On-line Fault Detection In Fermentation Development Facilities

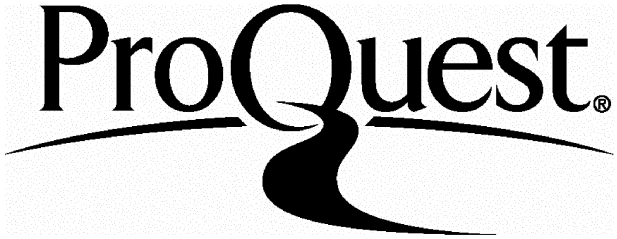
Ph.D. Thesis

Francis William Roche

Department Of Electronic And Electrical Engineering

University College London

January 1998



Acknowledgements

I would like to acknowledge and express my gratitude for the help, support, and expert guidance provided throughought this project by Mrs Nina Thornhill (UCL Dept. of Electronic and Electrical Engineering) and Dr.Steve Carleysmith (SmithKline Beecham plc).

Without their help and encouragement and that of Dr. Judi Sutherland (Ex SB), the completion of this thesis would not have been possible.

Abstract

This thesis reports the results of an investigation into the means which can be used to automate the supervision of a fermenter typical of the type used in antibiotic fermentation process development. The work uses detailed data sets obtained from experimental fermentations as a basis to evaluate possible on line supervision techniques.

The basis of this work is the use of multiple data sources to determine whether a fermentation is proceeding as planned. These data can be tested for internal consistency, and compared with data from previous fermentations to generate an automatic method of detecting faults. The use of an on line Kalman Filter as an observer to supply information to generate analytical redundancies was investigated.

As part of this work, an analysis is presented of the accuracy and precision of information available from a typical fed batch fermentation.

The report concludes with a recommendation as to which methodologies are most appropriate for monitoring large numbers of small scale development fermentations, and a review of potential areas for further work, including AI and Multivariate Statistical analysis is included.

Table Of Contents

Project Objectives

A large antibiotic fermentation development facility, in industry, typically can contain fifty to one hundred fermenters. These are employed on a range of activities from seed culture preparation to process and strain development. These fermenters pose a considerable monitoring and control problem, requiring skilled personnel to monitor the correct running of the vessels. The aim of this project is to investigate the means for automating the monitoring, and to improve the utilisation of information obtained in the course of routine sampling, of the fermentations.

The monitoring of large suites of fermenters has become an increasingly automated process, as control and instrumentation technology becomes cheaper and more reliable. Manual supervision is labour intensive, expensive and error prone. Typically, on-line information from sensors and frequently, data from off gas analysis using a mass spectrometer are monitored by control computers. This results in the availability of large amounts of information which can be used to perform more sophisticated supervision tasks. This work investigates the best means of using these data for the detection of process faults on-line.

The basis for these methods is that the data obtained by monitoring a fermentation can be combined with information derived from constraints on the system, and data from past fermentations, to produce redundant information. These sets of information can be checked statistically for inconsistencies, and against range and rate of change limits. If undue deviations are discovered, the presence of a fault can be inferred. Identification of the fault may then be attempted by comparing its characteristics with that of an on-line knowledge base containing descriptions of the likely consequences of faults.

The project was supported by SmithKline Beecham, and carried out in part in their fermentation research facility as well as in the UCL pilot plant. The project is therefore oriented to the practical needs of SmithKline Beecham, and utilises standard statistical process control ideas as well as investigating the applicability of state of the art methods. A key part of this Ph.D program was to look at the possibility of integrating conventional statistical process control ideas with more sophisticated fault detection techniques,

Project Outline

This section provides an overview of the important steps within the project and describes the key methods chosen to meet the objectives. It also summarises the outcomes of the work.

There are a number of steps in the development of a fault detection and identification system. Methods of obtaining information both available in real time, and with relatively short delays, must be devised. These different information sets must be integrated into a set of expected performance indices about the fermentation. Finally a method of identifying discrepancies between the observed information sets, and between these and the expected fermenter performance must be found.

The first step was to determine the type and accuracy of information that can be obtained out of a typical research fermenter. This required the running of 21 fed batch penicillin fermentations, and the detailed analysis of the data obtained from these batches. Monitoring and analysis techniques typical of those used at the project sponsor (Smithkline Beecham plc) were used throughout. This is because the intention of this project was to devise methods for improving the performance of an installed equipment base rather than to recommend hardware improvements.

The fermentation of penicillin was chosen as the working system principally because it was one of two fermentation systems of interest to the project sponsors, the other fermentation of potential interest to the sponsor being the production of clavulanic acid, which is another β-lactam fermentation. The reason for preferring penicillin was that there are available reasonable models in the literature for penicillin, which is an advantage since this project needs to make use of such models. In addition there was a group working on *Penicillium chrysogenum* fermentations in the Department of Chemical and Biochemical Engineering at UCL and this was of great assistance, both in setting up a practical system, and through collaborative fermentations, obtaining relatively high frequency samples.

As well as generating experimental data, a review was made of possible fault detection techniques. The most useful available techniques turned out to be variations on statistical process control methodologies. The function to be tested against

control limits could be raw process data, conditioned process data, or the output from more complex mathematical techniques such as Kalman Filters, neural networks, or multivariate regressions. The confidence limits themselves could be derived statistically from data from past fermentations.

In order to use on-line observers, it is necessary to have a model of the process of interest. An advantage of the penicillin fermentation process is that it has been the subject of extensive modelling effort; this project took advantage of that work. The model chosen was the one derived by Bajpai and Reuss (1980). This model has already been used as the basis of a Kalman filter (Montague *et al.* (1986a,b)). A predictor corrector form of the Kalman filter was implemented using this model along the lines of the one described by Lichfield (1979).

To be able to utilise a Kalman filter some work is required to determine the degree of confidence in the measurements used to generate feedback, such as carbon dioxide evolution rate, and also the quality of the model. Work was carried out to determine accurately parameters such as the variance of biomass measurements for a specific fermenter.

A series of 21 penicillin fermentations was carried out to provide a basic set of data of known or estimatable accuracy, on which the observer system could be tested. The important feature of these fermentations was that faults could be deliberately introduced. Some of these fermentations had step changes in conditions introduced, which provided data on how a real system would react after the occurrence of a faults. Some fermentations also developed unplanned faults, which generated useful additional data. The data obtained from these fermentations are presented in this thesis.

The analysis of data available from the fermentations carried out as part of this project shows that the most practical means of automating fault detection is by combining information from a range of sources. The single most useful source of data is previously conducted fermentations, which provide a benchmark against which a fermentation in progress can be assessed. The other key source of data is on line analytical redundancy. Where a process parameter can be inferred from more than one source, for example total carbon inputs vs. output and accumulation, the

difference between carbon accumulation as assessed from tracking inputs and outputs can yield data on whether the fermentation inputs are as planned.
A key conclusion is that quicker and more accurate fault detection could be obtained by using ratios of fermentation data to eliminate some of the batch to batch variation which can be expected in development fermentations. This method in effect enhances the available data by applying *a priori* knowledge about how fermentations work.

The ultimate purpose of this project is to define an algorithm to implement a fault detection scheme, backed with data to show that the individual components of such an algorithm will function. The algorithm, at its lowest level, will consist of a set of expected performance trajectories which are established by historical data. The next level will be on-line monitoring of a fermentation process, enhanced by use of an observer and off line data. This monitoring will test for faults based on inconsistencies in the fermentation data, and discrepancies between current and past performance trajectories.

Chronology

The work reported in this thesis took place between January 1988 and December 1990. At the time of the work, significant new inputs to the field of automatic fermentation supervision were developed.

Because this thesis has been delayed due to pressures of other work, there is a danger of presenting material which is no longer considered novel, or which has been overtaken by other means of meeting the projects objectives.

This thesis has been brought up to date by reference to key publications. Where publications after 1991 indicate an alternative method, this work is placed within the new context.

It is pleasing to note that, even with the introduction of completely new techniques such as the method of multivariate statistics, the work in this thesis can still claim novelty. The focus on this work of integrating practical fermentation monitoring techniques with knowledge of the biology of fermentation processes and advanced statistical algorithms provides a way forward which complements the more recently published methods.

Chapter 1: Introduction To The Penicillin Fermentation Process

There are a number of levels of information available around a fermenter. This thesis describes this information and shows how it can be combined to establish an automatic fault detection and identification scheme. This chapter reviews the following aspects of the penicillin process:

- The fermenter hardware system
- The penicillin fermentation process.
- On and off line measurements.
- The type and impact of process faults.

1.1 Description Of Fermenter Systems

Stirred tank fermenters, which are by far the most commonly used for industrial fermentations, have typically the configuration described in fig 1. overleaf

Information is available from these vessels from physical sensors such as temperature, pH, and dissolved oxygen probes which are mounted in the vessel itself. The composition of the exit gas stream from an aerobic fermentation can be determined, this is often done using a mass spectrometer or paramagnetic gas analysers. Samples of the contents of a fermentation can be withdrawn and analysed for biomass concentration and key substrate and product species. Typically these off-line data are available infrequently and after a time delay depending on the organisation of the assay used.

Additional information can be derived from the application of elemental mass balances (Heijnen *et al.* 1979, 1981, Stephanopoulos *et al.* 1986), and the use of simple kinetic models (Essner *et al.* 1981). The parameters for these models, such as maximum specific growth rate, require a good deal of effort to acquire (Nihtila *et al.* 1977). As an alternative, statistical descriptions of prior fermentations can be utilised, based on mean process variable trajectory and distribution.

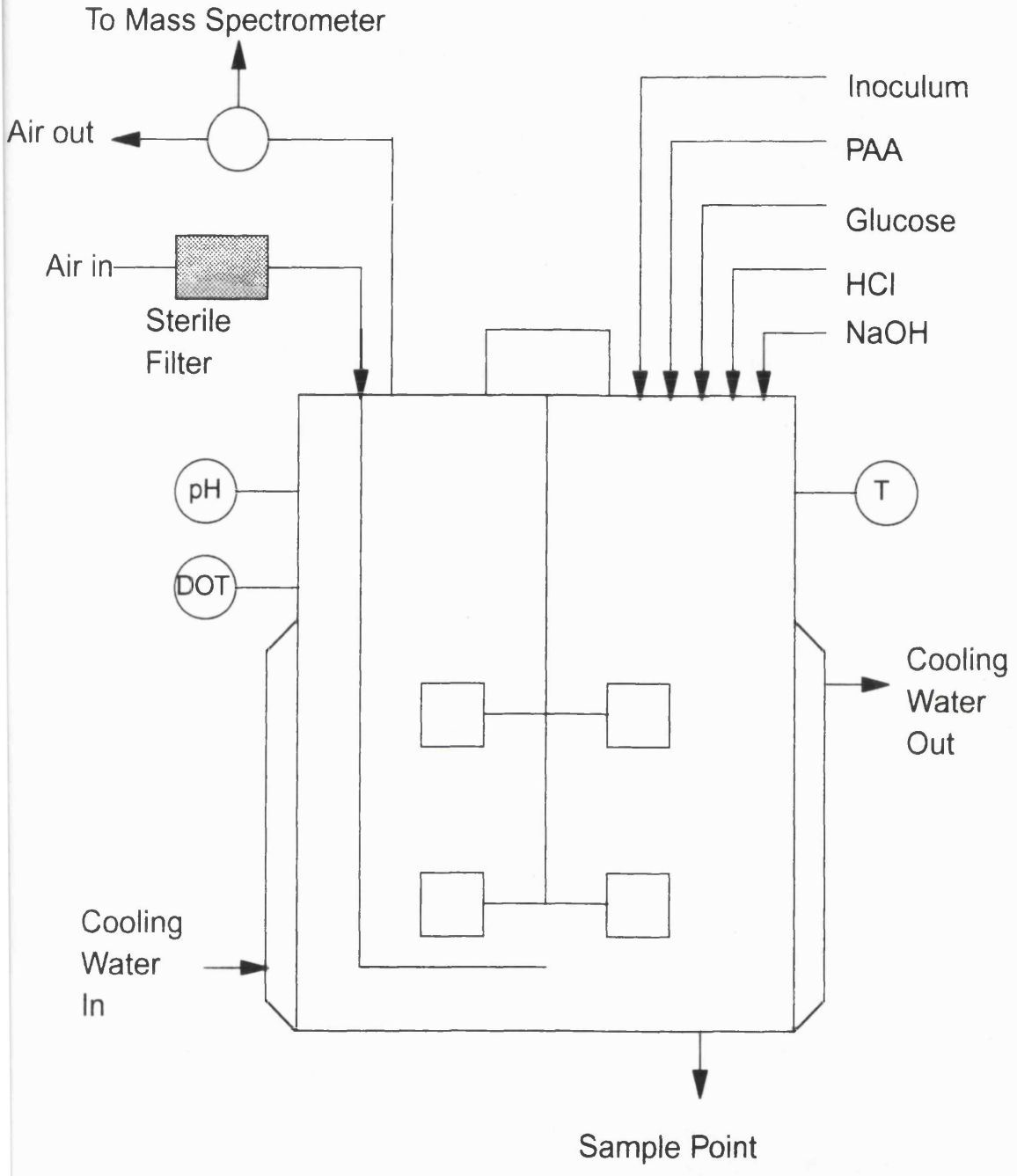
In large vessels, several probes of each type may be fitted, but this is far from a universal practice. This is because it can be considered that the risk of contamination posed by the extra fittings is a greater threat to the fermentation than the risk due to probe failures, particularly when on-line sterilisable ports are fitted, so that suspect instrumentation can be changed.

In small research fermenters (Typically 1.5 l to 200 l working volume), size limits the number of available probes, and in any case cost would probably preclude multiple probes. In addition, on this scale of fermenter, probes are likely to be of lower quality, and used for much longer than in production fermenters for cost reasons. In addition to the instrumentation in the fermenter, in most cases facilities exist for analysing the exit gas stream. This is usually by means of a mass spectrometer, or a combination of infra-red analysers for carbon dioxide, and paramagnetic analysers for oxygen. The analysis of effluent gas composition is of great importance, as, together with gas flow rate data, it gives an indication of the rate of respiration, and hence the level of metabolic activity in the culture.

Physical information about a fermentation, such as the rate of addition of substrates, is also usually readily available. In addition it is standard practice to sample the fermentation at regular intervals, the frequency of which is a function of the fermentation time, and the reason for which it is being carried out. At SmithKline Beecham it was standard practice to sample a 200 hour penicillin fermentation every eight hours. These samples are analysed for the concentration of biomass and the key components of the broth, such as substrate, and product composition.

The principal problem with this configuration, from a control point of view, is that several key process states remain unmeasured. In particular, it is not possible as yet to measure biomass, or the chemical composition of fermentation broths on-line. Off-line samples in addition provide accurate information only after a considerable time delay.

Work is ongoing to develop new techniques for on-line measurement of these parameters. In particular, Aber Instruments have developed a biomass probe, but this has not been proven as effective in mycelial fermentations as in all single cell fermentations. Cagney (1984) suggests that this is because of the presence of



Schematic Diagram Of Fermenter Configuration Used In This Work

ionisable salts. Given the design of the latest probes, there is some doubt whether they would function given the potential for fouling in a *Penicillium* fermentation.

1.2 Penicillin Fermentation Process

The production process for penicillin G has three distinct phases:

- Strain management: maintenance of live *Penicillium chrysogenum* spores.
- Inoculum production: conversion of spores into a population of mycelium which can be used to initiate a fermentation.
- Fermentation: use of the *Penicillium chrysogenum* micelles to make penicillin.

The production of penicillin is thought to be a defence mechanism for organisms in the wild, helping them to compete for scarce resources. Penicillin is only produced by the *Penicillium* fungi when there are insufficient nutrients present for unrestricted growth, hence the term secondary metabolite which is applied to penicillin. Therefore the fermentation strategy is to grow a quantity of biomass rapidly, and then keep this biomass in a state which encourages the production of penicillin, by limiting the rate of carbon addition.

The desired fermentation profile is achieved by charging the fermentation with an initial shot of carbon containing material, e.g. glucose, then feeding additional glucose throughout the fermentation at a relatively low rate. Typically the levels of glucose present in a fermentation broth sample are negligible. Such fermentations are called fed batch fermentations.

1.2.1 Strain Management

The production of penicillin G starts with spores of a specific (and usually highly optimised) strain of *Penicillium chrysogenum*. These spores are typically sourced from a spore bank, usually comprising of frozen slopes inoculated with spores of the desired strain. Very long term storage can also be achieved by freeze drying of spores. A sample of spores is inoculated onto the surface of a sterile agar slope of medium optimised for sporulation of the production strain. Typically such media, buffered to pH 5.5 to 6.5, would contain lactose, corn steep liquor, peptone and inorganic salts. These slant cultures are allowed to germinate, grow and sporulate under controlled conditions. Spores are harvested from mature slants by addition of a

standard volume of diluent (usually water plus a surfactant). The spore suspension thus generated is used to inoculate a seed fermentation.

1.2.2) Inoculum Production

The first step in the fermentation process is the inoculation of vegetative cultures. The vegetative cultures supply a sufficient concentration of fungal mycelium to the fermentation stage such that the generation of biomass in the full scale fermenter occurs over a reasonably short time period.

Inoculum development is typically conducted at 25°C in shake cultures and for large scale production in stirred vessels. A typical seed medium contains an organic nitrogen source such as corn steep liquor, and fermentable carbohydrate such as glucose together with organic salts.

1.2.3) Fermentation

The yield of penicillin per unit volume in a fermenter is the product of three parameters, biomass concentration (X), the specific rate of penicillin synthesis (q_{pen}) and the duration of the fermentation. The fermentation process is optimised to maximise the integral

$$\int_{t_1}^{t_2}[q_{pen}(t).X(t)]dt \qquad 1.1$$

where
t_1 = inoculation time
t_2 = harvest time

The first stage in the fermentation process is designed to achieve a high biomass concentration rapidly. Typically most of the biomass necessary to achieve high penicillin yields is grown within the first 30-40 hours of a fermentation. During this part of the fermentation, the organism is growing exponentially, with a biomass doubling time of about six hours. The fermentation at this stage is not producing significant amounts of penicillin. The end of the exponential growth phase is effected by limiting the amount of a key nutrient, usually the metabolisable carbon source. Once a sufficient biomass concentration has been achieved to support a satisfactory

penicillin yield, growth must continue at a certain minimum rate, at which new growth matches breakdown of mycelium, to achieve a high q_{pen}. The concentration of cells achieved by the time the end of the exponential growth phase occurs is targeted to permit this extra growth.
The key to a high yielding penicillin fermentation is the successful shift of a culture from exponential to slow or steady state growth with a high q_{pen} value. The fermentation is then run for 120-200 h (SmithKline Beecham's standard time is 186 h); during this phase most of the penicillin is produced. The exact harvest time is defined by the need to trade off rate of penicillin production with rate of degradation of existing penicillin stocks, and the cost of keeping the fermentation going. Care is taken to control the rate of biomass production by optimising the rate of addition of the limiting nutrient (usually the carbon source). The upper limit to growth is provided by the need to maintain this phase for as long as possible without running into oxygen limitation problems, as the biomass density and hence broth viscosity rises during the fermentation.

At the time of the shift from exponential to steady state growth, phenylacetic acid is added to the fermentation. The majority of this is not metabolised, but supplies the side chain which characterises Penicillin G from other penicillins. This side chain is typically cleaved off by an enzyme-catalysed reaction to provide a base β Lactam ring onto which alternative side chains can be added to produce the range of semi-synthetic penicillins. Once the level of oxygen in the broth drops below around 30% of saturated, penicillin production will decline. The precise lower limit for growth rate is strain dependent and is usually determined empirically.

Often the *Penicillium* mycelia in penicillin G fermentations form pellets; this has the advantage of reducing viscosity, thus helping oxygen mass transfer. The downside of pelleting is that organisms in the centre of the pellet become starved of oxygen and die.

At some time during the fermentation, usually because of oxygen limitation, the value of q_{pen} will decline to the point at which it is not economically worth running the fermentation further. In addition the rate of degradation plays a part in determining the end point of the fermentation, as penicillin concentration increases the rate of degradation will increase, decreasing the net rate of penicillin production to uneconomic levels.

1.2.4) Relationship Between Fermentation Parameters:

It is useful at this point to examine how the disparate data sources from a penicillin fermentation can be connected, as this shows the availability of redundancies in the data set produced from a monitored fermentation.

The following table summarises the key relationships which can be of use in fault detection:

Table 1.1. Relationship between fermentation parameters, and method of determination

Parameter	Method for Determination †	Associated Parameters††
Biomass concentration	• Packed volume. • Dry weight. • Mass balance • Model output • On-line observer output • Comparison with previous fermentations	• Specific Growth Rate. • Maintenance coefficient. • Yield of biomass on substrate. • Yield on oxygen. • Ratio of biomass production to penicillin production.
Glucose concentration	• Analysis of broth sample • Mass balance • Model output • On-line observer output • Comparison with previous fermentations	• Yield of biomass on glucose. • Yield of CO_2 on glucose.
Phenylacetic acid concentration	• Analysis of broth sample • Mass balance • Comparison with previous fermentations	
Penicillin concentration	• Analysis of broth sample • Mass balance • Model output‡ • On-line observer output • Comparison with previous fermentations	• Penicillin production rate. • Yield of penicillin on substrate. • Rate of penicillin decay.
Carbon dioxide production	• Measurement of off gas composition • Mass balance • Model output • On-line observer output • Comparison with previous fermentations	• Respiratory quotient. • Yield of CO_2 on glucose.
Oxygen uptake rate	• Measurement of off gas composition • Mass balance • Model output‡ • On-line observer output • Comparison with previous fermentations	• Respiratory quotient. • Yield of biomass on oxygen. • Yield of penicillin on oxygen.
pH	• Direct measurement using in-line probe • Analysis of samples	• Penicillin decay rate.
Temperature	• Direct measurement using in-line probe	• Penicillin decay rate.
$K_L a$ (Mass transfer coefficient)	• Infer from oxygen uptake and dissolved oxygen concentration.	• Dissolved oxygen concentration. • Dissolved carbon dioxide concentration. • Evaporation rate.

† For specific details of assays used, see chapter 3

†† Definitions of these parameters will be provided in chapter 2.
‡ Taking account of degradation.

1.3) Information Available Around A Typical Penicillin Fermentation

A key component of this project is to understand the type and quality of information available about real fermentations. This section discusses the measurements which can be obtained. Data about the accuracy and precision of that information is presented in chapter 3.section 3.5.

1.3.1) Off-line Data

Broth samples can be taken; typically this might be carried out 2 - 4 times a day by a process operator. From these samples, in a typical penicillin fermentation, the following measurements could be obtained:

- Sediment Volume (after centrifugation).
- Dry weight (solids content of sample).
- Substrate composition (e.g. glucose. phenyl acetic acid, sulphate, ammonium concentration etc.).
- Product (e.g. penicillin) concentration.
- pH
- Viscosity
- Appearance of samples on microscopic inspection.

An issue with broth sampling is that on a small scale a significant fraction of the total broth volume can be removed in the course of a 180 hour fermentation. This can distort the analysis of the fermentation kinetics unless accounted for, as the feed and carbon dioxide evolution rates are measured on a whole fermenter basis, while the biomass and other broth components are measured on the basis of concentration.

1.3.2) On-line Data

On small research fermenters, on-line data collection is often limited by the number of probes that can be fitted to the vessel used. The following data can be obtained:

- pH
- Temperature
- Dissolved oxygen
- Dissolved carbon dioxide.
- Aeration rate
- Agitation rate
- Volume (inferred from load cell data or measured directly)
- Exit gas composition
- Feed rates
- Heat input / output.
- Fermenter pressure.

For a detailed discussion of the instrumentation typically used in a typical fermenter suite see Carleysmith and Fox (1984) or Bailey and Ollis 1986. For an explanation of the means used to monitor exit gas see Heinzle (1991) who details the use of on-line mass spectrometry. In addition to the instrumentation described above, there is ongoing work to develop probes that can measure other fermenter parameters such as biomass (Cagney 1984) and secondary metabolite concentrations (Clarke 1985)

1.4 Fermentation Failure Mechanisms

Faults may be due to contamination of the fermentation media by competing organisms, to failures in the equipment used, or to operator errors. This section seeks to describe the consequences of these faults. This section was compiled with the help of penicillin fermentation experts at UCL and SB in interviews during 1989 and 1990.

1.4.1 Biological Failure

Biological failure is taken here to mean contamination of the fermentation broth by an organism other than the one which it is intended to grow. Other examples of biological 'failures' include the reversion to wild type in genetically engineered organisms, or the growth of fungi in the wrong morphology resulting in excessive pelleting.

1.4.1.1 Contamination

The impact of a contaminant is dependent upon when it appears in the fermentation and the nature of the contaminating organism. A number of scenarios are possible.

Fast Growing Contaminant Present at the Start of a Fermentation: The faster growing organisms outgrow the *Penicillium* fungus and become the dominant form in the fermentation. Upon depletion of the initial glucose supply, biomass production would be at approximately the same rate as for *Penicillium*, as it is unlikely that the yield of biomass on glucose would be radically different. Some extra biomass would probably be produced, reflecting the lack of penicillin formation.

At some point, if the organism is incapable of metabolising phenylacetic acid (PAA) (added as a necessary precursor for penicillin G production) the PAA will reach toxic levels and kill the contaminant. This form of contamination is readily detected, after a period of time, by observation of a sample of medium under a microscope.

The exponential growth phase of the fermentation can be expected to be more rapid since the contaminating organism must grow more rapidly than the *Penicillium* fungus to dominate the fermentation. A more rapid than usual build-up in carbon dioxide evolution, and a faster depletion of broth glucose can be expected. It is not necessarily the case that a greater than usual depletion in the dissolved oxygen level can be expected, as the fast growing organism is not likely to produce as large an increase in viscosity as mycelial *Penicillium* fungi. This reduction in viscosity can be expected to promote oxygen mass transfer and hence result in a higher than usual dissolved oxygen concentration.

Fast Growing Contamination After the Start of a Fermentation: If a non penicillin sensitive organism contaminates the fermentation after the initial glucose has been used up, it cannot take over the fermentation, as the level of available sugar will not allow rapid growth. Such an organism however is likely to be a *β-Lactamase* producer (*β-Lactamase* is the enzyme in bacteria which confers penicillin resistance, it functions by cleaving the *β-Lactam* ring at the heart of a penicillin molecule), otherwise it could not survive in the presence of the penicillin in the broth. The consequence of its presence will be an increase in the rate of penicillin degradation. This is likely to be the only significant effect of such a contamination. A similar effect

can be seen if a very low level of slowly growing contaminant is present from the start of a fermentation.

The presence of this type of contaminant can be established by adding additional sugar to the fermentation permitting both organisms present to grow at a rate unlimited by carbon availability. The specific growth rate observed in this event will be higher than the maximum specific growth rate which occurs in uncontaminated fermentations. This can easily be tracked by measuring the rate of increase of carbon dioxide evolution.

Effect Of A Slow Growing Contamination: A contaminant with growth rates similar to *Penicillium*, which does not produce a *β-lactamase* will not impact the fermentation, since it will be present in much lower concentrations that the *Penicillium* inoculation.

In any case, the presence of a contaminant, by competing for nutrients, is likely to lower penicillin titres.

1.4.2) Hardware Failures

Fermentation hardware periodically fails, or is set up incorrectly. The most sensitive components in most fermenters are dissolved oxygen and pH sensors. These rely on sensitive membrane barriers to maintain the electrochemical effects they use to effect their measurements. Also both types of probe are subject to drift due to depletion of electrolyte species over time. In addition the process of autoclaving or steam sterilising to ensure probe sterility can damage membranes. All the other mechanical and electronic equipment used to maintain and monitor a fermentation are of course also subject to faults from time to time.

It is worth pointing out that in an industrial context, these failures are likely to be more common on small scale research and development fermenters. These are present in large numbers in many industrial sites, they are likely to be fitted with lower quality equipment than is typical for a production fermenter, and this is likely to be changed after it fails rather than on a preventive maintenance schedule. In a production fermenter where the cost of charging with medium alone can be tens of

thousands of pounds, there is a much greater incentive to minimise hardware failure by using high quality instrumentation, or hardware redundancy.

1.4.2.1) Consequences of Possible Hardware Failures

This section discusses the impact of a fault in each of the systems used to maintain a fermentation.

Loss Of Temperature Control: As with any potential fault the effects are a function of the magnitude of the failure, and its nature. Temperature probes may fail by becoming completely inoperable in which case a default, or garbage value will be returned to the control computer. This can easily be detected by limit flags on the supervisory system. An alternative fault can occur where output of a probe drifts from the point at which it was calibrated. This will not be readily apparent as the temperature control system will attempt to return the fermenter to its set point which would now be incorrect.

A failure in the actual control system, such as a loss of cooling water, would be easily detectable as the control system would not be able to compensate, and temperature would deviate from the set point. It is a matter of routine to install low and high temperature alarms in modern control systems.

A failure in the temperature control system that was not swiftly detected and rectified would not immediately destroy the fermentation. The temperature at which a penicillin fermentation is run during the steady state part of its cycle trades off rate of penicillin production with rate of degradation. A rise in temperature would therefore alter the rate of penicillin accumulation. As the batch temperature approached 32-35°C, the *Penicillium* would die, ending the fermentation.

It is unlikely that a failure in temperature control which did not rapidly kill the *penicillium* organisms would impact noticeably on biomass, unless it occurred during the exponential phase of the fermentation. This is because the level of biomass in the linear growth phase is principally determined by the carbon fed into the fermenter. In the exponential growth phase, the observed specific growth rate will vary depending on whether the batch cooled or overheated, as maximum specific growth rate is a function of temperature.

A shift in temperature can be expected to impact on the level of dissolved oxygen and carbon dioxide. The solubility of these gases is inversely proportional to temperature (Perry 1985). The level of carbon dioxide present in solution is modified by its equilibrium relationship with carbonate ions. Oxygen concentration is likely to be a much more sensitive indicator of a temperature shift; the Henry's law constant for oxygen in water increases by 16% between 25°C and 35°C. Given that a temperature rise is likely to promote metabolic activity and decrease oxygen solubility, a drop in dissolved oxygen levels can be expected to be seen.

Failure in pH Control: Penicillium chrysogenum, when growing, produces lactic acid, so pH control is typically effected by using an alkaline solution (ammonium hydroxide) which is added automatically to keep pH within specification. The pH used is chosen as with temperature to optimise penicillin yield. Unless very high or very low pH conditions (i.e. pH is <4 or > 10) are encountered, the *Penicillium* will maintain growth. The main manifestation of pH changes within these bounds would therefore be a decrease in the rate of penicillin accumulation, either due to reduced synthesis or increased degradation.

A rapid change in pH results in a transient change in carbon dioxide evolution rate. This is because the ratio of dissolved carbon dioxide to carbonate ions is dependent on pH; a change in pH which shifted this equilibrium alters the amount of carbon dioxide in solution, which in turn changes the quantity stripped out by the sparged airflow.

Effect Of An Agitation Fault: A reduction in agitation rate will reduce the disruption of the incoming air stream which is aerating the fermentation. This will result in poorer mass transfer of oxygen from the incoming air and hence a lower dissolved oxygen concentration in the fermentation broth. As the *Penicillium* becomes starved of oxygen, metabolic activity declines, then ceases. This occurs at a dissolved oxygen concentration of 30-40% of the concentration of broth fully saturated with oxygen. If dissolved oxygen levels remain above this, no impact can be expected on the fermentation.

On larger scale equipment, a reduction in agitation will result in the formation of segregated zones in the fermenter, some of which will be completely unmixed and therefore starved both of glucose and oxygen. This will result in a reduction in the

viable biomass content of the overall fermentation. A consequence of this reduction is likely to be a drop in carbon dioxide production, and also of penicillin production and PAA utilisation. This tendency to produce dead zones on inadequate aeration is likely to be come more marked as the fermentation progresses, and the broth becomes both more viscous and more shear thinning. A property of a shear thinning fluid is its relative inefficiency in transmitting torque from an agitator.

An increase in agitation only disrupts the fermentation if the shear applied by the turbines became intense enough to disrupt the *Penicillin* mycelia. This is unlikely to be an issue on a small vessel where agitator tip speeds are low in comparison to production scale equipment.

Effect Of An Aeration Fault: As long as the rate of aeration remains above approximately 0.3 v/v/min there is unlikely to be any noticeable impact on the fermentation from an aeration reduction. Below this the fungi will become starved of oxygen and eventually die. A reduction in air flow will be apparent because the exit gas oxygen and carbon dioxide concentrations will be radically different from their expected values, with carbon dioxide higher and oxygen lower than expected (i.e. the respiratory quotient of the system will be magnified).

Feed Rate Fault: During the exponential growth phase of the fermentation an interruption in glucose feed results in an accelerated depletion of the batch carbon supply. After the end of the exponential growth phase of the fermentation an interruption in glucose supply will essentially halt metabolic activity as the level of metabolisable carbon in the fermenter during this part of the process is very low. A feed outside the planned feed rate will result in either more or less metabolic activity than expected, and hence a higher or lower level of carbon dioxide evolution and biomass production.

In the fermentation processes considered in this thesis, phenylacetic acid (PAA) was a secondary feed. This material was added as a precursor to penicillin G production. An overfeed of PAA would result in a build-up of PAA to toxic levels. an underfeed would result in a reduction in the rate of penicillin production and the production of unwanted variants of the penicillin molecule such as penicillin K or penicilloic acid, the so called natural penicillins.

The consequences of the principal possible fermentation faults are summarised in table 1.2.

Table 1.2 Impact Of Faults On Fermenter Parameters

Impact of possible fermentation faults on the concentration in the fermentation broth of biomass, glucose, phenylacetic acid, penicillin and oxygen, and the concentration of carbon dioxide in the exit gas stream. This assumes that the magnitude of the faults are large enough to influence the outcome of the fermentations but not large enough to kill them.

(⇧ = *increase*, ▫ = no change, ⇩ = decrease)

Fault	Biomass	Glucose	[PAA]	[Pen G]	$[O_2]$	$[CO_2]$
Fast growing contaminant: exponential phase	⇧	⇩	▫	▫	⇩	⇧
Fast Growing contaminant : steady state phase	▫	▫	⇧	⇩	⇩	⇧
Slow growing contaminant	▫	▫	⇧	⇩	▫	▫
Temperature Increase: exponential phase.	⇧	⇩	▫	▫	⇩	⇧
Temperature Decrease: exponential phase.	⇩	⇧	▫	▫	⇧	⇩
Temperature Increase: steady state phase.	▫	▫	⇩	⇩	⇩	▫
Temperature Decrease: steady state phase.	▫	▫	⇧	⇩	⇧	▫

Fault	Biomass	Glucose	[PAA]	[Pen G]	$[O_2]$	$[CO_2]$
pH decrease : steady state phase.	▫	▫	▫	⇩	▫	▫
Agitation decrease: exponential phase	⇩	⇧	▫	▫	⇧	⇩
Agitation decrease: steady state phase	▫	⇧	⇧	⇩	⇧	⇩
Aeration increase: exponential phase	▫	▫	▫	▫	⇧	⇩
Aeration increase: steady state phase	▫	⇧	⇧	⇩	⇩	⇧
Aeration decrease: exponential phase	⇩	⇧	▫	▫	⇩	⇧
Aeration decrease: steady state phase	▫	⇧	⇧	⇩	⇩	⇧
Overfeed of glucose: exponential phase.	⇧	⇧	▫	▫	⇩	⇧
Overfeed of glucose: steady state phase.	⇧	⇧	⇧	⇩	⇩	⇧
Underfeed of glucose: exponential phase.	⇩	⇩	▫	▫	⇧	⇩
Underfeed of glucose: steady state phase.	⇩	▫	⇧	⇩	⇧	⇩
Overfeed of PAA:	⇩	⇧	⇧	▫	⇧	⇩
Underfeed of PAA:	▫	▫	⇧	⇩	▫	▫

1.5 Summary

This chapter has described the fermentation hardware used in this project, and typically used in fermentation development departments throughout the pharmaceutical industry. The basis of fed batch antibiotic fermentations were presented. It was shown that the fed batch process used is optimised to deliver the best combination of raw materials usage and process throughput by manipulating the quantity of biomass produced during an initial unlimited or exponential growth phase, and a much longer production phase.

Analyses were then presented which show:

- Key fermentation states.
- Links between these states and measurable parameters.
- Information available around a typical fermentation
- An outline of possible fermentation faults.

The analyses show that it is possible to make predictions about what will happen if a fermentation experiences a fault.

Chapter 2: Introduction To Fault Analysis Methodologies

There is a range of fault analysis methodologies, ranging from purely analytical means of detecting and diagnosing faults, to purely heuristic. Most methods employ a degree of statistical analysis of the output from analytical models . The nature of the problem is such that no one approach will be able to detect all possible faults adequately, thus several techniques, employed complementarily, would provide the best supervisory system. The approach used in this project was therefore to identify the available data, and use this to identify feasible fault detection methodologies

It should be pointed out here that the detection of faults in instrumentation is widely described in the literature; the more complex problem of detecting faults in plant actuators and systems is much less well described. There is also some specialised literature aimed at fault detection due to biological problems (Stephanopoulos, 1984).

This section is intended to review available fault detection and identification methodologies. It must be said that only very simple applications to fermentation systems have been made to date. The problem with fermentation systems is that they are noisy, non-linear systems, in which key process states are unmeasurable. This greatly complicates many of the methodologies suggested in this section.

The methods for detecting faults can be differentiated by their level of abstraction from the real system (Isermann 1984). Isermann makes the point that fault detection can be based on measurable signals, non-measurable state variables, non-measurable process parameters, or non-measurable characteristics of the fermentation process. In the most simple case, the signals from the fermenter can be analysed directly, checks being made for the absolute value of a variable, or its rate of change, these being checked against pre-set limits.

In addition it is of interest to look at the characteristics of the output from a sensor. Sharp changes in, for example, the noise characteristics can be indicative of probe failure. Basseville (1988) shows how the application of spectral methods can give information on the state of a process, particularly with periodic process components,

such as the dosing of acidity or foam regulators, or substrate input in a fed batch fermentation.

2.1 Use of Process Models

At the level of unmeasurable states the results of process models are used to check the consistency of instrumentation output. Wang and Stepanopoulos (1983) give a methodology for applying statistical hypothesis testing for error detection. This technique involves using mass balance equations to describe a fermentation. The output from these balances, taken together with the output from an observer, overdefines the fermentation. One compares the output from this model with the output from sensors, and tests the resulting residuals to check for bias in their values. If the test is passed, the residuals can be used to update the balance equation.

A methodology for the identification of errors on the detection of inconsistent results is also given by Wang and Stepanopoulos. This method works by deleting one measurement at a time and checking the consistency of the remaining measurements and their associated equations. This type of book-keeping redundancy is also applicable to information which is initially produced by dynamic models, such as rates of material utilisation, and production.

Wang and Stepanopoulos's technique is interesting in that it incorporates all the elements of a basic fault detection and identification (FDI) system. It is relatively easy to implement and is based on directly established relationships. However, it does rely in part on off-line data such as biomass and product concentrations. Some of these may be estimatable but there is a limit to this before the system becomes unobservable. The method also requires unbiased models which can track the fermentation parameters of interest.

A range of techniques is available for using dynamic equations describing a process for fault analysis. The usual approach is to couch the set of state equations in some form of observer. In this project this was done through the use of a Kalman filter (Litchfield 1979, Montague *et al.* 1986) A detailed description of a Kalman filter as applied to a penicillin fermentation is given in chapter 3 section 3. Willsky (1976) suggests the use of failure sensitive filters, the basis of these being that the conventional optimal Kalman filter becomes in time over-reliant on old measurements

as the computed error covariances and filter gain becomes small. This effect can be overcome by either weighting new data, or fixing the filter gain. The consequence of these alterations is that the filter is more sensitive to abrupt changes, at the cost however of degrading its performance with respect to process noise.

Faults are detected by comparing the state estimates produced by a sensitised filter, with the estimates produced by an optimal filter. This technique, while identifying that a fault has occurred, will not however give information on the nature of the fault. In addition, since the penicillin fermentation process is non-linear, an extended Kalman filter is required, and it cannot thus be said to be optimal. This undermines the basis of Willsky's method as linearisation errors mean that both filters will be subject to some bias.

An alternative approach described by Willsky is the use of multiple hypothesis filter detectors. In this approach, a bank of filters based on various hypotheses concerning the behaviour of the system, given the occurrence of faults, is established, together with a filter which describes the fully operating system. The estimates from each filter are compared to obtain a probability that the models contained therein are correct. The need to run multiple filters across a large number of fermentation batches may impose unreasonable computation time constraints, and require a large modelling effort in complex systems such as penicillin fermentations. Also this technique is vulnerable to frequent process changes. Every time the process is changed, new models both describing correct operation and faulty operation will be necessary.

The use of multiple failure sensitive filters does solve the fault identification problem. This approach could be very valuable in an environment where accurate fault diagnosis was important. for example, it could be used to selectively boost the reliability of sub components of the fermentation process such as the aeration system. In this case a consistent pattern of events on the onset of a fault could be expected, such as a drop in dissolved oxygen, and evolved carbon dioxide. This would make possible the establishment of some filter models which were insensitive to process changes. The lack of general applicability of this technique resulted in its not being used in this project.

An alternative way of looking at the output from a Kalman filter is the analysis of the residuals, or innovations, defined as the difference between the actual process, and the

estimated output from modelling, and previous data. A property of a sequence of innovations is that its elements are independent, hence a whiteness test can be used for fault detection (Mehra 1970, Mehra and Peschon 1971). This test is based on an autocorrelation function for a stationary process, alternatively a sample correlation coefficient can be used as a whiteness test. A second test that may be employed is to compute the mean of the innovation sequence, which should be zero if there is no erroneous bias present. The third test that Mehra suggests is to test that the covariance of the innovation sequence has a Chi-Square distribution. This technique could usefully be extended by a comparison of unobserved states with either off-line data, or modelled states. An issue with the use of whiteness tests in the context of this work is that the process under consideration is non linear. the equations used to describe it have to be linearised to make the Kalman filter work, this means that the residuals will inevitably be biased.

Another approach to the use of innovations, suggested by Willsky, is the generalised likelihood ratio approach, which is based on the modelled effects of a fault on the innovations obtained. From these models, fault signatures are obtained, and these are used to compute the likelihood of a failure having occurred. This technique requires an ability to model the unsteady state behaviour of the system, and also suffers the problem of the requirement of a growing bank of signatures, with increasing system complexity.

Turning the problem around, Chow and Willsky (1984) suggest the use of parity vectors, which are formed by comparison of system inputs and outputs. In the absence of failures, in a noisy system, the parity vector is a zero mean random vector, which can be formulated independent of system states. Upon a failure the parity vector becomes biased, and moreover the bias is a function of the specific failure. Thus the parity vector may be used as the signature carrying residual for a fault detection system. The property of being independent of state estimation makes this an attractive approach where available models are poor. This method is theoretically elegant, but requires a greater degree of system data than is used in the fermentation system under consideration.

2.2 Use Of Non Measurable Process Characteristics

Parameters, such as maximum specific growth rate, are in effect the constraints on a process, and can be very sensitive to failures, if it is possible to identify them. Parameter identification on-line is surveyed by Young (1981) and Billings (1980). The general conclusion is that, if a reasonable model is available it is possible to identify parameters on-line. Ono et al (1987) suggests that fault identification could be carried out by the comparison of the identified model prior to failure, either on-line, or based on historical data, and the model identified post failure. Fault signature libraries would be required in this approach, as well as a trigger means of detecting faults. Molle et al (1987) suggest the use of confidence limits for fault detection, based on the variance of the parameter in question. As will be shown these ideas provide a workable means of fault detection based on real fermentation data.

Representative process characteristics, collected on-line, such as efficiency of substrate/product conversion, may also be used for fault analysis. In this case a static relationship, based on predecided limits, could be used, to detect faults in the same way that limit and trend checking are applied to measured system variables.

2.2.1) Heuristic Approaches

Both Hudson (1988) and Dohnal (1985) have shown that fermenters can be usefully described by heuristic models. A number of workers have applied these techniques to fault detection. These techniques can also be used to generate fault signatures instead of analytically computed results. This project did not implement an expert system, but the outcomes certainly lend themselves to that approach and form a recommendation of the work. An expert system would be useful in that it would generate a general approach to fault identification, with less computational and analytical effort than using numerical fault signatures. In particular, it could provide for simpler pattern recognition than statistical hypothesis testing techniques.

The development of an expert system to deal with process faults is described by Kramer (1987). He uses the material and energy balance equations associated with a process to establish a set of governing equations which limit a process. Having established these constraints the next step is to incorporate them into a rule base. An example of this would be that the flows into and out of a system must equal material

accumulation. If this is not the case then a fault can be diagnosed. The advantage of Kramer's approach is this combination of analytical and inferential techniques.

Kramer goes on to suggest that the problem of uncertainties caused by sudden jumps in the diagnostic conclusions reached, inherent in Boolean inference techniques, can be dealt with using either Bayesian inference or fuzzy logic. Bayesian inference is a statistical response to the problem of reasoning with uncertainty, however since it requires data on the probability of the occurrence of faults, its implementation in process fault detection is problematic.

Fuzzy logic uses the concept of membership grade to set variable degrees of inclusion of constraints into the rule base. The mathematical theory of evidence approach is suggested, this modifies the fuzzy logic approach in that it uses the concept that the probability of a hypothesis being true is not necessarily one minus the probability of its negation. This is the logic form used by Kramer (1987). The problem with such techniques is to assign likelihoods to the hypotheses or propositions of interest. Kramer suggests the use of the variances of the innovations of the constraint equations.

The great advantages of this approach are that it allows for the interpretation of analytical, and statistical data, and for the incorporation of non mathematically defined propositions. Also it gives an approach that will allow the ranking of a series of possible faults in their relative likelihoods. The need to acquire large amounts of data to supply statistically meaningful values to membership grades is the key weakness of this approach as applied to fermentations.

An approach to fault diagnosis using Signed Directed Graph method of representing process interactions is suggested by Kramer and Palowitch (1987). Fault simulation using the SDG involves the formation of a set of directed 'fault' trees branching from the information node at which the fault is hypothesised. This would describe the direction and order of events occurring following a fault, thus for example, if an alkali reservoir on a fermenter emptied, the alkali flowrate would become zero, and the pH of the fermenter would fall, this would eventually change the growth rate of the organism and ultimately kill it. This type of causal chain is used to describe and diagnose faults, which is analogous to the preparation of fault trees used in safety and loss prevention studies.

2.3 Statistical Process Control Techniques

An important approach to fault detection first became well known in 1991, Those methods known as Principal Component Analysis (PCA) and projection to latent structures (PLS), otherwise known as partial least squares regression, use multivariate statistical methods to establish a statistical regression model Future data are analysed using the model in such a way that they can be plotted onto a 2 D or 3 D chart, called a scores plot analogous to a statistical quality control (SQC) chart, Deviations from normal operations are detected through points appearing on the scores plot in an unusual position, and outside the cluster of plots indicating normal operations.

An early review of PLS is given by Geladi *et al* (1986). A summary of the mathematical description in their work is presented below to illustrate the concepts behind the PLS and PCA techniques:

A starting point is to mean centre (subtract the mean value of a variable from each datum) and variance scale (divide each datum of a given variable by the standard deviation of the full data set).

First order multiple linear regression is a technique for establishing the linear relationship between variable y and m variables x in the form:

$$y = b_1 x_1 + b_2 x_2 + \ldots\ldots + b_m x_m + e \qquad 2.1$$

Where

y = Dependent Variable
x = Independent Variables.
b = Dependencies
e = Error or residual.

Putting this into matrix form

$$y = x'b + e \qquad 2.2$$

where b is a column vector and x is a row vector.

This describes the situation in which there is only one sample; if multiple samples are available the following matrix system arises:

$$\begin{bmatrix} y_1 \\ \\ y_n \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,m} \\ \\ x_{n,1} & x_{n,m} \end{bmatrix} \bullet \begin{bmatrix} b_1 \\ \\ b_m \end{bmatrix} + \begin{bmatrix} e_1 \\ \\ e_n \end{bmatrix} \qquad 2.3$$

Typically m < n. In this case there is no exact solution for the equation system, so instead algorithms aim at minimising the error vector e usually using the least squares method summarised in the following well known equation:

$$b = (x'x)^{-1}x'y \qquad 2.4$$

The problem with multiple least squares(MLS) is its dependence on the existence of the x'x matrix.

It should be noted that MLS works as well for multiple dependent variables as for one. The matrix system simply becomes:

$$\begin{bmatrix} y_{1,1} & y_{1,p} \\ \\ y_{n,1} & y_{n,p} \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,m} \\ \\ x_{n,1} & x_{n,m} \end{bmatrix} \bullet \begin{bmatrix} b_{1,1} & b_{1,p} \\ \\ b_{n,1} & b_{n,p} \end{bmatrix} + \begin{bmatrix} e_{1,1} & e_{1,p} \\ \\ e_{n,1} & e_{n,p} \end{bmatrix} \qquad 2.5$$

Principal Component analysis (PCA) is a method of writing a data matrix X of rank r as a bank of r matrices of rank 1

$$X = M_1 + M_2 + ... + M_n \qquad 2.6$$

These M matrices can be written as the product of two outer matrices, known as a score vector (t) and a loading vector (p) thus equation 2.6 becomes:

$$X = t_1 p_1' + t_2 p_2' + ... + t_a p_a' \qquad 2.7$$

or

$$X = TP' \qquad 2.8$$

The key to this technique is the definition of t and p. The principal component is the line of best fit between two variables given minimal residuals. The loading vector contains the direction cosines or the projections of a unit vector along the principal component on the axis of a plot of the line of best fit. Essentially they define the slope of each fitted line. The scores vector is the co-ordinate (i.e. distance from origin) of each datum along that line.

What is required to identify T and P are operators that project the rows and columns of the X matrix onto a single dimension. The algorithm for achieving this is nonlinear iterative partial least squares (NIPALS). This can be summarised as follows:

1) take a vector x_j from X and call it t_h: $t_h = x_j$

2) calculate p_h': $ph' = t_h'X/t_h't_h$

3) Normalise ph' to length 1

4) Calculate t_h : $t_h = Xp_h/p_h'p_h$

5) Compare the result of step 5 with the result of step 1. If they differ, iterate between step 2 and 4.

Summarising, a data matrix X of rank r can be decomposed to a sum of r rank 1 matrices. The point of this is that the m.n data matrix can be reduced to an a n matrix comprising the scores and loadings. This T matrix has reduced rank and also has advantageous properties (orthogonality).

