0.112	0.067	0.048	0.019
	m	2.44	8.65	16.84	12.67	6.53	6.09	4.88	8.89
Start conditions (U mL^{-1})		113.80	111.50	83.43	128.00	124.00	118.00	134.00	134.00
ADH	a	0.219						0.124	0.129
	n	10.76						5.45	4.20
	b	0.224						0.125	0.134
	m	11.04						5.15	4.11

R^2 ADH	0.994						0.955	0.995
R^2 Cell Debris	0.979	0.971	0.977	0.993	0.998	0.984	0.985	0.986
R^2 RNA	0.968	0.989	0.978	0.991	0.982	0.991	0.981	0.981
R^2 Protein	0.839	0.995	0.987	0.994	0.971	0.977	0.994	0.959

Table 5.5. *Overview of batch flocculation runs through the 4-parameter model in terms of pH, ionic strength, start contaminant and ADH concentrations. Model parameters a, b, n and c were estimated through the Levenberg-Marquardt non-linear least squares method (100 iterations). The R^2 term is the coefficient of determination. The 4-parameter model is discussed in more detail in section 2.2.*

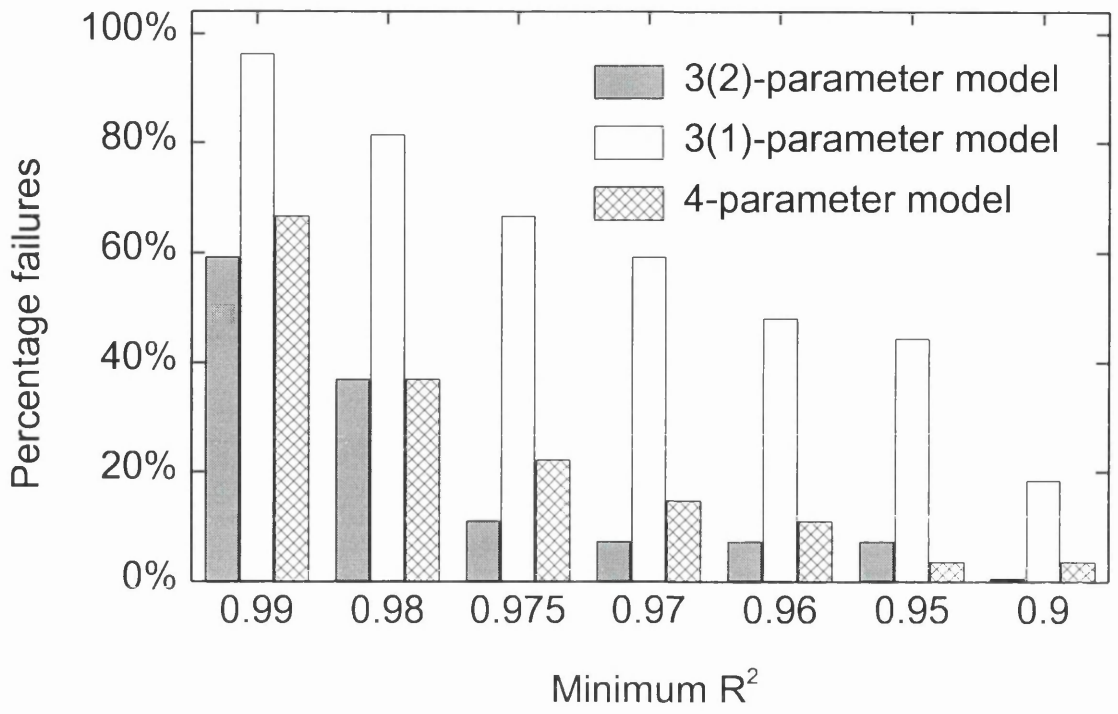


Figure 5.28. *Performance of the three flocculation models. The performance of the models were evaluated by counting the number of failed contaminant and product model fits for a specified minimum coefficient of determination (R^2) as a percentage of the total number of fits performed.*

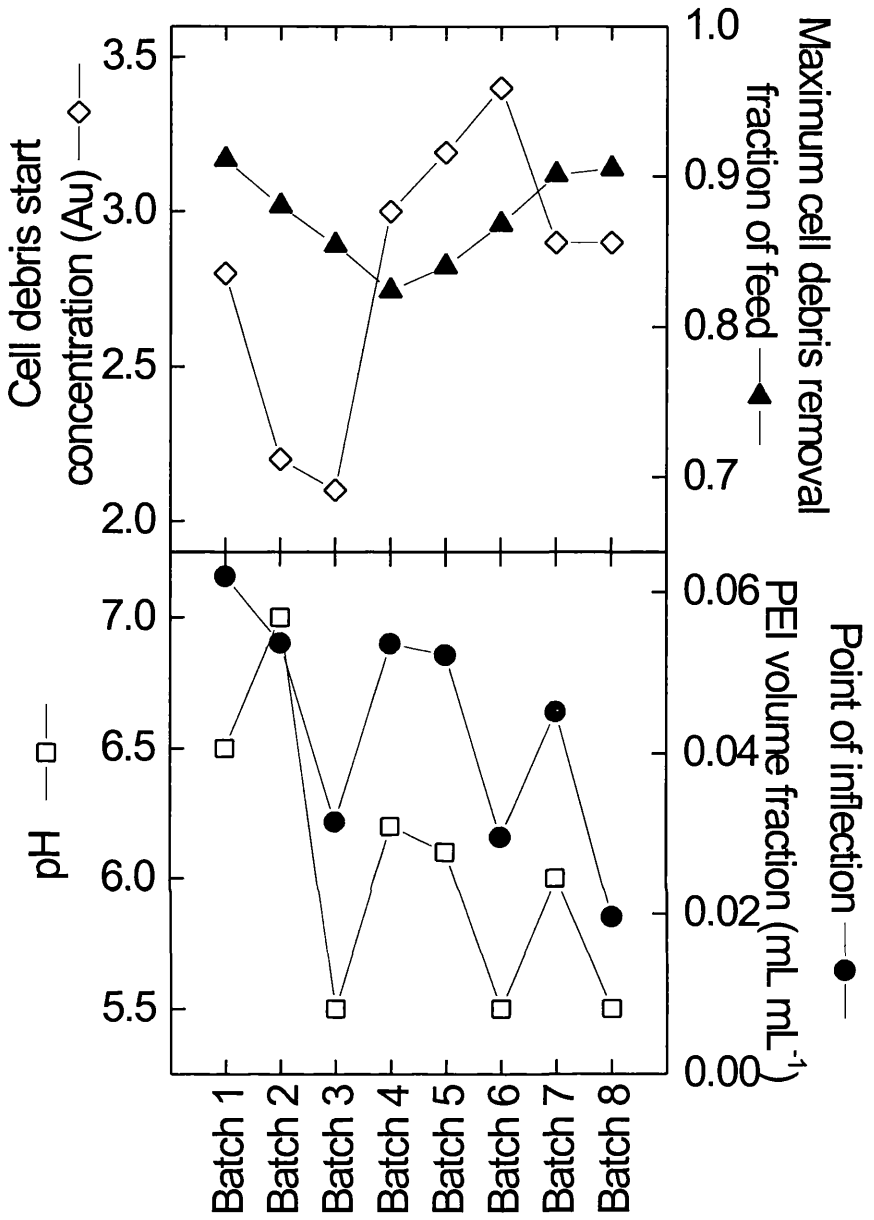


***Figure 5.29**. Effect of pH, feed contaminant concentrations on the batch flocculation cell debris profiles characterised by the 3(2)-parameter model. The inflection points* $(B\left(\frac{C-1}{C}\right)^{\frac{1}{C}})$ *for batch runs 4, 5 and 6 were divided by a factor of two due to the half strength PEI flocculant solution. The maximum removal of cell debris (fraction of feed) was determined through model parameter* A *(see section 2.2). The pH data is the average of the flocculant-homogenate mixture at the average point of inflection (table 5.3, row 3).*

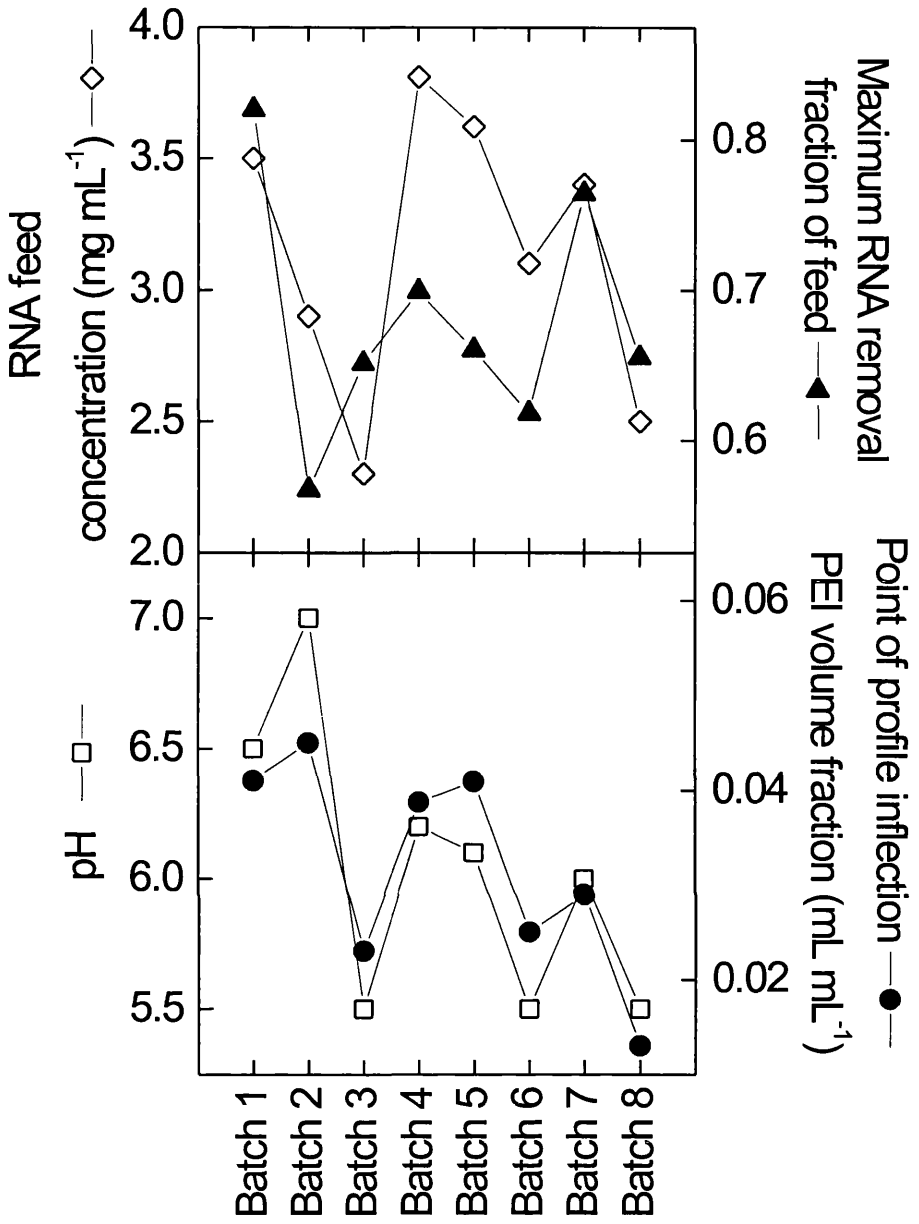


Figure 5.30. *Effect of pH, feed contaminant concentrations on the batch flocculation RNA profiles characterised by the 3(2)-parameter model. The inflection points* ($B\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$) *for batch runs 4, 5 and 6 were divided by a factor of two due to the half strength PEI flocculant solution. The maximum removal of RNA (fraction of feed) was determined through model parameter* A *(see section 2.2). The pH data is the average of the flocculant-homogenate mixture at the average point of inflection (table 5.3, row 3).*

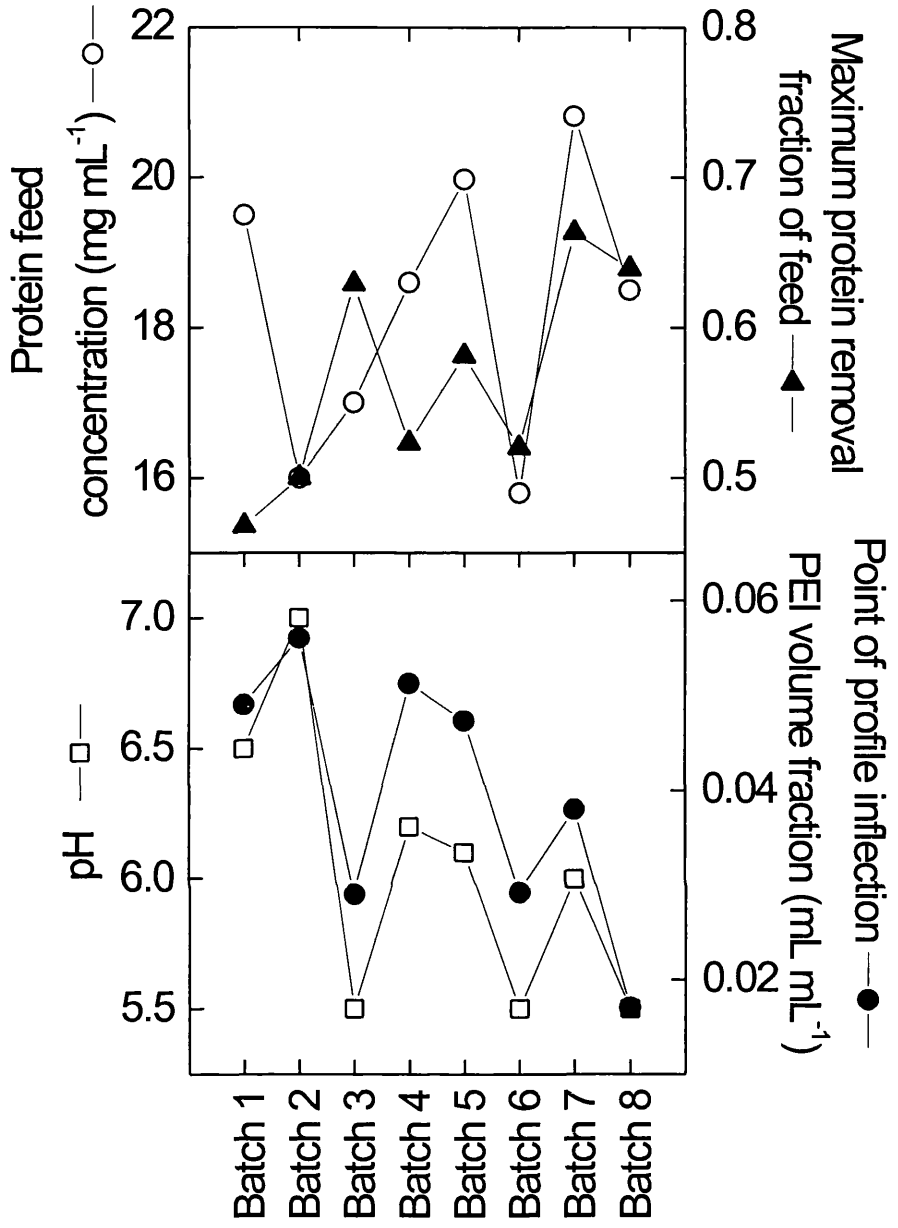


Figure 5.31. *Effect of pH, feed contaminant concentrations on the batch flocculation protein profiles characterised by the 3(2)-parameter model. The inflection points* $\left(B\left(\frac{C-1}{C}\right)^{\frac{1}{C}}\right)$ *for batch runs 4, 5 and 6 were divided by a factor of two due to the half strength PEI flocculant solution. The maximum removal protein (fraction of feed) was determined through model parameter* A *(see section 2.2). The pH data is the average of the flocculant-homogenate mixture at the average point of inflection (table 5.3, row 3).*

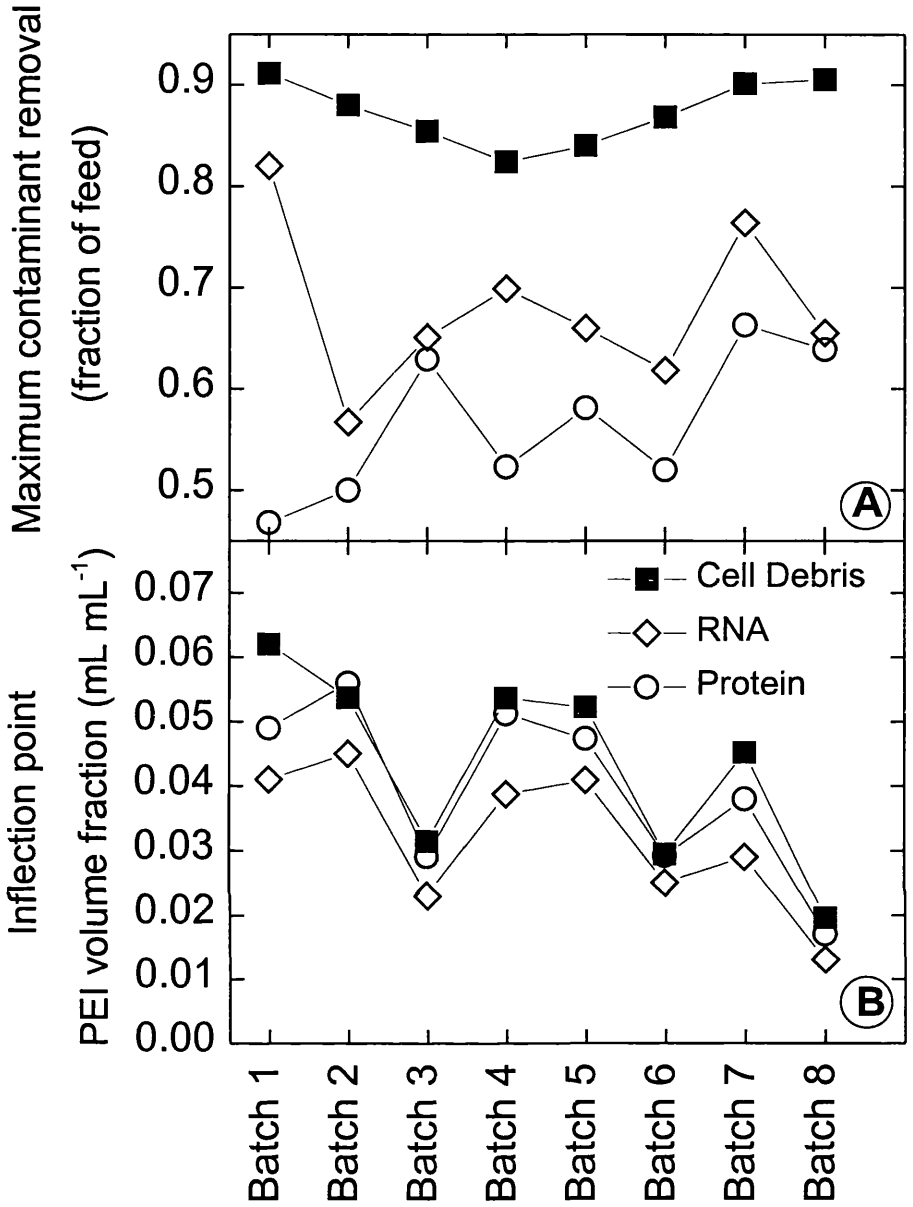


Figure 5.32. *Variation in the maximum fraction removed of RNA, cell debris and protein (plot A) and points of profile inflection (plot B) for eight batch flocculation runs, determined from model parameters* A *and* $B\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$ *(3(2)-parameter model). The inflection points for batch runs 4, 5 and 6 were divided by a factor of two due to the half strength PEI flocculant solution The symbols in Plot A are similar to those in plot B. The pH data is the average of the flocculant-homogenate mixture at the average point of inflection (table 5.3, row 3).*

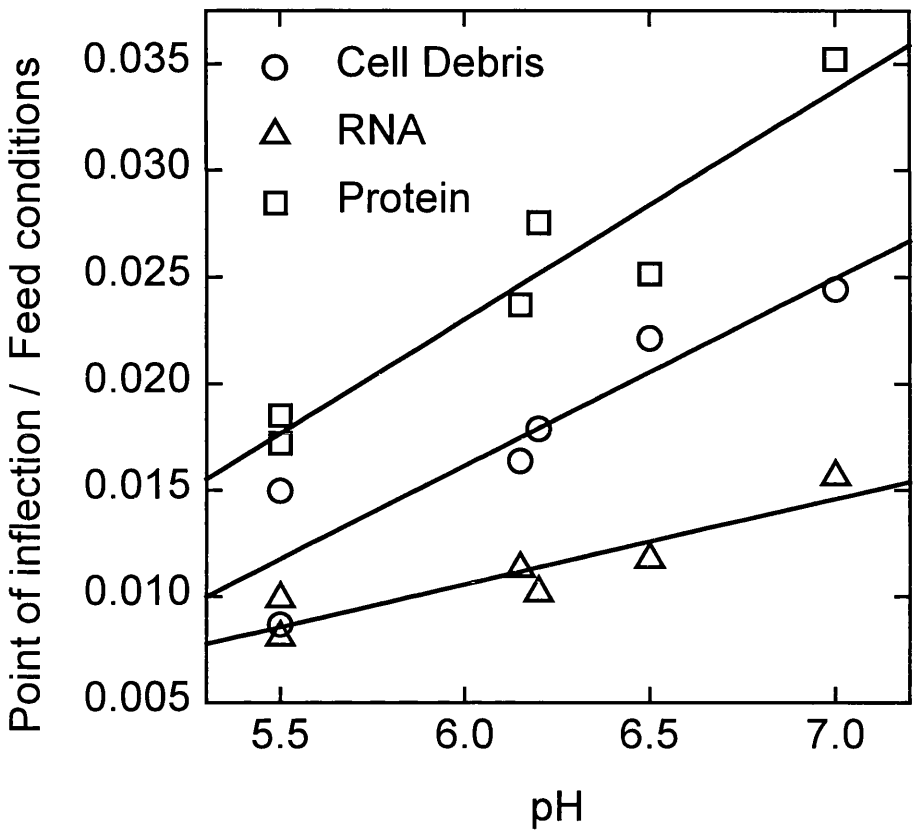


***Figure 5.33**. Normalised point of inflection of the contaminant flocculation profiles in terms of pH for batch runs 1 to 6. The inflection points were determined through* $B\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$*, from the 3(2)-Parameter model and divided by the feed concentrations of the respective contaminants. Batch runs 7 and 8 were not included due to their lower ionic strength levels. A linear fit of the data was performed. The inflection points for batch runs 4, 5 and 6 were divided by a factor of two due to the 2-fold dilution of the PEI stock solution compared to batch runs 1, 2 and 3. The linear fits resulted in coefficients of determination of 0.85, 0.84, 0.91 for the debris, RNA and protein linear correlations. The protein data was multiplied by a factor of 10. The pH data is the average of the flocculant-homogenate mixture at the average point of inflection (table 5.3, row 3).*

Chapter 6

Control and optimisation of the batch flocculation process

6.1 Summary

Real time measurements on both product and key contaminants were applied for optimisation and control of the batch flocculation process. Two control configurations were examined in terms of their ability to control the removal of contaminants and recovery of product during the batch flocculation process, where the flocculant dose was applied as sole manipulative variable. A model based feedback control configuration was implemented to optimise the batch flocculation process, in terms of maximising the removal of contaminants, recovery of product whilst minimising the utilisation of flocculant. This control setup also imposed a constraint on the maximum allowable cell debris concentration in the supernatant stream post flocculation. The adaptive nature of the control configuration was derived from the implementation of a model parameter identification technique. Two such estimation methods were studied and the most appropriate technique for real-time application was investigated. A less sophisticated control configuration was also investigated which utilised the cell debris data as the control variable.

6.2 Introduction

The ability to rapidly monitor the batch flocculation process in terms of both product and key contaminants was demonstrated in Chapter 5. The characterisation of the flocculation behaviour with respect to flocculant (PEI) volume fraction was performed implementing three mathematical descriptions, of which the 3(2)-7Parameter model was concluded to be most appropriate for real time application. This chapter will therefore only apply this model for the study of batch process characterisation, optimisation and control.

As discussed in chapter 1, several factors (input variables) affect the behaviour of the flocculation process (section 1.6.2). Which of these to implement for process control, is dictated by operational and equipment constraints. It is important to distinguish

between input variables which are/should be determined during the design stages and which factors can be implemented for control purposes. Most of the input variables such as batch tank volume, flocculant type and molecular weight, etc, are specified in the design stages of the flocculation process. The three main input variables applicable for process control are the flocculant dose, batch pH and salt concentration. This chapter will consider how the flocculant dose can be applied for process control.

This chapter will first evaluate the performance of two model parameter identification methods, namely the Levenberg-Marquardt non-linear least squares (NLLS) and extended Kalman filter (EKF) for at-line model prediction. The theoretical backgrounds behind these two estimation techniques have been discussed in chapter 2. The model parameter predictions will be implemented to form real-time graphical representations of contaminant and product flocculation profiles, which may be applied to facilitate the operation of the flocculation process. These estimation techniques will be tested on the eight batch runs discussed in chapter 5.

Various control configurations may be applied to control the batch flocculation process, from sophisticated control systems, which apply information on all the contaminants, product and flocculant dose, to relatively simple algorithms which apply the raw data of a single contaminant. Which method to implement is dependent on the complexity of the process and the objective which are set by the operator and/or by the feed requirements of the subsequent unit operation. This chapter will investigate the performance of two such control configurations. An adaptive model based optimisation control setup will be examined implementing the EKF and NLLS model parameter estimation techniques. A simple control configuration using the cell debris yield level as the sole control variable will also be examined. The performance of these two control systems will be tested in terms of eight batch runs illustrating their robustness and efficiency in controlling the batch operations to identify the lowest flocculant concentration to give satisfactory removal of contaminants and recovery of product.

6.3 Real time process characterisation, optimisation and control

This section will briefly discuss the rapid characterisation, optimisation and control configurations adopted in this chapter for the batch flocculation process.

Figure 6.1, illustrates an overview of the rapid process characterisation, optimisation and the two control configurations applied. Real time data was acquired through the rapid monitoring set-up discussed in chapter 3 and the at-line information on both

product and key contaminants was fed to two model parameter identification techniques; the Levenberg-Marquardt (LM) non-linear least squares (NLLS) method and extended Kalman filter (EKF) for model parameter prediction. The identification techniques implemented the 3(2)-Parameter flocculation model, the at-line data of product and contaminants as well as the flocculant volume fraction in order to estimate new model parameters for newly acquired data points. The estimated model parameters for each monitoring step were applied to create flocculation yield curves for both the product and contaminants. Based on this visual aid of the flocculation profiles the operator could verify whether appropriate flocculant conditions have been reached, and if so terminate the flocculant dosing (open loop control). For closed loop control, where the supervisory computer regulates the flocculant dose, an optimisation algorithm was adopted to determine the optimum flocculant volume fraction for the given batch operation which was applied as the ***primary control setpoint***. A ***secondary control objective*** was introduced to guarantee that the cell debris concentration post flocculation was suitable for the subsequent unit operation. The supervisory computer would verify whether the actual flocculant volume satisfied both control objectives and terminated the PEI dosing accordingly. This control configuration will be defined as ***control configuration 1*** for the rest of this chapter. Due to the nature of the batch flocculation process the control of the flocculant dosing was a simple ON/OFF system irrespective of the dosing recipe.

A more simplistic control setup is also illustrated in figure 6.1, where the raw measurements on the cell debris yield level was implemented as the control variable. As soon as the cell debris contaminant level had reached this minimum specification, which was set to a cell debris yield value of 0.2 (80% cell debris removal based on spun feed) the flocculant dosing would be terminated. This control configuration will be defined as **control configuration 2** for the rest of this chapter.

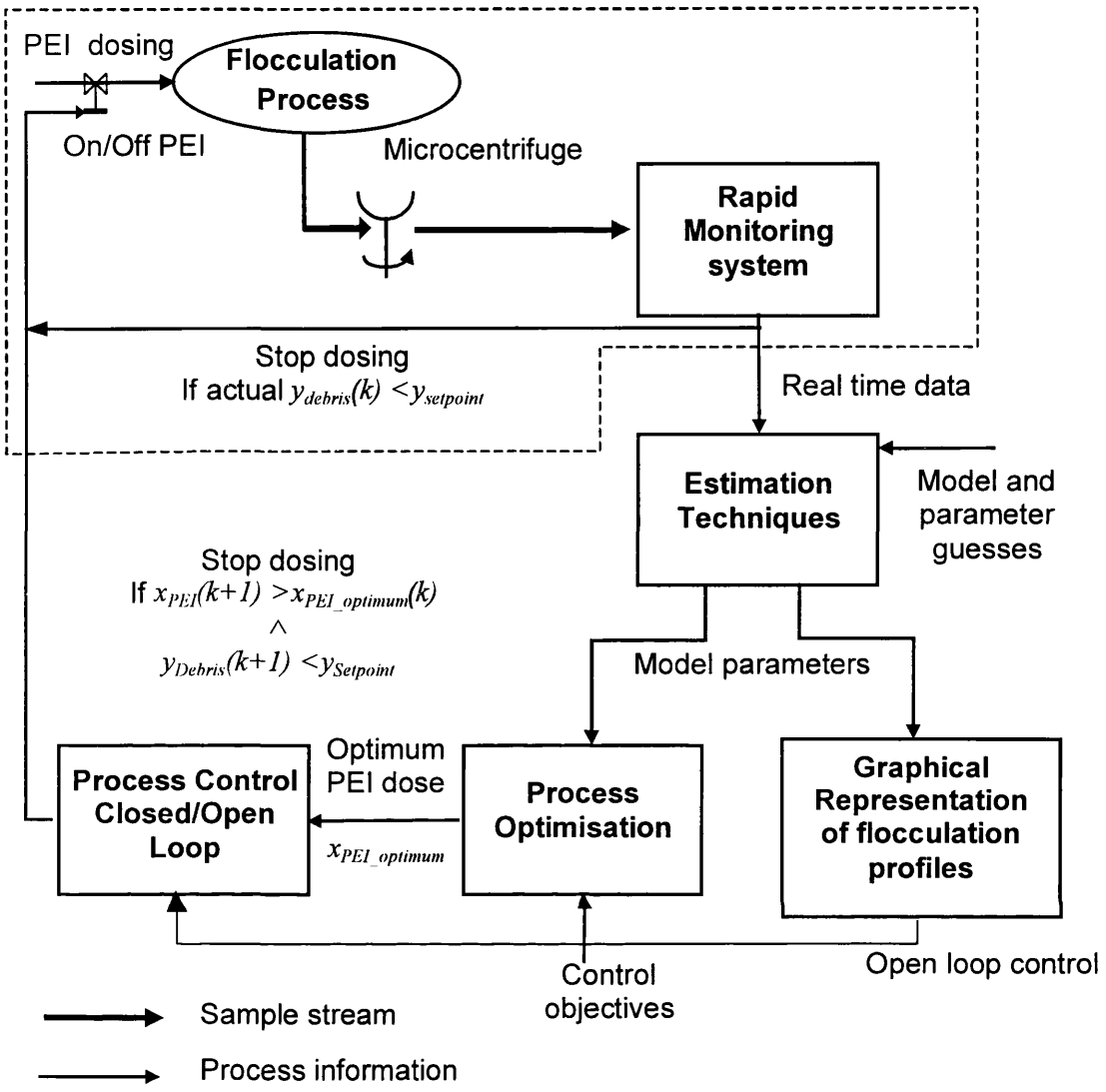


Figure 6.1. *Overview of the batch flocculation control configurations.*
The dashed box illustrates control configuration 2, i.e. using the cell debris yield data as the sole control variable. The flocculant dosing is terminated if the cell debris yield level (y_{Debris}) is smaller than the cell debris yield setpoint ($y_{Setpoint}$).
Control configuration 1 (the remainder of the above diagram) uses the estimated model parameters to optimise the flocculation process in terms of an optimum PEI volume fraction ($x_{PEI\_optimum}$). The flocculant dosing was terminated if the actual PEI volume fraction was greater or equal to the optimum predicted PEI volume fraction as well as insuring that the debris yield level was smaller than the its setpoint. The variable k refers to the measurement step.

A theoretical comparison of the two model parameter identification techniques (EKF and LM) was discussed in chapter 2. **Figure 6.2**, illustrates the flow of information from and to the two estimation methods. Both estimation techniques require a mathematical description of the flocculation behaviour and initial guesses of the model parameters. Given these starting requirements the estimation techniques would repeatedly find better and up to date combinations of the model parameter values. The initial guesses of the model parameters for both the contaminants and product were taken as the average parameter values of all the batch runs, listed in table 5.3. As seen from figure 6.2, the EKF requires additional start information to that of the NLLS technique, such as system and measurement noise terms and initial values for the combined system covariance term. The selection of these values enables the operator to tune the EKF to either put more or less weighting on the at-line data for the estimation of new model parameters and the level of model parameter change between iteration steps. From chapter 2 we have seen that the appropriate tuning of the EKF is vital for good model parameter prediction.

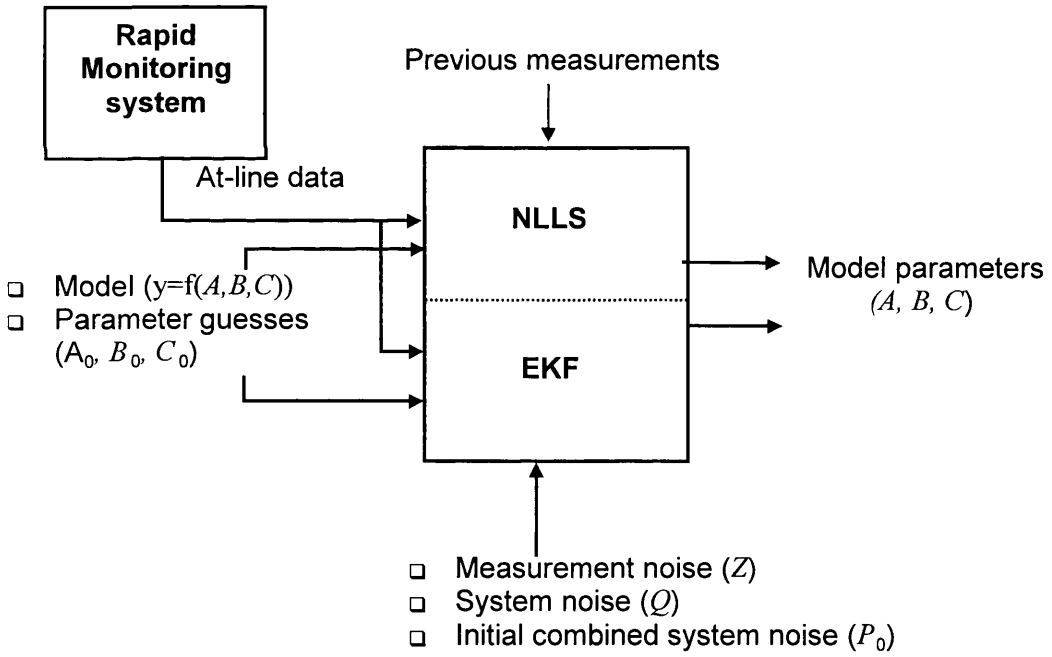


Figure 6.2. *Overview of the extended Kalman filter and Levenberg-Marquardt model parameter identification techniques. The terms NLLS and EKF refer to the Levenberg-Marquardt non-linear least squares and extended Kalman filter estimation methods.*

Table 6.1 lists the various additional constants applied in the EKF during model parameter estimation of all the eight batch runs. The simplest of these additional EKF constants to set was the measurement noise term (Z), which was determined as the square of the standard deviation of the yield measurements discussed in chapter 4. As

the 3(2)-parameter model was adopted the system covariance term (Q) for each contaminant and product was a 3 x 3 matrix, consisting of the variance of each model parameter (no co-variance terms between model parameters were applied). The individual variance terms for the system noise matrix are listed in table 6.1. These were based on the variance of the parameters for all the eight batch and the knowledge that during batch operation there is a gradual increase in the flocculant dose and hence gradual information of the flocculation yield. Due to the latter factor the initial combined system covariance (P) 3 x 3 matrix was set to zero for each term

	Debris	RNA	Protein	ADH
A-Average	0.87	0.68	0.57	0.14
B-Average	0.062	0.053	0.058	
C-Average	9.4	2.8	6.1	
q_{11}	0.002	0.005	0.003	0.0002
q_{22}	0.0002	0.0003	0.0003	
q_{33}	5.5	0.7	4.5	
$\%q_{11}$	5%	10%	10%	
$\%q_{22}$	22%	30%	30%	21%
$\%q_{33}$	25%	30%	35%	
Z	0.0064	0.0064	0.0064	0.0064

Table 6.1. *Extended Kalman filter constants.*

The parameters A, B and C-average refer to the average values of the model parameters over the eight batch runs (refer to table 6.4) implemented as the initial parameter guesses. The measurement noise is termed Z (= σ_m^2, where σ_m is the measurement standard deviation), the system noise term (3x3 matrix) consists of 3x3=9 terms where only three were non-zero, q_{11}, q_{22} and q_{33} which refer to the variance of model parameter A, B and C. The factors $\%q_{11} = \frac{\sqrt{q_{11}}}{A_{Average}} \times 100$, $\%q_{22} = \frac{\sqrt{q_{22}}}{A_{Average}} \times 100$ *and* $\%q_{33} = \frac{\sqrt{q_{33}}}{A_{Average}} \times 100$, *which give an indication of the allowable change between iterations.*

As discussed in chapter 2, if the initial guess for the combined system noise term was high, an initially high Kalman gain (weighting) would be the outcome, putting more weight on the initial measurements. During batch operation the first set of measurements were taken at low flocculant volume fraction levels, resulting in flocculant yield values close to unity (see chapter 5, figure 5.2 as an example). Hence

little information on the flocculation profile gradient, position and maximum removal was available to the estimation technique. By setting the combined system noise term initially to zero, no preconception was made that the EKF could determine all three model parameters based on a few measurements.

Figure 6.3, illustrates the flow of information during the optimisation and control of the batch flocculation process in more detail. The first step in attempting to optimise a process is to clearly define the optimisation objectives, put forward some kind of performance criteria, formulate the problem mathematically and finally attempt to solve it. The objective for the control and optimisation of the batch flocculation process is to **"maximise the recovery of product and removal of contaminants whilst minimising the use of flocculant"**. This control objective was met through the manipulation of solely the flocculant volume fraction, i.e. a one variable problem. To incorporate product, contaminant levels as well as the flocculant volume fraction into a single performance function, a weighting factor for each component was introduced. The choice of the weighting factors for product, contaminants and flocculant dose depend on the cost of the product, purity, and flocculant and the requirements which are imposed by the subsequent downstream unit operation. As illustrated by figure 6.3, up to date estimates of model parameters from the NLLS and EKF techniques were implemented to obtain yield profiles of both product and contaminants. These yield functions were multiplied by operator chosen weighting factors, which set a weight on the individual contaminants and product according to their importance during the recovery/removal of the flocculation process. The performance function was then formulated by subtracting the weighted contaminant and flocculant terms from the weighted product term, and maximised to obtain the optimum flocculant volume fraction. The solution of the maximisation problem was relatively simple due to the one variable nature of the function. A differential method could have been applied to reach a solution, however in this chapter a searching technique was implemented where 500 different PEI volume fraction values were examined, the one which rendered the highest value in terms of the performance function was selected as the optimum flocculant dose. The search range in flocculant volume fraction terms was between 0 and 0.5 mL mL^{-1}, hence a flocculant volume fraction resolution of 0.001 mL mL^{-1}.

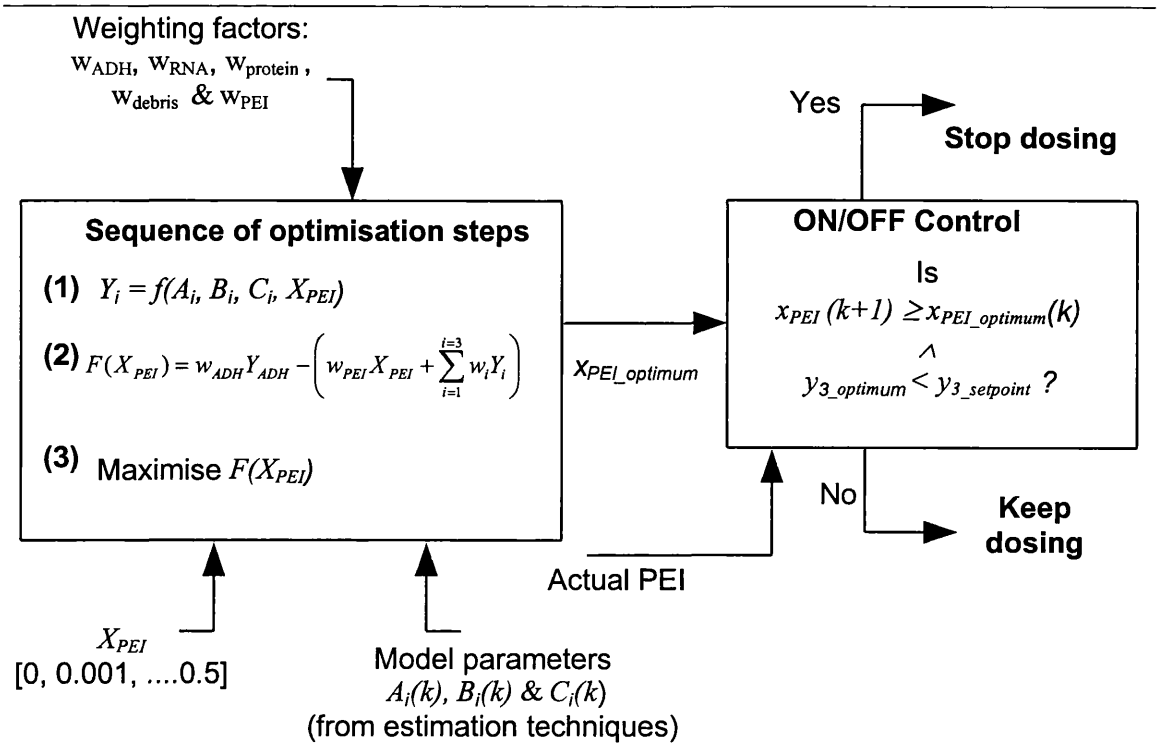


Figure 6.3. *Overview of the process optimisation and control set-up. The optimisation problem is a one variable problem with respect to PEI volume fraction (x_{PEI}). The weight factors for ADH, protein, RNA, cell debris and PEI dose are w_{ADH}, $w_{protein}$, w_{RNA}, w_{debris} and w_{PEI}. The optimum PEI volume fraction ($x_{PEI\_optimum}$) maximises the performance function (F). The Y terms refer to the flocculation yield array (spun flocculated sample as a fraction of spun feed) when applying the array of 500 PEI volume fraction values in the range 0 - 0.5 mL mL^{-1}. A, B & C refer to the model parameters. The subscript i = 1, 2, and 3 refer to protein, RNA, and cell debris. $y_{3\_optimum}$ and $y_{3\_setpoint}$ refer to the cell debris yield , when the optimum and actual PEI volume fraction values are applied. The above optimisation and control sequence was executed for each new measurement step k.*

The optimum flocculant volume fraction prediction was then fed to the control algorithm consisting of a simple ON/OFF switch. The control algorithm compared the latest actual flocculant volume fraction ($x_{PEI}(k+1)$) to the up to date predicted optimum flocculant volume fraction ($x_{PEI\_optimum}(k)$). If the actual flocculant volume fraction was greater than the optimum predicted flocculant volume fraction the dosing is terminated, given that the secondary control objective was satisfied. During the at-line analysis of a sample the dosing of PEI was not terminated. Hence when the information on a particular sample became available the actual PEI volume fraction had increased. Account was made for this measurement delay by comparing the optimum predicted

flocculant dose to the up to date actual PEI volume fraction. It should be noted that the secondary control objective (insuring a minimum debris removal) was based on the predicted debris yield profile.

The secondary control objective was introduced into the control configuration to make it more robust and applicable to a real system. The basic optimisation algorithm discussed above did not take into account the final level of contaminants remaining in solution after the flocculation step. The subsequent downstream operation after the flocculation process is usually a packed bed chromatographic step, which is sensitive to too high cell debris levels due to column fouling. By including a secondary control specification on the acceptable level of contaminant concentration the control configuration can be built more robust for actual application. A minimum cell debris yield level of 0.2 was set as the additional control objective.

6.4 Results

The eight batch runs discussed in chapter 5 will be presented in terms of their at-line model parameter predictions, optimisation and control. The first part of this section will present the results obtained from the at-line parameter estimation and the subsequent predictions of the flocculation profiles for Batch Run1. The Levenberg-Marquardt non-linear least squares (NLLS) and extended Kalman filter (EKF) will be applied for model parameter identification (the abbreviations for these two estimation techniques will be used in the rest of this chapter). The second part of this section will present the predicted flocculation profiles of batch runs 2 to 8, underlining the main points of interest. Optimum flocculant volume fractions were estimated after each measurement step and the profiles determined by both the EKF and NLLS for the eight batch operations will be shown. The results from two control configurations will be presented. The first control setup applies the optimisation data in combination with a minimum specification on cell debris removal as control objectives. The second control configuration applies the raw cell debris yield data as the sole control variable. Finally an overview of the efficiency of the control configurations for all the batch runs will be given.

6.4.1 Rapid model parameter and flocculation profile prediction

Figures 6.4 to **6.6** on page 188-190 illustrate the step by step real time estimation of model parameters for batch Run1 implementing both the NLLS and EKF techniques. The two identification methods implemented the same initial parameter guesses and at-line data of both product and contaminant levels. The additional start information

required by the EKF such as the system and measurement noise terms are listed in table 6.1. Note the initial combined system noise term was set to zero. These values were applied to all of the batch runs.

The trend in model parameter A prediction for Batch Run1 is illustrated in figure 6.4 on page 188. The NLLS estimations were initially more erratic than those from the EKF, and after 6-7 data points the predictions became more stable and levelled off to a constant value. It can also be observed that the parameter predictions for the contaminant curves plateau earlier than those for ADH.

Parameter A gives the operator an indication of the maximum contaminant/product removal level, hence at-line information around this area was required before a good estimation of this parameter could be made. As RNA, cell debris and protein were flocculated and removed earlier in the batch process, more relevant information in terms of contaminant yield values were initially available for the estimation of contaminant model parameters. At low PEI volume fraction levels very little information in terms of ADH removal was available and hence the levelling off of this parameter occurs much later after measurement step 25. The point of parameter A stabilisation was dependent on the quality of information rather than the quantity of data points.

The noisy NLLS estimations of parameter A may be attributed to the lack of sufficient information of the total flocculation profile. For the 3(2)-Parameter model three parameters need to be estimated. During the initial stages of the process run only a few measurements were available for the least squares method to perform model parameter estimation. For the batch operation these points were all in the same area of the flocculation profile and hence contained little information. The erratic nature of the NLLS parameter estimations was therefore due to that several combinations of parameter sets could fit the initial data points satisfactory.

As discussed in chapter 2, the behaviour of the Kalman filter in estimating model parameters is dependent on the correct tuning of Kalman gain and the size of the innovation term (difference between the measurement and the model prediction). The Kalman gain is a function of several factors, such as the system and measurement noise as well as the model derivatives, hence dependent on the position of the measurements on the flocculation profile. The system noise relative to the measurement noise term, sets a ceiling on the allowable change in model parameter prediction and hence prevents erratic estimation behaviour.

From table 6.1 it can be seen that the system noise term for parameter A was relatively small, compared to the two other model parameters (B and C), hence prevented the EKF from introducing any large changes in parameter A between estimation steps. The EKF behaviour will also change with its position on the flocculation profile, i.e. a function of the derivative term ($\delta y/\delta A$). The derivative term $\delta y/\delta A$ at low volume fraction values (measurement steps 0 to 4), will tend to zero which will force the Kalman gain weighting factor to approach zero (see appendix C, equation c.2). If the Kalman gain term approaches zero, one would reject the newly acquired measurement and simply use the nominal value of the model parameter as our estimate, resulting in little change in model estimation.

It can also be argued that the "innovation" term, during the first few measurements of Batch Run1 was very low. As discussed in chapter 2, if the measurement and the model prediction value are similar the innovation term will be small and little adjustment is made to the model parameters.

Model parameter B (3(2)-Parameter model), gives the operator an indication of where the flocculation profile inflection point occurs (see chapter 2). Figure 6.5, illustrates the estimation of model parameter B for both the EKF and NLLS techniques, which were similar in terms of the level of their initial erratic behaviour. This was not the case for the majority of the batch operations (see appendix B) where the model parameter B predictions based on the NLLS exhibited more erratic profiles. It can be argued that due to the higher system noise term associated with parameter B (see table 6.1 on page 171), larger variations were allowed during the Kalman filter estimation steps. The initial variation in the NLLS model parameter estimations were due to the sparse available information.

Although the EKF was slightly faster than the least squares technique in predicting model parameter B, after 6-7 measurement steps an offset between the two estimation methods was observed. It should be noted that the least squares approach applies all of the available measurements for parameter identification and attempts to identify parameter sets which result in an overall best fit. Due to the re-cursive nature of the EKF only the latest measurement is used and the previous data points are represented in the Kalman gain term, hence the Kalman filter is more likely to estimate model parameters which result in good local fits to the observed values. After 6-7 data points any additional measurements gave little information on the behaviour of parameter B (which influences the profile inflection point) as the flocculation profile was in the maximum removal region (figure 5.2). If model parameter A was estimated correctly,

the innovation term would approach zero. As seen before if the innovation term is zero the estimates of model parameters would not change, hence resulting in the offset illustrated in figure 6.5.

The estimation of parameter C, illustrated in **figure 6.6** follows the same patterns as that of parameter B identification, although the NLLS technique is quite unstable for ADH.

6.4.2 Flocculation profile prediction

The at-line graphical representations of Batch Run1 are illustrated in **figures 6.7** and **6.8** for the NLLS and EKF parameter predictions. These curves were created by implementing the estimated model parameters from the EKF and NLLS techniques and simulating the contaminants and product yield profiles. Figures 6.7 and 6.8 show the outcome of such graphical representation using 5, 7 and 10 at-line data sets. The operator can utilise such real time graphical representation to visually determine the performance of the batch flocculation process and act upon this information for open loop control.

During the NLLS graphical prediction (figure 6.7), one can see that initially (plot (a)), the least squares method was poor at predicting the overall flocculation profile, although there was a good fit of the data points. The prediction of the maximum removal level (related to parameter A) was especially poor for the cell debris profile, due to the lack of measurements in this region. After 7 and then 10 data points, plot (b) and plot (c), the predictions improved.

Figure 6.8, illustrates the graphical predictions using the EKF estimated model parameters. Although the predicted profiles describe the flocculation behaviour reasonably well even with initially just a few data points, the actual fitting of the 3(2)-parameter model to the data points was not as good as that of the NLLS method. Figure 6.8, plot(b) and plot(c), clearly illustrate that after 6-7 measurements the innovation term tended to zero, resulting in no variation or improvement to the model parameter estimations.

Figures 6.9 to **6.22** illustrate the graphical representation of batch runs 2 to 8, using both the NLLS and EKF estimations. Similar trends in the NLLS profile predictions to those observed for batch Run1 were seen, such as poor initial profile predictions due to the lack of data point sets through out the profile. However, as soon as there was

sufficient information on the lower part of the flocculation curve (the maximum removal level), the NLLS method became more effective in predicting the total flocculation profile. The EKF filter profile predictions were good with only a few measurement points due to their effective use of the model parameter guesses. Again the actual fitting of the data points was less efficient than that of the NLLS approach. It should be noted that the least squares technique in some cases did not converge (i.e. parameter prediction was zero), resulting in no model parameter prediction and hence no flocculation profile representation (for example during batch Run5 at measurement step 15).

Figures 6.13 to 6.16 illustrate the results from batch operations 4 and 5, which highlight the initial poor flocculation profile predictions estimated by the NLLS technique. As discussed in chapter 5, the stock solution for these runs was diluted 2-fold, pushing the apparent optimum flocculation dose to higher volume fraction levels. Although more data points were available between the start of the operations and the optimum flocculant dose, the NLLS estimation technique predicted correct values for the model parameters only after sufficient information on the total flocculation profile was available. Information on the maximum removal level was especially required before parameter A could be correctly estimated. Flocculation profile prediction for batch runs 7 and 8, illustrate a similar trend, just for a lower resolution of data points. The estimation techniques only had approximately 4-5 points available before correct profile predictions were required. The number of data points prior to the apparent optimum flocculant dose for Batch Run8 was so low that the estimation techniques had great difficulties in identifying the model parameters effectively.

The performance of the estimation techniques were also dependent on the initial model parameter guesses. For Batch Run8 the guesses for parameter B and C were respectively 200% and 130% off their actual values. With such poor initial parameter guesses the estimation techniques, especially the EKF had difficulties to converge to the correct parameter sets.

6.4.3 Batch optimisation and control

This section will present data obtained from the rapid optimisation and subsequent control of the batch processes. Results from the optimisation of the batch runs implementing both the EKF and NLLS estimation techniques will be shown. The performance of the two control configurations, control configuration 1 and 2 discussed in section 6.3 will be analysed. It should be noted that, control configuration 1 aims to

satisfy two control objectives. Its primary objective was to maximise product recovery and contaminant removal for the minimum flocculant dose, whilst the secondary objective imposed a minimum requirement on the removal of cell debris.

Figures 6.23 to **6.30** illustrate the behaviour of the at-line optimisation algorithm in predicting the optimum PEI dose during each batch run. The objective of the optimisation algorithm was to maximise the recovery of product (ADH) and removal of contaminants (RNA, protein and cell debris) for the lowest flocculant volume fraction. The predicted profiles would be implemented to determine optimum flocculant level for a given batch process, as discussed in section 6.3. Weighting factors on the product, contaminants and the flocculation volume fraction were introduced in order to group these factors together in a single performance function. For all the batch optimisation runs the same weighting factors were applied and are listed in the figure legends. The weights were chosen so that, ADH followed by the level of cell debris, RNA, the flocculant dose and then the protein had the greatest importance.

Figure 6.23, plot b illustrates the behaviour of the predicted optimum flocculant dose implementing the EKF and NLLS identification techniques during Batch Run1 for each measurement step. As the predicted optima values rely upon on the correct estimation of the flocculation profiles, the initial fluctuations seen in figure 6.23 were due to the variations in the model parameter estimations. The primary control objective put forward in section 6.3 did not take into account the actual level of contaminant removal. Hence the main driving force in determining the optimum floc dose was the correct estimation of model parameters B and C, as parameter A only informs the operator on the level of maximum contaminant removal and not where this position occurs. The introduction of the secondary control objective (a minimum acceptable cell debris concentration) imposed dependency on parameter A. The offset between the two estimation techniques in determining the optimum flocculant dose was due to the offset which was found in the estimation of model parameter B and C (figure 6.5 and 6.6).

Figure 6.23, plot a, illustrates the behaviour of the actual PEI dose at measurement step *k+1* divided by the predicted optima values at measurement step *k* during Batch Run1. This term gives the operator an indication of when to terminate the PEI dosing, and will be referred to as the *optimisation setpoint or primary setpoint*. If the primary setpoint was greater or equal to unity the dose rate would be terminated, given that the secondary control objective was met. As seen from figure 6.23 if the batch operation was controlled using the EKF the dose rate would have been terminated at a PEI volume fraction of 0.094 mL mL^{-1} (after measurement step 7). For the NLLS based

configuration the controlled PEI volume fraction would have been 0.105 mL mL^{-1} after measurement step 8. At these flocculant values the cell debris removal level was below a yield value of 0.2 (see figure 6.7), hence within the limits of the secondary control objective.

Figure 6.25, illustrates the optimisation results for batch Run3. The variation in the NLLS optimum dose prediction was again related to the fluctuations in the estimation of model parameter B and C, especially for RNA and cell debris (illustrated in appendix B). Due to this initial fluctuation the optimum PEI predictions using the NLLS estimator were at some points relatively low. Figure 6.25, plot b, illustrates that after measurement steps 6 ($x_{PEI}(k+1) = 0.045$ *mL* mL^{-1}) the optimum predicted PEI dose ($x_{PEI\_optimum}(k) \sim 0.044$ *mL* mL^{-1}) fell below the actual flocculant level, which resulted in the possible termination of the flocculant dosing. The secondary objective was verified and it can be seen that at the optimum predicted PEI volume fraction the predicted cell debris yield level was below the 20% setpoint (figure 6.11) and PEI dosing would be terminated at a PEI volume fraction of 0.053 mL mL^{-1} (measurement step 7). If the EKF optimum value was adopted as the control setpoint a flocculant volume fraction of 0.061 mL mL^{-1} would have been chosen.

Figures 6.26 and **6.27** illustrate the optimisation profiles for batch runs 4 and 5. Plot a for these two figures show that the optimisation setpoint predicted by the NLLS approach was first met after measurement step 5-6 for Run4 and after step 7 for batch Run5. However for both of these operations the secondary control objective was not satisfied and the dosing rate was continued (see figures 6.13 plot (a) and 6.15 plot (a)). After measurement step 13 for batch Run4 and step 14 for batch Run5, both control objectives were met and the PEI dosing could be terminated. The optimisation and subsequent control profiles for the EKF based approach were smoother, and the secondary control objective was always met if the optimisation objective was triggered. This may be due to the EKF's less erratic manner in estimating model parameters.

Figure 6.30, illustrates the optimisation profiles for batch Run8. Due to the fast dosing rate, low ionic strength and pH conditions the resolution of data points are low compared to the previous batch operations. The low number of data points prior to the apparent optimum flocculant dose and the poor initial parameter guesses made model parameter and hence flocculation profile predictions very difficult. The optimisation and control of this process was therefore relatively poor and as discussed below an overdose was seen for both estimation techniques.

6.4.4 Performance of the control configurations

Figure 6.31, illustrates the performance of the control configurations for the 8 batch operations in terms of their ability to target the optimum flocculant setpoint. The overall optimum PEI volume fraction setpoint was defined as the flocculant dose which maximised the recovery of product, removal of contaminants for the lowest flocculant dose as well as insuring that at least 80% of the cell debris has been removed. The optimum setpoints (—) were determined by implementing the optimisation algorithm using all of the available data points and the NLLS estimation method. As this identification technique takes into account all of the data and attempts to reach the best overall fit it was chosen as the overall target.

Control configuration 1 adopting the NLLS estimator (□) tended to regulate the PEI dose well, although a constant overdose was seen for the majority of the batch runs. This overdose was due to that the NLLS technique was slightly slower at reaching correct estimates of the flocculation profiles.

Control configuration 1 implementing the EKF (○) estimator performed similar to that of the NLLS based approach, however the controlled PEI dose was more scattered around the optimum setpoint.

The dotted horizontal line on figure 6.31, refers to the average optimum PEI value applying all eight batch runs. If such a flocculant dose was implemented a PEI overdose would be the outcome for batch runs 2, 3, 7 and 8 and for batch runs 4 and 5 too little flocculant would have been applied resulting in inadequate contaminant removal (see figure 6.33).

The *control configuration 2* (◊) (simple cell debris based control configuration) yielded in general lower flocculant dose control values to that of *control configuration 1*, especially during batch Run7. This was seen as the cell debris data tended to reach below 0.2 yield values prior to the optimum flocculant dose, which attempted to maximise the removal of all the contaminants. For batch Run8 it can be seen that *control configuration 2* yielded a better controlled flocculant dose. This can be attributed to the difficulty in parameter prediction for this run and hence the poor performance of the model based optimisation control algorithm.

Figure 6.32, illustrates the performance of the control configurations in terms of their ability to target the optimum flocculant setpoint quantitatively. Plot (a) shows the

percentage difference between the setpoints and controlled PEI values for four different control scenarios. Plot (b), illustrates the same difference however in terms of PEI volume fraction levels. The four different scenarios refer to the type of control configuration applied. From figure 6.32 it can be seen that if an average PEI dose was used as the flocculant control value, most of the batch runs would be operated unsatisfactory. For batch runs 4 and 5 such control resulted in PEI under dosing of 20% - 40% and hence a very poor removal of contaminants was seen (see figure 6.33). The average flocculant dose offset from the optimum setpoint was approximately 29% or a 0.028 mL mL^{-1} in terms of PEI volume fraction. Batch Run8 was not included in this average value. *Control configuration 1* implementing the EKF and NLLS performed similarly, with an average PEI dose offset of only 8% (not including Batch Run8) from the optimum setpoint and in flocculant volume fractions terms this was 0.008 - 0.009 mL mL^{-1}. Due to the relatively infrequent sampling during the batch operations, the PEI volume fraction step change between measurements was approximately 0.01 mL mL^{-1}. Comparing this value to the PEI offset for *control configuration 1*, based on the EKF and NLLS it may be concluded that good control was achieved. The PEI volume fraction terms controlled by *control configuration 2*, yielded in general an under dose in flocculant of approximately 16% or 0.017 mL mL^{-1} in terms of flocculant volume fraction. Compared to the optimum flocculant setpoint the performance of this control configuration was not as effective as *control configuration 1*. However compared with the use of an average flocculant dose the *control configuration 2* performed better.

During the control of Batch Run8 we see that *control configuration 1* implementing both the NLLS and EKF over estimated the flocculant dose by 120% and 60%. As discussed earlier the operational conditions applied during batch Run8, made model parameter estimation and hence optimisation and control difficult. The performance of a control configuration is very much influenced by the conditions under which one chooses to operate a given process. When a constant average PEI dose was applied as the controlled flocculant level for Batch Run8 the outcome was even worse than that of *control configuration 1*. *Control configuration 2* was the most effective control system for this batch operation as it did not rely on model parameter predictions and utilised the raw cell debris data directly.

Figure 6.33 illustrates the performance of the control configurations in terms of their ability to remove contaminants. When *control configuration 1* was applied implementing either the NLLS or EKF estimation techniques close to optimum contaminant removal was achieved for all the batch runs. Flocculant control based on

control configuration 2 saw close to optimum protein removal levels (apart from batch run 1 and 5). However in terms of RNA and cell debris yield values this simple control configuration performed sub-optimally for many of the batch runs. Although the specification on the allowable cell debris yield was set at 0.2 (80% removal based on spun feed), it can be seen that for batch runs 2, 6 and 7 further debris removal was possible. If an average PEI volume fraction was applied the performance in terms of contaminant removal was relatively good, apart from batch runs 4 and 5 where approximately 60% of the cell debris was still in solution.

6.5 Discussion

The behaviour of model parameter identification using the EKF or NLLS techniques was a function of several interrelated factors. For the NLLS method the number of measurements and their position on the flocculation curve was important for good at-line data fitting. During EKF model parameter predictions the tuning of the Kalman gain and the value of the innovation term dictated the behaviour of the Kalman filter estimations. The Kalman gain was in turn a function of multi-component interactions, including the relative magnitudes of the system and measurement noise levels, the initial guess of the combined system noise term and the position of the measurements on the flocculation profile (derivative terms).

We have seen that the prediction of model parameter A, which informs the operator on the optimum level of contaminant removal was directly dependent on the number of measurements available in this area and the accuracy of the initial parameter guesses. A similar trend was found for the prediction of model parameter B and C. During batch operation the first set of measurements were taken at low flocculant volume fraction levels, resulting in flocculation yield values close to unity. Hence little information on the flocculation profile gradient, position and maximum removal was available to the estimation techniques. The operating mechanism of a batch flocculation process therefore affected the estimation technique's ability to rapidly predict the model parameters.

The at-line estimation of model parameters was implemented to create a graphical representation of the overall flocculation profile for each contaminant and product. The behaviour of the flocculation profile predictions was directly dependent on the correct estimation of model parameters. Hence variations in parameter estimations yielded fluctuations in profile predictions. This was seen especially for the NLLS identification technique, where initially very poor predictions were seen due to poor estimates of model parameter A. The EKF based estimations were less erratic resulting in relatively

good predictions of the total flocculation profiles. However as more data points became available, the NLLS technique started to predict more accurate model parameter estimates, resulting in better profile predictions. The flocculation profile predictions may be applied as a graphical tool to visually inform the operator on the outcome of the batch flocculation process in real time. This information can be used to drive the process to critical operating areas for better process understanding or applied for open loop control whereby the operator regulates the process conditions in order to achieve certain process objectives, such as good contaminant removal.

One of the main advantages of implementing the Kalman filter is that the at-line computation load is low, compared to that required by a NLLS technique. If the EKF is well tuned the performance of such an estimator is ideal, as it accounts for both the in-accuracy of the model and measurement noise. The ability to tune the EKF for a particular system enables the operator more flexibility. For example the tuning of the EKF could be geared so that parameter changes are small between estimation steps. However for poorly defined systems, the advantages of the Kalman filter are less prominent. One of the main problems with applying such an estimation technique is that there are no straightforward guidelines for optimal tuning of the Kalman filter. Filter tuning is specific to individual processes and usually requires preliminary computer simulations. The tuning of the EKF is a function of many factors such as, the measurement noise, system noise, complexity of the model, the position of the measurements on the model profile, the level of expected change from batch to batch and within batch operations and the frequency of data points.

The main benefits of the NLLS are that it is a simple parameter identification technique to apply and it is included in most analysis software packages. Several authors (Ramamurthi *et al.* 1993, Holwill *et al.* 1997) have argued that one main disadvantage in implementing such a technique is the high computational load. For a process with relatively slow monitoring systems a few data points are acquired throughout a run and hence the computational load is not seen to be a constraint.

Closed loop control of the batch operations has the benefit of being fully automated and therefore not reliant on operator based decisions, which are susceptible to variation. Supervisory control will also enable the introduction of optimisation algorithms, in order to reproducibly regulate the batch process to ideal operating conditions. Two closed loop control configurations were examined and their ability to regulate the flocculant dosing during batch flocculation was compared in terms of contaminant removal and product recovery.

The two control configurations differed in their complexity and control objectives. The first control setup (*control configuration 1*) attempted to optimise the flocculant dose (*primary control objective*) by finding the lowest PEI volume fraction which would maximise product recovery and contaminant removal, given that at least 80% of the cell debris was removed (*secondary control objective*). The secondary control objective on the cell debris removal was introduced to guarantee that the product stream was suitable for loading on to for example a packed bed chromatography column. The primary control objective was determined through the maximisation of a performance function, which combined the product and contaminant flocculation profiles as well as the flocculant dose through the use of weighting factors. As the optimisation problem was a function of the flocculation profiles, the behaviour of the optimisation and subsequent control configurations were directly dependent on the trend in model parameter prediction implementing either the EKF or the NLLS estimation techniques. Due to the initially erratic nature of the NLLS estimation approach, the optimisation profile was slightly noisy especially for a low number of data points. We have seen that due to this initial fluctuations in the prediction of optimum flocculant levels, that in some cases the primary control objective would be met prematurely and only with the help of the second control objective would the batch operation be operated satisfactory.

The operator chosen weighting factors adopted during the optimisation algorithm, allowed the prioritisation of the removal of specific contaminants. The weighting factors in this chapter were chosen so that the optimisation algorithm put most weight on ADH, cell debris, RNA, flocculant dose and then protein. One of the benefits of the optimisation setup is its flexibility. If for example the subsequent downstream operation was a filtration step, the weighting factors could be chosen to suit this unit operation's ideal feed conditions. Additional manipulative variables could also be introduced into the optimisation algorithm allowing for the optimisation of the flocculation process in terms of the flocculant dose, pH and ionic strength. Real time control of these manipulative variables, may be a too difficult task during the batch process but for batch to batch operations this may be feasible. Although the optimisation based control configuration (*control configuration 1)*, may initially result in fluctuations due to variations in model parameter estimates, its ability to reproducibly optimise the batch flocculation process according to the requirements of the subsequent unit operation is a strong factor which promotes its use. Due to the good recovery of the product ADH during the batch flocculation the full benefits of this more complex control setup were not further highlighted. Which of the estimation techniques to apply for this control

configuration is a function of several factors as discussed above. For well known systems with small fluctuations between batch operations an EKF based system should be adopted. For systems which are less known and where within a batch, disturbances are minimal the NLLS will always eventually guarantee good predictions as more data points become available. However for a batch system where over dosing is crucial to avoid, the EKF's ability to rapidly predict the main trends in the flocculation profile with a few data points would be of great advantage.

The control configuration using the raw cell debris data as the sole control variable (*control configuration 2*), performed relatively well in terms of contaminant removal and did meet its 80% cell debris control objective for most of the batch operations. The benefits of such a control system is its ease of use and implementation. However with noisy measurements such a control system would be susceptible to instability. Although this control approach did not optimise a given batch operation in terms of the removal of all the contaminants this criteria need not be crucial for a unit operation so early in the recovery sequence. By comparing the above control configurations to the application of an average flocculant dose, it can be concluded that both control systems were superior to using a predetermined fixed value in terms of flocculant control for contaminant removal and product recovery given batch to batch disturbances. If the resolution of data points during batch operations become higher one should expect the performance of the control configurations to improve, especially for the cases which yielded a slight overdose in the controlled PEI volume fraction. The batch Run8 example illustrated that the manner of process operation can have a great influence on the performance of a control system, hence during the design of an unit operation this should be taken into account.

6.6 Conclusion

The at-line data for eight batch runs was used for real-time flocculation process characterisation, optimisation and control. Predicted model parameters were used to give the operator a graphical representation of the flocculation behaviour, which could be implemented as a visual aid for open loop control. Two model parameter estimation techniques were examined and it was concluded that the optimum method to apply was a function of several factors, such as the knowledge of the process, level of measurement noise, frequency of data points and level of batch to batch and within batch fluctuations.

Two closed loop control configurations were examined, which utilised different control objectives and level of complexity. A control configuration attempting to optimise the

removal of contaminants and recovery of product for the minimum flocculant dose, given that at least 80% of the cell debris in the spun feed was removed was examined. This approach utilised weighting factors to prioritise the importance of the individual contaminants, product and concentration of flocculant. These weighting factors could be chosen so that the control configuration attempted to control the batch flocculation in order to reach conditions ideal for the subsequent unit operation. The optimisation algorithm required a model of the flocculation process and hence both the EKF and NLLS methods were looked into for model parameter predictions for each newly acquired measurements. Good control to optimum operating conditions was achieved implementing both estimation techniques with on average 92% accuracy. The NLLS based approach tended to control the PEI volume fraction to a slight overdose, due to its slower ability to predict the correct model parameters than that of the EKF. A control configuration applying raw cell debris yield data as the sole control variable was also examined. This approach would terminate the flocculant dosing when the cell debris yield level was below a setpoint of 0.2, i.e. at least 80% removal of cell debris based on spun feed. The performance of this control setup was relatively good although it did not optimise the removal of all the contaminants. However for unit operations where tight control is not required and measurement noise is low such a control system would be adequate.

At-line time monitoring of both product and key contaminants was achieved in chapter 5. This chapter has shown how such real time information may be applied to improve the operation of the batch flocculation process either through open or closed loop control. Although the later control system required the use of relatively complex control algorithms, it would enable process automation and the possibility of reproducibly optimising the batch operation.

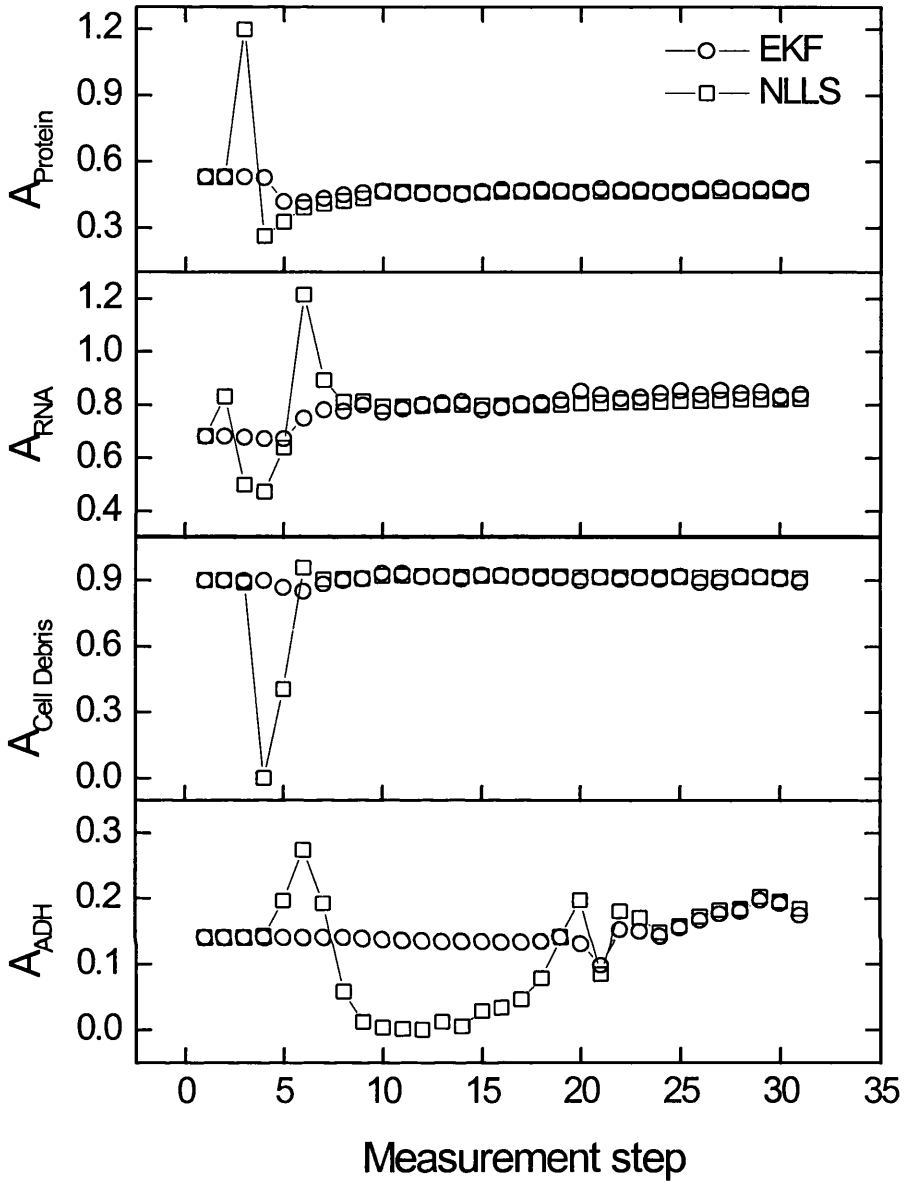


Figure 6.4. *Estimation of model parameter A for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Lavenberg-Marquant non-linear least squares techniques for Batch Run 1. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 (p. 171).*

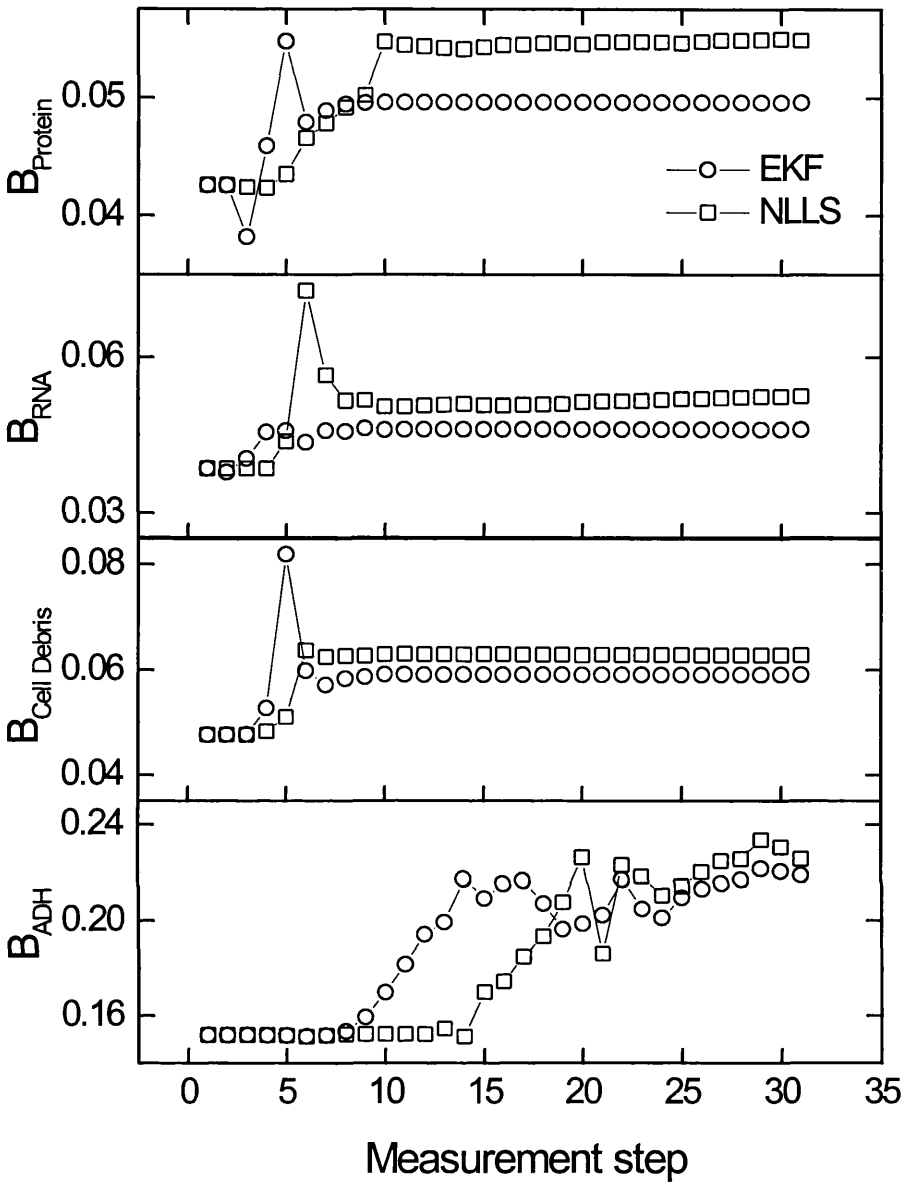


Figure 6.5. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Lavenberg-Marquant non-linear least squares techniques for Batch Run 1. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1*

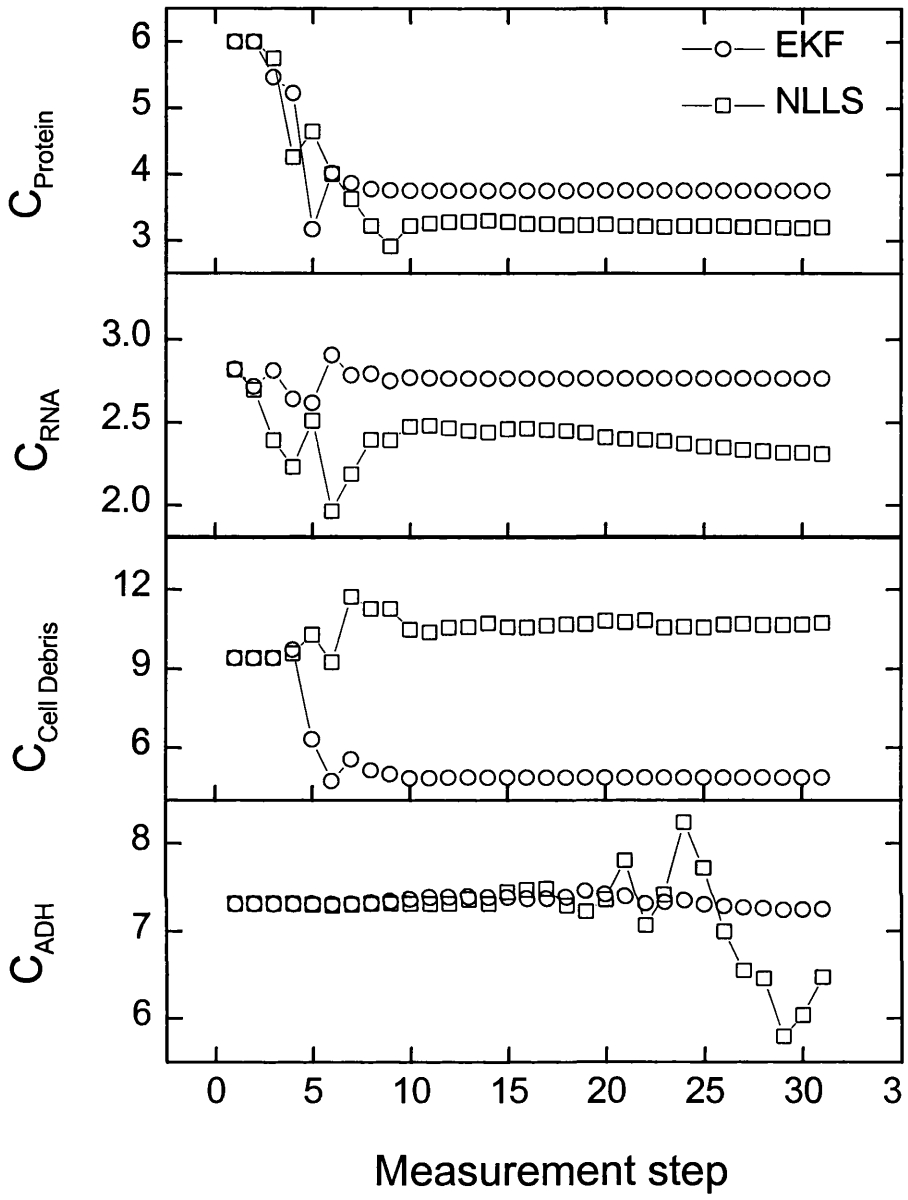


Figure 6.6. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Lavenberg-Marquant non-linear least squares techniques for Batch Run 1. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1*

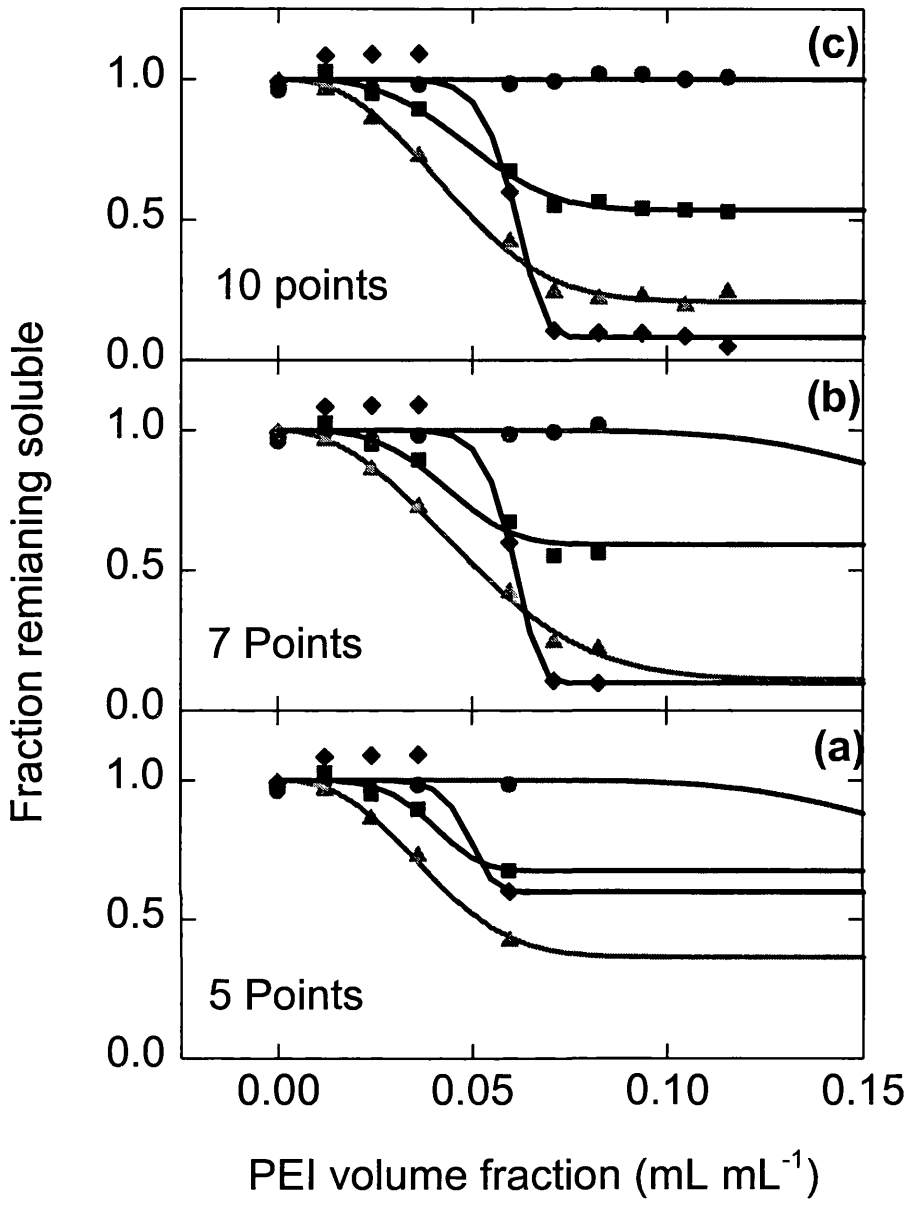


Figure 6.7. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run1. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

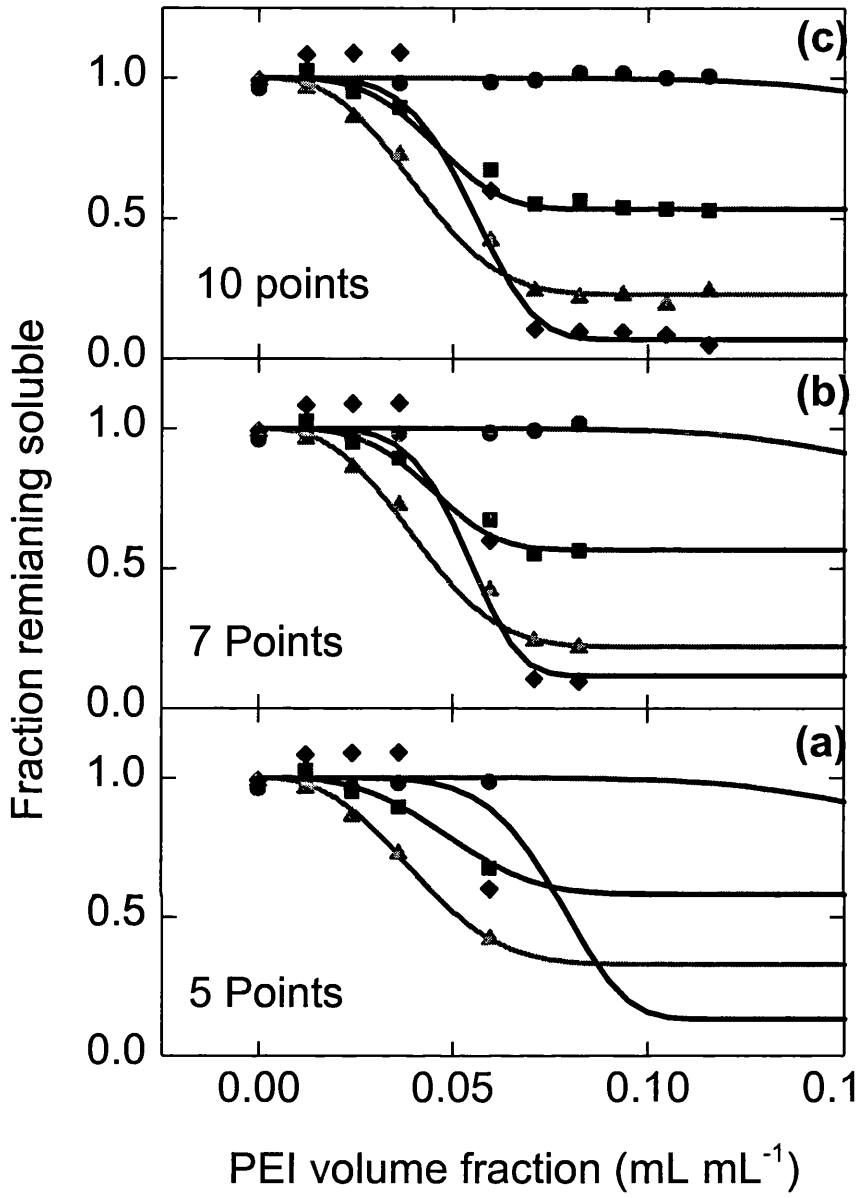


Figure 6.8. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run1. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

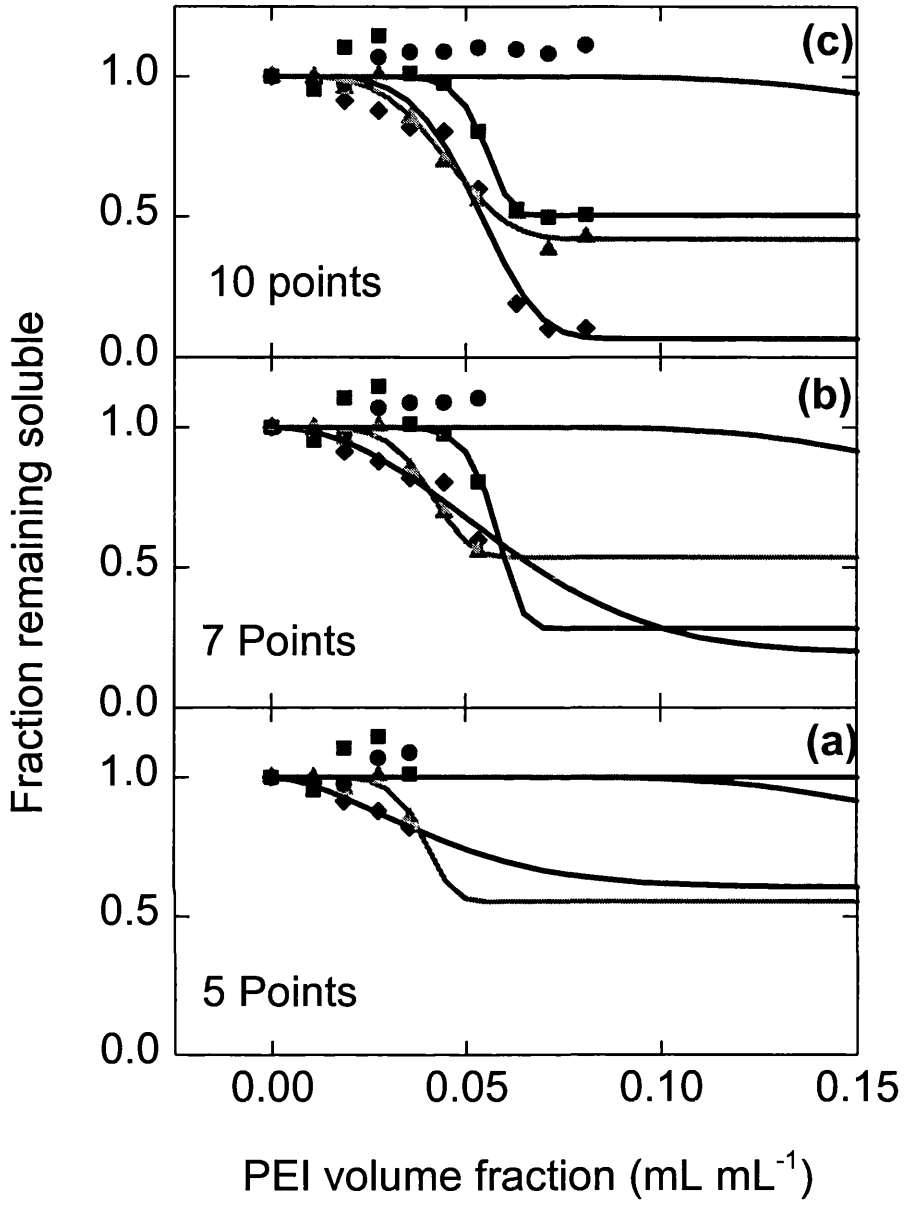


Figure 6.9. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run2. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

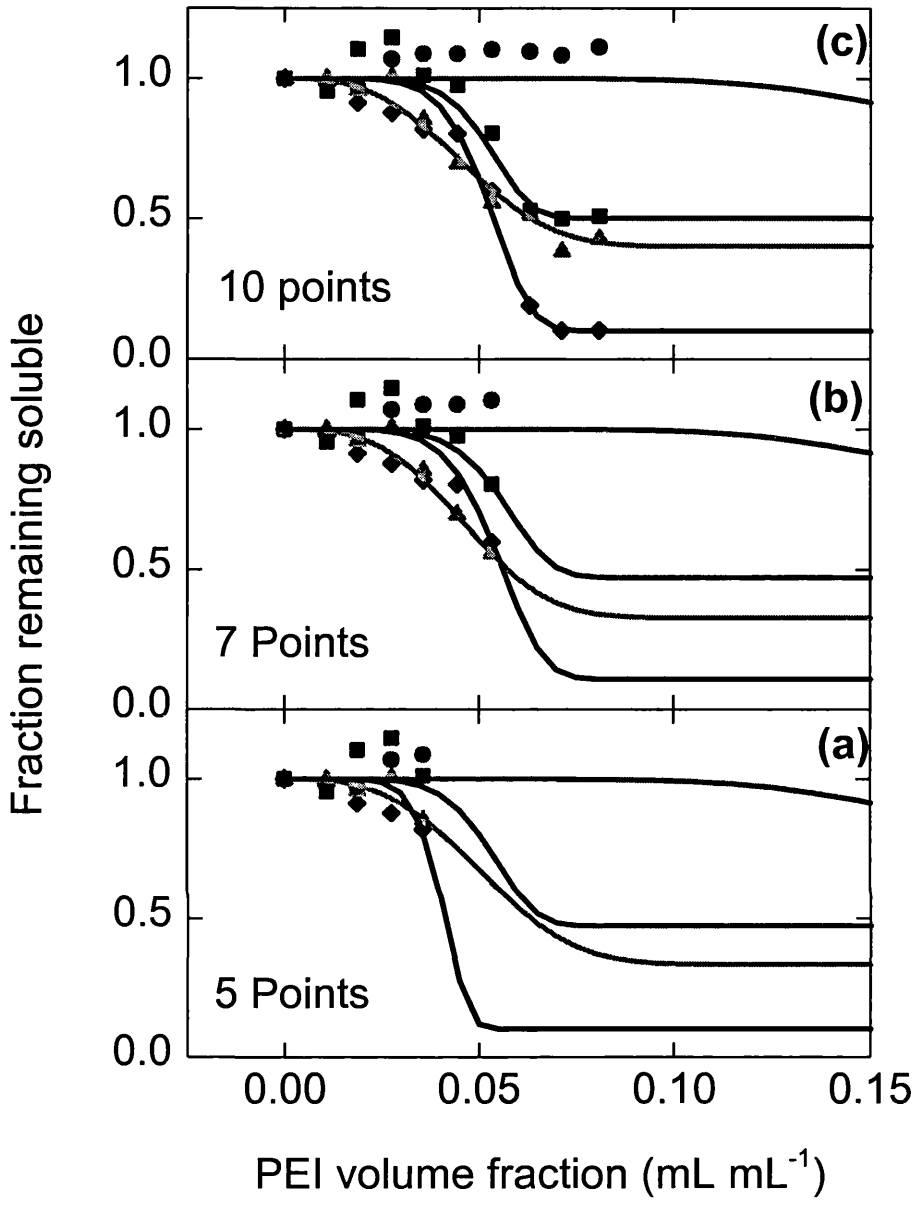


Figure 6.10. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run2. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

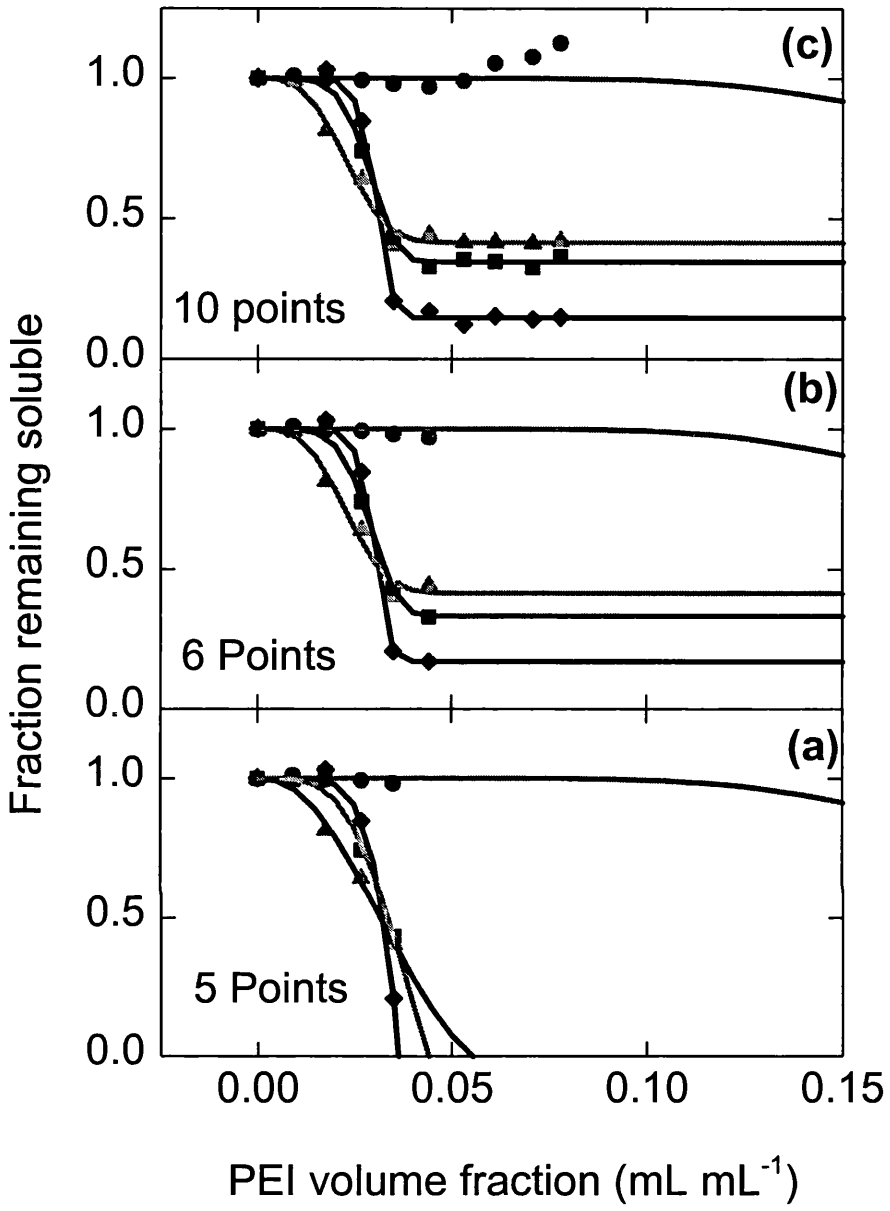


Figure 6.11. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run3. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

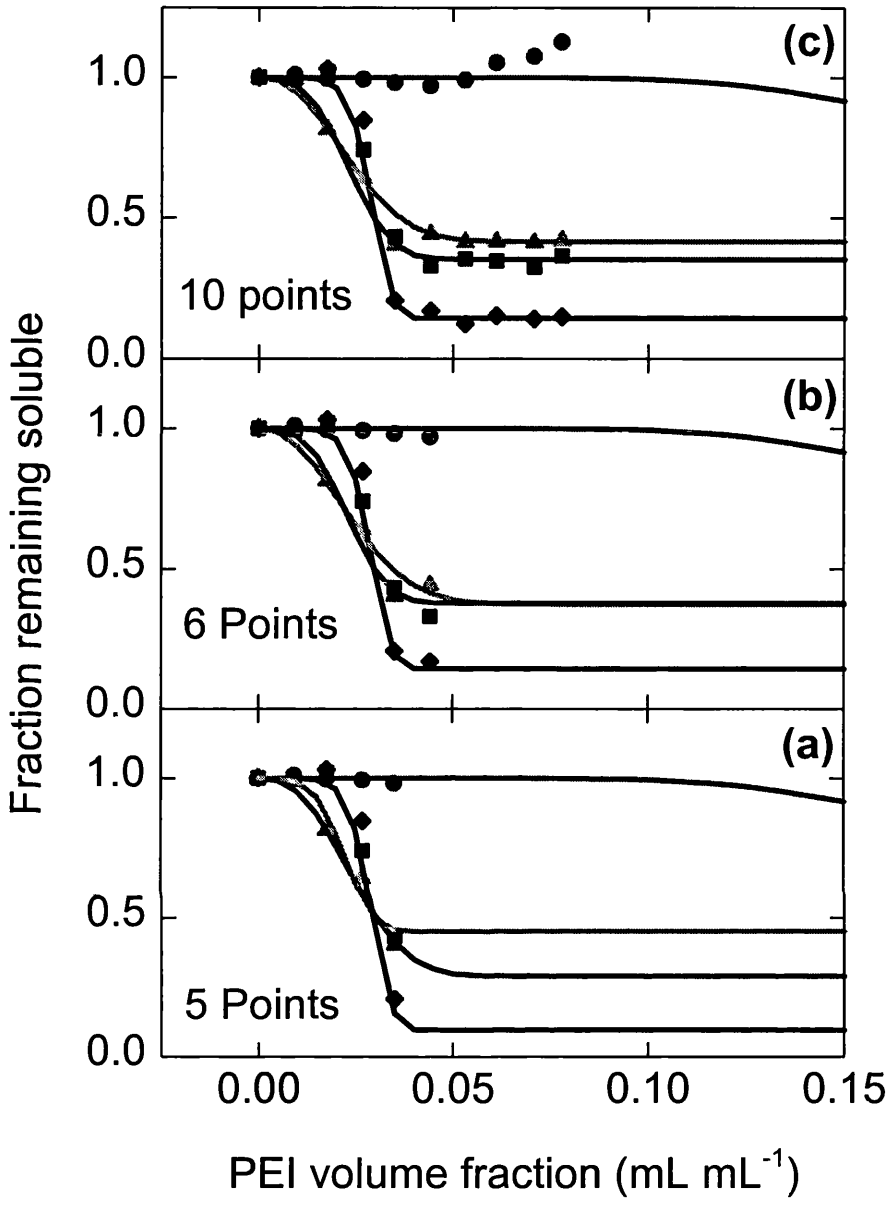


Figure 6.12. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run3. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 7 and 10 data points. The initial guesses for the model parameter are listed in table 6.1.*