The partial Least Squares (PLS) regression model can be obtained from the NIPALS algorithm. The X matrix becomes the data matrix. The PSL model essentially consists of the dependent and independent data (X and Y) matrices and an inner (U) matrix linking both blocks.

$$X = TP' + E \qquad 2.9$$

Similarly, dependent variables can be described

$$Y = UQ' + F \qquad 2.10$$

Expanded these become

$$\begin{bmatrix} x_{1,1} & x_{1,m} \\ y_{n,1} & x_{n,m} \end{bmatrix} = \begin{bmatrix} t_{1,1} & t_{1,m} \\ t_{n,1} & t_{o,m} \end{bmatrix} \bullet \begin{bmatrix} p'_{1,1} & p'_{1,p} \\ p'_{o,1} & p'_{o,p} \end{bmatrix} + \begin{bmatrix} e_{1,1} & e_{1,p} \\ e_{n,1} & e_{n,p} \end{bmatrix} \qquad 2.11$$

$$\begin{bmatrix} y_{1,1} & y_{1,m} \\ y_{n,1} & y_{n,m} \end{bmatrix} = \begin{bmatrix} u_{1,1} & u_{1,m} \\ u_{n,1} & u_{o,m} \end{bmatrix} \bullet \begin{bmatrix} q'_{1,1} & q'_{1,p} \\ q'_{o,1} & q'_{o,p} \end{bmatrix} + \begin{bmatrix} f_{1,1} & f_{1,p} \\ f_{n,1} & f_{n,p} \end{bmatrix} \qquad 2.12$$

The basis of the algorithm is to minimise F, and at the same time develop a relationship between X and Y. This relationship is developed in the same way as in PCA (equation 2.2). A simple linear model

$$\hat{u}_h = b_h t_h \qquad 2.13$$

relates u and hence Y and t. B in this context plays the role of the regression coefficient matrix in both the MLR and PCR models. The key to the PCA approach is an algorithm which exchanges information to the principal components (U,T) to permit a minimisation of the regression error. The starting point is the following algorithm, which is an extension of the NIPALS algorithm to permit interchange of scores for the T and U matrices thus:

1) Take u_{start} = some y_j

2) q'=u'X/u'u

3) $p'_{new}=p'_{old}/|p'_{old}|$

4) $t=Xp/p'p$

5) $q'=t'Y/t't$

6) $q'_{new}=q'_{old}/|q'_{old}|$

7) $u=Yq/q'q$

8) Compare t in step 4 with the previous iteration, if they are equal within acceptable rounding errors halt the procedure.

Summarising, PLS includes independent models of the X (independent variables) and Y (dependent Variables) matrices, known as the outer relationships. These modals are in the form of reduced order matrices (vs. original data) derived from data by means of least squares regression.

These models are themselves linked by a least squares regression derived model. This model is known as the inner relationship and has the form

$$\hat{u}_h = b_h t_h \qquad 2.14$$

Skagerberg et al (1991) provide an excellent description of the use of these techniques in control of a continuous low density polyethylene making reactor. They show how the PLS technique can be applied to data which is ill conditioned, due to variables being highly correlated and time differentiated. They show that given sufficient data, the PLS technique is a useful tool for process monitoring, although it does produce some problems in interpretation with multiple significant dimensions and interpretation of time dependent data.

Kresta et al (1991,1992) show some further steady state applications of PLS. They show some of the difficulties of combining high and low frequency data, and also comment on the difficulties of applying these techniques to unsteady state or rapidly changing processes. They conclude that for the technique to work, consistent process starting points, e.g feedstock composition are important, they also show that the technique is very robust to individual sensor failures.

Nomikos and MacGreggor (1994) show how PCA can be used to advantage in batch process monitoring. Their approach shows how closely PCA and traditional statistical process control techniques can be integrated. Nomikos's approach depends on the similarity of past batches to the batch under consideration. He also points out that fault events must be observable using the sensors applied to the process under consideration. His paper presents a form of control chart that is extensively used in this thesis.

2.4 Conclusions

This chapter has shown the main approaches available to an on line fault detection program:

- The use of statistical observers incorporating models of fault events.
- The use of heuristic or expert system techniques incorporating essentially non mathematical fault models.
- The use of statistical process control techniques, including principal component analysis.

The main disadvantages of the use of on line observers or expert system models is that a description of each of a large number of relatively infrequent faults is required to identify fault events.

The use of statistical process control methods only depend on statistical deviations from past experience. It will be remembered that the focus of this research is to identify techniques that will identify faults in a suite of research fermenters that will be continuously working on new processes and varying experimental conditions. The data presented in subsequent chapters will show an approach based on combining statistical process control methods with both mathematical descriptions and non mathematical knowledge of how fermentations work.

Chapter 3: Mathematical Description of a Penicillin Fermentation

The first requirement for an analytical redundancy system is an effective mathematical description of the fermentation dynamics. There are two aspects to consider:

- A model must be implemented, which is sufficiently descriptive to track a fermentation accurately enough to supply usable real time data.

- No model can account for all the potential variations in a real fermentation, so the model is built into an observer, which uses on-line observations from the real system to modify the trajectory the model predicts for the fermentation.

This chapter describes the basis for modelling penicillin fermentations. It goes on to describe the basis for using models in Kalman filters. Also the basis for neural network modelling is described.

3.1) Fermentation Modelling

One of the reasons for the choice of penicillin fermentations for the research work is that there is a number of well proven models available. In particular those derived by Bajpai and Reuss (1980, 1981), and Nestaas and Wang (1983) and subsequent developments of these models are widely used. The most suitable for the purpose of building an adaptive observer for penicillin is that described by Bajpai and Reuss (1981). It gives reasonable agreement with practically obtained fermentation data, without requiring the somewhat problematical parameters of the Nestaas and Wang model.

It should be noted that the authors of the model state that the model is expected to be strain dependent, therefore if specific industrial strains are to be tracked using this model, additional parameter identification is necessary. In addition it may be necessary to account for changes in the fermentation with time by varying some model parameters, as suggested by Bajpai and Reuss, although Montague (1986) suggests that this is not in practice necessary. There is a possible alternative to this, which is to use an adaptive observer based on on-line parameter identification, using

lower order models than that proposed by Bajpai and Reuss, such as has been described by Chattaway and Stephanopoulos (1989).

A fermentation is a very complex system to model. It is a population of cells that have distributions of physiological states, the behaviour of each cell is the result of many chemical reactions occurring both within it and in its environment, making a completely definitive model impossible to define. Models of varying degree of complexity have been applied to fermentation processes. Tsuchiya *et al.* (1966) have described a classification of fermentation models that is dependent on the detail with which they describe the biomass.

According to this classification, if no account is taken of distributions of size, structure, age, or other differentiating features of individual cells in a population, the model is classified as unsegregated. If such distributions are accounted for, the model is classified as segregated. If the model takes account of specific components of the cells metabolic machinery such as specific metabolic pathways, it is described as a structured model, otherwise it is described as unstructured.

Attempts to impose structure on models of the penicillin process frequently focus on macroscopic cell features, usually morphology (Nestaas and Wang 1983, Kluge 1992, Thomas 1993, Tiller 1994). The models take into account the activities of hyphal tips, which generate new biomass and produce penicillin, tubular sections of hyphae which do not grow but do produce penicillin, and vacuolised sections of hyphae which contribute to biomass measurements but are effectively inert. The simplest and probably most useful means of extending structure to the model described below would be to differentiate between live and dead biomass as described by Menszes *et al.* (1994).

Modelling of fermentation for applications such as monitoring and control, requires a compromise between accuracy and utility. The compromise is usually in favour of utility. It has been observed that the most widely used fermentation models are both unsegregated and unstructured (Dhurjati and Leipold, 1990). These models describe the fermentation in terms of the important macroscopic variables, i.e. those that can be measured and the relationships between which can be quantified. Usually they do utilise a priori knowledge about the biology of the fermentation system. This

differentiates models from input output correlation techniques such as neural networks.

3.2) The Process Model

A process model was sought that described the fed-batch penicillin fermentation in terms of the states of interest. The penicillin fermentation is a well studied system, it has been modelled for various purposes. Several of the models were simple, unstructured and unsegregated models in terms of linked kinetic equations describing at least the biomass concentration (Constantinides *et al.* 1970; Calam and Russell, 1973; Fishman and Biryukov, 1974; Bajpai and Reuss, 1980; Hegewald *et al,* 1981).

The kinetics used in these models reflected the particular application of the model. For example the model described by Constantinides *et al.* (1970) had temperature dependent parameters because the model was being used to find optimal temperature for the production of penicillin. Heijnen *et al.* (1979) described a complex and comprehensive model which was a hybrid of elemental balance and kinetic equations. Nestaas and Wang (1983) present a segregated model that held the biomass to exist as three types of cell: growing hyphal tips, penicillin-producing cells and inactive cells. Aynsley *et al.* (1990) also describe a segregated model for mycelial organisms, which was applied to a penicillin fermentation. Their model segregates the biomass by distinguishing productive tip cells from live and dead cells on the shaft of micelles .

Of these models, that developed by Bajpai and Reuss (1980) and later adapted by Bajpai and Reuss (1981) to ignore oxygen limitation was selected. There were three reasons for its selection. First, it describes the fed-batch penicillin fermentation in terms of the state variables of interest, biomass, glucose, and penicillin concentrations, and it includes the glucose feed and the consequent volume changes. Second, it has been successfully used to describe the fed batch penicillin fermentation for the evaluation of feed strategies by Bajpai and Reuss (1981), and optimisation of feed strategies by Tayeb and Lim (1986), who slightly modified the substrate concentration model, and San and Stephanopoulos (1989). Finally, it is typical of the process models used in statistical observers of the extended Kalman filter (EKF) type when these have been applied to fermentations. It is unsegregated and unstructured, and describes the relationships between the state variables using simple kinetics.

Montague *et al.* (1985; 1986 a and b) used it as the process model in their implementation of the extended Kalman filter.

The Bajpai and Reuss model is a Contois kinetic description of the growth (Contois, 1959) rather than the more usual Monod kinetic description. The basic Monod equation for specific growth rate takes the form of the Langmuir Absorption isotherm, and the Michaelis Menton equation for enzyme catalysis:

$$\mu_x = \frac{1}{x}\frac{dx}{dt} = \frac{\mu_{X\,max}.S}{K_s + S} \qquad 3.1$$

where

x	=	Biomass concentration g
S	=	Substrate concentration g/l
$\mu_{x\,max}$	=	Maximum specific growth rate g/(l.hr)
μ_x	=	Specific growth rate g/(l.hr)
K_s	=	Value of limiting nutrient at which the specific growth rate is half its maximum value.

In this equation, maximum growth is predicted as S becomes much larger than K_s. The Contois model develops the description of the fermentation by allowing for the diffusional limitations that occur at high biomass concentrations in mycelial fermentations, by making specific growth rate a reciprocal function of biomass:

$$\mu_x = \frac{1}{x}\frac{dx}{dt} = \frac{\mu_{X\,max}.S}{K_x x + S} \qquad 3.2$$

Where

K_x	=	Contois saturation constant for biomass g/g

The Contois equation therefore predicts that as biomass increases cell density becomes dominant, the equation then reduces to $\mu_x \propto x^{-1}$ Thus growth inhibition by biomass concentration is incorporated into the model.

The final form of the Bajpai and Reuss rate equation includes a term to account for the changes in broth volume that occur as a result of feeding and its effects on the predicted biomass concentration.

$$\frac{dX}{dt} = \frac{\mu_{X\max}.S.X}{K_X.X+S} - \frac{X}{V}.\frac{dV}{dt} \qquad 3.3$$

where

V = volume l
t = Time h

The rate of change of the penicillin concentration in the broth is modelled by Equation 3.4. This uses a substrate inhibition model to describe penicillin production, and a term to account for the hydrolysis of penicillin.

$$\frac{dP}{dt} = \frac{\mu_{P\max}.S.x}{K_P + S.(1+S/K_I)} - K.P - \frac{P}{V}.\frac{dV}{dt} \qquad 3.4$$

P = Penicillin concentration g/l
K_p = Saturation constant for Penicillin g/l
K_i = Substrate inhibition constant for penicillin synthesis g/l
K = Penicillin decay rate coefficient (h^{-1})

The rate of change of the broth substrate concentration is modelled by Equation 3.5. It is a mass balance using constant yield and maintenance terms, balancing the use of substrate for growth, production and maintenance with the substrate fed. This model assumes a single substrate; there is no consideration of the metabolism of carbon added to the fermentation in forms other than glucose feed. In use this equation is discretised, with the starting value of S set to take into account metabolisable carbon in the starting broth.

$$\frac{dS}{dt} = -\frac{1}{Y_{X/S}}.\frac{dx}{dt} - \frac{1}{Y_{P/S}}.\frac{dP}{dt} - m_X.x + F - \frac{S}{V}.\frac{dV}{dt} \qquad 3.5$$

where

F = Feed rate g/l/hr

$Y_{X/S}$ = Yield of biomass on substrate g/g

The rate of change of volume as a result of the substrate feed is given by Equation 3.6 This model does not account for the other causes of volume change such as evaporation, which are significant in both small and large scale fermentations. For the fermentations performed in this work, the volume change caused by the combined feeds of phenylacetic acid, ammonium hydroxide, and sulphuric acid was about the same as that caused by the glucose feed; and the total volume lost as sample could be as much as 40% of the final broth volume, because of the high level of data collection on these fermentations. Also, at times, it was necessary to increase sample size to maintain the broth volume within the working volume of the fermenter being used. The model also takes no account of evaporation, which is significant in a highly aerated system; evaporation could remove about 10% of total system water even with an exit gas stream condenser installed.

$$\frac{dV}{dt} = \frac{F.V}{S_F} \qquad 3.6$$

The Bajpai and Reuss model does not directly account for the effect of variations in the environmental variables: temperature, pH, and dissolved oxygen, on the growth and production of the culture. These can be taken into consideration, if data is available to permit modelling of the impact of environmental factors on the model parameters. The model also assumes a broth PAA concentration sufficient for penicillin G synthesis, but not at an inhibitory level.

3.2.1) The Relationship Between Carbon Dioxide Evolutuion (CER), Biomass and Penicillin Production.

Three versions of the relationship between the on-line measurable quantity (the carbon dioxide evolution rate) and the biomass, which is both the basis of the measurement equation for the extended Kalman filter are discussed in the literature:

$$\mathrm{CER} = \frac{1}{Y_{X/C}}.\frac{dX}{dt} + m_C.X + \frac{1}{Y_{P/C}}.\frac{dP}{dt} \qquad 3.7$$

$$CER = \frac{1}{Y_{X/C}} . \frac{dX}{dt} + m_C . X + k_P \qquad 3.8$$

$$CER = \frac{1}{Y_{X/C}} . \frac{dX}{dt} + m_C . X \qquad 3.9$$

CER = Carbon dioxide evolution rate l/l/hr

$Y_{X/C}$ = Yield of biomass on carbon g/g

All models explicitly include the carbon dioxide evolution that results from the growth and maintenance activities of the culture, but they differ in the inclusion of the carbon dioxide that results from the penicillin synthesis. The model shown in Equation 3.7 includes a term for carbon dioxide evolution that is dependent on the penicillin production rate. This type of model was used to describe the carbon dioxide produced in penicillin fermentations by Heijnen *et al.* (1979) and Cagney *et al.* (1984).

Calam and Ismail (1980) used the model shown in Equation 3.8 to estimate the carbon dioxide evolution rate in fed-batch penicillin fermentations. They stated that a penicillin production rate dependent term for carbon dioxide production was unnecessary because after an initial phase the penicillin production rate was constant. The penicillin production rate dependent term was replaced by a constant term to account for the carbon dioxide produced as a result of penicillin synthesis. The term was set equal to zero for the initial period of the fermentation during which no penicillin is detectable, and was invoked 15 hours after inoculation, on the onset of penicillin production.

This model was used for estimation of the biomass concentration in penicillin fermentations by Nelligan and Calam (1983) and Montague *et al.* (1985; 1986 a and b). The final model (Equation 3.9) does not include an explicit term for the carbon dioxide produced as a result of penicillin synthesis. The model therefore combines the penicillin synthesis and maintenance activities of the *Penicillium*. This model was used to estimate the biomass concentration in a penicillin fermentation by di Massimo et al. (1989) who considered carbon dioxide production as a result of penicillin

synthesis was insignificant. Tarbuck *et al.* (1985) also used this model to estimate the biomass concentration in a *Streptomyces clavuligerus* fermentation in which clavulanic acid was produced. The model shown in Equation 3.9 was selected for the work in this thesis because in the absence of a consensus it was the simplest model and therefore the easiest to implement.

3.3) Observer Systems For On-line Estimation Of Fermenter States

It is not practical to monitor most of the variables which are required to describe the state of a fermentation on-line, means have therefore been developed to infer these states from the measurements which can be made easily (Zabriski 1979, Montegue et al 1985, 1986). The most useful measurements to be made on-line are the off gas analyses, and the physical properties of the broth such as temperature, pH, and dissolved oxygen concentration. the level of CO_2 in the exit gas stream is particularly useful as it is a guide to the amount of energy being used by the cell population at any given time.

The object of an observer is to infer the values of states that previously were available off-line only, such as substrate, biomass or secondary metabolite concentration. These can then be used as a basis for a material and energy balance which can be compared with the one derived from the system inputs, or from off-line measurements, as the basis of a fault detection system. The estimates can also be used for process control purposes.

The particular observer used in this work was developed from the Bajpai and Reuss model together with the equation derived by Calam and Ismail (1980) which relates carbon dioxide production to biomass concentration. The principal feedback from the real system to the observer is via the carbon dioxide evolution rate as determined by a mass spectrometer. In addition since the model is based on concentrations, the fermenter broth volume is required. This is measured by monitoring the rate of input of process feeds, together with estimates of evaporation based on experience from prior fermentations and careful tracking of volume samples. A better solution would be to use load cell mounted fermenters, but none of the experimental equipment available to this project was equipped with this facility.

The biomass concentration and the growth rate are the most commonly estimated fermentation states because of their role in the understanding of a fermentation and therefore their common use as a control variable. If other states are estimated they are usually estimated in combination with the biomass. For fault detection purposes knowledge of the biomass concentration is useful because it can be independently measured off-line, and together with evolved carbon dioxide makes up the bulk of one side of an overall carbon balance, the other side being carbon inputs.

It has been observed by many workers (e.g., Zabriskie and Humphrey, 1978; Dekkers 1982; Wang and Stephanopoulos, 1986), that there are two basic methods of the indirect estimation of growth rate and hence biomass concentration in fermentation processes. Both of these methods are based on mass balances. The first of these methods was developed by Cooney *et al.* (1977) and Wang *et al.* (1977), and is based on the full stoichiometric balance of the fermentation. The other method, described by Zabriskie and Humphrey (1978), is a more simple balance based on the relationship between the growth rate and the rate evolution or uptake of a single substance, usually carbon dioxide or oxygen.

In real processes there is uncertainty, or noise, associated with both the on-line measurements and the relationship between the measurements and estimated states (Zabriskie, 1985). Stephanopoulos, (1984); and others (Johnson, 1987; Montague *et al.*, 1989 b) make a distinction between estimation methods for fermentations that is dependent on whether noise is considered. If the uncertainties are not considered the process is said to by deterministic if they are, the process is said to be stochastic. It is common practice in advanced process monitoring to consider the stochastic nature of the process and use estimation techniques designed to accommodate measurement and modelling uncertainties. The extended Kalman filter provides a convenient framework for accommodating these uncertainties.

3.3.1) On-line Estimation of Biomass Concentration by Mass Balances.

The method of on-line estimation using overall mass balances was developed by Cooney *et al.* (1977) and Wang *et al.* (1977); the description given here comes from Stephanopoulos and San (1984). A stoichiometric equation (Equation 3.10) is used to describe the overall reactions taking place in the fermentation. The stoichiometry in equation 3.10 describes a simplified fermentation system, a single substrate is

converted to biomass, water, and carbon dioxide, with ammonia as the only nitrogen source, and no secondary metabolite production:

$$\underbrace{a\,C_xH_yO_z}_{\text{substrate}} + b\,O_2 + c\,NH_3 \rightarrow \underbrace{C_\alpha H_\beta O_\gamma N_\delta}_{\text{biomass}} + d\,H_2O + e\,CO_2 \qquad 3.10$$

In this method it is assumed that all compositions are known and constant. The stoichiometric coefficients are normalised to those of biomass. Each of the four participating elements: carbon, hydrogen, nitrogen, and oxygen are balanced to give a set of solvable simultaneous equations. There is no assumption that the yield coefficients are fixed and therefore no use of kinetic models of the fermentation.

The problem with this approach, for an on-line system is the quantity of information that is required to solve the equations. To define the reaction, all significant, participating substances must be included in the description of the reaction and their chemical compositions accurately known. This information may be difficult to obtain for a practical fermentation process which may include complex nutrients of unknown composition, which are utilised at differing rates (for example most easily metabloisable carbon sources can be expected to be used up first). Typically, not all compositions are known, or even fixed throughout the fermentation.

In particular, the composition of biomass has been found to vary during some fermentations, e.g. the composition of *Penicillium chrysogenum* with respect to nitrogen and oxygen (Mou and Cooney, 1983 a), and the composition of *Saccharomyces cerevisiae* was found to vary with respect to nitrogen (Wang *et al.*, 1977). In the simple example of a fermentation reaction given no extracellular product formation, the measurements of carbon dioxide and oxygen are sufficient to close the system of equations.

If there is product formation, which was the case for the fermentations in this work, a term for the product has to be added to the right hand side of the reaction description. This adds another coefficient to the system of equations, and therefore an additional relationship is required for their solution. Wang and Stephanopoulos (1986) have suggested that this relationship may be obtained from measurements of the nitrogen

source, carbon source, product, or heat of fermentation with an enthalpy balance. In practice there are further complications in real systems such as penicillin fermentations, such as the presence of degradation products from penicillin, or phenyl acetic acid.

Mass balances are therefore difficult to perform accurately on non limiting species, such as nitrogen in a carbon limited fed batch fermentation. This is because the nitrogen reservoir in the fermentation broth would have to be quantified, while during steady state growth the level of metabolisable carbon in the broth is minimal. Fixed carbon in product, biomass and phenylacetic acid can be readily quantified, therefore carbon balances are used in this thesis as the basis of mass balancing techniques.

3.3.2) Mass Balances

An alternative way of describing a fermentation is by conducting a mass balance based on the fermenter inputs and outputs, and either off-line measured states or inferred states from the observer. With the information from the observer, from models, or from past fermentation runs, the system becomes overdefined and it is possible to test the consistency of the information set thus obtained. The methods for elemental balancing around a fermenter are described by Wang and Stepanopoulos (1983), together with a means of isolating the probable source of the inconsistency.

The fermentations on which this work was based were carbon limited, therefore this approach avoids the need to have an accurate method of measuring broth carbon, other than in product. To carry out a carbon balance, input feed rates and compositions must be known. Although not necessarily on-line, broth analyses must be performed at reasonable frequencies, and to be of any use the results must be available within a small fraction of the total fermentation time.

The practical use of mass balances in fault detection is therefore limited to continuous or fed batch fermentations. Account must be taken of the change in the elemental composition and structure of *Penicillium chrysogenum* over time. Details of these changes are available from the literature (Mou and Cooney 1983). They show that carbon concentration remained constant during a fermentation. Mou and Cooney used this element to carry out their mass balances for on-line growth monitoring. The second method for the indirect estimation of biomass is based on an empirical

correlation between the formation, or utilisation, of an on-line measurable parameter and an unmeasurable state such as biomass (Zabriskie 1978).

The obvious selections for the on-line measurable parameter are carbon dioxide formation or oxygen utilisation, these are linked as both are measures of energy being used in the fermentation, and most of the oxygen used ends up in the exit carbon dioxide. The value of oxygen uptake rate is usually a good measurement of the quantity of oxygen being used by the culture because of its low solubility in aqueous media. Carbon dioxide is much more soluble in aqueous media and the solubility is a function of the pH of the media. In most practical fed batch fermentation systems however, the pH is controlled at a fixed level, the hold up of carbon dioxide does not therefore change.

The carbon dioxide evolution rate was the on-line measurement used in this work, and, therefore the following description of the method is in terms of carbon dioxide and carbon dioxide evolution rate. However, in all cases an equivalent relationship in terms of oxygen and oxygen uptake rate could have been be formulated. In the simplest form of the method the growth rate can be estimated from the carbon dioxide evolution rate using the relationship shown in Equation 3.11, in which $Y_{X/C}$ is the yield of biomass on carbon dioxide, and is calculated as the ratio of total biomass accumulated to total carbon dioxide produced, in a given period. In this form it can be seen to be equivalent to a stoichiometric relationship between the biomass and carbon dioxide. After rearrangement, measurements of carbon dioxide evolution rate can be used directly to estimate the growth rate or integrated to give estimates of the biomass concentration.

$$CER = \frac{1}{Y_{X/C}} . \frac{dX}{dt} \qquad 3.11$$

CER	=	Carbon dioxide evolution rate l/l/hr
$Y_{X/C}$	=	Yield of biomass on carbon g/g
X	=	Biomass g(dry weight)
t	=	Time h

This simple method has been used to estimate biomass during the growth phase of a range of fermentations. Mou and Cooney (1983 a) used it successfully to estimate the

biomass concentration during the growth phase of fed-batch penicillin fermentation. The same method, but in terms of oxygen and the oxygen uptake rate, was used by Buckland et al. (1985) and Gbewonyo *et al.* (1989) to estimate the biomass concentration in batch avermectin fermentations. They used the oxygen version because of the large variation in pH that occurs during their fermentation. Kennedy *et al.* (1992) used the method in terms of carbon dioxide and carbon dioxide evolution rate to estimate the biomass concentration in a batch fermentation of *Bacillus subtilis.*

The constant relationship between the growth and the carbon dioxide formed only exists when growth is the only significant cellular activity, or where the ratio of carbon dioxide evolved from maintenance activity to that evolved due to growth is fixed, this occurs during the exponential growth phase of batch and fed-batch fermentations, when the fermenter is operating with no nutrient limitation. After the exponential growth phase has finished, fed-batch fermentations characteristically enter a lower growth rate phase during which it has been shown that the maintenance activities of the biomass predominate (Pirt, 1965).

To account for carbon dioxide formed as a result of the maintenance activities, Zabriskie and Humphrey (1978) suggested the form of relationship between carbon dioxide evolution rate and biomass in Equation 3.12, where m_C is the maintenance of biomass on carbon dioxide.

$$CER = \frac{1}{Y_{X/C}} . \frac{dX}{dt} + m_C . X \qquad 3.12$$

m_c = Maintenance coefficient g/g/hr

Zabriskie and Humphrey (1978) used a variant of equation 3.12 cast in terms of oxygen and oxygen uptake rate. The oxygen form was used both because of the insensitivity of oxygen uptake rate to pH fluctuation and because oxygen uptake is more closely related to energy production, and thus cellular activity, than the formation of carbon dioxide. The method was successful in estimating the biomass concentrations of simple batch fermentations of a *Thermoactinomycetes* sp. and a *Streptomyces* sp. It was less successful in estimating the biomass in a more complex *Saccharomyces cerevisiae* fermentation.

As with any model, to be useful values for the yield and maintenance parameters have to be available, these have to be obtained experimentally. The parameter values are also calculated using off-line measurements of biomass concentration. Wang and Stephanopoulos (1986) have criticised this aspect of the method, pointing out that the accuracy of the biomass estimates are heavily dependent both on the accuracy of the off-line biomass measurements and the ability to replicate the fermentation conditions. Nevertheless, this correlation method has found wide application because of its simplicity.

The two estimators described above are deterministic, they do not consider the uncertainties that are always associated with a real fermentation process. It has been observed by many authors that there are uncertainties associated with the measurements made on fermentation processes, including the on-line measurements that are used to drive estimators (Stephanopoulos and San, 1981; van der Heijden *et al.*, 1989).

To produce estimates of biomass concentrations, an initial value of the biomass concentration is required by both deterministic estimation methods. The initial value is often an estimate which also has an uncertainty associated with it. Montague *et al.* (1989 b) comment that an error in the initial estimate of the biomass concentration can be serious for the subsequent estimation, particularly if the growth rate is high.

Stephanopoulos and San (1981) showed that when a noisy measurement is integrated to produce an estimate, the variance of the estimate increases from that associated with the initial estimate at a rate dependent on the size of the measurement variance. The variance of the estimate may become so large that the estimate becomes useless. Wang *et al.* (1977) identified the build up of measurement error as the cause of the poor estimation of biomass towards the end of a *Saccharomyces cerevisiae* fermentation, using the full stoichiometric estimation method.

Another source of uncertainty is the relationship between the on-line measurement and estimated variables. For the methods already described, this is either the stoichiometric description of the fermentation reaction; or the model relating the carbon dioxide evolution rate to the growth and maintenance activities of the biomass.

Stoichiometric descriptions are subject to error due to inaccuracies in knowledge of component concentrations, or the composition of components. For the growth-maintenance model (Equation 3.12), as with most practical kinetic models applied to fermentation processes, a grossly simplified description of the production of carbon dioxide by the culture, and take no account of variations due to changes in the fermentation during its time course.

To take account of the stochastic nature of a process, an estimator is used. These estimators are called filters because they not only produce an estimate of an unmeasurable state, but they also extract the estimate from the various uncertainties (noises). The Kalman filter, developed by Kalman (1960) and Kalman and Bucy (1961), has been applied to fermentation processes. This filter is designed to produce optimal estimates of linear systems, but as observed by Stephanopoulos and San (1984) most systems of practical interest, which include fermentation processes, are non-linear. Jazwinski (1970) extended the Kalman filter to non-linear systems by the development of the extended Kalman filter. Other non-linear filters have been developed, e.g. Wishner *et al.* (1969), but these are usually more complicated than the extended Kalman filter.

3.3.3) The Extended Kalman Filter

Before the extended Kalman filter itself is explained it is necessary to describe the general, non-linear system on which the explanation is based. The system is assumed to have two components: the states, which include all variables that define the process; and the measurements that are available to drive the estimation. The state of the non-linear process is assumed to be modelled by the following equation

$$\mathbf{x}(k+1) = \mathbf{f}(k, \mathbf{x}(k)) + \mathbf{w}(k) \qquad 3.13$$

where $\mathbf{x}(k)$ is the state vector which contains all the state variables; $\mathbf{f}(k)$ is a vector that contains the relationships between the state variables, it is the model of the system; and $\mathbf{w}(k)$ is the process noise vector, which contains the uncertainties associated with the model of each state. The relationship between the available measurements and the process state vector are assumed to be modelled by the following equation:

$$\mathbf{z}(k) = \mathbf{h}(k, \mathbf{x}(k)) + \mathbf{v}(k) \qquad 3.14$$

where $\mathbf{z}(k)$ is the measurement vector which contains all the measurements that are used to drive the estimation; $\mathbf{h}(k)$ is vector containing the relationships between the available measurements and the state variables; and $\mathbf{v}(k)$ is the measurement noise vector, which contains the uncertainties associated with the measurements. Both $\mathbf{w}(k)$ and $\mathbf{v}(k)$ are assumed to be zero mean, Gaussian (normally distributed) noise processes, which are independent of each other.

The linear Kalman filter is designed to produce the minimum variance estimate of the state of a linear system. The equations for the extended Kalman filter are the same as those for the linear Kalman filter, but the extended Kalman filter applies them to linear approximations of a non-linear system. The series of equations that constitute the extended Kalman filter are given in Equations 3.15-3.20 in the 'predictor-corrector' format (the presence of a circumflex (^) over a vector indicate estimated quantities); and the linearisations of the state and measurement models are given in Equation 3.22 and 3.23.

The form of the equations presented here was taken from Lichfield (1979)

Prediction:

prediction of the state vector:

$$\hat{x}_{k/k-1} = f_{k-1}(\hat{x}_{k-1/k-1}) \qquad 3.15$$

prediction of the measurements

$$\hat{Z}_{k/k-1} = h_k \hat{x}_{k/k-1} \qquad 3.16$$

prediction of the estimation error covariance matrix

$$P_{k/k-1} = \Phi_{k-1} P_{k-1/k-1} \Phi^{\mathrm{T}}{}_{k-1} + Q_{k-1} \qquad 3.17$$

This step predicts

Correction:

calculation of the Kalman gain

$$K_k = P_{k/k-1} H^{\mathrm{T}}{}_k (H_k P_{k/k-1} H^{\mathrm{T}}{}_k + R_k)^{-1} \qquad 3.18$$

state estimation - correction of the prediction

$$\hat{x}_{k/k} = \hat{x}_{k/k-1} + K(k)(z_k - z_{k/k-1}) \qquad 3.19$$

correction of the estimation error covariance matrix

$$P_{k/k} = (I - K_k H_k) P_{k/k-1} \qquad 3.20$$

The linearisation of the system:

In the Extended Kalman Filter, the Jacobian Φ is a matrix of linearised process models

$$\mathbf{\Phi}(k-1) = \left.\frac{\partial \mathbf{f}(k-1)}{\partial \mathbf{x}(k-1)}\right|_{\hat{\mathbf{x}}(k-1|k-1)} \qquad 3.21$$

this may be expanded thus:

$$\mathbf{\Phi} = \begin{pmatrix} \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_n} \end{pmatrix}_{\hat{X}(k/k-1)} \qquad 3.22$$

The terms can be derived by taking the partial differentials of a Taylor series expansion of the equations comprising the process model. The measurement model is of a similar form:

$$\mathbf{H}(k) = \left.\frac{\partial \mathbf{h}(k)}{\partial \mathbf{x}(k)}\right|_{\hat{\mathbf{x}}(k|k-1)} \qquad 3.23$$

The first step in the estimation is to predict the state vector from the previous estimate of the state vector using the non-linear process model (Equation 3.15). The prediction of the state vector is then converted by the non-linear measurement model to a prediction of the measurement vector (Equation 3.16).

The next step in the estimation is to predict the state estimation error covariance matrix $P_{k/k-1}$. The value of the previous estimation error covariance $P_{k-1/k-1}$ is propagated over the interval using a linear approximation to the non-linear process (Equation 3.17). This requires the Jacobian Φ_{k-1}, which is the linear approximation of the process model, to be evaluated at the previous state estimate (Equation 3.20). The process noise is included at this stage by the simple addition of the process noise covariance matrix Q_{k-1} to the propagated estimation error covariance matrix.

The predictions are then corrected to produce the estimate of the state vector and the measurement error covariance matrix. The first step in the correction is to calculate the Kalman gain matrix K_k (Equation 3.18). The form of the Kalman gain is designed to minimise the mean square error between the estimate and the true value of the state.

The calculation of the Kalman gain requires the Jacobian H_k, which is the linear approximation of the measurement model evaluated at the predicted state. The value of the Kalman gain is dependent on the prediction of the estimation error covariance matrix, which includes a consideration of the process noise, and inversely proportional to the covariance of the measurement errors, represented by the measurement error covariance matrix R_k. Thus, high values of estimation error covariance and low values of measurement noise will tend to increase the weighting on the correction; indicating that the model based predictions are poor and in need of correction, and that the measurements are reliable and should be used in the correction.

The second step in the correction is the estimation of the state itself, which is a combination of predicted state vector and difference between the predicted values of the measurements and the measurements themselves (these are known as the residuals), weighted by the Kalman gain (Equation 3.17). At this stage both estimation and filtering occur. The Kalman gain is the mechanism which relates the measurements to the states. The weighting of the correction is also dependent on the covariances of the noises that were used to calculate the Kalman gain. The final step

in the estimation process is to correct the prediction of the estimation error covariance matrix using the Kalman gain (Equation 3.18).

To initiate the estimation process it is necessary to have an initial estimate of the state vector $\hat{X}$, with an associated error covariance ($\mathbf{P}_0$), which is independent of both process and measurement noise processes. The value of the initial estimate is, within reason, not critical but the uncertainty in it should be reflected in the associated $\mathbf{P}_0$ matrix.

The extended Kalman filter is termed a suboptimal filter (Jazwinski, 1970) because it produces the optimal estimate of the linearised system, which is just an approximation of the real non-linear system. In an effort to reduce the approximation errors inherent in the linearisation of the system, the non-linear nature of the system is exploited when possible: the non-linear models are used in the prediction steps. Furthermore, the process and the measurement models are linearised (the Jacobians $\mathbf{\Phi}_{k-1}$ and $\mathbf{H}_k$ respectively are evaluated) using the best estimates of the state available, which is the reason they are associated with different time indices. The filter equations are also the result of many simplifications that are heavily dependent on the assumptions made about the characteristics of the noise processes, i.e. that they are zero mean Gaussian and that they are not correlated with each other (Bozic, 1979).

The action of the extended Kalman filter can also be seen as that of combining system information available from a model with that available from measurements. The weighting of the correction reflects the relative confidence in the model and the measurements, as quantified by the covariance matrices. A poor description of the process by the model will have a process noise covariance matrix with elements that have large values, which will result in larger values of the Kalman gain and, consequently an increased weighting of the measurement information. Poor measurements will have a measurement noise covariance matrix with large elements, which will result in lower values of the Kalman gain and a decreased weighting of the measurement information.

3.3.3.2) The Use of the Kalman Filter In Fermentation Systems.

The observer used in this work is therefore an extended Kalman filter of the predictor corrector form. This is the most common form of the extended Kalman filter

described in the literature (Jaswinski 1970, Lichfield 1979), it has the advantages of ease of implementation and interpretation.

The process model used was the Bajpai and Reuss model (Equations 3.3-3.6) less equation 3.4, the term for penicillin. This is because there is no adequate connection between available on line measurements (volume and carbon dioxide evolution) and penicillin production. Thus the four states (parameters the filter is designed to track or observe) included in the state vector X are biomass, substrate concentration, volume and carbon dioxide evolution rate. The components of the observation vector Z are volume and carbon dioxide evolution rate. Of these, carbon dioxide evolution rate in practice contains the available information about metabolic activity used to correct the filter output.

The extended Kalman filter linearises its model equations about the last state estimate, by means of a truncated Taylor expansion. In order to cope with a discrete time based system they were discretised using first order Euler equations. These have the following form:

$$x_{k+1} = x_k + \Delta t.\frac{dx_k}{dt} \qquad 3.24$$

where

x = Generic parameter
t= Time

Applying these to the Bajpai and Reuss model (equations 3.3-3.5) and the associated volume term (3.6), the following equations are obtained:

$$X_{k+1} = X_k + \Delta t.\left[\frac{\mu_X S_k X_k}{K_X X_k + S_k} - \frac{X_k F}{S_F}\right] \qquad 3.25$$

$$S_{k+1} = S_k + \Delta t.\left[-\frac{\mu_x S_k X_k}{Y_{x/s}(K_x X_k + S_k)} - \frac{\mu_p X_k S_k}{Y_{p/s}\left[S_k + S_k\left(1 + {}^{S_k}\!/_{K_I}\right)\right]} - m_X X_K - \frac{S_k F}{S_F}\right] \qquad 3.26$$

$$V_{k+1} = V_k + \Delta t.\left[\frac{FV}{S_F}\right] \tag{3.27}$$

$$CO_{2k+1} = CO_{2k} + \Delta t.\left[\frac{\mu_x S_k X_k}{K_4(K_x X_k + S_k)} + m_c X_k + K_5\right] \tag{3.28}$$

The process model Jacobian (Φ) comprises the partial differentials of the Eulers with respect to each state:

$$\Phi_{k+1} = \begin{bmatrix} \frac{\partial X_{k+1}}{\partial X_k} & \frac{\partial X_{k+1}}{\partial S_k} & \frac{\partial X_{k+1}}{\partial V_k} & \frac{\partial X_{k+1}}{\partial CO_{2k}} \\ \frac{\partial S_{k+1}}{\partial X_k} & \frac{\partial S_{k+1}}{\partial S_k} & \frac{\partial S_{k+1}}{\partial V_k} & \frac{\partial S_{k+1}}{\partial CO_{2k}} \\ \frac{\partial V_{k+1}}{\partial X_k} & \frac{\partial V_{k+1}}{\partial S_k} & \frac{\partial V_{k+1}}{\partial V_k} & \frac{\partial V_{k+1}}{\partial CO_{2k}} \\ \frac{\partial CO_{2k+1}}{\partial X_k} & \frac{\partial CO_{2k+1}}{\partial S_k} & \frac{\partial CO_{2k+1}}{\partial V_k} & \frac{\partial CO_{2k+1}}{\partial CO_{2k}} \end{bmatrix} \tag{3.29}$$

The elements of the Jacobian are as follows (example derivations indicated).

$$\Phi_{1,1} = \frac{\partial X_{k+1}}{\partial X_k}$$

$$\Phi_{1,1} = 1 + \Delta T.\left[\frac{\mu_x S_k K_x X_k + \mu_x S_k^2 - K_x S_k X_k \mu_x}{(K_x X_k + S_k)^2} - \frac{F}{S_F}\right]$$

$$\Phi_{1,1} = 1 + \Delta T.\left[\frac{\mu_x S^2}{(K_x x + S)^2} - \frac{F}{S_f}\right] \tag{3.30}$$

$$\Phi_{1,2} = \frac{\partial X_{k+1}}{\partial S_k}$$

$$\Phi_{1,2} = \Delta t\left[\frac{\mu_x X_k^2 K_x + \mu_x X_k S_k - \mu_x S_k X_k}{(K_x X_k + S_k)^2}\right]$$

$$\Phi_{1,2} = \Delta t.\left[\frac{\mu_x X^2 K_x}{(K_x X + S)^2}\right] \tag{3.31}$$

The following equations were similarly derived:

$$\Phi_{1,3} = 0 \qquad 3.32$$

$$\Phi_{1,4} = 0 \qquad 3.33$$

$$\Phi_{2,1} = \Delta t.\left[-\frac{\mu_x S^2}{Y_{x/s}(K_x x + S)^2} - \frac{\mu_p S}{Y_{p/s}(K_p + S + S^2/K_i)^2} - \frac{F}{S_F}\right] \qquad 3.34$$

$$\Phi_{2,2} = 1 + \Delta t.\left[-\frac{\mu_x x^2 K_x}{Y_{x/s}(K_x x + S)^2} - \frac{\mu_p S}{Y_{p/s}(K_p + S + S^2/K_I)^2} - \frac{F}{S_F}\right] \qquad 3.35$$

$$\Phi_{2,3} = 0 \qquad 3.36$$

$$\Phi_{2,4} = 0 \qquad 3.37$$

$$\Phi_{3,1} = 0 \qquad 3.38$$

$$\Phi_{3,2} = 0 \qquad 3.39$$

$$\Phi_{3,3} = 1 + \Delta t.\frac{F}{S_F} \qquad 3.40$$

$$\Phi_{3,4} = 0 \qquad 3.41$$

$$\Phi_{4,1} = \Delta t.\left[\frac{\mu_x S^2}{K_4(K_x X + S)^2} + m_c\right] \qquad 3.42$$

$$\Phi_{4,2} = \Delta t.\left[\frac{\mu_x X^2 K_x}{K_4(K_x X + S)^2}\right] \qquad 3.43$$

$$\Phi_{4,3} = 0 \qquad 3.44$$

$$\Phi_{4,4} = 1 \qquad 3.45$$

Since volume and CER are states in this system, the measurement Jacobian is:

$$H = \begin{vmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \qquad 3.46$$

The relative performance of deterministic model, and the filter is discussed in chapter 7. There are several points to be made about the Kalman filter. First, it requires knowledge of the accuracy of the sensor information being fed back into the model, the level of noise inherent in the process, and the accuracy of the model, in terms of parameter variances. An assessment of the inputs to the filter is presented in section 4.5.

The extended Kalman filter has been commonly applied to the estimation of the biomass concentration from on-line measurements of carbon dioxide evolution rate or oxygen uptake rate. In some cases additional on-line measurements are used such as ethanol production rate and nitrogen consumption for pH control. The exceptions, those reports that do not estimate biomass, either have, or assume in simulated fermentations, some measurement of biomass allowing them to estimate other states and parameters.

The extended Kalman filter equation system requires a process model that describes the fermentation in terms of the states of interest. These are usually the concentrations of biomass, substrate, and, in some cases, the product. The process model is in the form of several linked differential equations, one for each state, which describe the rates of growth, substrate uptake, and product synthesis, in terms of the states of interest and appropriate parameters. These equations (3.25-3.28) are the elements of the vector **f**

As discussed above, the model is a simplification of what is really happening in a fermentation. The whole fermentation is described by lumped, macroscopic variables with no consideration of intra-cellular activity or the distribution of properties within a population of cells of various ages, i.e. the model is unstructured and unsegregated.. This is typical of models used in estimation, Jazwinski (1970), commented that lower

order approximations of complex processes would have to be used with an extended Kalman filter. Staniškis and Simutis (1986) recommend an upper limit of six estimated states, fewer than this are usually seen in application of the extended Kalman filter to process problems.

In addition to the process model the extended Kalman filter requires a measurement model that relates the state variables of interest to the available on-line measurements, i.e. makes the state variables observable. An observable state variable is one that can be determined from the information present in the on-line measurements. If there is no relationship between a state variable and possible on line measurements, then that state variable is unobservable.

In this thesis, where biomass is the most important estimated state and the carbon dioxide evolution rate is the principal on-line measurement, there are two available options that are related to the two deterministic estimators of biomass described above. One option is to convert the measurements to be comparable with the process model predictions. The carbon dioxide evolution rate is converted into a value of the growth rate using the full stoichiometric balance that can be compared with the process model prediction of growth rate. This method was developed and used by Stephanopoulos and San (1984) and San and Stephanopoulos (1984), and subsequent examples of its use have been reported by Shioya *et al.* (1986), Náhlik and Burianec (1988), and Dubach and Märkl (1992).

The other option is to convert the process model predictions to be comparable with the on-line measurements. The process model predictions of growth rate and biomass concentration can be converted to an estimate of the carbon dioxide evolution rate using a model of the type shown in Equation 3.12. Examples of the use of this method include Montague *et al.* (1986 a, b, and c), Tarbuck *et al.* (1985), Tarbuck *et al.* (1986) and Yu *et al.* (1987).

Both the process model and the measurement model, if they are in the form of the differential equation relating biomass to the carbon dioxide evolution rate, include various kinetic parameters such as growth rates, yields, and maintenance coefficients. It has been observed (van der Heijden *et al.* 1989) that one of the most significant problems associated with implementing an extended Kalman filter on a fermentation process is poor estimation caused by erroneous parameter values. Poor parameter

values may be the result of inadequate identification methods or, commonly in fermentations, variable parameter values.

Box *et al.* (1978) say that model parameters vary when some of the dynamics in systems of interest are not accounted for, they vary to compensate for the undescribed processes. The parameters are functions of fermentation states (e.g. yield on carbon is a function of biomass and substrate concentration). A complication is that some parameters are likely to be variable with time, such as the ratio of active penicillin producing hyphal tips to essentially inactive biomass). The presence of variable parameters in a system will make that system non-linear. The distinction between truly non linear and time varying parameters is of little practical importance in parameter estimating filters.

As a non-linear estimator the extended Kalman filter can be used to identify variable parameter values on-line as well as perform state estimation. Parameter identification has been cited as major advantage of using an extended Kalman filter in the estimation of fermentation processes (Stephanopoulos and Park, 1991), and Náhlik and Burianec (1988) see it as necessary to the estimation of a fermentation process. To perform parameter identification, the state vector is augmented to include the variable parameters to be identified.

When no model for the parameter variation is available, it can be modelled as random noise process. This was done successfully by Stephanopoulos and San (1984) in the estimation of the specific growth rate and yield of biomass on substrate. This method was also followed by Shioya *et al.* (1986), Yu *et al.* (1987), Náhlik and Burianec (1988), and Dubach and Märkl (1992). Leigh and Ng (1984) and later Tarbuck *et al.* (1985) and Tarbuck *et al.* (1986) report the adaptation of a parameter in their extended Kalman filter, however they do not give details. The augmentation of the state vector increases the computational burden, and Stephanopoulos and Park (1991) have suggested that the sensitivity of the estimation to the parameter values should be determined to indicate the most influential parameters and therefore those most usefully identified.

The extended Kalman filter requires the process and measurement noise covariance matrices (Q and R respectively) to be specified. This can be done by either determining the process and measurement noise processes or by adjusting the values

until satisfactory performance is observed, using historical data or simulated fermentations. In reports of the extended Kalman filter, where the specifications of these matrices are discussed, there appears to be a consensus that the measurement noise and, therefore the R matrix should be determined by experiments on the measurement instruments or assays, and the Q matrix found by experimentation. This is an acceptance that the filter is not really operating as an optimal filter. This non optimality is implied in any case by the impact of the linearisation of the model equations, which invalidates the assumptions around the Gaussian distribution of errors.

The determination of the uncertainties associated with typical fermentation measurements is difficult. This is because measurements made on fermentation processes are commonly either manual assays or composite quantities, which do not lend themselves to the statistical analysis required to characterise and quantify their distributions. It is probably as a result of this that is common to find reports that, in fermentation applications, both the Q and R matrices are determined by 'experience', 'trial and error', or by 'simulation', i.e. they are tuned. No guidelines exist for the tuning of the extended Kalman filter and this has been identified as a major problem by Stephanopoulos and Park (1991).

Methods have been developed for the on-line tuning, or adaptation, of the noise covariance matrices (e.g. Sage and Husa, 1969), which vary the elements of matrices, on-line, in response to process performance. These methods have not been widely applied to fermentation process, only Stephanopoulos and San (1984) and Swinarski et al. (1982) report their use.

3.4) Neural Networks

An alternative way of observing a fermentation is by the use of neural networks. The use of neural networks in observing chemical processes including fermentations is described by a number of authors (Montague 1994, Thompson 1994, Bakshi 1993, Di Massimo 1992, Stepanopoulos 1990). This technique has been criticised by Montague (1994) as being unstable in the presence of significant process and strain variations. It was thought that for this work, observers which incorporated *a priori* knowledge in the form of process models would be a more useful tool for fermentation tracking.

A method for incorporating process models for fed batch penicillin fermentations was presented as a case study by Thompson and Kramer (1994). In this work it is shown that a high degree of correlation can be obtained between observed fermentation data and a hybrid model / neural network estimator. They also show that pure neural networks perform relatively badly as fermentation state estimators. In this case the neural network is being trained to determine a correction factor for the difference between the model and observed data.

In this thesis the error between modelled and observed fermentation parameters was itself treated as a subject for statistical process control. The additional complexity of going to Thompson and Kramer's approach could not be adopted for time reasons. The use of their approach to generate further data which could be used in an analytical redundancy scheme would be a useful next step.

3.5 Conclusions

The development of a model which fully describes a fermentation system, but which is of use for process monitoring and control is not feasible for other than model systems. This is because of the very large effort in parameter identification which would be required. For most on line operations therefore, relatively simplistic non segregated models have been deployed.

The accuracy of these models in predicting the state of a specific fermentation can in principle be improved by comparing model predictions with predictions from the fermentation, and using any discrepancies to correct the values of states in the model prior to its next iteration. This provides a feedback mechanism for the correction of the model's predictions for unmeasurable states such as biomass concentration based on measurable states such as carbon dioxide evolution. This approach was shown to be feasible for varied penicillin processes (chapter 7). The efectiveness of filtering is analysed in chapters 7,8 and 9.

An alternative approach, employed by neural networks, is to eliminate any attempt to describe biological processes within the model and instead use purely statistical means to relate system inputs and outputs over time. Neural networks require a consistent

process, so it was not used in this project as this is not consistent with the experimental programs likely to be used in research fermenters.

Chapter 4 Materials And Methods

The micro-organism used in this work, *Penicillium chrysogenum PC2*, was chosen because it is typical of many of the organisms used by the project sponsor (SmithKline Beecham plc.). The method used to carry out penicillin fermentations was based on the process described by Mou and Cooney (1983). This enabled the project to benefit from the work that Mou and Cooney carried out to identify the parameters in the model described by Bajpai and Reuss (1980, 1981).

This chapter describes in detail the fermentation processes used, and the analytical methods used to generate data for this project. These methods were chosen to give comprehensive data sets, using analytical methods typically employed in research and development fermenter suites of the type used by the project sponsor. In addition data were obtained from fermentations in which changes simulating faults, and in some cases real faults occurred.