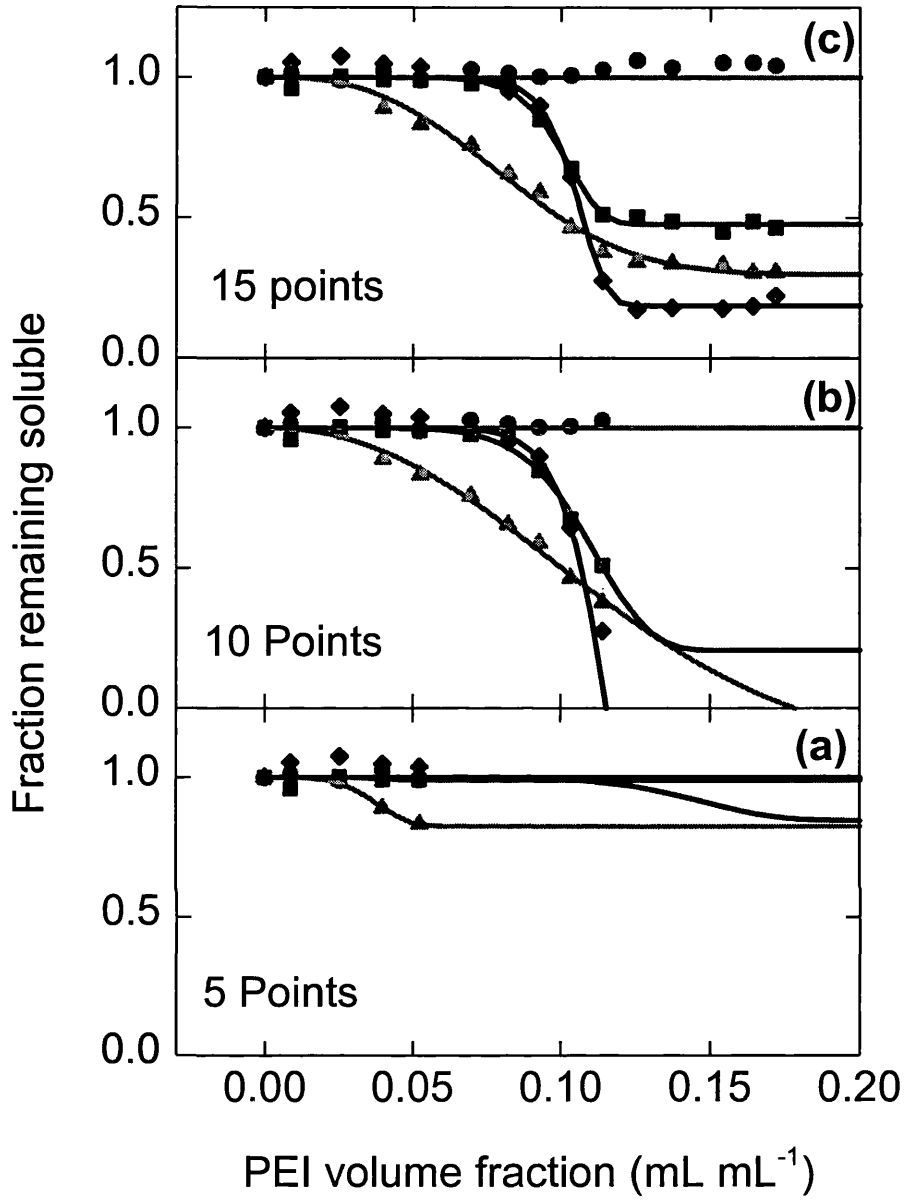


Figure 6.13. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run4. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

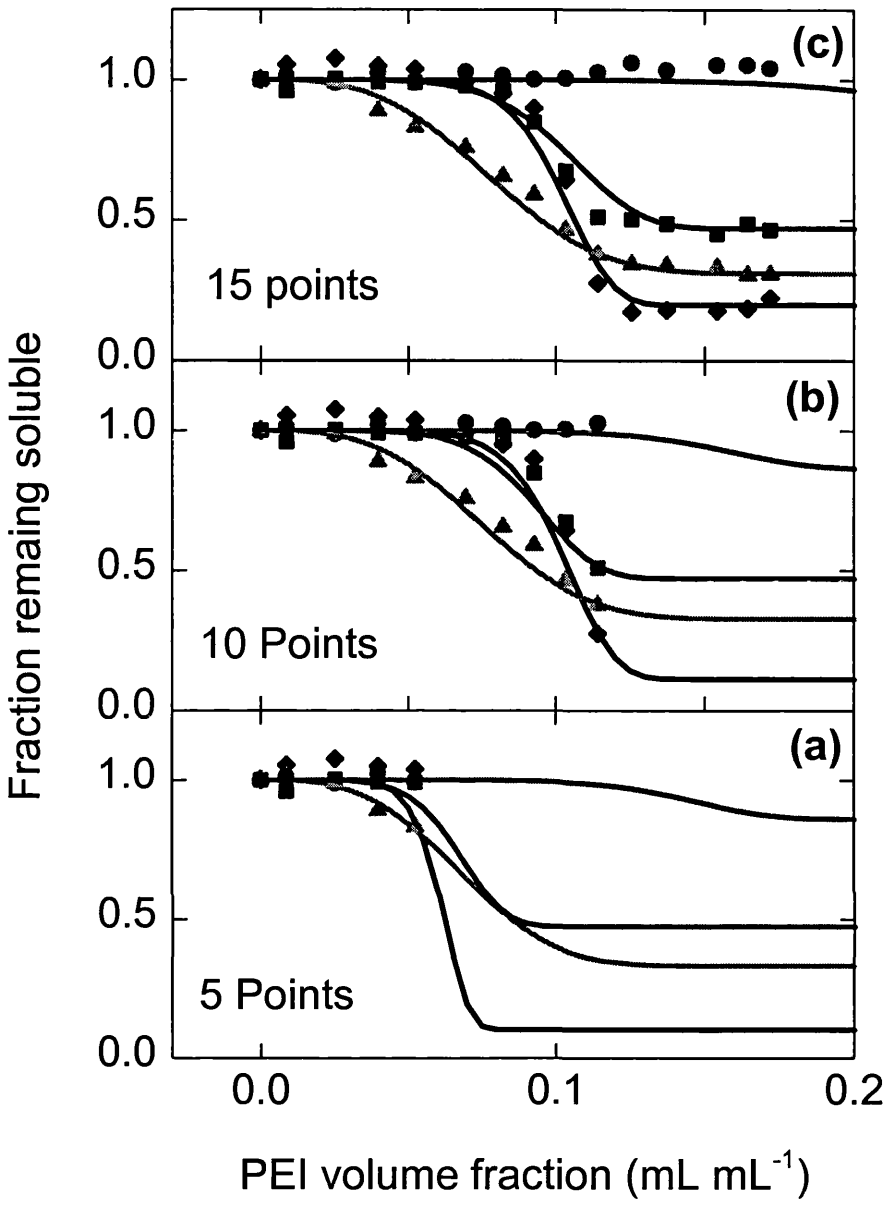


Figure 6.14. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run4. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

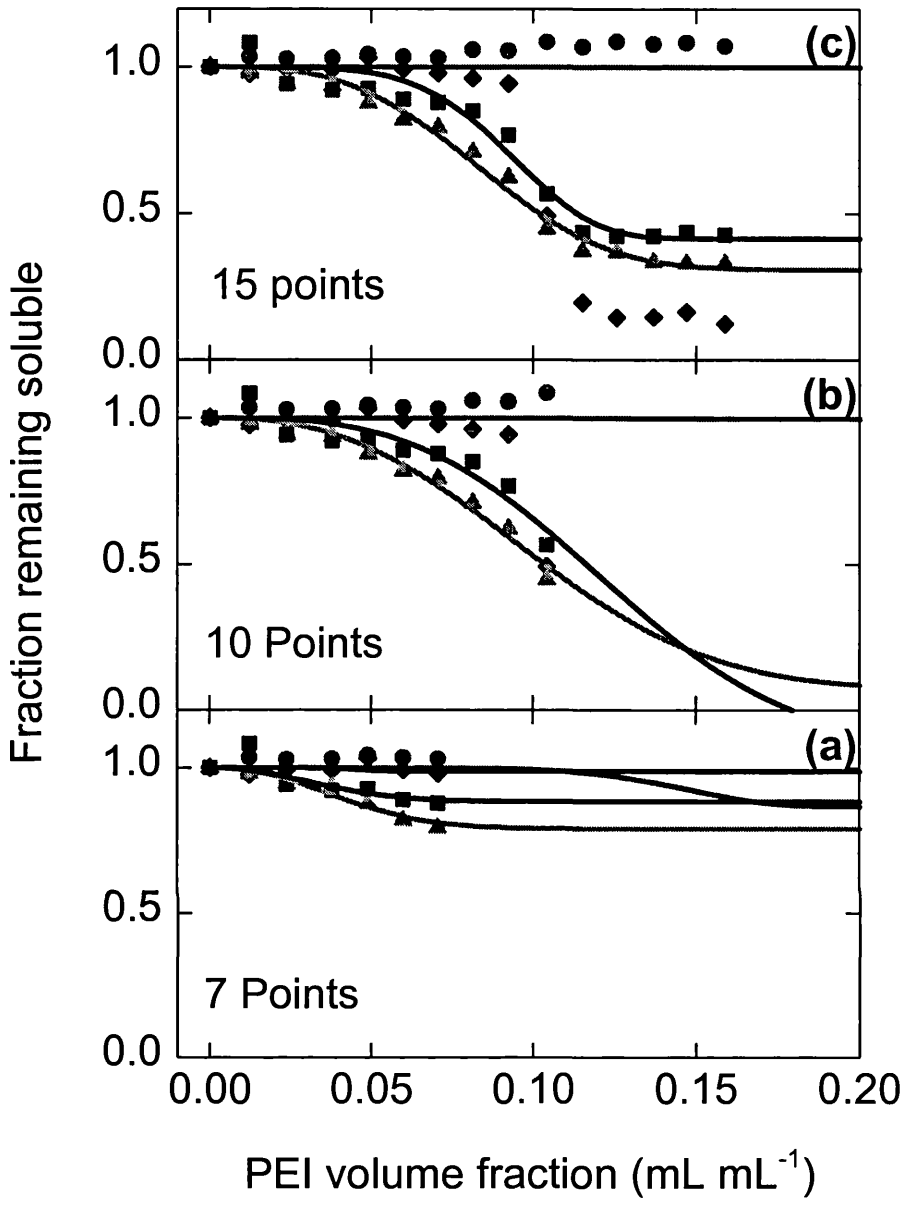


Figure 6.15. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run5. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 7, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

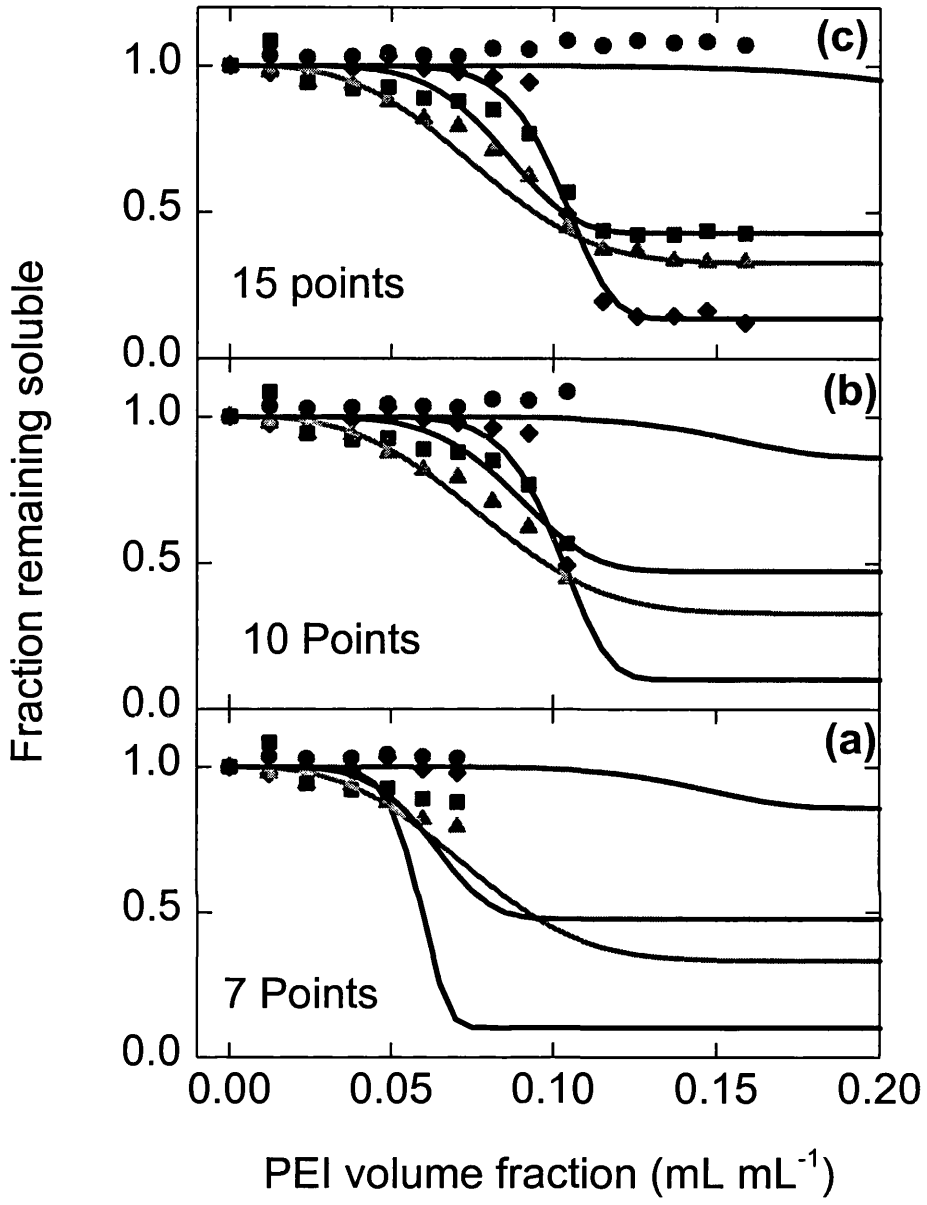


Figure 6.16. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run5. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

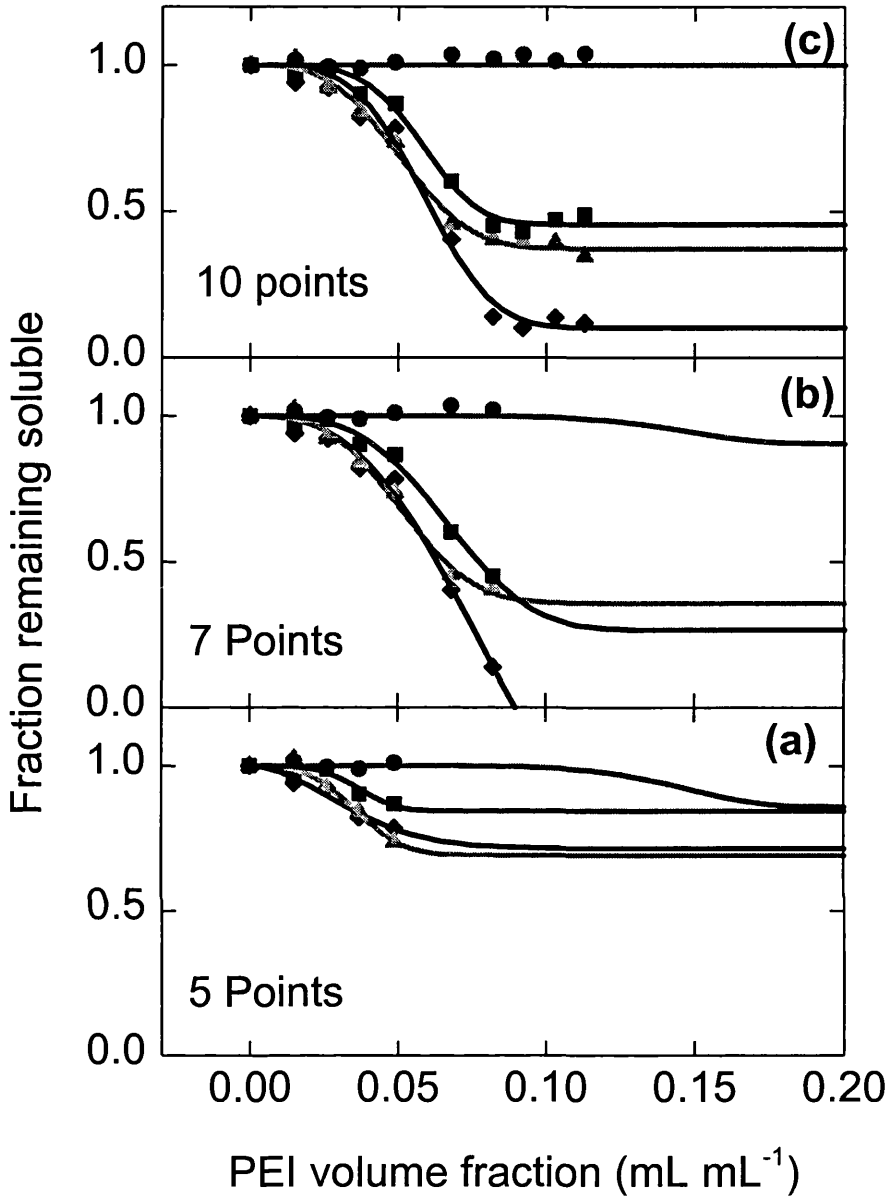


Figure 6.17. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run6. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 7, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

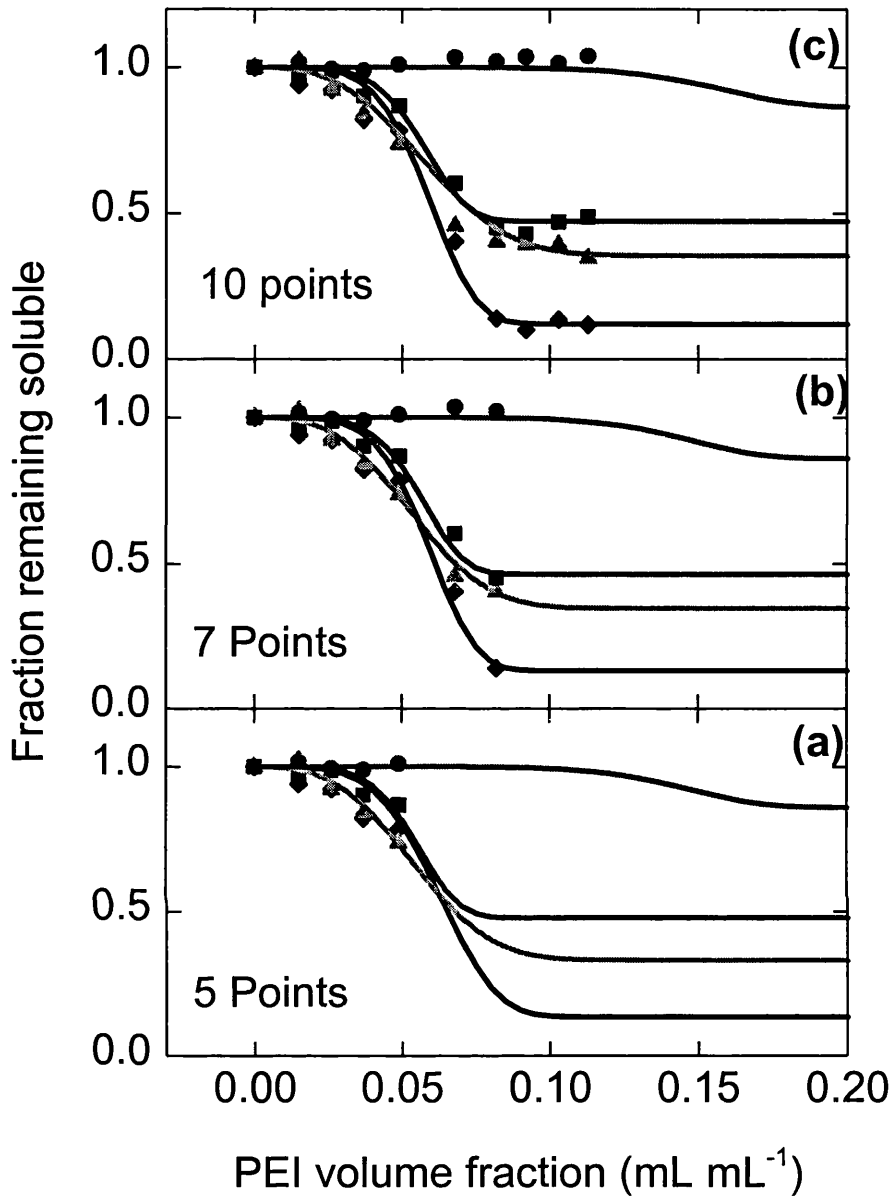


Figure 6.18. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run6. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

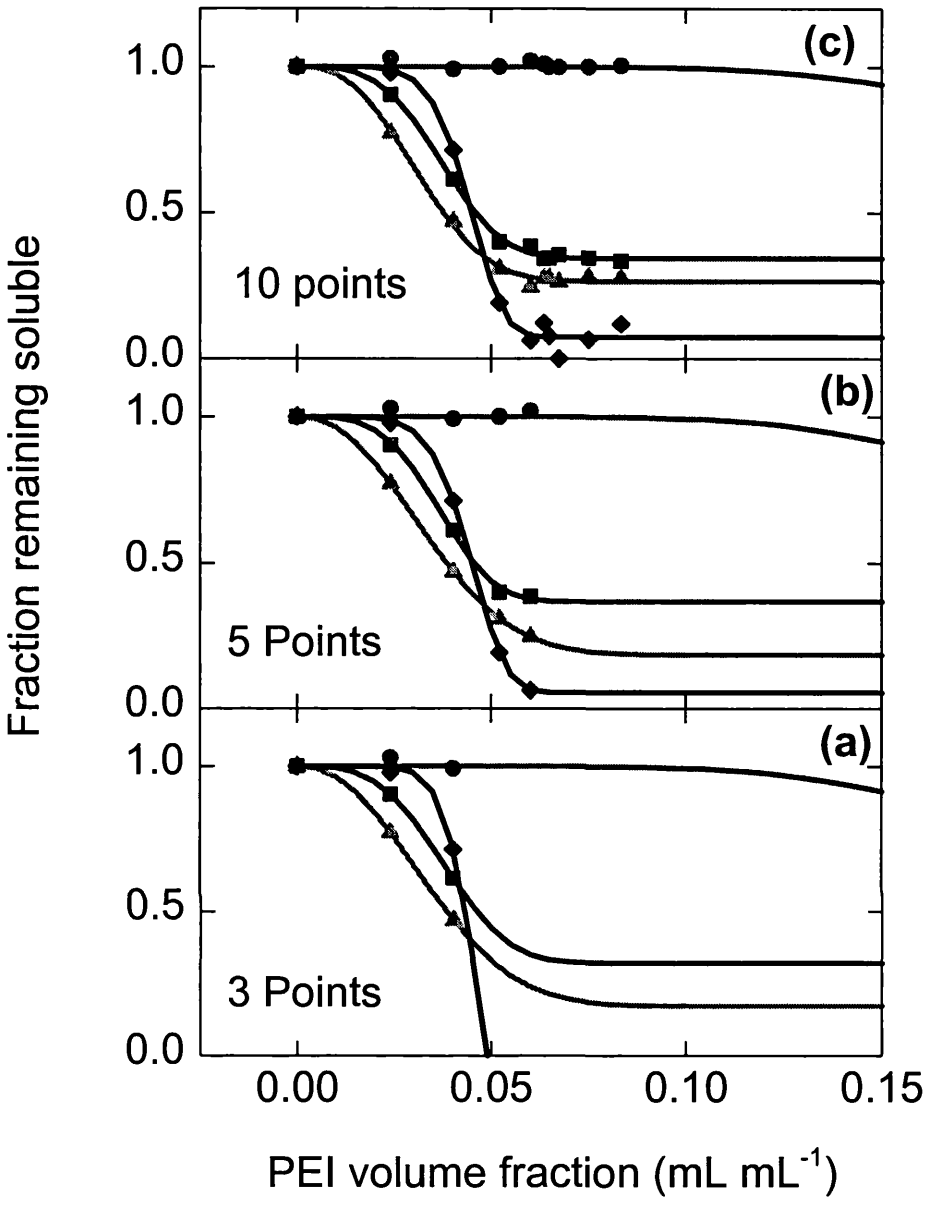


Figure 6.19. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run7. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 7, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

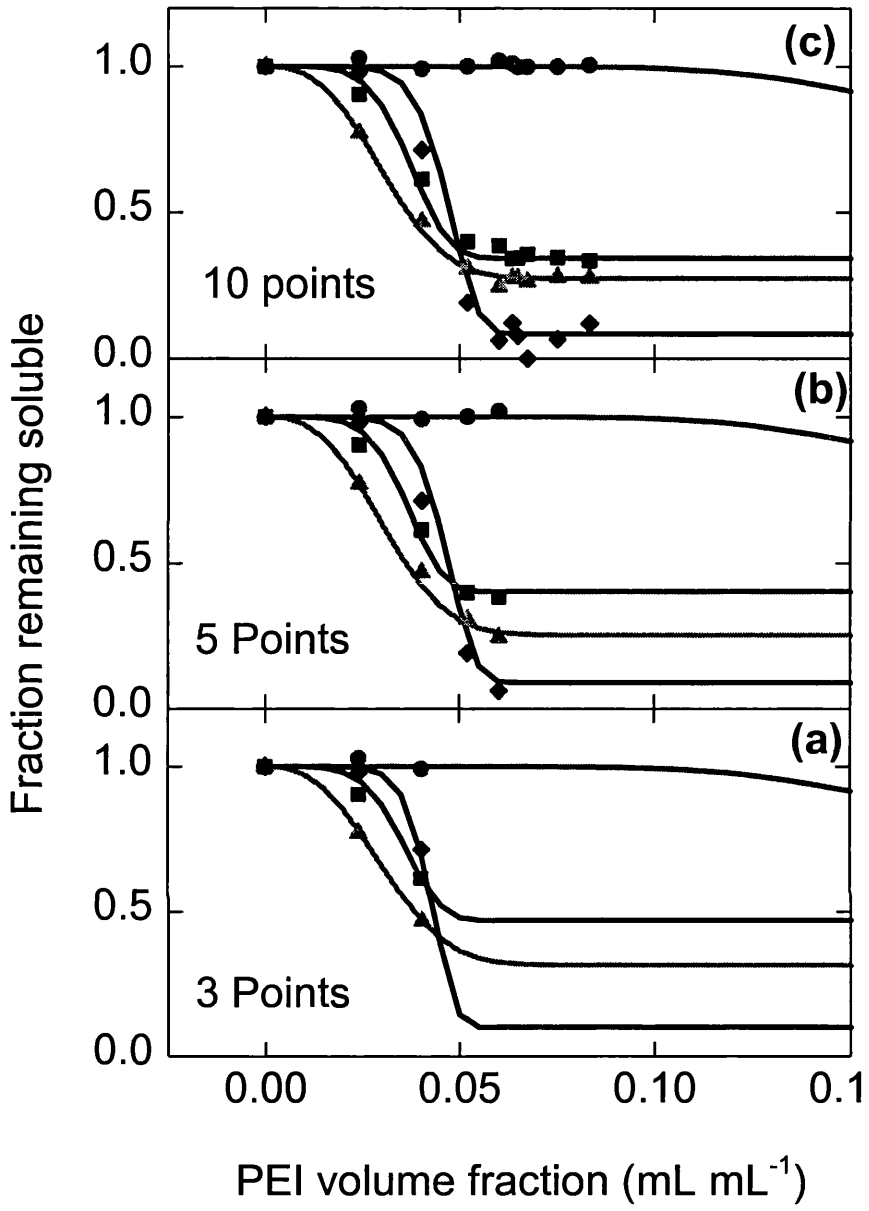


Figure 6.20. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run7. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

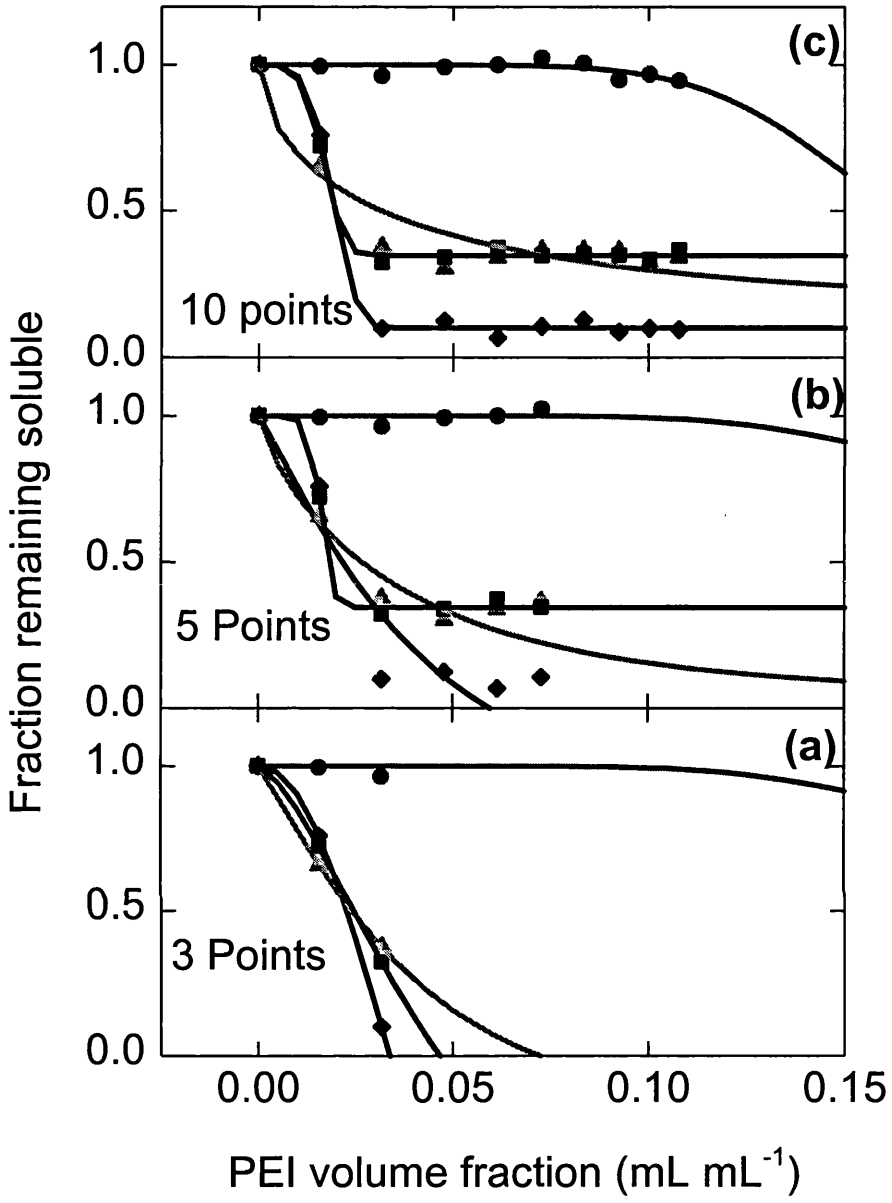


Figure 6.21. *Rapid characterisation of the flocculation profiles using the Levenberg-Marquardt non-linear least squares technique for Batch Run8. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 7, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

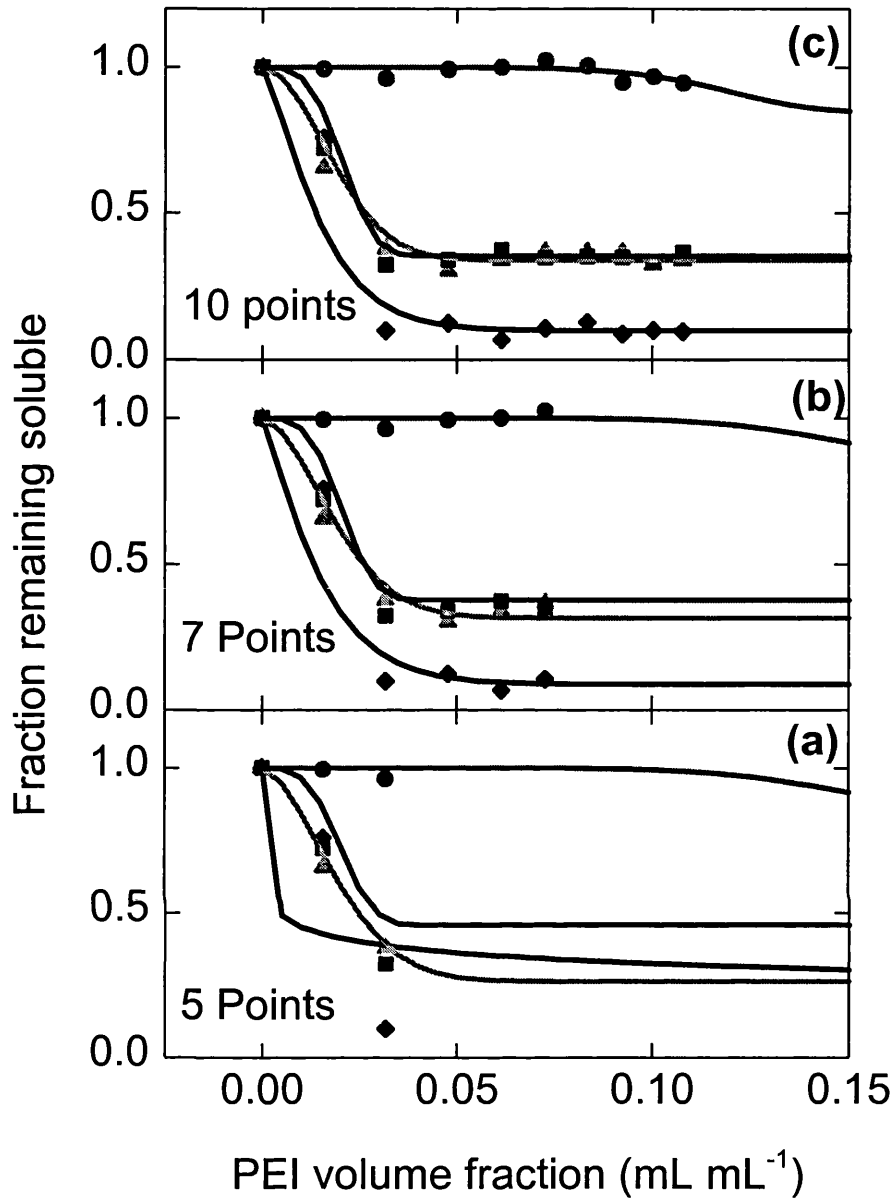


Figure 6.22. *Rapid characterisation of the flocculation profiles using the extended Kalman filter for Batch Run8. The symbols refer to RNA (▲), protein (■), cell debris (◆) and ADH (●). Plots a, b and c illustrate the profiles (—) of contaminants and product predicted using the first 5, 10 and 15 data points. The initial guesses for the model parameter are listed in table 6.1.*

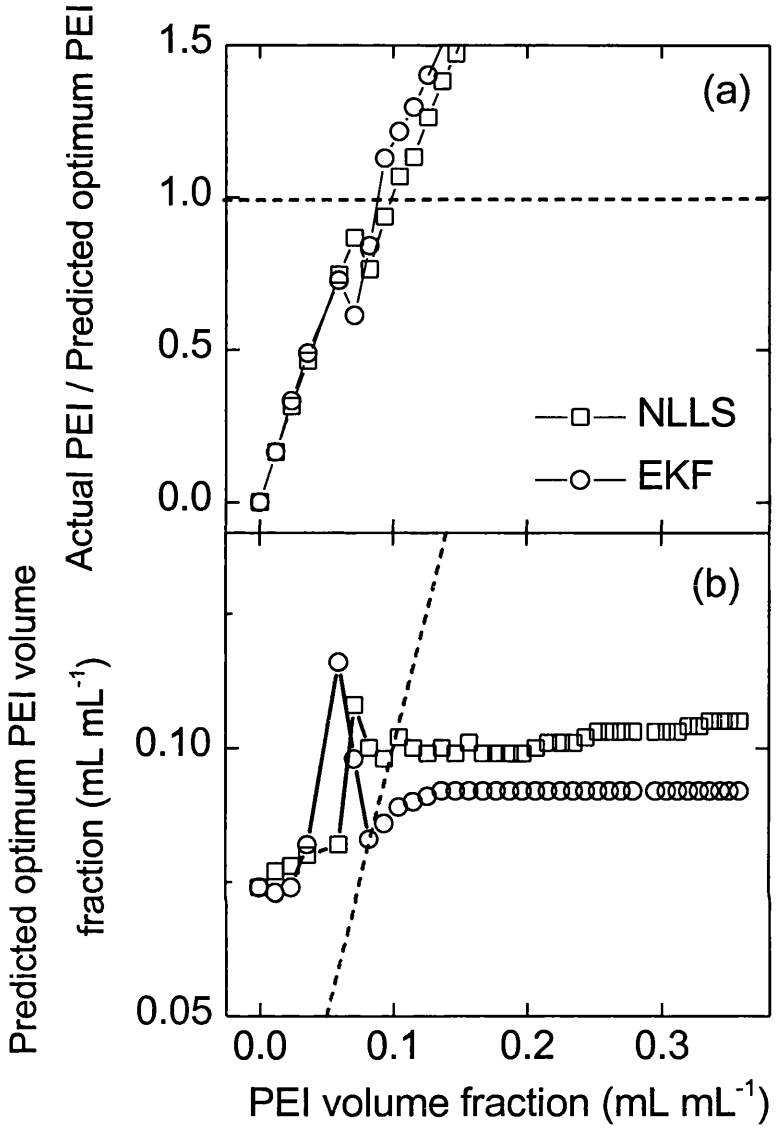


Figure 6.23. *Real time optimisation of Batch Run1, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a. *Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b. *Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3 on page 173. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

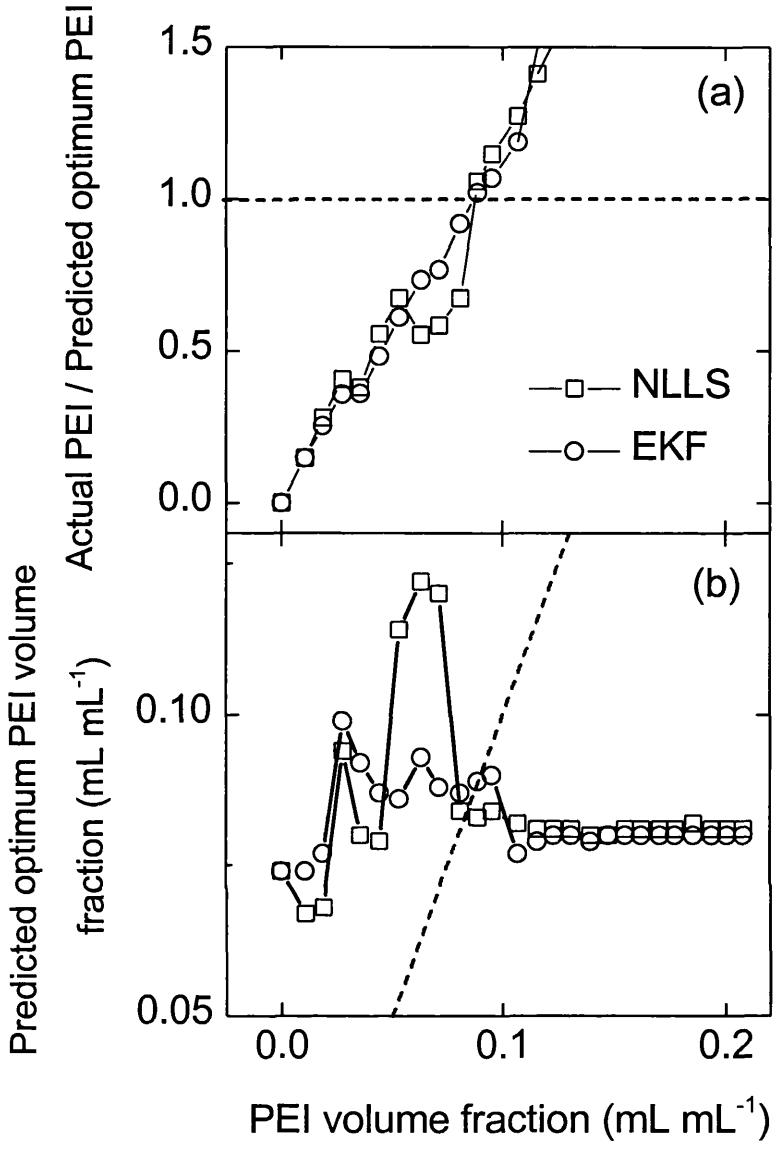


Figure 6.24. *Real time optimisation of Batch Run2, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a. *Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}$ (k)) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}$ (k).*

Plot b. *Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

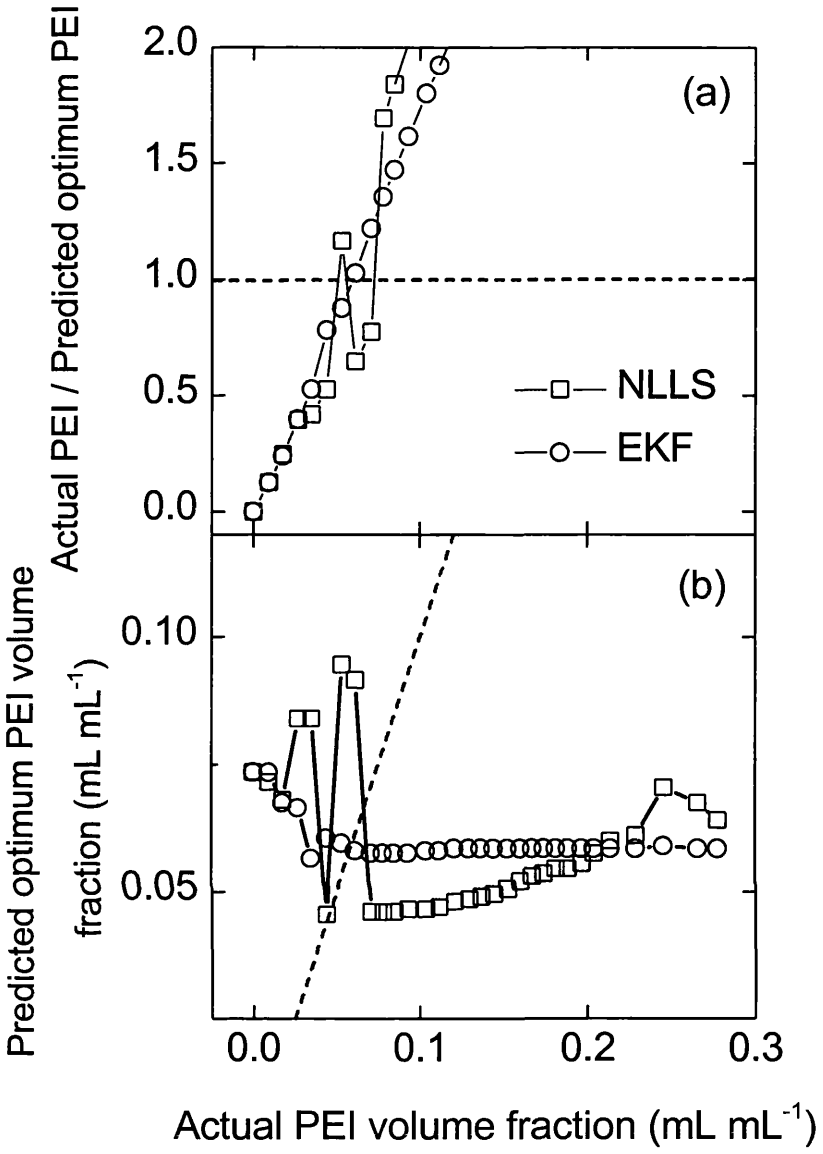


Figure 6.25. *Real time optimisation of Batch Run3, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a. *Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b. *Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

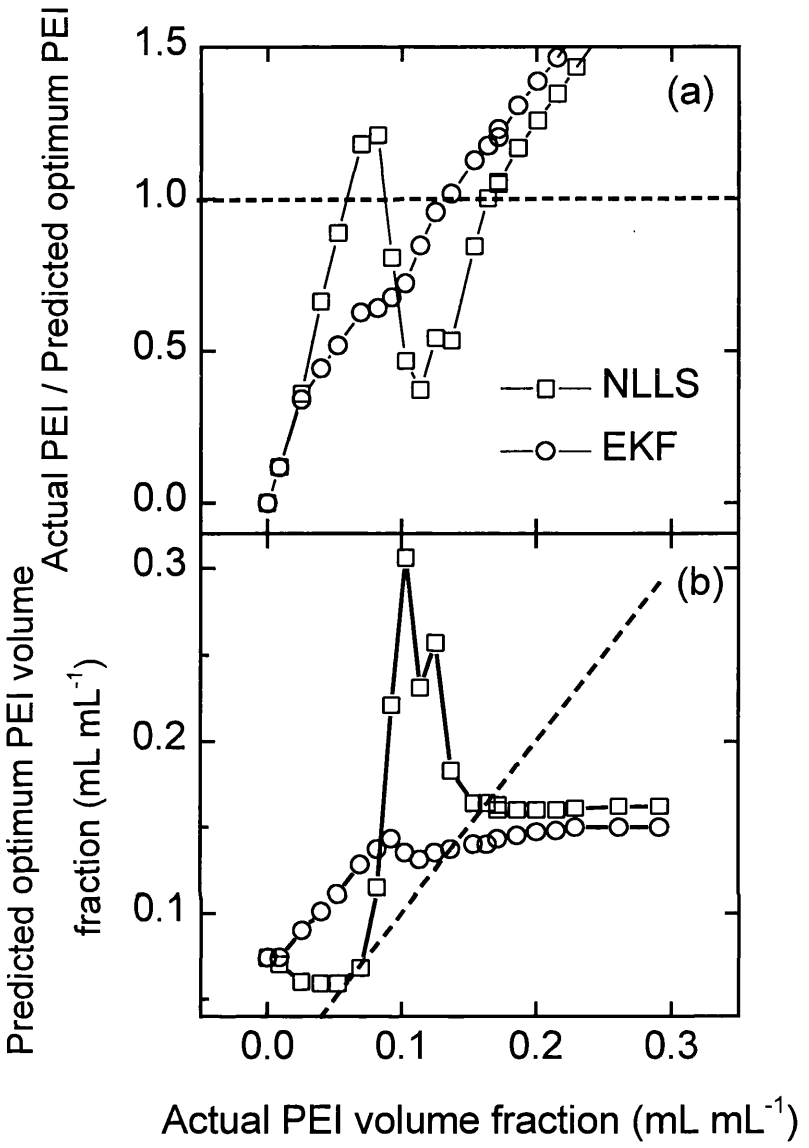


Figure 6.26. *Real time optimisation of Batch Run4, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a*. Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\ Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b*. Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

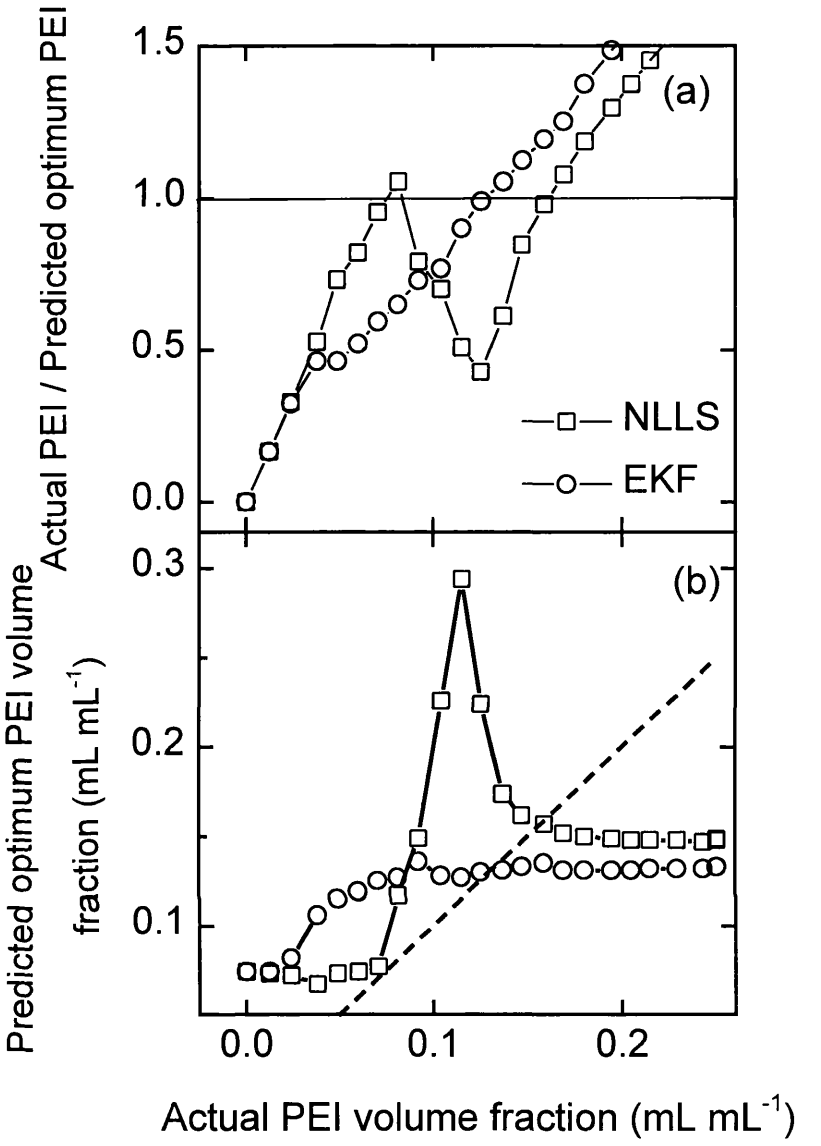


Figure 6.27. *Real time optimisation of Batch Run5, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a*. Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b*. Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

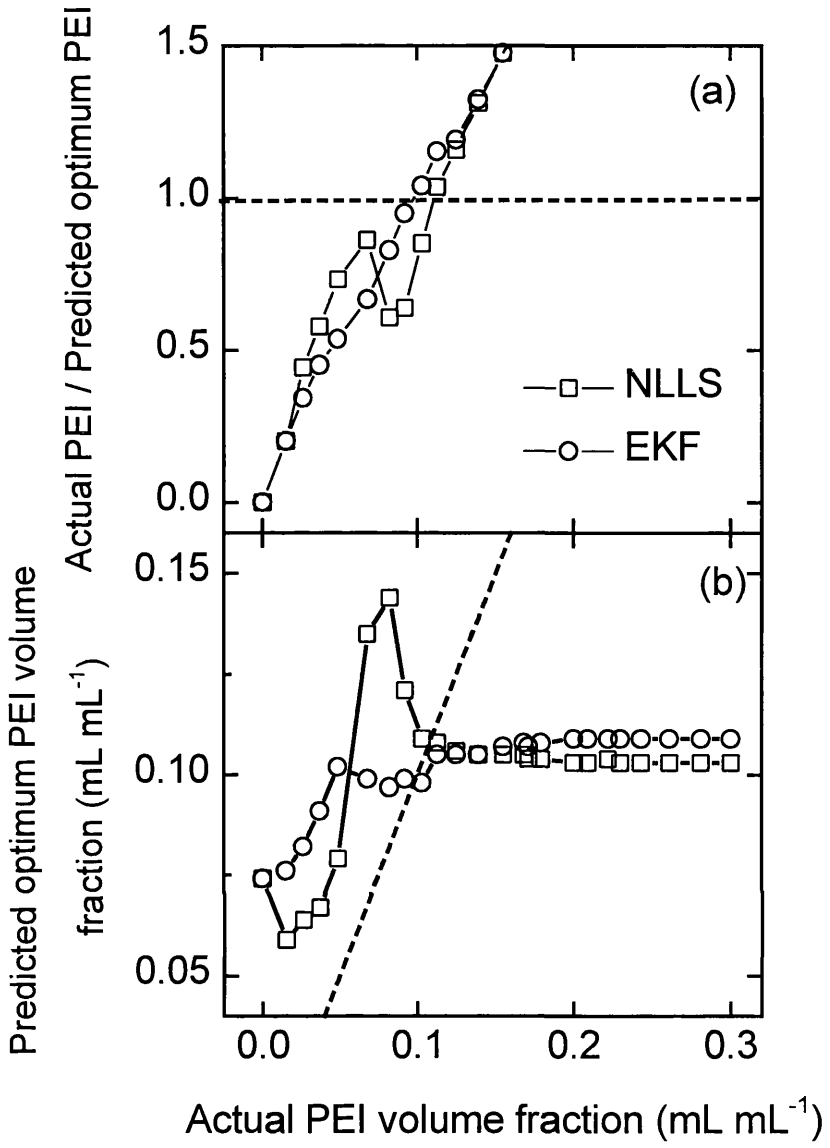


Figure 6.28. *Real time optimisation of Batch Run6, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a*. Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b*. Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