An evaluation of the variability of data generated from the analytical methods chosen was made, this analysis is presented in section 4.5 of this chapter. This information is not typically available in the literature.

4.1 Materials

4.1.1 Micro-organism:

A single microbial strain, *Penicillium chrysogenum PC2*, was used in all fermentations. The original spore stock was provided by Dr.T.Keshavarz (University College London). This organism was chosen because of the similarity of its fermentation process to industrially useful fed batch fermentations.

4.1.2 Media

The media used were drawn from the paper by Mou and Cooney (1983), modified somewhat by advice from fermentation experts at UCL (Dr T Keshavarz, and Mr.R.Elgin). The medium ingredients used are listed on the following page.

In addition, materials used for analytical purposes, for spore suspension and for pH control of fermentations are listed.

Table 4.1 Raw Materials Used In fermentation Experiments

Item	Origin	Conc. in Spore Media g/l	Conc. in Seed Media g/l	Conc. in Ferm. Media. g/l
NH_4 OH	Aldrich Chemical Co. Ltd. Gillingham, Dorset, UK.			Fed
KH_2PO_4	Aldrich			0.453
K_2HPO_4	Aldrich	0.06		0.4
$CaSO_4.2H_20$	Aldrich			0.05
KOH	Aldrich			Fed
Molasses	Community Foods, London, UK	2.5		
$MgSO_{4(anh)}$	FSA Lab Supplies, Loughborough, UK.	0.05	0.24	0.115
$(NH_4)_2SO_4$	FSA Lab Supplies			0.227
Glycerol	FSA Lab Supplies	7.5		
$MnSO_4.4H_2O$	FSA Lab Supplies			0.005
$ZnSO_4.7H_2O$	FSA Lab Supplies			0.0034
NaCl	FSA Lab Supplies	10.0		0.057
H_2SO_4 (98.8%)	FSA Lab Supplies			Fed
$FeSO_4.7H_20$	FSA Lab Supplies	0.003		0.057
$CuSO_4.5H_2O$	FSA Lab Supplies	0.001		0.001
PAA	FSA Lab Supplies			
Henkel Nopco TDBI Antifoam	Supplied by SmithKline Beecham			0.1
No.3 Agar	Oxoid Ltd UK	20.0		
Nutrient Agar†	Oxoid Ltd			
Yeast Extract powder	Oxoid Ltd	1.0		

Item	Origin	Conc. in Spore Media g/l	Conc. in Seed Media g/l	Conc. in Ferm. Media. g/l
Peptone (bacteriological)	Oxoid Ltd	5.0		
Item	Origin	Conc. in Spore Media g/l	Conc. in Seed Media g/l	Conc. in Ferm. Media. g/l
Spray Dried Corn Steep Liquor	Roquette Freres, Lestrom, France via Glaxo Group Research, UK.		1.28	0.227
Na_2SO_4	Sigma Chemical Co UK		0.30	
Glucose $_{(anh.)}$	Sigma		2.0	0.8
Tween 80	Sigma			
Water	Deionised at UCL	Balance	Balance	Balance

† Used in master spore slope preparation, and for contamination testing.

4.3 Equipment

4.3.1 Fermentation Equipment

All fermentations were carried out in 7 litre or 1.5 litre vessels, supplied by LH Fermentation, Stoke Poges, U.K.. The 7 litre vessel was used at UCL, three 1.5 litre vessels were used at SmithKline Beecham.

These vessels were baffled and equipped with Rushton Turbine agitators to permit good distribution of sparged air. Both fermenters were equipped with electric heating and water cooling systems. The 7 litre fermenters were sterilised by use of their electric heating system. The 1.5 litre systems could be placed in an autoclave for sterilisation.

4.3.1.1 Fermenter Instrumentation And Control System

The fermenters were equipped with TCS 6358 eight loop PID controllers. This system was used to control and log air flow rate, pH, temperature, impeller speed. The system at UCL could also log the speed of the Watson Marlow peristaltic pumps used to feed glucose and PAA. The smaller scale of the fermenters at SmithKline Beecham necessitated the use of syringe pumps for glucose and phenylacetic acid feed, as peristaltic pumps with a low enough flowrate were not available at SmithKline Beecham. The reason for this was that syringe pumps were more accurate than peristaltic pumps for very low flowrates.

Air flow into the fermenter was measured using a thermal mass flow meter (HI Tech F100 Bronkhorst High Tech B.V., Ruurlo, Netherlands). Broth temperature was measured using a resistance thermometer fitted into a thermowell. pH and dissolved oxygen tension (DOT) were measured using steam sterilisable instruments mounted in the fermenter. The dissolved oxygen tension was measured using a polarographic probe (Ingold, Urdorf, Switzerland). The equipment used at SmithKline Beecham and at UCL was similar in design, versions differed to accommodate the differing fermenter vessel size.

4.3.2 On Line Mass Spectrometer

Both the fermenter suites used at UCL and at SmithKline Beecham are equipped with on-line mass spectrometers. The equipment used in both sites were supplied by VG Gas Analysis Ltd, (Middlewich UK). These devices could analyse the inlet and exit gas streams in the fermenter aeration system for nitrogen, oxygen and carbon dioxide and argon. The mass spectrometers were calibrated using a reference gas once every 24 hours. Both systems were multiplexed to provide data on a number of fermenters so the frequency of data recording was a function of the number of fermenters being tracked. At SmithKline Beecham data recording frequency could drop as low as one point every 15 minutes. At UCL, owing to the smaller number of fermenters in use at any given time, frequency never dropped below 1 analysis every 6 minutes.

4.3.3 Off Line Determination of Penicillin and Phenylacetic Acid Concentrations..

At UCL, a Perkin Elmer HPLC was used to determine concentrations of penicillin and phenylacetic acid in solution in the fermentation broth. The HPLC system comprised of a Perkin Elmer ISS-100 autosampler and a Perkin Elmer System 10 Constant Flow Pump unit. Separation was achieved on a Spherisorb S5C8 column (Hichrom Ltd UK). A Milton Roy Spectromonitor 3 U.V. spectrophotometer set to a wavelength of 220 nm was used as the detector. The output from this was fed into a Perkin Elmer LCI-100 integrator.

At SmithKline Beecham, analysis for penicillin was carried out using a Cobas Fara centrifugal analyser (Roche Diagnostics). This detected penicillin-G by using a spectrophotometric assay following a reaction between the penicillin-G and a mercuric chloride based Azole reagent. The Cobas Fara analyser was also used to determine broth glucose concentration. phenyl acetic acid concentration was measured using HPLC as at UCL. In this case the column used was a Whatman 125 mm packed column containing Partisphere C18.

4.3.4 Fermenter Nutrient Feed System

A key part of this work was to ensure that the rate of glucose feed was accurately known. To ensure this, at UCL feed reservoirs were calibrated perspex measuring cylinders, to which side arms had been fitted by the UCL glassblower. Recording the rate of volume loss from these reservoirs permitted verification of the calibration of the peristaltic pumps used. The same check on pump calibration could be obtained at SmithKline Beecham by noting the rate at which the syringes became emptied.

4.4 Fermentation Procedure

Fermentations were started by the transfer of a seed culture containing *Penicillium chrysogenum PC2* organisms into pre-sterilised fermentation medium. This transfer was timed to occur when most but not all of the glucose in the seed culture broth had been used up, giving a high biomass loading of rapidly growing mould. The seed culture was inoculated from spores maintained on Agar slopes stored in a refrigerator. These slopes had a shelf life of around three months, they were replenished from a

master culture stock held in a deep freeze. The master culture was also in the form of spores grown on an Agar slope.

4.4.1 Spore Preparation

The cultures of *P.Chrysogenum PC2* moulds used in these fermentations were grown from spore stocks kept on Agar slopes, the composition is shown in table 4.2.

Table 4.2 Agar Slope Composition

Component	Concentration % w/w
No.3.Agar	2.0
Sodium Chloride	1.0
Glycerol	0.75
Peptone	0.50
Molasses	0.25
Yeast Extract	0.10
$KH_2PO_{4(anh)}$	0.006
$MgSO_{4(anh)}$	0.005
$FeSO_4.7H_20$	0.0003
$CuSO_4.5H_2O$	0.0001
Water	Balance

This formula was in routine use at UCL amongst a group working on penicillin fermentations. The formula was supplied in a personal communication by Mr. R.Eglin, it worked completely reliably.

Sporulation for use in fermenter seed culture preparation was carried out in 500 ml 'Mediflat' glass bottles. A slope volume of 100 ml was used which permitted a surface area for growth of about 100 cm^2. Slopes were seeded from master slopes which were held in 25 ml glass bottles.

4.4.2 Slope Preparation

The medium ingredients were added to each of a batch of bottles, each of which was then capped with a gas permeable foam bung and aluminium foil. The bottles were then sterilised by being held at 121°C in an autoclave for at least 20 minutes. The contents of the bottles were then mixed gently to ensure homogeneity and laid on their sides to cool and set.

4.4.3 Slope inoculation

A suspension of spores was prepared by adding 10ml of a sterilised 0.1% v/v Tween 80 solution containing 3mm diameter glass beads into a master slope. The master slope was gently rocked to remove spores from its surface without disturbing the underlying Agar gel. When the liquor was judged sufficiently concentrated, it was transferred aseptically to a sterile bottle. The sterilised slopes in the 500 ml bottles were each inoculated with approximately 0.8 ml of the resultant suspension. This volume was chosen because from experience it was just sufficient to cover the surface of the slope while leaving unwetted islands. These islands were a useful test for contamination in that in uncontaminated slopes they remained free of any visible growth.

When preparing new master slopes, a sterile wire loop was used to inoculate the surface rather than free liquid. The presence of growth only on the streak marks was also a useful sterility check, showing that at least the media was sterile. A contamination check was also carried out on residual spore suspensions after use. A sample was streaked onto a sterile Agar plate using a sterilised wire loop, the plates were then incubated at 37°C for two to three days. Contaminants were indicated by the presence of fast growing colonies, slow growing contaminants would have been dominated by *penicillium*. Any slopes prepared from suspensions from which a positive contamination test was obtained were discarded.

4.4.4 Slope Growth And Storage

The slopes were incubated at 26°C for 9-10 days. A dark green surface colouring typically developed as growth and then sporulation took place. After sporulation the

slopes were made air tight by replacing their foam bungs with screw caps. The slopes were stored in a cold room at 4°C for use within three months.

4.4.5 Seed Culture Preparation

Seed cultures were prepared from slopes in conical flasks of either 2.5 l or 1 l depending on the scale of the fermentation to be carried out. The medium used for the seed culture was as specified by Mou and Cooney 1983. Its composition is given in table 4.3.

Table 4.3 Seed Culture Medium

Component	Concentration % w/w
$Glucose_{(anh.)}$	2.00
Corn Steep Liquor (Spray Dried)	1.28
$Na_2SO_{4\ (anh.)}$	0.30
$MgSO_{4(anh.)}$	0.24
Water	Balance

4.4.6 Medium Preparation.

In order to avoid undesirable reactions between glucose and proteins present in the corn steep liquor (Maillard's reactions), glucose was sterilised separately from the rest of the medium components. The glucose was sterilised in a 100g/l solution, the remaining seed medium was sterilised *in situ* in the shake flask. The means of sterilisation was heating to 121°C and holding temperature for at least 20 minutes. On cooling the glucose solution was aseptically added to the rest of the seed medium in the shake flask, and the flask was capped with a gas.permeable foam bung. The resultant liquid occupied approximately 10% of the volume of the flask. This provided a high surface to volume ratio, aiding oxygen mass transfer during culture growth.

As a sterility test, the flasks were incubated for 48 h prior to use. Typically six seed flasks would be prepared for a fermentation, of which two would be used. The spares were prepared to cover the need to discard some flasks due to contamination.

4.4.7 Seed Growth

Each shake flask was inoculated with spores prepared in the same way as described in section 4.1.2. As described, a contamination test was performed on the residual spore suspensions. Shake flasks seeded from these suspensions were discarded if the sample was not shown to be clear of contaminants. The volume of spore suspension required to inoculate each shake flask was defined by the need to produce a concentration of at least 10^8 spores/l in the shake flask medium. The concentration of spores in the suspension was estimated using a haemocytometer.

Once inoculated, the flasks were placed in an orbital shaker (in a temperature controlled cabinet at UCL, and in a temperature controlled room at SmithKline Beecham). The flasks were maintained at 25-26°C for 44 hours and then transferred aseptically to the fermenter. An inoculum volume of 10% of the total fermentation volume was used.

4.4.8 Fermentation Medium Preparation.

A low solids semidefined medium was used in this work. This made possible an estimate of biomass from dry weight measurements (see section 4.5.2). The medium was adapted from that described by Mou and Cooney (1983) Its composition is given in table 4.4.

Table 4.4 Initial Medium Composition

Component	Concentration % w/w
Glucose $_{(anh.)}$	0-0.80
KH_2PO_4	0.453
K_2HPO_4	0.40
Corn Steep liquor	0.227
$(NH_4)_2SO_4$	0.227
$MgSO_{4(anh.)}$	0.115
Henkel Nopco Antifoam	0.10
NaCl	0.057
$FeSO_4.7H_2O$	0.057
$CaSO_4.2H_2O$	0.050
$MnSO_4.4H_2O$	0.0050
$ZnSO_4.7H_2O$	0.0034
$CuSO_4.5H_2O$	0.0010

The medium, with the exception of glucose which was sterilised separately for the reasons discussed above, was made up and adjusted to a pH of 6.5. It was then transferred to the fermenter. The fittings and instrumentation were added to the fermenter and it was sealed. The fermenter and its medium at UCL was then sterilised by means of heating to 121°C using an internal heating coil. The fermenters used at SmithKline Beecham were placed in an autoclave and heated to 121°C. In both cases sterilisation time was 30 minutes. The glucose solution was independently sterilised in an autoclave, the quantities being varied to produce differences in fermentation results (see table 5.2 for details). At the same time demineralised water was sterilised for flushing through the contents of the seed flasks.

After sterilisation, the fermenter instrumentation was checked. The dissolved oxygen tension probe was tested by first sparging with air via a sterile filter, then sparging with nitrogen. This gave a reading for the top and bottom of the instrument's range. A sample of broth was drawn and its pH measured. The set point of the pH probe amplifier was adjusted on the basis of an independent measurement of the pH of this sample.

4.4.9 Fermenter Inoculation

Seed medium was transferred to an inoculation vessel for 5 litre fermentations at UCL. Typically the contents of two shake flasks were combined to give an inoculum volume of approximately 10% of the batch volume. The full 5 litre starting volume for the fermentation was achieved by adding the appropriate amount of sterilised deionised water to bring fermenter level up to a predetermined point in the vessel. For this operation it was necessary to shut down the fermenter's agitation and aeration system.

The work at SmithKline Beecham required only one shake flask per fermenter due to the lower working volume, otherwise similar inoculation procedures were followed.

4.4.10 Feed Preparation

Glucose

In the penicillin fermentation process, growth rate is controlled by the rate at which glucose is added to the fermentation vessel. Glucose was therefore fed to all of these fermentations continuously. In order to achieve differentiated fermentations, the rate of glucose feeding was adjusted, for details see table 5.4.
The required quantity of glucose for a fermentation was gradually added to hot deionised water under agitation. Once the glucose was fully dissolved, the required volume was achieved by the addition of further quantities of water. The glucose solution was then transferred to the graduated measuring beaker that was modified to act as a feed reservoir. It was then sterilised by heating to 121°C and holding for 20 min.

Phenylacetic Acid

Phenylacetic acid was fed to the fermentations in 100g/l solutions. These were prepared by adding phenylacetic acid slowly to deionised water in a stirred beaker. It was necessary to maintain a pH of between 5 and 7 to permit phenyl acetic acid solubilisation. This was achieved by adding potassium hydroxide pellets to the solution, a ratio of 3 parts phenyl acetic acid to 2 parts potassium hydroxide by weight was found to be effective. On completion of the solubilisation step, the

volume of the solution was adjusted to give the desired concentration, and the final pH was adjusted to 6.5 using some additional potassium hydroxide solution. The solution was then transferred to the feed reservoir, and sterilised by heating to 121°C for 20 min.

4.5 Fermentation Monitoring

4.5.1 Fermentation Broth Sampling

The LH fermenter used at UCL was fitted with a sterile sample port through which aliquots of broth could be withdrawn. The fermenter used at SmithKline Beecham was sampled using a hypodermic syringe, the needle of which was passed through a sterile membrane or septum. Samples of approximately 25 ml were drawn, and placed in pre weighed bottles The samples were immediately placed in ice and analysed as quickly as possible to minimise the impact of any further metabolic activity. The volume withdrawn was recorded, this change in broth volume is taken into account in subsequent calculations.

4.5.2 Biomass Determination.

The concentration of biomass in each fermentation broth sample was measured by drying samples of pre-weighed then filtered broth. There is some error in this procedure, as the initial medium contains insoluble solids. This level of insoluble solids typically found in fermentation media prior to innocculation was of the order of 1.1-1.2g/l . Since the same concentration of solids was found in each batch, this did not prevent valid batch to batch data comparisons from being made. An alternative approach would have been to measure the level of solids in a sample of uninnocculated broth taken after sterilisation, when most soluble material would have been dissolved. This could have been taken as a blank, the observed solids quantity being subtracted from the dry weight measurement. This requires the assumption that no broth solids are metabolised in the course of the fermentation. The validity of this assumption is examied in section 5.5.

As no accounting was made for broth non biomass solids, it is more correct to refer to total broth solids rather than biomass, when discussing dry weight data presented in this thesis. This approach will be adopted where appropriate throughout the thesis.

The dry weight concentration was measured by pipetting approximately 5 ml of well-mixed sample onto a predried and pre weighed filter paper. (Whatman GF/C 1.2 μm pore size filter papers were used.) The sample was coarsely dried by application of a vacuum (using a Sartorius vacuum filtration rig), and then washed with 10 ml of deionised water to remove associated sugars and other water soluble components. The samples were then dried in an 80-90°C oven for 24 h, and then weighed.

In order to be able to calculate the dry weight as a concentration, the weight of the sample pot was measured before and after each sample was pipetted onto a filter paper. An assumption was made that the density of the broth did not significantly differ from that of water. In order to avoid bias due to single erroneous results, four filter papers were loaded from each sample drawn from the fermenter and the final dry weight measurement was taken as the mean of the dry weight results thus obtained.

4.5.2.1 Dry Weight Concentration Measurement Uncertainty

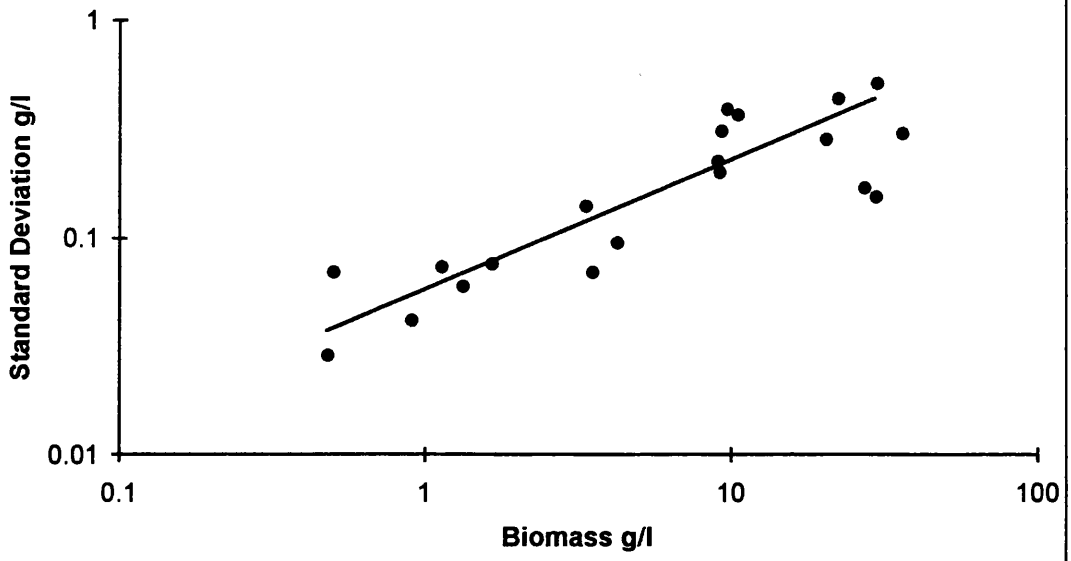
Work was done to determine the precision of the dry weight data obtained. The following table reports the data obtained by multiple measurement of samples from fermentations. Data were also obtained for this study from spare shake flasks used to grow up seed cultures.

Table 4.5 Dry Weight Variance Data

Source of sample	Media Solids g/l	Mean dry weight conc. (g/l)	Sample standard deviation (g/l)	99% Confidence Limit	Set size n
combinations of uninoculated medium components	2.3 g/l CSL	0.4824	0.0288	0.074	9
	2.3 g/l CSL	0.9037	0.0419	0.107	9
	5.7 g/l CSL	1.3339	0.0600	0.154	10
	Medium salts excluding CSL	0.5031	0.0695	0.178	10
	2.3 g/l CSL and medium salts (uninoculated medium)	1.1392	0.0733	0.188	9
	5.7 g/l CSL and antifoam	1.6590	0.0759	0.194	9
shake flask culture		3.5350	0.0695	0.178	10
		4.2540	0.0955	0.244	10
		9.0600	0.2254	0.577	9
		9.3044	0.3109	0.796	9
		10.5378	0.3684	0.943	9
		3.3717	0.1399	0.358	9
		9.1799	0.2016	0.516	9
		9.7102	0.3915	1.002	9
Fermentation samples	FERM08	30.1951	0.5150	1.318	9
	FERM09	29.8853	0.1538	0.394	9
	FERM11	22.4736	0.4352	1.114	9
	FERM12	20.4685	0.2841	0.727	9
	FERM12	27.3931	0.1694	0.434	9
	FERM13	36.6436	0.3020	0.773	10

Plotting this data on logarithmic axes, a power law relationship can be observed:

Graph 4.1 Relationship Between Broth Solids Measurement and Broth Solids Measurement Standard Deviation



The extremes of dry weight concentration measured during the fermentation work were 2.93 $g.L^{-1}$ and 38.65 $g.L^{-1}$ the above graph gives uncertainties in terms of sample standard deviation of 4.63% and 1.03% respectively, or in terms of worst case limits (±3s limits) ±0.32 $g.L^{-1}$ and ±1.19 $g.L^{-1}$ respectively. .

Some reported measurement uncertainties for dry weight and biomass concentration measurements are summarised in table 4.6. Although none of the reported values are for the same fermentation system as that used in this work, they were useful because they indicate what are considered to be reasonable values for the uncertainty associated with this measurement.

Table 4.6 Variance or range reported for biomass data.

Reference	Biomass Range g/l	Range (max/min or ± 2 σ)
Svrcek *et al.* (1974) Reported for a simulated fermentation.	1.3	±0.050 g/l
San and Stephanopoulos (1984 b) Reported for a fermentation of *Saccharomyces cerevisiae*. The concentration was measured by optical density with a subsequent conversion to dry weight using a calibration curve.	0.68 to 0.87	±0.034 g/l
Staniškis and Simutis (1986) Reported for a simulated fermentation of *Saccharomyces cerevisiae*.	16 to 24	±0.5 to 1.0
Stephanopoulos (1986) The same fermentation and measurement system as for San and Stephanopoulos (1984 b).	1.2 to 1.8	±0.08
Chattaway and Stephanopoulos (1987) Reported for a mixed culture of *Saccharomycopsis lypolytica* and *Escherichia coli*. Total biomass was measured by an unspecified method of dry cell weight.	0 to 30 g/l	±0.5
Yu *et al.* (1987) Reported for a fermentation of 'yeast'; no practical details were given.	3 to 45 g/l	±0.2L
Pigott (1989) Reported for a fermentation of *Streptomyces clavuligerus*. The dry weight concentration was measured by a similar method to that used in this work.	2 to 3	±0.088

From the above data it can be seen that the uncertainties in biomass measurement in this work are similar to those encountered by other researchers. The data also shows the level of initial bias in measurements due to the presence of insoluble components of the corn steep liquor.

4.5.3 Glucose Concentration Determination

At SmithKline Beecham, the concentration of glucose in the fermentation broth was measured using a proprietary analytical technique. At UCL, it was detected spectrophotometrically. A sample of broth was centrifuged at 4000rpm for 10 min., the supernatant was then filtered through a Whatman 0.45μm pore size cellulose nitrate filter pad. The filtrate was used both for glucose determination, and for analysis by HPLC for phenylacetic acid and Penicillin-G. The packed volume of cell debris in the centrifuge tubes gave a quick approximation of the level of biomass in the fermenter.

20 μl of the filtrate was added to 2.5 ml of Glu-cinet reagent in a spectrophotometer cuvette and allowed to stand at room temperature for one hour. The absorbance of light at 510 nm wavelength was then measured. The absorbance, minus the absorbance of a Glu-cinet sample to which no glucose has been added, is proportional to glucose concentration within the range 0 - 5 g/l glucose in the 20 μl sample. Ten replicates were performed for each glucose determination. The mean result was converted into a broth glucose concentration by comparison with the absorbance obtained from a set of standard solutions.

4.5.4 Glucose Concentration Measurement Uncertainty

The assay used to measure broth glucose concentration was investigated for a relationship between the measured value and its associated uncertainty. The large manual component in preparation of the sample and performing the assay suggested that precision would probably be the most significant component of the measurement uncertainty; accuracy was expected to be less significant because frequent calibration of the assay was normal practice during the fermentations.

Routinely ten replicate measurements were carried out for each glucose measurement during the fermentations. This was to permit easy rejection of outlier data caused for

example by incomplete filtration of samples, and to achieve a reasonable degree of precision (± 3%) in an assay dependent on a number of manual operations. Measurement sets were also available from glucose concentration measurements made on samples from the shake flask cultures used to investigate dry weight concentration measurement uncertainty and fermentations that were terminated because of contamination.

4.5.5 Penicillin And Phenylacetic Acid Concentration Measurement

The remaining filtered supernatant from the filtration described above was frozen. It was then thawed out at the end of the fermentation, permitting the analysis of all samples from one fermentation in a single run on a HPLC.

The composition of the mobile phase for the HPLC is given in Table 4.7. (Personal communication, R.Eglin UCL):

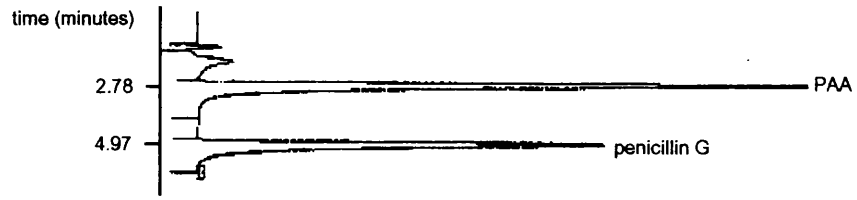
Table 4.7 Mobile Phase Used For HPLC Analysis

Component	Concentration % w/w
0.15 M KH_2PO_4 Soln	50
Deionised water	30
Acetonitrile (CH_3CN)	20

After mixing, the mobile phase liquor was filtered using 0.45μm pore size solvent resistant filter pads, and then degassed using a sonicator.

Using this system, phenylacetic acid and penicillin G could be separated, and detected using the absorbance of 220 nm wavelength light. The peaks thus obtained were compared to peaks produced using samples of phenylacetic acid and the potassium salt of penicillin G of known concentration. The following HPLC absorbance trace illustrates typical output from this technique:

Fig 4.2 Output Of HPLC Detector



4.5.6 Measurement of Fermentation Broth Volume.

The broth volume over time was calculated once each fermentation was complete. This allowed a measure of the quantity of water lost to evaporation during the course of a fermentation. This was calculated by summing all the additions to and abstractions from the fermenter, (i.e. initial broth volume, glucose, and phenylacetic acid feed, acid and base additions for pH control, and sample removals), and subtracting the final broth volume from this total, the difference being assumed to be due to evaporation. The fermenter volume could then be reconstructed, by assuming the rate of evaporation was constant over the course of the fermentation. This assumption is justified given constant temperature in the fermenter, and that the air supply used was dried and filtered, and fed into the fermenter at a constant rate (i.e. it was not corrected for broth volume changes to give a constant v/v/min of air).

This analysis combines volume changes due to metabolic activity such as the conversion of glucose into carbon dioxide with evaporation.

4.5.7 Measurement Of Carbon Dioxide Evolution Rate.

The means of monitoring carbon dioxide evolution rate (CER) was by on line mass spectrometry, the equipment used is described in section 4.3.2., The value of carbon dioxide evolution rate was calculated using the following equation,

$$CER = \left[\frac{\text{Air Flow (l/hr)}}{\text{Broth Volume (l)}}\right]\left[\left(\frac{N_{2\,In}}{N_{2\,Out}}\right)(CO_{2\,Out} - CO_{2\,In})\right] \qquad 4.1$$

The $N_{2_{in}}/N_{2_{out}}$ term is included to correct for the difference in air flow into and out of the fermenter due to for example water vapour pickup, and the difference between oxygen absorption and carbon dioxide evolution rates due to some oxygen being becoming incorporated into biomass or metabolic products.

This calculation requires a knowledge of broth volume. Ideally, this would be calculated on line using a load cell mounted into the fermenter support. For this work, such equipment was not available. This was not a problem in that all the fermentation data analysis was carried out after the fermentations were completed.

4.5.8 Uncertainty In Carbon Dioxide Evolution Rate Measurements

Knowledge of uncertainty associated with the measurement of the rate of carbon dioxide evolution was required to determine how meaningful a carbon balance could be obtained, and for running the extended Kalman filter. The carbon dioxide evolution rate is a derived quantity, which, in this work, was calculated from six separate measurements: the broth volume; the air flow rate to the fermenter; and the proportions of nitrogen and carbon dioxide in the fermenter's inlet and exhaust air, using Equation 4.1.

The uncertainties of the component measurements were investigated separately and combined to produce the overall uncertainty in the carbon dioxide evolution rate data. This section describes the investigations of the measurement uncertainties of the component measurements and their combination to produce the uncertainty in the carbon dioxide evolution rate separately.

4.5.8.1 The Uncertainty Of The Broth Volume Measurement

The investigation of the measurement uncertainty for the broth volume calculation has already been described (Section 4.5.6). The worst case limits of the measurement uncertainty were fixed at ±100 ml for the range of broth volumes measured for the fermentations in this work.

4.5.8.2 The Uncertainty Of Air Flowrate Measurement

An examination of the time courses of the air flow rate to the fermenter during the fermentations showed that precision of the air flow rate meter was good: there was no drift and the measurement fluctuation was confined within ±0.01 $l.min^{-1}$ (at the level of the air flow meter discrimination) of an apparent fixed level, No reference air flows were available to calibrate the air flow rate meter, therefore, the accuracy claimed by the manufacturer of ±1% of full scale, which is typical for this type of instrument (Flynn, 1990), was used.

The full scale air flow rate for this instrument is 10 $l.min^{-1}$: at 2.5 $L.min^{-1}$; the air flow rate used for all the fermentations, the accuracy is ±0.1 $l.min^{-1}$, or ±4% of the set value. In this work the uncertainty in the air flow rate measurement due to accuracy was more significant than due to precision, therefore, accuracy was used to define the measurement uncertainty. The measurement accuracy was assumed to define worst case limits of measurement uncertainty, within which the distribution of the measurements is unknown.

4.5.8.3 The Uncertainty Of The Gas Analysis

The uncertainties associated with measurements of the proportions of carbon dioxide and nitrogen in the fermenter inlet and exhaust gas streams were investigated. To do this, gas sources that were constant and typical of those of the fermenter inlet and exhaust gas streams were required. During the fermentations the measured value of the proportion of carbon dioxide varied from approximately 0.03%, in the inlet gas (air), to a maximum of 3%, in the exhaust gas, and the proportion of nitrogen varied from 74%, in the exhaust gas, to 78% of the inlet gas. Three constant composition gas sources with proportions of carbon dioxide and nitrogen that spanned the ranges measured during the fermentations were used in this investigation.

Gas source 1: a cylinder of compressed air (approximate composition: 1% inert gases, 0.03% CO_2, 20% O_2, and 78% N_2) was used to provide air with a composition similar to that of the inlet air. Ambient air was unsuitable for this investigation because it had a slightly

variable composition depending on the use of the equipment in the adjacent fermenter hall.

Gas source 2: a cylinder of gravimetrically mixed air (stated composition: 1% Ar, 3% CO_2, 20% O_2, and 76% N_2, all proportions with an accuracy of ±1% (BOC a-grade, BOC, Twickenham, UK)) was used to provide air with the maximum level of carbon dioxide found during the fermentation. This was the same specification as the reference gas used to calibrate the mass spectrometer during the fermentation work.

Gas source 3: a cylinder of volumetrically mixed air (stated composition: 1% Ar, 5%, CO_2, 20% O_2, and 73% N_2 , all proportions with an accuracy of ±2% (BOC b-grade)) was used to provide air with a low level of nitrogen.

The cylinders were connected to the mass spectrometer by available inlet ports. The mass spectrometer and analysis conditions were the same as those for the fermentations, i.e. the air flow from the cylinders to the mass spectrometer was adjusted to approximately 200 $ml.min^{-1}$; the frequency of gas analysis was determined by the number of users on the mass spectrometer; and the mass spectrometer was calibrated every 24 hours. Repeated analysis of the air sources were made over periods of about 100 hours, which allowed investigation of both short term and long term performance. The analysis results were logged by the mass spectrometer system's independent computer.

The measurement precision was investigated using gas analysis data. Each analysis was divided into inter-calibration episodes, which eliminated the effects of calibration adjustments, which are discussed below, and limited the influence of drift. The precision of the measurements within each episode was calculated as the standard deviation of the measurements. The best performance of the mass spectrometer that was found for the measurement of nitrogen and carbon dioxide from each of the sources is presented in Table 4.8 in terms of lowest standard deviation. The highest standard deviation calculated for an episode in the same analysis is also presented in Table 4.8; these values were more typical of the mass spectrometer performance.

Table 4.8 Mass Spectrometer Variability

Species	Gas source	Measured value (approx.)	Lowest standard deviation		Highest standard deviation	
			value	% of measured value	value	% of measured value
CO_2	1	0.03%	$1.36\%10^{-4}$	$4.53\%10^{-1}$	$2.85\%10^{-4}$	$9.50\%10^{-1}$
	2	3%	$1.15\%10^{-3}$	$3.83\%10^{-2}$	$1.37\%10^{-3}$	$4.57\%10^{-2}$
	3	5%	$1.26\%10^{-3}$	$2.52\%10^{-2}$	$1.68\%10^{-3}$	$3.36\%10^{-2}$
N_2	1	78%	$8.75\%10^{-3}$	$1.12\%10^{-2}$	$1.88\%10^{-2}$	$2.41\%10^{-2}$
	2	76%	$8.32\%10^{-3}$	$1.09\%10^{-2}$	$9.31\%10^{-3}$	$1.22\%10^{-2}$
	3	74%	$5.68\%10^{-3}$	$7.68\%10^{-3}$	$9.02\%10^{-3}$	$1.22\%10^{-2}$

The absolute value of the precision worsens with increasing measured value, confirming the observation made above of an apparent increase in noise amplitude with an increase of measured value, but its significance with respect to the measured value decreases. Considering the highest standard deviation, the worst case limits ($\pm 2\sigma$ limits) for the range of proportions of nitrogen and carbon dioxide measured during the fermentations are approximately: ±0.03% (at 74% N_2) to ±0.05% (at 78% N_2); and ±2% (at 0.03% CO_2) to ±0.1% (at 3% CO_2).

The accuracy of the analysis available from the mass spectrometer was limited to that of the reference gas source used to calibrate it. The specification of the reference gas source was the same as that for Gas source 2, i.e. all proportions were stated with an accuracy of ±1%. Therefore all measurements made by the mass spectrometer had an uncertainty of ±1% of the measured value associated with them. This was backed up by observation of the analysis results from the gas sources with known compositions (gas source 2 and gas source 3): the measured proportions of nitrogen and carbon dioxide were never outside the possible limits, which could be as much as ±2% for gas source 2 (±1% from the calibration and ±1% from the stated composition of gas source 2) and ±3% for gas source 3 (±1% from the calibration and ±2% from the stated composition of Gas source 3).

This investigation found that, except for the proportion of carbon dioxide typical of the inlet air (approximately 0.03%), the contribution to the measurement uncertainty associated with the measured proportions of nitrogen and carbon dioxide was more significant from accuracy than from precision. The accuracy of ±1% was used to define the uncertainty associated with the measured values of $\%N_{2\,in}$, $\%N_{2\,out}$, and $\%CO_{2\,out}$, and the worst case uncertainty of ±2%, from the investigation of precision, was used to define the uncertainty of the measured value of $\%CO_{2\,in}$.

4.5.8.4 The Uncertainty Of The Carbon Dioxide Evolution Rate Data

The measurement uncertainties for the component measurements of the carbon dioxide evolution rate are summarised in Table 4.9. These uncertainties are in the form of worst case limits because, with the exception of the measurement of $\%CO_{2\,in}$, they were determined by the uncertainty associated with the reference used in their calibration; or, in the case of the measurement of air flow rate, by the manufacturer's statement of accuracy, which were in terms of worst case limits.

Table 4.9 Carbon Dioxide Evolution Rate Uncertainties

Measurement	Uncertainty
broth volume	±0.1 L (fixed)
air flow rate	±0.1 l/min (fixed)
$\%N_2$ in	±1%
$\%N_2$ out	±1%
$\%CO_2$ in	±2%
$\%CO_2$ out	±1%

The upper and lower limits of the worst case uncertainty for the carbon dioxide evolution rate were determined by combining the uncertainties of the component measurements. Taking upper and lower limits of uncertainty, and substituting them into equation 4.1 it was calculated that the maximum possible error on carbon dioxide evolution rate is ± 8%. The most significant contribution to the uncertainty was from the measurements of air flow rate and volume. This is supported by Stephanopoulos and Wang (1983) who commented that the uncertainty in the air flow rate measurement is a major contribution to uncertainty in the carbon dioxide evolution

rate measurement. The uncertainty of the carbon dioxide evolution rate in this work was determined by systematic factors, such as the uncertainty in the calibration of all component measurements.

4.6 Conclusions

In this chapter, the methods that were used for running and monitoring fed batch penicillin fermentations have been presented. An analysis of the confidence limits for the data obtained, together with the means used to identify these limits is also presented.

Table 4.10 Confidence limits on key fermentation states

Parameter	Variability
Broth Solids	3±0.32g, 40±1.0g
Glucose	±0.15g/l
Penicillin	±0.1 g/l
Broth Volume	±100 ml
Carbon Dioxide Evolution	±12 ml/l/hr

The limits are specified either as three standard deviations, or where possible represent absolute limits.

This data is novel, detailed analyses of fermentation measurement system accuracy is not commonly available.

Chapter 5: Fermentation Data

In the previous chapter, the protocol for conducting penicillin fermentations was described. Using those methods, a total of 31 penicillin fermentations were conducted, of which 21 provided useful data. Thirteen of these fermentations were carried out with an associate, Mr (now Dr.) K.Stone, at UCL. The primary causes of unused fermentations were:

- failures in the data recording system
- failures in process services early in the fermentation
- contamination by organisms which prevented the growth of any significant levels of *Penicillium chrysogenum.*
- Non-standard fermentation operations conducted while experimental methods were being developed.

This chapter reports the outcome of the 21 fermentations which yielded useful data.

These fermentations essentially fall into two series. The first series was conducted to provide a database of fermentations that could be considered to have been successful. These were conducted to provide data on which a fermentation monitoring package could be developed. They were run with some batch to batch variations to provide a spread of data, as will be described later.

The second group of fermentations, conducted at SmithKline Beecham's Biotechnology Development facility, was conducted to provide data on the effects of faults on fermentation processes. To this end, fault events were deliberately introduced to these fermentations. This chapter is intended to describe the fermentations carried out, and to present in a manageable form the data obtained. The next chapter will show the analysis of this data to produce data that is directly derivable from the on line fermentation data. This chapter also reports the data gathered to validate the monitoring techniques used.

The output of data from these fermentations, and the associated work to measure the accuracy and precision of data obtained from these fermentations is a key part of the results from this project. They show under controlled conditions what quantity and

quality of data is obtainable from typical research fermenters of the type widely used in industrial fermentation strain improvement. For this reason the key experimental data is reported here at length, and its accuracy analysed.

The data was manipulated in one significant way to make it useful for this project. The data was curve fitted using cubic splines as described by Thornhill et al (1994). This permitted comparisons between fermentations which were not sampled at precisely the same time interval.

5.1 Starting Conditions

The initial values of state variables for each fermentation were important in that they defined how the initial fast growth phase of the fermentation proceeded, and they provided a set of initial values for the differential equations used to model the fermentation and to drive the Kalman filter used to simulate an on line observer. The initial set of values in this case were one of the key controllable parameters in these fermentations.:

Table 5.2 reports the initial values of state parameters for each fermentation from which useful data was obtained. Table 5.2 shows the variation in initial glucose used throughout this program to generate diverse fermentation data. This is to simulate the variability to be expected in a development fermentation suite.

The initial dry weight represents contributions from the inoculum and from solids in the spray dried corn steep liquor. The dry weight of samples of uninoculated fermentation broth were typically 1.2 ± 0.2 g/l. The true level of biomass therefore was in the order of 2-3 gdw/l at the start of each fermentation. This does not constitute a significant issue for fault detection, as the carbon content of the corn steep liquor was within 2% of the carbon content of the dried cells. These two components can therefore be combined in an overall carbon balance with negligible error. Biomass data reported here should be taken as biomass plus undissolved corn steep liquor solids, as it is likely that little of the undissolved solids would have been metabolised during the course of the fermentation.

Data on initial glucose concentration was obtained on the basis of measuring the quantity of glucose added. Inevitably during the glucose sterilisation process, some

caramelisation may have ocurred. Some of the glucose would therefore not have been metabolisable. Table 5.2 shows the starting conditions of all the fermentations.

Table 5.1 Fermentations carried Out as Part Of This Study.

Fermentation	Location	Purpose
Ferm 4	UCL	Standard run
Ferm 8	UCL	Repeat standard run
Ferm 9	UCL	Repeat standard run
Ferm 11	UCL	Repeat standard run
Ferm 12	UCL	Low glucose feed run
Ferm 13	UCL	High glucose feed run
Ferm 1a	SB	Mass Spec fault
Ferm 1b	SB	Air filter fault
Ferm 1c	SB	Agitation fault
Ferm 2a	SB	Contamination†
Ferm 2b	SB	Temperature control system fault
Ferm 2c	SB	Repeat standard run
Ferm 3a	SB	Contamination
Ferm 3b	SB	Mass spectrometer fault†
Ferm 3c	SB	Mass Spectrometer fault†
Ferm 5a	SB	Standard run
Ferm 5b	SB	Standard run
Ferm 5c	SB	Temperature control fault
Ferm 6a	SB	Standard run
Ferm 6b	SB	Standard run
Ferm 6c	SB	Phenylacetic Acid overfeed

† unplanned fault.

Table 5.2 Fermentation Start Conditions

Fermentation	Initial Broth solids g (dry weight)/l	Initial Glucose Conc. g/l†	Initial Broth Volume l
Ferm 4	3.42	7.90	5.00
Ferm 8	2.93	4.36	5.00
Ferm 9	3.23	3.90	5.00
Ferm 11	3.73	0.21	5.00
Ferm 12	3.55	3.54	5.00
Ferm 13	3.08	3.28	4.90
Ferm 1a	3.14	1.99	1.55
Ferm 1b	3.13	2.16	1.42
Ferm 1c	3.13	2.22	1.43
Ferm 2a	3.11	8.80	1.42
Ferm 2b	4.04	7.40	1.50
Ferm 2c	3.37	8.30	1.52
Ferm 3a	3.67	7.7	1.44
Ferm 3b	3.77	7.1	1.50
Ferm 3c	3.34	6.0	1.53
Ferm 5a	3.84	7.5	1.42
Ferm 5b	4.36	7.7	1.50
Ferm 5c	4.06	8.6	1.55
Ferm 6a	2.97	6.88	1.48
Ferm 6b	2.82	6.24	1.50
Ferm 6c	4.08	9.90	1.50

† Variation in glucose levels is due to a planned attempt to introduce variations into the fermentation trajectories obtained from these batches. As will be seen later, glucose feed rates also differed between batches.

5.2 Process Conditions

For all fermentations, the following control parameters were used:

Table 5.3 Control Parameters

Parameter	Set Point
Temperature °C	25
pH	6.5
Agitation Rate rpm	600-1300
Air Flow v/v/min	2.5

Accurate control was achieved, temperature was controlled to within ± 0.2°C, and pH was controlled to within ± 0.15 units. The calibration of the agitator control unit was not checked, because there was no issue with maintaining dissolved oxygen above the level at which it would become a constraint on the fermentation. Given the small scale on which these fermentations were occurring, it is not surprising that control was precise, and that oxygen mass transfer was not limiting, except where non-standard processing conditions occurred.

5.3 Carbon Feed

The key feed for controlling these fermentations was glucose; growth was limited by the amount of glucose fed into the fermentation. Glucose was the sole source of metabolisable carbon added to these fermentations except for carbon in the spray dried corn steep liquor solids added at the start of the fermentation. Phenylacetic acid is not generally metabolised, it is utilised to produce the side chain for penicillin G.

The glucose feed rate was varied in these fermentations between 0.4 g/l/hr and 0.9 g/l/hr (see Table 5.4) to generate a disparate set of fermentation data. This was to simulate the range of fermentation conditions which would be expected to be used in a fermentation process development facility.

Table 5.4 The glucose feed pattern for each fermentation:

Fermentation	Total Glucose Feed g	Total Glucose Feed g/l	Mean Glucose Feed Rate g/l/hr
Ferm 4	560	112	0.70
Ferm 8	550	110	0.69
Ferm 9	604	121	0.76
Ferm 11	601	120	0.63
Ferm 12	525	105	0.52
Ferm 13	718	144	0.90
Ferm 1a	62	41	0.51
Ferm 1b	63	42	0.52
Ferm 1c	60	40	0.50
Ferm 2a	79	53	0.57
Ferm 2b	81	64	0.57
Ferm 2c	80	64	0.57
Ferm 3a	83	65	0.50
Ferm 3b	82	55	0.50
Ferm 3c	83	55	0.50
Ferm 5a	63	42	0.47
Ferm 5b	63	42	0.47
Ferm 5c	61	41	0.46
Ferm 6a	87	59	0.42
Ferm 6b	88	59	0.42
Ferm 6c	86	57	0.41

Glucose was fed using peristaltic pumps which were calibrated prior to use. These calibrations were checked periodically using graduations on measuring cylinders which were used as glucose reservoirs. There was no significant problem with this method of feeding, the peristaltic pumps rairly required adjustment.

5.4 Phenylacetic Acid Feed

In normal operation, addition of phenylacetic acid was manually adjusted over the course of a fermentation to maintain concentrations within reasonable bounds (i.e. 1000-2000 ppm by weight). The exception to this was after faults occurred, or were simulated, where the trajectory of the phenyl acetic acid concentration was allowed to vary uncontrolled.

Table 5.5 The Pattern of Phenylacetic Acid Additions

Fermentation	Total PAA Feed g	Total PAA Feed g/l	Mean PAA Feed Rate g/l/hr
Ferm 4	97.6	19.5	0.15
Ferm 8	168	33.6	0.26
Ferm 9	135	27.0	0.21
Ferm 11	165	33.0	0.21
Ferm 12	159.3	31.9	0.19
Ferm 13	112	22.4	0.17
Ferm 1a	9.75	7.5	0.15
Ferm 1b	9.82	7.6	0.16
Ferm 1c	9.63	7.5	0.15
Ferm 2a	5.6	4.3	0.09
Ferm 2b	11.8	8.8	0.10
Ferm 2c	11.6	8.7	0.09
Ferm 3a	9.1	7.0	0.11
Ferm 3b	15.3	11.8	0.11
Ferm 3c	15.7	12.0	0.12
Ferm 5a	13.5	9.8	0.23
Ferm 5b	16.0	12.3	0.24
Ferm 5c	15.7	11.9	0.22
Ferm 6a	9.8	6.2	0.14
Ferm 6b	22.8	17.8	0.16
Ferm 6c	22.3	17.3	0.16

Phenylacetic acid was fed using peristaltic pumps or syringes mounted on stepper motors. The rate of addition of phenylacetic acid was adjusted on the basis of analysis of samples by HPLC.

The only other feeds were ammonium hydroxide and sulphuric acid, used for pH control. The volume of sulphuric acid fed to the fermentations was minimal, serving only to correct overfeeds of ammonium hydroxide solution. The added ammonium hydroxide was a significant proportion of the nitrogen fed to the fermentation. If an effective means of tracking dissolved nitrogen had been available, the nitrogen utilisation would have been a useful extra element of analytical redundancy. This facility was not, unfortunately available for this work.

5.5 Fermentation Data

The following section reports the trajectories of the fermentation outputs. These outputs are the key data which will be used to demonstrate that there is a potential for analytical redundancy, which is useful for on line fault detection.

5.5.1 Dry Weight Data

Throughout the course of these fermentations, samples of broth were withdrawn, filtered and dried. The difference in weight between the predried filter paper, and the dried filter paper plus filtrate was used to infer the level of broth solids present in the fermenter. All of the fermentations show the same broad pattern, that is an exponential growth phase which ended at about 30 hours, and then a longer steady state growth phase, during which penicillin production ocurred. The low frequency of the broth solids data does not permit detailed observations to be made about the lag phase prior to the onset of fast growth, or about what happens between the end of the exponential growth phase, and the onset of penicillin production.

The dry weight data are a key piece of information showing the course of the fermentations, however for fault determination purposes, they have the disadvantage of being time delayed. It takes 24 h to sufficiently dry mycelia in a natural convection drying oven. The data presented here represent broth solids time courses for all the fermentations carried out at UCL and SB which produced usable data.