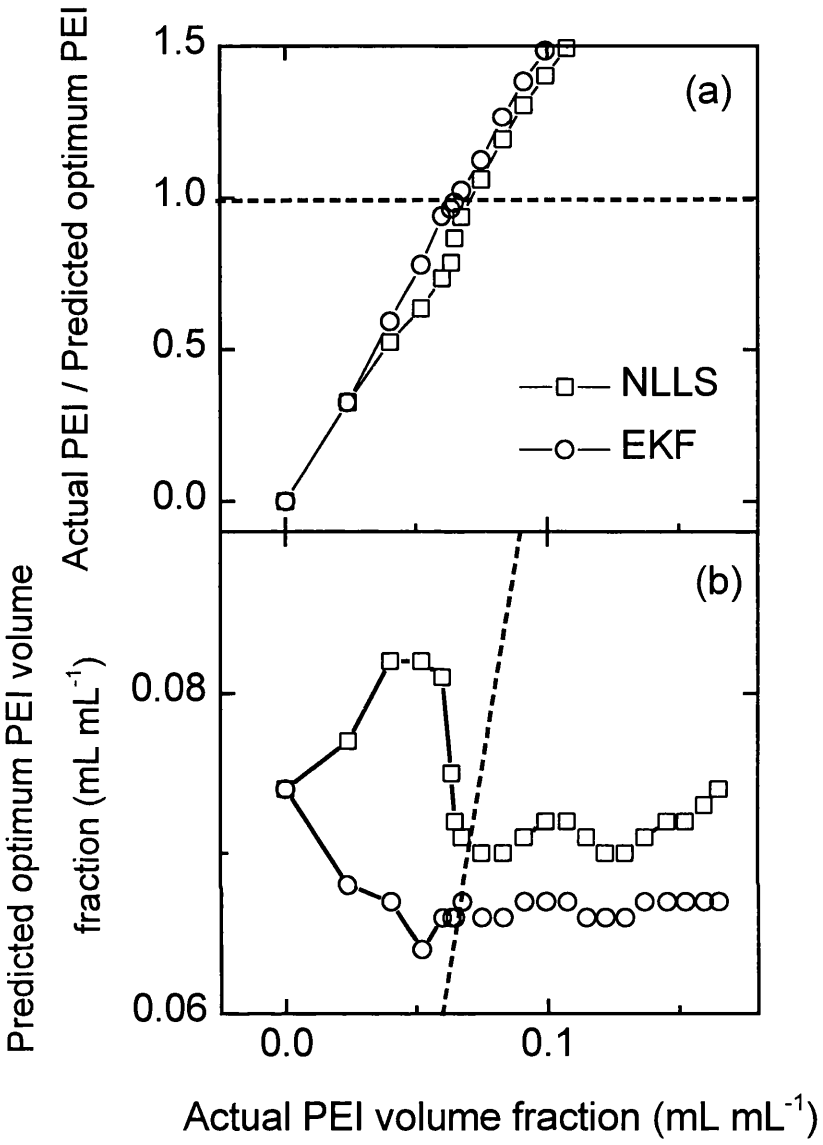


Figure 6.29. *Real time optimisation of Batch Run7, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a*. Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\_Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b*. Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

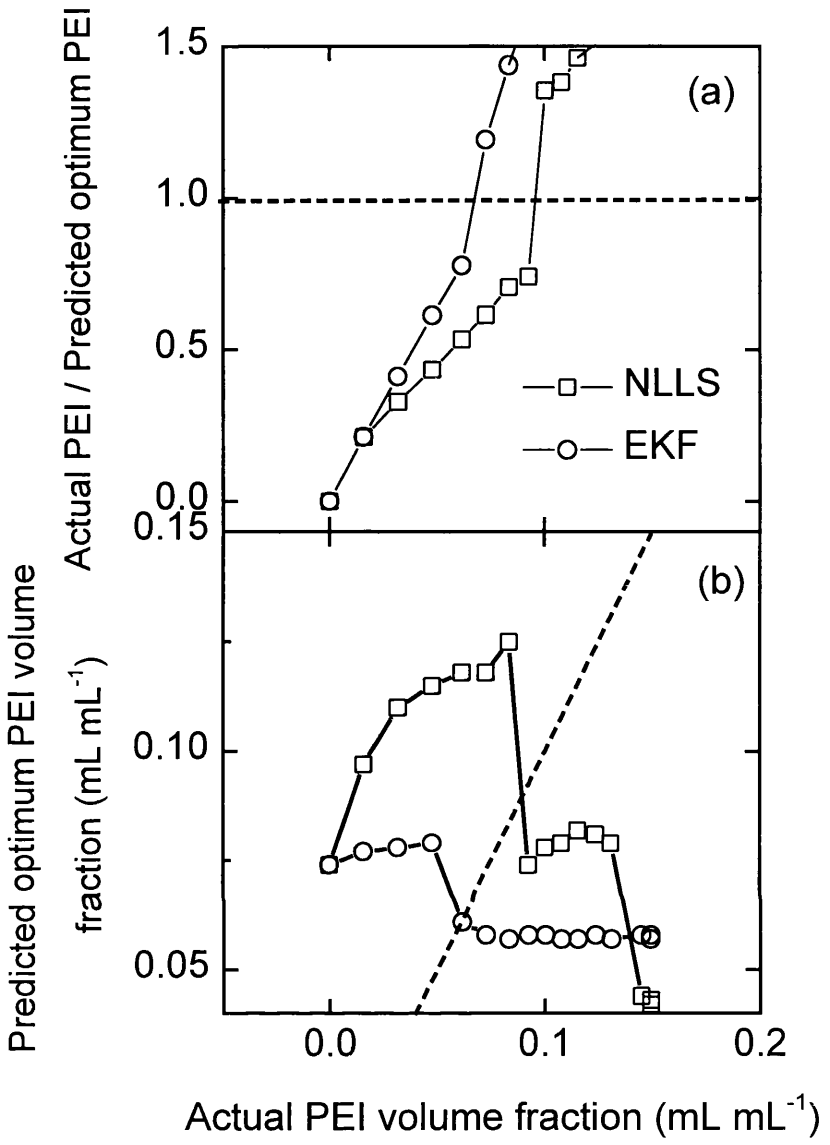


Figure 6.30. *Real time optimisation of Batch Run8, implementing EKF and NLLS estimation techniques, in terms of PEI volume fraction.*

Plot a*. Actual PEI volume fraction at step k+1 ($x_{PEI\_Actual}(k+1)$) divided by the predicted optimum PEI dose ($x_{PEI\_optimum}(k)$) at step k versus of the actual flocculant volume fraction, where k refers to the measurement step. The dashed line (- - -) indicates where $x_{PEI\ Actual}(k+1) = x_{PEI\_optimum}(k)$.*

Plot b*. Predicted optimum PEI volume fraction levels implementing the EKF and NLLS estimation techniques for each newly acquired measurement set. The optimisation procedure was illustrated in figure 6.3. The weighting factors imposed on the product, contaminant and PEI dose were $w_{ADH} = 0.5$, $w_{RNA} = 0.1$, $w_{Protein} = 0.05$, $w_{Debris} = 0.25$ and $w_{PEI} = 0.1$. The initial guesses for the model parameter and covariance constants for the EKF are listed in table 6.1. The dashed line (- - - -) refers to the actual PEI volume fraction level.*

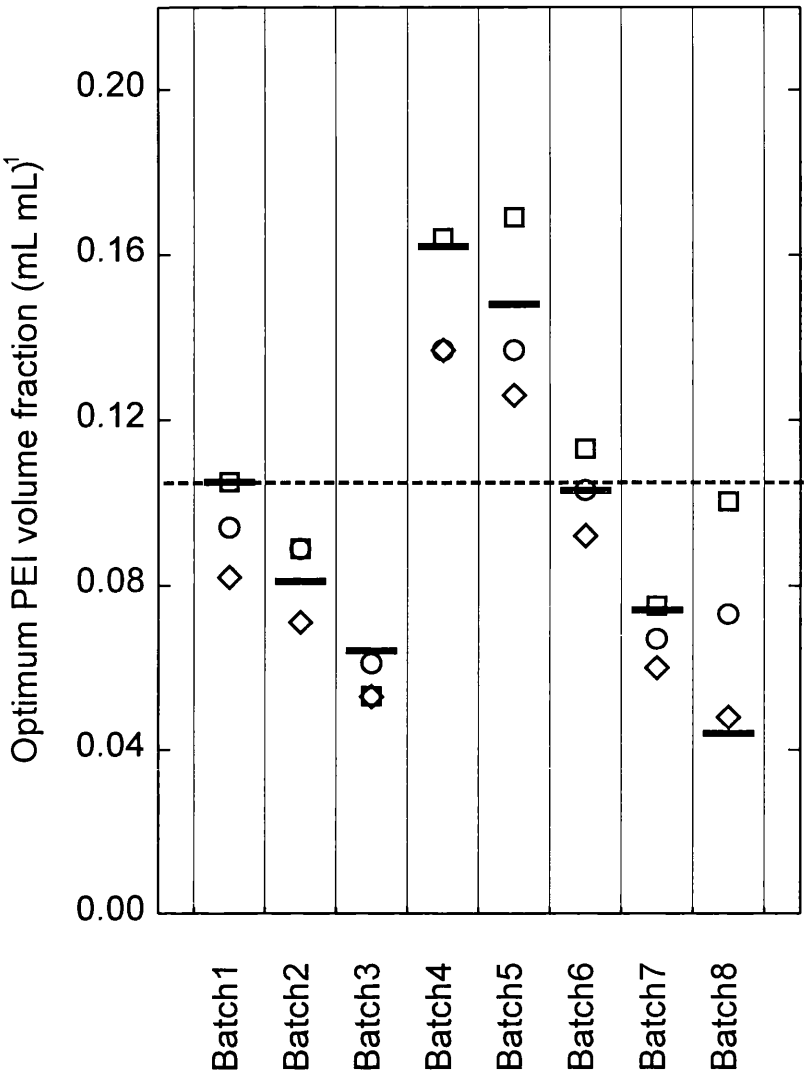


Figure 6.31. *Performance of the batch control configurations in terms of controlling the PEI dose to the optimum flocculant level. The below symbols refer to controlled PEI levels implementing a specific control configuration and the PEI dosing was terminated when their control objectives were met for each newly acquired measurement set.*

(—) Is the optimum flocculant setpoint determined using the optimisation algorithm illustrated in figure 6.3, applying all the available measurements and the NLLS estimator.

(◊) Controlled to a setpoint of at least 80% cell debris removal based on spun feed (control configuration 2).

(○) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 6.3, implementing the EKF (control configuration 1 with EKF).

(□) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 6.3, implementing the NLLS (control configuration 1 with NLLS).

(----) The horizontal dotted line refers to the average optimum volume fraction of all eight batch runs.

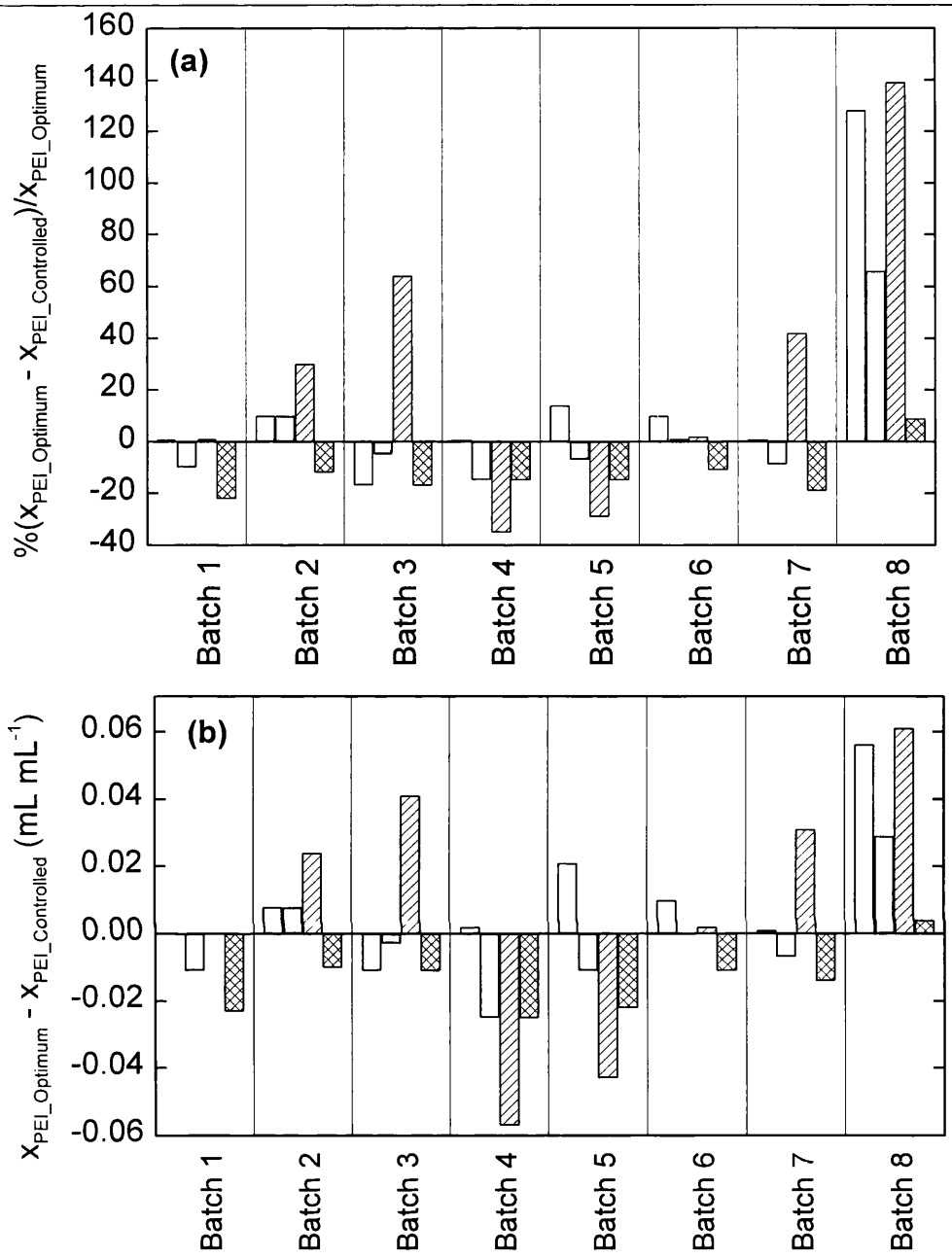


Figure 6.32. *Quantitative performance of the batch control configurations in terms of controlling the PEI dose to various optimum flocculant setpoints.* ***Plot a****. Percentage difference between the controlled and optimum flocculant setpoint.* ***Plot b****. Difference between the controlled and optimum flocculant setpoint in terms of PEI volume fraction. The below symbols refer to the different control configurations applied.*

(▭) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 6.3, implementing the NLLS estimator.

(▭) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 6.3, implementing the EKF estimator.

(▨) Controlled to a setpoint of at least 80% cell debris removal based on spun feed.

(▩)Controlled to a constant flocculant dose based on the average optimum PEI dose of all eight batch runs.

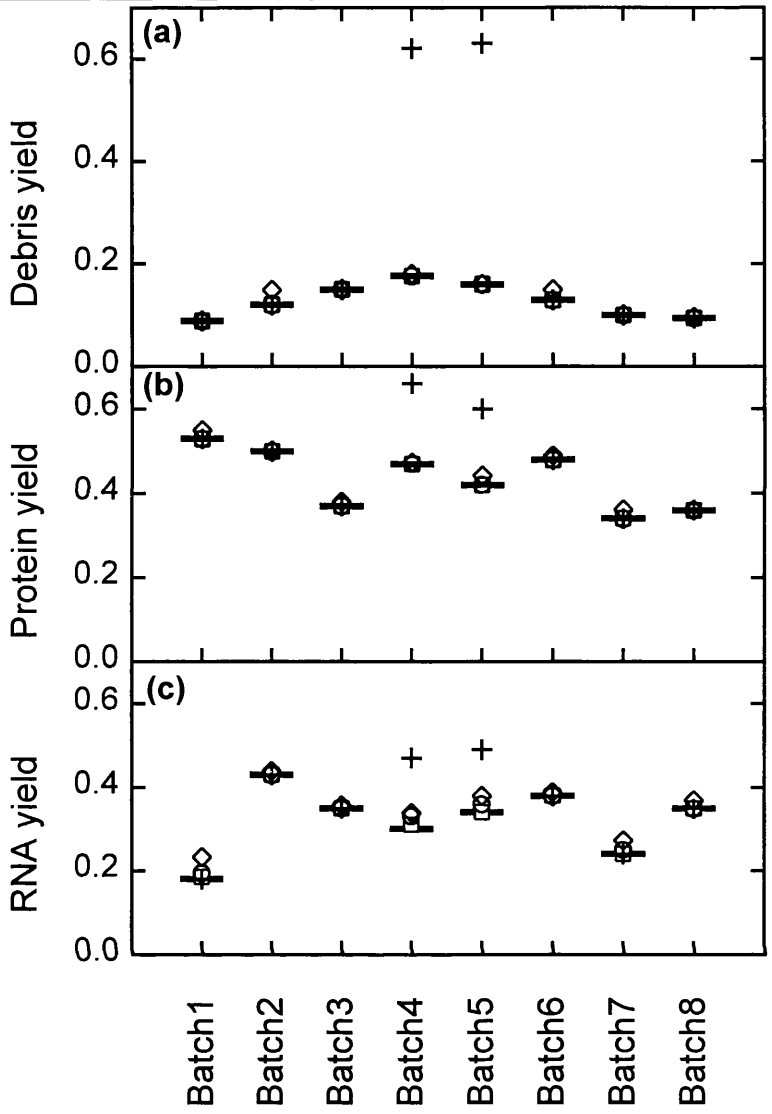


Figure 6.33. *Performance of the batch flocculation control configurations in terms of contaminant removal. The below symbols refer to the different control configurations applied. The contaminant yield values were calculated by substituting the controlled PEI levels determined in figure 6.31 into the 3(2)-parameter model. The parameters applied are listed in table 5.3 (p. 157).*

(—) Optimum removal level based on model parameter A (see table 5.3).

(□) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 7.3, implementing the NLLS estimator (control configuration 1).

(○) Controlled to optimise the batch operations according to the optimisation algorithm illustrated in figure 7.3, implementing the EKF estimator (control configuration 1) .

(◊) Controlled to a setpoint of at least 80% cell debris removal based on spun feed (control configuration 2) .

(+) Controlled to a constant flocculant dose based on the average optimum PEI dose of all eight batch runs.

Chapter 7

Rapid monitoring and characterisation of the continuous flocculation process

7.1 Summary

Real time measurements of product and key contaminants for three continuous flocculation processes will be presented. The rapid measurements were implemented for flocculation profile characterisation adopting the Levenberg-Marquardt non-linear least squares (NLLS) and extended Kalman filter (EKF) model identification techniques. The primary objectives put forward were the speed and stability in which the continuous flocculation process could be characterised. Two searching methods were investigated in terms of how different sequences of at-line information would influence the behaviour of model parameter estimations. A sequential and controlled search technique was examined and it was concluded that by applying a more structured and intelligent use of the at-line data the characterisation of the flocculation process could be improved.

7.2 Introduction

In the previous chapters the ability to monitor the batch flocculation process in terms of both product and key contaminants and the subsequent characterisation using a mathematical description (3(2)-parameter model) was demonstrated. It was also shown that by implementing at-line model parameter estimations, flocculation profile predictions could be used to optimise the batch processes reproducibly.

This chapter will evaluate the performance of two estimation techniques (NLLS and EKF) for at-line model parameter identification for the continuous flocculation process. During the continuous flocculation operating mode the level of flocculant may be varied freely by altering the flowrate ratio between the feed homogenate and PEI streams. Hence flocculant overdosing was not seen as a major problem as the possibility of reducing the PEI volume fraction was present. Emphasis was put on the speed and stability of process characterisation.

7.3 Searching methods for flocculation characterisation

This section will discuss the rapid characterisation methods adopted in this chapter.

Figure 7.1 illustrates the rapid process monitoring and characterisation configuration applied. Real time data was acquired through the rapid monitoring set-up described in chapter 3. The at-line information on both product and key contaminants was fed to two model parameter identification techniques; the Levenberg-Marquardt non-linear least squares (NLLS) method and extended Kalman filter (EKF) for model parameter prediction. The identification techniques implemented the 3(2)-parameter flocculation model (equation 7.1), in order to estimate model parameters for each newly acquired data point.

$$y_{3(2)} = \frac{E_i}{E_0} = 1 - A\left(1 - Exp\left(-\left(\frac{x}{B}\right)^C\right)\right) \qquad (7.1)$$

where E_i and E_O are the component (product and contaminants) concentrations before and after flocculation. In both cases the suspension has been treated by a solid-liquid separation step. The variable x is a measure of the flocculant dose and the model parameters A, B and C are dependent on system conditions (chapter 2).

After each monitoring step the estimated model parameters were applied to create flocculation yield curves and the operator could select the appropriate flocculant dose to either control the process to reach optimum conditions or to investigate additional operating areas to further characterise the process. This operating choice was dependent on the accuracy of the flocculation profile predictions. This work judged the prediction performance on the level of parameter estimation change. If this was low between iterations it was assumed that appropriate characterisation was achieved and no additional data points were required. Process optimisation and control could proceed.

The optimisation procedure applied was similar to that used during the batch flocculation operation (chapter 6). The objective of the optimisation algorithm was to ***"maximise the recovery of product and removal of contaminants whilst minimising the use of flocculant"***. This control objective was met through the manipulation of the flocculant volume fraction.

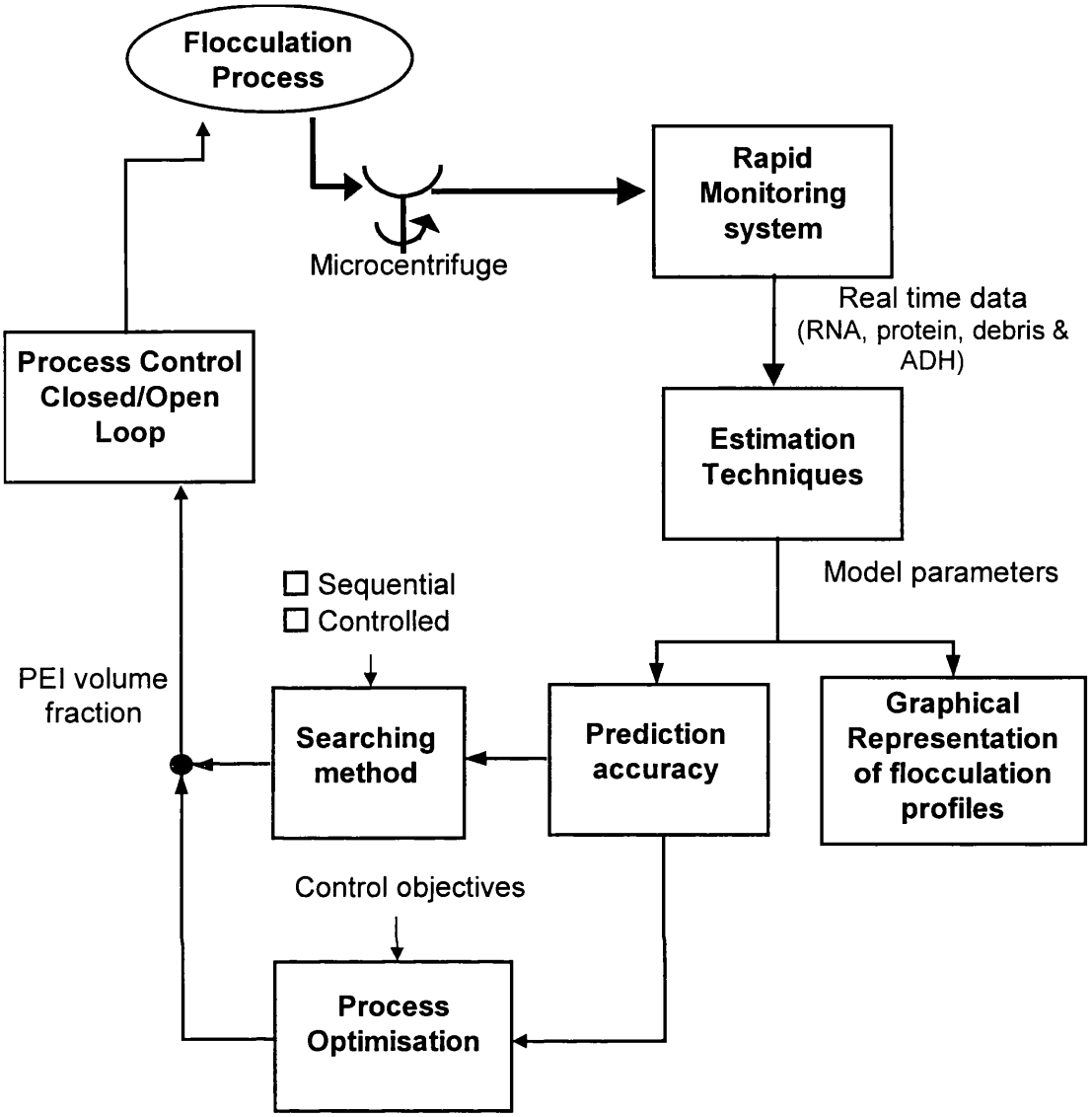


Figure 7.1. *Overview of rapid process characterisation and control configuration for the continuous flocculation process adopting either the sequential or controlled search methods.*

If the accuracy of the flocculation profile estimations were not satisfactory, more data points would be acquired through a ***"search method"*** to improve the prediction of the process characteristics. Figure 7.1 represents two different searching scenarios. The first searching technique was termed the ***sequential searching method***. Here the continuous flocculation process was operated in a batch fashion by increasing the PEI dose incrementally. The second searching technique was defined as the ***controlled searching method***. During this search mode the flocculant level was chosen so that the estimation techniques would be given sensitive flocculation information for rapid and effective model parameter identification, this procedure is illustrated in figure 7.2.

The searching procedure starts by implementing the initial parameter guesses (chosen by the operator) and the at-line data of the spun feed homogenate at measurement step $i = 0$. This will give the operator an initial guess of the cell debris flocculation profile and feed conditions which would be implemented to determine the subsequent flocculation yield values. Based on the initial prediction of the flocculation profile the next searching step ($i = 1$) was to target a relatively large PEI volume fraction level in order to obtain a good estimation of the maximum contaminant removal level. This enabled the effective prediction of model parameter A. The next step ($i > 1$) was to target the apparent flocculation inflection point based on the up to date flocculation profile prediction. The searching method would repeat this step until parameter B had stabilised sufficiently indicating that a relatively good estimate of the position of the inflection point was obtained. The subsequent step was to select flocculant levels around the profile inflection point as this region is vital for effective estimation of the optimum PEI dose. By acquiring data points around the inflection point good estimates of model parameter C were possible as this parameter influences the slope of the flocculation profile. The searching technique could be terminated when the parameter C estimates stabilised. It can be observed from figure 7.2 that after each decision step the PEI dose was controlled to the wanted flocculant level, the subsequent response was then measured by the at-line monitoring set-up and model parameters estimated adopting the NLLS identification technique.

The initial guesses for model parameters B and C for both the contaminants and product were chosen so that there was a 50% deviation from their actual values, whilst parameter A was taken as the average of all the continuous flocculation runs (listed intable 7.1). The choice of initial parameter guesses were based on knowledge of how the parameters deviated between runs during the batch operation observed in chapter 5. Parameter A tends to be relatively constant for all the flocculation runs despite variation in feed and PEI stock solution whilst parameters B and C had a deviation (coefficient of variation) of 44% and 52% from their actual values.

The EKF required additional start information to that of the NLLS technique, such as system and measurement noise terms and initial values for the combined system covariance term. The selection of these values enabled the operator to tune the EKF to either put more or less weighting on the at-line data for the estimation of new model parameters and the level of model parameter change between iteration steps.

Table 7.1 lists the various additional constants applied in the EKF during model parameter estimation of the continuous flocculation runs. The measurement noise term (Z), which was determined as the square of the standard deviation of the yield measurements discussed in chapter 4. As the 3(2)-parameter model was adopted the system covariance term (Q) for each contaminant and product was a 3 x 3 matrix consisting of the variance of each model parameter (no co-variance terms between model parameters were applied). The individual variance terms for the system noise matrix are listed in table 7.1, for the sequential (table 7.1(b)) and controlled (table 7.1(c)) searching methods.

During batch operation there is a gradual increase in the flocculant dose and hence gradual information of the flocculation yield. Hence little information on the flocculation profile gradient, position and maximum removal was available to the estimation technique initially. As discussed in chapter 2, if the initial guess for the initial combined system noise term was high, an initially high Kalman gain (weighting) would be the outcome, putting more weight on the initial measurements. By setting the initial combined system covariance (P) 3 x 3 matrix to zero no preconception was made that the EKF could determine all three model parameters based on a few measurements. The constants for the sequential search mode are listed in table 7.1(b). During the controlled search method the initial combined-system-covariance term was non-zero (table 7.1(c)) while the system-covariance term was fixed at zero.

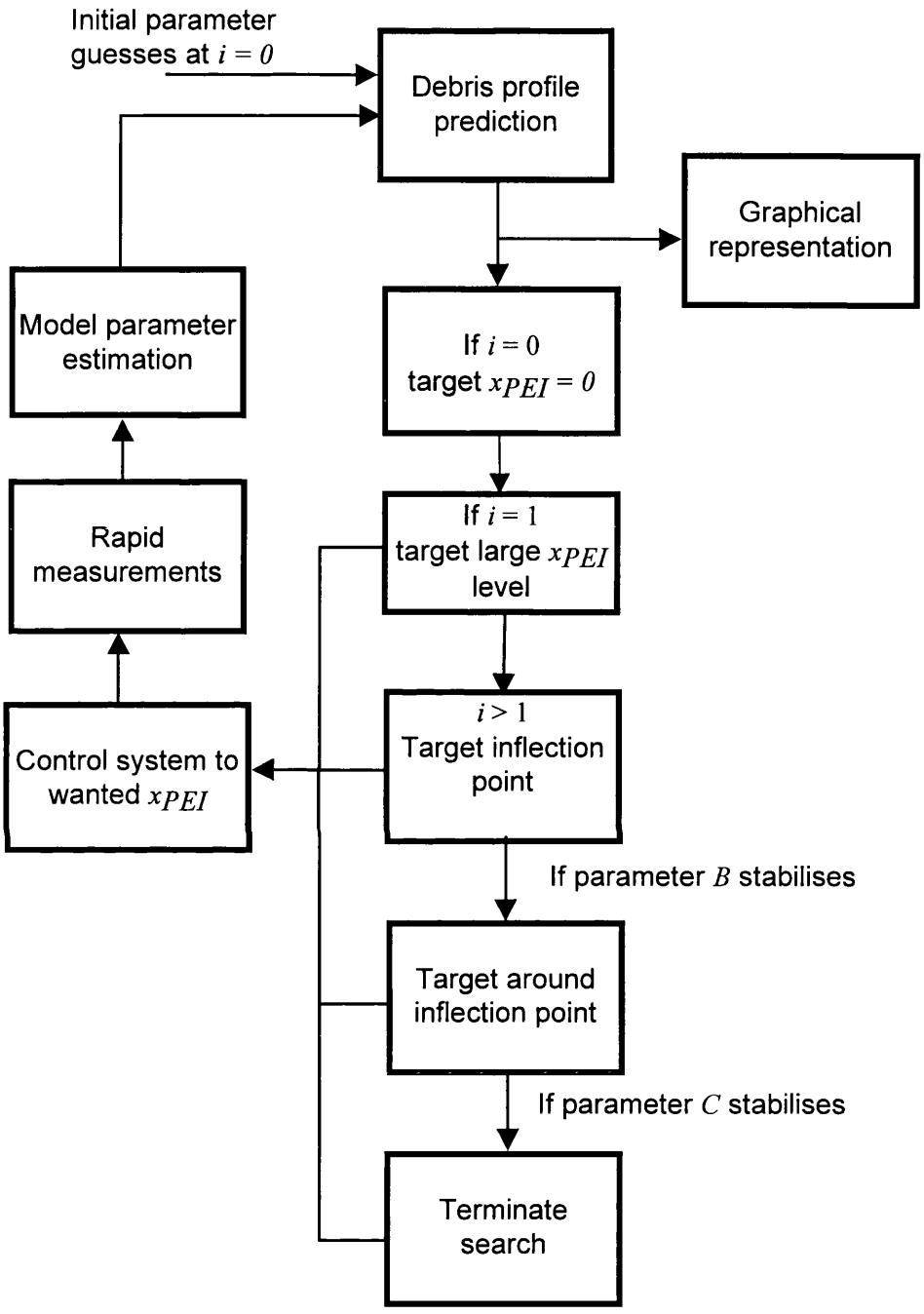


Figure 7.2. *Overview of the controlled search method.* x_{PEI} *refers to the PEI volume fraction and* i *refer to the measurement step.*

Table 7.1(a)

	Run1				Run2				Run3			
	Debris	**RNA**	**Protein**	**ADH**	**Debris**	**RNA**	**Protein**	**ADH**	**Debris**	**RNA**	**Protein**	**ADH**
$A_{Initial}$	0.82	0.71	0.63	0.1	0.82	0.71	0.63	0.1	0.82	0.71	0.63	0.1
$B_{Initial}$	0.068	0.038	0.047	0.25	0.09	0.06	0.069	0.25	0.19	0.2	0.18	0.25
$C_{Initial}$	2.7	1.9	2.3	3	4.3	1.9	2.5	3	9.5	5.5	3.7	3
Z	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007

Table 7.1(b) Sequential mode

q_{11}	0.002	0.004	0.004	0.002
q_{22}	0.0002	0.0003	0.0003	0.0003
q_{33}	0.5	0.5	0.5	0.5
$p_{11}=p_{22}=p_{33}$	0	0	0	0

Table 7.1(c) Controlled mode

$q_{11}=q_{22}=q_{33}$	0	0	0	0
p_{11}	0.002	0.004	0.004	0.0002
p_{22}	0.0002	0.0003	0.0003	0.0003
p_{33}	0.5	0.5	0.5	0.5

Table 7.1. *Overview of initial parameter guesses and tuning constants for the extended Kalman filter for the continuous flocculation runs. Table (a), lists the initial parameter guesses for each run and Z refers to the measurement noise term applied during the EKF. Table (b), EKF constants applied during the sequential search mode. Table (c) EKF constant applied during the controlled search mode. The constants p_{11}, p_{22}, p_{33}, q_{11}, q_{22} and q_{33} refer to the initial combined system noise terms and system noise terms specific to model parameters A, B and C.*

7.4 Results

The first part of this section will present three flocculation processes in terms of their at-line measurements of product and key contaminants as well as model fits implementing the Levenberg-Marquardt non-linear least squares (NLLS) estimation technique. The second part of this section will present the results obtained from the at-line model parameter estimations for the continuous flocculation Run1 in detail applying both the NLLS and an extended Kalman filter (EKF). A step by step example of how the controlled search method was implemented will be shown. Flocculation profile predictions will be presented, underlining the main points of interest. Optimum flocculant volume fractions were estimated after each measurement step and the profiles for each run presented in terms of prediction technique and searching mode applied.

Continuous run	1	2	3
ADH in feed (U mL^{-1})	122	132	158
Protein in feed (mg mL^{-1})	17.8	18.3	20
RNA in feed (mg mL^{-1})	3.3	3.3	3.5
Debris in feed (Au)	2.5	2.5	2.9
PEI stock solution (mg (PEI) mL^{-1})	10	10	5
Ionic strength in PEI and feed (mM)	50	50	50
Feed and PEI pH	6.5	6.5	6.5

Table 7.2. Overview of feed and PEI stock solution conditions for the three continuous flocculation runs.

7.4.1 At-line measurements of the continuous flocculation process

Figures 7.3 to **7.5** on page 233, plot b illustrate the at-line measurements of the continuous flocculation runs 1, 2 and 3 and their respective NLLS fits. The 3(2)-parameter model described in chapter 2 was applied to characterise the flocculation profiles. The model parameter estimates are listed in the figure legends. The sequence of flocculant volume fractions applied during the controlled searching mode is also given (plot a).

For all three runs it can be observed that a relatively high resolution of data points of both product and key contaminants was acquired giving the operator a detailed at-line characterisation of the flocculation profiles. It can also be seen that the noise

associated with the at-line data was slightly higher than those observed during the rapid monitoring of the batch processes (chapter 5). This is due to the instability of the flocculant pump especially at low flowrates resulting in noisy PEI volume fraction values, which was discussed in chapter 4, section 4.3.5. Due to the large fluctuations during low flocculant levels the minimum acceptable PEI volume fraction was 0.04 mL mL^{-1}.

The slight change seen in the flocculation profiles between runs 1 and 2 may be due to the small differences in feed conditions observed in table 7.2. A larger shift in flocculation profiles was observed in Run3 (figure 7.5). As table 7.2 indicates a 2-fold dilution was made to the PEI solution during Run3, hence approximately double the PEI dose in terms of volume fraction was required for the same level of contaminant removal. The dilution of the flocculant solution enabled lower PEI concentration to be examined hence increasing the resolution of data points around the initial part of the flocculation profile.

It can be observed that the maximum removal level of the contaminants (parameter A) for the three runs are relatively similar (given that the salt concentration is similar), confirming that small variations of parameter A are usually seen between runs. In all three flocculation runs product activation can be observed. Activation of ADH has been seen by several authors (Smith (1997) and Holwill *et al.* (1997)) and was discussed in more detail in chapter 5. The presence of PEI and/or the removal of a large number of contaminants may be the cause of this product activation.

7.4.2 Model parameter and flocculation profile prediction

Figures 7.6 to **7.8**, illustrate the step by step real time estimation of model parameters for the continuous flocculation Run1, adopting the EKF (circles) and NLLS (squares) identification techniques applying the sequential (open symbols) and controlled searching (closed symbols) methods, i.e. four different scenarios. Figure 7.6, shows the real time estimation of parameter A. The initial guesses of this parameter (table 7.1) were relatively good compared to their actual values due to the known low variation between runs. As described in section 7.3 the initial guess for the model parameters B and C were chosen so that they were 50% of their actual values. For runs 1 and 2 a too high a guess was deliberately made whilst for Run3 too low initial parameter values were chosen.

Although relatively good guesses were available for parameter A it can be observed that the NLLS identification technique adopting the sequential searching mode yielded erratic estimations compared to the three other cases.

Parameter A gives the operator an indication of the maximum contaminant removal level, hence at-line data around this area was required before good estimation of this parameter could be made. During the sequential searching method no information on the maximum removal contaminants levels was obtained during the first few measurements. However the EKF still rendered good parameter A predictions even when this searching mode was applied. As described in chapter 2 and 6 the behaviour of the Kalman filter in estimating model parameters is a function of several factors, such as the tuning of the Kalman gain and the size of the innovation term. In table 7.1 it can be observed that the system noise term (Q) was for parameter A (q_{11}) relatively small as the known variation of this model parameter was low. By selecting a low system noise term the EKF estimations was constrained to introduce only small changes between prediction steps. Hence as the initial guesses for parameter A were relatively close predictions adopting the Kalman filter were good.

Parameter A estimations were improved when the controlled search method was applied in combination with either of the estimation techniques. By initially searching in the high PEI volume fraction region a relatively good estimate of model parameter A was obtained. One of the conclusions from chapter 6 was that the point of parameter A stabilisation was dependent on the quality of information rather than the quantity of data points. By adopting the controlled search mode appropriate operating conditions could be chosen so that this parameter was predicted rapidly and effectively.

Figures 7.7 and 7.8 illustrate the at-line estimations of model parameter B and C, for the 4 cases described above, i.e. two estimation and search techniques. Parameters B and C give the operator an indication of the point and slope of the profile inflection, hence estimations of these model parameters were sensitive to the available information around these points. It can be observed that the EKF applying either search mode yielded less erratic model parameter predictions than those of the NLLS technique. Again this was the result of Kalman filter tuning which prevented large fluctuations in parameter estimations.

Although the EKF was slightly faster than the least squares technique in predicting model parameters B and C, an offset between the two estimation methods was observed. It should be noted that the least squares technique approach applies all of the available measurements for parameter identification and attempts to identify parameter sets which result in an overall best fit. Due to the re-cursive nature of the EKF only the latest measurement was used and the previous data points were represented in the Kalman gain term, hence the Kalman filter was more likely to estimate model parameters which result in good local fits to the observed values. Hence during the least squares estimates no matter which search method was applied eventually when the identification technique had all of the available data points its predictions yielded similar values. During the EKF estimation the final parameter levels need not be similar due to the method in which the at-line information was applied. From figure 7.3, it can be seen that after only 1-2 measurements information on the inflection point and slope was obtained. This explains the relatively fast response of the EKF estimation technique, using the sequential search method.

The parameter estimation behaviour for the two different searching modes is best shown by comparing the NLLS technique applying the two search methods. It can be seen that when the controlled search mode was implemented estimation of model parameters stabilised sooner than when the sequential search method was adopted. Parameter estimate stability was reached after 5 for the controlled and 7-8 measurements for the sequential search method.

Figures 7.9 and **7.10** illustrate the step by step procedure adopted during the controlled search method for Runs 1 and 3 using the least squares model parameter estimation and subsequent flocculation profile prediction. These figures only represent the first 6 at-line cell debris measurements (-♦-) and their respective flocculation profile predictions (—). Figures 7.9 and 7.10 give a graphical representation of how the controlled search method (section 7.3) was applied for Runs 1 and 3. The continuous flocculation Run2 is not shown as its operating conditions were very similar to those of Run1. The step by step search procedure is illustrated by plots (a) to (f). The arrows on the plots indicate the target PEI volume fraction based on the up to date profile prediction.

Plot (a) shows the flocculation profile prediction adopting a single measurement of the spun feed homogenate and the initial parameter guesses. As described in section 7.3, first a relatively high flocculant dose was selected to obtain a good estimate of

parameter A. The next step in the controlled searching method was to focus on the apparent inflection point of the debris flocculation curve as indicated by the arrow in plot (b). Plot (c) illustrates the subsequent outcome, where for Run1 (figure 7.9) and Run 3 (figure 7.10) the third at-line measurement was just below and above the apparent inflection point. Plot (d) illustrates that focus was still on characterising the flocculation profile in terms of its inflection point. Plot (d) and (e) for Run1, illustrates that some stability has been reached in determining the inflection point (~ model parameter B). The controlled search technique then focused on the region close to the bottom of the flocculation profile. Plot (d) to (e) during Run3 (figure 7.10) illustrates that the profile inflection point was still not characterised fully, PEI levels close to the inflection point were still targeted. Plot (f) shows the flocculation profile predictions implementing the initial parameter guesses (···) as well as the up to data profile predictions given 6 at-line data points. A clear shift can be observed from the initial profile guesses to the actual profile predictions.

At-line graphical representation of the continuous flocculation runs 1 and 3 are illustrated in **figures 7.11** and **7.12** for the NLLS and EKF estimation techniques adopting the sequential and continuous searching modes. The curves were created by implementing the estimated model parameters and simulating the contaminants and product yield profiles. Figures 7.11 and 7.12 show the outcome of such graphical representation using 5 at-line data sets. The operator can utilise such real time graphical representation to visually determine the performance of the continuous flocculation process and act upon this information for open loop control. During the sequential searching technique (plot a and b), the at-line data were fed to the estimators with increasing PEI volume levels. For the controlled searching method (plot b and d) the sequence of the first five at-line data were fed according to the figures attached to the ADH data.

It can be observed that during the NLLS estimations (Plot A) adopting the sequential searching method the least squares technique was poor at predicting the overall flocculation profile, although the actual fit to the data points was good. The estimation of the maximum removal level (parameter A) was especially poor for the cell debris profile due to the lack of available information in this region. The predictions using the EKF and sequential searching technique are illustrated in plot c. Although the profile predictions are slightly better than those of the NLLS estimation method the actual fitting of the 3(2)-parameter model to the data points was not as good. During the

controlled search mode both estimation techniques performed better in predicting the correct flocculation profiles as model sensitive information was fed to the parameter identification methods allowing for rapid and effective estimations.

Figures 7.13 to **7.15** illustrate the behaviour of the at-line optimisation algorithm in predicting the optimum PEI dose during each of the continuous runs. The objective of the optimisation algorithm was to maximise the recovery of product (ADH) and removal of contaminants (RNA, protein and cell debris) for the lowest flocculant volume fraction. A similar optimisation procedure was applied in chapter 6. The setpoint optimum PEI volume fraction was defined as the final optimum flocculant dose determined by the NLLS estimation technique using all of the available data points.

Figure 7.13, illustrates the behaviour of the predicted optimum flocculant dose implementing the EKF and NLLS identification techniques applying both the sequential and controlled searching modes during Run1. As the predicted optima values were based on the correct estimation of the flocculation profiles the initial fluctuations seen in figure 7.13 were due to variations in the model parameter estimations (figure 7.6 to 7.8). It can be observed that both estimation techniques adopting the controlled searching method (-●- EKF -■- NLLS) yielded more stabile optimisation results close to the optimum PEI setpoint after 5-6 measurements. The NLLS approach adopting the sequential searching technique (-□-) exhibited poor predictions of PEI optimum values initially and only after measurement step 8 was it able to achieve within 90% accurate estimates. Due to the poor information available for the NLLS approach when adopting the sequential searching technique, poor parameter and hence optimum PEI dose estimates were achieved. It can be observed form figure 7.13, that although the EKF estimates implementing the sequential searching method (-o-) stabilised after a few measurements an off-set existed whereby the predictions were approximately 10% below the optimum PEI setpoint.

Figures 7.14 and 7.15 illustrate the estimated PEI optima values for the 4 scenarios discussed above for the continuous flocculation runs 2 and 3. By combining the estimation techniques with the controlled searching method more stable and faster prediction responses were seen. After 6-7 measurements the optimum PEI estimates were within the 90% accuracy range. By selecting the type of information fed to the estimation techniques more effective parameter estimations where obtainable, yielding better flocculation profile and optimum PEI level predictions. Although the EKF yielded

more stabile optimisation results than that of the NLLS approach a constant offset was observed when the Kalman filter was adopted.

7.5 Discussion

The effect of two data point scouting methods (sequential and controlled) in conjunction with the two estimation techniques (NLLS and EKF) was examined in terms of their ability to characterise the continuous flocculation process. The subsequent prediction of optimum flocculant volume fractions was investigated. As the optimisation problem was a function of the flocculation profiles, the behaviour of the optimisation estimations was directly dependent on the trend in model parameter prediction. It was seen that during the controlled search method both estimation techniques would predict optimum flocculant levels more efficiently both in terms of speed and stability. It was observed that although the EKF model parameter estimations stabilised earlier than those of the NLLS technique an offset was seen from the optimum PEI setpoint. Hence for 100% accurate model predictions of the optimum PEI dose the least squares approach should be implemented.

Which estimation technique is most suitable for a given process is a function of several factors. The correct tuning of the EKF, computational load, prior knowledge of process behaviour in terms of product and contaminant characteristics and fluctuations, batch to batch and within batch disturbances, are some of the factors which will influence the behaviour of these two estimation techniques. It was concluded in the previous chapter that with just a few data points during the batch flocculation process the EKF had superior performance. By implementing the correct tuning constants the Kalman filter would use the initial parameter guesses more effectively and introduced fewer fluctuations during parameter prediction. By introducing the controlled search method the advantages of the EKF were less prominent due to the possibility of examining process sensitive areas rapidly. By adopting the technique discussed in section 7.3, the model parameters would be estimated in a logical and efficient fashion. The controlled search technique did not only improve model parameter estimation using the NLLS technique but also made the EKF more efficient by reducing the offset between actual and predicted model parameters. It should be noted that the main disadvantage of the EKF was the need for correct tuning. As illustrated by table 7.1 two different sets of tuning constants were needed for the two search technique highlighting the time and effort required when applying the EKF.

7.6 Conclusion

At-line monitoring of both product and key contaminants was demonstrated allowing for an automatic and efficient method to obtain real time information of the continuous flocculation process without the need for time consuming manual off-line assays. The real time information was implemented to characterise the process rapidly and present the operator with a graphical representation of the flocculation behaviour. Two model parameter identification techniques in combination with two scouting methods were examined. It was demonstrated that by applying an intelligent method of searching, model sensitive information could be fed to the estimation technique facilitating model parameter predictions resulting in more efficient characterisation and optimisation of the continuous flocculation process.

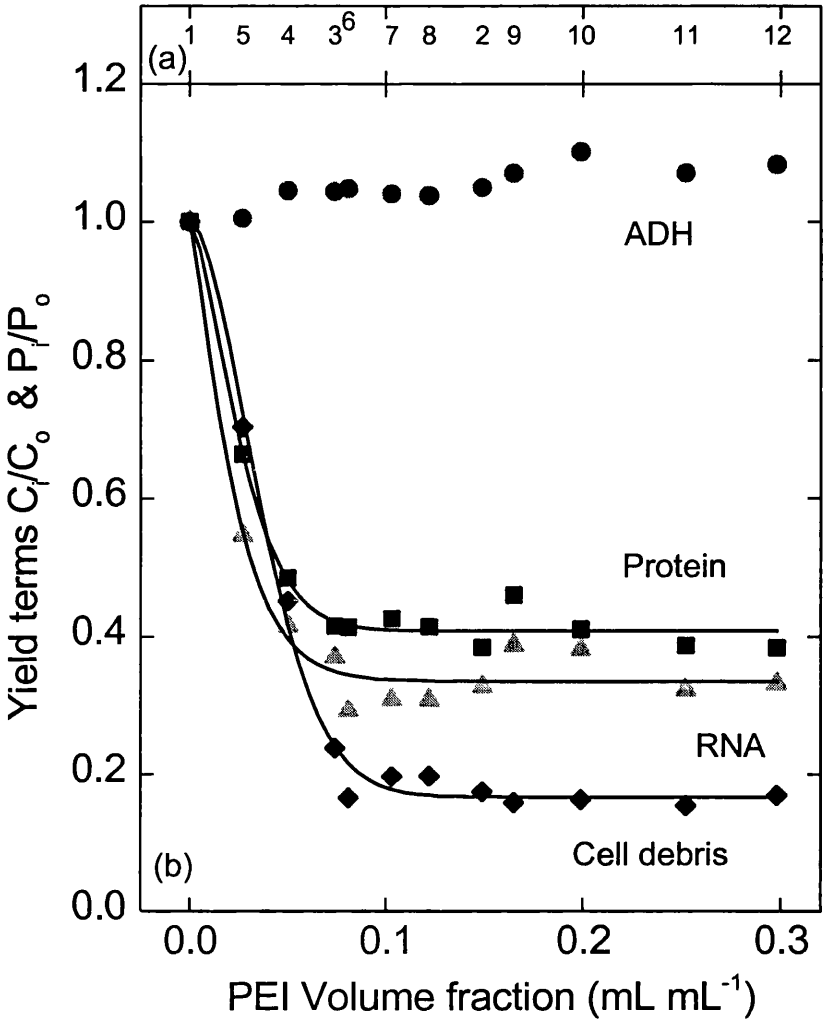


Figure 7.3. *Non-linear least squares fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-Parameter model for Run1.*
Plot(b), Baker's yeast homogenate (125 g (wet weight) L^{-1}) was mixed continuously with PEI flocculant (0.01 g mL^{-1} stock solution) at different ratios and a total flowrate of 22.8 mL min^{-1}. The feed and PEI conditions are listed in table 7.2. Samples (5.6 mL) were automatically taken every 130-140 seconds for rapid solid-liquid separation and measurements. The NLLS resulted in the below model parameter estimations.
Plot(a), Indicates the sequence of PEI volume fraction levels tested during the continuous searching mode.

Parameter	A	B	C
Debris	0.83	0.045	1.77
RNA	0.66	0.025	1.25
Protein	0.59	0.031	1.5

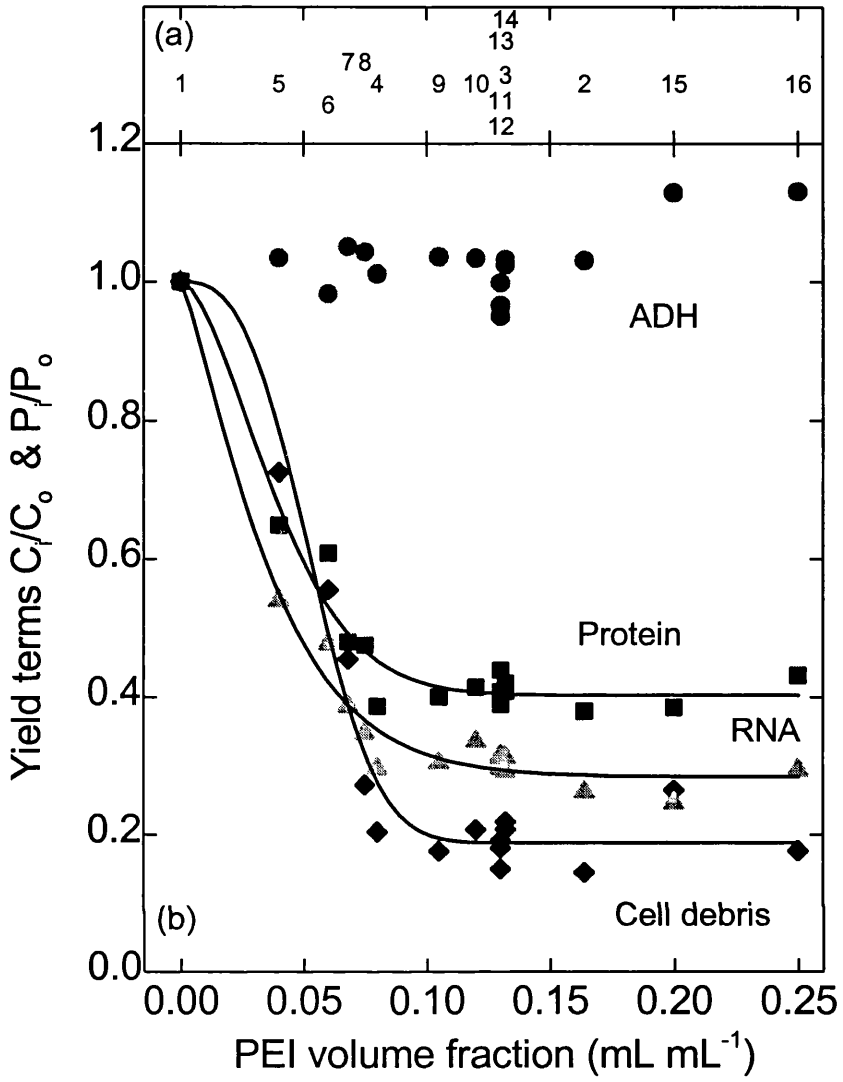


Figure 7.4. *Non-linear least squares fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-Parameter model for Run2.*
Plot(b), Baker's yeast homogenate (125 g (wet weight) L^{-1}) was mixed continuously with PEI flocculant (0.01 g mL^{-1} stock solution) at different ratios and a total flowrate of 22.8 mL min^{-1}. The feed and PEI conditions are listed in table 7.2. Samples (5.6 mL) were automatically taken every 130-140 seconds for rapid solid-liquid separation and measurements. The NLLS resulted in the below model parameter estimations.
Plot(a), Indicates the sequence of PEI volume fraction levels tested during the continuous searching mode.

Parameter	A	B	C
Debris	0.81	0.06	2.85
RNA	0.72	0.04	1.24
Protein	0.60	0.05	1.66

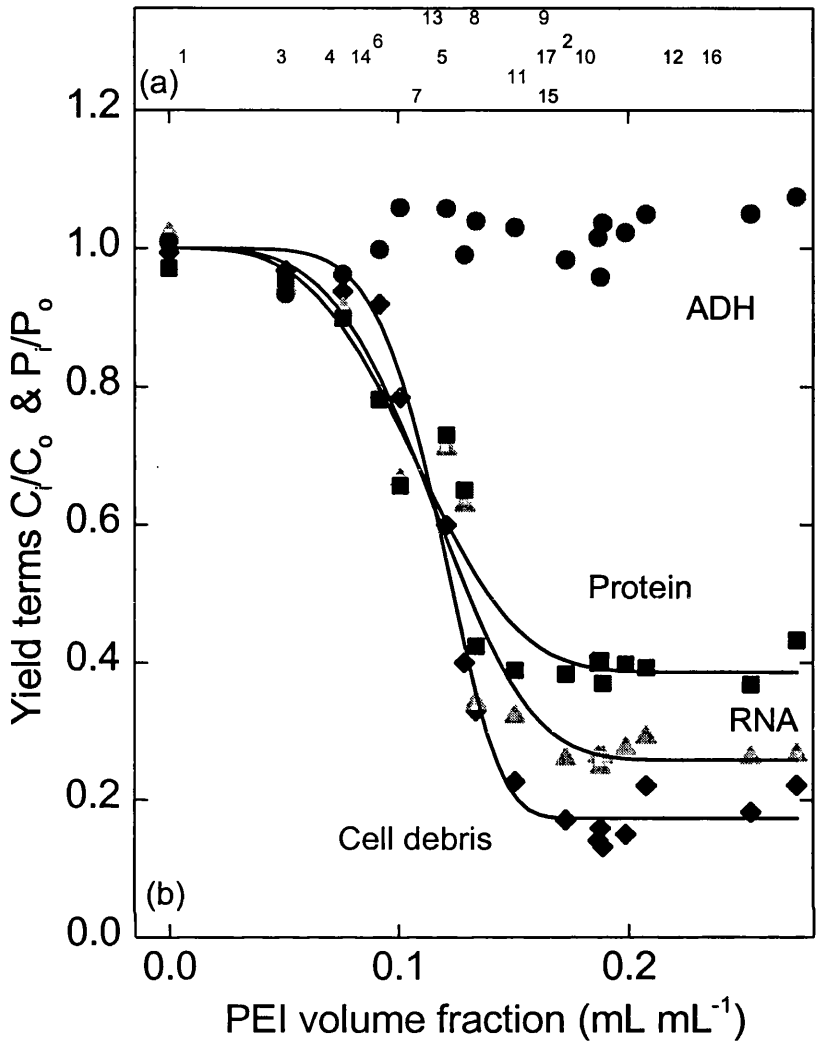


Figure 7.5. *Non-linear least squares fit of product and contaminant yield values in terms of PEI volume fraction implementing the 3(2)-Parameter model for Run3.*
Plot(b), Baker's yeast homogenate (125 g (wet weight) L^{-1}) was mixed continuously with PEI flocculant (0.01 g mL^{-1} stock solution) at different ratios and a total flowrate of 22.8 mL min^{-1}. The feed and PEI conditions are listed in table 7.2. Samples (5.6 mL) were automatically taken every 130-140 seconds for rapid solid-liquid separation and measurements. The NLLS resulted in the below model parameter estimations.
Plot(a), Indicates the sequence of PEI volume fraction levels tested during the continuous searching mode.

Parameter	A	B	C
Debris	0.83	0.125	6.3
RNA	0.75	0.131	3.6
Protein	0.63	0.121	2.5

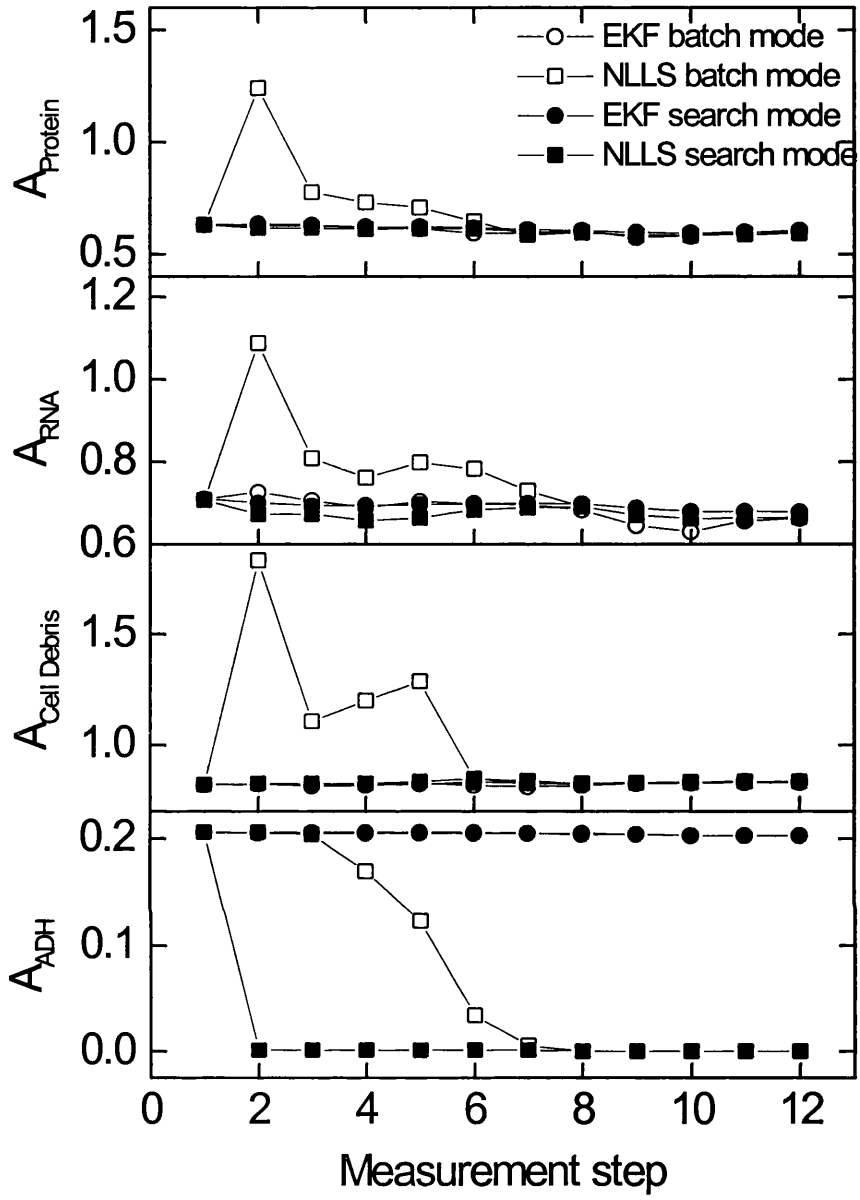


***Figure 7.6**. Estimation of model parameter A for protein, RNA, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt techniques adopting the continuous and batch searching techniques for Run1. The sequence of data points during the continuous searching method is shown in figure 7.3, plot (a). For the batch search model the sequence followed an ascending PEI volume fraction order. The model parameter initial guesses and EKF tuning constants are listed in table 7.1.*

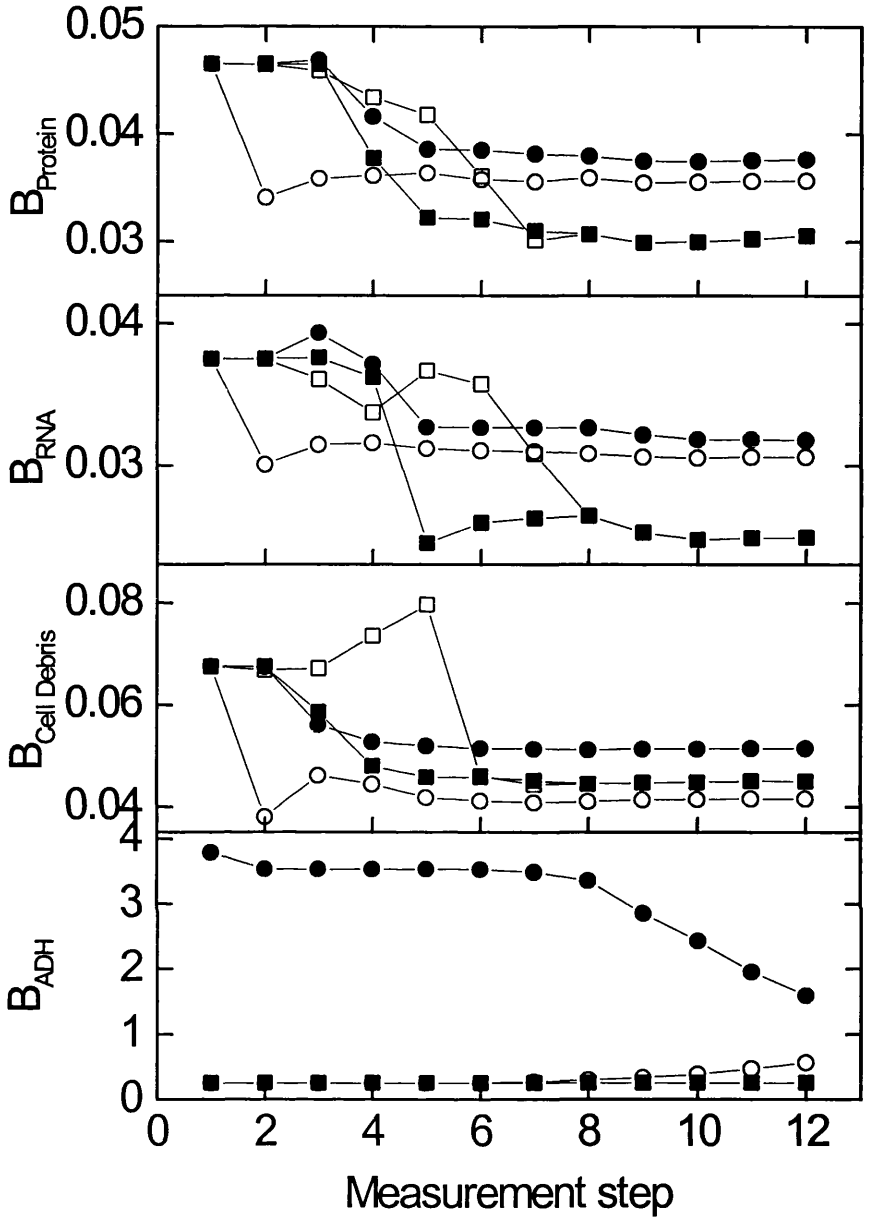


Figure 7.7. *Estimation of model parameter B for protein, RNA, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt techniques adopting the continuous and batch searching techniques for Run1. The sequence of data points during the continuous searching method is shown in figure 7.3, plot (a). For the batch search model the sequence followed an ascending PEI volume fraction order. The model parameter initial guesses and EKF tuning constants are listed in table 7.1. The symbols (-●-) and (-■-) refer to the EKF and NLLS estimation techniques applying the continuous search method, whilst (-o-) and (-□-) refer to the EKF and NLLS using the batch search mode.*

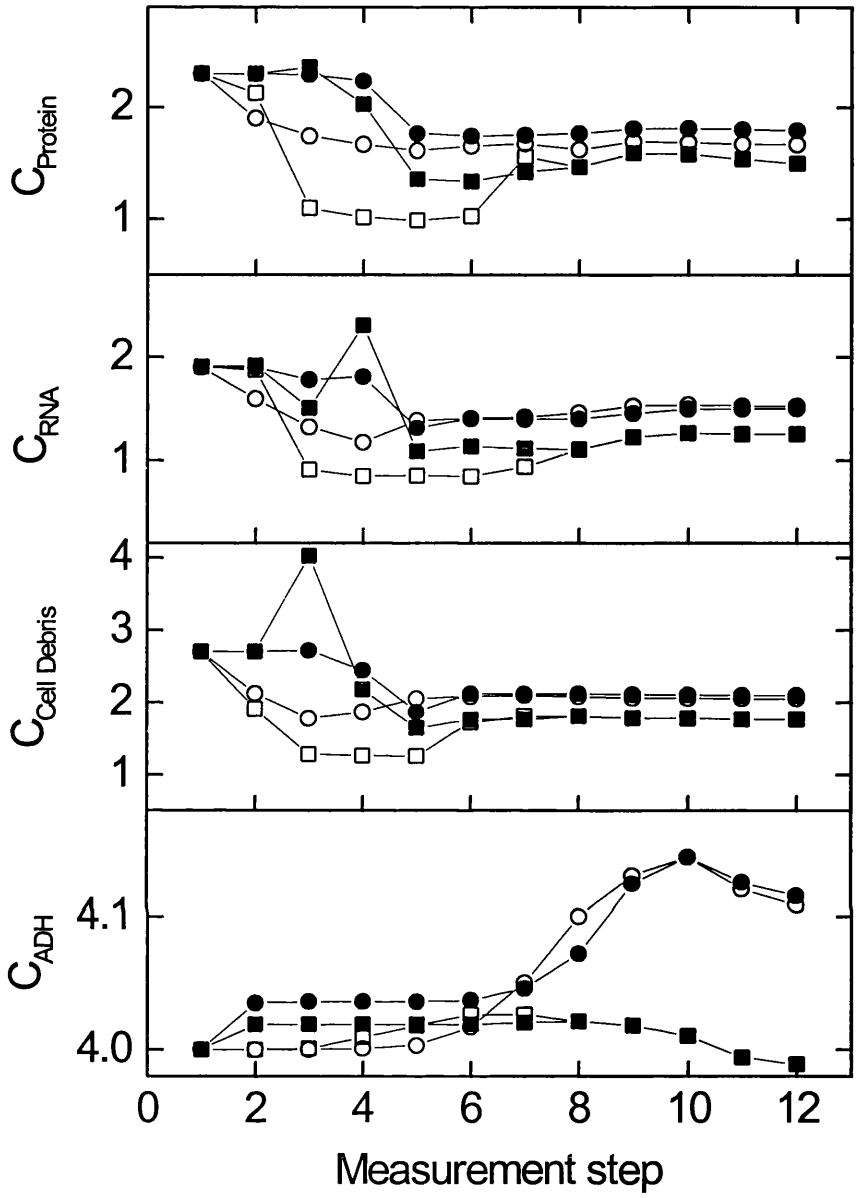


Figure 7.8. *Estimation of model parameter C for protein, RNA, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt techniques adopting the continuous and batch searching techniques for Run1. The sequence of data points during the continuous searching method is shown in figure 7.3, plot (a). For the batch search model the sequence followed an ascending PEI volume fraction order. The model parameter initial guesses and EKF tuning constants are listed in table 7.1. The symbols (-●-) and (-■-) refer to the EKF and NLLS estimation techniques applying the continuous search method, whilst (-○-) and (-□-) refer to the EKF and NLLS using the batch search mode.*

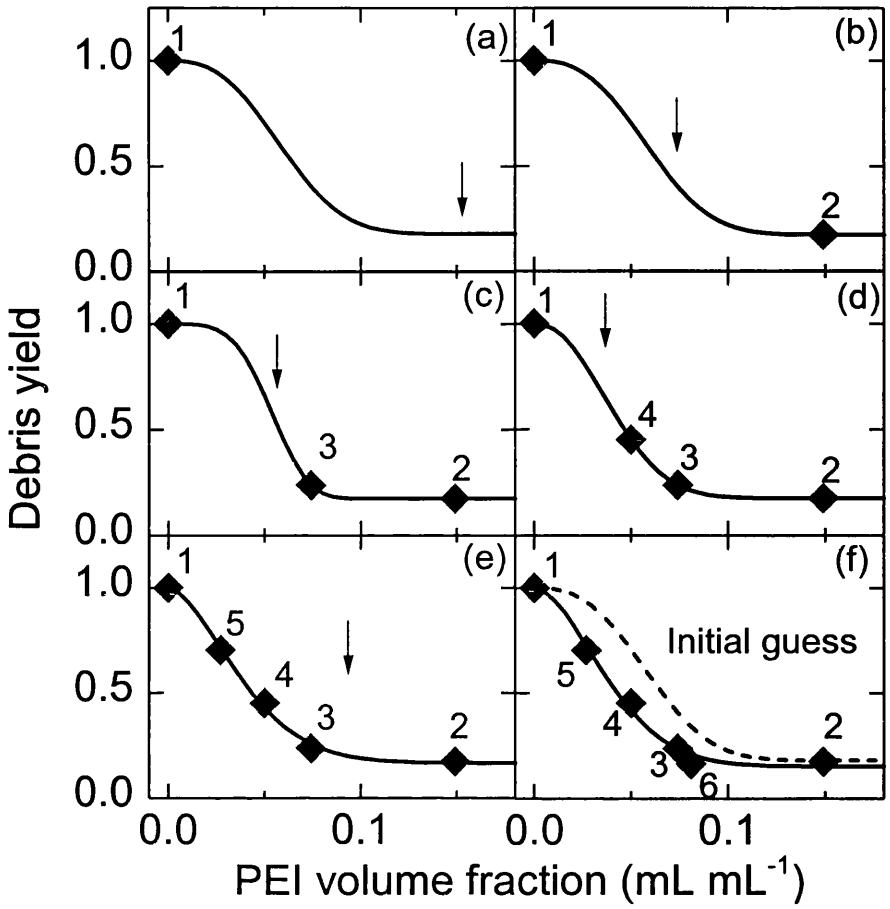


Figure 7.9. *Debris flocculation profile prediction implementing the Levenberg Marquardt estimation technique for the continuous flocculation Run1. The profile predictions (—) were carried out using 1 to 6 data points (-◆-) illustrated in plots (a) to (f). The sequence of at-line debris data are indicated by the numbers by the points. The arrows indicate the next area of interest based on the up to date predicted flocculation profile and the decision making procedure illustrated in figure 7.2. The dotted curve (···) in plot (f) refers to the flocculation profile based on the initial parameter guesses listed in table 7.1. The initial profile prediction (plot (a)) was derived from the initial parameter guesses.*

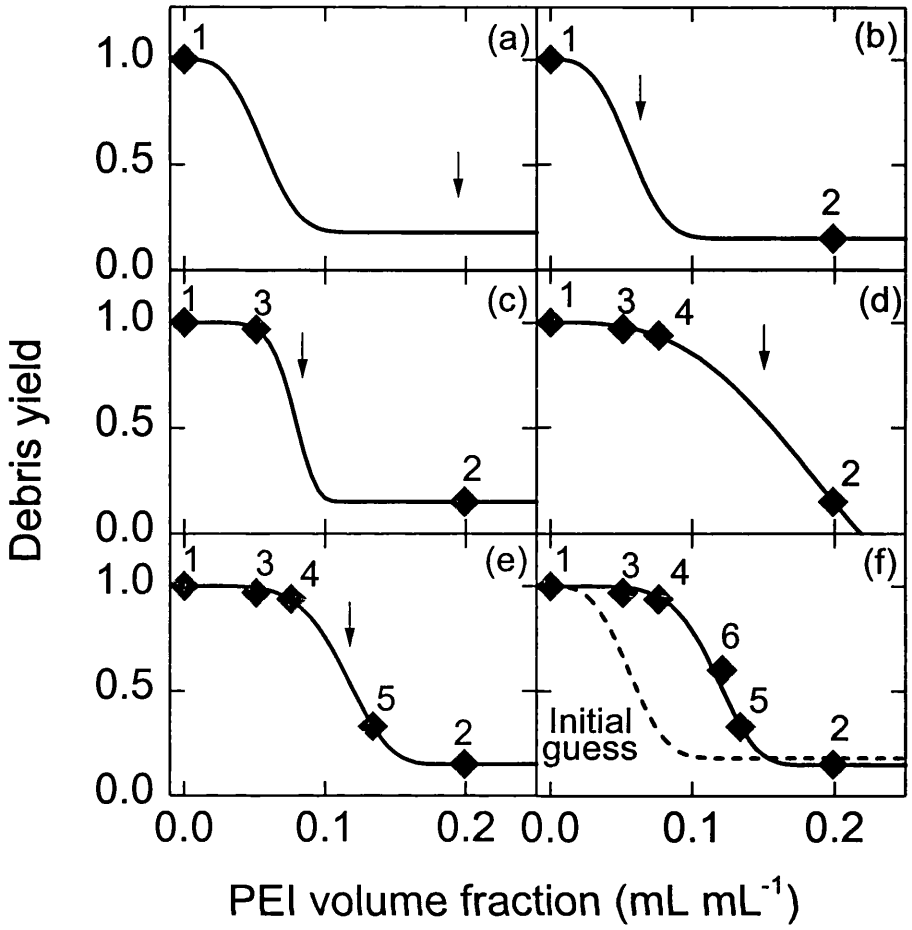


Figure 7.10. *Debris flocculation profile prediction implementing the Levenberg Marquardt estimation technique for the continuous flocculation Run3. The profile predictions (—) were carried out using 1 to 6 data points (-◆-) illustrated in plots (a) to (f). The sequence of at-line debris data are indicated by the numbers by the points. The arrows indicate the next area of interest based on the up to date predicted flocculation profile and the decision making procedure illustrated in figure 7.2. The dotted curve (···) in plot (f) refers to the flocculation profile based on the initial parameter guesses listed in table 7.1. The initial profile prediction (plot (a)) was derived from the initial parameter guesses.*

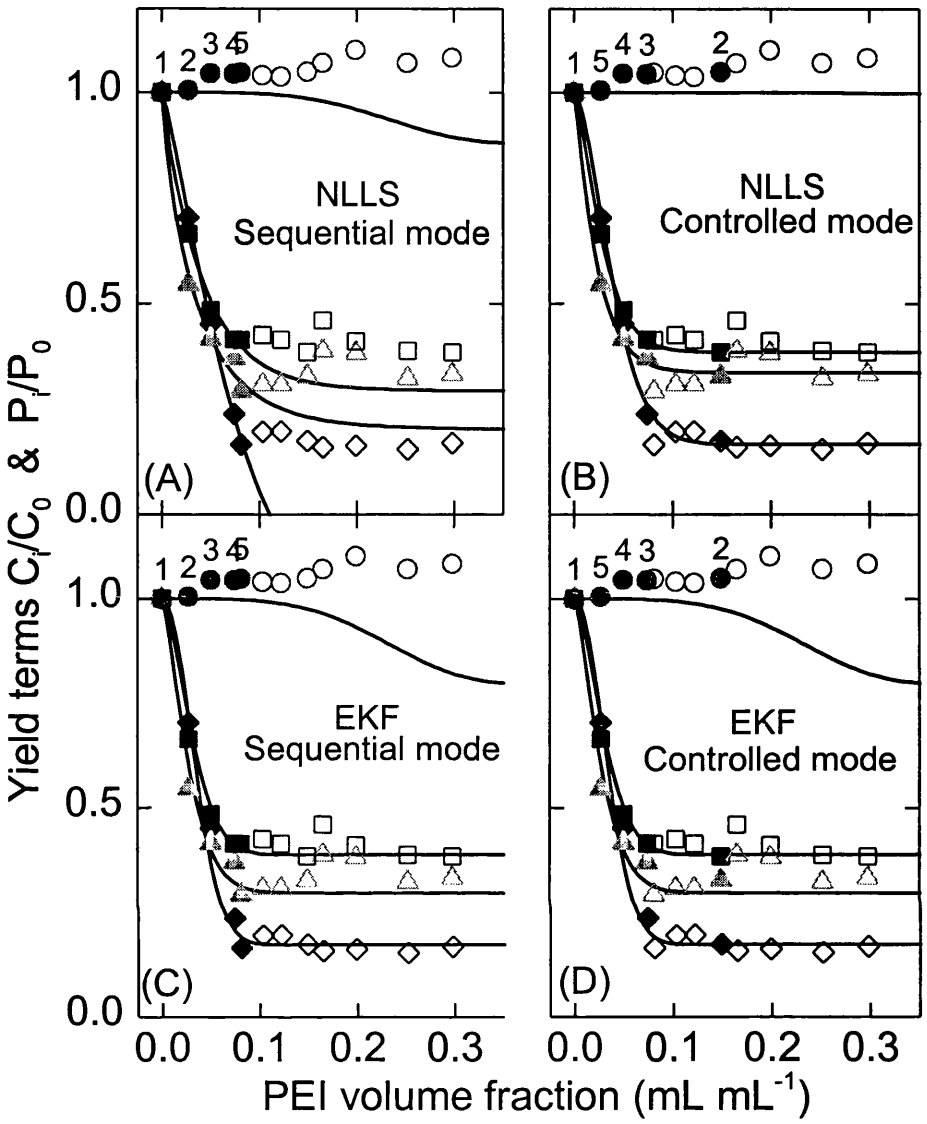


Figure 7.11. *Flocculation profile prediction of the continuous flocculation Run1, implementing 5 at-line data points. The symbols (-●-), (-■-), (-◆-) and (-▲-) refer to ADH, protein, debris and RNA data made available to the estimation techniques. The corresponding open symbols refer subsequently acquired data points. The numbers attached to the ADH data represents the sequence of at-line data fed to the estimation techniques.*

Plot (a). The Levenberg-Marquardt (LM) non-linear least squares (NLLS) technique adopting the batch searching method.

Plot (b). The NLLS technique adopting the continuous searching method.

Plot (c). The extended Kalman filter (EKF) technique adopting the batch searching method.

Plot (d). The EKF technique adopting the continuous search method.

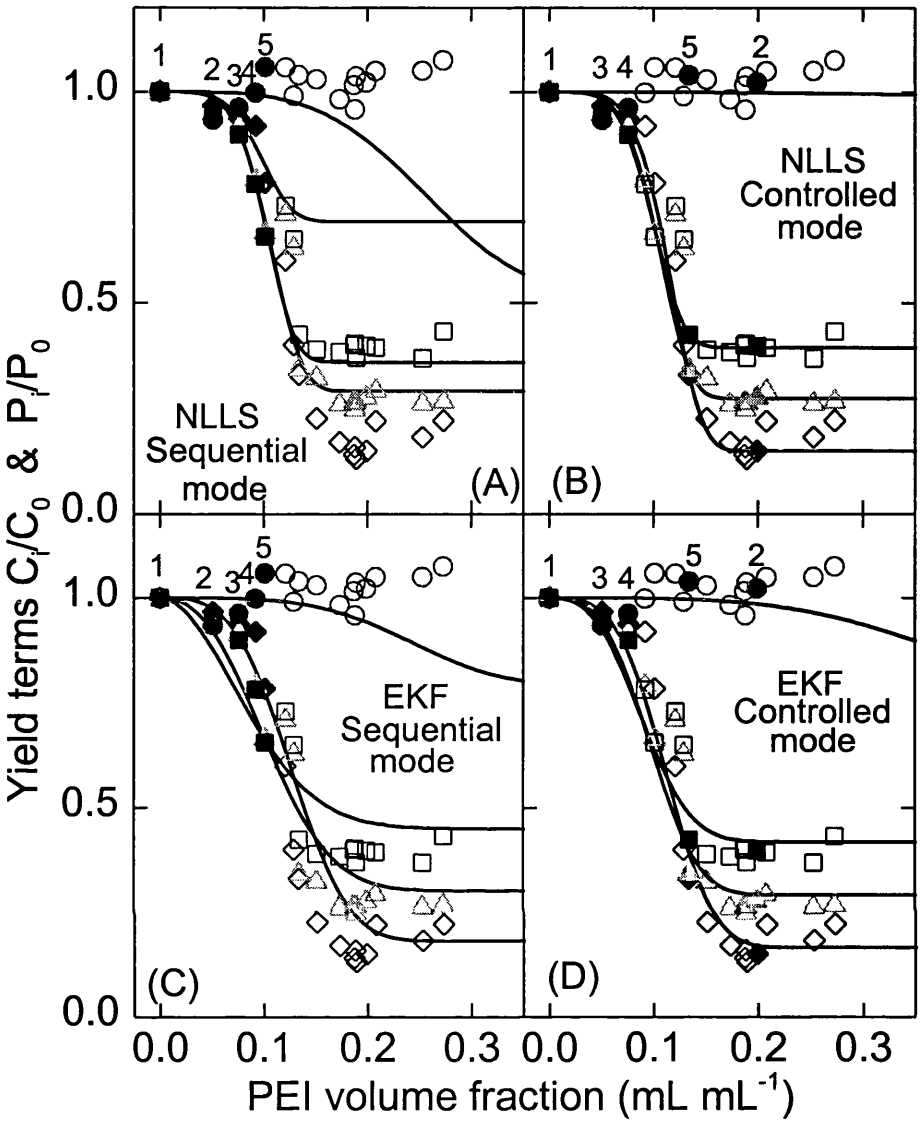


Figure 7.12. *Flocculation profile prediction of the continuous flocculation Run3, implementing 5 at-line data points. The symbols (-●-), (-■-), (-◆-) and (-▲-) refer to ADH, protein, debris and RNA data available to the estimation techniques. The corresponding open symbols refer subsequently acquired data points. The numbers attached to the ADH data represents the sequence of at-line data fed to the estimation techniques.*

Plot (a). The Levenberg-Marquardt (LM) non-linear least squares (NLLS) technique adopting the batch searching method.

Plot (b). The NLLS technique adopting the continuous searching method.

Plot (c). The extended Kalman filter (EKF) technique adopting the batch searching method.

Plot (d). The EKF technique adopting the continuous search method.

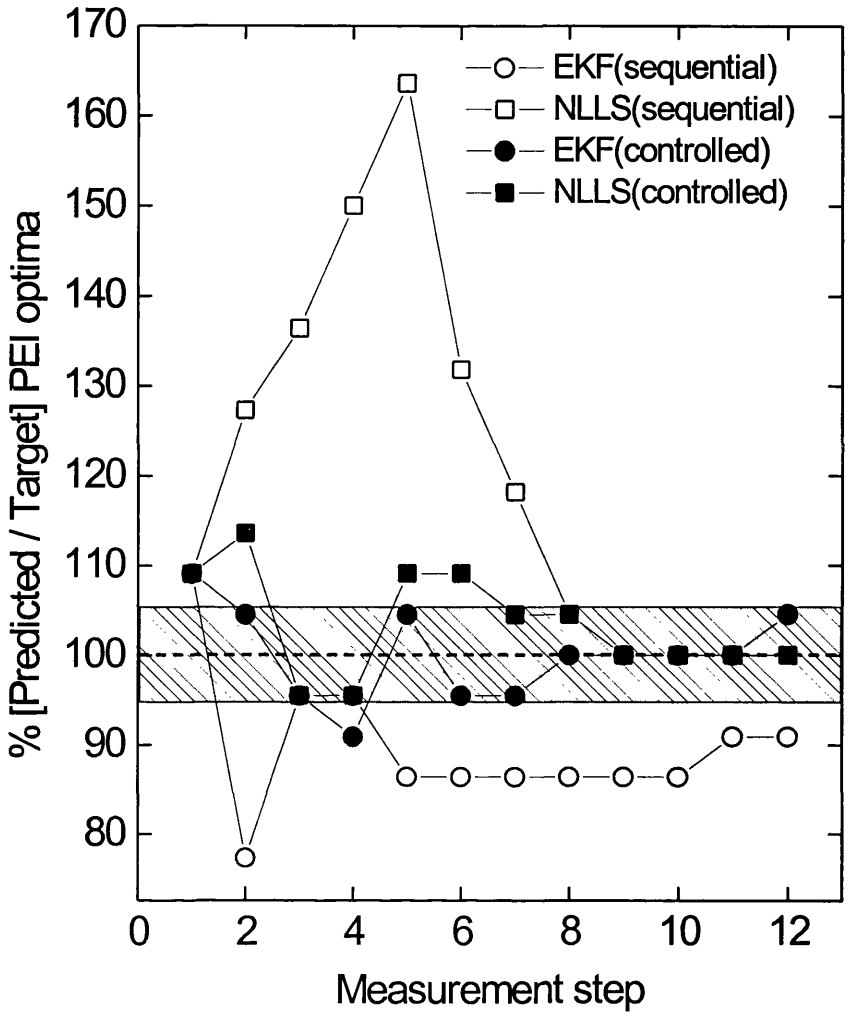


Figure 7.13. *Real time optimisation of the continuous flocculation Run1, implementing the EKF and NLLS estimation techniques adopting the batch and continuous search methods. The predicted optimum PEI values were plotted as a percentage of the target optimum PEI setpoint, which was defined as the optimum PEI level determined by the NLLS estimation technique with all available measurements. The data sequence for the continuous search method is shown in figure 7.3, plot (a). The dashed box indicates a 90% (± 5%) accuracy range. The optimum PEI setpoint was 0.11 mL mL^{-1}. The weighting factors imposed on the product, contaminants and PEI dose were w_{ADH} = 0.5, w_{RNA} = 0.1, $w_{protein}$ = 0.05, w_{Debris} = 0.25 and w_{PEI} = 0.1. The initial parameter guesses for the model parameters and EKF tuning constants are listed in table 7.1.*

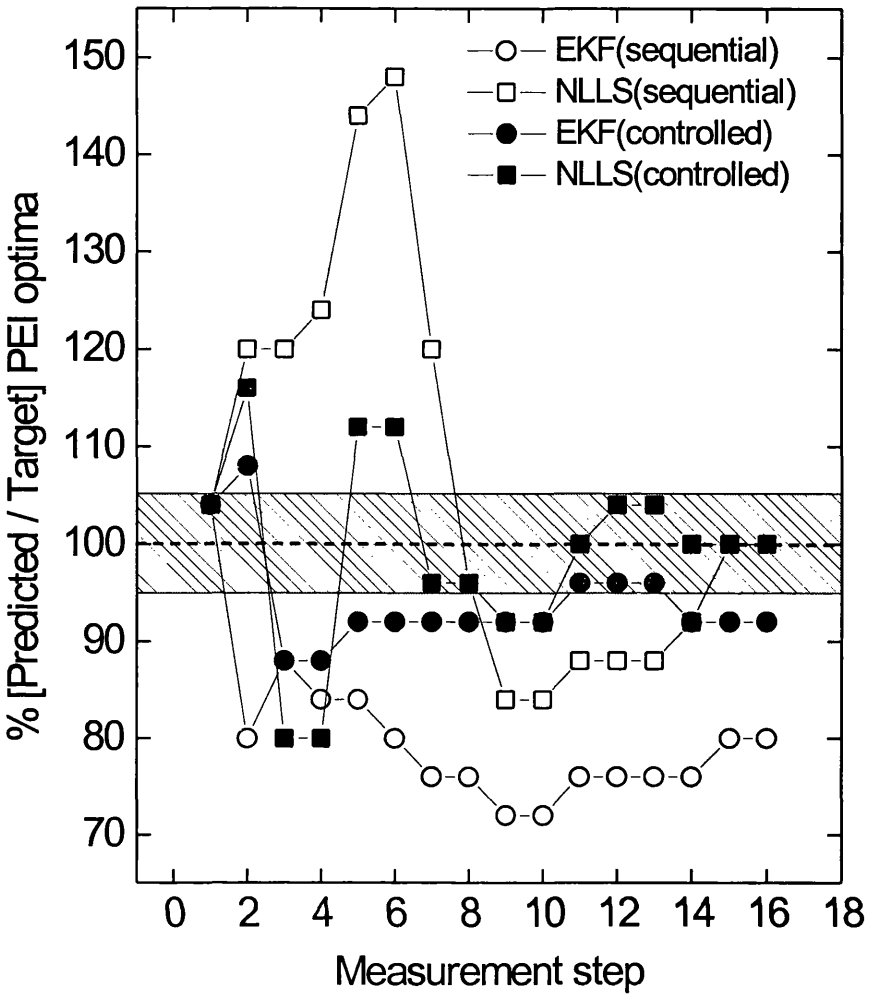


***Figure 7.14**. Real time optimisation of the continuous flocculation Run2, implementing the EKF and NLLS estimation techniques adopting the batch and continuous search methods. The predicted optimum PEI values were plotted as a percentage of the target optimum PEI setpoint, which was defined as the optimum PEI level determined by the NLLS estimation technique with all available measurements. The data sequence for the continuous search method is shown in figure 7.4, plot (a). The dashed box indicates a 90% (± 5%) accuracy range. The optimum PEI setpoint was 0.125 mL mL^{-1}. The weighting factors imposed on the product, contaminants and PEI dose were w_{ADH} = 0.5, w_{RNA} = 0.1, $w_{protein}$ = 0.05, w_{Debris} = 0.25 and w_{PEI} = 0.1. The initial parameter guesses for the model parameters and EKF tuning constants are listed in table 7.1.*

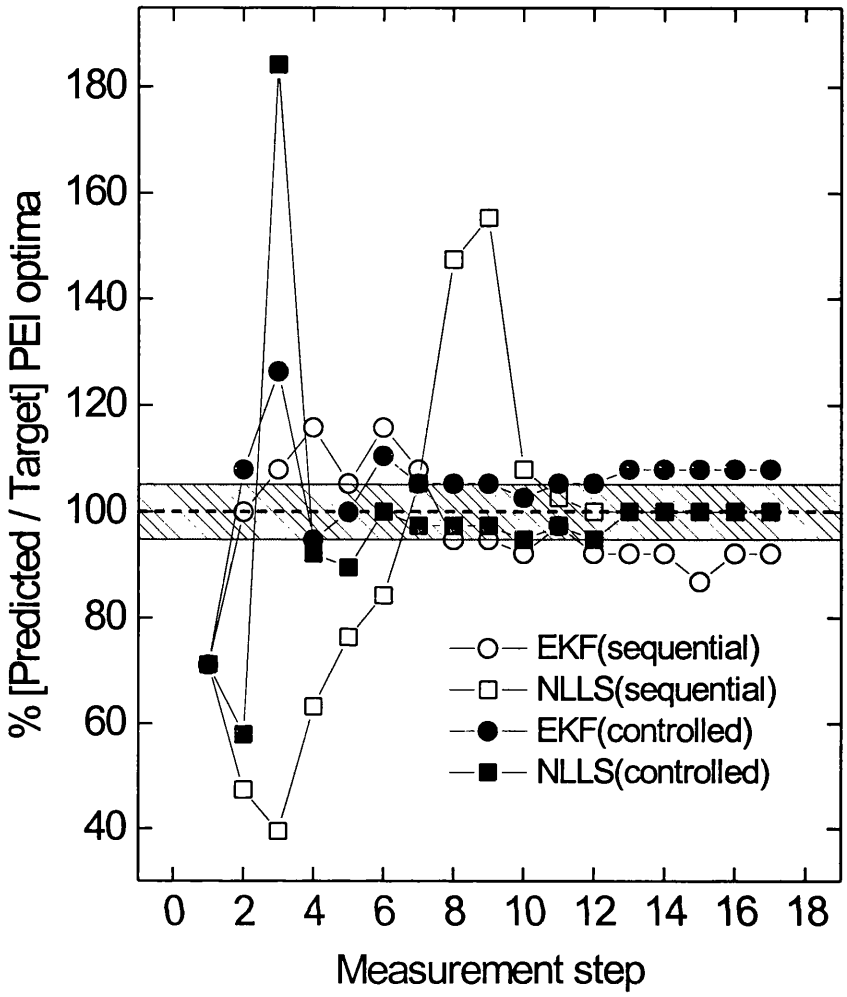


***Figure 7.15**. Real time optimisation of the continuous flocculation Run3, implementing the EKF and NLLS estimation techniques adopting the batch and continuous search methods. The predicted optimum PEI values were plotted as a percentage of the target optimum PEI setpoint, which was defined as the optimum PEI level determined by the NLLS estimation technique with all available measurements. The data sequence for the continuous search method is shown in figure 7.5, plot (a). The dashed box indicates a 90% (± 5%) accuracy range. The optimum PEI setpoint was 0.19 mL mL^{-1}. The weighting factors imposed on the product, contaminants and PEI dose were w_{ADH} = 0.5, w_{RNA} = 0.1, $w_{protein}$ = 0.05, w_{Debris} = 0.25 and w_{PEI} = 0.1. The initial parameter guesses for the model parameters and EKF tuning constants are listed in table 7.1.*

Chapter 8

Rapid monitoring and control of packed and expanded bed chromatography

8.1 Summary

Real time monitoring of both product and key contaminants was applied for the characterisation, process analysis and control of the packed and expanded bed chromatographic operations. Rapid at-line measurements enabled a higher resolution of data points to be acquired resulting in improved chromatographic characterisation compared with the use of traditional time consuming and laborious off line assays. Further to this, real time process analysis was achieved accelerating the development cycle by allowing the design engineer to prepare subsequent experiments rapidly in an intelligent and informed manner. Control of the chromatographic operations was investigated. For the packed bed system at-line product measurements were applied in a simple feedback control arrangement to regulate the load cycle. For expanded bed operation a predictive model based control configuration was implemented in order to compensate for the relatively large dead volume associated with this unit operation. In both cases good control of the load cycle was demonstrated despite the variation in feed material.

8.2 Introduction

This chapter will demonstrate the ability to rapidly monitor the hydrophobic interaction packed bed and expanded bed chromatographic operations. At-line measurements of both product (alcohol dehydrogenase) and key contaminants (RNA, cell debris, protein) will be presented.

The at-line monitoring setup applied for the flocculation process was implemented for both packed and expanded bed monitoring according to the procedure described in chapter 3. The calibration model for contaminant predictions derived in chapter 4 may be applied for the packed bed system due to that its contaminant levels are within the calibration range. For the expanded system although the cell debris concentration is higher than that calibrated for, it will be assumed that the calibration model is still applicable.

Rapid monitoring enables the design engineer to demand a higher resolution of process data compared to when utilising laborious time consuming off-line measurements. Further to this at-line monitoring allows for rapid process analysis, enabling the process performance to be studied in real-time. This allows the operator to rapidly plan and set-up the subsequent experiment in an informed manner accelerating the process development cycle. Real time monitoring also enables the possibility of process control, automation and for improved process reproducibility. Process control may be implemented to accelerate process development, by enabling specific operating conditions (breakthrough levels) to be met and hence allowing the process performance to be studied around sensitive operating areas.

In chapter 1 (section 1.6.4) it was discussed that several factors influence the behaviour of a hydrophobic interaction chromatographic process, such as the type of packing matrix, feed conditions, pH, type and concentration of salt, temperature, flowrate, sample load, etc. Hydrophobic interaction chromatography may be divided into three operating cycles; the loading of feed on to the column, the washing of contaminants off the column and the elution of the target product. Given that most of the input variables are fixed during the design stages the only manipulative variable applicable for control was the loading time, the wash time and the start and end of elution peak collection, all of which may be regulated implementing ON/OFF valve control.

Rapid monitoring, analysis and control of the three chromatographic cycles will be investigated. For the packed bed systems rapid measurement of protein and ADH were acquired, whilst during expanded bed operations RNA and cell debris were also monitored. This chapter will focus on the effective characterisation and control of the loading stage as this step has great impact on the two subsequent cycles (wash and elution). The performance of the overall chromatographic operation in terms of overall process yield, purity, productivity (amount of product per volume of matrix and unit time) and process economy are partly dictated by the correct loading of feed. If the loading cycle is too long product losses during breakthrough and the wash cycle will be seen. However, too short a load cycle would result in poor utilisation of the column capacity. A trade off exists between product yield and column capacity, which is most important for a given system is a function of operating costs (matrix, buffers, CIP) and productivity. Additional factors such as regulatory issues also play a part in determining the optimum running conditions.

Due to the importance of correct column loading, monitoring and control of the level of product breakthrough to a specific setpoint was investigated. The breakthrough setpoints chosen were not due to their ability to optimise the chromatographic process but to illustrate the effectiveness of the control configuration. During development work focus is on specifying input conditions to purify the product in interest to the highest yield and the required purity as quickly, cheaply and easily as possible. Therefore the development step will specify the majority of input variables. The task of process control and automation is to reproducibly operate the chromatographic system to meet these specified operating conditions given disturbances in the feed material and column performance.

Three packed bed operations, two bench and one pilot scale as well two bench scale expanded bed processes will be studied. A schematic representation of these processes is illustrated in section 3.7 and 3.8. A feedback control configuration will be investigated for breakthrough profile control both for packed and expanded bed operations. During the control of the expanded process a predictive technique was examined to compensate for the relatively large dead volume associated with its expanded state.

8.3 Rapid characterisation and control

This section will discuss the feedback control configuration adopted to control both the packed and expanded bed load cycle to a specific operator chosen breakthrough setpoint.

Figure 8.1 illustrates an overview of the two control configurations applied. Real time data was acquired through the rapid monitoring set-up discussed in chapter 3, sections 3.7 and 3.8.. For the packed bed systems, indicated by the dashed box, the raw data on product levels would be implemented as the control variable. If the product level had reached the setpoint breakthrough level the load cycle would be terminated and the wash cycle initiated. This may be implemented by the operator for open loop control or by a supervisory computer in a closed loop configuration.

For the expanded bed (EXB) operation a prediction based control configuration was adopted due to the relatively large dead volume associated with this operation. Due to this large dead volume, a relatively long lag time (t_d) is associated with measurements of the exit stream. Hence any real time data acquired at time i should be associated

with an input condition at time (i - t_d). If a control system similar to that applied for the packed bed operation was used for the EXB column any change in a manipulative variable would be registered in the exit steam minutes later.

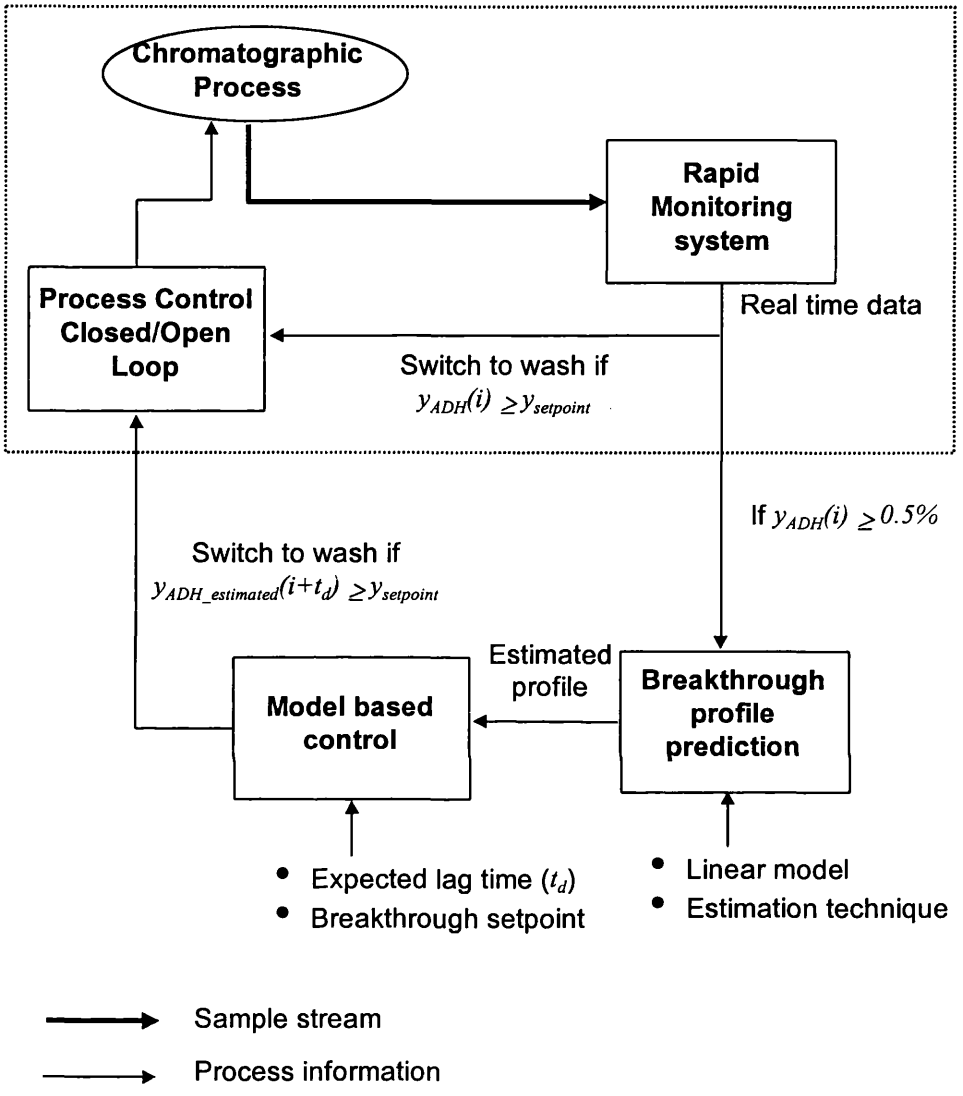


Figure 8.1. *Overview of breakthrough control configuration for the packed and expanded bed chromatographic operations.*

The dashed box illustrates the control configuration applied during the packed bed loading cycle. The product yield y_{ADH} = ADH/ADH_{feed}, is defined as the concentration of ADH at measurement step i as a fraction of ADH in the feed stream. The load cycle is terminated if the product yield is greater or equal to the product yield breakthrough setpoint ($y_{setpoint}$). The expanded bed control configuration (remainder of the above diagram) was based on a linear breakthrough profile prediction technique. The load cycle was terminated when the estimated breakthrough value ($y_{ADH\_estimated}$) based on the expected lag time (t_d) was greater or equal to the breakthrough setpoint.