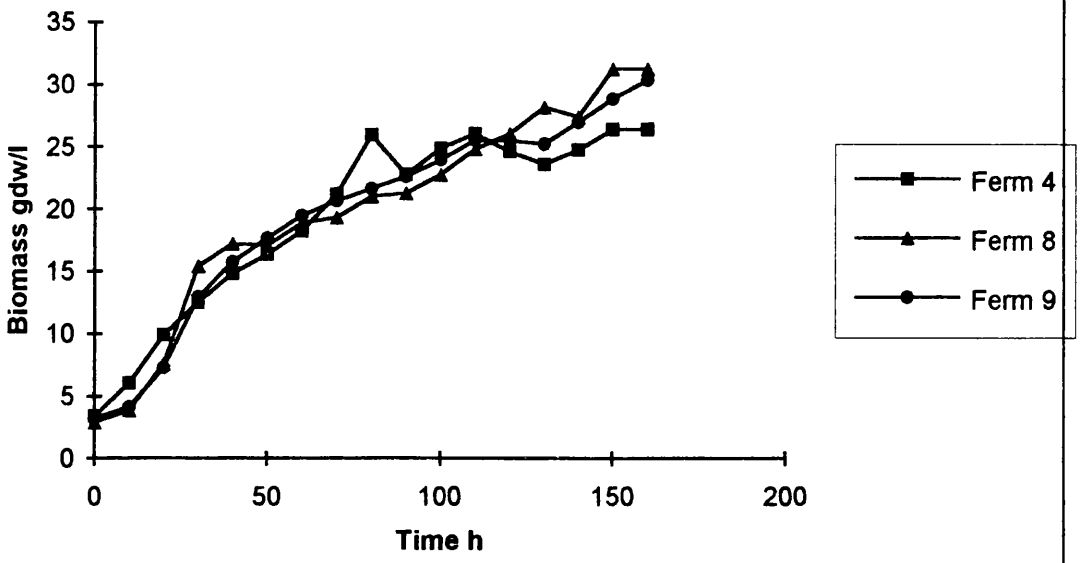
Unfortunately, there is a complication in interpretation of dry weight data. This is that the media contains some insoluble solids from the corn steep liquor used. Mou and Cooney (1983) present a study of fed batch penicillin fermentations using both corn steep liquor and glucose in the starting fermentation broth. In this study they report that after the utilisation of available glucose, and then lactose, which is a soluble component of corn steep liquor, the carbon containing components of the corn steep liquor solids is digested. They report that this comprises 34 - 42% of the CSL solids.

An issue with the determination of biomass is that it is not known at what rate these CSL solids were utilised. This means that, given a CSL total spray dried solids level of 2.2 g/l in the starting fermenter broth. the error in biomass measurements due to the presence of undigested CSL solids will be 0.6 - 0.75 g/l on completion of any solids digestion, based on a starting insoluble solids level of 1.1 g/l. Because of this lack of clarity, dry weight data will be referred to as total dissolved solids. For perspective this additional solids component represents an error of around 3% on the total biomass level in a 160 hour fermentation, however dissolved solids could distort biomass measurement by up to15-20% in the early part of the fermentation process.

A second complicating factor is that the CSL insolubles are being continually diluted during the course of a fermentation. Typically, the total volume of added liquid in a 5 litre fermentation was in the region of 3 litres, as glucose, phenyl acetic acid and ammonium hydroxide were added during the course of the fermentation. this was compensated by evaporation and the removal of samples. Dilution would have reduced the proportion of corn steep solids in a given sample by 40% during a batch. Therefore, it is probable that by the end of a fermentation, starting solids would only contribute 0.4 - 0.5 g/l out of a total dry weight of 30 g/l.

In order to make possible the comparison of batches, raw data was interpolated using cubic splines. The conversion involved only interpolation, no effort at smoothing was made. The interpolation period used was 10 hours. This makes batch to batch comparison possible without having to precisely synchronise sampling points. The use of splines for interpolating fermentation data is discussed by Thornhill *et al.* (1994).

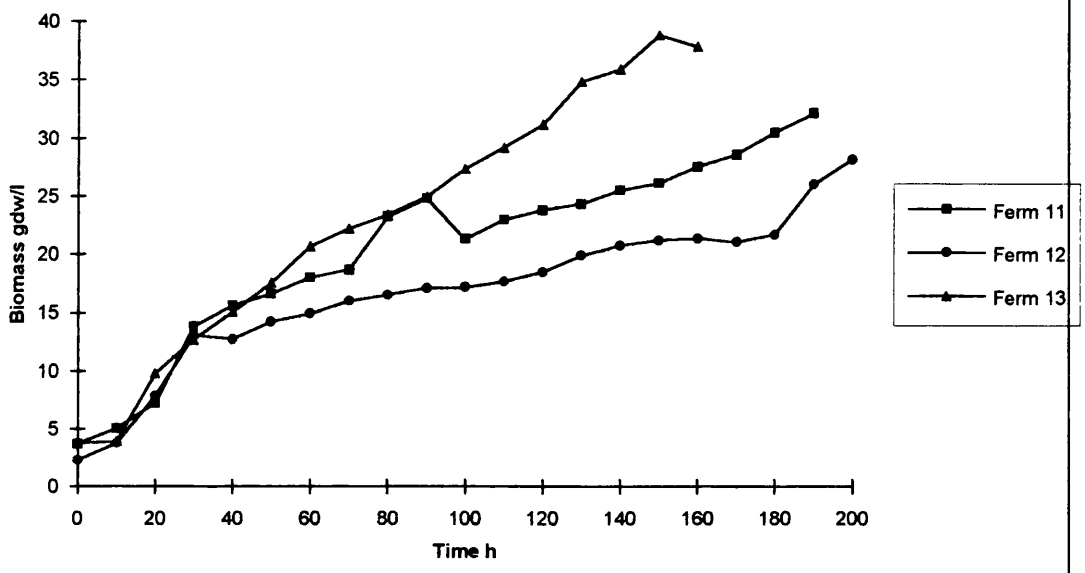
Fig. 5.1 Fermentations 4,8,9 Broth solids data



This figure shows the broth solids from the first three fermentations carried out in this project. These fermentations were intended to be consistent and to yield data for developing simulation and tracking software. The apparent event in fermentation 4 at 80 hours is believed to be a consequence of measurement error. This event highlights the risk of false alarms in fault detection systems.

Fermentations 4, 8 and 9 were run in as close to possible identical conditions. The fermentations were intended to identify the inherent variability in the fermentation process employed at UCL.

Fig 5.2 Fermentations 11,12,13, Broth solids Data



This graph shows the remaining three fermentations carried out at UCL. The fermetations differ in the rate of glucose feed, fermentation 11 getting 0.63 g/l/hr, fermentation 12 getting 0.52 g/l/hr and fermentation 13 getting 0.90 g/l/hr. This accounts for the differing broth solids trajectories. These differences were deliberately introduced to generate differing data sets, to simulate the type of differences typically seen in research fermenters used in industry. A fault detection system which could not cope with typical variability introduced in the course of an experimental program would be of little use.

Fermentation 11 was run under the same conditions as fermentations 4,8 and 9.

Fermentations 11,12, and 13 are less homogeneous than fermentations 4, 8 and 9 because variations in feed patterns (detailed above) were deliberately inserted to provide variable data.

Broth solids data from the fermentations carried out at SmithKline Beecham is less homogeneous, because a number of deliberate or naturally occurring faults ocurred. These faults are summarised in table 5.6 below.

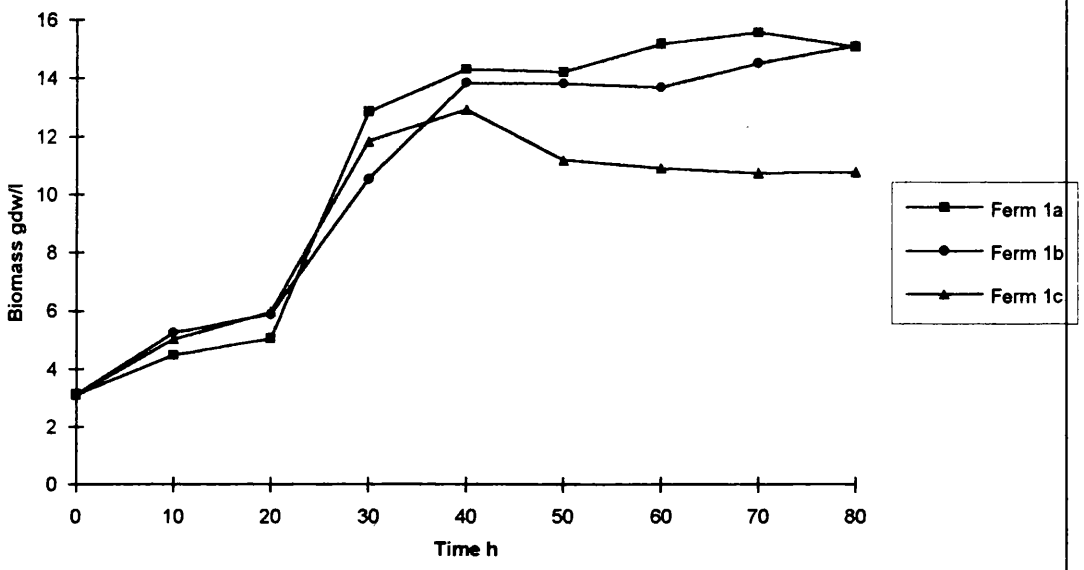
Table 5.6 Fermentation Faults

Fermentation	
1a	Mass Spectrometer fault
1b	Air filter blockage
1c	Stirrer shut down
2a	Cocci contamination
2b	Temperature shifted from 26°C to 36°C
3a	Contamination / mass spec fault
3b	Mass spectrometer fault
3c	Mass spectrometer fault
4a†	Power failure: no growth
4b†	Power failure: no growth
4c†	Power failure: no growth
5c	Temperature shifted from 26°C to 36°C
6b	Phenyl acetic acid overfeed

† These fermetations yielded no data and are not discussed further.

The broth solids data obtained from these fermentatins is shown in figs 5.3-5.7:

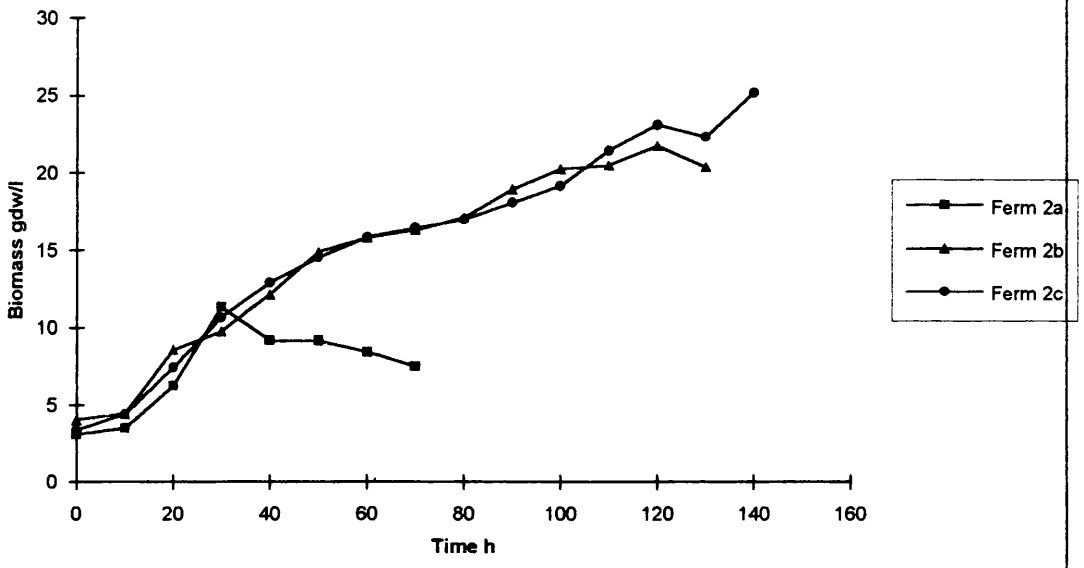
Fig 5.3 Fermentations 1 a,b,c : Broth solids Data



This figure shows the broth solids trajectories from the first fermentations carried out at SB, Fermentation 1c is interesting because it shows the impact of an agitation shutdown.

The lower rate of biomass production vs. the fermentations carried out at UCL is due to the lower rate that glucose was added to the fermenters (See table 5.4). This was a deliberate change to generate a set of significantly differing fermentation trajectories.

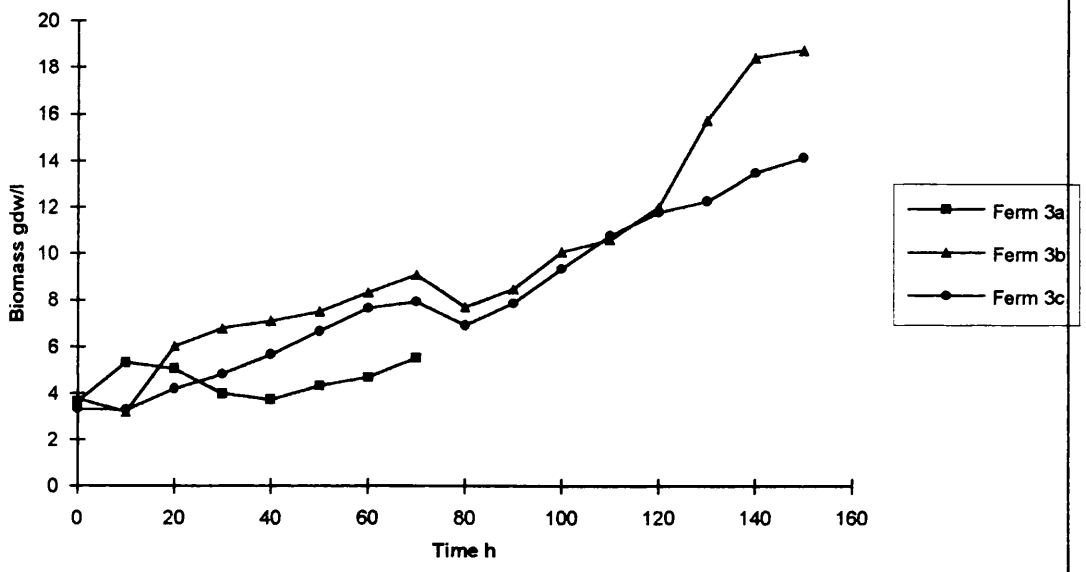
Fig 5.4 Fermentations 2a,b and c : Broth solids Data



This figure shows the broth solids trajectories for the second set of fermentations carried out at SB. Fermentation 2a shows the impact of a cocci contamination on dry weight, possibly due to poor retention of coccal biomass on the filters used.

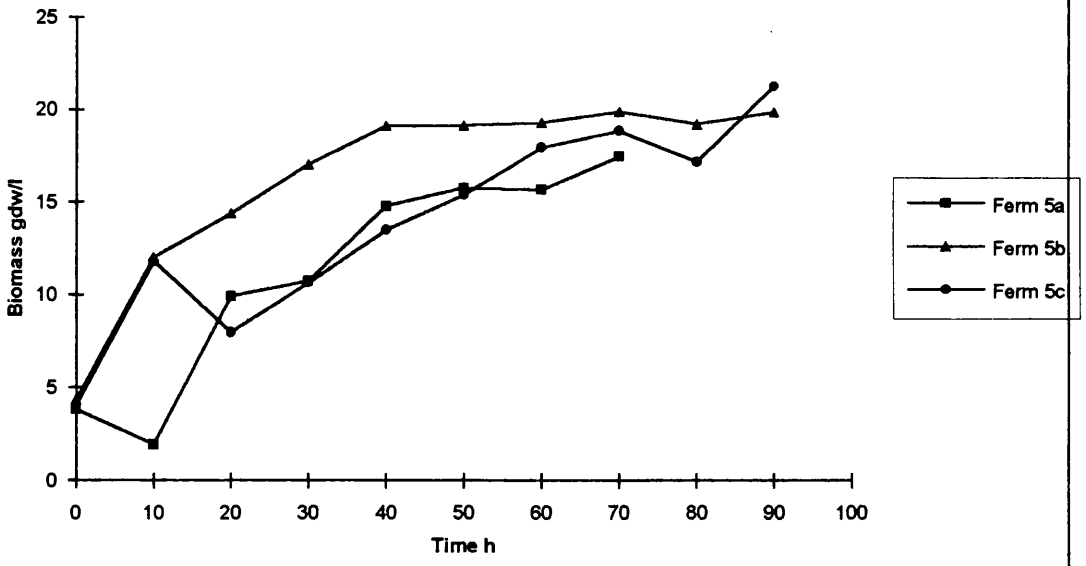
The other impact of contamination on broth solids is likely to be an increase in rate of penicillium lysis as the fungi become starved of glucose.

Fig 5.5 Fermentations 3 a,b,c : Broth solids Data



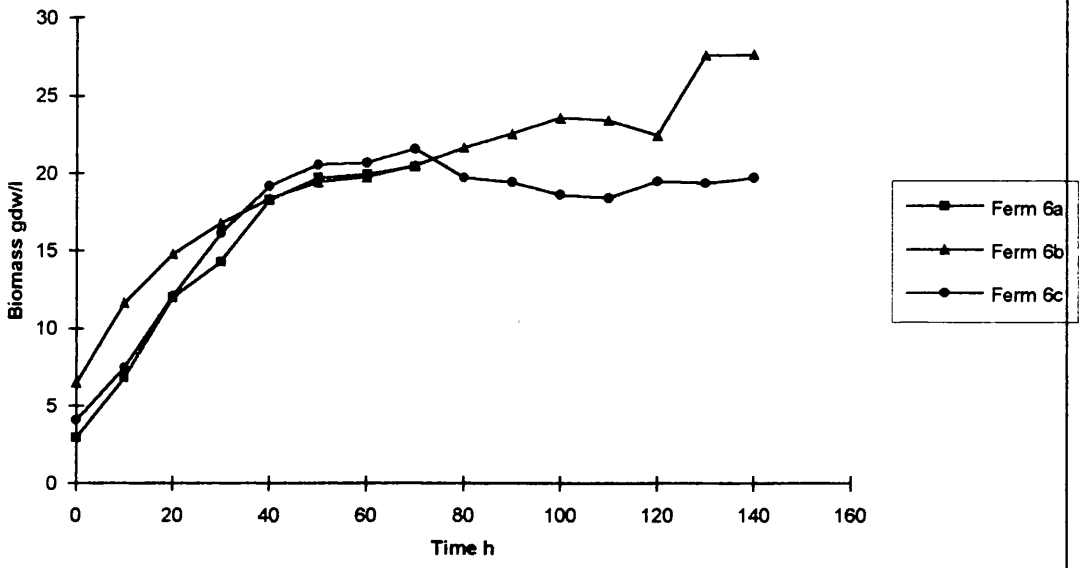
The data from these fermentations is interesting, in that fermentation 3a was subject to a fast growing contamination. It can be seen that this fermentation starts growing more quickly than the other fermentations.

Fig 5.6 Fermentations 5a,b,c : Broth solids Data



These broth solids trajectories are the closest to a group of three fermentations run under identical conditions achieved at the SKB facility. Even under these conditions there were significant differences in practice.

Fig 5.7 Fermentations 6a,b,c: Broth solids Data



The interesting feature on this fermentation was the impact of an overfeed of PAA on the development of fermentation 6c

The broth solids curves differ in the early part of the fermentations due to a somewhat higher innocculum concentration added into fermentation 6b. This difference effectively disappears as broth carbon reservoirs become exhausted and the fermentations depend on fed carbon, i.e. after 30 h.

The broth solids in fermentations 6b and 6c diverge after 7 hours as increased PAA levels inhibit further broth solids production.

5.2.2 Glucose data

Broth Glucose was tracked. This data shows that, after 20 - 30 hours, there was no discernible glucose in the fermentation broth. This permitted the completion of a carbon balance, with some qualifications about unquantified reservoirs of carbon such as corn steep liquor solids and penicillin breakdown products.

Fig 5.8 Glucose Concentration In Fermentations at UCL

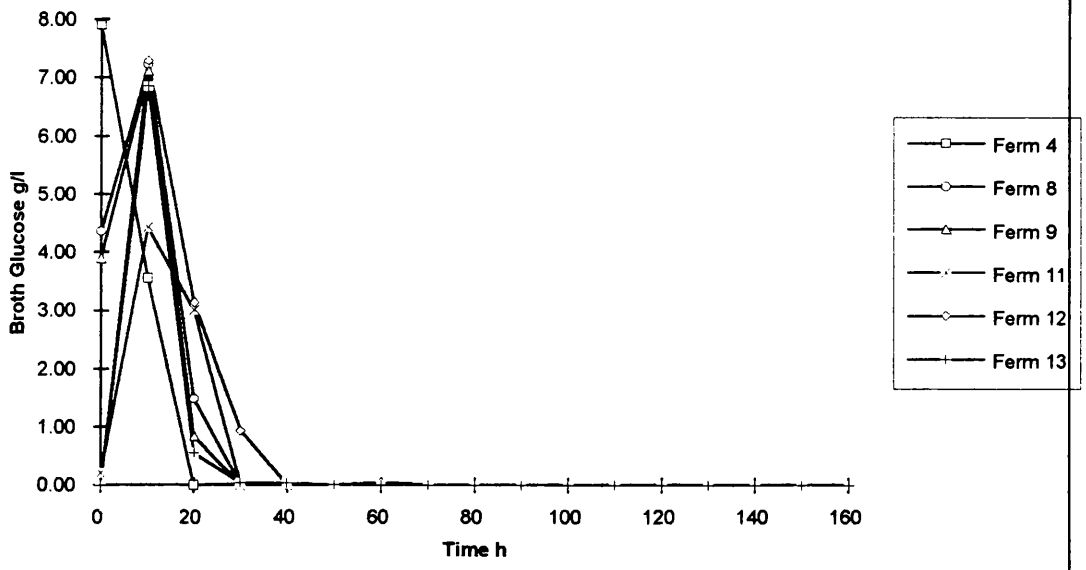
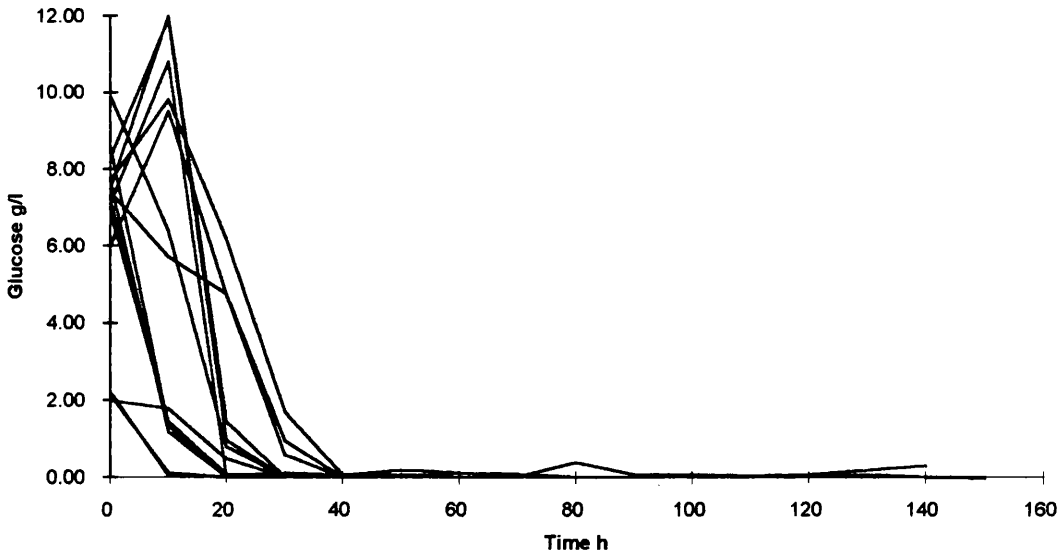


Fig 5.9 Glucose Concentration in Fermentations At SKB.



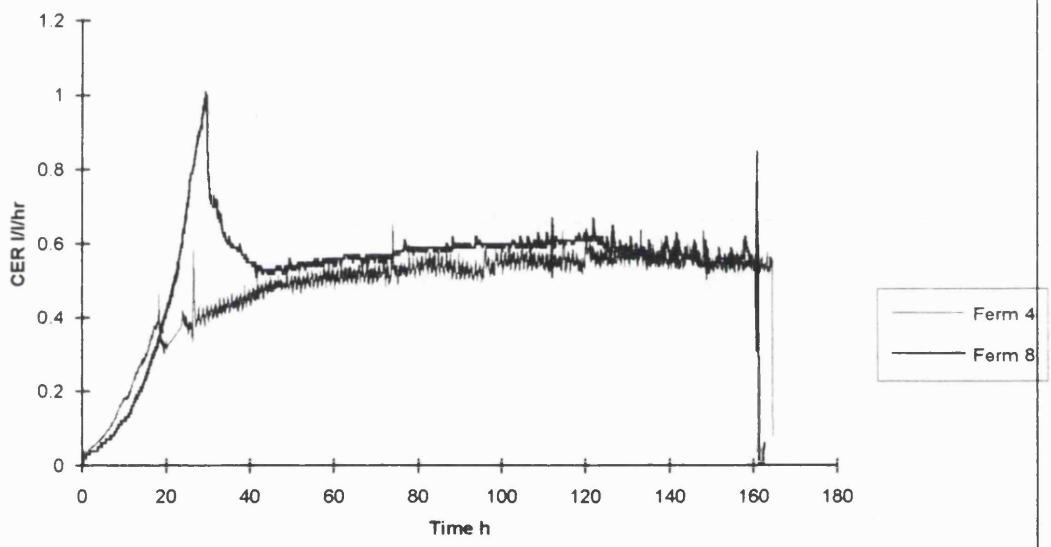
These graphs show that glucose reserves in the fermentation broth are largely exhausted between 20 and 30 hours into the fermentations. The exhaustion of glucose typically happens before the end of the exponential growth phase of the fermentations due to the presence of lactose from the spray dried corn steep liquor solids which are also added to the fermentation.

5.5.3 Carbon Dioxide Data

The key piece of on line data from the fermentations was the carbon dioxide evolution rate. variations in CER are strong indicators of disturbances in fermentations. This is because CER represents the only rate variable (as opposed to integrated variables such as total broth solids or penicillin concentration) which can be tracked accurately.

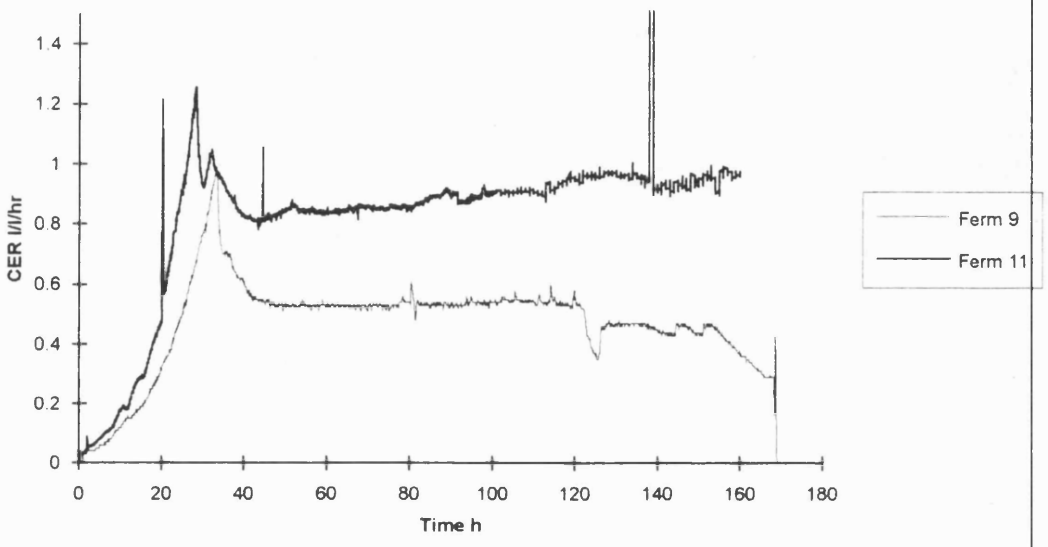
This data is shown in Figs 5.10 - 5.17

Fig 5.10 Fermentations 4 and 8: CER Data



The CER data from fermentation 4 is interesting because of its unusually high level of noise. This was due to the feeding of phenylacetic acid at 15 minute intervals rather than continuously. This lead to shifts in pH which changes carbon dioxide solubility.

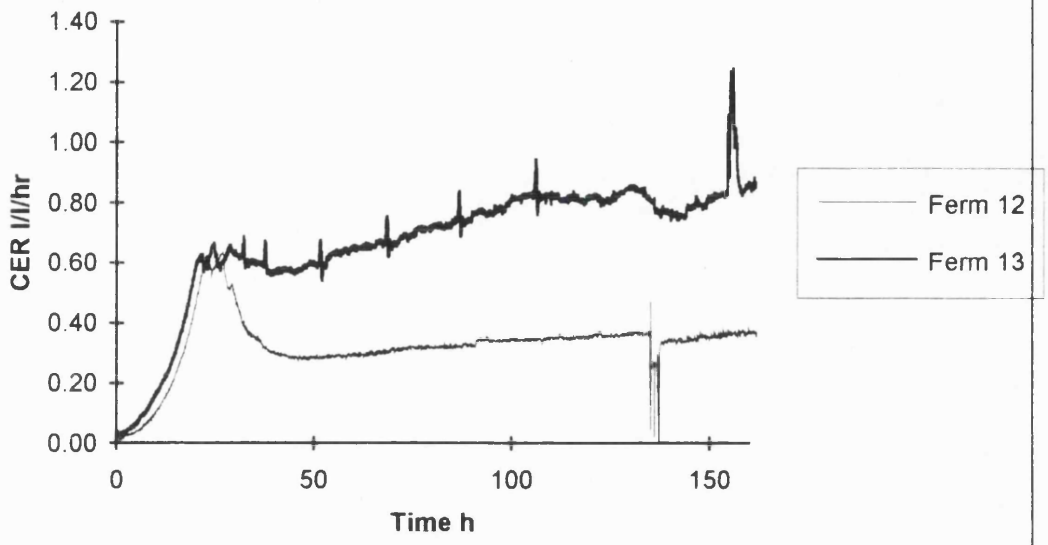
Fig 5.11 Fermentations 9 and 11: CER Data



The data from fermentation nine incorporates two fault events, one occurring at around 80h in which a filter became temporarily blocked, producing a minor disturbance in the carbon dioxide evolution rate trajectory. A more substantial fault became apparent after 120 hours. This was due to the depletion of the glucose reservoir used to feed the fermentation. This event is more fully discussed in chapter 8.

The spikes evident in fermentation 11 were entirely due to faults within the calibration software used to maintain the accuracy of the mass spectrometer.

Fig 5.12 CER data from Fermentations 12 and 13

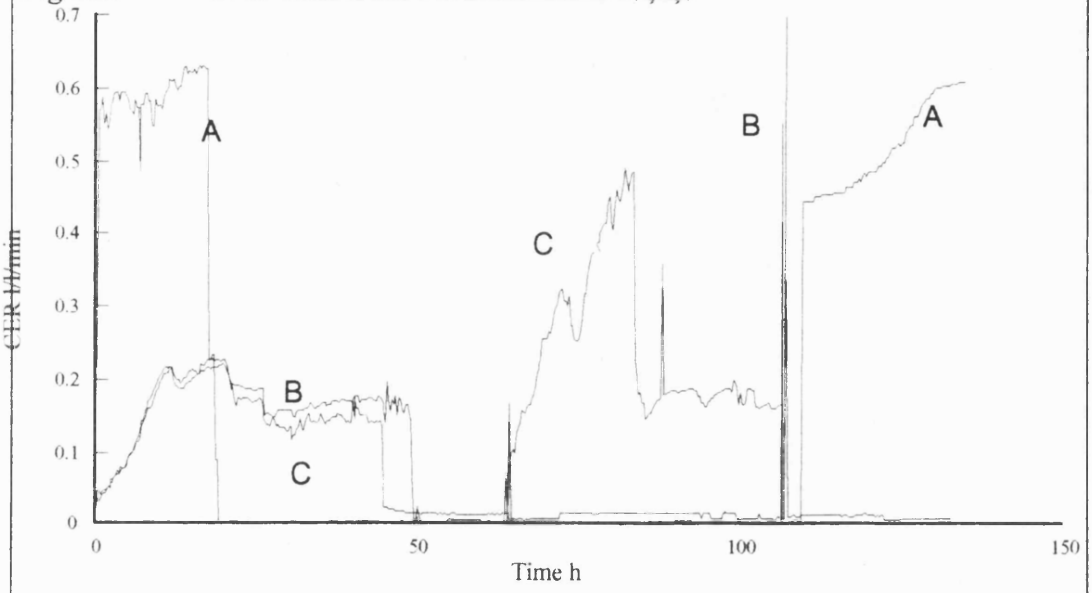


A feature of the data from fermentation 13 was that owing to a minor software fault in the mass spectrometer, the recalibration carried out automatically every 24 h shows up as spikes.

Fermentation 12 is interesting in that an exit filter became blocked at 135 hours. This resulted in an unusual fault signature which is fully analysed in chapter 8

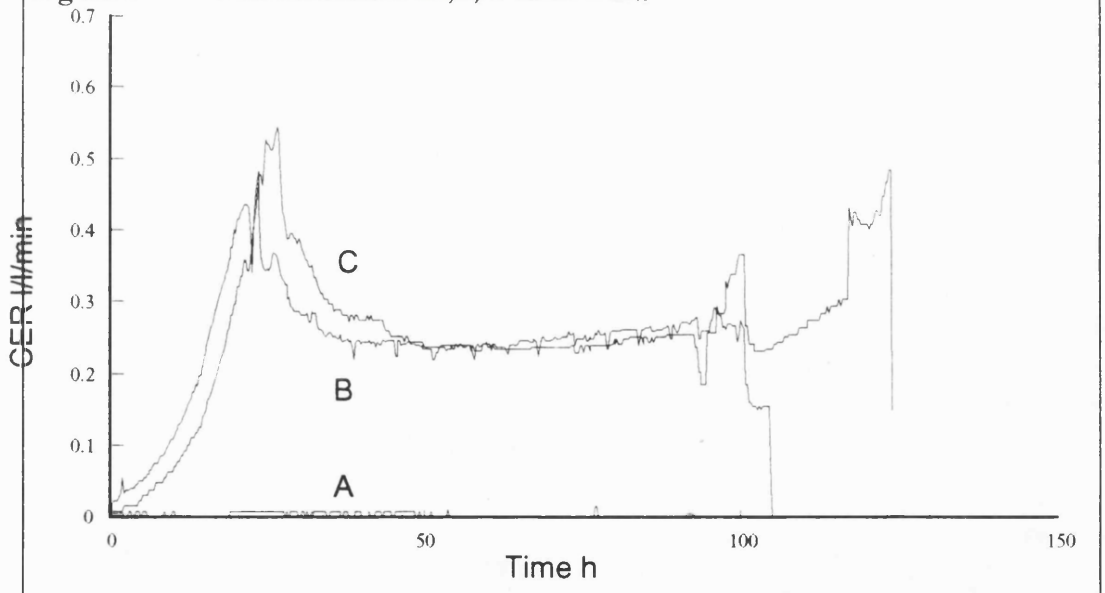
Unfortunately, as can be seen from the following graphs (Figs 5.13 - 5.16), the on line mass spectrometer used to track the fermentations at SmithKline Beecham was less reliable than the equipment available at UCL:

Fig 5.13 CER data from Fermentations 1a,b,c



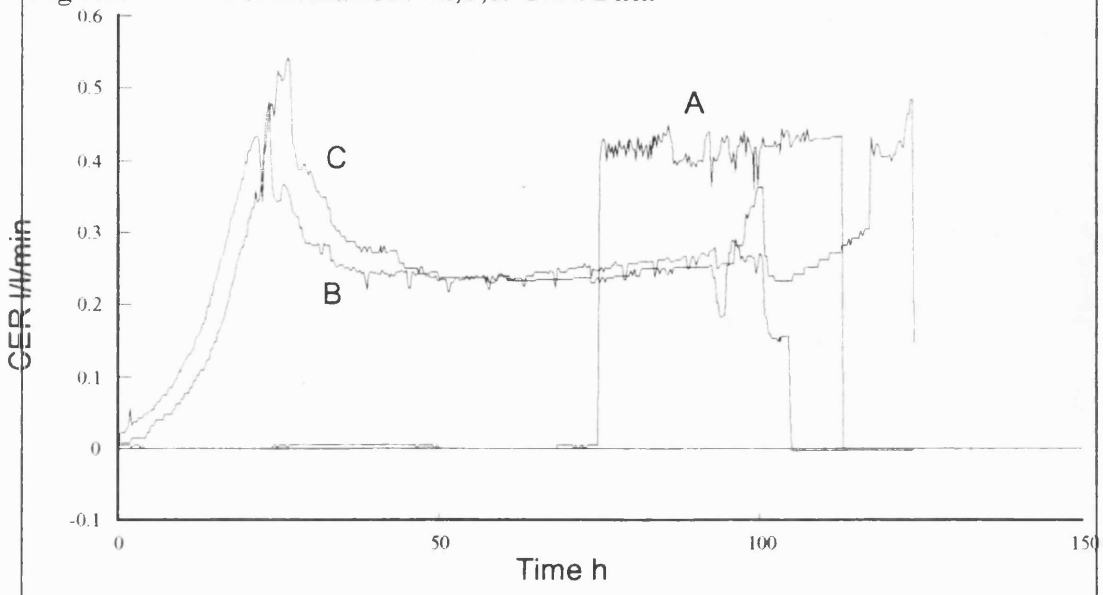
The data collected from fermentation 1A was distorted by a blocked filter in the line from the fermenter to the mass spectrometer.

Fig 5.14 Fermentations 2a,b,c: CER Data



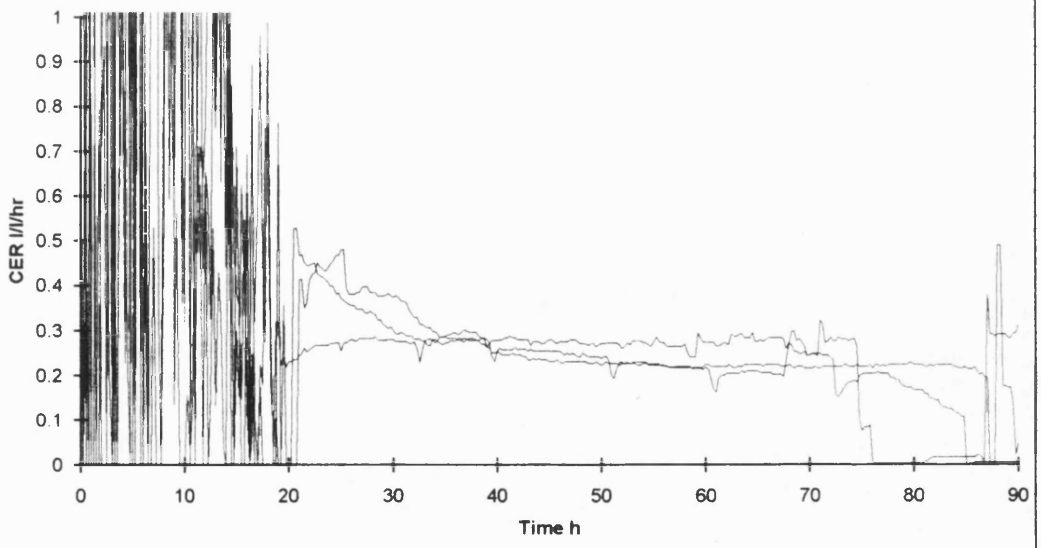
In this set of fermentations, one of the inlet lines to the mass spectrometer became detached, so no data was collected.

Fig 5.15 Fermentations 3a,b,c: CER Data



Similarly, in fermentation set 3, a fault ocurred in one of the mass spectrometer inlet lines.

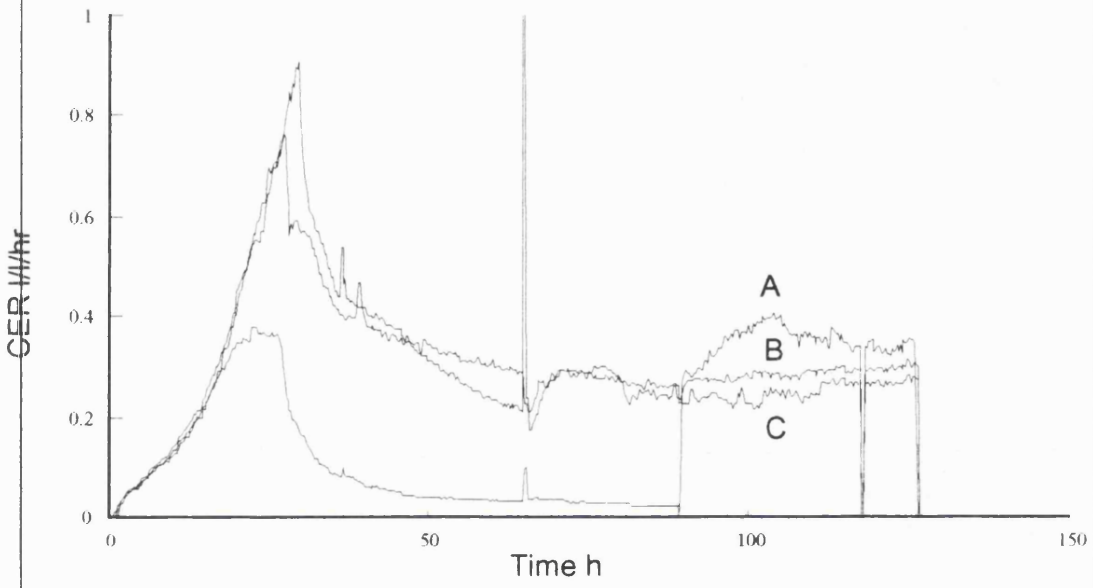
Fig 5.16 Fermentations 5a,b,c: CER Data



The mass spectrometer was functioning correctly during the first 24 hours of this fermentation but the data reporting and recording system failed.

An issue with the collection of gas data at SB at the time this work was carried out was that the systems used were of low reliability. This led to the result that the data was not used much, resulting in a fix of the reliability issue being accorded low priority!

Fig 5.17 Fermentations 6a,b,c: CER Data

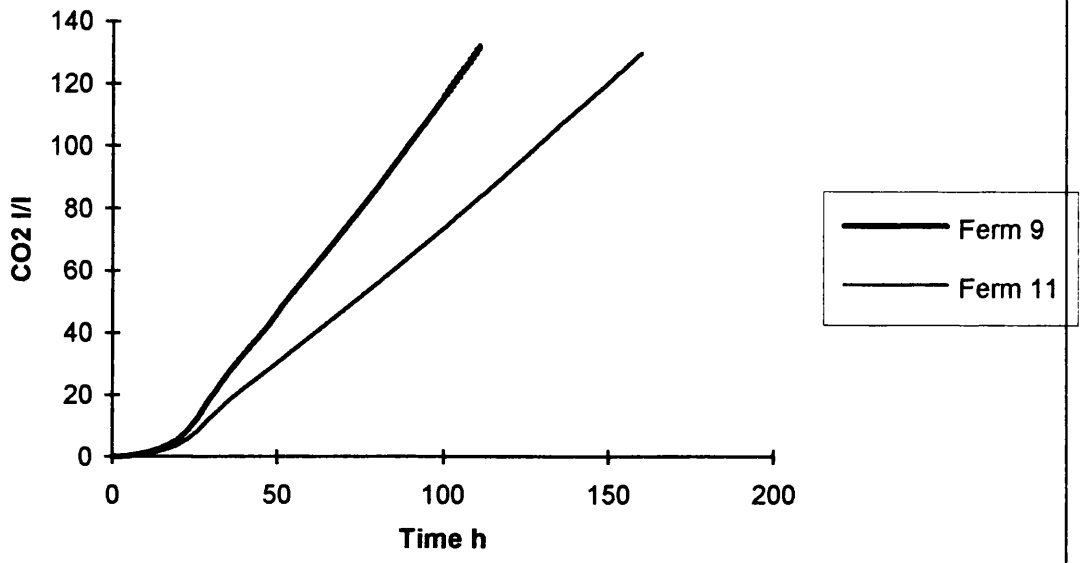


In fermentation set 6, one of the fermenters suffered an aeration fault, which resulted in the low apparent rate of carbon dioxide evolution. The calculation for CER assumed that the air flow rate was the programmed rate.

It is interesting to note the spike in the observed data at 65 hours. This ocurred at the time that the instrument was going through an automatic recalibration.

It is worth presenting a graph (Fig 5.16) showing total carbon dioxide production. This demonstrates the difference between integrated and rate variables.

Fig 5.18 Typical Carbon Dioxide Production During a Fermentation, (Ferms 9,11)



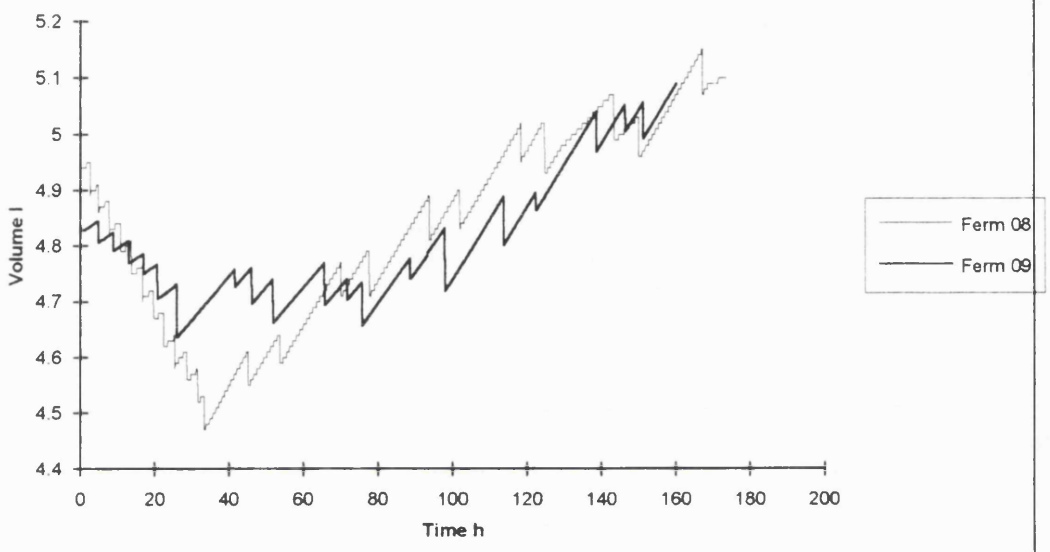
The above graphs clearly show the phases of the fed batch process, in that there is a well defined exponential growth phase, which typically lasts for 30 hours. In this part of the fermentation, the organism is growing unrestricted by nutrient limitations. The rapid increase in carbon dioxide evolution is matched by a rapid increase in broth solids. This process continues after the exhaustion of the initial glucose load (Typically this happened within 10 hours), clearly carbon is being taken up from spray dried corn steep liquor added at the start of the fermentation.

An interesting point from the CER data is that the end of the exponential phase in most fermentations happens approximately 5 hours after the exhaustion of glucose reservoirs. This is largely due to the presence of 1.5 g/l of dissolved material coming from the spray dried corn steep liquor solids. This is largely lactose, which would be expected to be metabolised once all the glucose was exhausted. In addition, there is some evidence of digestion of up to 0.5 g/l of corn steep liquor solids which would also extend the duration of the exponential phase of the fermentation.

5.5.4 Volume Data

Other key variables monitored include volume. Typical volume trajectories, for fermentations 9 and 11 are presented in fig 5.17:

Fig 5.19 Typical volume data (Fermentations 8,9)

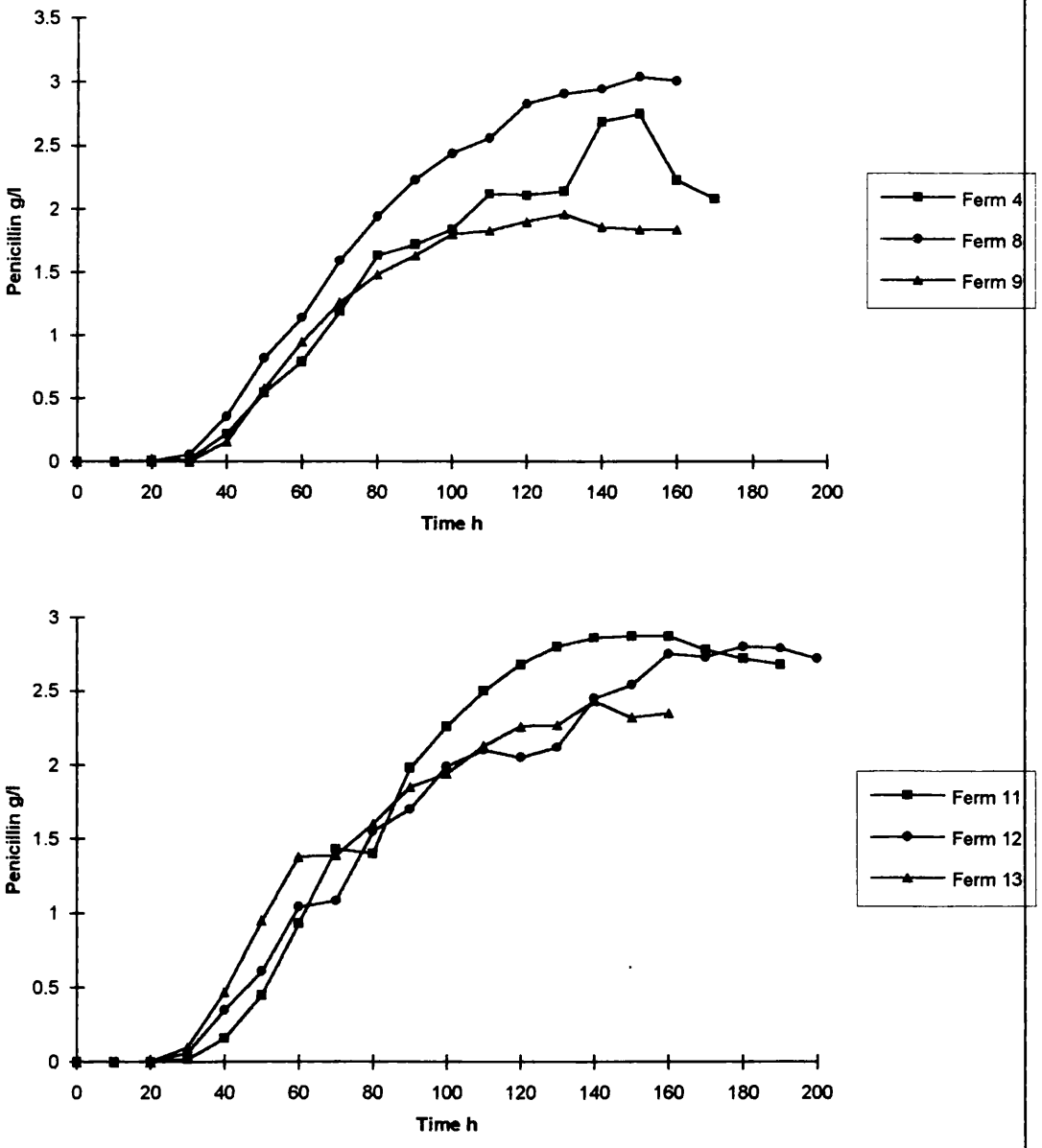


This saw tooth pattern is due to the periodic removal of samples, followed by steady accumulations due to addition of nutrients.