By characterising the breakthrough profile mathematically forward predictions can inform the operator on what is actually taking place within the column. Hence the control action can be based on the actual predicted rather than the apparent column performance. As illustrated by figure 8.1, breakthrough profile characterisation would be initiated as soon as a significant level of product was registered in the eluate stream. For each newly acquired data point a prediction of the breakthrough profile would be performed. Knowing the expected lag time (function of the flowrate and column dead volume) and given the breakthrough setpoint chosen by the operator the model based control algorithm could estimate the appropriate time to terminate the load cycle.

The real time process analysis performed in this work are represented by the equations below (8.1 ,8.2 and 8.3), which define the yield and purity terms in the load, wash and elution cycles.

$$Yield_{load\ \&\ wash} = \frac{loaded\ on\ to\ column\ -\ lost\ in\ breakthrough\ and\ wash}{loaded\ on\ to\ column} \qquad (8.1)$$

$$Yield_{elution} = \frac{eluted\ off\ the\ column}{loaded\ on\ to\ column\ -\ lost\ in\ breakthrough\ and\ wash} \qquad (8.2)$$

$$Purity_{elution} = \frac{sum\ of\left(\frac{ADH_i}{ADH_0}\right) eluted\ off\ the\ column}{sum\ of\left(\frac{protein_i}{protein_0}\right) eluted\ off\ the\ column} \qquad (8.3)$$

8.4 Results

The results of both packed bed and expanded bed HIC chromatographic operations will be presented in terms of off-line and on-line data of product and key contaminants. Rapid monitoring was achieved implementing the at-line sampling and measurement apparatus described in chapter 3. The first part of this section will show packed bed chromatography data, both for bench (0.05 m diameter column) and pilot scale (0.2 m diameter column) operations. The second part of the results section will present data from two bench scale (0.05 m diameter column) expanded bed runs. A comparison between off-line and at-line data will be given for both chromatographic operations and the potential use of rapid measurements will be shown in terms of process control and at-line process analysis.

Run	ADH ($U\ mL^{-1}$)	Protein ($mg\ mL^{-1}$)	Specific ADH activity ($U\ (ADH)\ mg^{-1}$ (protein))	Dilution level (buffer : feed)
PB Run1	185	11	16.8	none
PB Run2	120	8	15	1.6 : 1
PB pilot	173	11.2	15.3	none
EXB Run1	145	10.8	13.4	NA
EXB Run2	92	10.5	8.8	NA

Table 8.1. Feed conditions for the chromatographic processes. PB and EXB refer to packed bed and expanded bed columns. The dilution level refers to the level of dilution required to bring the column feed total protein concentration to approximately 10 mg mL^{-1}. A batch flocculation operation was carried out prior to the packed bed system, hence the dilution level will give an indication if any excess PEI has been diluted.

8.4.1 Rapid monitoring of packed bed chromatography

The feed material for the packed and expanded bed processes were the outcome of an upstream PEI flocculation and subsequent centrifugal operation as described in section 3.7. The feed conditions are listed in **table 8.1** in terms of ADH, total protein, specific ADH activity and dilution levels prior to loading on to the columns.

Figures 8.2 to **8.4** illustrate at-line (plot a) and off-line (plot b) data of the bench scale packed bed operation Run1 in terms of ADH and total protein. Figure 8.2 shows the chromatograms for these two components indicating a relatively good match between the off-line and at-line measurements. The faster at-line measurements allowed for a higher resolution of data points for the characterisation of the ADH profile in the load, wash and particularly elution cycles. This is more evident in figures 8.3 and 8.4 where detailed plots of the different chromatographic operating stages are presented. During the load and wash cycle the rapid monitoring apparatus was able to perform 122 ADH assays whilst only 65 measurements were obtainable through the off-line procedure, i.e. approximately half the resolution. For fast elution cycles this is an important feature as efficient characterisation of the elution profile is crucial for effective collection of eluate fractions. It should be noted that an additional operator would be

required, if off-line assays and the operation of the pack bed were to be performed effectively.

Slight deviation between the off-line and at-line ADH data can be observed in figure 8.3, which can be attributed to several factors. Most importantly, the off-line data was taken of pooled fractions of column eluate, whilst at-line measurements were taken of approximately 0.5 mL of eluate sample at a specific time point. Dilution errors both in the stopped flow analyser (SFA) and off-line may also have contributed to the above deviation, especially during the elution cycle where dilution levels for ADH measurements were approximately 1:24400 off-line and 1:4913 at-line. The use of a 1 mm pathlength flowcell instead of a standard 10 mm pathlength flowcell applied during off-line assays allowed for the lower at-line dilution level.

As described in section 3.7, the rapid protein measurements were acquired post packed bed operation, however the stopped flow analyser and rapid spectral analysis techniques (chapter 4) were applied. This explains the lower resolution of protein data points compared to the at-line ADH data, as a manual link was required to perform the former analysis. Despite the lower resolution of protein information it should be noted that the off-line and rapid protein measurements are very similar in terms of profile trends and absolute values.

Figure 8.5 shows the breakthrough profiles of ADH for three packed bed operations, namely two bench scale (Run1 —O— and Run2 —□—) and a pilot scale (—▽—). As the superficial liquid velocities for the packed bed runs were similar (section 3.7) the breakthrough curves have been plotted in terms of process time to allow for their comparison. Note the initial feed conditions are listed in table 8.1.

A shift in the breakthrough profiles for the three operations was observed. This could be due to the difference in the concentration of ADH and protein in the feed streams as contaminant protein competes in binding onto the HIC matrix. Other factors such as the efficiency of the column between runs and the level of dilution of the supernatant stream prior to loading are also important factors to consider. The level of dilution of the supernatant will alter the level of PEI remaining in solution after the flocculation and centrifugal separation operations. No study has been conducted investigating the influence of PEI (cationic polyelectrolyte) on the performance of HIC matrix. The concentration of other contaminants such as RNA and cell debris would also affect the performance of the packed bed in terms of column capacity. It can be concluded that

several factors affect the behaviour of a packed bed column in terms of loading and breakthrough. Many of these can be predicted qualitatively, however the correct quantitative estimation of how the packed bed system would behave in terms of a setpoint breakthrough level is not yet possible.

Figure 8.5 illustrates that the at-line data enabled the operator to load the packed bed columns reproducibly to an approximate 4 % ADH breakthrough setpoint. If regulation of the loading was carried out based on prior runs, for example Run2 was implemented to operate the two other chromatographic runs, the load cycle would have been terminated after ~39 minutes and the breakthrough level would be ~0.023 (ADH_i/ADH_0) for the large scale system and for Run1 very poor utilisation of the column capacity would be the outcome. Predictions from prior data to determine the duration of the loading cycle may be improved by taking into account the concentration of product and contaminants (protein) in the feed. However a deviation from ideal breakthrough will always be present due to poor quantitative knowledge of how the column would behave under the influence of an array of input variables.

Figure 8.5, illustrates that real time monitoring of product during the loading and subsequent breakthrough of a packed bed operation enables the operator to effectively and reproducibly run the chromatographic process to a pre-specified breakthrough setpoint despite the possible fluctuations in feed conditions and column efficiency.

Rapid measurements of product and key contaminants may also be applied for real time process analysis, giving the operator an up to date indication of the column performance. This is illustrated in **figure 8.6**. Plot A, shows the possible attainable yield (-Δ-) during the load and wash cycles for packed bed Run1. The yield term was determined in real time after each measurement step by taking the apparent amount of product retained on the packed bed column after losses in the breakthrough and wash, as a fraction of the total amount of product loaded on to the column (equation 8.1). On termination of the wash step it can be seen that approximately 2% of the ADH loaded on to the column was lost during breakthrough and the wash cycle.

Figure 8.6, plot B, illustrates the at-line purity and ADH yield values during the elution step of packed bed Run1. The purity term (-◊-) was determined from at-line data on ADH and predicted protein levels giving the operator an indication of the level of purity after each measurement step (equation 8.3). The at-line yield term during the elution stage was a function of the amount of ADH eluted as a fraction of the amount of ADH

remaining on the column after the load and wash cycle (equation 8.2). Note that the final ADH yield for the packed bed Run1 was approximately 135 %. This may be attributed to a mass balance error, activation / de-activation of the ADH enzyme or due to errors in the at-line dilution. Retrospective off-line analysis of the process showed that the total protein mass balance was within 100% ±10% accuracy indicating that the volumes used during ADH mass balancing were not the cause of the 35% increase in the ADH total balance. A >100% ADH yield was also seen using off-line data indicating that ADH activation was probably the cause for the balancing deviation. As discussed in chapter 5 ADH activation has been seen by several researchers (Smith 1997, Holwill *et al.* 1997).

8.4.2 Rapid monitoring of expanded bed chromatography

Figure 8.7, shows at-line (plot A) and off-line (plot B) measurements of the HIC expanded bed chromatography Run1. As seen from the figure the off-line and at-line data are a relatively good match in terms of their total protein and ADH chromatograms. The deviations found between the at-line and off-line measurements may be attributed to the same factors which were put forward during the rapid monitoring of packed bed systems. An additional problem involved in the real time measurements of the expanded bed system is the complexity of the process stream. During expanded bed operation disrupted cells were directly loaded on to the column prior to any clarification step, hence the feed stream contained a large number of contaminants such as protein, RNA, lipids and cell debris. The close similarity between the off-line and at-line protein levels indicated that the spectral scan prediction technique developed in chapter 4 was seemingly applicable for the monitoring of the expanded bed operation.

It can be observed from figure 8.7 that the at-line measurements resulted in an approximately 2 times higher resolution in terms of ADH data thereby enabling a better characterisation of the product chromatogram. The rapid monitoring setup also yielded RNA and cell debris data, of which the former contaminant would take up to 24 hours to measure through traditional off-line chemical techniques.

A more detailed representation of expanded bed Run1 is shown in **figure 8.8**, which illustrates the load and wash cycle in plot A and the elution step in plot B. During the

elution cycle no significant RNA or cell debris concentration was detected and hence these two contaminants were not plotted.

An important characteristic of the expanded bed operation was its large dead volume due to its expanded state which resulted in a relatively large lag phase between a change in an input variable (such as the switching from feed to wash) and subsequent registration in the eluate stream. This is illustrated in plot A. The vertical dashed line indicates when the wash cycle was initiated. It can be seen that although the load was terminated after approximately 1300 mL (20 min), the breakthrough of ADH continued until an approximate output volume of 1900 mL (29 min), i.e. an apparent lag volume of 600 mL or lag time of 9 minutes. Hence for expanded bed Run1 the *apparent breakthrough* was approximately 0.1 (ADH_i/ADH_0), whilst the *actual breakthrough* level was 0.22 (ADH_i/ADH_0).

The rapid measurements of contaminants and product were utilised for rapid process analysis for two expanded bed (EXB) runs (**figure 8.9)**. The attainable yields for Run1 and Run2 were 77% and 87%. This difference is due to the shorter loading time of Run2.

A shift to the right of the Run1 data was observed which was due to its longer wash cycle (see figure 8.10)). The final purification level for EXB runs 1 and 2 were ~5.8 and 6.2 U (ADH) mg^{-1} (protein), whilst the final yield values were ~80% and 100%. For both runs the off-line total protein mass balance were within 100% ± 12% accuracy, indicating that the <100% ADH yield deviation in Run1 may due to dilution errors in the at-line ADH assays or ADH activation / deactivation. Retrospective off-line analysis of EXB Run1 confirmed that there was a below 100% (68%) yield value during this operation, suggesting that the perceived errors in at-line dilution are less than those in the off-line measurements.

Similar to the packed bed operations, real time monitoring may be implemented for process control and automation for more reproducible operation of the expanded bed column. **Figure 8.10**, illustrates the at-line product breakthrough profiles for EXB Run1 (◊ & ♦) and Run 2 (○ and ●), during the load and wash cycle. The lag phase is again illustrated by the delay in the ADH breakthrough peak (time between end of solid symbols to the top of the breakthrough profile). The load cycle for Run2 was terminated earlier than that of Run1 resulting in a higher ADH breakthrough level for Run1.

Although the feed material for the two expanded bed runs were prepared in a similar method (chapter 3) it can be seen in table 8.1 that the ADH and total protein levels in the feed were different (probably due to the age of the baker's yeast). The variation in feed conditions could explain the slight deviation in the breakthrough profiles for the two expanded bed runs.

Breakthrough of product during packed bed operations discussed in section 8.4.1 could be controlled by implementing the at-line information on ADH directly in a feedback manner due to the relatively short lag time involved with that unit operation. By contrast the expanded bed system such an approach is not effective due to the larger lag time, i.e. what is monitored in the exit stream has taken place approximately 9 minutes prior within the column. Hence some form of prediction was required for effective regulation of the expanded bed system especially during the load and wash stages.

Figure 8.11, illustrates the EXB breakthrough profiles for runs 1 and 2 in plot A and B. Linear fits of the breakthrough profiles were performed and their coefficients of determination were R^2_{Run1} = 0.95 and R^2_{Run2} = 0.97 for EXB Run1 and Run2 indicating a relatively good fit. A least squares approach was used for parameter estimation. Three lines are shown on each plot. The dashed line represents the overall linear fit for the whole breakthrough profile, whilst the two other lines illustrate the linear fits using just the first two data points (after $ADH_i/ADH_0 > 0.005$) and all of the data points during the load cycle. The linear fits were initiated when a significant ADH level ($ADH_i/ADH_0 > 0.005$) was observed during the load cycle. From figure 8.11, it can be seen that the linear fits give the operator a relatively good prediction of the breakthrough profile, which may be implemented for control purposes. This is described in more detail below in figure 8.12.

The control of feed onto the column may be performed using prior data. Due to the slight variations in feed conditions and possible fluctuations in column capacity a control configuration based on such an approach would be susceptible to disturbances which could lead to poor regulation of breakthrough levels. If for example EXB Run2 was operated on the basis of Run1's ADH breakthrough profile given the 9 minute dead time, to achieve a 0.1 (ADH_i/ADH_0) breakthrough setpoint, Run2 should have been terminated after ~ 10.7 minutes (see plot A or implement the linear equation for

Run1 in the figure legend 8.10). A 10.7 minute loading time for EXB Run2 would have lead to a ~ 0.06 ADH_i/ADH_0 breakthrough, i.e. 40% off target (see Plot B or implement the linear fit for Run2 in the figure legend 8.10).

The effectiveness of breakthrough prediction using a straight line is illustrated in **figure 8.12**. Plot A, illustrates the outcome of the linear fits adopting latest acquired ADH data and subsequent predictions of the actual breakthrough level for the two EXB runs in terms of the number of at-line measurement used. These predictions were compared to the breakthrough levels expected by the overall linear fits for both EXB runs 1 (-■-) and 2 (-●-). After approximately 12 at-line measurements close to 100 % accuracy was achieved. The poor breakthrough predictions during measurement steps 7 to 10 for EXB Run1 were due to the noisy at-line measurements seen at 18 - 20 minutes into the load cycle (figure 8.11, plot A).

The dotted and dashed vertical lines in plot A, indicate the number of at-line measurements available for predictions given the duration of the actual experimental load cycles for the two EXB runs. For EXB Run1, 11 at-line data points were available for actual breakthrough level prediction, which was approximately 8% off target. In other words after 11 at-line measurements (17 actual at-line data points were acquired however 6 of these were during the lag phase and were therefore not used for the linear prediction) the loading cycle was terminated after an apparent breakthrough level of 0.1 (ADH_i/ADH_0). After approximately 9 minutes the actual breakthrough was detected at a 0.22 (ADH_i/ADH_0). The predicted product breakthrough level using the linear fit was ~0.24 (ADH_i/ADH_0), i.e. a 92% accuracy. For EXB Run2, the predicted ADH breakthrough level was 0.086 (ADH_i/ADH_0) compared to actual final breakthrough level of 0.105 (ADH_i/ADH_0), i.e. a 82% accuracy (predicted / actual). The prediction accuracy for Run1 was worse due to the fewer available at-line measurements prior to the termination of the loading.

Figure 8.12 plot B, illustrates the effectiveness of the linear prediction technique in terms of estimating the actual breakthrough profiles for EXB runs 1 and 2. The predicted data points were generated by implementing the available at-line measurements and the up to date linear fit to estimate the outcome of the breakthrough profile 9 minutes into the EXB runs. It can be observed from plot B that the overall estimates of the predicted values are relatively close to the overall linear fits. The poor breakthrough profile predictions around 28 minutes for EXB Run1 were due to the use of noisy at-line data 9 minutes prior to these values. Similarly for EXB

Run2 the relatively poor predictions around the 26 minute mark were amplifications of the noisy data 9 minutes prior.

If the linear prediction technique was implemented in a control configuration for regulation of the breakthrough profile the load cycle would be terminated when the ADH_i/ADH_0 term was equal to or above the actual breakthrough setpoint. For EXB Run1 if the actual breakthrough setpoint was 0.22 (ADH_i/ADH_0), based on the predicted data the load cycle would have been terminated after ~ 26 minutes resulting in an actual breakthrough of 0.181 (ADH_i/ADH_0), i.e. 18% off target. For EXB Run2 if the breakthrough setpoint was 0.1 (ADH_i/ADH_0) the control configuration would terminated the load cycle after ~28 minutes yielding an actual breakthrough of 0.126 i.e. 26 % off target.

8.5 Discussion

At-line monitoring of both product and key contaminants was demonstrated for the packed and expanded bed chromatographic operations. The benefits of real time monitoring put forward in this chapter were real time characterisation of the chromatographic systems, allowing for at-line process analysis and the potential of process control and automation to improve process reproducibility.

It has been shown that with the higher frequency of available at-line ADH measurements better characterisation of the chromatographic systems were possible. During the elution step this is especially important as only a few off-line measurements were available for the correct characterisation of this cycle, which affects process analysis issues such as chromatographic mass balancing and the practical problem of correct cutting and collection of the elution peak. Conventional off-line measurements of the chromatographic systems in terms of product and contaminants are usually performed post-experimental operation. For time critical products, assays are carried out parallel to the process requiring an additional operator. Another benefit of the at-line monitoring system was the measurements of key contaminants such as RNA, protein and cell debris allowing for real time process characterisation in terms of both product and contaminants.

A difficulty in setting up the at-line system for the chromatographic systems was the wide range of product concentrations that were required to be monitored, from 0.5% to 600% of the feed level. This was achieved by physically changing the tube diameters

during the stopped flow analysis sequence to alter dilution levels. An alternative approach would be to automatically change the flowcell pathlength applied in the spectrophotometer, by implementing a flowcell carousel, i.e. switching from a 10 mm to a 1 mm pathlengh flowcell would be equivalent to a 10 fold-dilution. The resolution of contaminant data could also be increased by performing the pre-dilution and rapid spectral analysis at-line.

The potential benefits of rapid automatic assays have been shown in terms of real time analysis of the chromatographic process. Real time process analysis gives the operator up to date information on the performance of the chromatographic systems allowing their operation to be conducted in an intelligent and informed manner. By performing the process analysis in real time the operator can focus on planning the subsequent chromatographic run based on the up to date analysis and hence accelerate the process development cycle. Correct process analysis and mass balancing of the chromatographic operations was a difficult task using either off-line or at-line data, especially when this was performed with respect to a product's level of activity. A mass balance of a unit operation should balance whilst in terms of product activity this is not necessarily the case. Other factors that complicated the analysis of the chromatographic operation were its dynamic nature and ability to concentrate product up to 60 fold. The vast concentration range of product in the elution cycle gives room for the potential of high measurement errors due to the large level of dilution required. Dynamic disturbances may also effect the mass balancing of the process. An increase or decrease in the flowrate during the load, wash or elution cycle would have a great impact on the final volumes and hence component balances.

Rapid monitoring of the chromatographic operations also enabled the potential for process control and automation. Traditionally chromatographic operations are operated using knowledge of prior runs and/or the use of indirect monitoring techniques such UV280. As feed material to the chromatographic operations are usually the outcome of a series of upstream operations the composition of this feed is susceptible to fluctuations. Additionally if a column is being re-utilised for several chromatographic runs its performance will change with time. Due to these possible process fluctuations any control configuration based on the use of prior data would be very sensitive to process disturbances. Account may be made to operating conditions based on the characteristics of the feed material, however an exact quantitative prediction of how the column would behave due to feed composition variations is a complex problem. Indirect monitoring methods can pick up general changes in

contaminant protein levels, however a direct indication of product and key contaminant levels is not possible making efficient chromatographic control virtually impossible especially during the load cycle.

Rapid monitoring of product and key contaminants enables all stages of the chromatographic process to be controlled to specific loading, washing and elution criteria. In the load cycle the product breakthrough was controlled to a specific operator chosen setpoint by applying real time information on product in the eluate stream. Appropriate control to terminate the wash cycle could be performed by using the real time information on the contaminant levels (protein). In the elution step real time measurements on product could trigger the start and end of peak collection.

Correct loading of the chromatographic systems is vital for the optimisation of the column performance in terms of yield, purity and productivity. If the loading cycle is too long product losses during breakthrough and the wash cycle will be seen. However, too short a load cycle would result in poor utilisation of the column capacity. Due to the importance of correct column loading, regulation of this chromatographic cycle has been investigated in terms of controlling the level of product breakthrough to a specific setpoint.

For the packed bed chromatographic system a simple feedback control configuration could be applied directly controlling the termination of the load step triggered by real time measurements of the product level in the eluate. As the dead volume associated with the packed bed column is relatively small tight control could be achieved. For the expanded bed system some form of breakthrough profile prediction was required due to the relatively large dead volume associated with this type of chromatographic operation. Although several models exist to characterise the breakthrough profile this chapter sought to implement a straight line to estimate breakthrough characteristics due to its ease of use and relatively good fitting performance.

It has been shown that the performance of the linear breakthrough predictions was a function of the number of at-line measurements available for fitting and the level of measurement noise. Relatively good prediction of product breakthrough profiles for two expanded bed runs was achieved, with 87% accuracy on average. When the linear prediction technique was utilised within a control configuration the system would on average only be 22% off the target breakthrough setpoints. This should be compared to the use of prior data where the accuracy of regulating the loading to a

setpoint product level was 40% off target. Improvements to the breakthrough profile predictions can be made by either increasing the frequency of at-line data or by reducing the noise associated initial breakthrough measurements. An alternative approach to improve the control of the expanded bed operation would be to eliminate the large dead volume associated with this operation. By sampling from within the column bed one can rapidly get an indication of the actual breakthrough profile characteristics. Work in this area is currently being done at UCL.

Variation in feed conditions during expanded bed operation is a function of fermentation and homogenisation conditions. The feeds for packed bed systems is subject to larger fluctuations due to the larger number of unit operations upstream. Tight control during packed bed operation was therefore crucial and was shown to be achievable. During the expanded bed operation breakthrough regulation was improved with the aid of rapid measurements and at-line breakthrough profile prediction.

8.6 Conclusion

Rapid monitoring of both packed and expanded bed chromatographic operations has been demonstrated in terms of key contaminants and product. Real time process analysis was demonstrated in order to accelerate the process development cycle by enabling the operator to initiate the design of the subsequent experimental run without the need for time costly post process analysis. At-line monitoring was also implemented for the control of the loading cycle in order to insure the breakthrough of product was regulated to a specific setpoint. For packed bed systems tight control was achieved despite large fluctuations in feed conditions. For the expanded bed column due to the intrinsic large dead volume associated with this operation a linear based prediction technique was used in combination with at-line measurements to achieve reproducible control of the loading cycle.

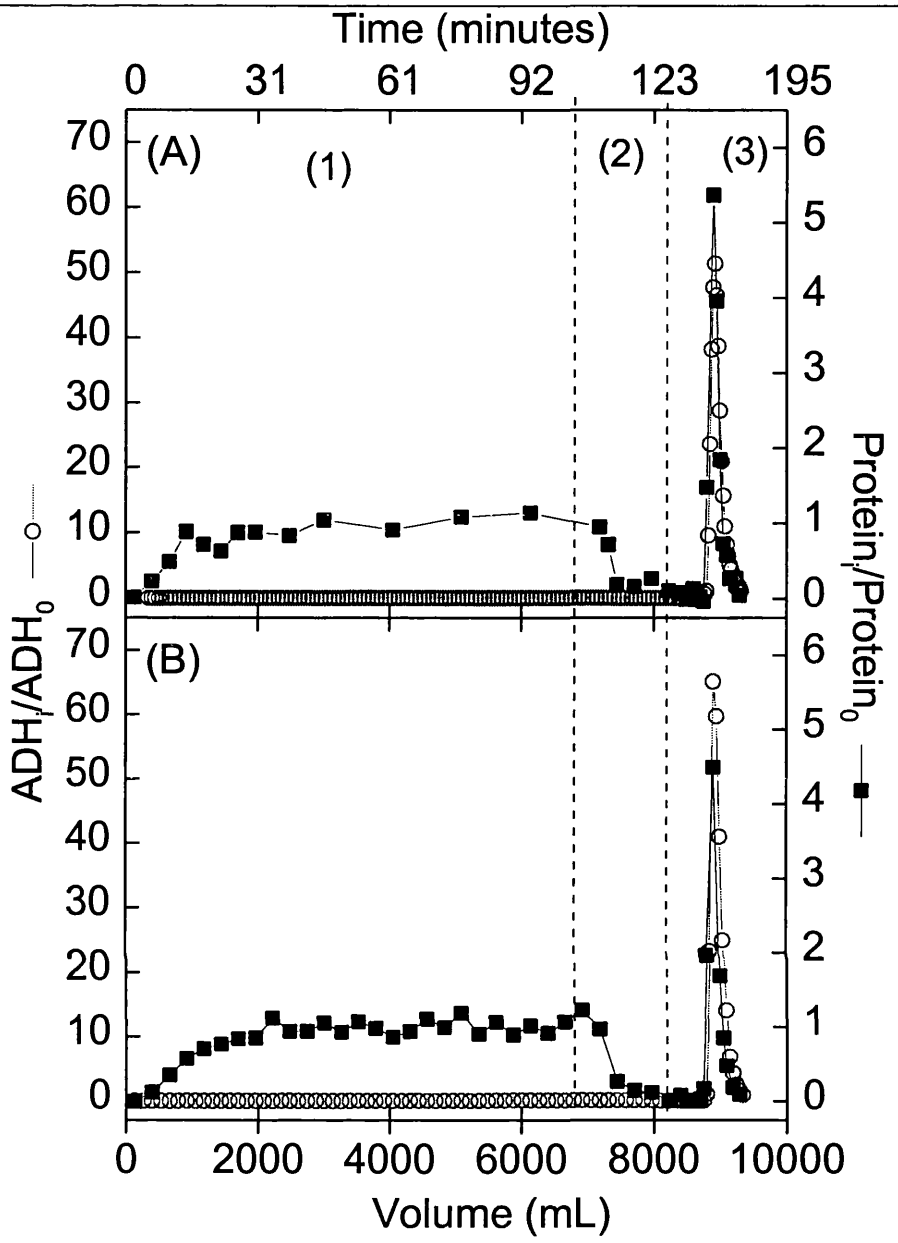


Figure 8.2. *At-line (Plot A) and off-line (Plot B) measurements of the 0.05 m diameter HIC packed bed chromatography Run2. Recovery of ADH from a 0.125 m bed height XK50/40 column packed with Phenyl Sepharose FF (low sub) loaded to ~ 5% breakthrough of ADH.* ADH_i *and* ADH_0 *refer to the concentration of ADH at measurement step* i *and the feed. Column washing was performed in the same direction to loading and elution. Clarified Baker's yeast homogenate was obtained from the supernatant of a batch flocculation step. The concentration ADH and protein in feed was approximately 120 U mL^{-1} and 8.3 mg mL^{-1}. Off-line ADH measurements were taken every 120 s of pooled fractions, whilst at-line measurements were acquired every ~50 s (non pooled samples). Protein measurements for both at-line and off-line were taken of pooled fractions every 240 s in critical process areas (load and elution). (1) refers to the load, (2) the wash and (3) the elution cycle.*

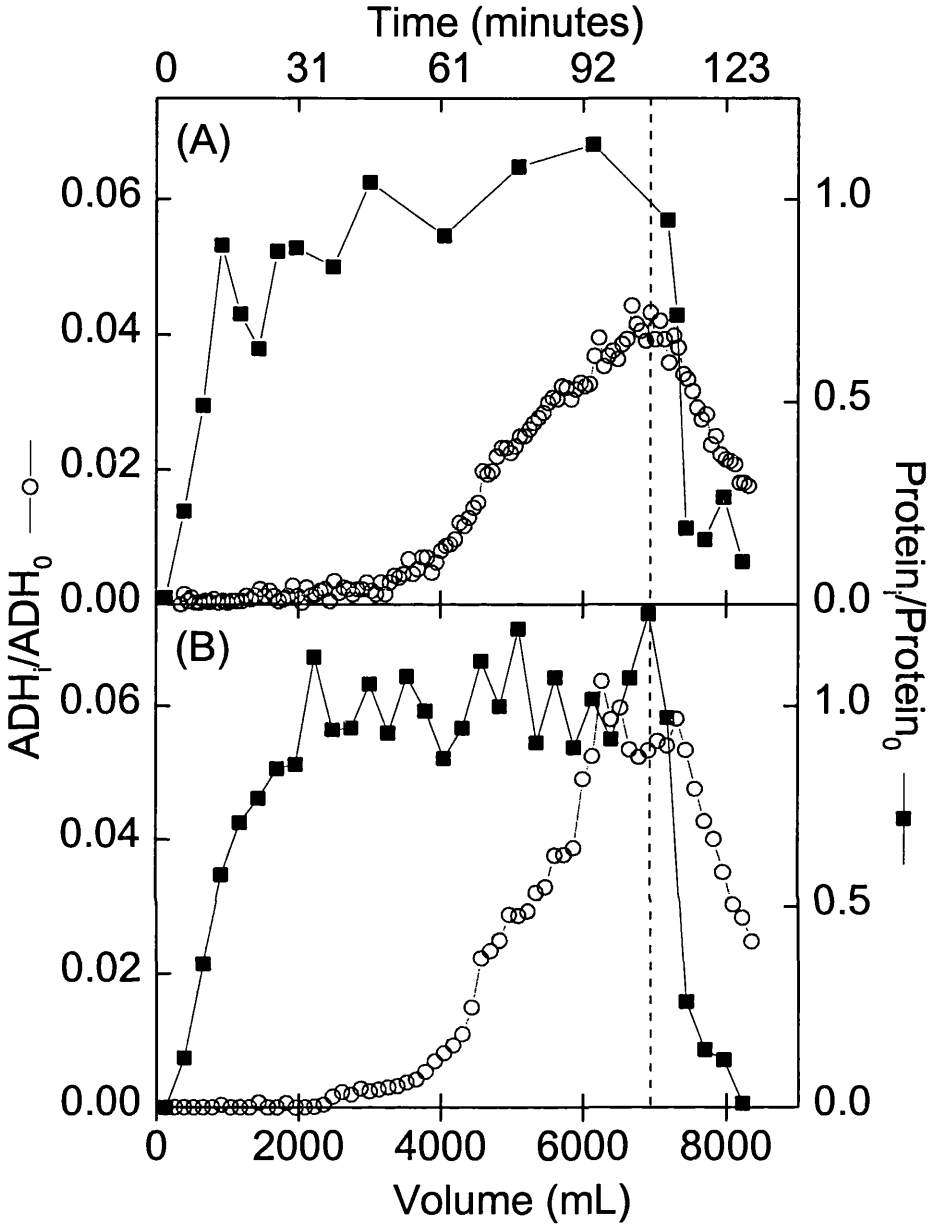


Figure 8.3. *At-line (Plot A) and off-line (Plot B) measurements of the 0.05 m diameter HIC packed bed chromatography Run1 load and wash cycles. Operating and feed conditions are listed in the legend of figure 8.2. The dashed vertical line indicates the termination of the load cycle.*

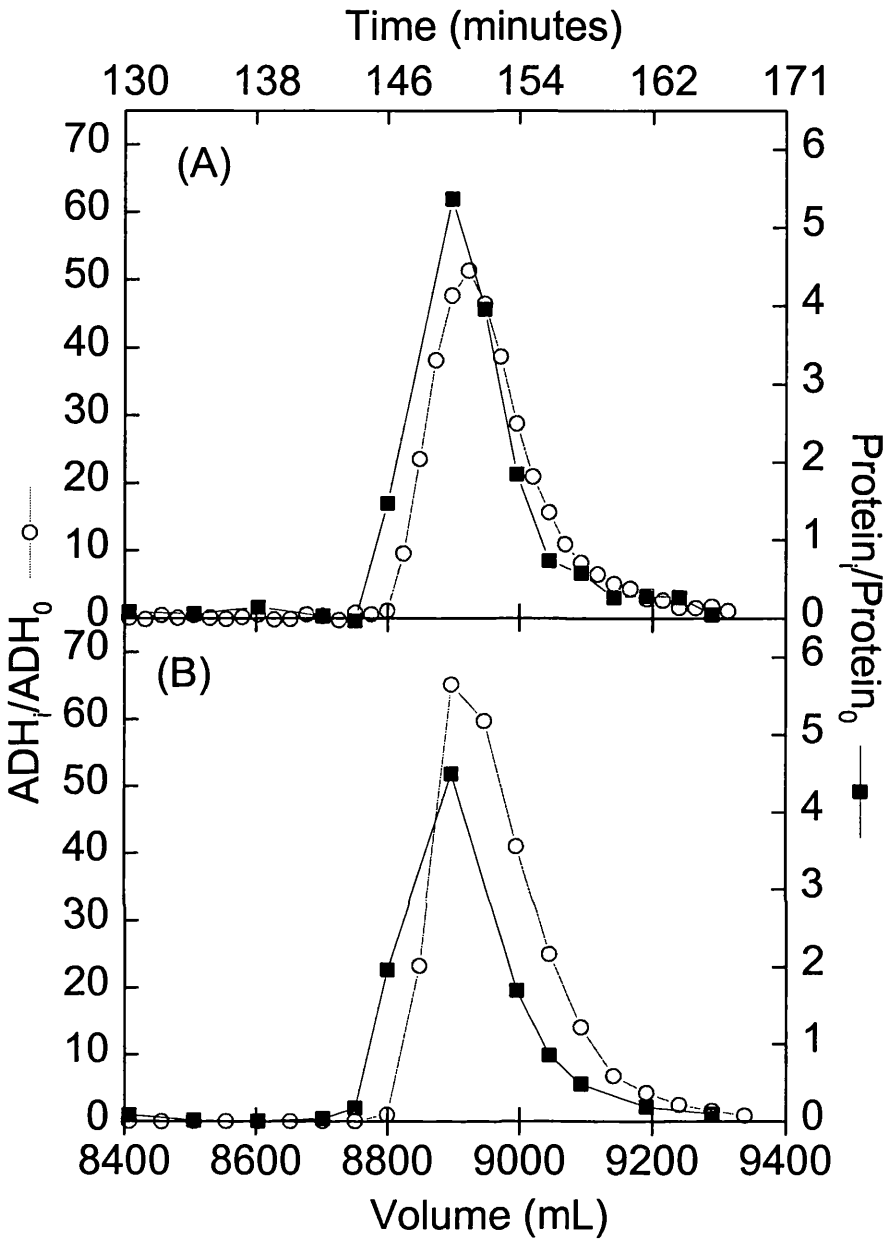


Figure 8.4. At-line (Plot A) and off-line (Plot B) measurements of the 0.05 m diameter HIC packed bed chromatography Run1 elution cycle. Operating and feed conditions are listed in the legend of figure 8.2.

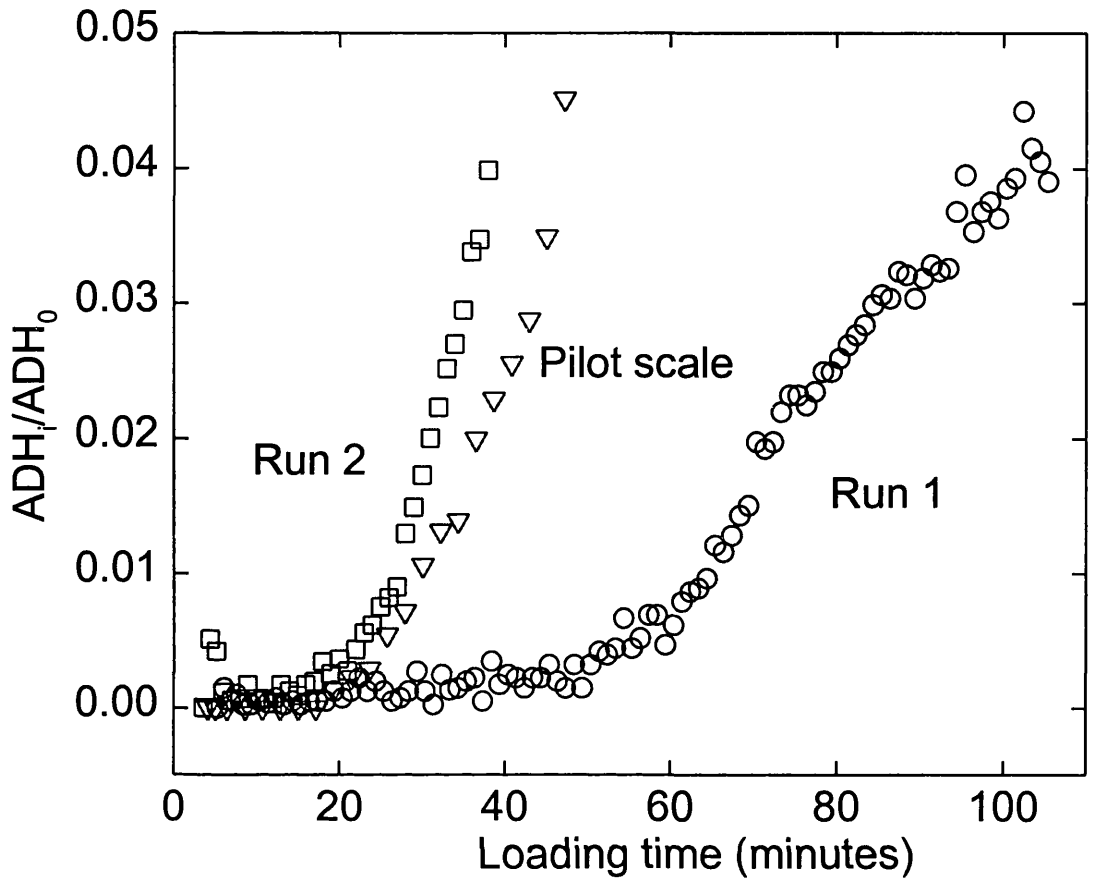


Figure 8.5. *At-line measurements of ADH breakthrough profiles of two bench scale and a pilot scale HIC packed bed chromatography processes in terms of time. Operating and feed conditions for Run1 are described in figure legend 8.2.*

The large scale chromatography run was performed on a 0.129 m bed height BPG200/500 0.2 m diameter column packed with Phenyl Sepharose FF (low sub) loaded to ~ 5% breakthrough of ADH. The column feed was clarified Baker's yeast homogenate from a continuous PEI flocculation run. The concentration of ADH and protein in the feed was approximately 173 U mL^{-1} and 11.2 mg mL^{-1}.

The bench scale Run2 was performed on a similar column to that of Run1. The concentration of ADH and protein in the feed was approximately 120 U mL^{-1} and 8 mg mL^{-1}

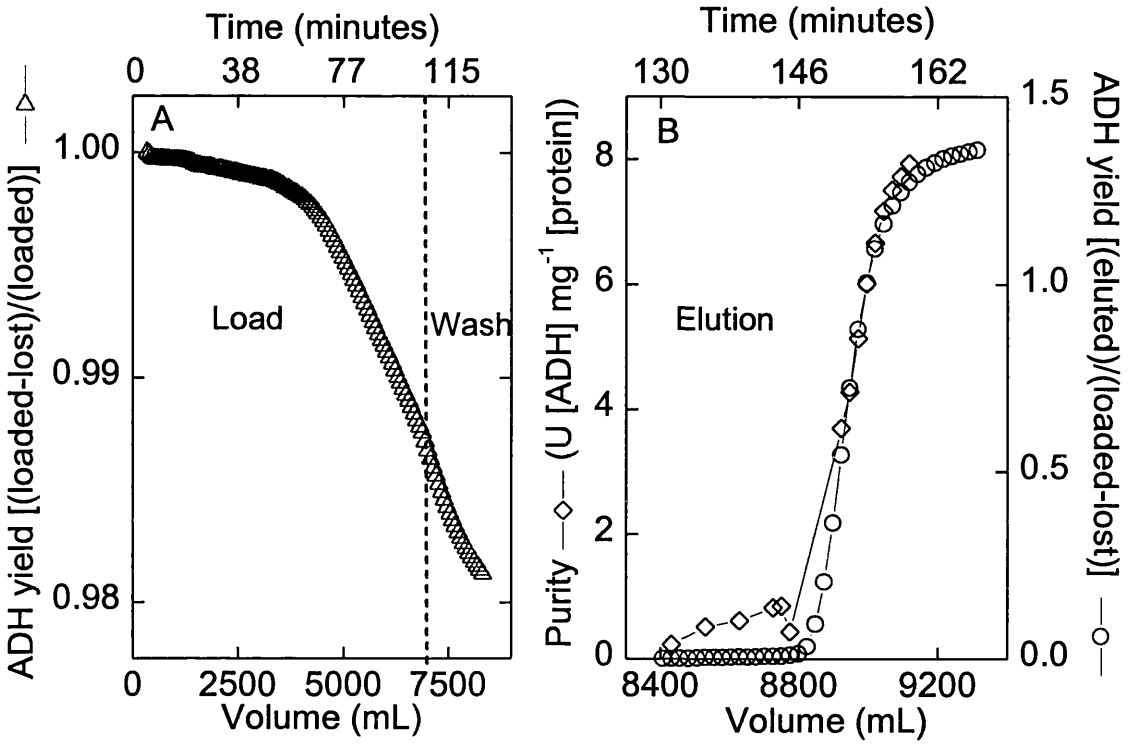


Figure 8.6. *At-line purity and yield data for packed bed Run1. The feed and operating conditions for Run1 is listed in the figure legend 8.2.*

Plot A, the yield term was determined by subtracting the amount of ADH lost in the load and wash cycle from the amount loaded on the column and taking this value as a fraction of the total ADH loaded on to the column. The dashed vertical line indicates the step change between the load and wash cycle.

Plot B, the elution cycle. The purity data was determined by taking the total ADH level as a fraction of the total protein amount of protein at each measurement step. The yield term was determined as the amount of ADH coming off the column as a fraction of the total ADH remaining on the column after the load and wash cycle. The ADH yield during the elution cycle was greater than unity. This was discussed in section 8.4.1, page 252.

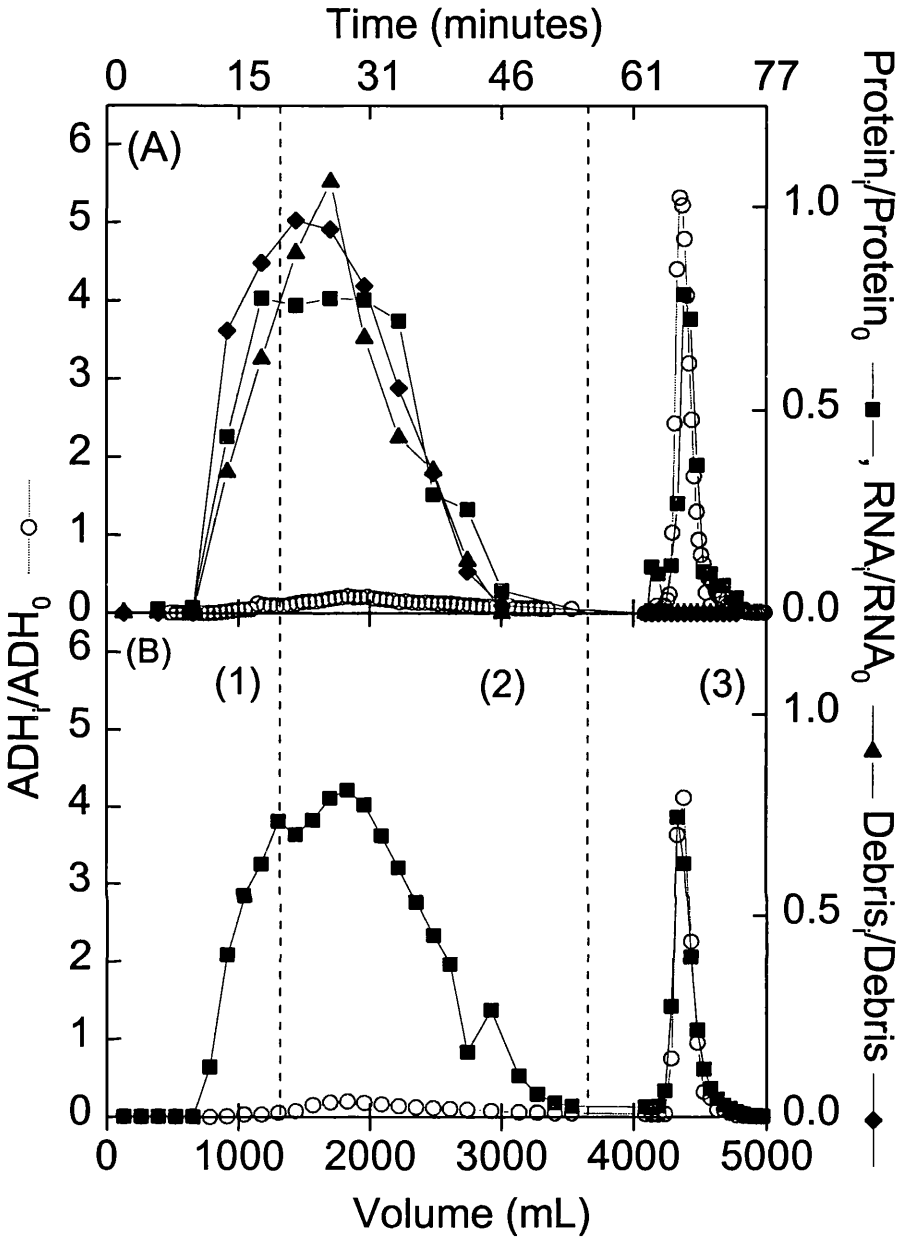


***Figure 8.7**. At-line (Plot A) and off-line (Plot B) measurements of the 0.05 m diameter HIC expanded bed chromatography Run1. Diluted homogenate was loaded onto a ST-50 expanded bed containing 300 mL of STREAMLINE-Phenyl (low sub) matrix. The settled bed height was 0.148 m and expanded bed height was 0.47 m. The concentration of ADH and protein in the feed was approximately 145 U mL^{-1} and 10.8 mg mL^{-1}. (1) refers to the load, (2) the wash and (3) the elution cycle.*

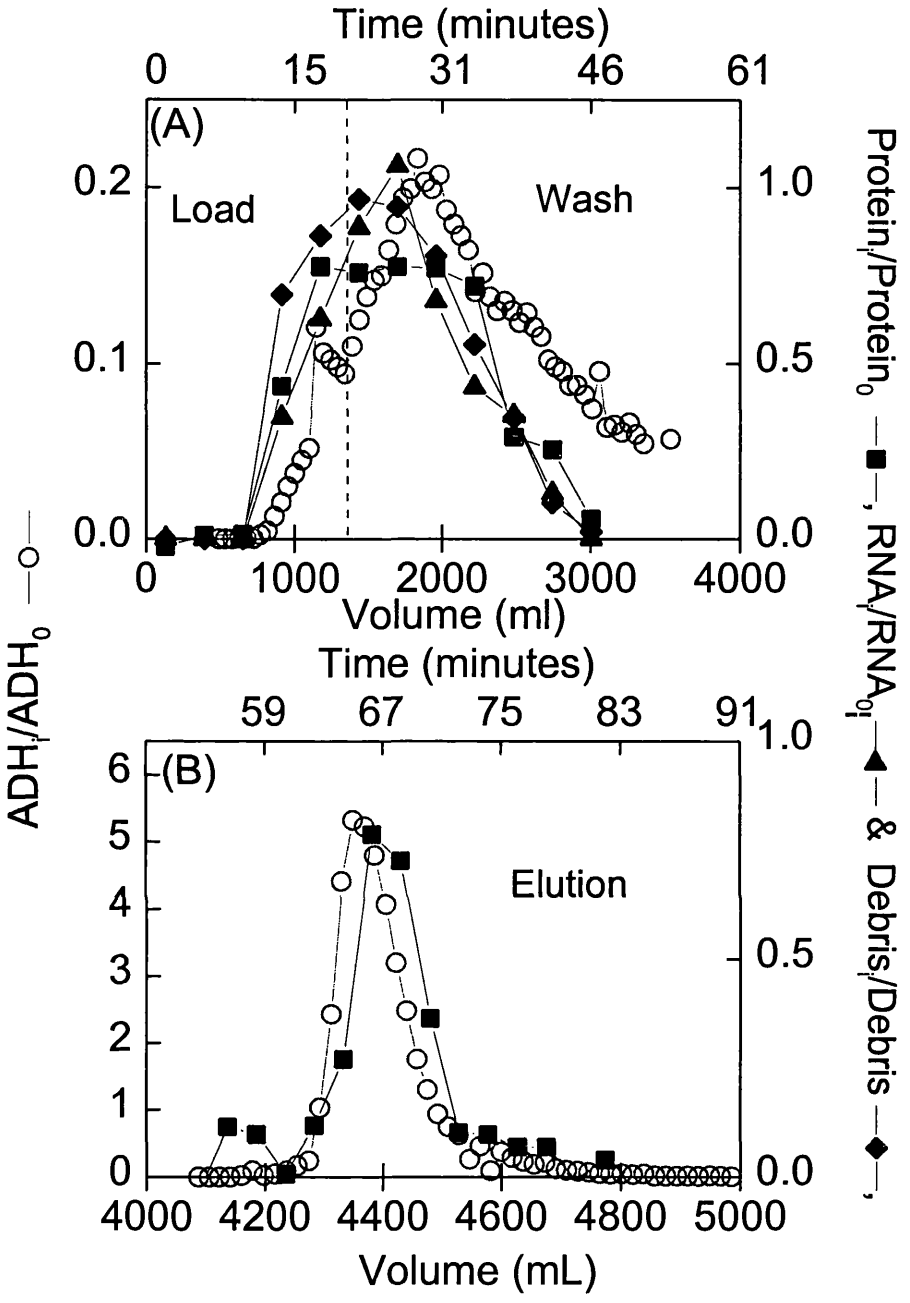


Figure 8.8. *At-line measurements of HIC expanded bed chromatography Run1 load (Plot A) and elution cycles (Plot B). Feed and operating conditions are listed in figure legend 8.7.*

The vertical dashed line indicates in Plot (A) refers to the start of the wash step.