5.5.5 Penicillin Production

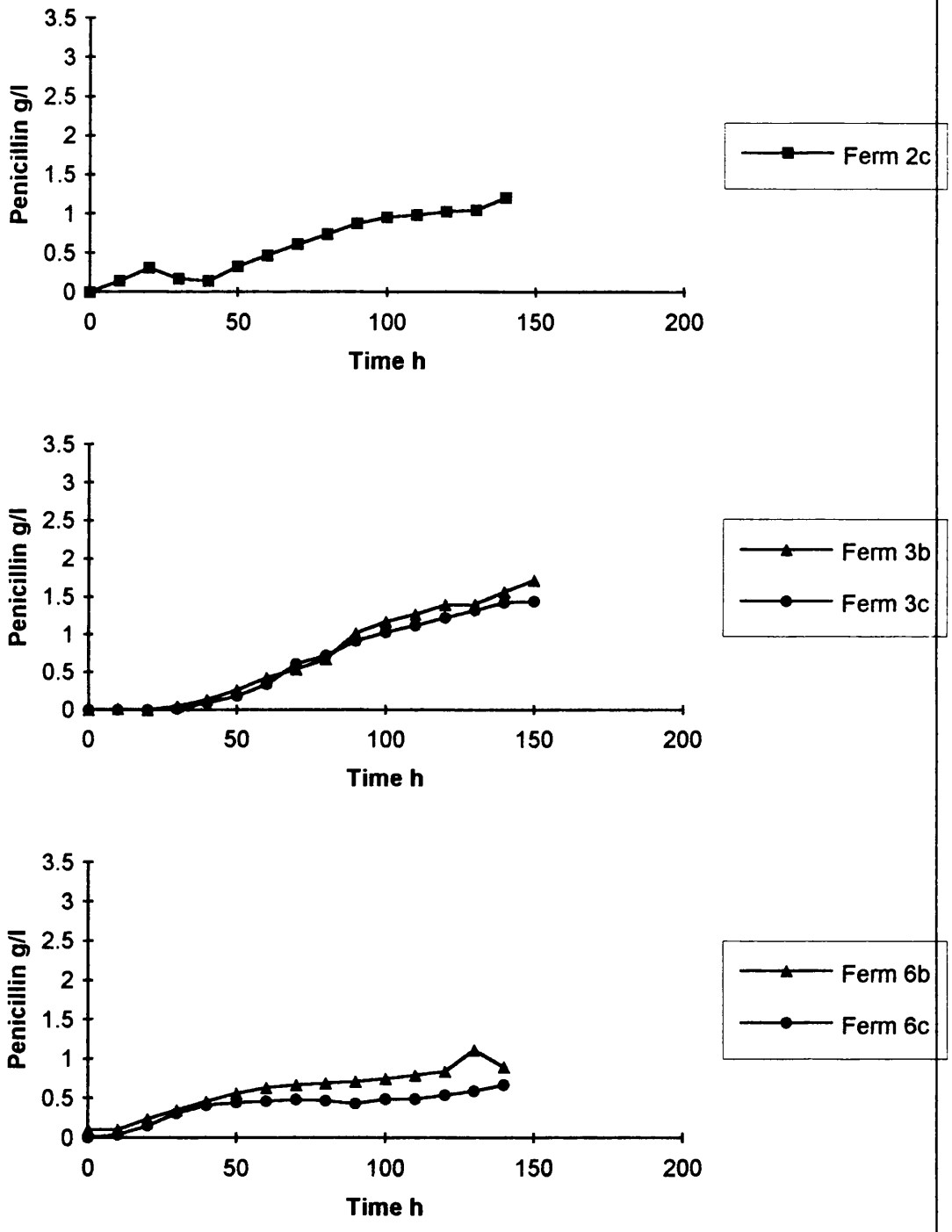
The following penicillin production was obtained:

Fig 5.20 Penicillin production from UCL fermenter, (Fermentations 4,8,9,11,12 and 13)



The fermentations carried out at SmithKline Beecham produced the following results.

Fig 5.21 **Penicillin production from fermentations carried out at SmithKline Beecham.**

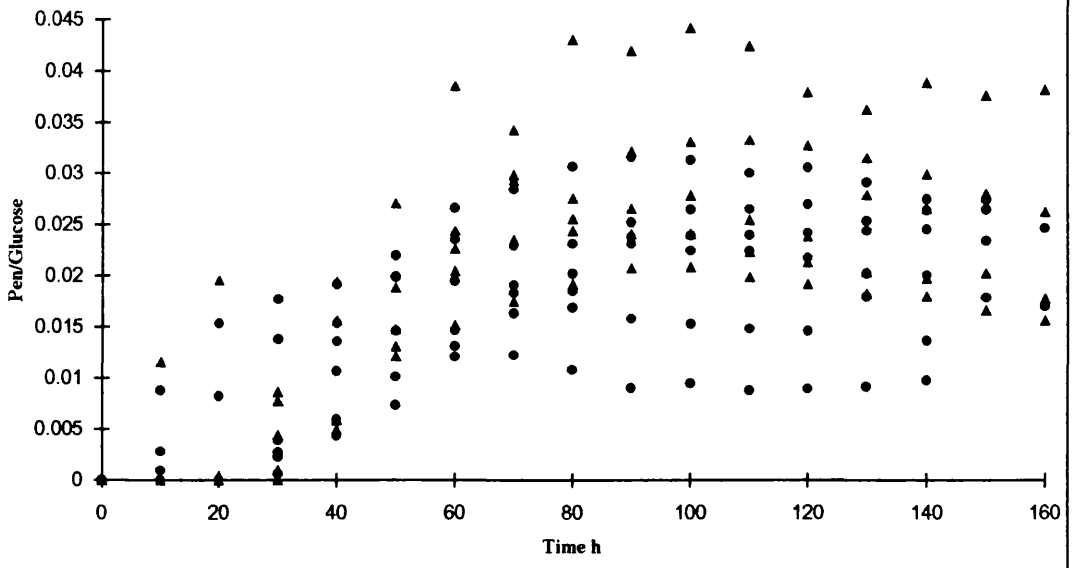


It is clear that different levels of penicillin production / unit volume were obtained in fermentations carried out at UCL and SKB. This is due to the differing feed patterns used, Fig 5.22 shows the relative yields of penicillin on glucose obtained at each site.

Only the fermentations which produced substantial penicillin titres are reported.

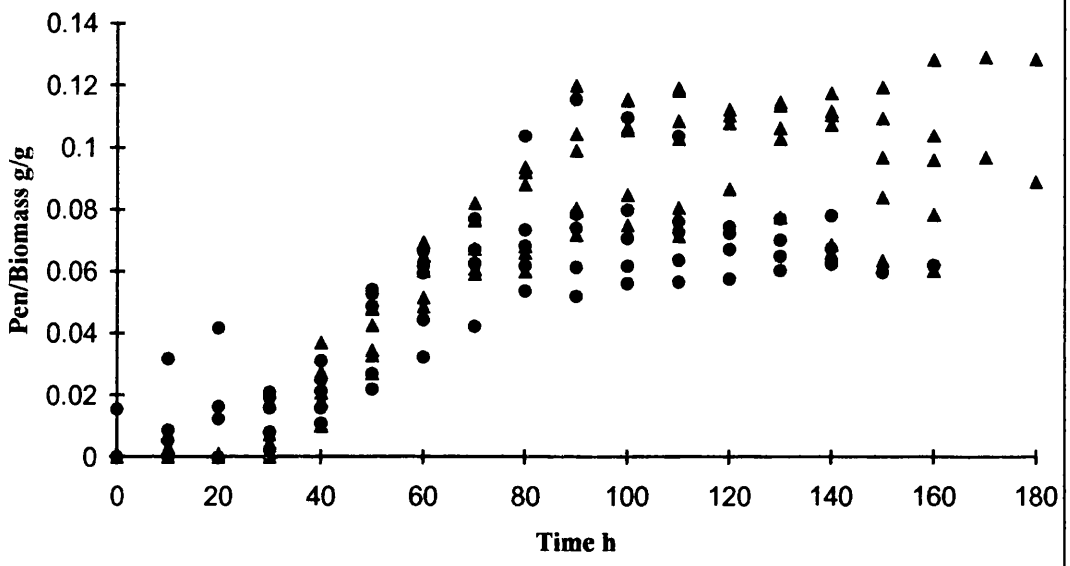
It is interesting to compare the yield of penicillin on glucose, and the ratio of penicillin produced to biomass obtained at each site.

Fig 5.22 Comparison of Penicillin Yield on Glucose Between batches produced at UCL (▲) and SmithKline Beecham (•)



This chart shows a comparison between the penicillin yields achieved on glucose in fermentations carried out at UCL and SKB. It can be seen that there tended to be a lower yield at SKB. This is probably because the fermentations carried out at SKB used a lower rate of addition of glucose. The result of this would be that a greater proportion of this glucose would be used up in cell maintenance activities and broth solids production. It can be seen in graph5.23 that typically fermentations carried out at SKB produced less penicillin per gram of broth solids produced than fermentations at UCL.

Fig 5.23 Ratio of penicillin production to broth solids production in fermentations carried out at UCL (▲) and SmithKline Beecham (•)



This chart shows that the feed conditions used at UCL generally resulted in a more efficient trade off between biomass and penicillin production. This accounts for the lower yields of penicillin found in the fermentations carried out at UCL vs. those carried out at SKB.

Figs 5.22 and 5.23 show that more of the available energy was used to produce and maintain biomass in the relatively restricted glucose availability conditions available at SKB.

Note: Although penicillin titres were measured at UCL and SKB using different analytical methods (see section 4.3.3 for details) both methods were calibrated frequently (daily at SKB and prior to use at UCL) using penicillin standard solutions of known concentration covering the range in which measurements were being made. This provides reassurance that the penicillin assays were reliable at both sites.

5.6 Conclusions

The data presented in this chapter show that real fermentation data is subject to all sorts of process events generating noise and one off discontinuities. This data is typical of real world fermentation output in the case where the fermentations are being used to generate experimental data on for example the productivity of organisms.

This data contains uncertainties due to the difficulties of measuring key data. This is a particular issue in identification of biomass. The practical measurement available is that of dry weight. This does not differentiate between solids remaining from the nutrients added at the start of the fermentation, dead cell mass and active biomass. This picture is complicated further by uncertainties about how much of the initially insoluble CSL solids are fermentable.

The other uncertainty in this data is the tracking of evaporation. In the data presented, evaporation was accounted for by trapping exiting moisture in silica gel crystals. The rate of moisture loss was assumed to be linear, as aeration rates were constant. Therefore, it was possible to retrospectively account for water removed by evaporation into the aeration gas stream. Tanks mounted on load cells would have permitted on line accounting for evaporation.

It is data of this quality on which any on line monitoring system will have to work.

Chapter 6 Data Analysis

It is possible to extract additional information from a body of raw fermentation data that can add to what is known about an ongoing fermentation. This data can be at three levels of abstraction.

1) Data collected on line and by off line measurements on broth samples can be organised into a format in which it can be cross checked for consistency.

2) Parameters that are not directly observable can be inferred, such as the specific growth rate of the micro-organisms in a fermentation, the yield of biomass, product, and carbon dioxide on carbon sources. This represents a useful method of normalising data from fermentations that differ in feed patterns.

3) Other information can be incorporated. The most useful source of data was that obtained from prior fermentations. Information from this source is in the form of expected fermentation trajectories directly obtained from prior data, or models, whose parameters are essentially a means of recording data from past fermentations. Information from current fermentations can be compared against data from past fermentations and tested for consistency.

As will be seen in chapter 8, the most important information from a fault detection perspective is the ratios of process parameters. When the correct parameters are expressed as a ratio, variations between fermentations run under different conditions can be minimised. This means that the fermentation of interest can confidently be compared with past fermentations.

It should be noted that in this chapter, a distinction is made between biomass, and the measured level of solids in the fermenter broth. Although the material recovered in a dry weight assay is mainly biomass, there is 0.6-0.75 g/l of solids left over from the innocculum nutrients, largely due to insoluble and indigestible components of corn steep liquor. Dry weight data is therefore discussed as broth solids rather than biomass level. Changes in broth solids concentrations over time however will be exclusively due to changes in biomass concentration.

This chapter is intended to report the organisation of the fermentation data presented in the previous chapters into such useful information streams.

Table 6.1 Summary of processed data presented in this chapter.

Parameter	Discussion
Direct Measurements	Broth solids concentration Penicillin concentration
Rate Data	Carbon dioxide Evolution Rate Broth solids Production Rate Penicillin Production Rate
Yield Data	Yield of broth solids on glucose Yield of penicillin on glucose Yield of penicillin on phenylacetic acid
Production Ratios	Penicillin / broth solids Carbon dioxide evolution / glucose addition Broth solids / carbon dioxide
Mass Balance	Overall carbon balance

The first step in the analysis of the fermentation data was to use cubic splines to interpolate the off line data to give estimates of fermentation data at ten hour intervals. This permitted the employment of statistical analysis to establish a trajectory within which a properly functioning fermentation could be expected to develop.

For the purposes of fault identification, the most useful trajectory is that provided by the on-line reconciliation of carbon fed into the fermentation, carbon accumulated, and carbon removed. This reconciliation is dependent on the rate of addition of glucose, phenyl acetic acid and air to the fermentation, and is also affected by fermentation failures such as contamination by micro organisms that produce β-lactamase. As will be seen (Chapter 8) other fermentation faults such as temperature shifts will also affect the mass balance.

For the purpose of compiling the graphs presented in this chapter, data was drawn from fermentations known to have performed correctly. That is that they were not subject to faults such as hardware failures or contamination.

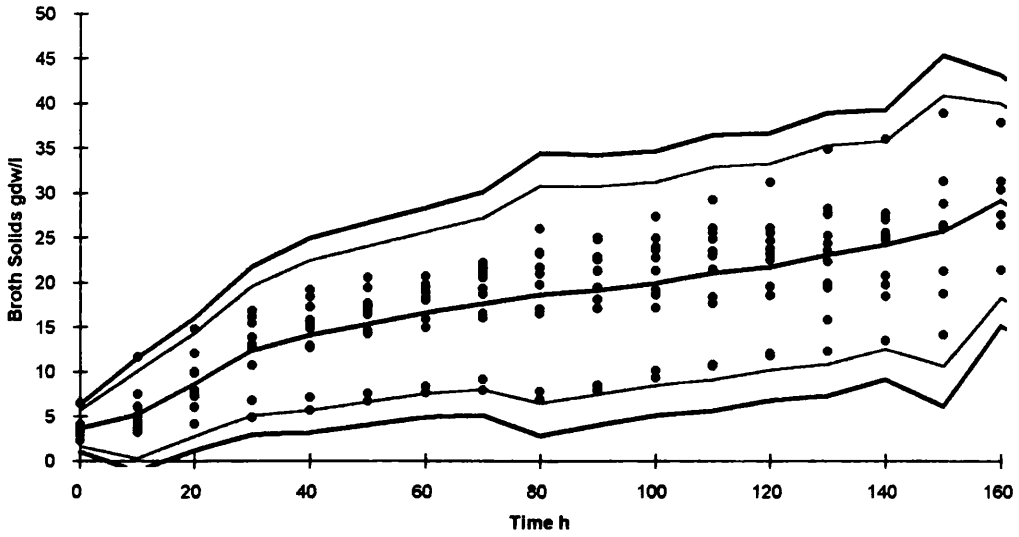
6.1 Data Obtained By Direct Measurement

Direct measurement by the methods described in chapter 4, (Materials And Methods) produced the data presented in the previous chapter. All the data presented here is from fermentations, or portions of fermentations which were running to plan, and in which no fault had as yet occurred.

In this chapter graphs are presented showing mean, 2σ and 3σ confidence limits for each of the data reported. The choice of 2 and 3 standard deviations reflects typical industrial statistical process control practice. If the data is normally distributed, 2 standard deviations would encompass 95% of the data, and 3 standard deviations 99.5% of the data. The actual data are shown as fainter broken lines.

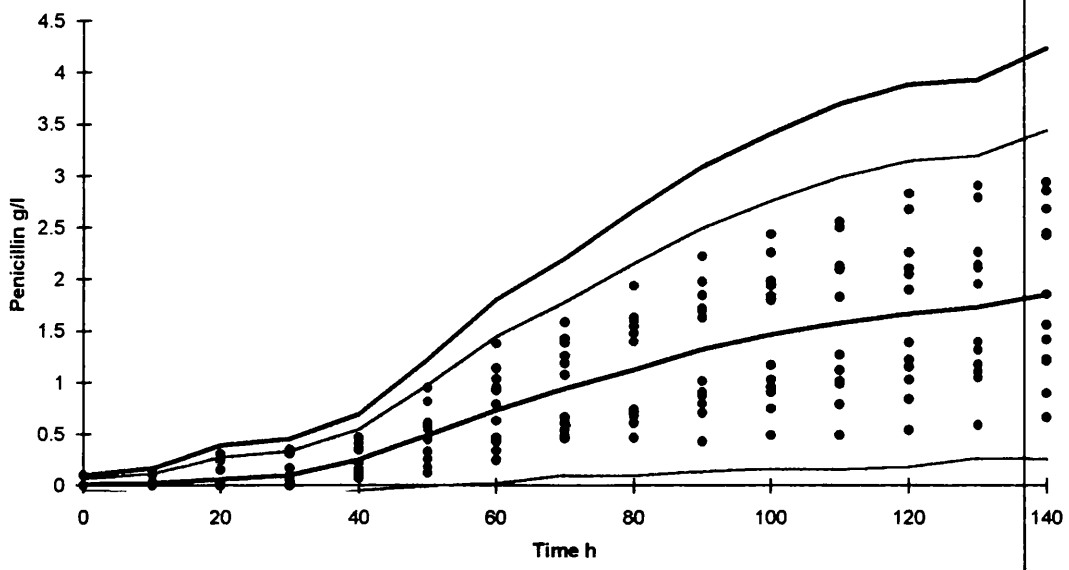
In a fault detection scheme, defective fermentations should show up by having broth solids concentrations in samples outside the confidence limits at the point in the fermentation at which the sample was drawn.

Fig 6.1 **Expected Broth solids Trajectory Limits, showing mean, 2σ and 3σ confidence limits.**



Note: In this chapter bold lines are used to denote mean, 2σ, and 3σ trajectories. The central bold line denotes mean, the next bold lines out represents the 2σ trajectory, and the outer bold lines represent the 3σ trajectory. Raw data from all fermentations yielding useful information is shown as points on the graph.

Fig 6.2 Expected Penicillin Production Limits



This graph and the previous broth solids data (fig 6.1) illustrate the level of scatter from fermentations run to plan using different starting and feed conditions. This data is virtually unusable for fault detection purposes because the limits of detection which it yields are so broad.

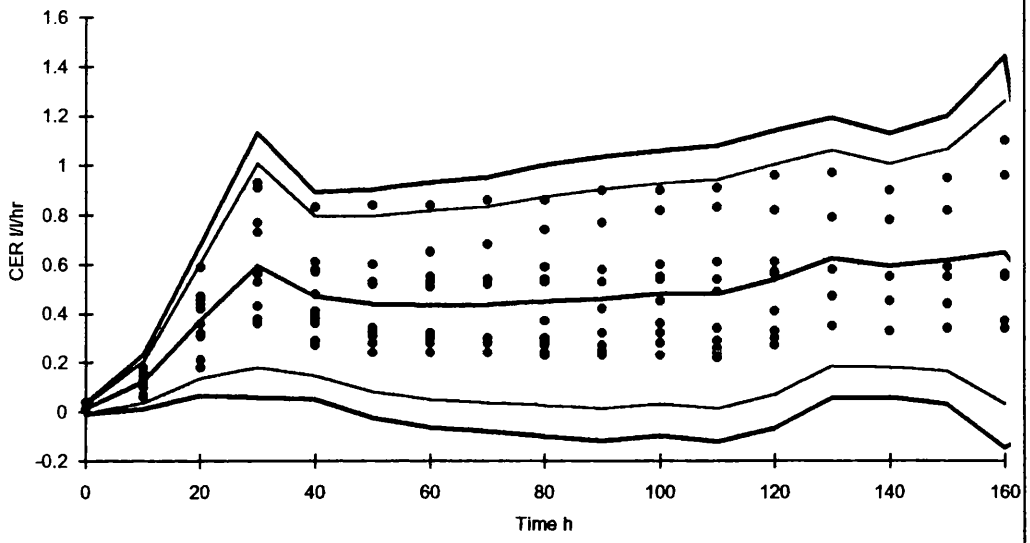
Directly measured fermentation parameters can be expected to vary widely especially in fermentations conducted for R&D purposes, such as the fermentations conducted in the small scale fermenter suites in pharmaceutical companies' research and development laboratories, which is the subject of this thesis. This is because these fermentations are used to investigate the impact of changes in operating conditions, medium components and new organism strains. A useful extension to this work could be to group fermentation data from experiments with similar objectives or characteristics.

6.2 Rate Data

In addition to the production of broth solids and penicillin basic rate data was derived. This is of interest in that, in a correctly functioning fermentation, the rate of broth solids production and carbon dioxide evolution is a function of the rate of nutrient addition and the maintenance of suitable growing conditions in the fermentation. Such rate data is likely to be more reproducible than direct

parameter data. A rapid change in rate data is indicative of a number of possible faults as will be seen in chapter 8.

Fig 6.3 **Carbon Dioxide Evolution Rate Data : Mean, 2σ and 3σ confidence limits.**



Carbon dioxide evolution per litre of fermentation, based on data from all batches, unless obvious or planned faults were in progress. As with biomass and penicillin data, the raw CER data is not useful in establishing that a fault has ocurred.

6.2.1 Specific Growth Rate

Specific growth rate is defined as

$$\mu = \frac{dX/dt}{X} \qquad 6.1$$

μ = Specific Growth Rate g/g/hr

X = Broth solids g

t = Time h

The following equation was used to estimate specific growth rate:

$$\mu_k = \frac{(X_k - X_{k-1})/\Delta t}{(X_k + X_{k-1})/2} \qquad 6.2$$

The time increment in the above equation is taken as the time interval over which splines were calculated (10 hours).

The differential nature of the calculation and the sparse nature of the available dry weight concentration measurements resulted in rather disturbed time courses of specific growth rates. However, two phases of growth can be distinguished in all time courses, an initial period of rapid growth followed by a decline in growth rate to a low value for the remainder of the fermentation, which corresponds to the production phase. The disturbed nature of the specific growth rates makes it difficult to determine the values of maximum specific growth rate (μ_{max}), and the specific growth rate during the production phase, accurately. However, the values of μ_{max} and the mean values of μ during the production phase, identified for each successful fermentation are presented below.

Table 6.1 Observed Specific Growth Rate

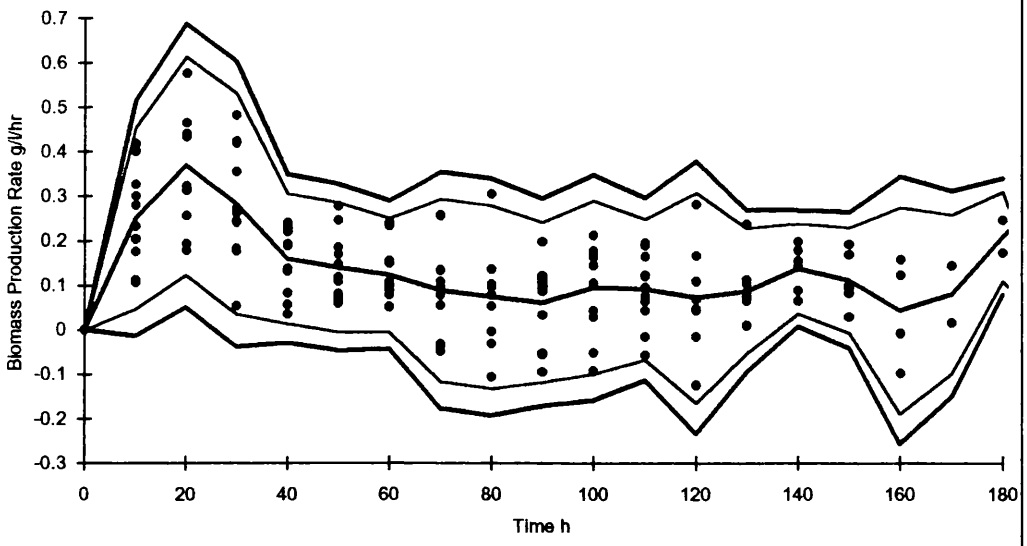
Fermentation	μ_{max} (h^{-1}) †	Mean μ during the production phase (h^{-1})
FERM04	0.068	0.016
FERM08	0.086	0.015
FERM09	0.065	0.014
FERM11	0.063	0.013
FERM12	0.070	0.012
FERM13	0.086	0.015
FERM2C	0.051	0.013
FERM3B	0.061	0.014
FERM3C	0.053	0.015
FERM5B	0.057	0.012
FERM5C	0.067	0.014
Overall Mean	0.065	0.014

† μ estimated using equation 6.2

The calculated values of μ_{max} are lower than, for example, those of Mou and Cooney (1983a) or Nestaas and Wang (1981) who reported values of 0.11 h^{-1} and 0.17 h^{-1} respectively for *Penicillium chrysogenum* P2 grown under similar conditions to those used in this work. This could be because of the presence of non-biomass solids in the medium, which made the calculation of specific growth rate inaccurate during the growth phase in this work. This is unlikely however, since such solids are unlikely to significantly change in concentration from sample to sample (although they will decline over the course of the fermentation as they are diluted by input streams of glucose, PAA and ammonium hydroxide streams)

The mean specific growth rates during the production phase are qualitatively consistent with the glucose feed rates. In practice, the rate of production of broth solids varied widely due to the differing feed rates used.

Fig 6.4 Specific Growth Rate (μ_x) Trajectory Limits

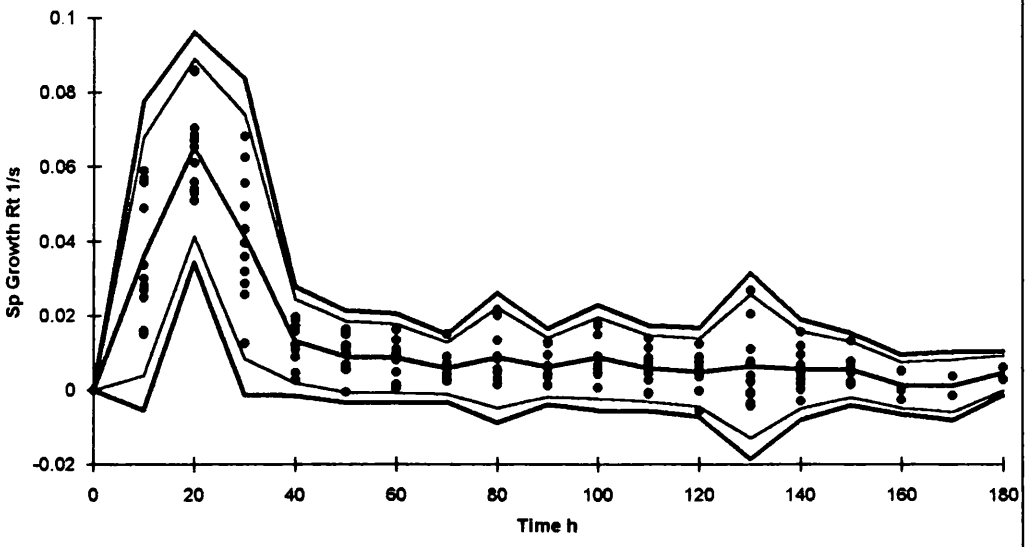


Specific growth rate data. This is based on broth solids accumulation / litre of broth per hour. The presence of solids from corn steep liquor does potentially impact specific growth rate figures because the concentration of these solids may be decreasing over the first 30-40 hours, for a discussion of this issue, see section 5.5.1. The data is somewhat depressed by the fact that the broth is continuously diluted. This is because the rate of material addition due to feeds and pH control exceeds the rate of material lost into the exit gas stream and sampling.

The relatively larger errors seen in the first 30 hours on this chart is largely due to differences in timing of peak growth rates, as some fermentations went from lag to exponential growth phase earlier than others, and some depending on glucose feed regimes continued to grow exponentially for a few hours more than others.

A better comparison of fermentations can be obtained by dividing specific growth rate by carbon fed. This reduces the impact of differing feed profiles as can be seen in fig 6.5.

Fig 6.5 **Normalised specific growth rate (μ_x /S) trajectory limits showing mean, 2σ and 3σ confidence limits.**



This figure illustrates the usefulness of normalising fermentation data for batch to batch comparison. In effect, the basis for comparison is now between the relative yields of broth solids on glucose $Y_{x/s}$, rather than specific growth rates. The parameter $Y_{x/s}$, is likely to be significantly more stable than μ_x in fermentations where glucose addition rates are varied.

6.2.2 Penicillin Production Rate (m_P)

Penicillin production rate can be defined as

$$\mu_P = \frac{dP/dt}{X} \qquad 6.3$$

P = Penicillin concentration g/l

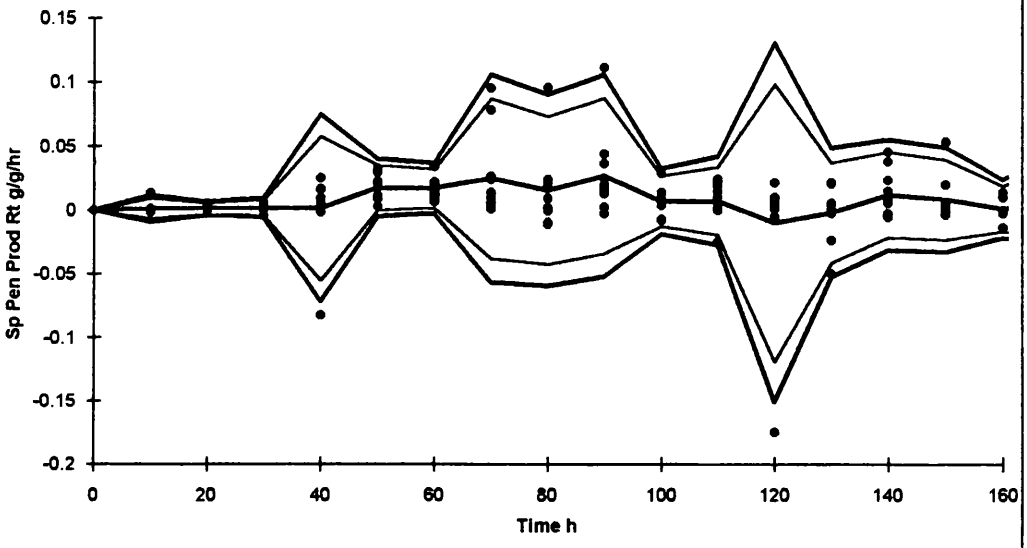
μ_p = Specific penicillin production rate g penicillin / g broth solids (dry wt) / hr

The specific penicillin production rates (m_P) were calculated for all fermentations using Equation 6.4

$$\mu_{P(k)} = \frac{(P_k - P_{k-1})/\Delta t}{(X_k + X_{k-1})/2} \qquad 6.4$$

The specific production rate for all fermentations are presented below

Fig 6.6 Specific Penicillin Production Rate, showing 2σ and 3σ confidence limits.



This graph neatly illustrates why Gaussian distributions cannot be applied, the distribution is distorted by outliers. This is the reason that 2 and 3 standard deviation confidence limits are shown rather than 95% and 99% confidence limits.

This data also shows the difficulties of working with limited data sets. More linear confidence limits can be expected given a larger data base

The m_p data is of limited value for fault detection purposes as it is highly sensitive to measurement noise. Each data point is obtained from a combination of four measurements, two of broth solids and two of penicillin concentration.. The maximum values specific penicillin production rate, excluding obvious outliers was 0.03-0.04 h^{-1}. These values are comparable with those reported by Cooney (1979) of 0.04 h^{-1} (reported as 6000 units.$(g.cell)^{-1}.h^{-1}$, where 10^6 units = 0.667 g penicillin G)

The conclusion from this section is that it is necessary to process the data with a view to compensating for batch to batch process variations if an effective fault detection scheme is to be constructed.

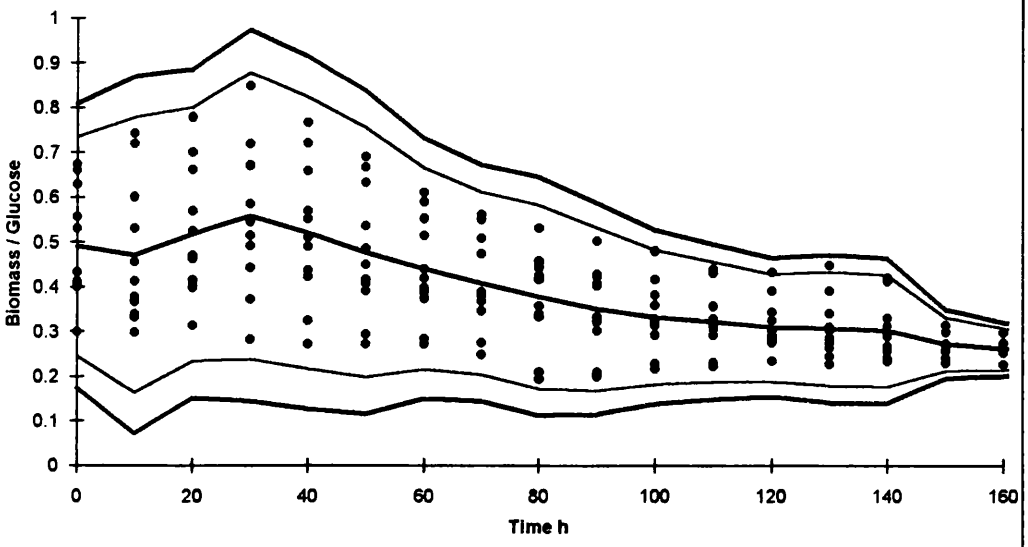
6.3 Normalised Data

Clearly the variation on directly measured parameters such as broth solids, carbon dioxide evolution rate, specific growth rate and specific penicillin production rate is wide. This reflects differing fermentation conditions due to both deliberate process variability, for example variability caused by different feeding patterns, and natural variability due to factors such as inoculum titre.

6.3.1 Yield Of Broth solids On Glucose ($Y_{X/S}$)

Dividing the broth solids concentration by the amount of glucose added, the following graph can be obtained.

Fig 6.7 **Broth Solids : Glucose addition ratio. Shows mean, 2σ and 3σ Confidence limits.**



This graph shows the ratio of broth solids generated to glucose added. The shape reflects the uncertainties at the start around the seed inoculation composition, and also the impact of adding varying levels of spray dried corn steep liquor solids and other initial carbon containing compounds. As these become fully metabolised, the ratio of glucose / broth solids stabilises.

The key driver of the graph shape is however the differing fermentation starting points. Initial glucose levels vary from 0 - 8 g/l. When these reservoirs have been used up, after the first 30 hours confidence limits begin to converge.

The overall yield of broth solids on glucose averages at 0.38g/g in this work, compared with a figure of 0.47g/g quoted by Mou and Cooney (1981). While it is difficult to make comparisons between fed batch secondary metabolite fermentations unless similar conditions are used, it is possible that the lower value reflects the extended steady state phase of the fermentation, in which glucose will have been converted to carbon dioxide and penicillin without extensive production of new biomass. Mou and Cooney report fermentations lasting for 110 to 140 hours, while this work considers fermentations lasting for up to 200 hours.

6.3.2 *Penicillin Production / Glucose Fed*

The ratio of penicillin production to glucose feed, unlike the specific penicillin production rate described above produces data sets which remain within the 3σ confidence limits in all cases in which the fermentations were believed to have proceeded to plan, as can be seen in figure 6.9.

Fig 6.8 **Ratio of Penicillin Production to Glucose Feed (i.e. cumulative yield of penicillin on glucose). Shows mean, 2σ and 3σ confidence limits.**

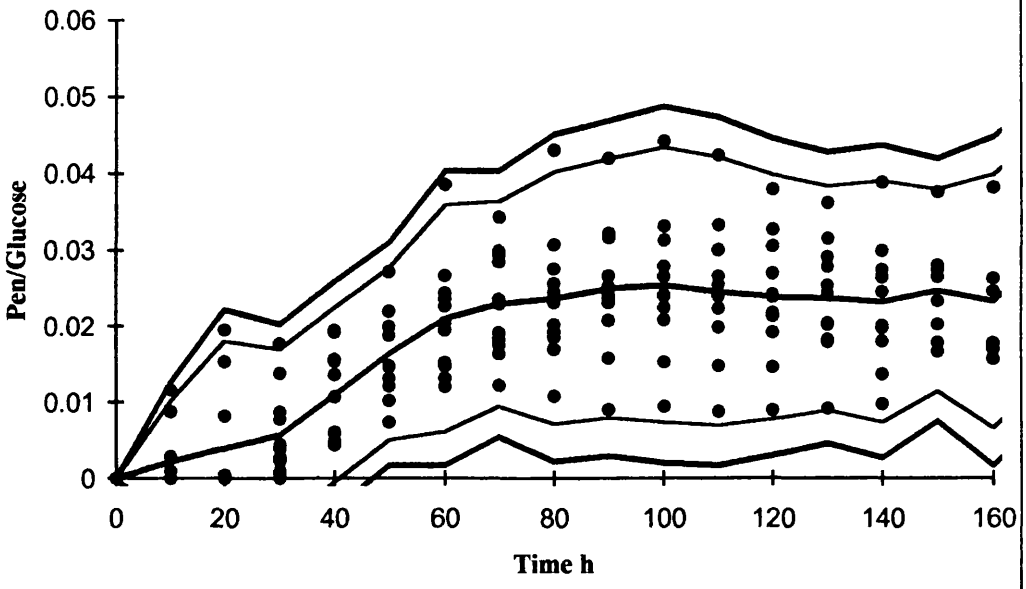


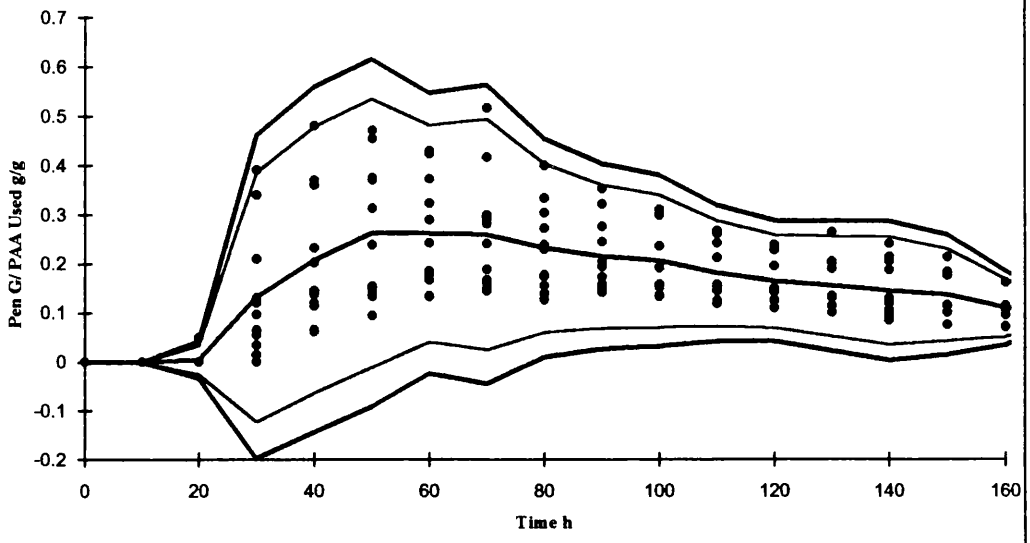
Fig 6.8 shows the benefit of normalising fermentation data, the 3σ confidence limits contain all penicillin data which is from fault free fermentations. This is in marked contrast to fig 6.6 which reports penicillin production rate. The implication of this is that when a datum falls outside the expected (3σ) range, a confident fault identification can be made.

Visual inspection shows that the ±3σ limit trajectories are much smoother than is the case for penicillin data.

6.3.3 Penicillin Produced / Phenylacetic Acid Utilised

On a mass basis, the relationship between the rate of phenylacetic acid utilisation to penicillin production was found to be as shown in fig 6.9.

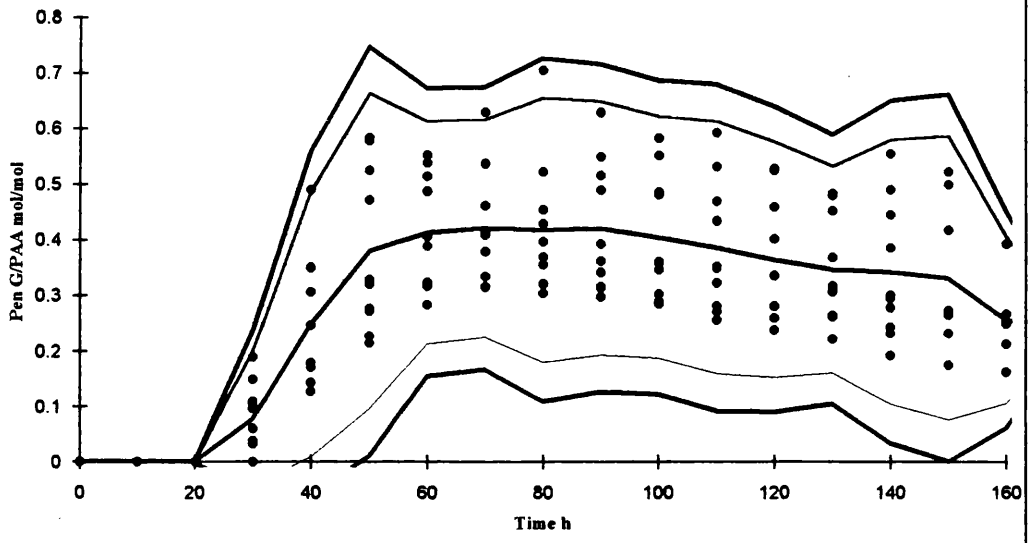
Fig 6.9: Ratio of Penicillin produced to Phenyl Acetic Acid used (i.e. yield of Penicillin on Phenyl Acetic acid) Shows mean, 2σ and 3σ confidence limits.



The ratio of penicillin produced to phenylacetic acid utilised indicates the efficiency with which phenylacetic acid is converted to penicillin G. Because phenylacetic acid was a major potential source of carbon, it was necessary to investigate its fate. The yield was calculated as the molar ratio of total penicillin G produced by the fermentation to the amount of phenylacetic acid that was used. The total amount of penicillin G produced was calculated as the total amount of penicillin G measured at the end of the fermentation corrected for the amount of penicillin lost in the samples and by an estimate of the amount of penicillin G hydrolysed to penicilloic acid (using the hydrolysis constant for penicillin G of 0.0014 h^{-1} for the pH and temperature conditions of this fermentation (Benedict *et al.*, 1945 and 1946)).

The total amount of phenylacetic acid used was calculated as the total amount of phenylacetic acid fed during the fermentation, corrected for the amount of phenylacetic acid remaining in the broth at the end of the fermentations and the amount lost in samples. On a molar basis, the following trajectories were obtained:

Fig 6.10: Molar Yield of Penicillin on phenylacetic acid. Shows mean, 2σ and 3σ confidence limits.



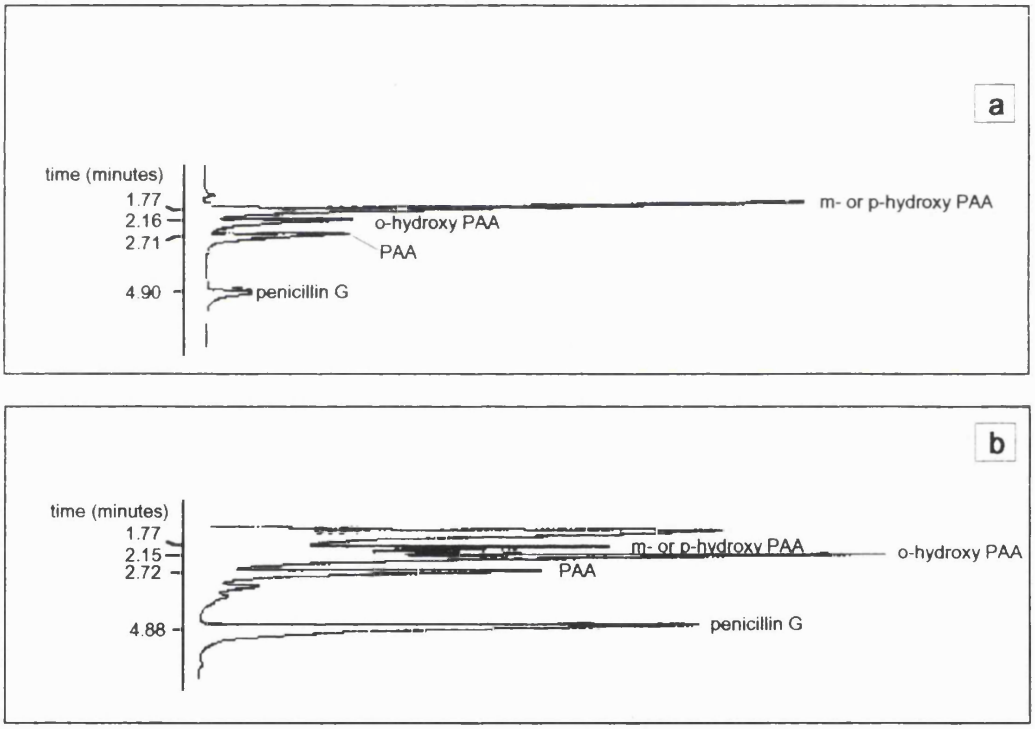
The graph shows the wide variation in yield of penicillin on phenyl acetic acid. This could in some cases be due to metabolism of phenylacetic acid as a carbon source for the fermentation, under some fermentation conditions. The mean conversion rate of approximately 0.4 seen in this work falls within the range of 0.3 - 0.5 reported by Mou and Cooney (1983 a).

The yields of penicillin G on phenylacetic acid are much lower than both the theoretical 1 mole of phenylacetic acid required to produce 1 mole of penicillin, (Cooney and Acevedo, 1977) and practical values: Perlman (1970) reported that conversion efficiencies of greater than 90% are possible; Mou and Cooney (1983 a) present data showing conversion rates of 0.3 mols/mol and 0.5 mols/mol for their fermentations, the fate of the phenylacetic acid which was converted but did not appear as penicillin G was not discussed. Calam (1987) has said that

phenylacetic acid can be used as a carbon source; and Swartz (1985) reported that phenylacetic acid can be hydroxylated by the producing organism.

The possibility that phenylacetic acid was being hydroxylated to any of its three forms: o-, m-, and/or p-hydroxy phenylacetic acid was investigated by HPLC analysis of broth samples with the same method used to measure phenylacetic acid and penicillin G . The output from the HPLC integrator for a mixed standard of 0.5 $g.L^{-1}$ o-, m-, p-hydroxy phenylacetic acid, phenylacetic acid, and penicillin G is presented below.

Fig 6.11 HPLC Output



The HPLC integrator output from: (**a**) 0.5 $g.L^{-1}$ standard of o-, m-, and p-hydroxy phenylacetic acid, phenylacetic acid and penicillin G; (**b**) a typical sample late fermentation broth sample (sample 23 fermentation FERM11) showing peaks for m- or p-hydroxy phenylacetic acid that are not resolved with this method, o-hydroxy phenylacetic acid, phenylacetic acid, and penicillin G.

It can be seen from fig 6.10 a that using techniques available to this project, it was not possible to discriminate between the peaks for m- and p-hydroxy phenylacetic

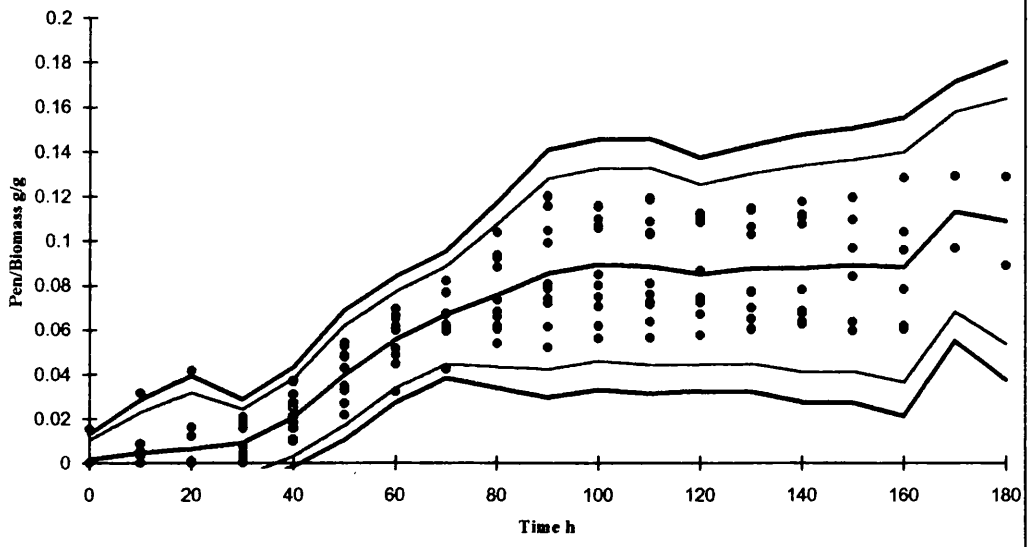
acid. Graph b shows the output of the HPLC integrator for sample 23, fermentation FERM11, which was typical of a sample taken late in a fermentation, when it can be expected that breakdown products have accumulated in the broth.

The presence of o-hydroxy phenylacetic acid is indicated by the peaks at 2.15 minutes and the presence of m- and/or p-hydroxy phenylacetic acid is indicated by the peak at 1.77 minutes. Other workers have detected hydroxy phenylacetic acid in the broth of penicillin G fermentations: Lenz *et al.* (1985) detected the presence of o-hydroxy phenylacetic acid and Adlard *et al.* (1990) detected the presence of both p- and o-hydroxy phenylacetic acid. The conclusion from these findings is that phenylacetic acid was being hydroxylated which explains the low yields of penicillin G on phenylacetic acid. It is not however possible to show what proportion of the phenylacetic acid went to degradation products.

6.3.4 Ratio Of Penicillin / Broth solids

As described in chapter 1, the growth of biomass is carefully controlled in penicillin fermentations to optimise the yield of penicillin. Therefore the ratio of broth solids produced to penicillin produced is a key performance indicator for penicillin fermentations.

Fig 6.12 **Yield of Penicillin on Broth Solids. Shows mean, 2σ and 3σ confidence limits.**



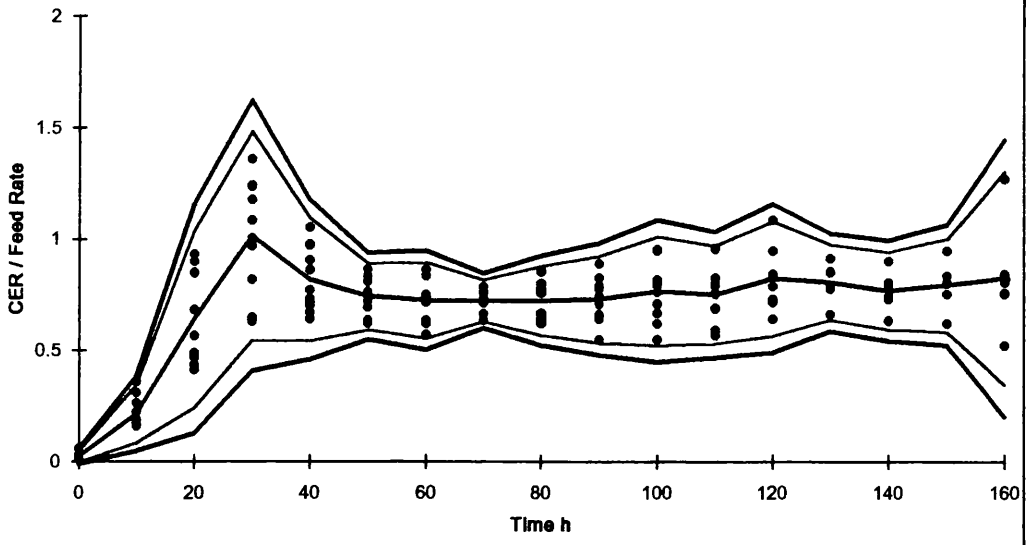
This graph shows the trajectory of the 2σ and 3σ limits of penicillin yield with time. It can be seen that the mean value of this trajectory remains close to zero for 30 hours, as during the exponential growth phase of the fermentation virtually no penicillin is produced.

This performance indicator is important because, in the event of a contaminated fermentation it is highly likely that the ratio of penicillin / broth solids will drop below the bottom end of the confidence limit.

6.3.5 CER / Glucose Addition Rate

A comparison can be made between CER and glucose added, this ratio is effectively a yield of CO_2 on glucose:

Fig 6.13 **Yield of Carbon Dioxide on Glucose. Shows mean, 2σ and 3σ confidence limits.**



The ratio of carbon dioxide evolved to glucose added is important in a fault detection scheme as it gives an approximation to the overall carbon balance. All of the data used in deriving this ratio is either gathered on line (airflow rate, [CO2], etc.) or is known from control settings on the system (glucose feed rate, sample removal etc.). This permits a check on the performance of a number of fermenter subsystems, including feed regime, and aeration regime. Also malfunctions in other subsystems such as the pH controller will result in changed metabolic activity thus changing the value of this ratio.

Also, because of its similarity to an overall mass balance, this ratio is relatively insensitive to fermentation conditions except during the exponential phase when carbon is freely available from sources other than glucose.

The confidence limits obtained from the ratio of carbon dioxide evolution to glucose addition are of great use in a fault detection scheme. It allows a comparison between two independently measurable process parameters, both of which can be effectively measured.

All of these normalised data represent a more precise measure of the correct functioning of a fermentation than is given by the basic on-line and off-line data. The fermentation envelopes represented by the 3σ confidence limits will be shown

in chapter 8 to represent boundaries which the trajectories of faulty fermentations transgress.

6.3.6 Yield of Broth solids on Carbon Dioxide ($Y_{X/C}$)

The parameter $Y_{X/C}$ is the true yield of broth solids on carbon dioxide, or the amount of carbon dioxide that is produced as a result of growth alone. During the growth phase of these fermentations, it is assumed that all carbon dioxide is produced as a result of growth (Zabriskie and Humphrey, 1978; Alford, 1978; and Mou and Cooney, 1983 a), i.e. that cell maintenance and other activities, such as secondary production, are insignificant. Therefore, the measurements of dry weight concentration and total carbon dioxide evolved, made during this phase, can be used to identify an approximate value of $Y_{X/C}$. For each fermentation, $Y_{X/C}$ was calculated as the gradient of the linear least squares fit to a plot of the total carbon dioxide evolved against total dry weight during the growth phase. The mean value of $Y_{X/C}$ calculated by this means was 1.52 gdw/L(CO_2)

When compared with other values of $Y_{X/C}$ reported for *Penicillium chrysogenum* P2, the mean value of 1.52 g(broth solids)/L$_{(CO_2)}$ is similar to that reported by Mou and Cooney (1983 a) of 1.46g(broth solids)/L$_{(CO_2)}$ (reported as 35.8 g(dwt).mol$_{(CO_2)}^{-1}$), but it was much less than that reported by Cagney *et al.* (1984) who reported a value of 5.1 g(broth solids).L $_{(CO_2)}^{-1}$ (reported as 0.122 g(dwt).mmol$_{(CO_2)}^{-1}$).

6.4 Overall Carbon Balance

Fig 6.13 summarises the data obtained over all the fermentations for which good carbon dioxide evolution data exists.

These fermentations were carbon limited, therefore it is possible to establish an overall mass balance on the basis of carbon. Carbon is fed to the fermentation in the form of glucose, phenylacetic acid, corn steep liquor and inoculum biomass. It is converted into biomass penicillin, penicillin degradation products, and carbon dioxide. The carbon content of the biomass and of the corn steep liquor was measured using gas chromatography with a mass spectrometer detector. The results are shown in table 6.2

Table 6.2 Carbon Content Of Broth solids And Corn Steep Liquor Samples.†

Broth solids Samples	%C	%H	%N
Sample 12 Ferm 12	42.9	6.3	7.5
Sample 27 Ferm 12	40.4	6.5	6.0
Sample 12 Ferm13	40.4	6.4	8.0
Sample 24 Ferm 13	42.7	6.7	5.2
Mean Broth solids Composition	41.7	6.5	6.7
Corn Steep Liquor Samples			
Sample 1	40.6	5.8	7.7
Sample 2	40.4	6.0	7.5
Sample 3	41.0	5.4	7.8
Mean CSL Composition	40.7	5.7	7.7

† Measured by GC MS, by UCL Dept. of Chemistry.

The mean broth solids carbon content (41.7%) was close to the 42.9% figure reported by Mou and Cooney (1983a).

Fig 6.14 Carbon Balance. Shows mean, 2σ and 3σ confidence limits.

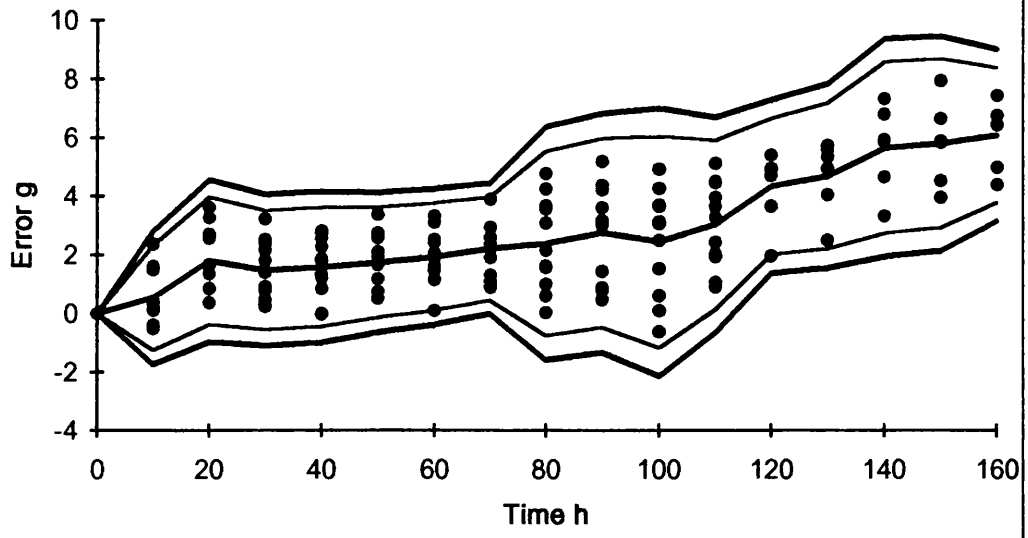
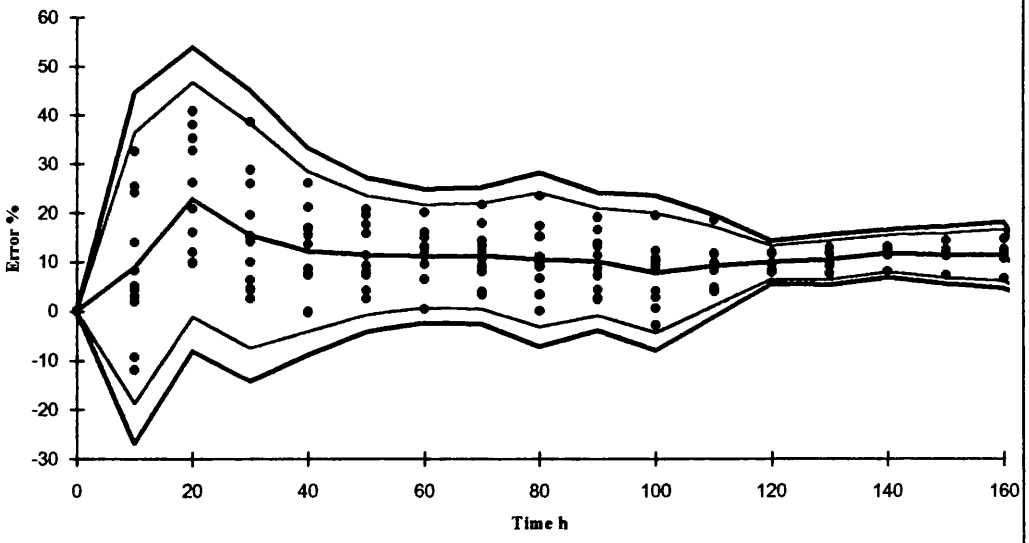


Fig 6.15 Ratio of Error in Carbon Balance To Carbon Additions. Shows mean, 2σ and 3σ confidence limits.



Figs 6.14 and 6.15 show that as a fermentation progresses, greater confidence can be placed in the overall carbon balance, on a % error basis but that the absolute expected error range remains close to constant at approximately ± 3g/l. Also, the data shows that there is a bias in the mass balance estimates, indicating that some broth carbon is unaccounted for. This is probably in the form of material lost in preparation for dry weight analysis of broth solids, and metabolite products unmeasured in the broth.

It can be seen from this data that a consistent 0-20% of the carbon in the overall mass balance is not being accounted for. It is known that some carbon is not accounted for in the form of dissolved carbon dioxide, unmetabolisable phenylacetic acid degradation products and penicillin breakdown products (see fig 6.11). Probably the process of establishing cell dry weights also results in the loss of some material by leaching of cell contents when the cells are being washed, and by evaporation during the drying process.

In addition there are biases in some of the measuring systems used to establish flowrates, and concentrations of broth components (See chapter 4). None of this is important to a fault identification process, providing batch to batch errors are consistent. All that is of interest is the behaviour of the system when no equipment failures or contaminations have occurred.

The key data derived from this section of the work are the 2σ and 3σ confidence limits. These will be defined as bounds outside which a fermentation can be considered to no longer be functioning correctly. It can be seen from the above data that, in order to have a robust fault detection system, the use of 3σ confidence limits will be preferable. This is because, due to the distribution of the data, which often contains time dependent features, would give a high frequency of false alarms if a fault detection system was to adopt a 2σ standard. An analysis of which of these measurements is of use in fault detection will be presented in chapter 8.

6.5 Conclusions

From the above data it can be seen that, even over the limited number of broadly similar fermentations presented here, there is considerable batch to batch variation in data. The utility of direct measurements in fault detection is therefore limited. This is because of the very high divergence from expected trajectories which a parameter would have to go through before a fault was flagged.

Scaled parameters are a more promising basis for fault detection. Therefore, the recommendation is to use the carbon dioxide production / glucose added and specific growth rate / glucose added data presented above.

Chapter 7: Mathematical Modelling Applied To On Line Mass Balancing

It is possible to incorporate data from prior fermentations into a mathematical model. This can then be used to estimate state variables in current fermentations. A review of the models available for penicillin fermentation modelling was presented in the introduction. The use of the P2 *Penicillium chrysogenum* strain and the media and methods described in chapter 3 was adopted for this work because of the availability of published model parameters used on specifically that system. The model used was that derived by Bajpai and Reuss (1980, 1981). This was described in chapter 3.

In this chapter, the Bajpai and Reuss model is used in two ways, the first way is to use the output from the model directly to replace biomass measurements in fermentation mass balances. The advantage of using model outputs is that they are available immediately, whereas biomass measurements are time delayed with respect to the fermentation. The other outputs of the Bajpai and Reuss model, penicillin yield, and substrate concentration were not used in mass balancing. The penicillin titres in the fermentations carried out as part of this study were highly variable, and not therefore susceptible to modelling. The level of carbon-containing substrate in the fermentation beyond the exponential growth phase was negligible from a mass balance perspective.

The second way of using the Bajpai and Reuss model was in an on line estimator, based on the Extended Kalman Filter (Jaswinski 1970, Lichfield 1979). The use of the Bajpai and Reuss model as the basis for a Kalman filter was described by Montague *et al* (1985). The use of the filter was intended to determine whether better estimation of biomass could be obtained by feeding back a correction term based on the difference between an estimate of the carbon dioxide evolution rate and the observed carbon dioxide evolution rate.

The usefulness of both techniques will be assessed by comparing the errors in on line mass balances obtained using them against the errors obtained using measured biomass in fault free fermentations. Their usefulness in fault detection will be discussed in chapter 8.

7.1 Comparison Between Mass Balances Obtained Using Measurements And Mass Balances Obtained Using The Bajpai And Reuss Model

The Bajpai and Reuss model was described fully in section 3.2. Its equations can be stated as follows (Equations 7.1)

$$\frac{dX}{dt} = \frac{\mu_{X\max}.S.X}{K_X.X+S} - \frac{X}{V}.\frac{dV}{dt}$$

$$\frac{dP}{dt} = \frac{\mu_{P\max}.S.X}{K_P+S.(1+S/K_I)} - K.P - \frac{P}{V}.\frac{dV}{dt}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{X/S}}.\frac{dX}{dt} - \frac{1}{Y_{P/S}}.\frac{dP}{dt} - m_X.X + F - \frac{S}{V}.\frac{dV}{dt}$$

$$\frac{dV}{dt} = \frac{F.V}{S_F}$$

The following parameters obtained from Mou and Cooney (1983) were used in the model

Table 7.1 Model Parameters.

Parameter	Nomenclature	Value	Units
Biomass	x	-	gdw
Penicillin concentration	P	-	g/l
Culture volume	V	-	l
Substrate feed rate	F	-	g/l/hr
Substrate concentration	S	-	g/l
Substrate concentration in feed	S_f	-	g/l
Maximum specific growth rate	μ_x	0.11	h

Parameter	Nomenclature	Value	Units
Maximum specific penicillin prod'n rate	μ_p	0.004	g/g/hr
Maintenance coefficient	m_x	0.02	g/g/hr
Penicillin decay rate coefficient	K	0.01	1/h
Substrate inhibition constant for penicillin synthesis	K_i	0.1	g/l
Parameter	Nomenclature	Value	Units
Contois saturation constant for biomass	K_x	0.15	g/g
Saturation constant for penicillin	K_p	0.0001	g/l
Yield of biomass on substrate	$Y_{x/s}$	0.47	g/g
Yield of penicillin on substrate	$Y_{p/s}$	1.2	g/g
Time	t	-	h

7.2 Usefulness Of Biomass Predictions From The Bajpai And Reuss Model

Two illustrations of the fit between modelled and measured biomass are shown below:

Fig 7.1 Comparison of Model and Measured Broth Solids data. Fermentation carried out at UCL, on a 5 l scale)

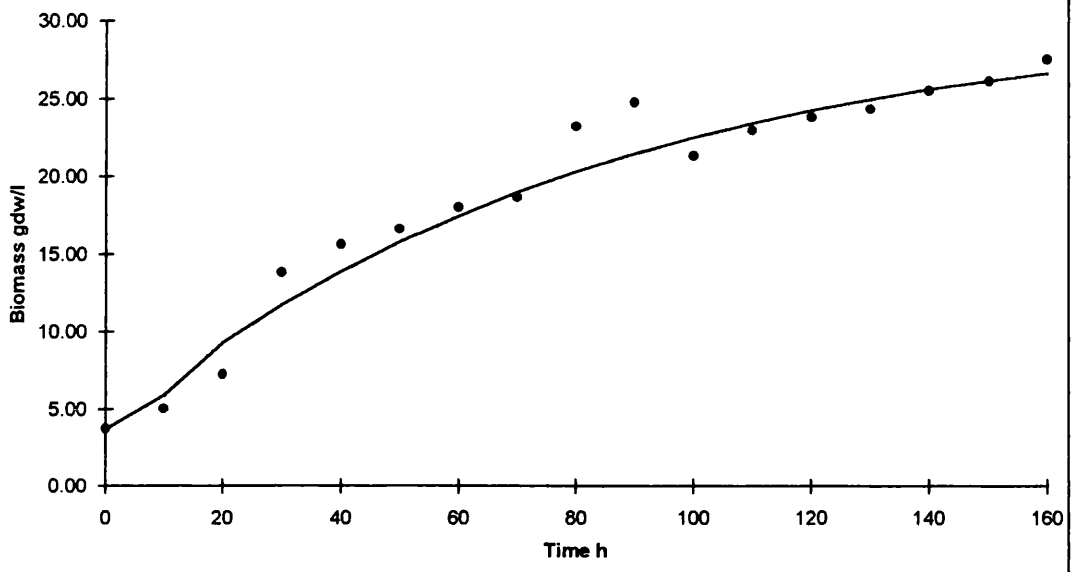


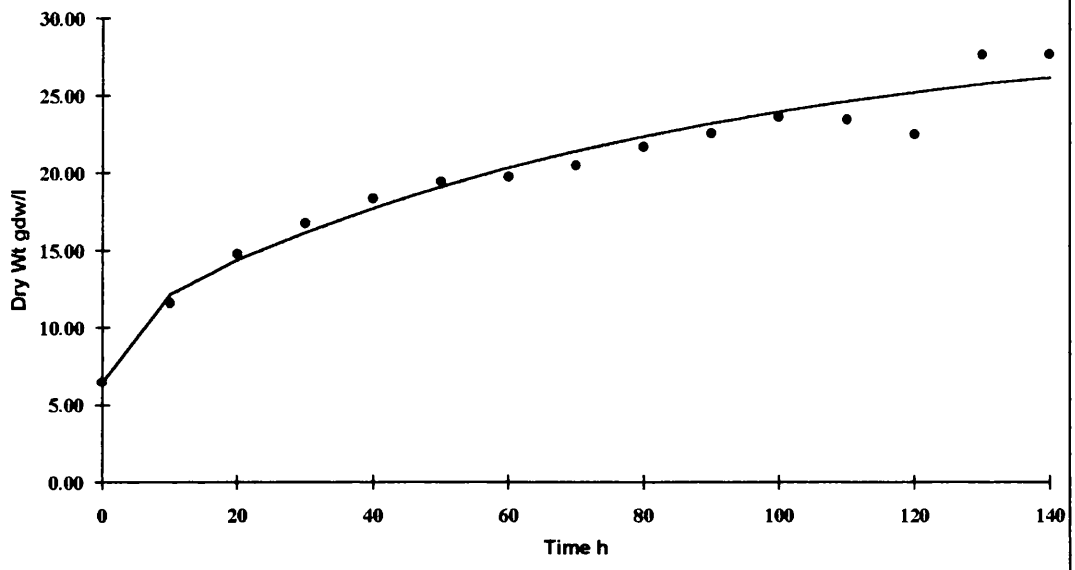
Figure showing the agreement between the Bajpai and Reuss model and experimental fermentation data, from fermentation 11. From this data, it can be concluded that samples at 80 and 90 hours are outliers.

The model was given as its starting point a value of biomass which included a proportion of undissolved solids from the starting broth. Therefore the model is providing an estimate of total broth solids rather than biomass.

Figure 7.1 shows an advantage of using modelling in that, being independent of the specific batch to which automatic supervision is being applied, it can be used to differentiate between outliers and accurate fermentation data.

The second example of the efficacy of biomass modelling using the Bajpai and Reuss model is taken from Fermentation 6b, which is one of the fermentations carried out at SmithKline Beecham.

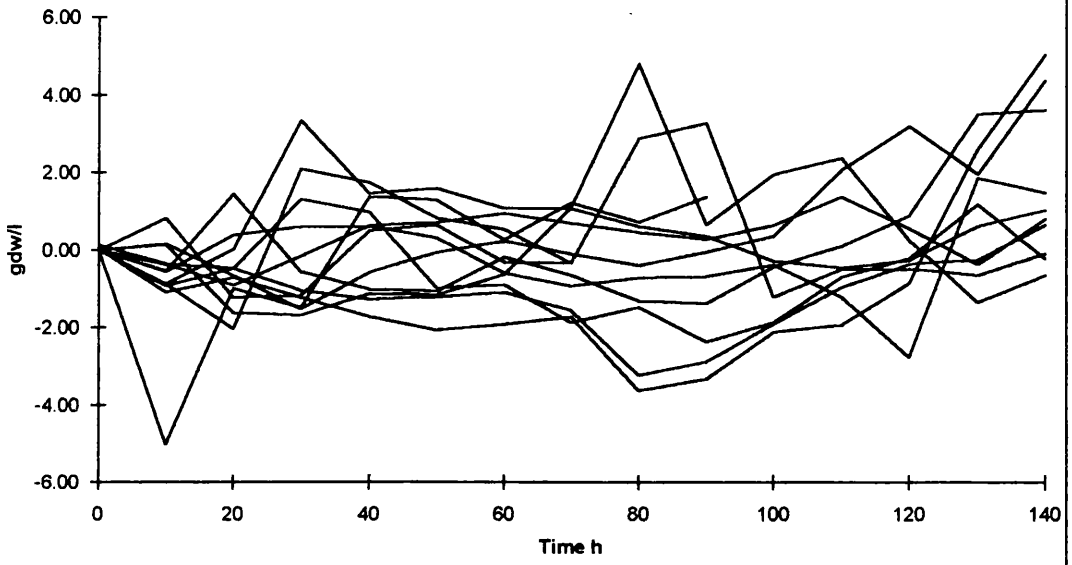
Fig 7.2 Comparison between model and Experimental data (Fermentation carried out at SB on 1.5 l scale)



Figures 7.1 and 7.2 show that the Bajpai and Reuss model, with Mou and Cooneys parameters predicts broth solids concentrations effectively.

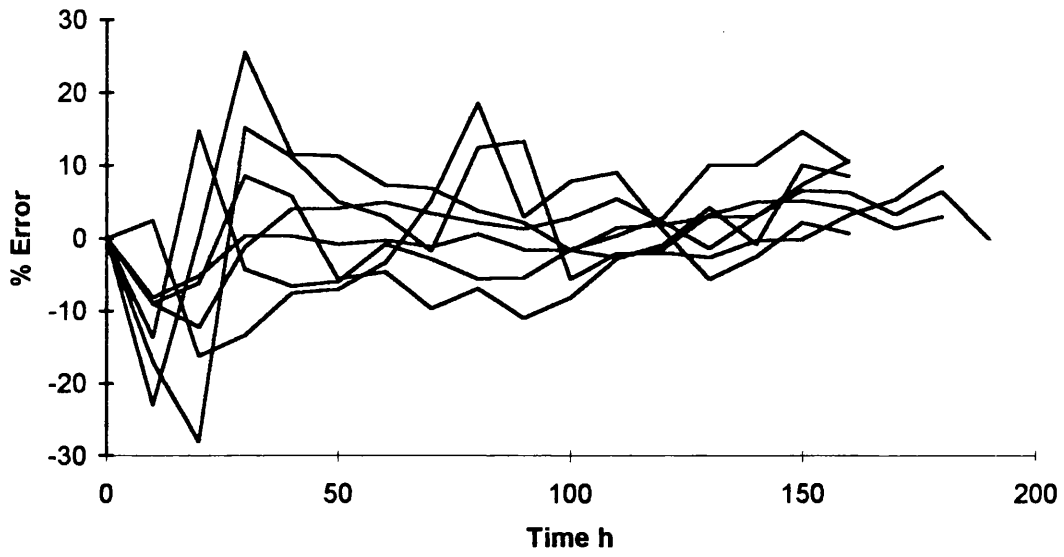
The usefulness of the model for determining an expected broth solids trajectory for fault free fermentations is shown by graphs 7.3 a - 7.3c:

Fig 7.3a **Difference between modelled and observed solids data, all fermentations (Absolute difference g/l).**



This figure illustrates the absolute accuracy if the Bajpai and Reuss model in predicting broth solids (i.e. largely biomass) observations across all fermentations. It can be seen that there is no evidence of significant bias in these data.

Fig 7.3b **Difference between modelled and observed biomass data, all fermentations (relative difference %).**

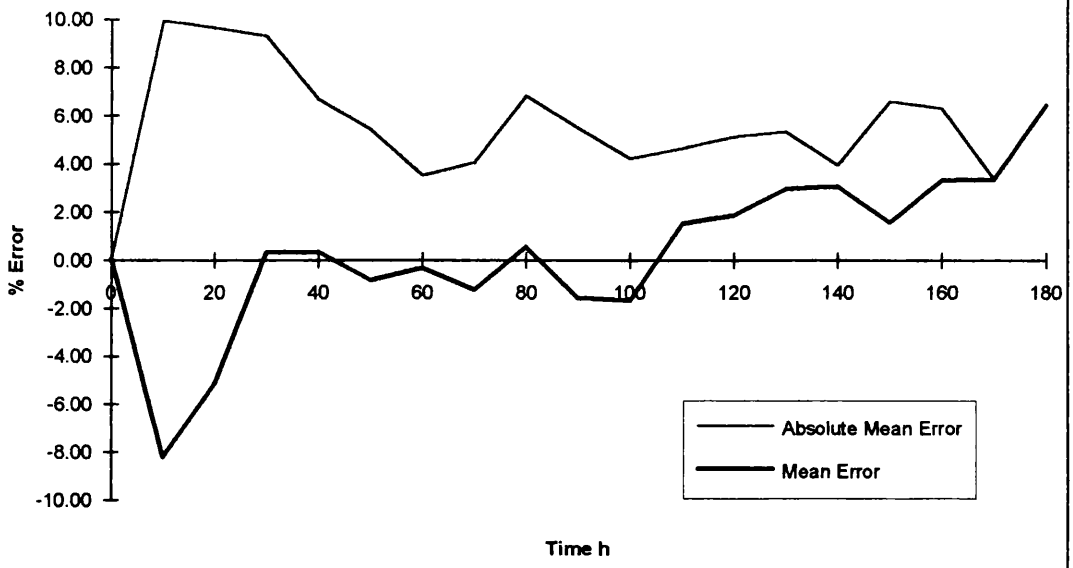


This figure shows the relative error on Bajpai and Reuss model predictions. It can be seen that the model in practice predicts the exponential phase of the fermentation less well than the steady state phase (typically starting at 30 hours). This is to be expected because the it is much easier to model steady state carbon limited growth, where activity is closely tied to the quantity of glucose being fed. In the exponential phase, where there is no glucose limitation, other factors such as the time taken for the onset of exponential growth become important.

This may vary depending on the status of the seed fermentation, for example whether the seed fermentation was harvested before or after the end of its own exponential growth phase, or how long the seed fermentation was held without aeration between removal from its orbital shaker and addition to the main fermentation. Another factor is the degree of caramelisation of sugars in the fermenter during the sterilisation process, varying the quantity of readily available energy.

The mean error shown above is of the same order (approximately 10%) as the mean error seen in the overall mass balance (section 6.4), it is therefore possible to conclude that the model can estimate biomass in the absence of a fault at least as well as the techniques available for measuring biomass.

Fig 7.3 c **Average model error as percentage of total biomass, reported both as $\frac{1}{n}\sum_{i=1}^{n}|error|$ and $\frac{1}{n}\sum_{i=1}^{n}error$**



This graph summarises the performance of the Bajpai and Reuss model in predicting the concentration of biomass in the fermentations carried out as part of this project. It can be concluded that the model was unbiased and on average correct to within 6%.

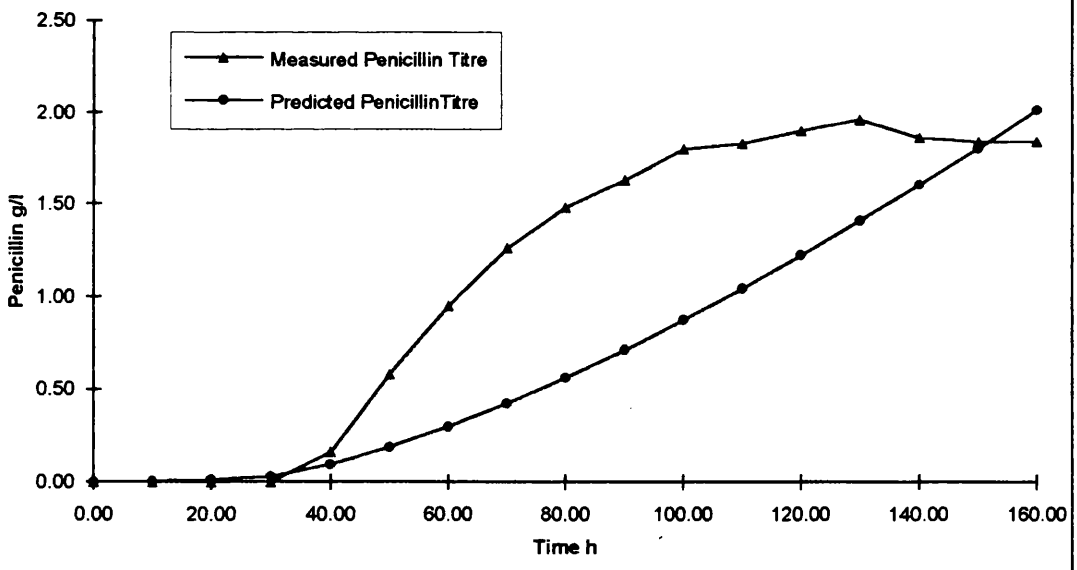
This data shows that there is an average error of less than 10% throughout the course of all the fault free fermentations between data obtained directly from the fermentations and model outputs. The absolute difference between the two data sets rarely exceeded 4gdw/l, and on the occasions that it did, the subsequent samples showed a difference of less than 4g/l. As will be seen in the next chapter this was not the case for fermentations in which faults were induced or in which they were known to have occurred. It can be concluded therefore that the Bajpai and Reuss model is a useful tool in detecting faults in ongoing fermentations.

7.3 Usefulness Of Penicillin Predictions From the Bajpai And Reuss Model.

The model makes predictions about parameters other than biomass. The second most potentially useful prediction from the point of view of completing a mass balance in real time as opposed to using off-line sample analysis is that of penicillin production itself. Using the model parameters described above and the fermentation data presented in chapter 4 model penicillin trajectories were prepared for all the fermentations not subject to known faults. The outcome was that the penicillin predictions were inaccurate to around ±100% and that the nature of the inaccuracy differed between the 5 litre fermentations and the 15 litre fermentations.

Taking as an example fermentation 9, the divergence between the model output and the measured penicillin concentration can be seen in the following graph:

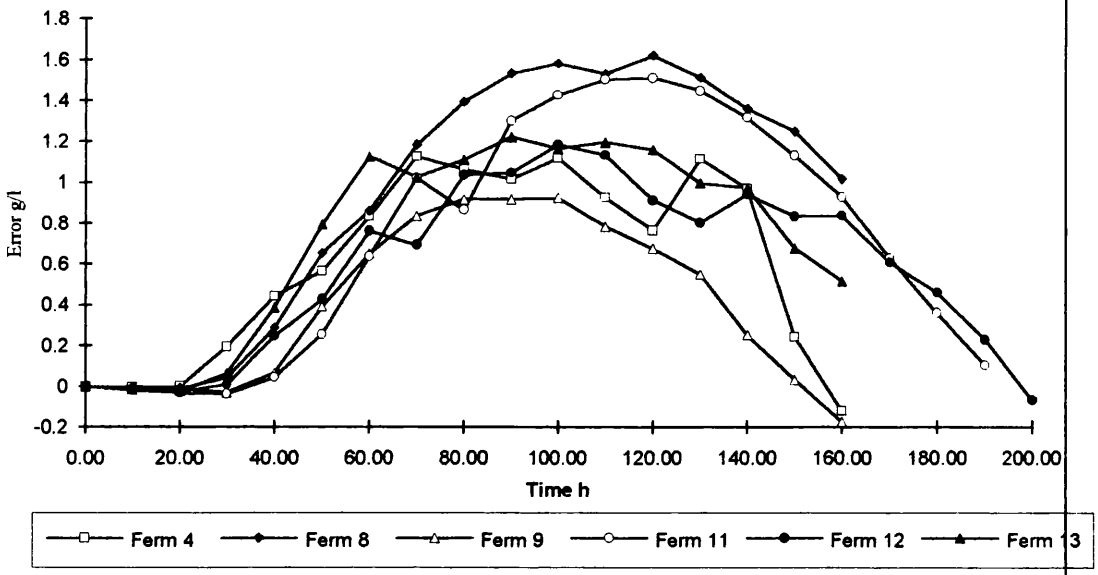
Fig 7.4 Comparison between actual and modelled penicillin titre.



In contrast to the prediction of biomass, the Bajpai and Reuss model turned out to predict Penicillin titre poorly.

If the errors (difference between actual and predicted penicillin titre) for all the 5 litre fermentations are plotted the following graph is obtained.

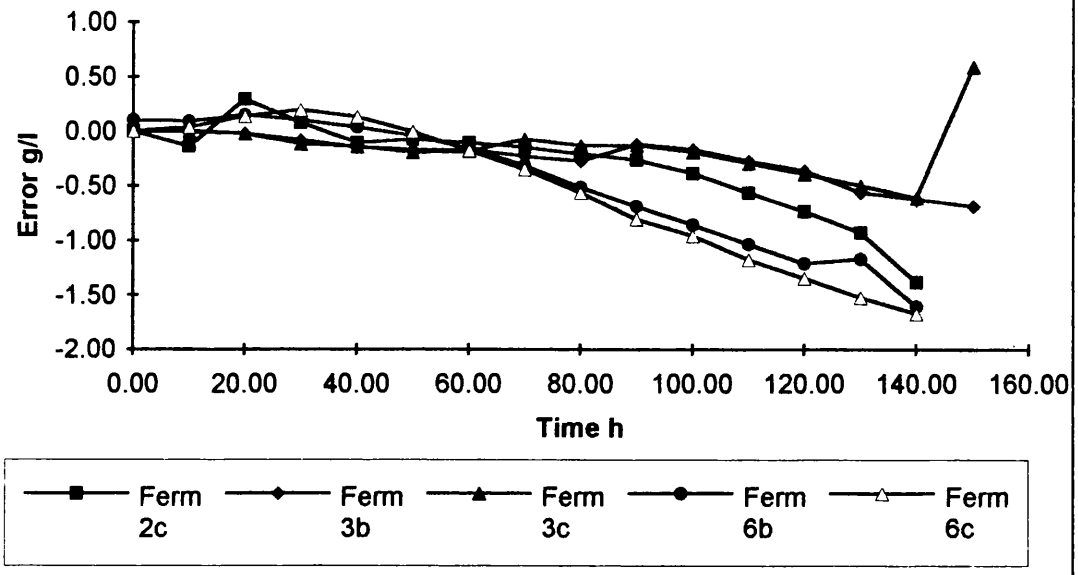
Fig 7.5 Fermentations carried out at UCL : Penicillin Model Errors.



This figure shows that the Bajpai and Reuss model is a poor predictor of penicillin titre. For the purposes of this project, this encouraged a focus on biomass as the leading fault indicator.

This would not necessarily invalidate penicillin predictions as a potential fault detection tool but data from the 1.5 l fermentations shows that the error pattern is inconsistent. This is shown in the following graph:

Fig 7.6 Prediction of Penicillin titre : Fermentations carried out at SmithKline Beecham



It is clear that the fermentation conditions used at SmithKline Beecham produced lower penicillin titres than predicted by the model, while the fermentations at UCL produced higher than predicted titres. This difference is probably due to the generally lower glucose feed rates used at SB.

This differing error pattern from fermentations carried out on different equipment and with different feed profiles is not surprising. Since this work has in mind fermentation systems being used to carry out strain and process improvement studies it is unrealistic to expect a model to predict accurately penicillin titres across the range of conditions applied. This is because the work is aimed at identifying conditions which optimise the rate of penicillin production.

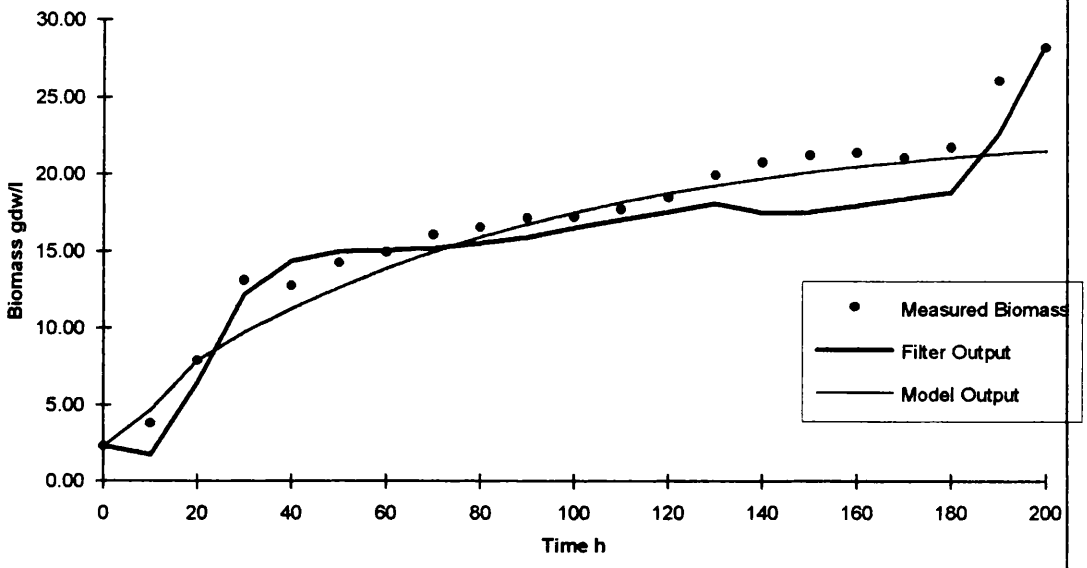
7.4 Application Of An On Line Observer.

Considerable literature exists (Montague et al 1985, 1986, Stephanopolous et al 1986, Tarbuck 1985, 1986) to support the idea of using an on line estimator to track fermentation parameters, in particular penicillin concentration. The mathematics of

the observer used in this work, the Extended Kalman Filter, have been discussed in chapter 3. This section shows how the filter can be used to track biomass in penicillin fermentations and how the output from the filter differs from the model on which it was based.

The difference between the output from the filter and the model can be illustrated by fermentation 12.

Fig 7.7 Fermentation 12, biomass as measured using dry weights, as modelled and as estimated using an extended Kalman filter.



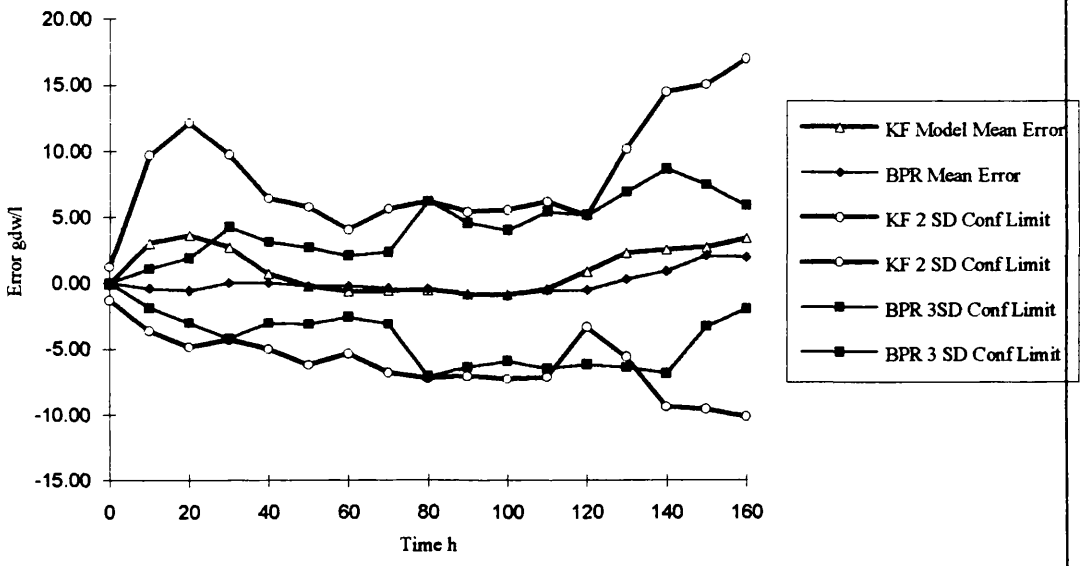
This fermentation illustrates the differing output of the filter and the model in the event of an unplanned event or fault. In this case, the rate of glucose feed was increased after 178 hours, to the extent that free glucose was available in the fermentation broth at measurable levels. The result of this was that the fermentation went back into unconstrained growth. The filter was capable of tracking this event due to the inputs from additional evolved carbon dioxide.

There are a number of interesting features on this graph. The first is that the filter output would seem to indicate that biomass is falling in the first 10 hours of the fermentation. This is typical of the Kalman Filter outputs from these fermentations. It reflects the fact that the dry weight method of measuring biomass is biased by the presence of insoluble solids in the fermentation medium, and that the culture exhibits a lag phase which is not accounted for by the model. The filter can clearly be seen to more closely match the shape of the fermentation trajectory than the model, in

particular the filter output matches this particular fermentation data after 178 hours when an additional aliquot of glucose was added to the system to simulate a glucose addition fault. The implications of this difference in response to fermentation events between the filter and the model will be discussed in the next chapter.

Generalising, the following patterns of differences can be seen between filter and model output, and observed biomass:

Fig 7.8 Comparison of the difference (or error) between model output, filter output and measured biomass

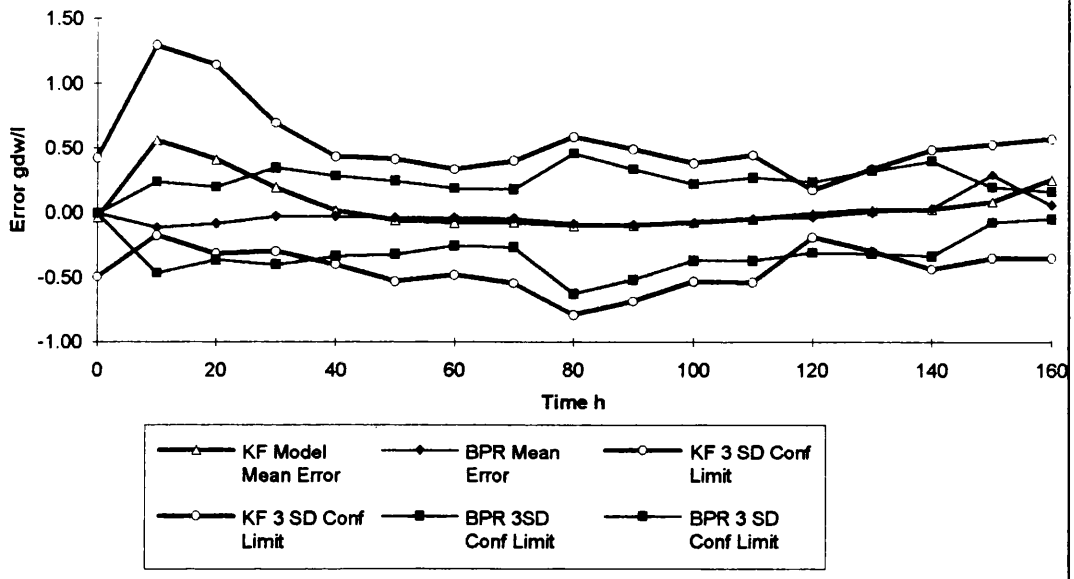


This graph shows the relative performance of the Bajpai Reuss model, and a Kalman filter based on that model. The graph shows the mean difference between both model estimates and dry weight measurements, and filter estimates vs. corresponding dry weight measurements. It can be seen that after the first 30 hours the performance of the model and the filter are similar. The difference before 30 hours is probably due to the presence of undissolved solids in the starting broth, which form part of the biomass estimate used to initiate both the filter and the model.

The filter, by utilising on line carbon dioxide evolution data can converge on an accurate biomass estimate, the model has no means of eliminating this error. Over time, as more carbon is added to the fermentation the importance of the remaining dissolved solids becomes lower and the filter and model estimates converge.

Normalising for biomass the errors shown in fig 7.9 are obtained:

Fig 7.9 Comparison of model and filter errors as a fraction of total biomass



This graph shows the relative performance of a model based on starting conditions and feed pattern data vs. an Extended Kalman Filter using on line carbon dioxide evolution rate data in addition to the information supplied to the model. The filter (Solid symbols) provided a less precise estimate of the biomass in all the fermentations than was provided by the model. Accuracy was identical outside the exponential phase of the fermentation.

It can be seen from these graphs that the filter and the model produce similar error patterns, with on average the model providing a somewhat better prediction of biomass than the filter. This reflects the greater complexity of the filter and the problem that the filter outputs can be irrecoverably destabilised by an unexpected fermentation event, in particular a short sharp transient..

In addition, the accuracy of the filter is a function of the accuracy of the information with which it is fed. Data such as carbon dioxide evolution rate are obtained by the combination of a number of pieces of information such as the concentration of CO_2 in the airstream exiting the fermenter, the fermentation volume and the air flow through

the fermenter, each of which is subject to bias. Also the tuning of the filter was conducted on a pragmatic basis, rather than by rigorous determination of all the necessary covariance data as this was judged to be impractical. A useful next step in this work would be to look at filter optimisation.

For the purposes of this project the usefulness of the filter is not that it can give precise biomass estimates but that it can provide a means of using on line CO_2 measurements to provide information on biomass which can be compared with off line biomass measurements and model output. This provides analytical redundancy which can be used to identify faults. For this purpose it is necessary only that errors are consistent.

The accuracy of the filter used here was enhanced by limiting the minimum value of the correction term along the lines suggested by Mehra (1970). Another approach would be to generate filter output corrections by the use of off-line data (Stone K.M. 1992); this option was considered too complex for use in this project, as the existing filter structure was already fit for the purposes of fault detection.

7.5 Conclusions

This chapter has shown that modelling can successfully describe the course of a fermentation. The weakness of modelling in a fault detection scheme is to select a model general enough to cope with varying organism types, and process conditions. A potential solution to this is to use the Kalman filter described in chapter 2. This provides a means to feed back the difference between model outputs the performance of the real system, generating more representative estimates. As can be seen from fig 7.7, the filter, using carbon dioxide data, is able to capture fermentation data features following disturbances, which is not possible for a model.

In fault detection, the Kalman Filter and the model which drive it contain subtly different information about past fermentations. Model parameters capture trends from prior data and *a priori* knowledge about the fermentation process. The Kalman filter takes this information, and incorporates additional relationships between fermentation states and outputs. Taking account of process and measurement noise data also obtained from past batches, it permits updates of fermentation states in the model in response to events specific to the fermentation under observation.

The model / filter approach can potentially therefore bring additional information to the consistency checks to be used in fault detection. The issue discussed in chapter 8 is whether this information stream is of more use than other simpler data streams available to the fault detection system, such as carbon dioxide evolution rate.

Chapter 8 Fault Detection By Analytical Redundancy

The information which has been shown in the previous chapters to be available from a fermentation can be combined to form the basis of a fault detection scheme. This chapter will analyse examples of major faults which can occur in a fermentation and, where data exists, show how these can be identified by the comparison of multiple data sources.

Information about a fermentation typically comes from sources ranging from qualitative operator observations, to on line instrumentation and the results from analysis of samples. Finally it is possible to make comparisons between the current and past fermentations. This chapter is organised to look at potential faults by type, the classes being:

- Contamination
- Hardware Failures
- Fermentation Process Failures
- External Faults

Experimental data for all these classes are available and are compared with confidence limits for normal batches.

8.1 Contamination

This section explains the impact of contaminations on penicillin fermentations. The information presented here was obtained from experimental work, the literature, and most importantly by talking to fermentation experts from SmithKline Beecham and UCL.

The impact of a contaminant is dependent on the time when it appears in a fermentation, and the nature of the contaminating organism.

- The effect of a fast growing contaminant present from the start of the fermentation.

Biomass: The faster growing organism will outgrow the *Penicillium* fungi and become the dominant form in the fermentation. Upon depletion of the initial glucose supply, biomass production would occur at approximately the same rate as for *Penicillium*, assuming carbon remained the limiting nutrient, as it is unlikely that the yield of biomass on glucose would be radically different between the *Penicillium* and the contaminant. Some extra biomass would probably be produced reflecting the lack of penicillin formation, unless the organism produced a secondary metabolite of its own in significant quantities. The level of biomass typically found in the steady state phase of a penicillin fermentation would be reached earlier in the event of a faster growing contaminant.

Under some conditions, a fall in broth solids content was observed (see fig 5.4). This could have been due to *penicillium* cell lysis when there is insufficient glucose available to support maintenance activities in the *penicillium* population.

Phenylacetic Acid Concentration: At some point, if the organism is incapable of metabolising phenylacetic acid, this will reach toxic levels and kill the contaminant. This form of contamination is readily detectable, after a period of time, by observation of a sample of fermenter broth under a microscope. Thus either very high or, if the contaminant is capable of metabolising it, very low levels of phenylacetic acid can be expected to be present in the broth.

Glucose And Penicillin Concentration: It could be expected that the initial glucose reservoir in the broth would be depleted more quickly than usual (otherwise the contaminant would be outgrown by the *penicillium*), and that less or no penicillin would be produced under these circumstances.

Carbon Dioxide Evolution: Initially a higher than usual carbon dioxide evolution rate, and oxygen uptake can be expected, reflecting higher biomass production rates. Due to faster glucose depletion the steady state phase of the fermentation can be expected to start earlier than usual. Somewhat higher carbon dioxide evolution can be expected during the steady state phase of the fermentation due to the likely presence of some additional biomass, and the lack of a diversion of cell resources into penicillin production.

Broth viscosity: Most contaminants being non mycelial, the change in cell morphology away from a mycelial culture will decrease broth viscosity. Occasionally, the contamination is by a mould which results in very high viscosity broth.

It can be concluded therefore, that in the event of a penicillin fermentation being contaminated early on, all the parameters routinely monitored in such a fermentation would move from their expected trajectories.

- Fast Growing Contamination introduced after the start of the fermentation.

If a non-penicillin-sensitive organism contaminates the fermentation after the initial glucose has been used up, it cannot take over the fermentation as the level of available glucose will not allow rapid growth. However an organism surviving in a penicillin containing environment is likely to produce a *β-lactamase* thus degrading existing penicillin. The competition for resources between the contaminant and the existing culture would in any event reduce the penicillin production rate leading to a fall in penicillin concentration due to hydrolysis and dilution, which was not outweighed by additional production as normal.

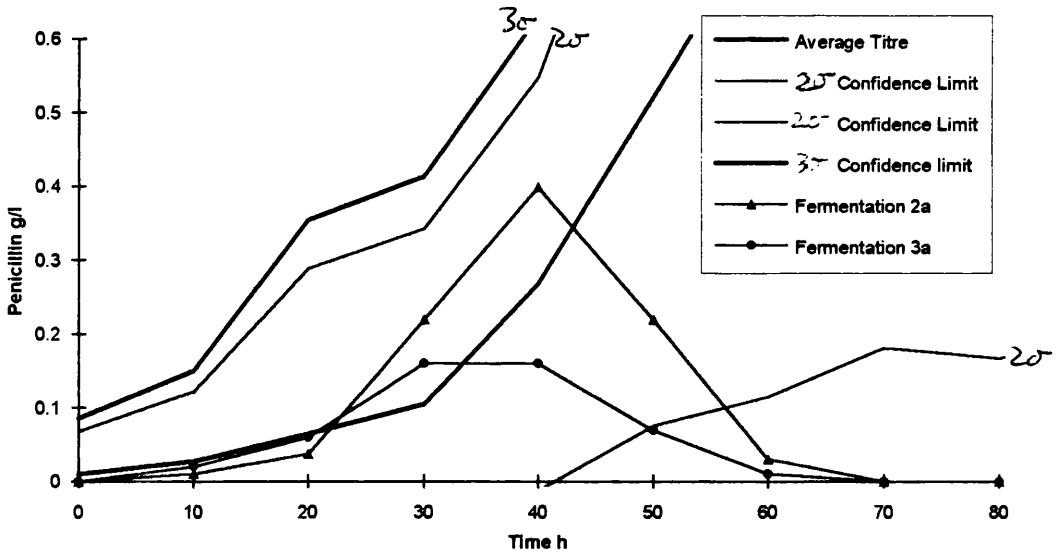
A fast growing contaminant, by utilising available glucose, will reduce the rate of penicillin formation, thus leading to an increase in the level of phenylacetic acid.