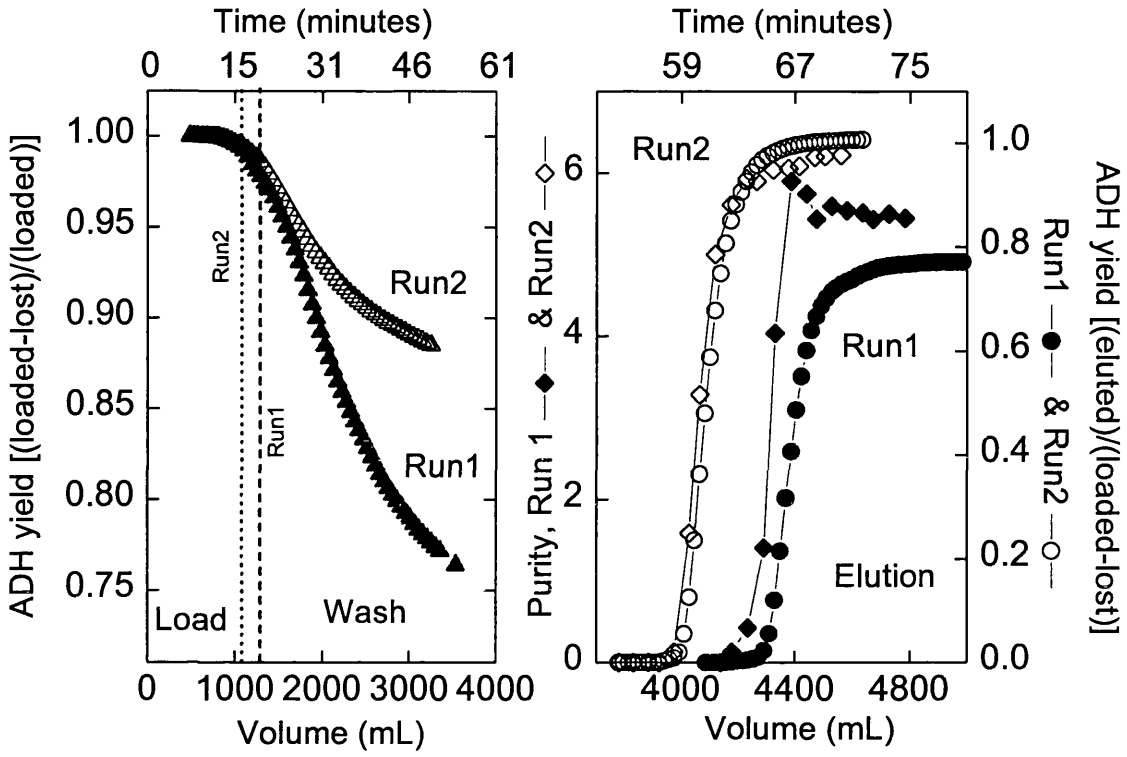


Figure 8.9. *At-line purity and yield data for expanded bed Run1 and Run2.*

The feed and operating conditions for Run1 are listed in the figure legend 8.7.

The expanded bed (EXB) Run2 was performed on a similar column to that of EXB Run1. The concentration of ADH and protein in the feed was approximately 92 U mL^{-1} and 10.5 mg mL^{-1}

The ADH yield and purity terms in both plot A (load and wash cycles) and plot B (elution cycle) were determined in similar manner those in figure 8.6.

The dotted and dashed vertical lines in plot A refer to the start of the wash cycles for EXB runs 2 and 1.

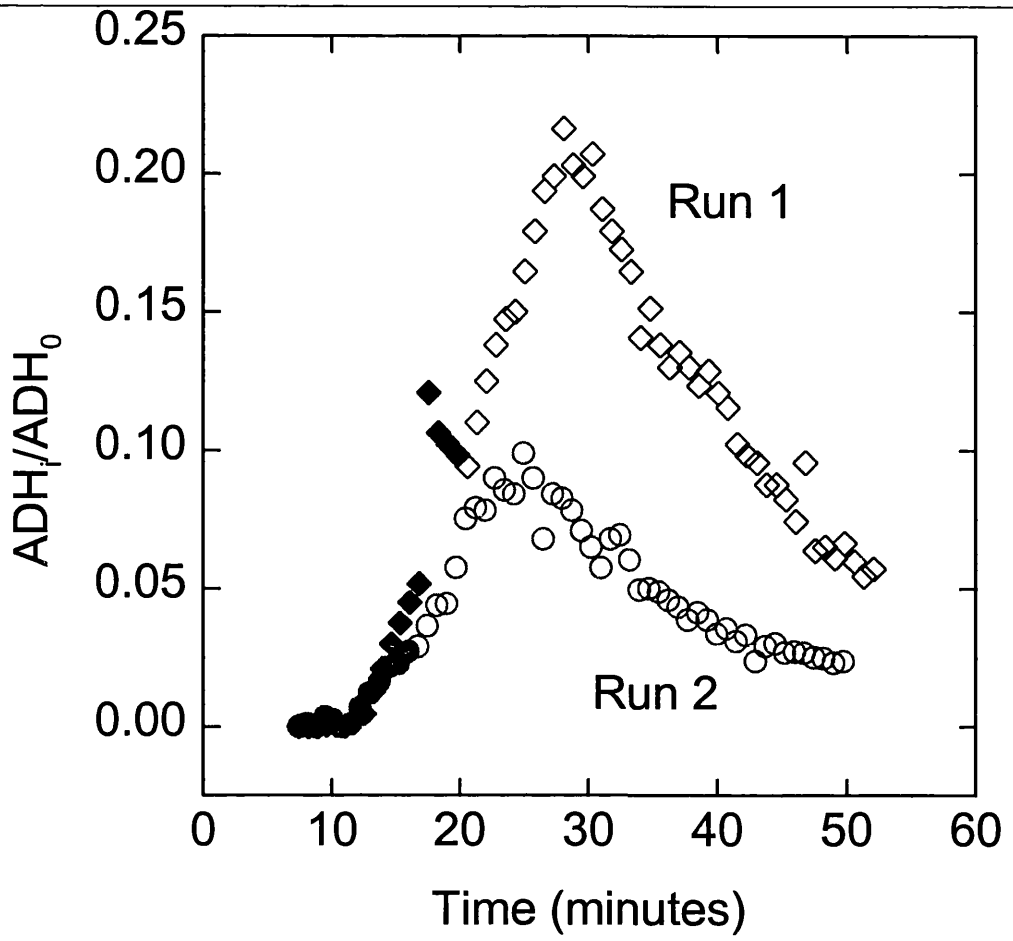


***Figure 8.10**. At-line ADH breakthrough profiles for EXB Run1 and Run2.*
The feed and operating conditions are described in figure legend 8.7 and 8.9. The solids symbols (♦ & ●) represent at-line ADH measurements acquired during the load cycle. The open symbols (O & ◊) refer to the at-line ADH measurements acquired during the wash cycle.

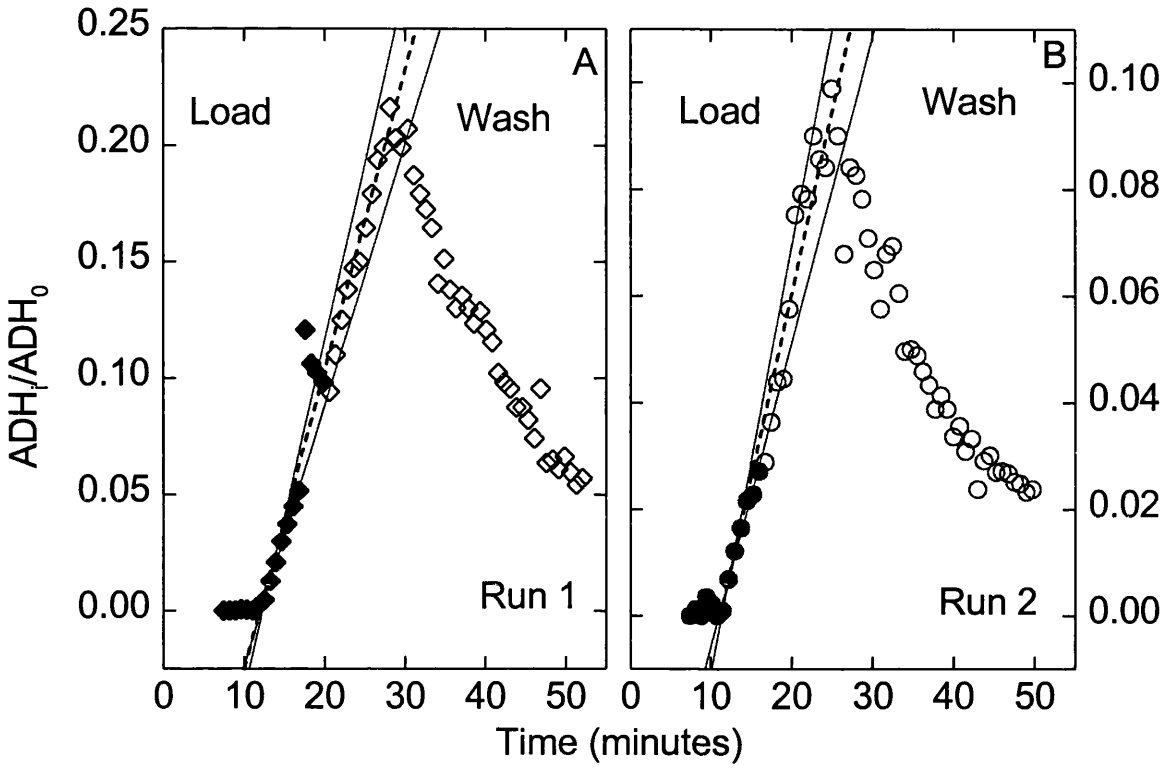


Figure 8.11. *At-line measurements and linear predictions of ADH breakthrough profiles of EXB Run1 and Run2. Operating and feed conditions for Run1 and Run2 are described in figure legend 8.7 and 8.9. The solids symbols (♦ & ●) refer to at-line ADH measurements acquired during the load cycle. The open symbols (O & ◊) refer to the at-line ADH measurements acquired during the wash cycle. The dashed lines refer to the overall linear fits of the apparent breakthrough profile and their regression coefficients were* R^2_{Run1} *= 0.95 and* R^2_{Run2} *= 0.97 for EXB runs 1 and 2. The two additional lines (per plot) are from the linear fits of the first two at-line data points (after* $ADH_i / ADH_0 > 0.005$*) and all of the measurements acquired during the load cycle.*

The linear equations for the total breakthrough profiles (---):

$$y_{ADH\_Run1} = 0.01295 \text{ x time} - 0.15544$$

$$y_{ADH\_Run2} = 0.00742 \text{ x time} - 0.08744$$

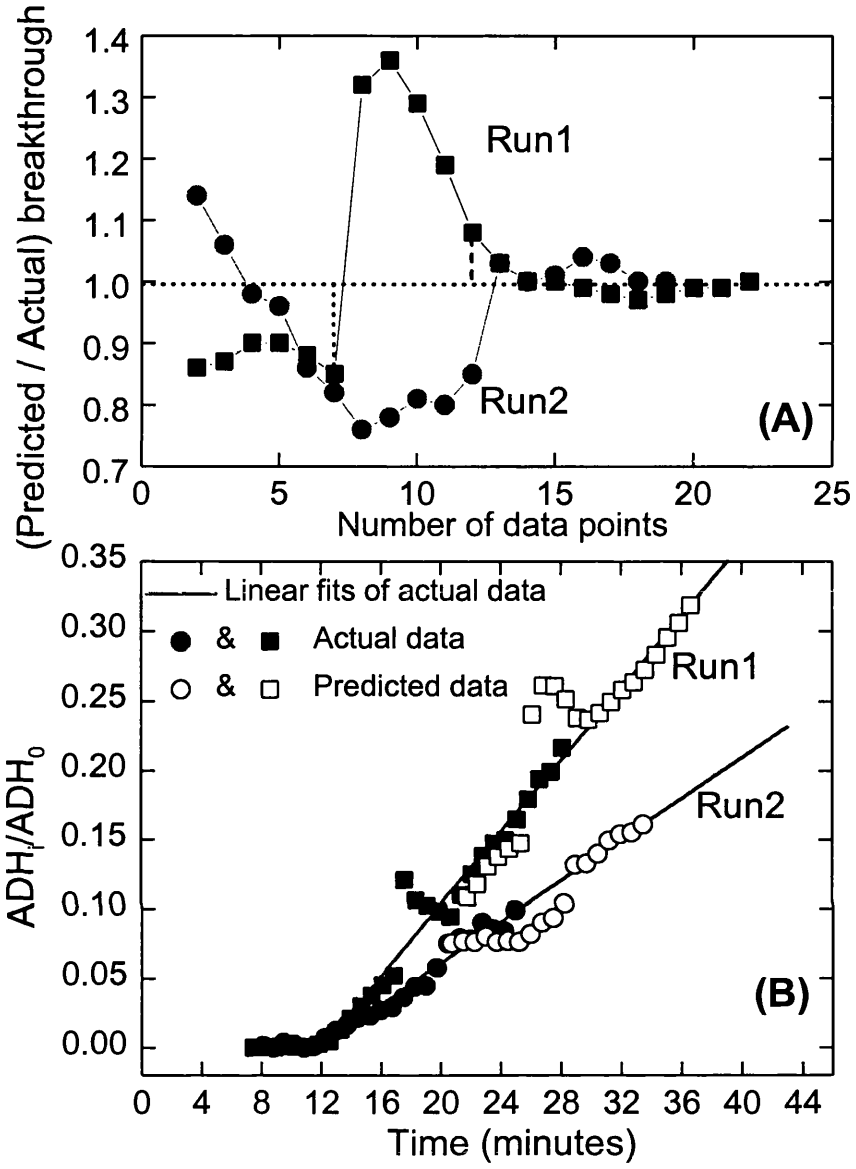


***Figure 8.12**. Linear prediction versus actual breakthrough product levels for expanded bed runs 1 and 2.*

Plot A. Linear breakthrough prediction for EXB Run1 and Run2 given a 9 minute lag time as a percentage of the actual breakthrough values for an increasing number of at-line measurements. The actual breakthrough data are based on the overall linear fits (see equations in figure legend 8.11).

Plot B. Breakthrough profiles for Run1 and Run2 using the at-line data, the overall linear fits of the actual measurements and the predicted data based on a 9 minute lag time, against process time. ADH_i and ADH_0 refer to product levels at measurement step i and in the feed.

Chapter 9

Conclusion and future work

9.1 Conclusion

The aim of this thesis was to apply rapid measurement of product and key contaminants during the recovery of alcohol dehydrogenase from *S.cerevisiae* cell homogenate, for enhanced process definition, rapid characterisation and control of the batch and continuous flocculation process as well as the packed and expanded bed chromatographic operations.

The rapid monitoring methods applied were of an at-line nature, i.e. a sample was drawn of the process, prepared and transported to a detection instrument for product and contaminant measurements automatically. Chapter 4 demonstrated the applicability of adopting an existing off-line enzymatic alcohol dehydrogenase (ADH) assay and converting it for at-line purposes through the use of a stopped flow analyser (SFA). The high reproducibility rate of the SFA allowed for a reduced assay time of 10 s compared to the traditional 60 s applied during off-line analysis. The total ADH at-line assay time was approximately 45 s which included sample dilution, reagent mixing, transportation to a detection instrument as well as a wash cycle.

Protein and RNA absorb light readily in the ultraviolet (UV) spectra especially between 260 nm and 280 nm due to the presents of tyrosine and tryptophen, whilst cell debris particulate absorbs light in the visible (VIS) region. This concept was used to build a calibration model in order to predict the level of contaminants (RNA, protein and cell debris) from the analysis of the their spectral scans. A factorial experiment consisting of 36 samples of different known concentrations of RNA, protein and cell debris were used for the calibration of the UV-VIS spectral data. A multivariate statistical technique called partial least squares (PLS) enabled the efficient and effective calibration model building by regressing the known contaminant concentration levels to the main UV-VIS spectral features in the range between 230 - 500 nm. The calibration model was subsequently applied for the prediction of contaminant concentrations in Baker's yeast homogenate for unknown samples from their UV-VIS scans. The spectral scans were

performed by an in house built diode-array spectrophotometer capable of acquiring UV-VIS scans every second. In chapter 4 it was demonstrated that the prediction efficiency was very good with standard errors of prediction for RNA, protein and cell debris of 0.11 mg mL^{-1}, 0.41 mg mL^{-1} and 0.136 Au. This includes any errors involved in sample handling and dilution. The use of the SFA in combination with the diode-array spectrophotometer enabled automatic acquisition of contaminant measurements in a rapid, effective manner with little effort compared with the complex, time consuming variable nature of the existing off-line chemical techniques.

For rapid monitoring of the flocculation process described in chapters 5 to 7, a solid-liquid separation step was required to evaluate the performance of polyethyleneimine (PEI) flocculant's ability to selectively flocculate contaminants whilst maintaining the product (ADH) in solution. The at-line monitoring setup therefore consisted of an automated microcentrifuge followed by a SFA and two spectrophotometers for the detection of an enzymatic reaction and the rapid scanning of diluted samples for contaminant prediction. The measurement cycle was approximately 135 s with measurement errors of ±8% most of which was due to the variability of the microcentrifuge.

Chapter 5 demonstrated the ability to consistently characterise eight batch flocculation runs in terms of their at-line product and contaminant measurements. The rapid measurements allowed the design engineer to demand a higher resolution of data points without excess effort improving process characterisation. Rapid monitoring techniques could substitute the laborious time consuming off-line assays eliminating the variability associated with manual measurements. This chapter investigated the use of three mathematical descriptions to describe the flocculation behaviour in terms of the flocculant volume fraction, were the model parameters were related to the environmental conditions, such as pH and ionic strength. A three parameter empirical model, the 3(2)-parameter model was concluded to be most appropriate for real time characterisation due to its effectiveness to describe the flocculation behaviour with the least number of model parameters. Analysis of the environmental conditions in terms of the model parameters yielded a linear relationship between pH and the flocculation profile inflection point and therefore optimum flocculant conditions. Such a linear correlation may be applied during control of the flocculation process and the pH level could be implemented as an additional manipulative variable to that of the flocculant dose.

The at-line data for the eight batch flocculation operations was implemented for real-time flocculation process characterisation, optimisation and control in chapter 6. The at-line data of both product and key contaminants in combination with a process model and a model parameter identification technique enabled up to date characterisation of the flocculation behaviour. The real time data and up to date model fits can be presented graphically to facilitate an immediate and comprehensive overview of process conditions and performance. This visual representation of the flocculation performance could be implemented to drive the process to ideal operating conditions. Two model parameter identification techniques were examined, the Levenberg-Marquardt non-linear least squares and extended Kalman filter. The optimum model parameter estimation method to apply was a function of several factors, such as the knowledge of the process, level of measurement noise, frequency of data points and level of batch to batch and within batch fluctuations. By adopting a model parameter identification technique data reconciliation was performed enabling the any decisions made to be based on filtered data.

Chapter 6 also examined the potential of applying the at-line data on product and contaminant level for process control whereby the flocculant dose was implemented as the sole manipulative variable. Two closed loop control configurations were examined, which utilised different control objectives and level of complexity. A control configuration attempting to optimise the removal of contaminants and recovery of product for the minimum flocculant dose, given that at least 80% of the cell debris in the spun feed was removed was examined. This approach utilised weighting factors to prioritise the importance of the individual contaminants, product and concentration of flocculant. These weighting factors could be chosen so that the control configuration attempted to control the batch flocculation in order to reach conditions ideal for the subsequent unit operation. The optimisation algorithm required a model of the flocculation process and hence both the EKF and NLLS methods were looked into for model parameter predictions for each newly acquired measurements. Good control to optimum operating conditions was achieved implementing both estimation techniques with on average 92% accuracy. The NLLS based approach tended to control the PEI volume fraction to a slight overdose, due to its slower ability to predict the correct model parameters than that of the EKF. A control configuration applying raw cell debris yield data as the sole control variable was also examined. This approach would terminate the flocculant dosing when the cell debris yield level was below a setpoint of 0.2, i.e. at least 80% removal of cell debris based on spun feed. The performance of this control setup was relatively good although it did not optimise the removal of all the

contaminants. However for unit operations where tight control is not required and measurement noise is low such a control system would be adequate. Chapter 6 also illustrated that the design of the batch flocculation process in terms of dosing rate, pH and salt concentration has a great impact on the performance of a control configuration. By choosing extreme operating conditions robust and efficient control was difficult to achieve.

Chapter 7 examined the characterisation of the continuous flocculation process using at-line data. The primary objectives put forward were the speed and stability of process characterisation. Two searching methods were investigated in terms of how different sequences of at-line information would influence the behaviour of model parameter estimations. A sequential and controlled search technique was examined and it was concluded that by applying a structured and intelligent use of the at-line data more sensitive at-line information could be given to the estimation techniques improving the characterisation of the flocculation process in terms of speed and stability. Rapid and stabile process characterisation was subsequently applied for fast and robust process optimisation.

Chapter 8 demonstrated the real time monitoring of both product and key contaminants for the characterisation, process analysis and control of the packed and expanded bed chromatographic operations. The at-line data enabled a higher resolution of data points to be acquired resulting in improved process characterisation compared with the use of traditional time consuming and laborious off-line assays. Further to this real time process analysis was achieved accelerating the process development cycle by allowing the design engineer to prepare subsequent experiments rapidly in an intelligent and informed manner. During the control of the packed system at-line product measurements were applied in a simple feedback control arrangement to regulate the load cycle. For the expanded bed operation a predictive based control configuration was implemented in order to compensate for the relatively large dead volume associated with this unit operation. In both cases good control of the load cycle was demonstrated despite the variation in the feed material. For the expanded bed operation the feedback control system resulted in approximately 80% accuracy in regulating the load cycle to the breakthrough setpoint, whilst for the packed bed system the performance of the control configuration was even better.

The above chapters have shown that by adopting at-line data on product and contaminant levels, rapid process definition, characterisation and control of

bioseparation processes is achievable. Real time process information may be represented in an optimum visual manner giving the development engineer a powerful tool to facilitate the running of a process. Process information around critical operating areas can be targeted to enhance the knowledge of a unit operation. Furthermore at-line data can substitute laborious time consuming off-line assays allowing for a higher resolution of data points to be acquired and more ambitious factorial experiments to be performed with less effort and in a shorter time frame. At-line information of a unit operation is usually acquired in a discrete format, contains some time delay and associated with relatively large measurement noise compared to traditional measurements such as temperature and pH. This work sought to implement model based systems to compensate for the slow and noisy nature of at-line data. The implementation of model based systems enabled the whole process performance to be represented in a graphical manner from just a few at-line data points by implementing model parameter identification technique. Based on these visual aids the operator would rapidly and effectively obtain an overview of the process performance.

The applicability of at-line data for process control was also demonstrated for the flocculation (batch and continuous) and chromatographic (packed and expanded) operations given possible disturbances in feed material. The model based approach was vital for robust control as it enhanced the method of applying the available at-line data through noise reduction, possibility of process performance prediction and process optimisation. During the flocculation and expanded bed chromatographic operations the model based approach enabled the prediction of how the process would behave at a particular flocculant dose or at a specific time in the future. This information would be applied during the control configuration in order to control an event which could not be detected fast enough due to monitoring or process lag delays. For process optimisation a process model is crucial and by implementing a model parameter identification technique after each acquired at-line measurement optimum operating conditions may be determined in real time, batch to batch.

The control examples given in this work demonstrated that product and contaminant levels could be defined and subsequently controlled in real time through the manipulation of input variables. Bioseparation processes are inherently variable due to the fluctuations in upstream conditions, process additives and sub-optimal performances of processing equipment. Rapid monitoring allows a process to be followed in real time and operating conditions to be altered in order to compensate for any disturbances. Process control of directly related to products and contaminants will

reduce the variability of bioprocesses performances yielding a more reproducible and hence better quality product.

9.2 Future work

The benefits to be gained through real-time bioprocess monitoring have been highlighted in terms of enhanced process definition, characterisation and control. Although the at-line monitoring setup adopted in this work was specific for the rapid measurement of a target enzyme and key contaminants in *S.cerevisiae* cell homogenate the advantages of rapid monitoring may be applied to other systems such as *Escherichia coli* and mammalian cell systems. The rapid data will differ both in terms of speed, accuracy and measurement methods. The availability of numerous rapid monitoring techniques exists such as biosensors, immunoassays or HPLC systems which would accommodate most systems in terms of rapid product measurements. Rapid spectroscopy is seen to have a great potential as a non-invasive method for the acquisition of process information. Although the system examined will have a different analysis technique the rapid data will posses similar constraints such as discrete measurements associated with relatively high noise. Hence the techniques examined in this work will be applicable for many systems in terms of achieving enhanced process definition and rapid process characterisation. Rapid measurements in combination with a process model and model parameter estimation techniques offer the development engineer an intelligent technique to implement scarce data associated with measurement delays and noise. The at-line data can be shown in an optimum visual manner to give the operator an immediate and comprehensive overview of the process behaviour.

One of the primary objectives of pharmaceutical companies is to reduce the time spent to bring a drug to market. Time consuming events such as screening of downstream protocols and definition of the unit operations may be carried out in a faster and more efficient manner through the implementation of rapid measurements. The availability of automated and rapid measurements on product and key contaminants will yield a higher resolution of data points, allow for more thorough factorial experiments hence enhance the definition and characterisation of a bioprocess. Rapid monitoring is becoming a necessity in bioprocess development as demand increases for time efficient ways of extracting information from each experiment (Olsson 1998). Real time process information can give the development engineer an intelligent technique to facilitate the running of a process to gain optimum information around critical operating

areas to enhance the knowledge of a unit operation. Rapid measurements are therefore fundamental for the improvement of the process development cycle in terms of speed and efficiency.

An important aspect of rapid bioprocess measurements is their relatively large noise values and delays hence some form of structural data reconciliation is required to handle the at-line statistical uncertainty. Model based systems have an advantage as prior knowledge of a particular process and system may be incorporated and the resultant data may be represented in terms of a process profile giving the operator an overview of the process performance. However, in many cases a model of the bioprocess for a given system does not exists or is too complex to be implemented in real time. In such cases an empirical model may have to be constructed.

Due to the complexity involved in monitoring the level of a specific product accurately in real time regulatory authorities traditionally impose strict regulations on keeping the operating conditions of a process to specific levels. The regulatory authorities will therefore tend to define the process rather than the product quality directly. Currently control of a bioprocess is achieved by controlling the input variables to pre-specified levels and thereby defining setpoints on process operating conditions rather than setpoints on product and contaminant levels. This may be classified as simple automation. In the eyes of the regulatory authorities definition of a process is seen to be equivalent to defining the product quality. However with variable feed material, process additives and equipment performance it is often difficult to define accurately the process input conditions hence product quality will invariably fluctuate. Process control can be implemented to maintain product quality within pre-specified bands.

In the future process control can be seen to regulate input variables to refine the operating conditions still keeping within regulatory specified settings in order to reduce for example the number of failed runs or to fine tune a well known system. Control to a specific product quality may be introduced if a rapid chemical definition of the product becomes available. The control engineer will have to prove to the regulatory authorities that by implementing a control configuration improvement to a product quality in terms of reproducibility will be the outcome. Control to optimise a process will have an impact on existing processes where reducing costs is vital for economic sustainability.

Chapter 10

References

Asenjo, J., (1991), Cell disruption and removal of insolubles, *Separation processes in biotechnology*, Marcel Dekker, New York, USA, pp.11-20.

Astroem, K., Wittenmark, B., (1984), *Computer-controlled system*, Prentice Hall, New Jersey, USA, pp.416-430.

Beebe, K., Kowalski, B., (1987), An introduction to multivariate calibration and analysis, *Anal. Chem.,* **59**, pp.1007A-1017A.

Bergmeyer, H., Gabl, M., Walter, H., (1933), Methods of enzyme analysis, ed. Bergmeyer H., Vol 2, Verlag Chemie, Weinheim.

Bonnerjea, J., Oh, S., Hoare, M., Dunnill, P., (1986). Protein purification: the right step at the right time, *Bio/Technology,* **4**, pp.954-958.

Bracewell, D., Gill, A., Hoare, M., Love, P., Maule, C., (1998), An optical biosensor for real-time chromatography monitoring: Breakthrough determination, *Biosensors and Bioelectronics*, **13**, pp.847-853.

Bradford M., (1976), A rapid and sensitive metho for the quantitation of microgram quantities of protein utilizing the principal of protein-dye binding, *Anal. Biochem.*, **72**, pp.248-254.

Brimmer, P., Hall, J., (1993), Determination of nutrient levels in a bioprocess using near-infrared spectroscopy, *Canadian journal of spectroscopy*, **38**, pp.155-162.

Bulmer, M., Hoare, M., Hinton, J., (1994), Optimisation of selective flocculation for processing of cell homogenates, *The 1994 IChemE Research Event*, pp.141-143.

Catlin, D., (1981), Estimation, control and the discrete Kalman filter, Applied mathematical sciences, New York, USA.

Cavinato, A., Mayes, D., Ge, Z., Callis, J., (1990), Noninvasive method for monitoring ethanol in fermentation processes using fiber-optic near infrared spectroscopy, *American Chemical Society*, **62**, pp.1977-1982.

Chang, Y., Chase, H., (1994), Protein purification in expanded beds using pupose designed ion exchangers, *The 1994 IChemE Research Event.*

Chard, S., Holwill, I., Hoare, M., (1994), On-line rapid monitoring for the enhanced development and control of protein separation processes, *Presented at third Asia-Pacific Biochem. Eng. Conference.*

Charm Work, On-line manual version 1.1, Process analysis and automation limited, 1995, Hampshire, UK.

Cheruy, A., (1997), Software sensors in bioprocess engineering, *J. Biotech.,* **52**, pp.193-199.

Christensen, L., Marder, J., (1996), Semi-on-line analysis for fast and precise monitoring of bioreactor processes, *Biotechnol. Bioeng.*, **52**, pp.237-247.

Clarkson, A., I., Lefevre, P., Tichener-Hooker, N., J., (1993), A study of process interactions between cell disruption and debris clarification stages in the recovery of yeast intracellular products, *Biotechnol. Prog.*, **9**, pp.462-467.

Cohn, E., (1925), *Physiol. Rev.*, 5, pp.349-437.

Cordes R., Sims, W., Glatz, C., (1990), Precipitation of nucleic acid with PEI, *Biotechnol. Prog.,* **6**, pp.283-285.

Cowan, C., Optimum estimation techniques, *Adaptive filters*, ed. P. M. Grant, Prentice-Hall, Englewood Cliffs, New Jersey, USA, pp.21-28.

Creighton, T., (1984), Proteins structure and molecular properties, W. H. Freeman, New York, USA.

Du Plessis, R., (1967), Poor man's explanation of Kalman filtering, *North American viation, Inc. Autonetics Division.*

Dunnill, P., (1983) Trends in downstream processing of proteins and enzymes, *Process. Biochem.*, **18** (5), pp.9-13.

Ehresmann, B., Imbault, P., Weil, J., (1971), Spectrophotometric determination of protein concentration in cell extracts containing tRNA's and rRNA's, *Analytical Biochemistry*, **54**, pp.454-463.

Filipini, C., Sonnlitnber, Fiechter, A., (1992), Intelligent analytical subsystems for on-line control and monitoring of bioprocesses, *Anal. Chemica Actca*, **265**, pp.63-69.

Fish, N.M., Lilly, M. D., (1984) The interaction between fermentation and protein recovery, B*io/Technology,* **2** (7), pp.623-627.

Fisher, R., Glatz, C., (1987), Polyelectrolyte precipitation of proteins: 1. The effects of reactor conditions, *Biotech. Bioeng.*, **32**, pp.777-785.

Frej, A., Johansson, S., Leijton, P., Expanded bed adsorption at production scale: *Scale-up verification, process example and sanitization of column and adsorbent, Bioproc. Eng.,* **16**, pp.57-63.

Geladi, P., Komalski, B., (1986), Partial least-squares regression - a tutorial*, Anal., Chim. Acta.,* **185**, pp.1-17.

Gelb, A., (1974), Applied optimal estimation, M.I.T. Press, Cambridge, USA.

Glassey, J., Ignova, M., Ward, A., Montague, G., Morris, J., (1997), Bioprocess supervision; neural networks and knowledge based systems, *J. Biotech.*, **52**, pp.201-205.

Gregory, J., Nelson, D., (1986), Monitoring of aggregates in flowing suspensions, C*olloids and Surfaces*, **18**, pp.175-177.

Griffiths, D., Hall, G., (1993), Biosensors, what real progress is being made?, *TIBTECH.*, **11**, pp.122-130.

Gritis, D., Tichener-Hooker, N., (1989), Biochemical Process Simulation, The 1989 *IChemE Research Event*, **114**, pp. 69-77.

Habib, G., Holwill, I., Hoare, M., (1997), Rapid piloting of a selective flocculation process for product purification, *J. Biotech.*, **59**, pp.91-101.

Heijden, R., Hellinga, C., Lyben, K., (1989), State estimators (obsevers) for the on-line estimation of non-measurable process variables, *TIBTECH*, **7**, pp.205-209.

Hetherington, P., Follows, M., Dunnill, P., Lilly, M., (1971), Release of protein from Baker's yeast by disruption in an industrial homogeniser, *Trans. Instn. Chem. Engrs.*, **49**, pp.142-148.

Holwill, I., Bulmer, M., Hoare, M., (1998), Fault detection during homogenisation, *in press.*

Holwill, I., Chard, S., Flanagan, M., Hoare, M., (1997), A Kalman filter algorithm and monitoring apparatus for at-line control of fractional protein precipitation, *Biotech. Bioeng.*, **53**, pp.58-70.

Horn, D., (1980), Polyethylenimine - physicochemical properties and applications, *Polymeric Amines an Ammonium Salts*, ed. Goethals, E., Pergamon Press, Oxford, pp.333-355.

Huang, C., Chen, G., (1996), Use of the fiber-optical monitor in evaluating the state of flocculation, *Wat. Res.*, **30**, pp.2723-2727.

Huddleston, J., Veide, A., Kohler, K., Flanagan, J., Matshushima, T., and Saito, Y., (1986), The molecular-basis of partitioning in aqueous 2-phase systems, *TIBTECH*, **9**, pp.381-386.

Jendrisak, J., (1987), The use of polyethyleneimne in protein purification, *Protein Purification: Micro to Macro*, ed. Burgess, R., Proceedings of a Cetus-UCLA symposium held at Frisco, Colorado, pp.75-97.

Kalb, V., Bernlohr, W., (1977), A New spectrophotometric assay for protein in cell extracts, *Analytical Biochemistry*, **82**, pp.362-371.

Kalman, R., (1960), A new approach to linear filtering and prediction problems, *Journal of Basic Eng.*, pp.35-45.

Kelley, B., Hatton, T., (1991), The Fermentation / Downstream Processing Interface, *Bioseparation*, **1**, pp.333-349.

Konstantinov, K., Chuppa, S., Sajan, E., Tsai, Y., Yoon, S., Golini, F., (1986), Real-time biomass-concentration monitoring in animal-cell cultures, *TIBTECH*, **12**, pp.324-333.

Lim, H., Lee, K., (1991), Control of bioreactor systems, *Biotechnology*, Measuring, modelling and control, ed. Rhelm, H., Reed, G., VCH, Cambridge, UK.

Lindquist, G., Stratton, R., (1975), The role of polyelectrolyte charge density and molecular weight on adsorption and flocculation of colloidal silica with polyethyleneimine, *Journal of Colloid and interface Science*, **55**, pp.45-59.

Locher, G., Sonnleiter, B., Fiechter, A., (1992), On-line measurement in biotechnology: techniques, *Process Biotechnology*, **25**, pp.23-53.

Lubbert, A., Simutis, R., (1994), Using measurement data in bioprocess modelling and control, *TIBTECH*, **12**, pp.304-311.

Macaloney, G., Draper, I., Preston, J., Anderson, K., Rollins, M., Thompson, B., Hall, J., McNeil, B., (1996), At-line control and fault analysis in an industrial high cell density Escherichia coli fermentation, using NIR spectroscopy, *Trans IChemE,* **74**, pp.212-219.

Martens, H., Naes, T., (1984), Multivariate calibration. Concepts and distinctions, *Trends in Anal. Chem.,* **3**, 204-210.

Mattiasson, B., Borrebaeck, K., (1978), Non-equilibrium, isoelectric enzyme immunoassay of insulin using reversible immobilised antibodies, *Enzyme labelled immonoassay of hormones and drugs*, ed. S., Pal Water de Gryuter and Co., Berlin, Germany, pp.91-105.

Mattiasson, B., Haakanson, H., (1991), Measurement and control in down-stream processing, *Measurement and Control in Bioprocesses*, ed. K. Carr-Brian, Elsevier, Applied Science, pp.221-249.

Merbel, N., Linegeman, U., Brinkman, U., (1996), Sampling and analytical strategies in 0n-line bioprocess monitoring and control, *J. Chrom.,* **725**, pp.13-27.

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

Montague, G., Morris, A., Tham, M., (1992), Enhancing bioprocess operability with generic software sensors, *J. Biotech.,* **25**, pp.183-201.

Moudgil, B., Behl, S., (1991), A model of the selective flocculation process, *Journal of Colloid and interface Science*, **146**, pp.1-8.

Myers, M., Luecke, R., (1991), Process control applications of an extended Kalman filter algorithm, *Computers chem. Engng*, **15**, pp.853-857.

Naveh, D., (1990), *BioPharm*, May, 28.

Niktari, M., Chard, S., Richardson, P., Hoare, M., (1990), The monitoring and control of protein purification recovery processes, *Separation for Biotechnology 2*, ed. Pyle DL, pp622-630.

Nilsson, M., Haakanson, H., Mattiasson, B, (1992b), Process monitoring by flow-injection immuniassay, *J. Chrom.,* **596**, pp.383-389.

Nilsson, M., Mattiasson, G., Mattiasson, B., (1992a), Automated immunochemical binding assay (flow-ELISA) based on repeated use of an antibody column placed in a flow-injection system, *J. Biotech.,* **31**, pp.381-394.

Noui, L., Habib, G., Hill, J., Yeung, K., Keay, P., Hoare, M., (1998), Rapid monitoring of contaminants in a flocculation process using UV-VIS spectroscopy, *in press.*

Ogez, J.R., Hodgdon, J.C., Beal, M. P., Builder, S., (1989) Downstream process of proteins: Recent advances, *Biotech. Adv.,* **7**, pp.467-488.

Olsson, L., Schulze, U., Nielsen, J., (1998), On-line bioprocess monitoring – an academic discipline or an industrial tool?, *Trends in anal. Chem.,* **17**, pp.88-95.

Paliwal, S., Nadler, T., Regnier, F., (1993), Rapid process monitoring in biotechnology, *TIBTECH.*, **11**, pp.95-101.

Pfund, N., (1987), Environmental service laboratories, *Environmental science & technology*, **21**, pp.951-961.

Plugge, W., Van Der Vlies, C., (1992), The use of near infrared spectroscopy in the quality control laboratory of the pharmaceutical industry, *Journal of Pharm. and Biomed. Analysis*, **10**, pp.797-803.

Ramamurthi, Y., Sitstu, P., Bequette, B., (1993), Control relevant dynamic data reconciliation and parameter estimation, *Computer Chem., Eng.,* **17**, pp.41-59.

Ransohoff, T., Murphy, M., Levine, H., (1990), Automation of biopharmaceutical purification processes, *BioPharm*, pp.20-26.

Reiner, T., Van Der Heijden, Hellinga, Luyben, Honderd, (1989), State estimators for the on-line estimation of non-measurable process variables, *TIBTECH*, **15**, pp.205-209.

Rhelm, H., Reed, G., (1991), Monitoring, modelling and control, *Biotechnology*, 2nd edition, VCH, Cambridge, UK.

Richardson, P., Hoare, M., Dunnill, P., (1990), A new biochemical engineering approach to the fractional precipitation of proteins, *Biotech. Bioeng.*, **36**, pp.354-366.

Richardson, P., Molloy, J., Ravenhall, R., Hoare, M., Dunnill, P., (1996), High speed centrifugal separation for rapid on-line sample clarification in biotechnology, *J. Biotech.*, **49**, pp.111-118.

Ritzka, A., Sosnitza, P., Ulber, R., Scheper, T., (1997), Fermentation monitoring and process control, *Biochem. Eng.*, **12**, pp.106-164.

Royce, P., (1993), A discussion of recent development in fermentation monitoring and control from a practical perspective, *Critical Reviews in Biotechnology*, **13**, pp.117-149.

Ruzicka, J., Hansen, E., (1988), Flow injection analysis, 2nd ed., Wiley, New York, USA.

Ruzicka, J., Marshall, G., (1990), Sequential injection: A new concept for chemical sensors, process analysis and laboratory assay, *Anal. Chem. Acta*, **237**, pp.329-343.

Sadana, A., (1998), Bioseparation of proteins, *Separation science and technology*, Academic press, California, USA.

Salt, D., Hay, S., Thomas, O., Hoare, M., Dunnill, (1995), Selective flocculation of cellular contaminants from soluble proteins using polyethyleimine: A study of several organisms and polymer molecular weights, *Enzyme and Microbial Tech.*, **17**, pp.107-113.

Scawan, M., Atkinson, A., Derbyshire, J., (1980), *Protein biochemistry*, ed. Grant, R., Applied Science Publications, London, UK, pp.281-325.

Scawan, M., Hammond, P., (1989), Fractionation techniques in process biotechnology, *J. Chem. Techn. Biotech.,* **46**, pp.85-103.

Scheper, T., Hilmer, J., Lammers, F., Muller, C., Reinecke, M., (1996), Biosensors in boprocess monitoring, *J. Chrom.,* **725**, pp.3-12.

Schugerl, K., (1993), Which requirements do flow injection analyszer / biosensor systems have to meet for controlling the bioprocess?, *J. Biotech.,* **31**, pp.241-256.

Schugerl, K., Hitzmann, B., Jurgens, H., Kullick, T., Ulber, R., (1996), Challenges in integrating biosensors and FIA for on-line monitoring and control, *TIBTECH*, **14**, pp.21-31.

Scopes, R., (1974), Measurement of protein by spectrophotometry at 205 nm, *Analytical Biochemistry*, **59**, pp.277-282.

Seber, G., Wild, C., (1989), Nonlinear regression, John Wiley and Sons, New York, USA.

Siddiqi, S., Clarkson, A., Kesharavz, M., Tichener-Hooker, N., (1991), Modelling of process interactions in biochemical processes, The 1991 *IChemE Reseach Event*, pp.155-158.

Skidmore, G., Horstmann, B., Chase, H., (1989), Modelling single-component protein adsorption to the cation exchanger Sepharose® FF, *J. Chrom.*, **498**, pp.113-128.

Smith, M., (1997), An evaluation of expanded bed adsorption for the recovery of proteins from crude feedstocks, *Thesis*, University College London.

Sonnleitner, B., (1996), New concepts for quantitative bioprocess research and development, *Advances in Biochem. Biotech.,* **54**, pp155-188.

Sonnleitner, B., (1997), Bioprocess automation and bioprocess design, *J. Biotech.*, **52**, pp.175-179.

Sonnleitner, B., Locher, G., Fiechter, A., (1991), Automatic bioprocess control. 1. A general concept, *J. Biotech.*, **19**, pp.1-18.

Sonsnitza, P., Farooqui_M, Saleemuddin_M, Ulber_R, Scheper (1998), Application of reversible immobilization techniques for biosensors, *Anal. Chim. Atca*, **368**, pp.197-203

Swerin, A., Odberg, L., Waagberg, L., (1996), An extended model for the estimation of flocculation efficiency factors in multicomponent flocculant systems, *colloids and surfaces*, **113**, pp.25-38.

Takashashi K., Taniguchi, S., (1989), Automated micro stopped-flow / continuous-flow apparatus for serial measurement of enzyme reactions and its application as a real-time analyser for column chromatography, *Anal. Chimica Acta*, **220**, pp.13-21.

Thornhill, N., Royce, N., (1991), Modelling fermenters for control, *Measurement and Control in Bioprocesses*, ed. Carr-Brian, Elsevier, Applied Science, pp.221-249.

Tranford, C., (1997), How protein chemists learned about hydrophobic factor, *Protein science*, **6**, pp.1358-1366.

Turner, C., Thornhill, N., Fish, N., (1993), A novel method for the on-line analysis of fermentation broth using a sampling device, microcentrifuge and HPLC, *Biotechnol., Tech.*, **7**, pp.19-24.

Vallino, J., Stephanopoulos, G., (1987), Intelligent sensors in biotechnology – applications for the monitoring of fermentations and cellular-metabolism, *Annals of the New York academy of sciences*, **506**, pp.415-430.

van Brunt, J., (1985), Fundamentals of biochemical engineering, *Biotechnology*, eds. H., Rhem and G., Reed, 3, Cambridge, UK.

Virkar, P., Narendranathan, T., Hoare, M., Dunnill, P., (1981), Studies of the effects of shear on globular-proteins – extension to high shear and to pumps, *Biotech. Bioeng.*, **23**, pp.425-429.

Warburg, O., Christian, W., (1942), Isolierung und Kristallisation des Garungsferments Enolase, *Biochem.Z.*, **310**, 384-421.

Weeseling, J., (1994), Third international conference on separations for biotechnology, society of chemical industry, university of Reading, Reading, UK, September 12-15.

Wheelwright, S., M., (1989), The design of downstream processes for large-scale protein purification, *J. Biotech.*, **11**, pp.89-102.

Yeung, K., Hoare, M., Thornhill, N., Williams, T., Jeetendra, V., (1998), Near infrared spectroscopy technique for bioprocess monitoring and control, *in press.*

Yoshimura, K., Monoshi, R., Kiyozuri, Y., Soeda, T., (1980), Identification of unknown parameters by a modified EKF, *Int., J., System, Sci.,* **11**, pp.97-105.

Yu, K., Philips, J., (1992), The us eof infrared spectroscopic techniques in monitoring and controlling bioreactors, *IFAC Modeling and control of Biotechnical processes*, Colorado, USA,

Zhou, Y., Holwill, I., Titchener-Hooker, (1997), A study of the use of computer simulations for the design of integrated downstream processes, *Bioproc. Eng.,* **16**, pp.367-374.

Appendix A

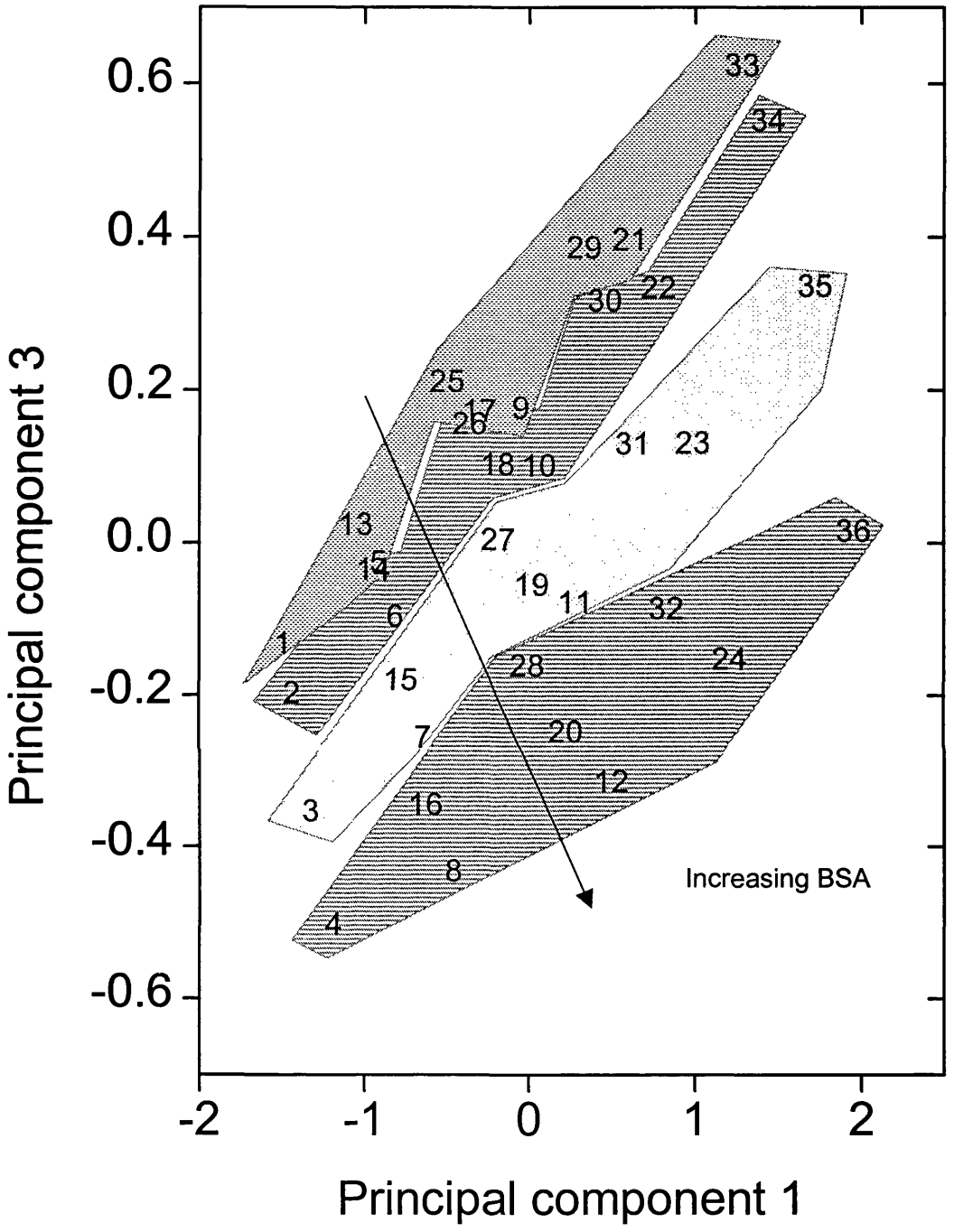


Figure A.1. Scores plot of UV-VIS spectral scan calibration data in terms of principle components 1 and 3. The numbers on the plot refer to the spectral scan taken of the particular sample number in the factorial experiment (table 4.1). Sample clusters (samples which have a geometrically close proximity in a scores plot) have been circled. Trends due to BSA variations are indicated by the arrows on the plot.