In the case of contamination later in the process, and therefore at a lower level than would be expected from a contamination which occurred from the start, fewer of the observed parameters would be impacted. In the context of this work, detecting such a contaminant is more critical than detecting a gross contamination. This is because such a contamination could give false negative experimental results for experiments aimed at optimising penicillin production.

8.1.1 Case Studies of A Contaminated Fermentation:

Two fermentations in this work were contaminated by micro-organisms which produced *β Lactamase*. The impact of these contaminants on penicillin titre was as follows:

Fig 8.1 Pattern of penicillin production in contaminated fermentations



This graph shows the impact of contaminant organisms on the level of penicillin found in two contaminated fermentations (2a,2b) carried out at Smithkline Beecham.

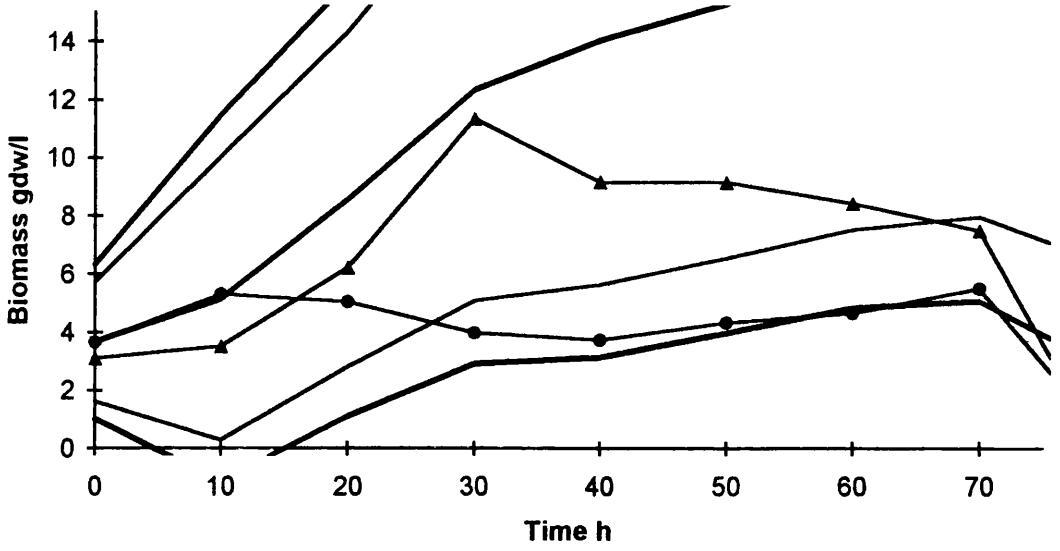
The graph indicates that, given sufficiently high data collection rates negative rates of production of penicillin could be used for fault detection.

It can be clearly seen that as the faster growing contaminants outgrew the *Penicillium*, the concentration of penicillin was decreased. The titre dropped below the lower 2σ confidence limit after between 50 and 60 h, although before this at up to 40 hours, the penicillin titre of both fermentations had been typical of normal production.

The scattered nature of penicillin titres in the early parts of a fermentation means that there are fairly wide confidence intervals for penicillin titre, however it can be said that at a time greater than 50 hours into the fermentation, a zero penicillin titre is indicative of a fermentation failure.

In the event of contamination other fermentation parameters remain within the range of typical experimental variation. As can be seen from the following chart the observed dry weight declined at about the same point in the fermentations as the penicillin titre declined.

Fig 8.2 Broth Solids in contaminated fermentations.



Graph showing the biomass trajectories corresponding to the penicillin trajectories presented in fig 8.1. This shows that while one of the fermentations went outside the 2σ confidence limit after 30 h, the other did not until 70 h. The presence of a fault could well have been missed simply by looking at biomass concentrations. At best the fermenter would have been running unproductively for two days .

It is probable that the decline in observed biomass is due to lysis of *Penicillium* mycelia due to the shortage of nutrients caused by the presence of competing organisms.

This contaminant was also observable under a microscope, given careful observation. There was no impact on the viscosity, appearance or odour of the fermentation. The reason for this is that the *Penicillium* fungi were still the dominant life form in the broth. This is because the contamination did not take hold until after the initial carbon reservoir in the broth was used up. Therefore, the growth of both the contaminant and the *Penicillium* would have been limited by available glucose.

The impact of this contamination is summarised in the following table:

Table 8.1 Impact of Contamination on Fed Batch Penicillin fermentations:

Parameter	Contamination From Start	Later Contamination
Biomass	Increases more rapidly than usual. Reaches somewhat higher equilibrium level as the batch moves to steady state.	Little change if the contamination occurs late in the batch. There can be a decline in observed biomass as the level of nutrients in the broth is no longer sufficient to support the existing *Penicillium* level.
Penicillin	Little or no penicillin produced	Penicillin level will decline either slowly through dilution and hydrolysis, or rapidly if the organism produces β-lactamase.
Glucose	Declines more rapidly than usual.	No measurable impact, as there is a very low level of glucose present even in uncontaminated fermentations.
Phenylacetic Acid	Will not be metabolised, so will attain higher than usual levels in the fermentation.	Will be used up at a lower level than is typical, so will build up slowly in the fermenter.
pH	No change in pH, as it is controlled on a feedback loop. The rate of addition of alkali will be atypically high.	No significant impact on pH
Temperature	Higher growth rates will put some additional demands on coolant supply.	No impact, as metabolic activity will be fixed by glucose addition rate.

8.2 Hardware Failures

Hardware fails on fermenters from time to time, the items most subject to failure being sensors measuring pH and dissolved oxygen as these rely on sensitive membrane barriers to maintain the electrochemical gradients used to effect their measurements. Both pH and dissolved oxygen probes are subject to drift due to depletion of ionic species, principally chlorides in their electrolyte. Drifting in a pH probe will not be readily apparent as it is usually a component of a feedback control loop. This will result in the addition of acid or base to maintain the fermentation at the now erroneous set point.

In an industrial context, the failure of sensors is much more frequent on small scale fermenters used for research and development. There are likely to be large numbers of these in a fermentation development establishment, they are likely to be equipped with lower cost instruments, which are used for prolonged periods between maintenance activities. In a production fermenter, where the cost of charging with media can be tens of thousands of pounds, there is a great incentive to use high quality, frequently checked instrumentation, and hardware redundancy. The techniques described in this thesis are therefore perceived to be primarily of use in R&D fermenter facilities.

The principal measurements made on a fermenter for control purposes are temperature, pH, air sparge flow rate, dissolved oxygen concentration, and off gas analysis and agitator rotation rate.

The control system on a fermenter can also fail because an actuator develops a fault. For example a reservoir of acid or base could become depleted, or cooling water flow could fail.

8.2.1 Loss Of Temperature Control.

Temperature control problems due to faults in the temperature probe could take two forms. In the first case, the probe could fail catastrophically, in which case it would return a meaningless output. Such output can be easily detected by a supervisory computer system simply by having an upper and lower limit test on the probe output. Alternatively the probe can start to drift, in which case an apparently correct

temperature will be maintained because the control system will trim temperature on the basis of the faulty data. This is difficult to detect except by the use of redundant sensors.

8.2.1.1 Effect of Temperature Faults On Biomass:

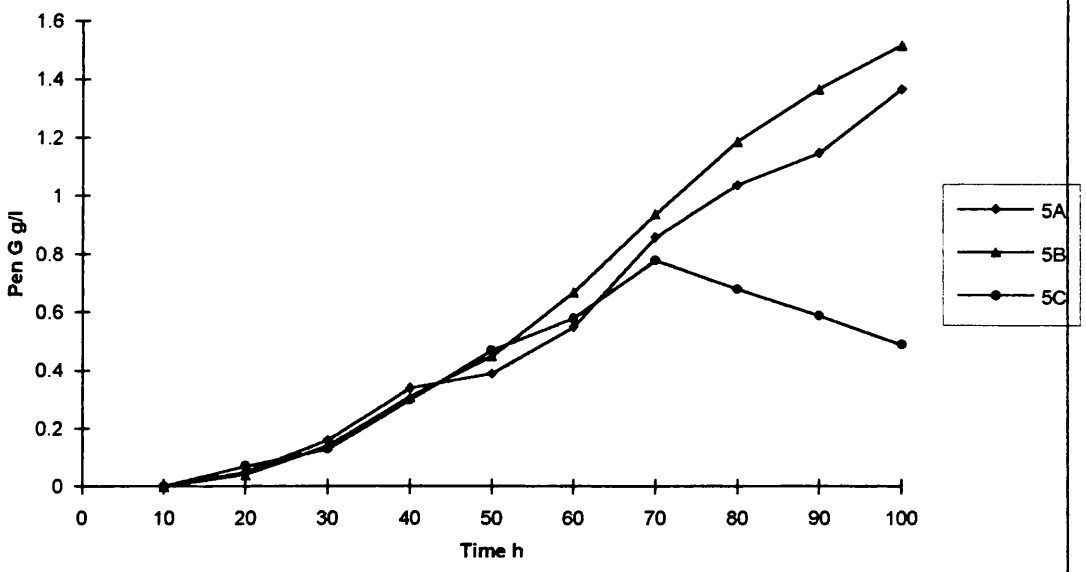
If the fault occurred in the exponential growth rate phase of a fermentation, the growth rate would be expected to change in line with the change in temperature of the system following Arrhenius's equation.. During the fed batch phase of the fermentation however, where growth is limited by carbon availability, biomass would be expected to change less dramatically. As the yield on glucose decreases with temperature, biomass could be expected to be lowered in the later stages of a fermentation if temperature of the system was to rise. As temperature approached 35°C the organisms would start to become adversely affected and a small further increase would kill the organisms.

Heat production from the organisms and mechanical energy dissipated from agitation mean that in a small scale fermenter, heating rates of 1°C in ten minutes can be attained. Typically, a penicillin fermentation operates at around 25°C, the organisms will be destroyed by temperatures above 40°C. Faults in temperature control systems can therefore destroy a fermentation in 2-3 hours.

In addition to the effect on biomass a fault in temperature control will have a significant impact on penicillin concentration. The rate of decay of penicillin is sensitive to both temperature and pH. As will be seen below, the concentration of penicillin in the fermentation is a useful indicator of a temperature high fault.

The fifth set of three fermentations carried out at Smithkline Beecham was used to obtain data on the impact of a heater set point fault. In order to illustrate the effects, a step change in temperature set point was imposed in one of the fermentations, from 26°C to 36°C after 70 hours. The temperature in the 1.5 l fermenter rose to 36°C within a matter of minutes. The resultant data can be used to demonstrate which of the possible impacts of an unplanned temperature change are significant.

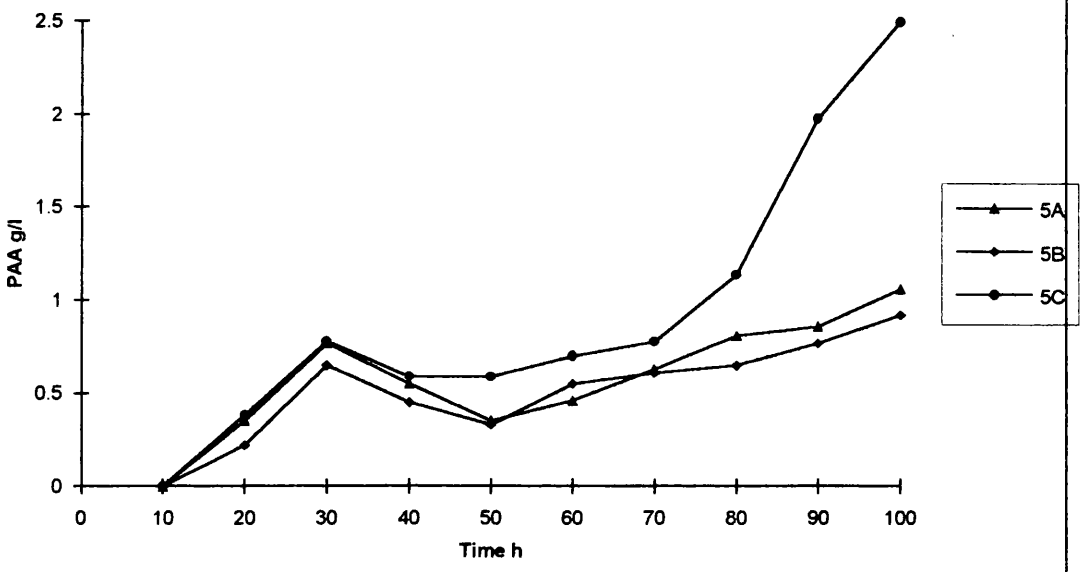
Fig 8.3 Fermentations 5a, 5b, 5c penicillin concentration



This graph shows what happens to the level of penicillin when the temperature of a fermentation goes from 26°C to 36°C. The higher temperature in 5c results in hydrolysis of the penicillin, producing a dramatically different penicillin concentration profile from the controls (5a and 5b).

The impact of the temperature rise on phenylacetic acid uptake also turned out to be significant.

Fig 8.4 Phenylacetic Acid concentration

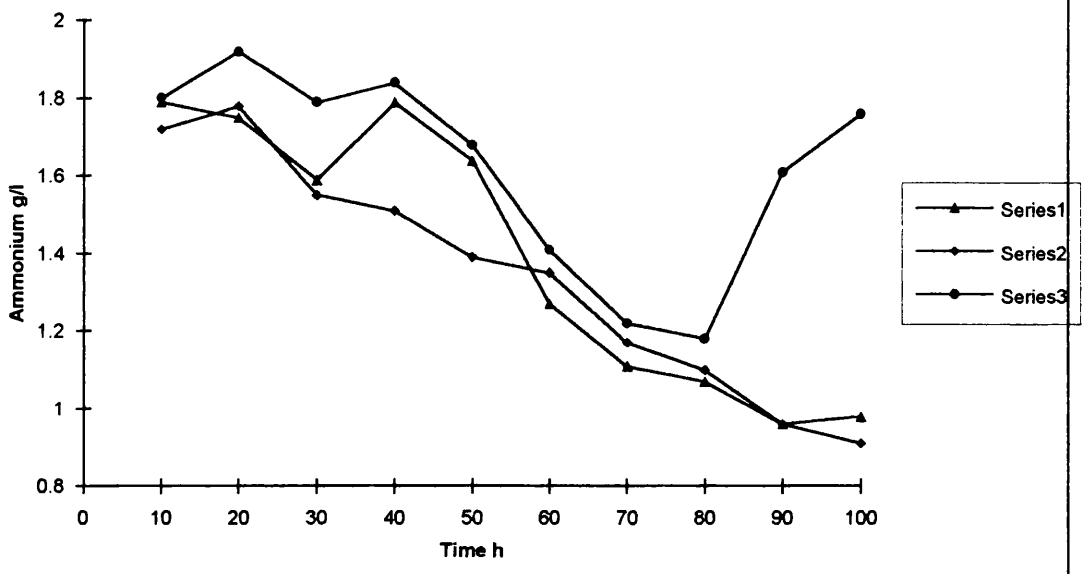


This graph shows the impact of a sudden increase in temperature from 26°C to 36°C on the concentration of phenylacetic acid. In all these fermentations, the feed rate of phenylacetic acid was identical (They were all fed using a common peristaltic pump, capable of handling multiple tubes). The temperature rise in fermentation 5C resulted in a cessation of phenylacetic acid uptake, resulting in the increase in phenylacetic acid concentration.

In fermentation 5C, the temperature rise was enough to halt penicillin production, and hence halt the uptake of phenylacetic acid.

An interesting side affect of the increase in phenylacetic acid concentration and higher temperature was a concurrent increase in broth ammonia concentration. This reflected greater feeding of ammonium hydroxide which was used to establish pH control and the impact of a sudden temperature change on carbonate solubility, resulting in the release of carbon dioxide into the broth.

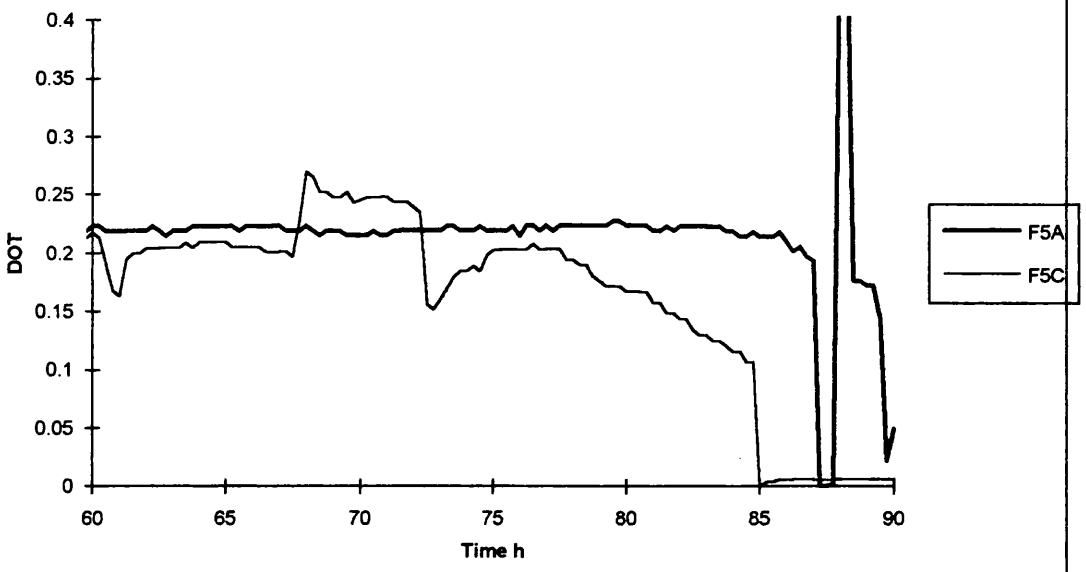
Fig 8.5 Ammonium concentration g/l



This graph shows the impact of a temperature shift on broth ammonia concentration. In fermentation 5C the increase in temperature resulted in an increase in ammonium hydroxide addition to maintain pH control. This is due to the release of carbon dioxide into solution as the equilibrium between carbonate and carbon dioxide shifts. The required level of ammonium hydroxide is also probably responding to the increased level of phenylacetic acid.

The other parameter affected by temperature is dissolved oxygen tension (DOT). Oxygen solubility is reciprocally related to temperature, the Henry's law constant increasing by 16% between 25°C and 35°C (Perry 1983) hence DOT will fall as temperature rises.

Fig 8.6 Dissolved Oxygen as Fraction of Saturation Concentration (DOT)



This figure shows the decline in dissolved oxygen with temperature after the temperature is shifted. This is due to a decrease in oxygen solubility with temperature.

The impact of loss of temperature cooling and therefore temperature control is summarised as shown overleaf.

Table 8.2 Impact of Loss Of cooling on Fed Batch Penicillin fermentations in the steady state part of the process.:

Parameter	Impact
Biomass	If the loss of cooling is during the steady state part of the fermentation, no noticeable additional biomass is produced. This is because growth is limited by the availability of carbon.
Penicillin	Penicillin titre declines as degradation speeds up.
Glucose	No noticeable impact as broth glucose levels are below threshold of detection. In principle, as metabolic activity rises with temperature, glucose levels will decline.
Phenylacetic Acid	Will not be metabolised, so will attain higher than usual levels in the fermentation.
pH	No change in pH, as it is controlled on a feedback loop. The rate of addition of alkali will be atypically high.
Dissolved oxygen	Oxygen solubility and hence dissolved oxygen levels will decline with temperature rises. The level of dissolved oxygen will also be impacted by higher meatbolic activity levels.
Ammonium concentration	Ammonium concentrations will increase as the control system reacts to phenylacetic acid accumulation.

8.2.2 Failure in pH Control

Penicillium chrysogenum, when growing produces acid. pH control is typically effected using ammonium hydroxide, which neutralises the acid and supplies nitrogen to the fermentation. Control systems sometimes also include an acid feed to correct pH controller overshoot. pH probes are subject to drift, or failure because they rely on a membrane across which an electrochemical potential can be measured to determine pH, this membrane can become fouled. Alternatively the electrolyte reservoir on the probe side of the system can become depleted either of water or electrolyte. The result of these effects would be a gradual drift of the probe output away from the true pH of the system.

Unfortunately, a shortage of experimental time prevented data from being gathered in this area.

8.2.3 Impact Of Aeration System Faults

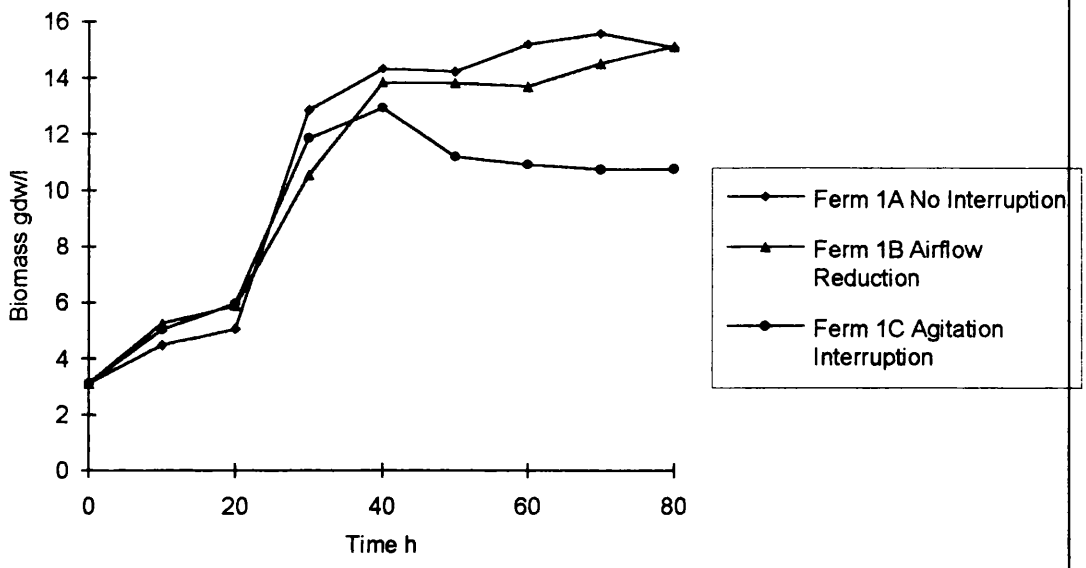
The oxygen supply to a penicillin fermentation is critical. Oxygen is only sparingly soluble in water (1.26 mmol/l at 25°C (Bailey and Ollis 1986)). Given that typically during the steady state component of a fermentation the dissolved oxygen level is at approximately 40% of its saturated value, and neglecting all oxygen utilisation other than for respiration, it can be estimated from CER data that the fermentations in this work would run out of oxygen in 2 -3 minutes if the air supply was completely cut off.

Because of the sensitivity of the penicillium fermentation system to aeration faults, this section provides an opportunity to review key fault detection algorithms in detail. This section makes a comparison between the usefulness of simple data, ratios of key data, Kalman filters and principal component analysis based on both major faults and short term transient process disturbances.

Aeration interruptions can take the form of a reduction or cessation of aeration air, or a shutdown or slowdown of the agitator, resulting in reduced oxygen mass transfer. The transfer of oxygen from the air supply to the bulk of the fermentation is dependent on the creation of a large interfacial area for mass transfer by the agitator. In the first set of fermentations carried out at SmithKline Beecham, the effects of an interruption in agitation and a restriction in air flow were studied.

The relative importance of an interruption to agitation, and a restriction in air flow are apparent from the change in biomass concentration after 30 hours:

Fig 8.7 Biomass - Aeration interruption after 45 h (Fermentation 1A,B,C)

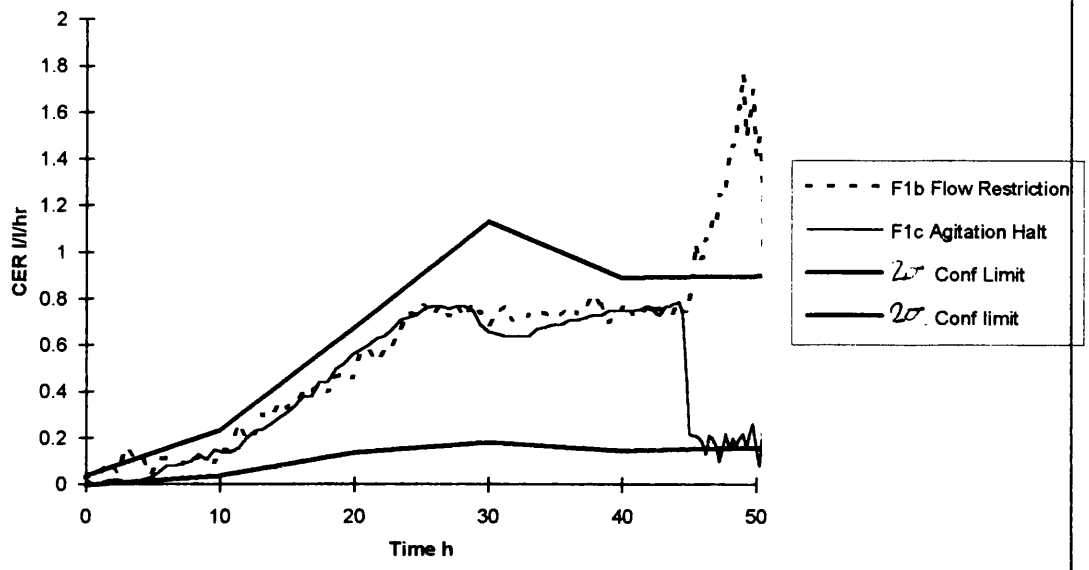


Impact of aeration and agitation interruption on observed biomass. It is interesting to note the differing impact of a reduction in aeration across the fermentation, and a cessation of agitation, which would effectively mean that the broth near the sample port was both static and unaerated.

This type of fault signature would be very useful in an expert system used to identify faults.

The decline in biomass in fermentation 1C is not sufficient for the 2σ confidence limit for the expected biomass range to be exceeded (See fig 6.1). The other aspect of these faults however is a sharp change in apparent carbon dioxide evolution. As would be expected a restriction in air flow results in an increase in carbon dioxide concentration in the off gas, while an agitation stoppage results in a decrease. The data gathered was as follows:

Fig 8.8 Impact Of Aeration Fault On Apparent CER



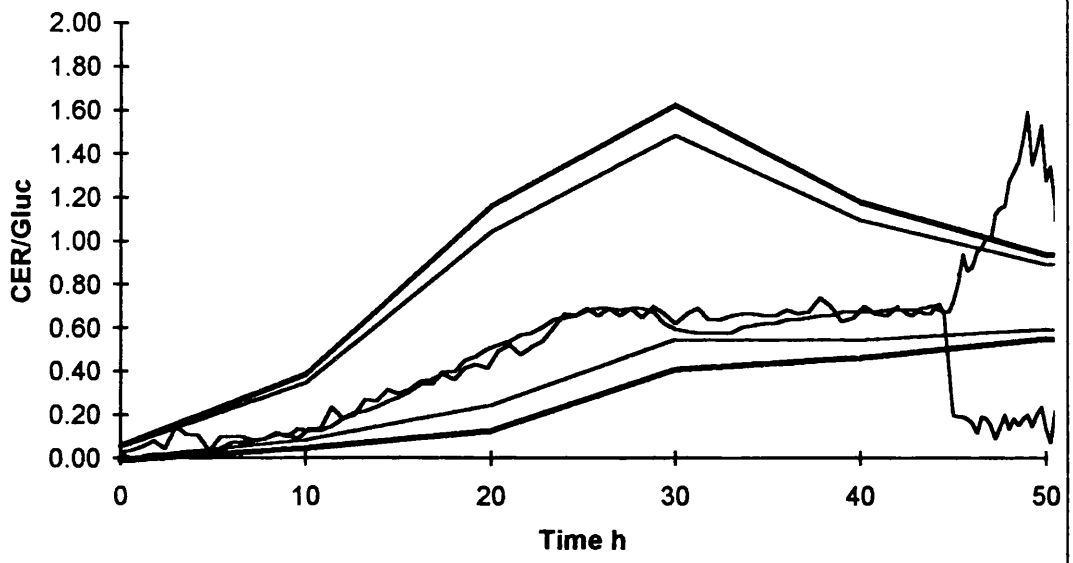
Reduction in air flow rate causes an apparent rise in CER, as the concentration of carbon dioxide in the exit air stream goes up, this is due to the fact that the calculated CER depends on an assumption that the airflow is as pre-programmed. Therefore, an increase in carbon dioxide concentration as detected by the mass spectrometer will indicate that the quantity of carbon dioxide produced in the fermentation is higher than it really is.

The impact of a halt in agitation is that the CER goes down. This is because the biomass is producing less carbon dioxide, because it is able to receive less oxygen, in this case a reduction in CER is observed, this is in part due to the fact that with less air / water interfacial area mass transfer of carbon dioxide out of the broth will be reduced, and in part due to a reduction in carbon dioxide production due to the non availability of oxygen to the bulk of the *penicillium* organisms.

As can be seen in fig 8.8, both of these events produced distinctive fault signatures which would transcend expected 2σ CER limits. The key point to note is that each type of fault gives a very different signature vs. the expected CER trajectory. The signatures from these incidents can be seen much more clearly by comparing data

from the current fermentation with limits adjusted for glucose feed rate. This gives the following output:

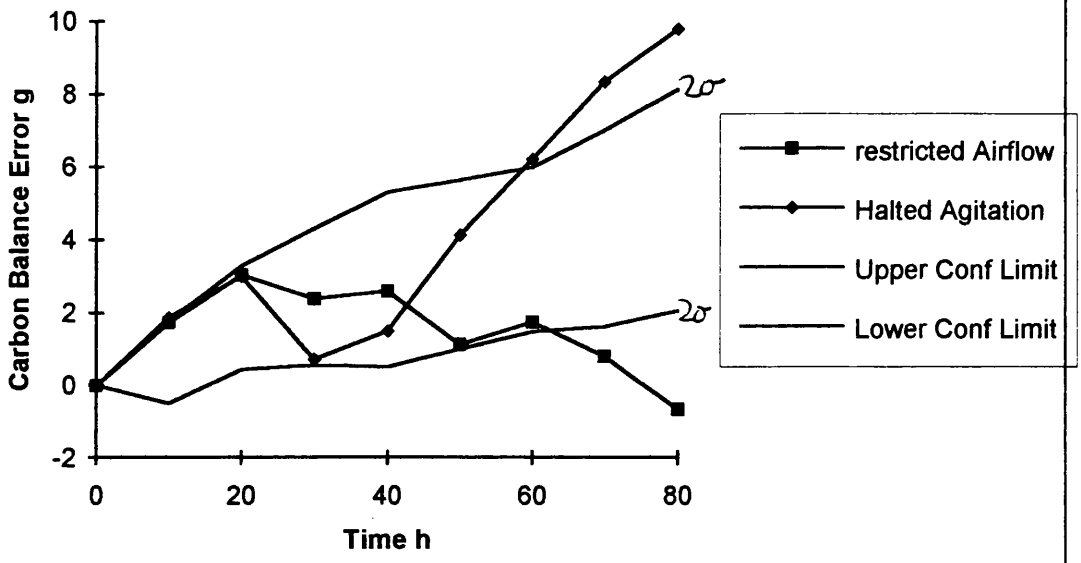
fig 8.9 Apparent CER/Glucose Addition Rate: Ferms 1b, 1c



This graph shows the same fault event as in fig 8.8. In this case the trajectory displayed is apparent CER/Glucose Feed vs. Time. Both fault incidents produce trajectories that go outside the 3σ confidence limits. This is a significant advantage in a fault detection scheme because the use of 3σ limits will reduce the false alarm rate.

The change in off gas analysis will also result in unusual mass balance reconciliation results. This can be seen in fig 8.10

Fig 8.10 Impact Of Aeration Faults On Carbon Balance Error: Ferms 1b, 1c



CER data is important in establishing running carbon balances in the fermentation. Any event which changes the calculated CER value will result in a change in the reconciliation between carbon inputs and carbon outputs, which is what this graph shows.

The mass balance, in the event of the agitation fault is biased by the incomplete utilisation of glucose as oxygen becomes the limiting nutrient. The assumption in the carbon balance calculation is that all glucose is utilised as it is added after the end of the exponential phase. In this case the bias is in the direction of insufficient exit carbon in comparison to input carbon.

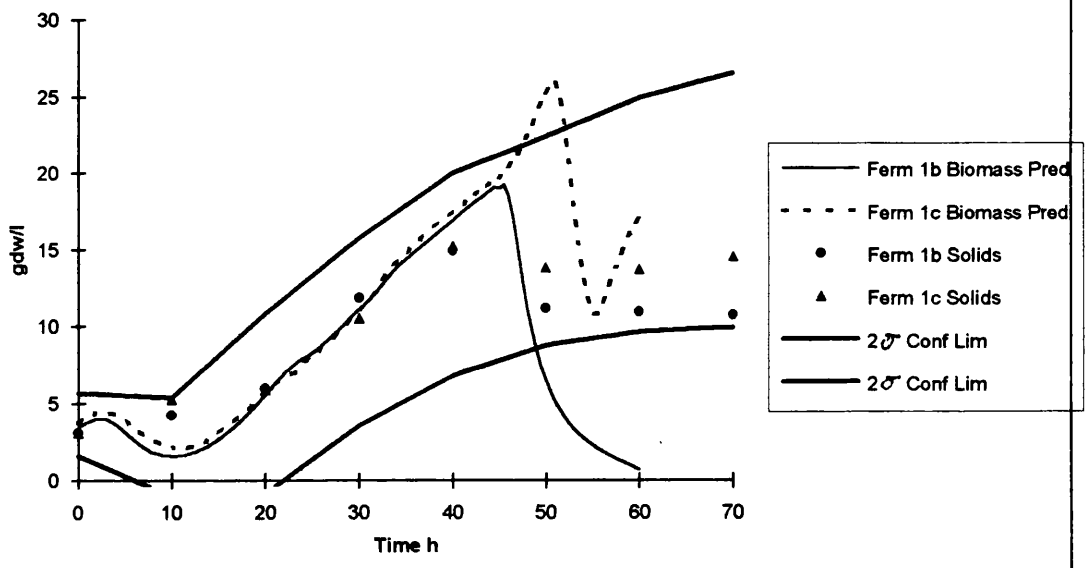
In the event of the aeration fault, the mass balance is biased in the opposite direction. The apparent quantity of carbon leaving the fermentation is the product of programmed air flow rate and observed carbon dioxide concentration. Given a reduction in air flowrate, the carbon dioxide concentration will increase. As the outlet air flow rate is not measured, but assumed to be as pre programmed, the outcome is an overestimate of carbon outputs.

Both of these faults resulted in detectable levels of glucose in the broth, this could have potentially been identified on line, if a probe capable of tracking glucose levels was installed in line. Such probes are not however typically installed on the type of fermenter considered in this thesis.

The mass balance data gives a much slower indication of the occurrence of a fault than the direct calculated CER data (or the raw CO_2 output data). In addition, being dependent on sampling there would be additional delay in obtaining a mass balance result which would detect either of these faults.

The mass balance is dependent on biomass and it is possible to estimate biomass using the Kalman filter. The following results were obtained.

Fig 8.11 Kalman Filter Output, Ferms 1b, 1c.



Fermentations 1 b and 1c, broth solids measurement vs. biomass estimates. Limits determined from Kalman Filter biomass estimates.

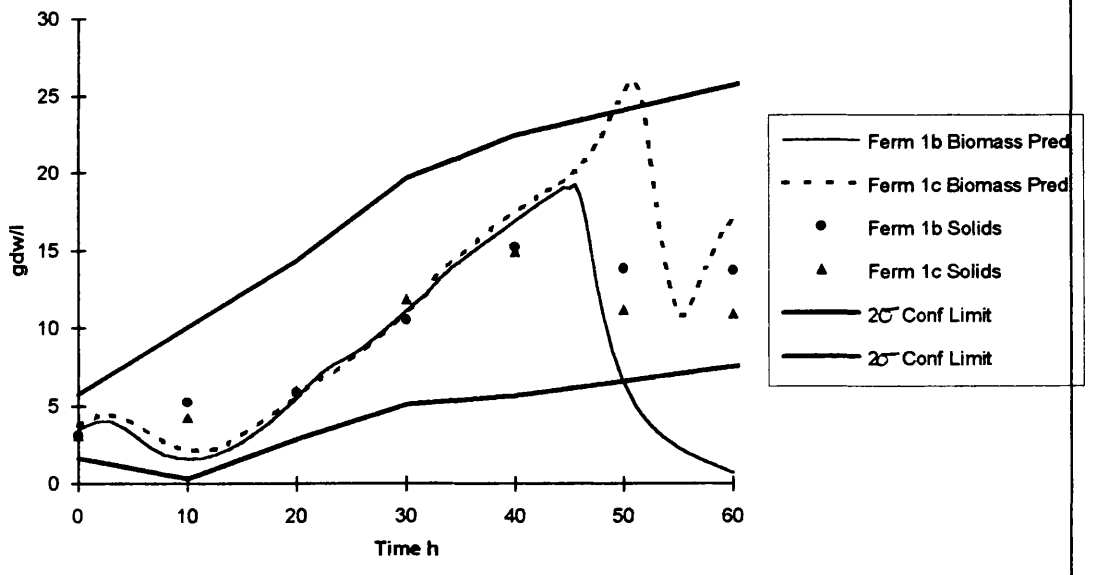
In effect, the Kalman filter is being driven by erroneous CER data, so it is producing an inaccurate biomass estimate.

The confidence limits quoted above are based on the population of successful fermentations. The Kalman filter outputs respond to the change in observed CER, giving clear indications that a fault has occurred. This happens more slowly however than the direct use of CER data (fig 8.8). The response of the trajectories to the change in CER however differs from the CER output and thus builds a useful extra element into a fault detection system.

The filter output is a compromise between modelled biomass, and biomass estimated from carbon dioxide evolution, and it acts as a damper on the fault detection system. This is likely to eliminate false alarms due to transient CER spikes. The data presented in section 5.5.3 shows that these spikes are common features of practical CER data. The use of a Kalman filter to achieve this effect is however not the most practical approach. A simple moving average of carbon dioxide data would yield the same data smoothing benefit.

It is interesting to l3 at the comparison between the CER estimate of biomass and the expected biomass trajectory limits determined from measured biomass.

Fig 8.12 Kalman filter output Ferms 1b, and 1c : Confidence limits based on broth solids measurement.



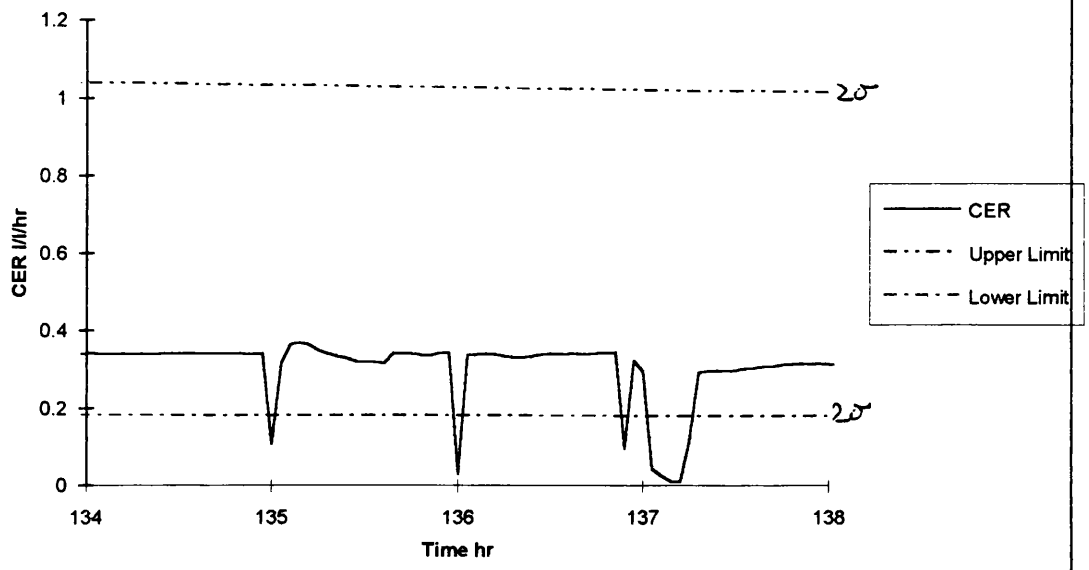
Fermentations 1b and 1c Kalman Filter biomass estimates vs. measured biomass, limits based on measured biomass data.

As can be seen in fig 8.12 above, the narrower spread of the 2σ confidence limits for Kalman filter estimates, based on measured biomass (vs. biomass estimated using Kalman Filters in other fermentations) gives a narrower range of 'on spec.' biomass estimates and therefore makes for a more sensitive fault indicator.

The advantage of using the Kalman filter in this way is that it incorporates more information about how a fermentation can be expected to behave than simple CER data, and thus in the case of the agitation failure gives a more clear-cut indication of failure than simple CER data, or mass balance data depending on off line samples. (Biomass data by itself would not have flagged a fault in either of the case studies followed in this section).

Another case in which airflow became blocked, this time on the exit to the fermenter, occurred by accident in fermentation 12. The exit air line became intermittently blocked as foam was transported up the exhaust air line. This problem was easily fixed by the addition of antifoam, allowing the fermentation to continue. The problem was resolved in this case because the fermentations were highly monitored, this level of monitoring is unlikely to be found in industrial applications. The CER data obtained from this fault event are as follows:

Fig 8.13 Exit Filter Blockage : Impact on CER: Ferm 12

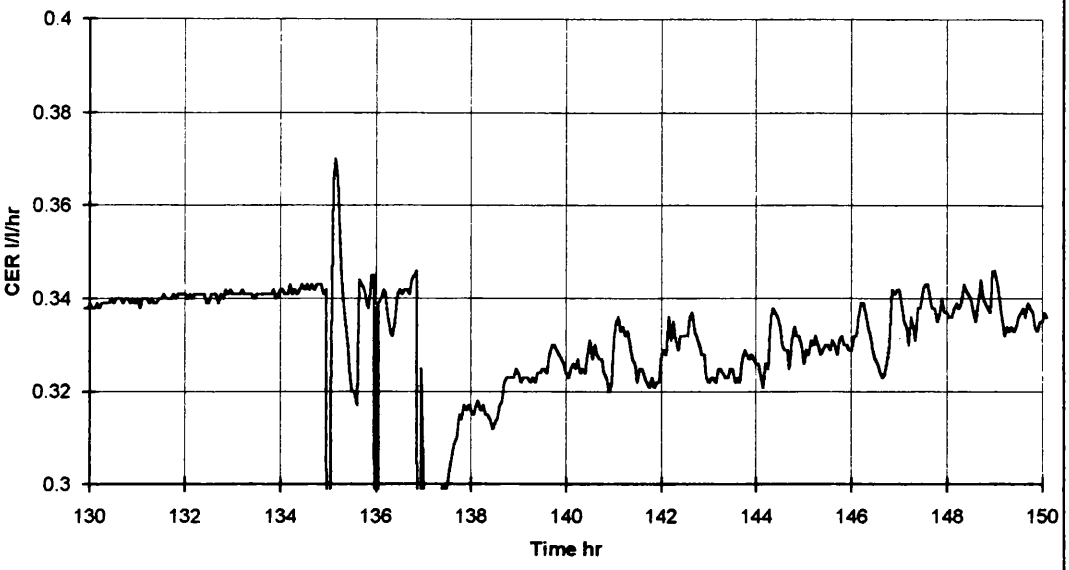


It can be seen from this data that the faulty exit gas filter could have been detected at 135 hours with an on line fault detection system. It was actually picked up at 137 hours after it had caused significant changes to the fermentation trajectory.

As can be seen, the fault would have been picked up when it was incipient, at 135 hours, several hours before the filter became fully blocked. In this case, as the filter was downstream of the fermentation, the blockage resulted in an apparent reduction in carbon dioxide production.

This blockage did marginally influence the course of the fermentation as can be seen in fig 8.14, this could have been avoided if the filter was changed after 135 hours.:

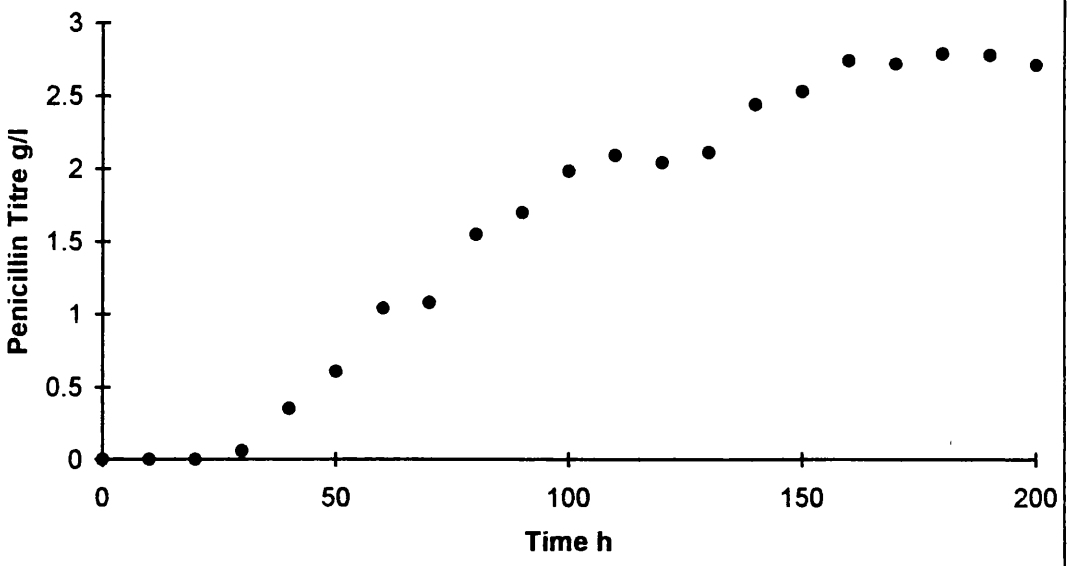
Fig 8.14 Exit Filter Blockage, Effect on CER: Ferm 12 over 20 hours.



Graph showing the recovery of carbon dioxide production in fermentation 12. It can be seen that carbon dioxide production is not restored to its original level for 10 - 12 hours after the fault was corrected, and even then has probably not reached the point where it would have been if the fault had not ocurred.

If this fermentation was part of a wider experimental design, the impact of the blockage event would have to be taken into account when assessing the overall results of the experiment. In this case, there was no obvious impact on penicillin production as the following graph shows:

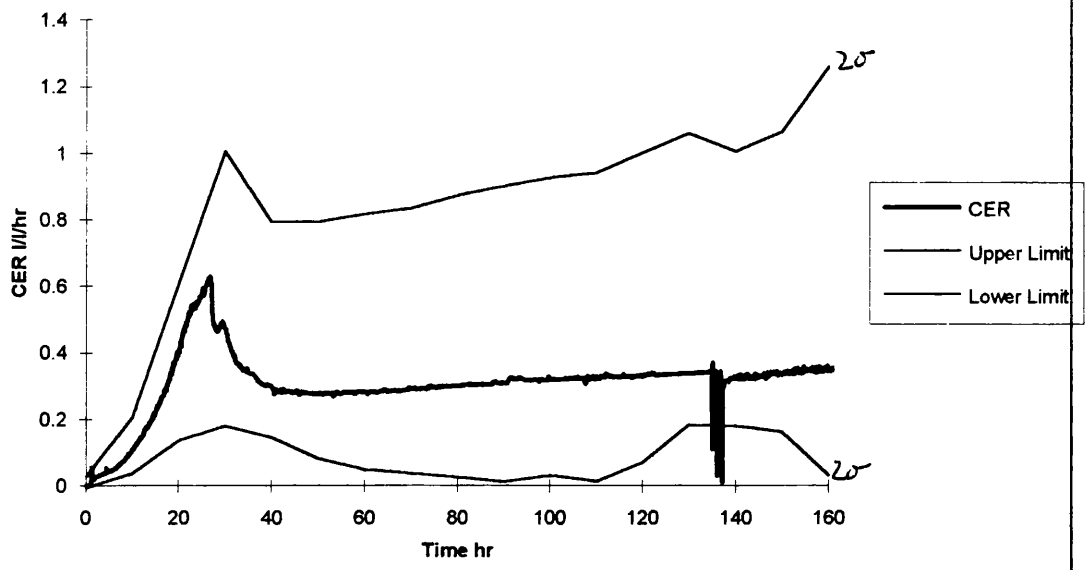
Fig 8.15, Penicillin Production : Ferm 12



however, the fault could have resulted in sub-optimal penicillin production, resulting in the distortion of an experimental program..

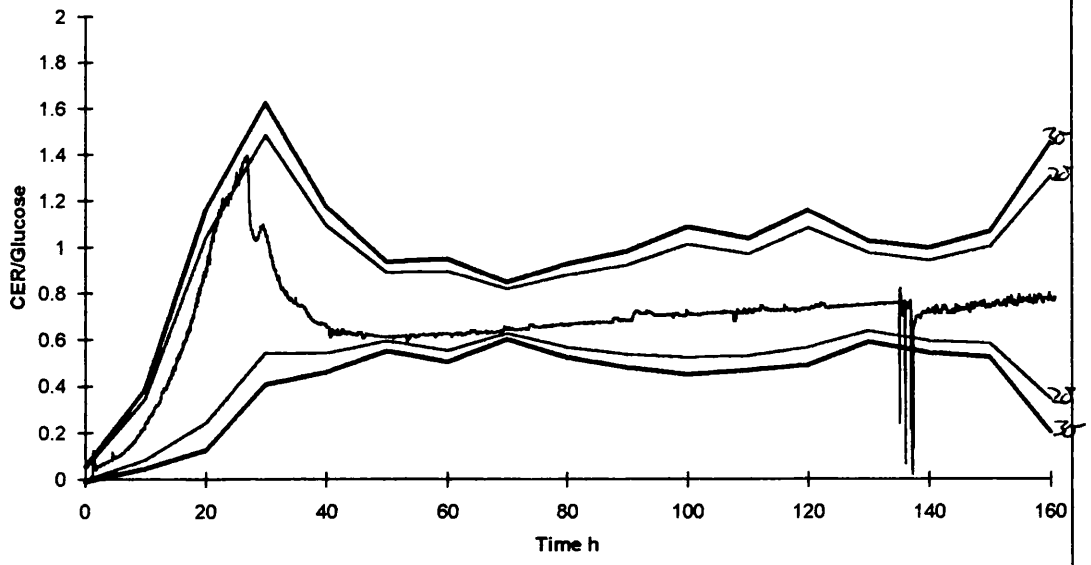
This fault case study shows the value of going beyond a simple statistical process control strategy for fault detection. The 2σ confidence limits for CER give a very wide range for variation:

Fig 8.16 CER: Ferm 12



If CER is divided by glucose feed rates, the following output and control limits are obtained:

Fig 8.17: CER/Glucose Feed : Ferm 12.