Appendix B

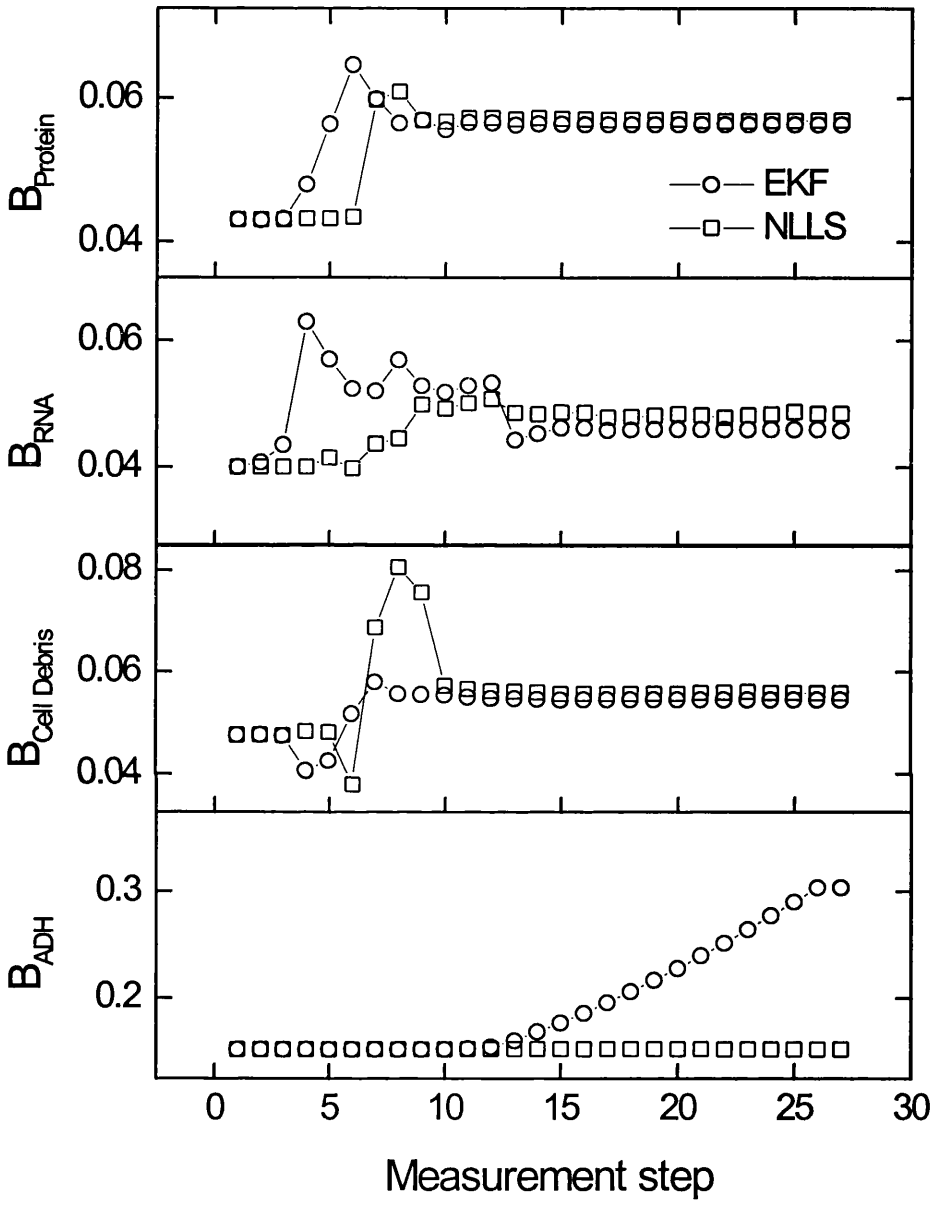


Figure B.1. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run2. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

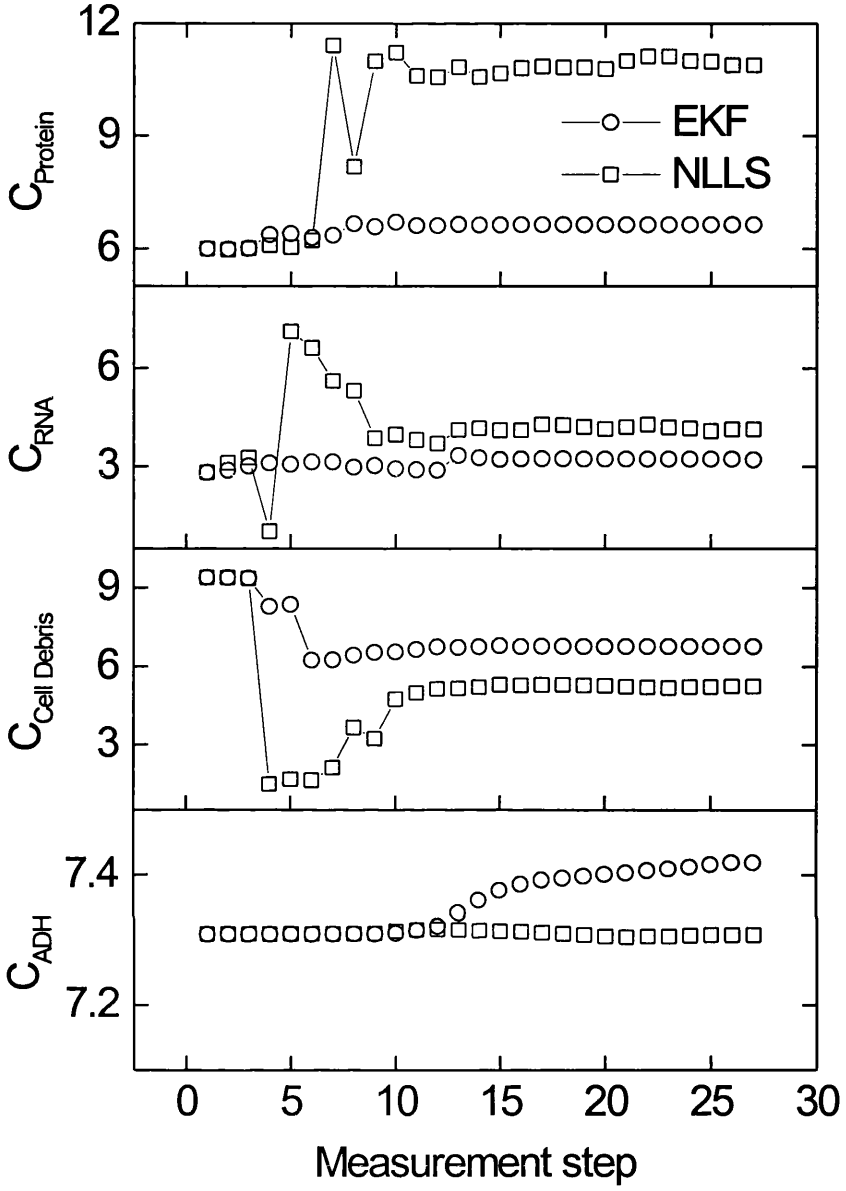


Figure B.2. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run2. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

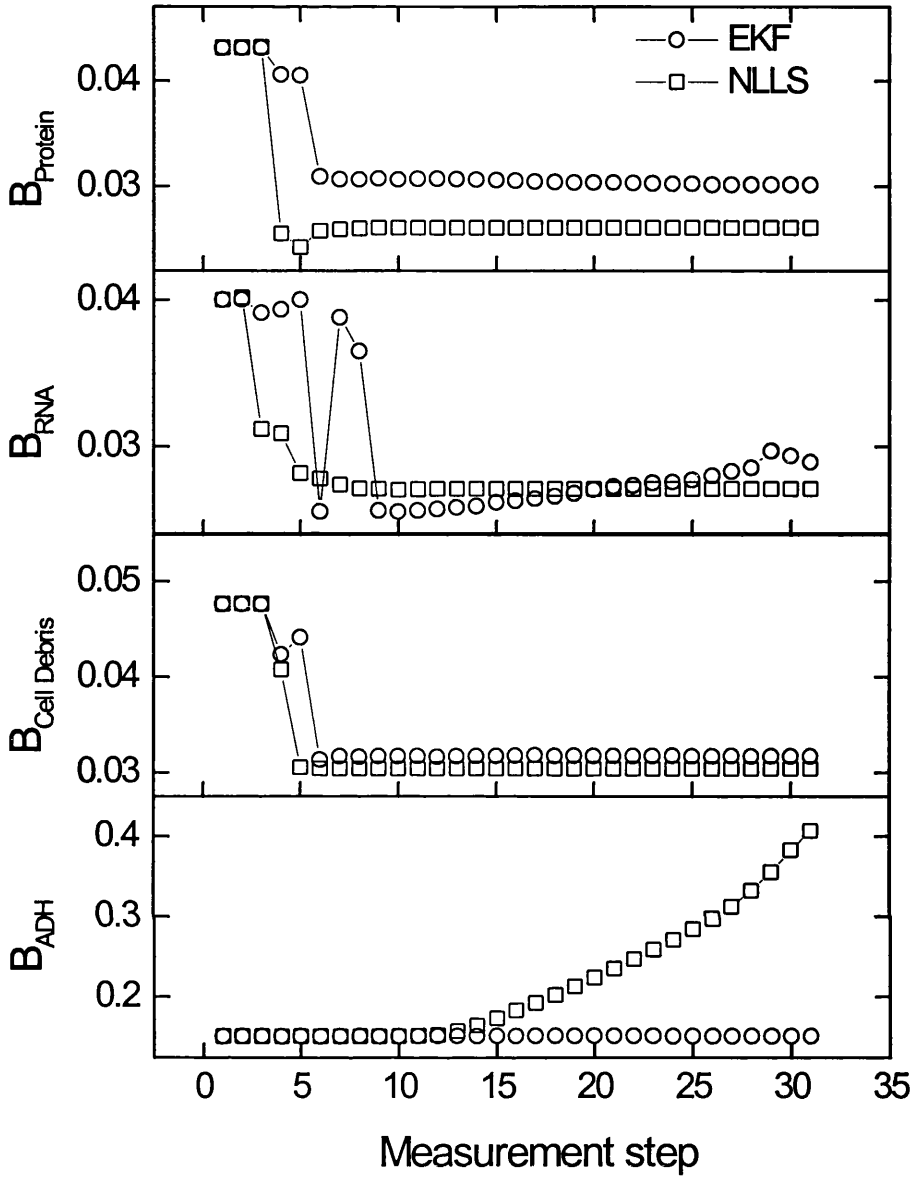


Figure B.3. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run3. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

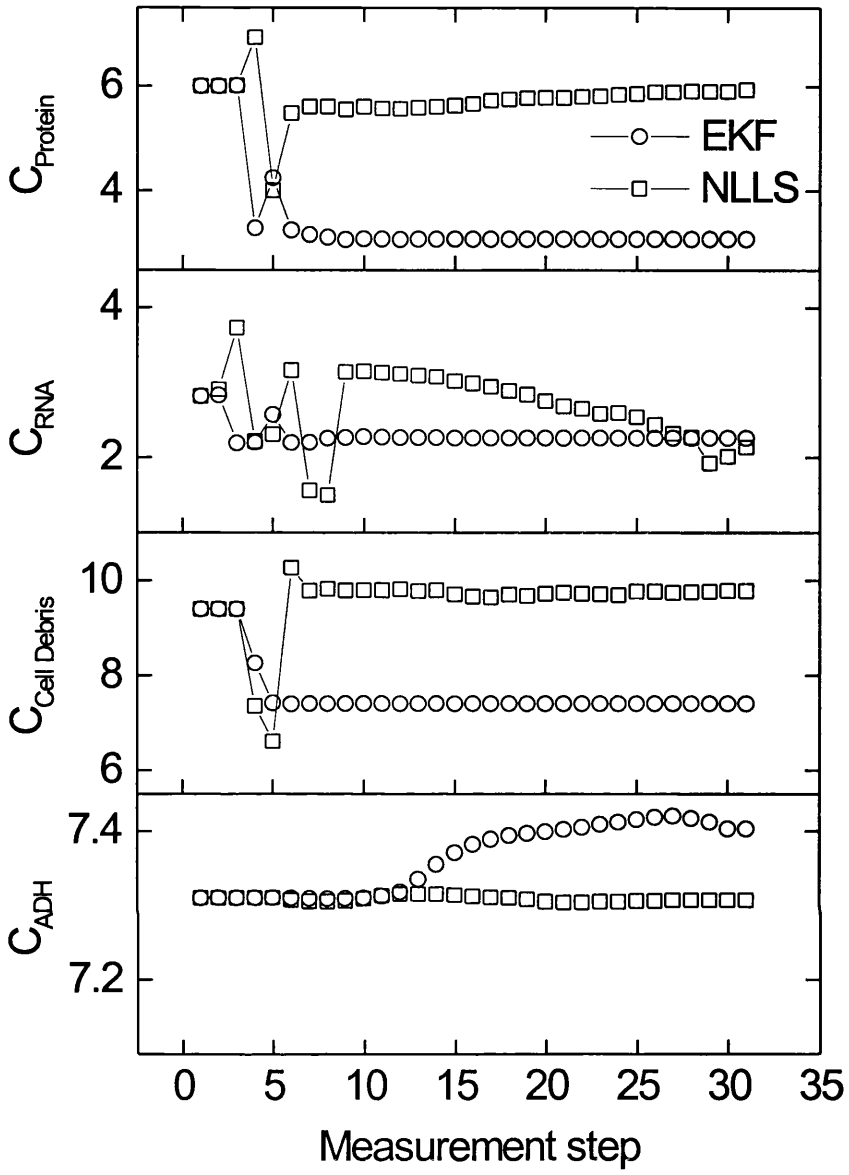


Figure B.4. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run3. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

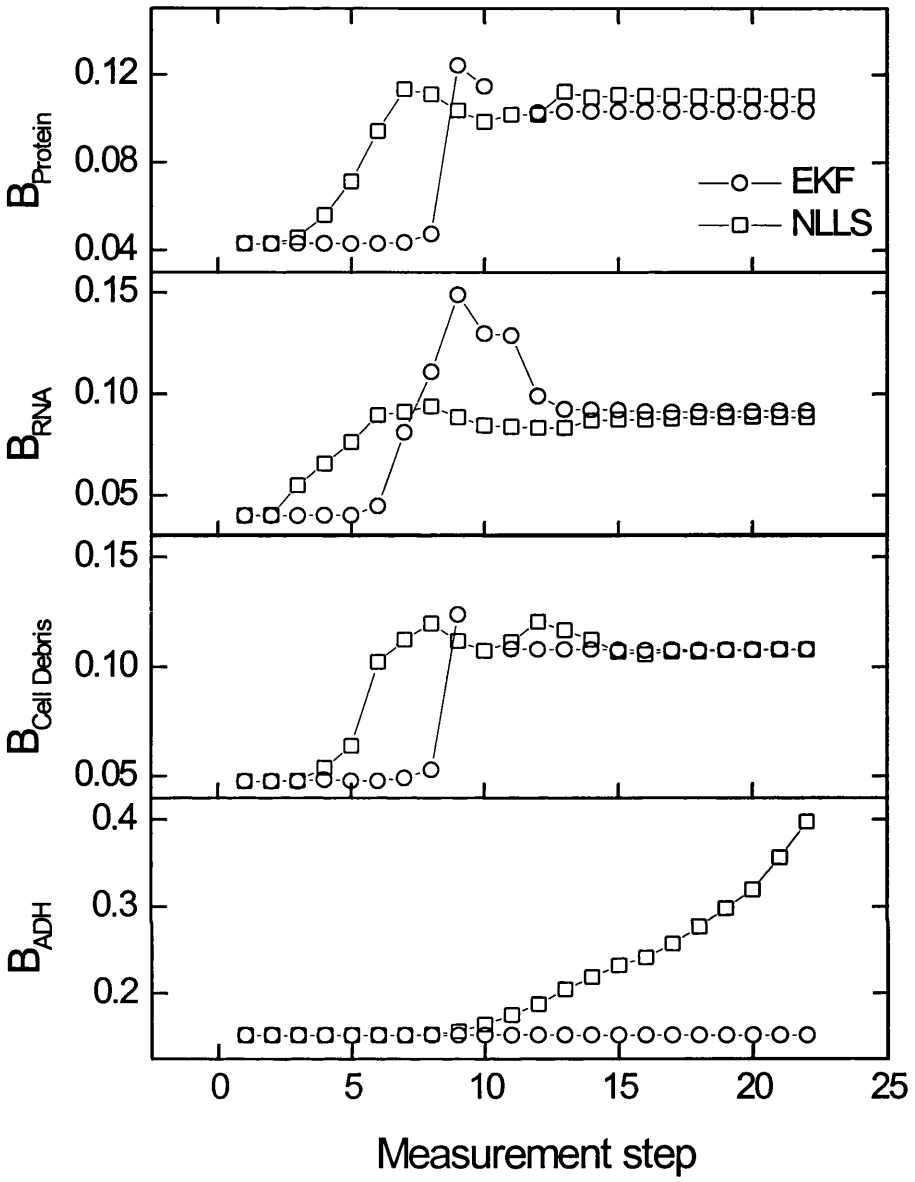


Figure B.5. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run4. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

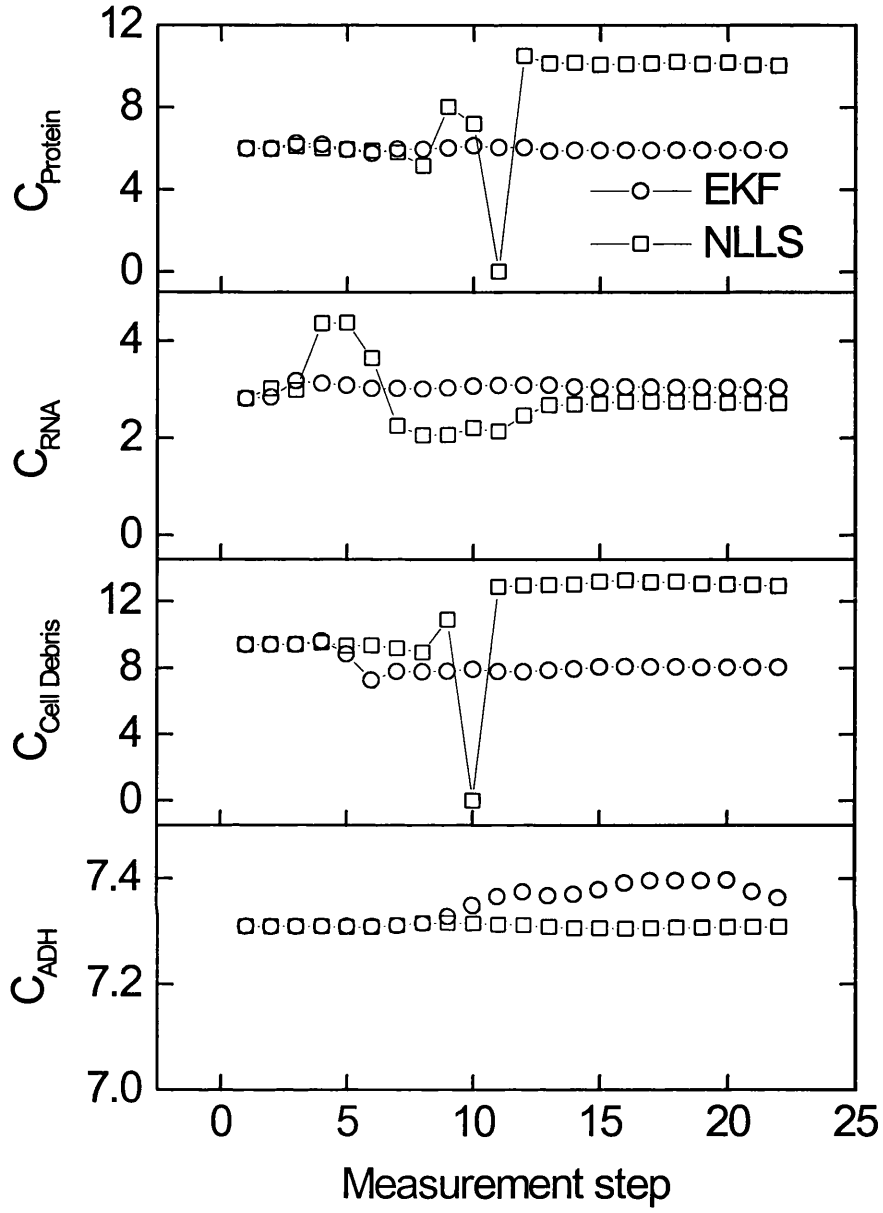


Figure B.6. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run4. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

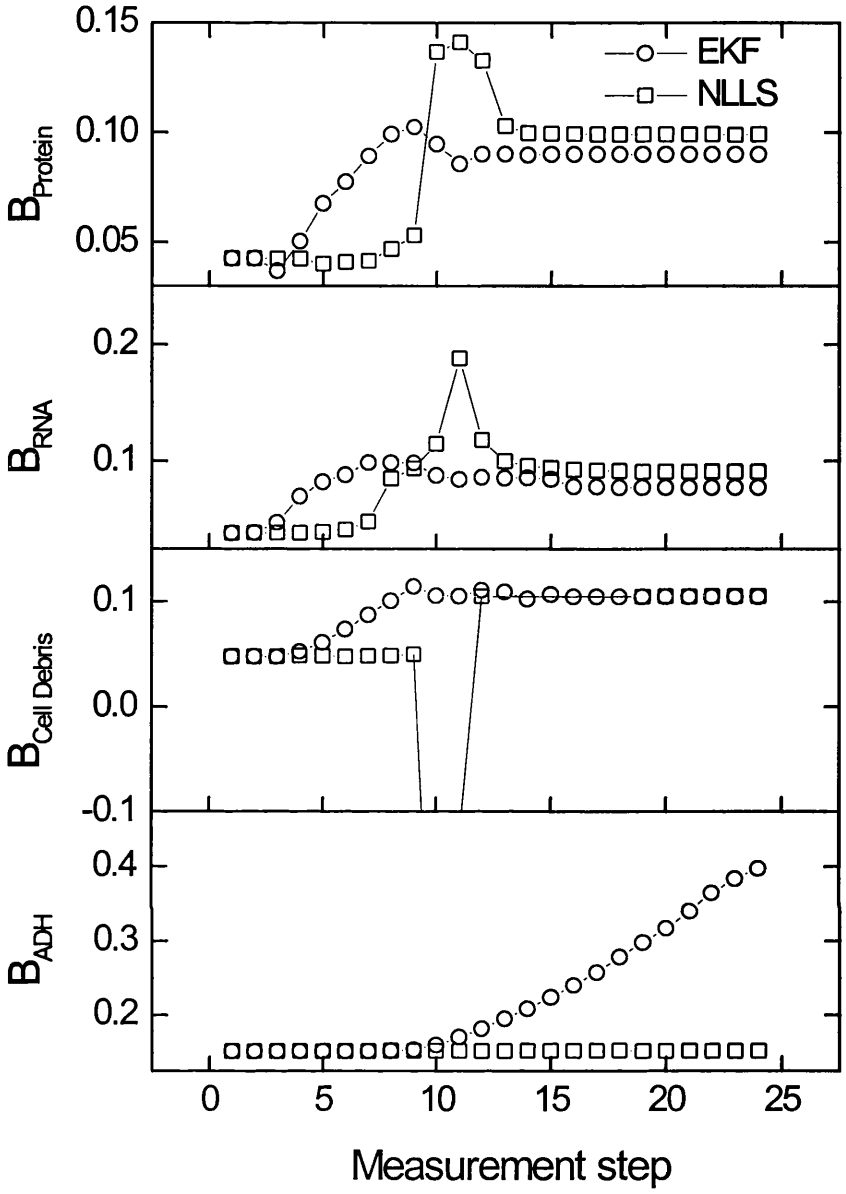


Figure B.7. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run5. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

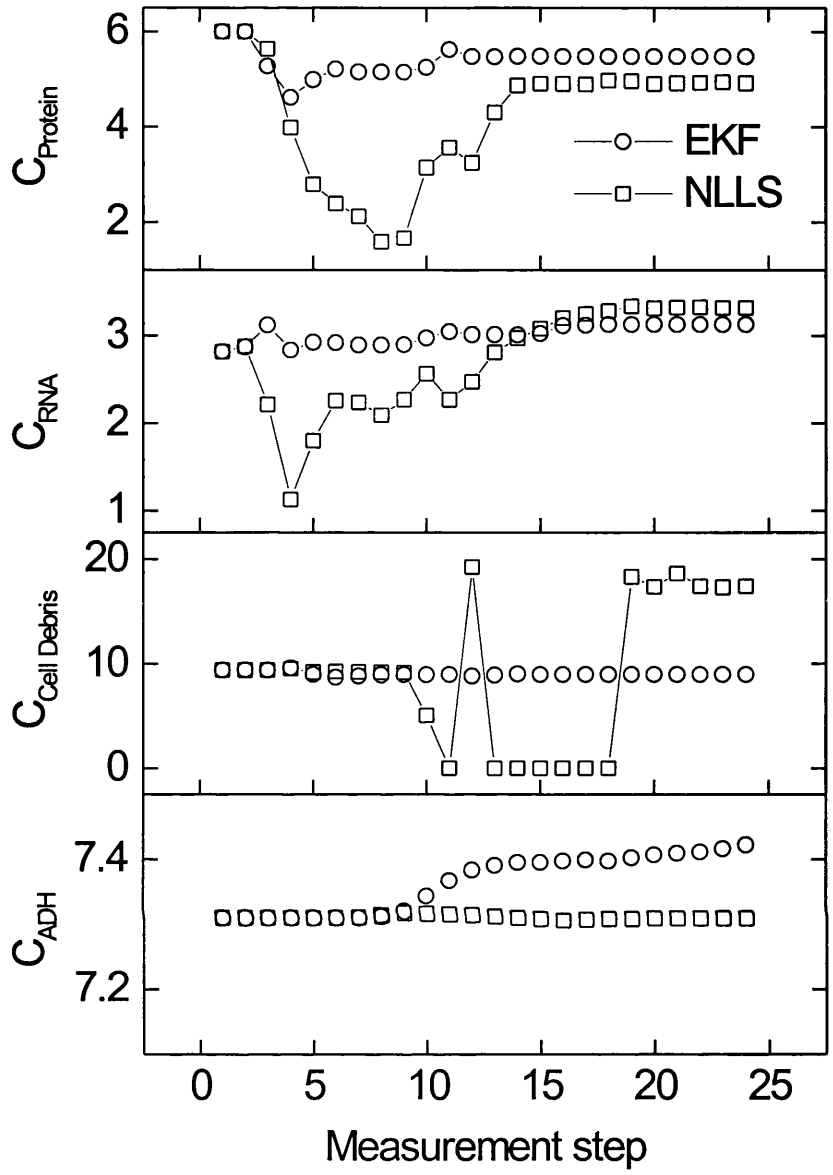


Figure B.8. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run5. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

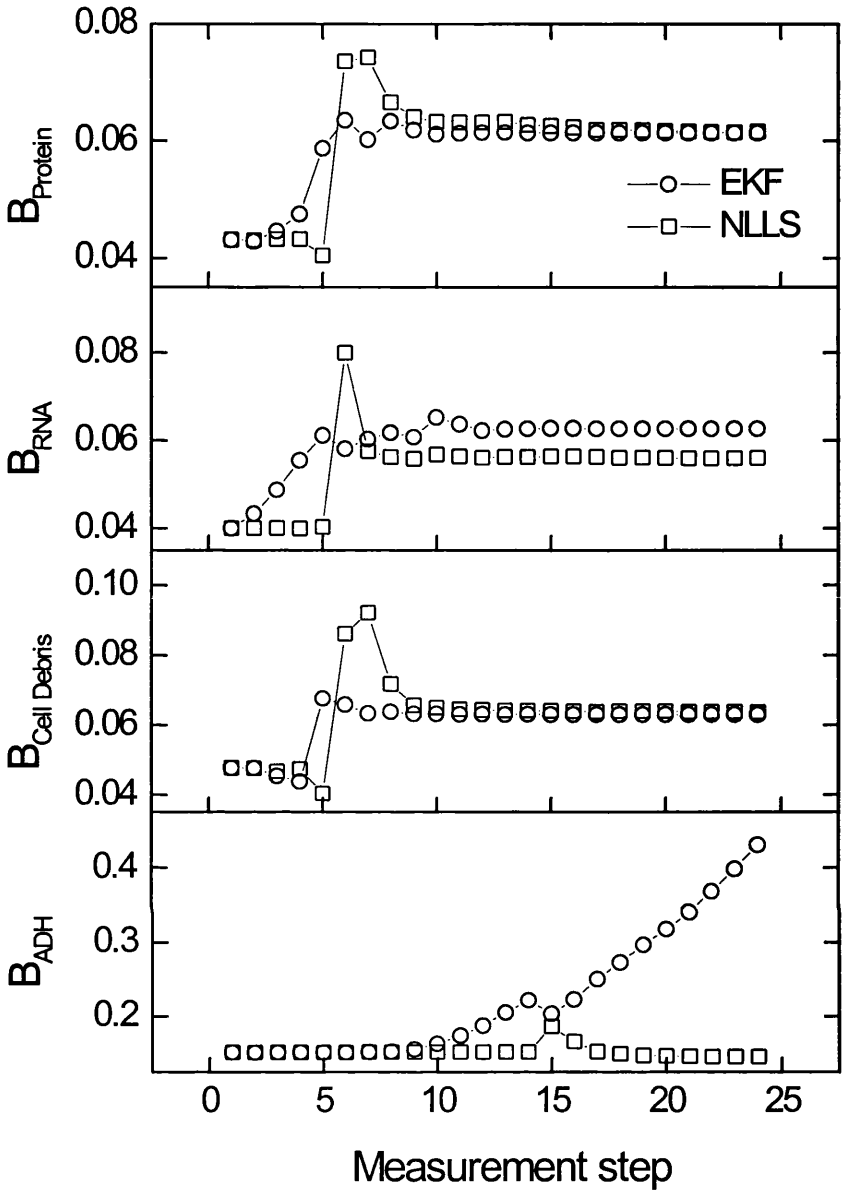


Figure B.9. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run6. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

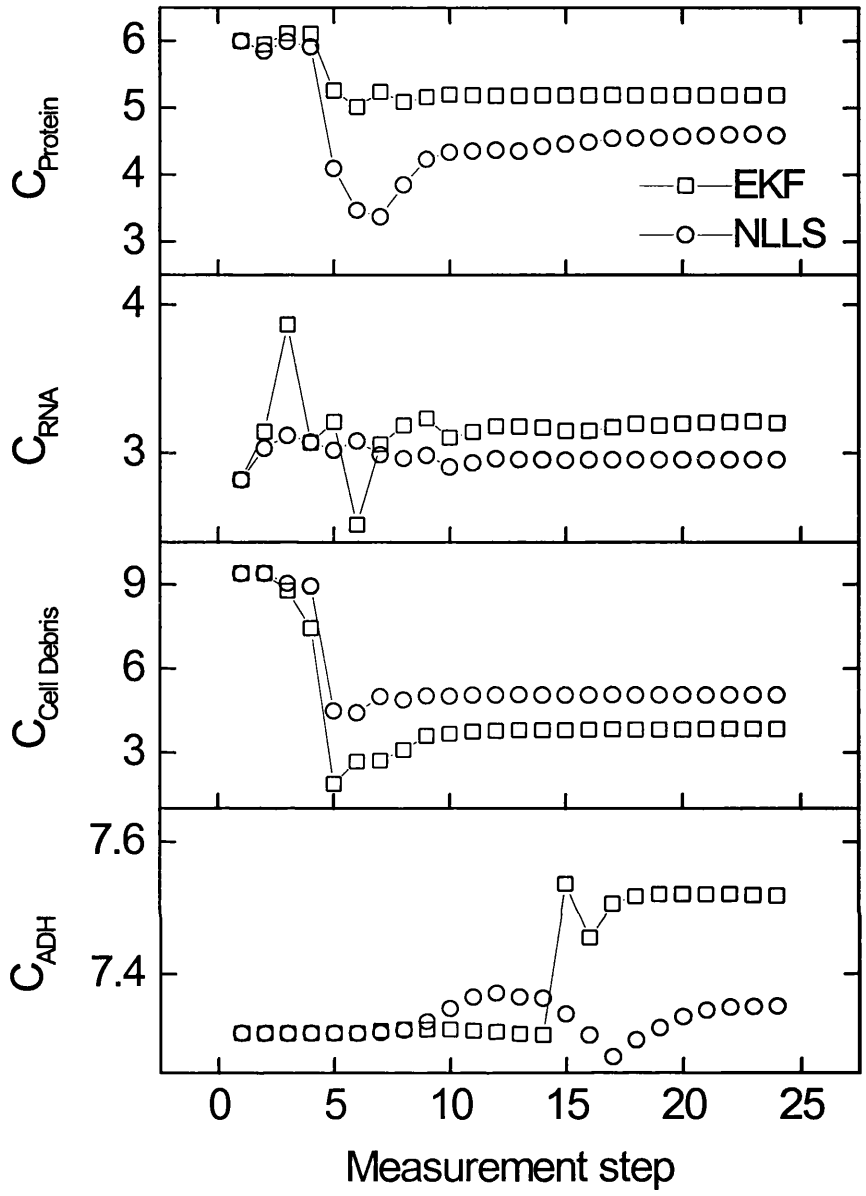


Figure B.10. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run6. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

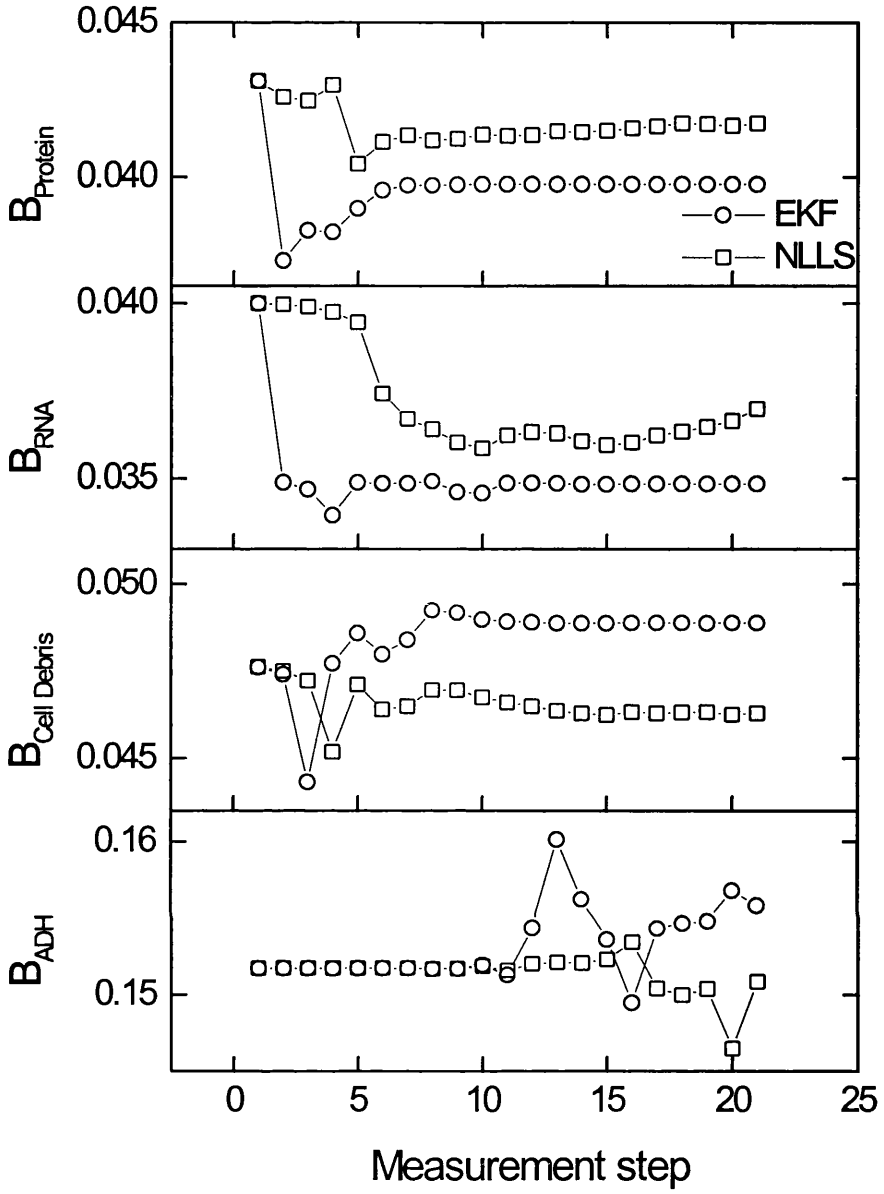


Figure B.11. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run7. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

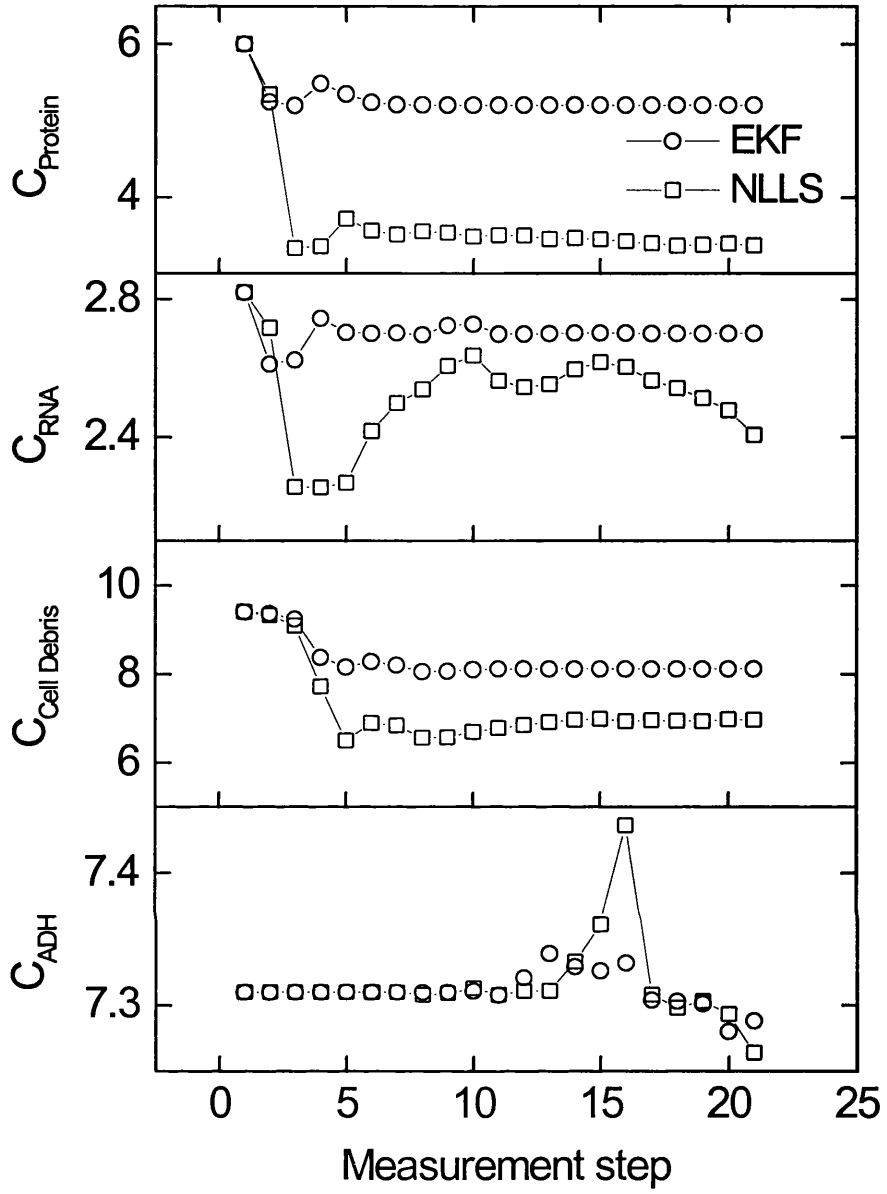


Figure B.12. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run7. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

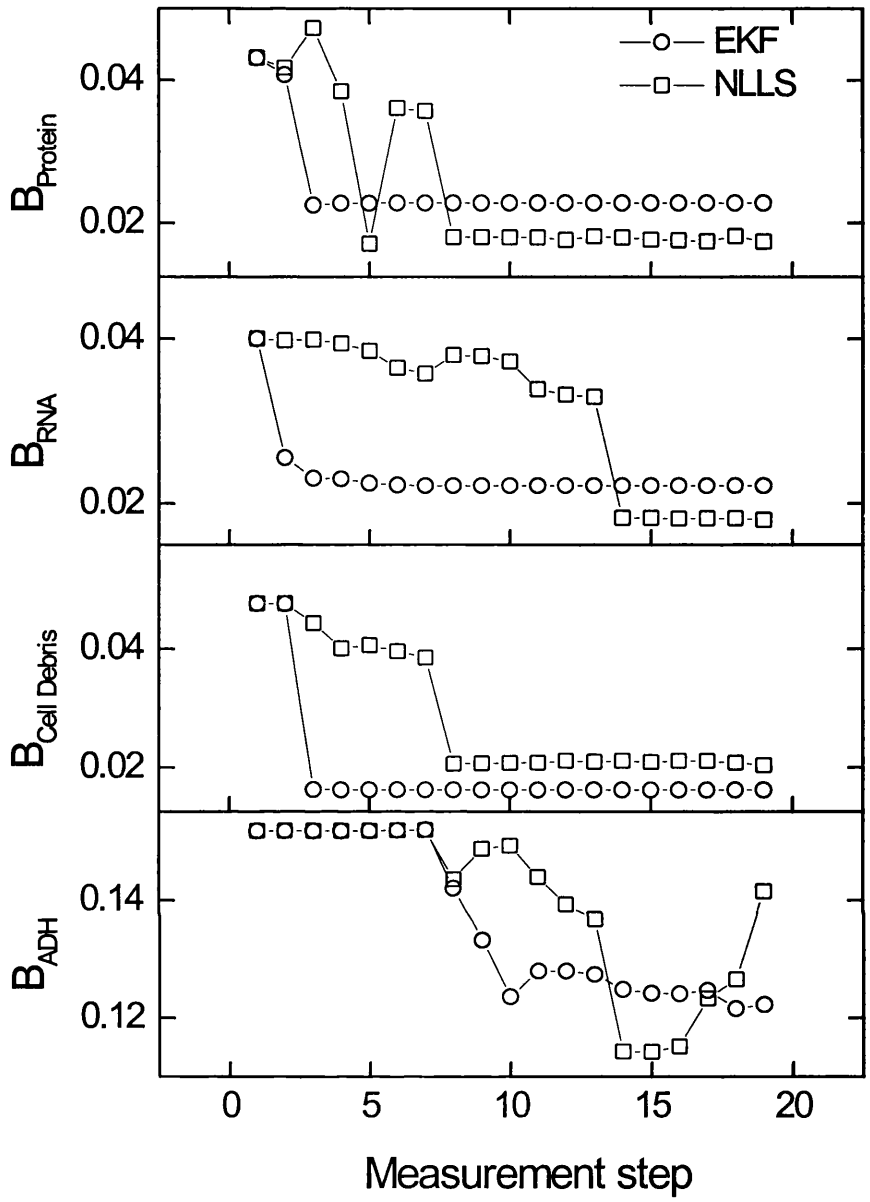


Figure B.13. *Estimation of model parameter B for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run8. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

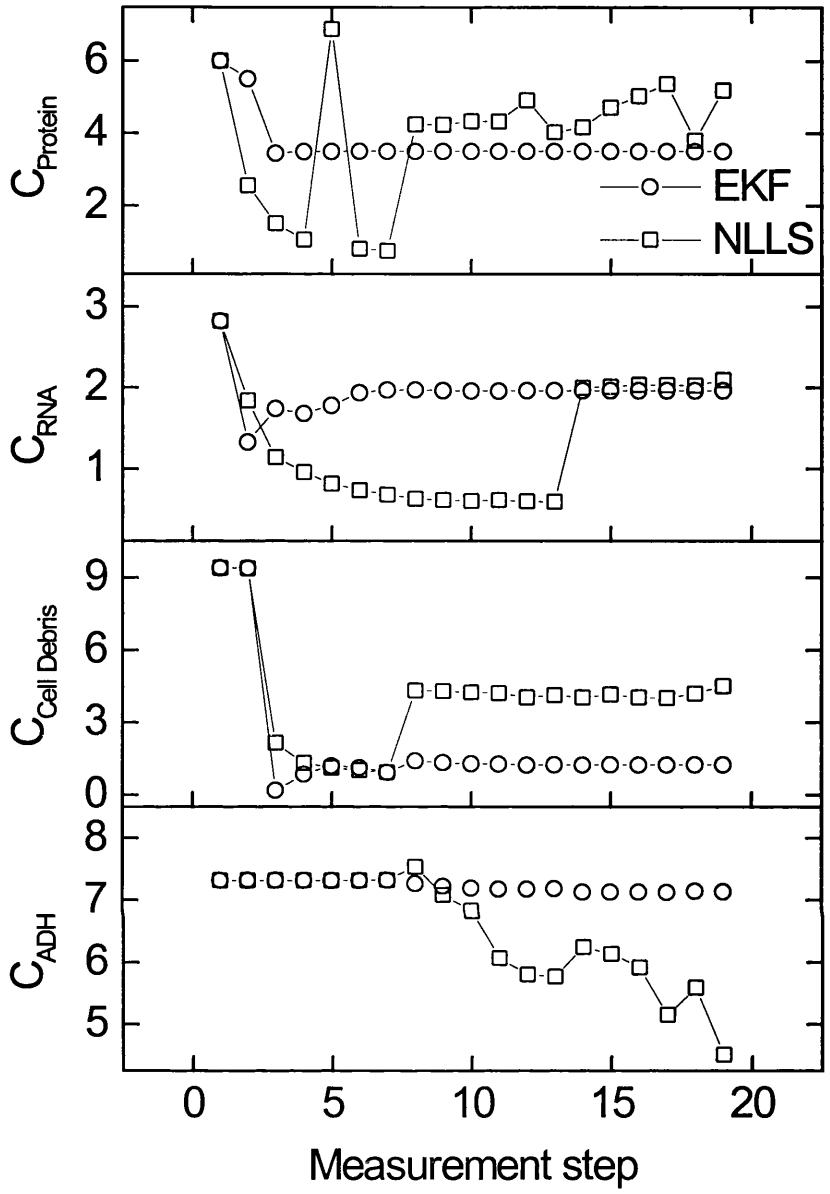


Figure B.14. *Estimation of model parameter C for RNA, protein, cell debris and ADH flocculation profiles through the extended Kalman filter and Levenberg-Marquardt non-linear least squares techniques for Batch Run8. The model parameter initial guesses, measurement and system noise terms are listed in table 6.1 and batch operating conditions are listed in table 5.2.*

Appendix C

3(1)-Parameter Model (equation 2.7)

$$\frac{dy_{3(1)}}{dx} = 0 \Rightarrow x = a\left(\frac{n-1}{n+1}\right)^{\frac{1}{n}} \qquad \text{(c.1)}$$

where $y_{3(1)}$ is the 3(1)-Parameter model, a and n are model parameters, and x is the measure of flocculant dose.

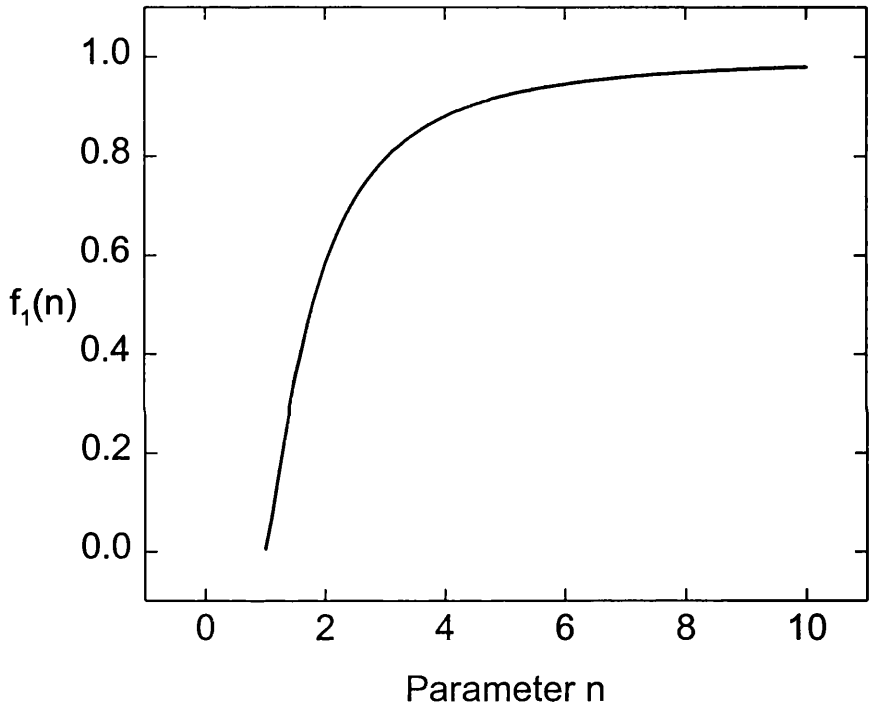


Figure C.1. *Change in function $f_1(n)$ with respect to n. Where $f_1(n)$ is defined as* $f_1(n) = \left(\frac{n-1}{n+1}\right)^{\frac{1}{n}}$. *When* $n \geq 4.5 \Rightarrow f_1(n) \geq 0.9$.

3(2)-Parameter model (equation 2.10)

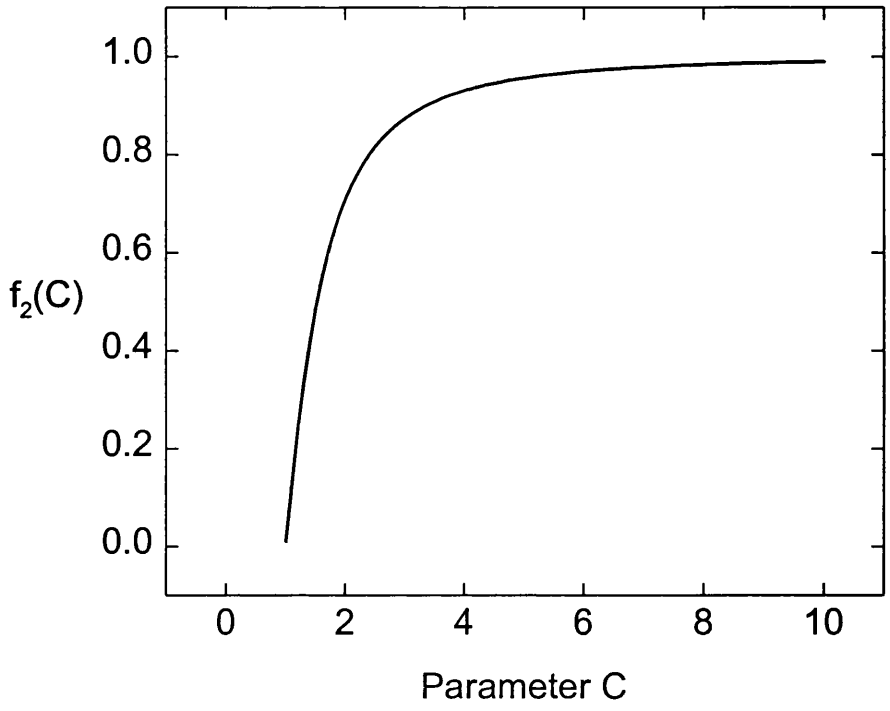


Figure C.2. *Change in function $f_2(C)$ with respect to C. Where $f_2(n)$ is defined as $f_2(C)=\left(\frac{C-1}{C}\right)^{\frac{1}{C}}$. When $C \geq 3.5 \Rightarrow f_2(C) \geq 0.9$.*

3(2)-Parameter model derivatives

$$\frac{dy_{3(2)}}{dA} = \exp\left(-\left(\frac{x}{B}\right)^{C}\right) - 1 \qquad \text{(c.2)}$$

$$\frac{dy_{3(2)}}{dB} = ACB^{-C-1}x^{C}\exp\left(-\left(\frac{x}{B}\right)^{C}\right) \qquad \text{(c.3)}$$

$$\frac{dy_{3(2)}}{dC} = -A\left(\frac{x}{B}\right)^{C}\ln\left(\frac{x}{B}\right)\exp\left(-\left(\frac{x}{B}\right)^{C}\right) \qquad \text{(c.4)}$$