Graph shows the impact of normalising CER data using glucose feed data. By comparison with Fig 8.16 it can be seen that much tighter confidence limits can be deployed. This will produce more sensitive fault detection vs. limits based on unnormalised data.

Using this adjustment, it becomes possible to use 3σ confidence limits because of the much smaller range over which the data varies. The use of 3σ confidence limits is preferred in an automated fault detection system as this will reduce the rate of false alarms.

It is interesting to look at the impact of short transient disturbances on Kalman Filter output:

Fig 8.18: Comparison of Kalman Filter Output to CER

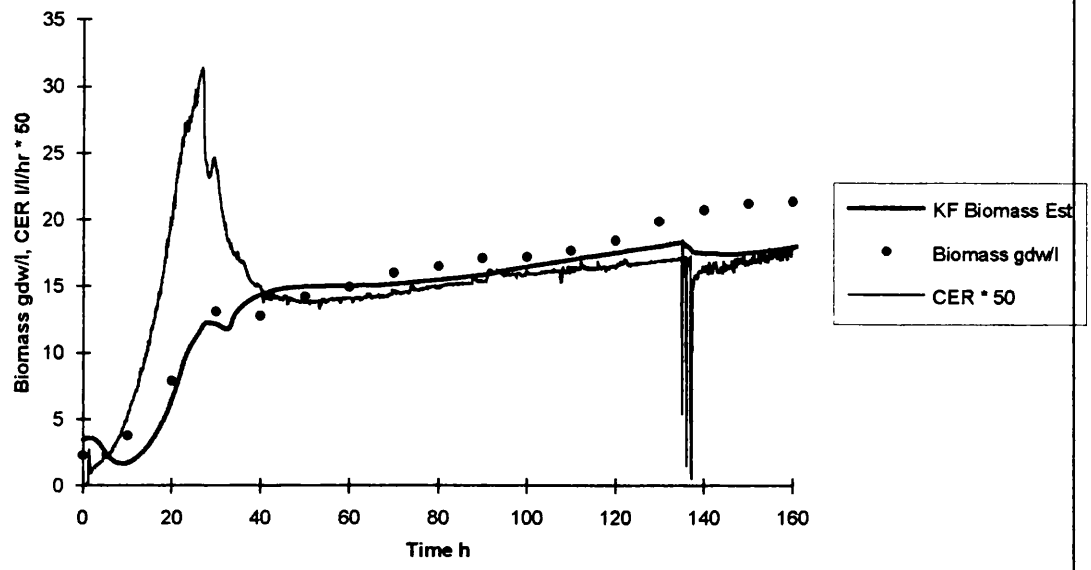
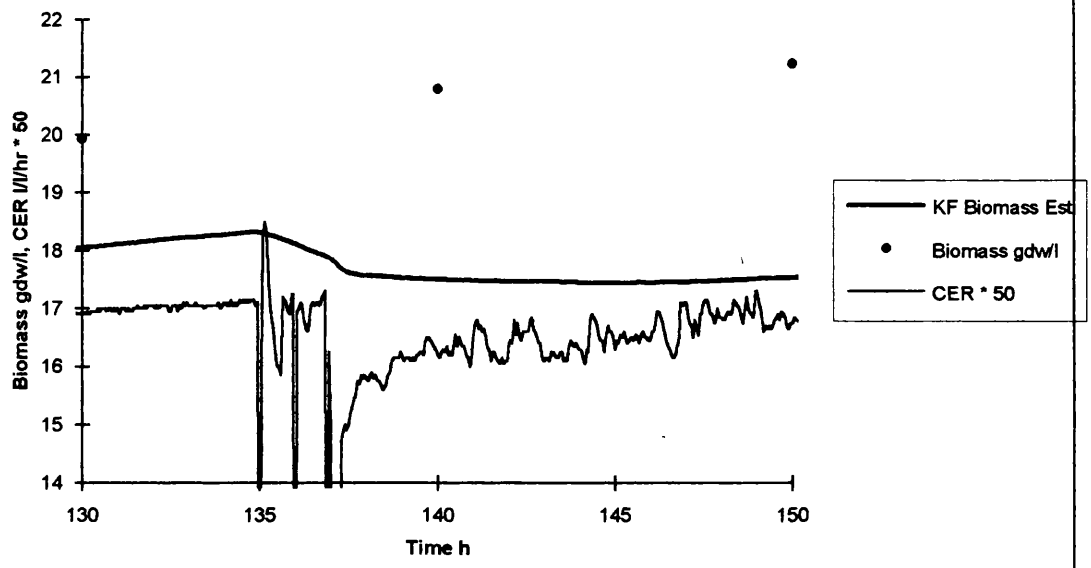


Fig 8.19 Comparison of Kalman Filter Output to CER

Kalman filter biomass estimate vs. CER vs. broth solids measurement over the fault event.



These two figures show the overall Kalman filter trajectory, and the fault event in close up.

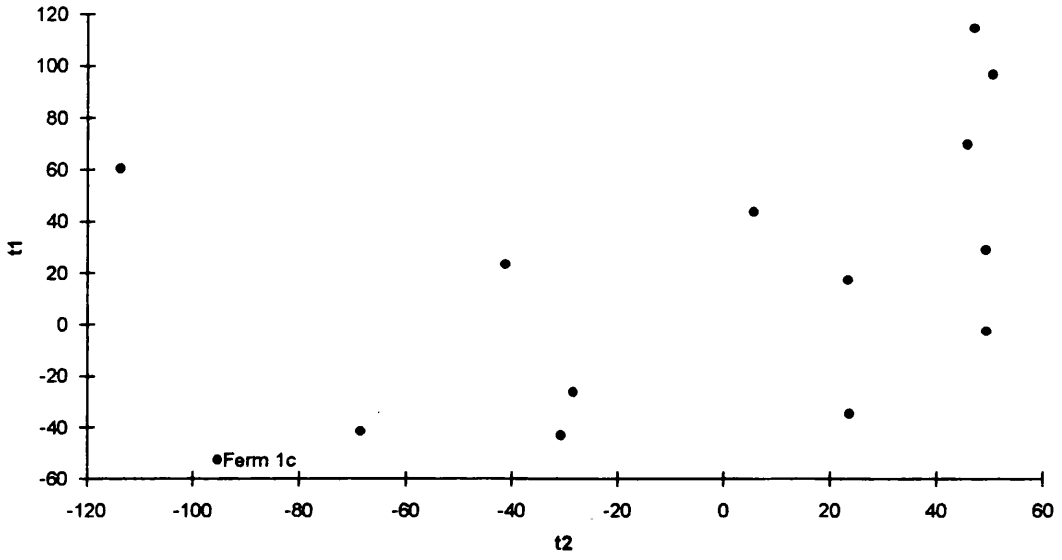
One can see that the filter only responds quickly to the fault in that it changes gradient at 135 hours. The absolute deviation from the pre fault trajectory is however much

smaller than that of the CER trajectory. This is ultimately a facet of how the filter is tuned. Filter tuning essentially implies making a compromise between the model used in the filter, in this case the Bajpai and Reuss model, and the error between estimated and observed outputs. If the filter was tuned so that there was no error between estimated and observed CER, the biomass trajectory predicted by the filter would closely match in shape the CER curve.

In tuning the filter to give accurate biomass estimates, one reduces its sensitivity to rapid CER transients. This gives it a potential advantage over raw CER data as a fault indication tool, as it will be less responsive to short time length disturbances which are in any case unlikely to disturb a fermentation, at the cost of a loss in sensitivity to significant changes.

An attempt was made to use Principal Component Analysis to identify these aeration faults. The following result was obtained:

Fig 8.20: Principal component analysis output for the fermentation set used in this project. Data from good fermentations only is considered with the exception of fermentation 11.



Principal component analysis output. Principal components generated using the algorithm reported by Nomikos et al. As it can be seen the fermentation with the significant aeration problem produced principal component values on the edge of the population of values obtained from satisfactory fermentations based on biomass, CO2 production, glucose feed rate, and penicillin production data.

It was concluded from this effort to employ PCA that the data available to this project was insufficient. Nomikos bases his work on an initial data set of 50 satisfactory batches (in his case of emulsion polymerisation reactions).

The output from PCA was unhelpful in this instance as the relatively low level of background data available did not permit the appliance of sufficiently tight confidence limits to detect the faulty fermentation.

In conclusion, aeration faults can be detected by their impact on CER. This section shows that the most effective test for aeration faults is a comparison between expected CER/Glucose Feed Rate and observed CER/Glucose Feed Rate.

Kalman filter estimates of biomass are useful in detecting major faults, but do not respond to short term transients if tuned appropriately. The Kalman filter can therefore be helpful in fault detection by acting as a digital filter, preventing alarms due to very short term incidents such as automatic mass spectrometer recalibrations. However, it is rather a complex algorithm for achieving this, analysing CER data on a moving average basis would give similar benefits.

Table 8.3 Impact of Agitation and Aeration Faults on Fed Batch Penicillin Fermentations:

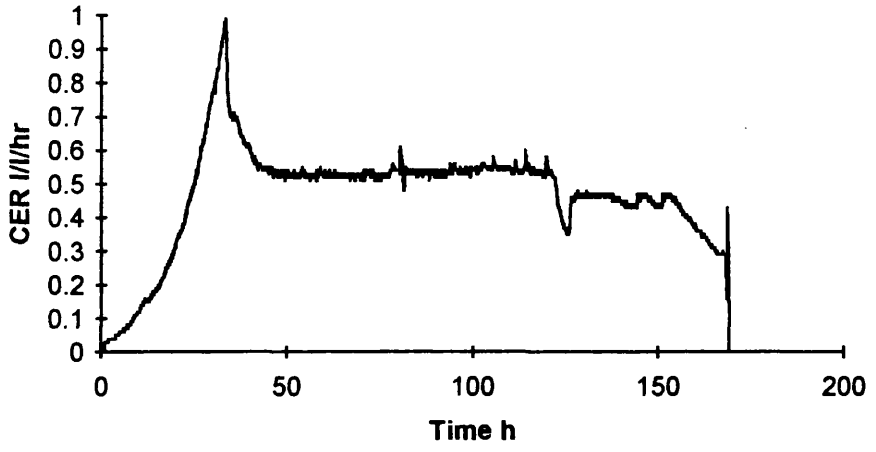
Parameter	Aeration fault	Agitation fault.
Biomass	Reduction in air flow produced no lasting impact on biomass production. enough oxygen must have been distributed to prevent significant cell death Clearly the agitator distributed oxygen uniformly avoiding dead zones.	A halt in agitation produced a significant reduction in biomass levels. This was probably due to areas of the fermenter being completely starved of oxygen,
CER	A restriction in air flow produced an unexpectedly high carbon dioxide concentration in the exit gas stream.	Loss of agitation resulted in lower than expected levels of carbon dioxide in the exit gas stream.

8.2.4 Glucose Feed Faults

An interruption in the feeding of glucose would quickly impact on biomass, penicillin and carbon dioxide production. As carbon is the rate limiting nutrient in this system, the reserve of glucose is small and would quickly be exhausted. This will produce a rapid impact on the fermentation.

An example of this was seen in fermentation 9, where a feeding fault occurred as the glucose reservoir became empty at around 125 hours. This showed up on the CER trajectory as follows:

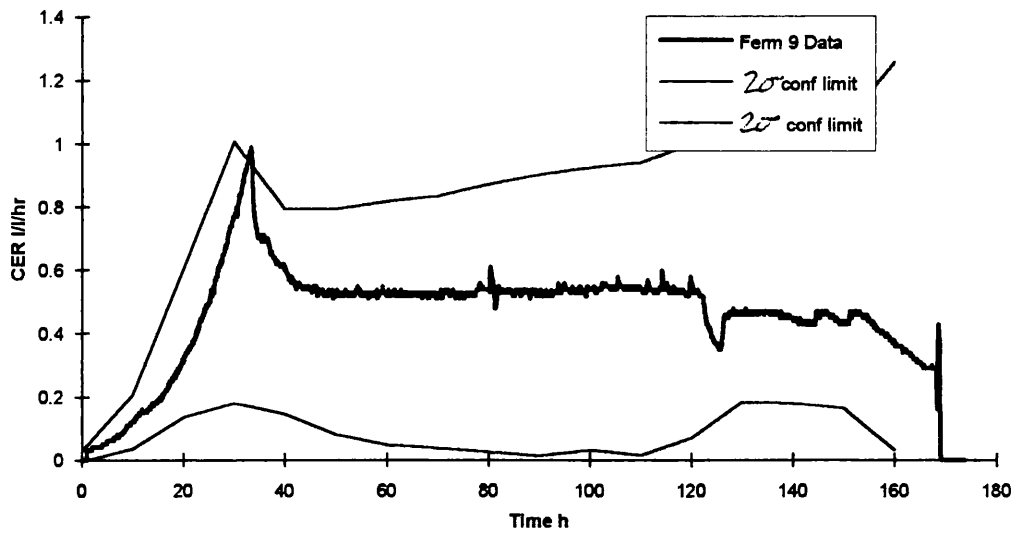
Fig 8.21 Fermentation 9 CER Trajectory



This graph shows the impact on CER of an interruption in glucose feed, which was rectified 8 hours after it occurred..

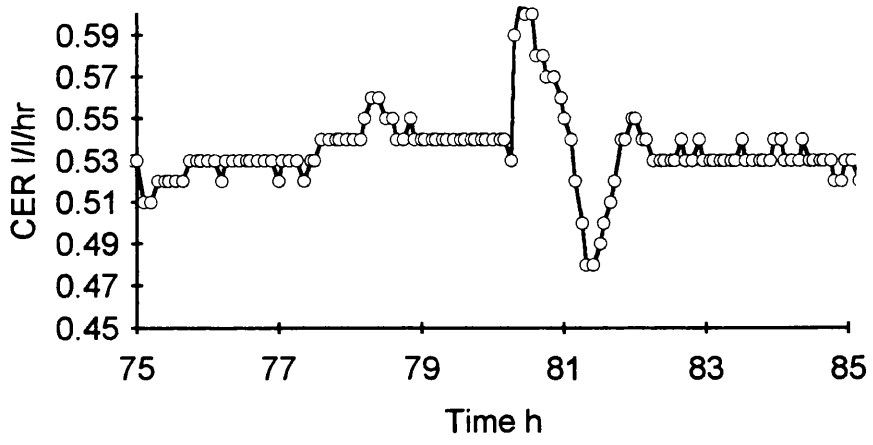
Comparing this trajectory with the 2σ confidence limits for CER one gets the following result:

Fig 8.22 Fermentation 9 CER Trajectory including 2σ Confidence Limits



One can see that simply comparing CER data is inadequate as a means of fault detection. The limitations of using rates of change of CER are also apparent if an earlier event occurring at 80 hours is investigated :

Fig 8.23 Fermentation 9 CER Trajectory During Fault Event



This fluctuation in CER over about two hours was due to a depletion of the hydrochloric acid reservoir in the fermenter pH control system, leaving the fermenter unable to compensate for an overfeed of ammonium hydroxide, except by production of acid by the *Penicillium.* This fault could also be detected by tracking pH, however

unable to compensate for an overfeed of ammonium hydroxide, except by production of acid by the *Penicillium.* This fault could also be detected by tracking pH, however this would be complicated by the production of acid by the fermentation process (which is why many small scale fermenters are not fitted with acid addition systems). Without an acid addition system return to set point after over addition of ammonium hydroxide would still occur, the return would be somewhat slower.

The much larger impact fault changed penicillin production as can be seen by the change in the gradient of the penicillin production curve after 130 hours in fig 8.23.

Fig 8.24 **Fermentation 9 Penicillin Titre.**

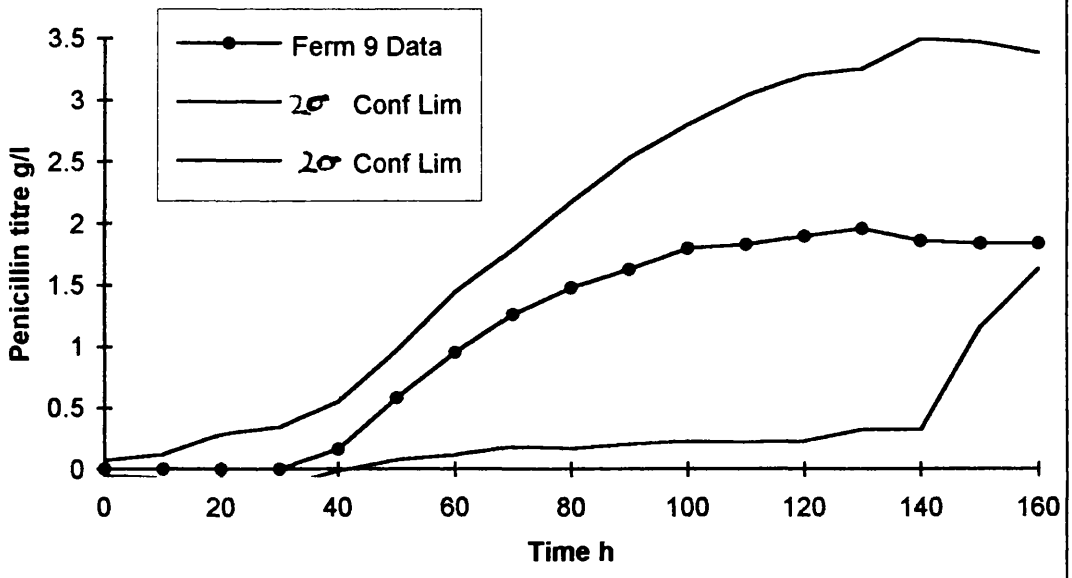


Figure shows the impact on penicillin titre trajectory of an interruption in glucose feed after 128 hours. The level of penicillin starts to decline due to a reduction in active cell population. This results in the rate of degradation and dilution becoming higher than the rate of production/

whereas the event at 80 hours had a lesser effect, although the decline in penicillin production does start at this point. Similar data was obtained from biomass measurements:

Fig 8.25 Fermentation 9 Biomass Titre

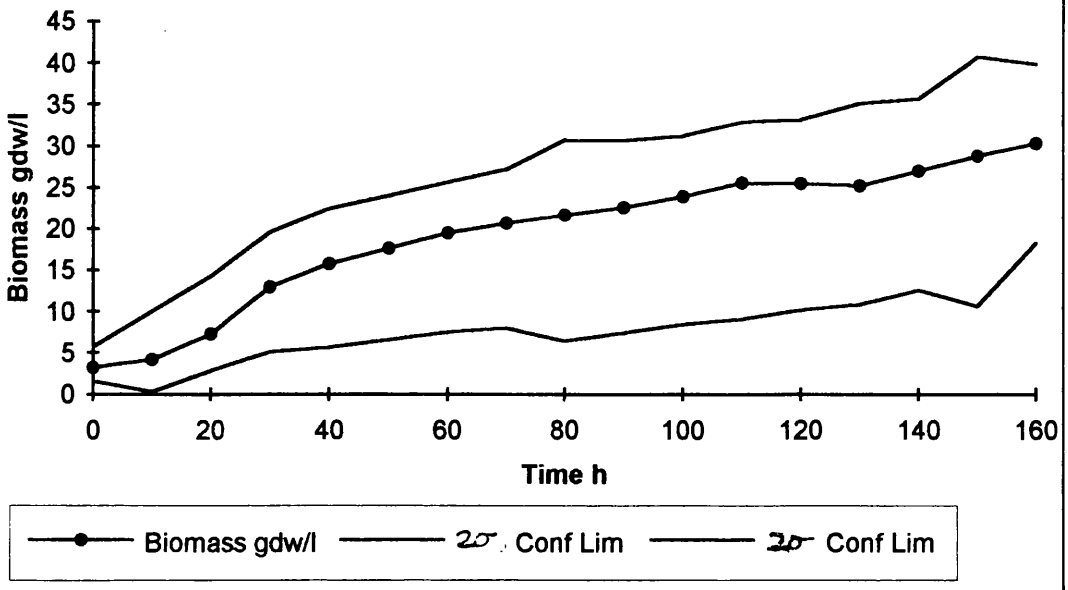


Figure shows the mild effect on biomass measured on a dry weight basis of the interruption in glucose feed.

Looking at the basic data, there has been a fault, but none of the measured data would have revealed this by comparison with a database of other similar fermentations.

If the CER is related to glucose feed, the following is obtained:

Fig 8.26 Fermentation 9, CER/Glucose Feed Rate

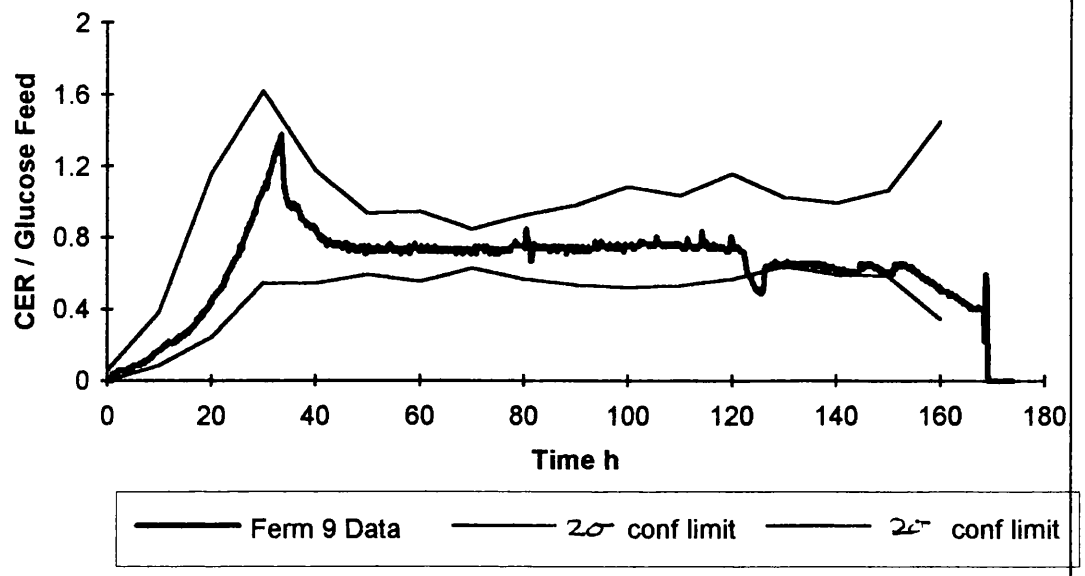


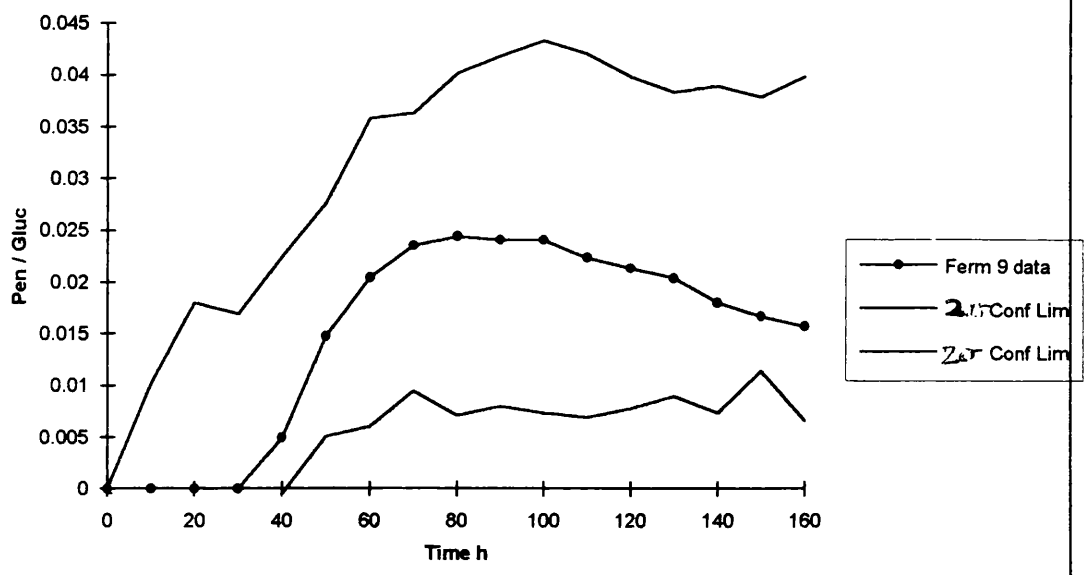
Figure shows the impact of the glucose feed fault on the ration of CER / Glucose feed.

This graph shows the benefit of using normalised data. The CER/Glucose trend clearly goes outside its expected range. By comparison (see graph 8.22) the CER trend does not go below its expected lower limit.

One can see that this measure gives a much more effective means of fault detection, because the confidence limits are tighter. In this case, the minor fault occurring at 80h is not detected but the major fault occurring at 125 h is detected. In addition, the peak CER during the end of the exponential phase remains within the range of expected data.

Comparing other data when ratioed with glucose feed rate does not produce similar results:

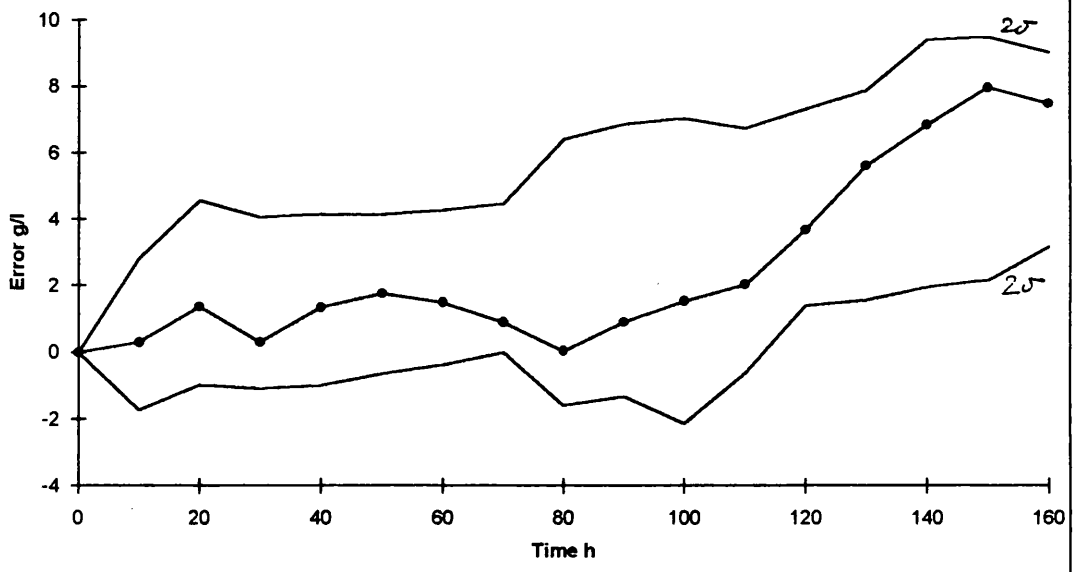
Fig 8.27 Fermentation 9 Penicillin Produced/Glucose added



In this figure, it can be seen that the ratio of penicillin produced / glucose consumed gives no useful indication that a fault has ocurred.

Another way of combining data is to look at an overall mass balance:

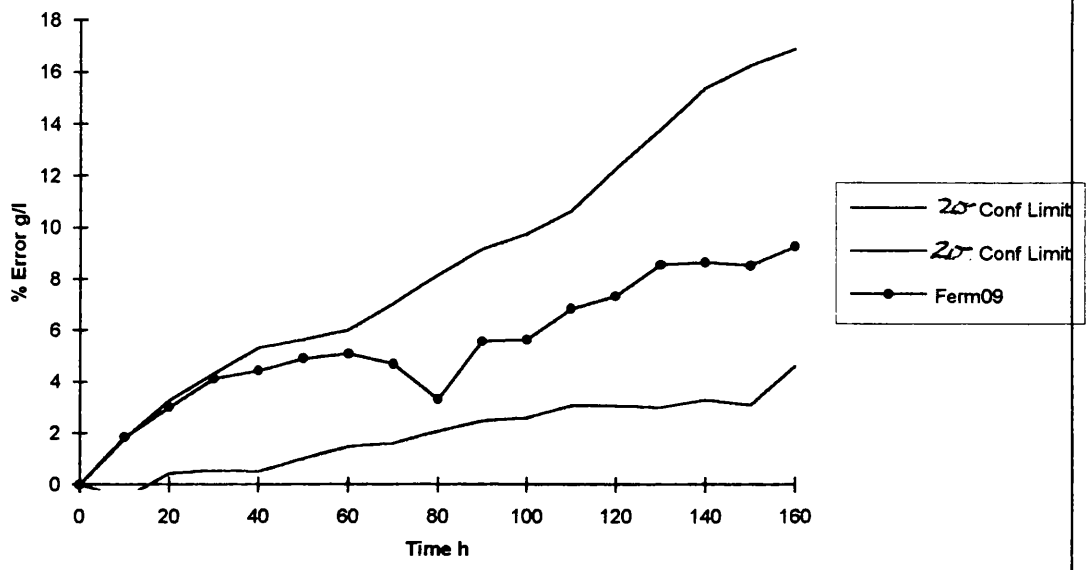
Fig 8.28 Ferm 9 Carbon Balance: Carbon Unaccounted For



This figure shows that the fault event is not detectable in the overall mass balance.

Looking at the error in the carbon balance as a % of carbon input, fig 8.29 is obtained:

Fig 8.29 Unaccounted Carbon as % of Total Carbon Inputs

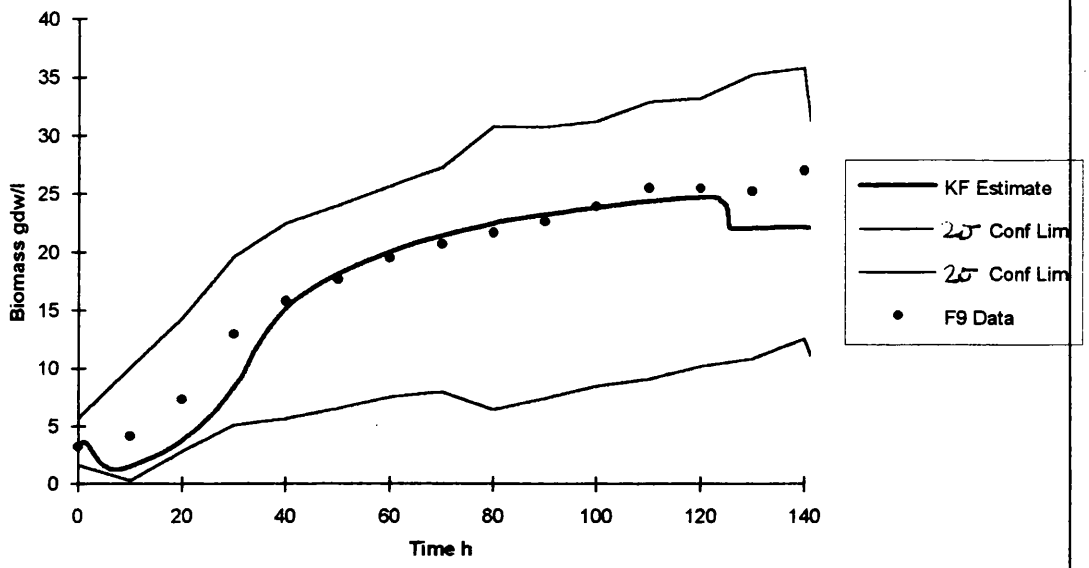


Similarly, normalising the mass balance error with respect to overall carbon input, no indication is given of a fault event at 130 hours. The reason for this is that the proportion of overall carbon lost during the feed interruption is a small fraction (~ 2%) of overall carbon inputs and outputs. This is masked by the error on the mass balance which is of the order of ±10%.

In this case, overall mass balances do not provide any warning of a fault having occurred. This is because the batch to batch variability is simply too wide, the 2σ confidence limits are simply never crossed.

It is interesting to look at Kalman filter biomass estimates vs. measured biomass.

Fig 8.30 Kalman Filter Biomass Estimate.

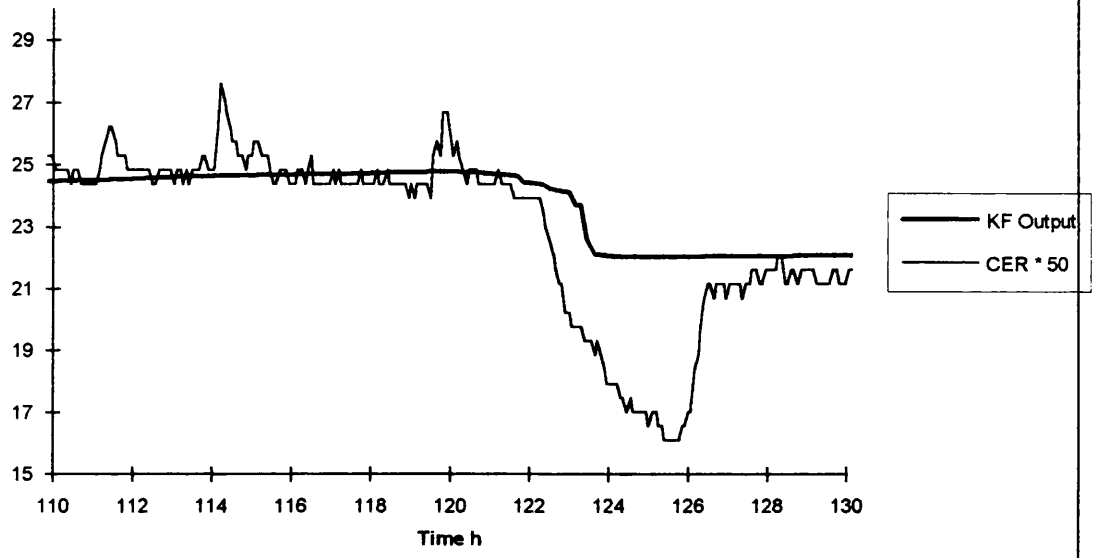


Kalman filter estimate of biomass. While the filter reacts quickly to the fault, the biomass estimate produced in this way does not fall outside the expected distribution of biomass estimates.

The filter attempts to compensate for the reduction in CER by adjusting down the estimate of biomass. It is interesting to compare this result with the previous case, in which a much shorter duration but more complete interruption in CER had little impact on filter biomass estimates.

In terms of fault detection however, the filter to be of use must provide data additional to the CER trajectory. Comparing the filter output at the point of the fault with the CER data the following graph is obtained:

Fig 8.31 Fermentation 9 Comparison Between CER and Kalman Filter Output



In this figure, the CER and Kalman filter output are superimposed, by a scaling factor of 50 being applied to the filter biomass estimate. It is striking how close the two trajectories are before and after the fault event, but that during the fault event, they become separate. This illustrates how the filter effectively produces a smoothed version of CER data.

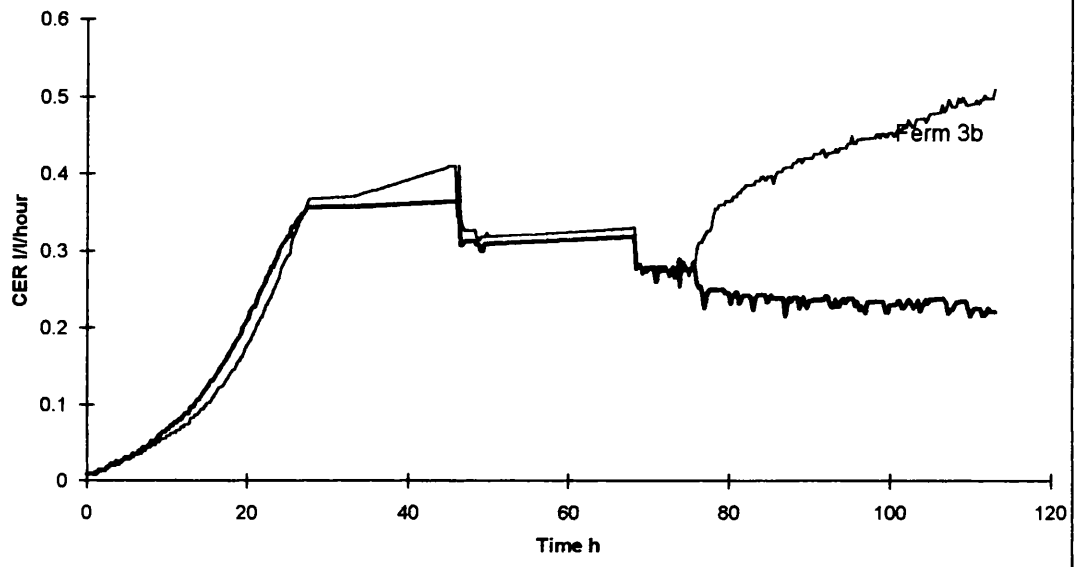
It is difficult, on the basis of this data, to justify the complexity of using a Kalman Filter in a fault detection system. In this case the CER changed by a greater amount in proportion to its total value. The filter output would not have added anything to the detection of this particular fault.

The conclusion from the examples above is that relatively simple ratios are often much more effective means of detecting faults than either raw data or complex means of tracking a fermentation such as overall mass balance. This is because the variation on

both raw data and data which is derived from a combination of information sources turns out to have an excessive variance for it to be of use in fault detection.

A second feed fault case study can be made by comparing fermentations 3b and 3c. In fermentation 3b an increase in feed rate of 50% was made at 75 h. The impact of this on carbon dioxide evolution can be seen by comparison with fermentation 3c which was identical up to the point at which the feed change was made:

Fig 8.32 Ferm 3 b and 3 c Effect Of Increase In glucose Feed On CER

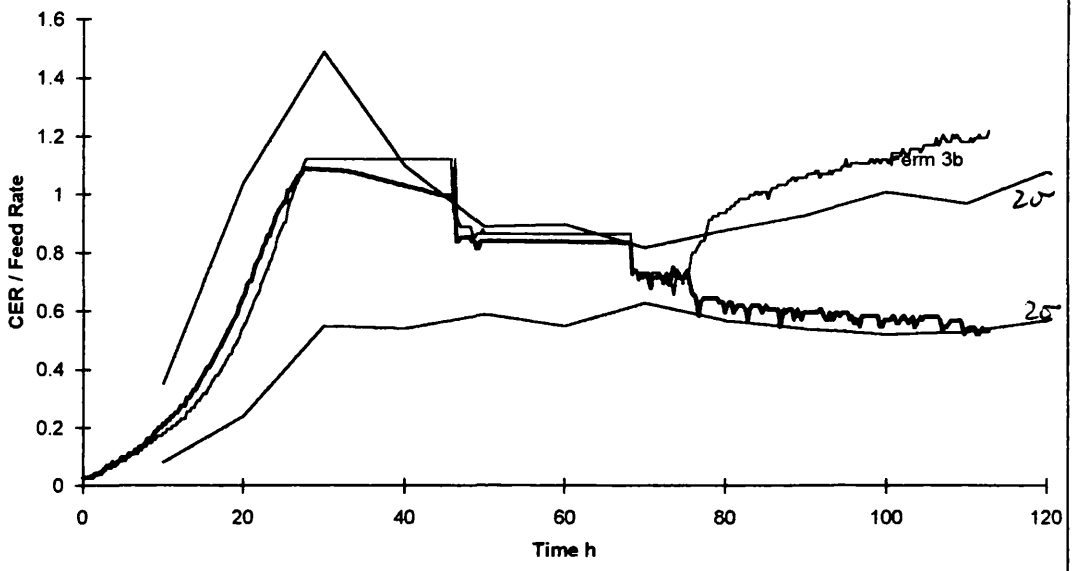


Note, between 30 and 70 hours, mass spectrometer faults ocurred, producing the unusual CER profile.

Note: the CER monitoring system in this fermentation was subject to intermittent failures. This accounts for the somewhat unusual look of the above graph between 25 and 70 h.

The addition of extra glucose was to fermentation 3b. The impact of this can be easily detected when looking at the ratio of CER to apparent glucose feed rate (i.e. glucose feed rate prior to increase in addition rate, the increase was carried out to simulate the impact of a mistake in feed addition:

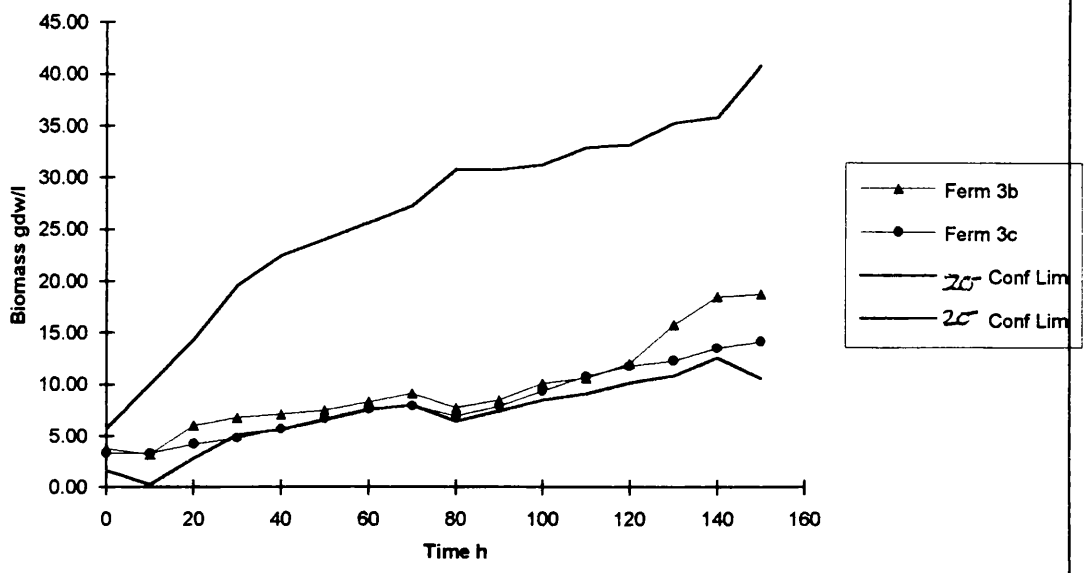
Fig 8.33: Fermentations 3b and 3c, CER/Planned Glucose Feed Rate



This figure illustrates once again the effectiveness of the CER/Glucose ratio in combination with expected trajectory limits. The advantage of using CER / Glucose ratio is that this ratio permits much tighter limits than if CER data only was used giving more rapid and more sensitive fault detection.

It is interesting to note that there was a 40 hour lag before any significant difference became apparent in the biomass concentration in these two fermentations:

Fig 8.34 Fermentations 3b and 3c Biomass



This figure shows that there was a considerable lag between adding extra glucose at 80 h, to fermentation 3b, and any divergence in measured biomass vs. the control fermentation (3c)

These figures show that biomass measurements or derived measurements such as specific growth rate would not be effective in identifying this glucose overfeed fault. In this case, because of the failure of the CER analysing system at several points in the fermentation an overall mass balance cannot be used.

Penicillin production in this fermentation was also not effectively indicative of a fault as can be seen in fig 8.34:

Fig 8.35 Fermentations 3b,3c Penicillin Titre

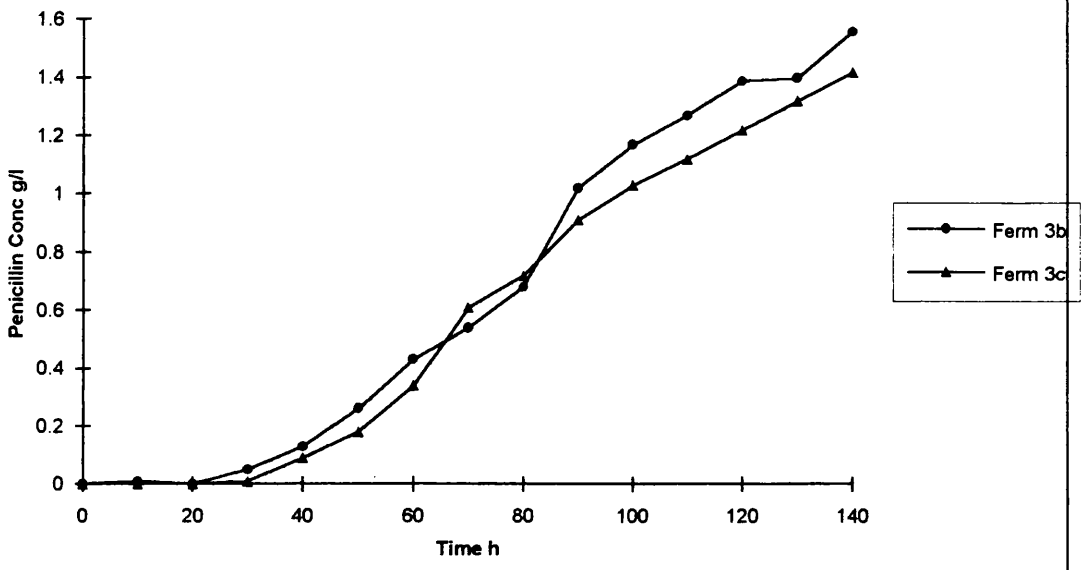


Figure shows that adding extra glucose after 80 h produced only a small extra quantity of penicillin. This extra penicillin production would have been well within the expected penicillin titre change, and so would not have show up as an anomaly.

8.3 Conclusions

The key conclusion to be drawn from these case studies is that the most useful source of information available for fault detection is that derived from past fermentations, providing it is correctly conditioned. In particular, figs 8.9, 8.17, 8.26, and 8.33 all illustrate for aeration and glucose feed faults, the utility of the ratio of carbon dioxide evolution to glucose (or more generally, carbon) feed. Direct comparison of process variables such as penicillin concentration, biomass and phenylacetic acid concentration also provided useful fault markers.

A useful confirmatory test is to determine an overall mass balance on a key element, in this case carbon. If the carbon balance reconciles to within usual error levels, assumptions being made about the rate of feed, and aeration are verified.

Another conclusion to be drawn from all these case studies is that there is very little value added to the information available from a fermentation by using complex on-line observer structures. The most practical way of developing an on-line fault detection system is to use relatively simple numerical relationships, such as the ratio of the rate of carbon dioxide being produced to glucose added.

Thus most of the benefits of an on-line fault detection system are attainable without complex algorithms being required. The simple limit test techniques can operate with data from relatively few batches. It will be remembered that the fermentations were deliberately varied to simulate the output from a research and development fermenter suite.

8.3.1 Summary of Fault Detection Methodologies.

In chapter 2 a number of fault detection approaches were discussed. It is useful now to review their usefulness in the light of the cases reviewed in this chapter:

- Statistical process control: Statistical process control is likely to be the basis of any practical fault detection scheme. The use of statistically derived acceptable operating ranges for data must provide the test by which faults are indicated. The information which can be tested against predefined limits can be basic process data such as pH or temperature or complex functions derived from on-line observers or multivariate statistical analysis.

- Analytical Redundancy: Analytical redundancy can bring additional information to the statistical process control scheme. This is illustrated by the carbon balance, in which inputs vs. outputs and accumulations provide a redundant information set. In practice this provides another parameter for the statistical process control scheme because there is always an error in the calculation of the carbon balance. This difference between measured inputs and measured outputs is not important for fault detection as long is it is consistent, the error term simply becomes another measured variable.

- Use of On-line Observers: The Kalman filter was shown to be able to provide effective estimates of biomass (chapter 7). Its usefulness as a fault detection tool

versus using carbon dioxide evolution data directly was not shown. This is because it does not provide additional data vs. its inputs.

- Multivariate Analysis: There was not sufficient data available to test Principal Component Analysis properly in the cases presented in this chapter. Initial results (graph 8.20) show that the technique is worth following up.

- Neural Networks: The use of neural networks (described in chapter 3 section 4) was not considered appropriate for the purposes of this project because of the variability of the fermentation data.

8.3.2 Summary of Useful Data and Interactions

Table 8.4 below summarises the parameters which showed significant changes on the onset of a fault:

Table 8.4: summary of fault detection information.

Fault	Indicators	Impact of fault on indicator
Contamination	Penicillin concentration	Declines
	phenylacetic acid Concentration	Rises
	Biomass	Declines
	Visual Inspection	Contaminants are readily apparent under a microscope. .
Temperature Rise	Penicillin titre	Declines
	Phenyl Acetic Acid concentration	Rises
	Ammonium concentration	Assuming ammonium hydroxide is the pH control agent, broth ammonium concentration rises.

Fault	Indicators	Impact of fault on indicator
	Dissolved Oxygen tension.	Declines or rises dependant on the relative sensitivity of the process to temperature on the solubility of oxygen vs. the effect of changes in metabolic activity. Rising temperature up to the point at which the temperature starts adversely impacting the organisms in the batch will reduce dissolved oxygen concentration.
Aeration Failure	Biomass	Declines
	Apparent CER	Increases rapidly.
	Apparent CER / Glucose feed.	Increases rapidly
Agitation Failure	CER	Declines rapidly
	CER/Glucose feed.	Declines rapidly.
	Biomass	Declines slowly.
	Dissolved oxygen	Declines.
Glucose Feed Interruption	CER	Declines rapidly
	Dissolved oxygen	Increases.
	Penicillin titre	Decreases.
	CER/Glucose Feed	Decreases.
Glucose Overfeed	CER	Increase
	CER/Glucose Feed	Increases.

The above tables the key raw data for the implementation of an expert system which can be used to identify faults.

Chapter 9: Conclusions And Recommendations For Implementation of An On-line System

The analysis of data available from fermentations reported in this work shows that the most practical method of automating fault detection is by combining information from a range of sources. The single most useful source of data is previously conducted fermentations, these provide information which can be compared with current sensor outputs.

The other source of fault detection data is on-line analytical redundancy. Where a process parameter can be inferred from more than one source, for example total carbon inputs vs. output and accumulation, the difference between the two values of the parameter can be monitored and the value of this error used to infer faults. In practice two means of identifying the same parameter will inevitably yield somewhat different results. This difference in effect becomes another parameter for the statistical process control scheme.

The key to making a statistical process control system work is to make it robust to the large batch to batch variability which can be expected in a small scale fermentation process development facility. The best means to do this is to seek to scale the data taking advantage where possible of the basic relationships which underlie any biological process. A good example of this is the use of the ratio of glucose feed to carbon dioxide evolved.

9.1 Available Information.

In the work reported in this thesis, an assumption was made that the desire to install a fault detection system would not justify the installation of extensive extra instrumentation, or the monitoring off-line of fermentation parameters at a high frequency. On this basis the data presented (in this thesis) was obtained from routinely used sensors, in particular exit gas analysis (by mass spectrometer or other gas analysis devices), pH probes, and temperature probes.

Additional information was obtained from accurately monitoring material flows into and out of the fermentation. In particular, the rate of glucose addition and the flow of gasses through the fermenter were tracked. The outcome of this monitoring effort was summarised in the data presented in chapter 5. An immediate conclusion from the analysis of this data, presented in chapter 6 is that if disparate fermentations are to be monitored on one system, data must be normalised. A good illustration of this is the comparison between figs 6.4 (specific growth rate) and 6.5 (specific growth rate / substrate feed rate)

The most complete summary of all the data obtained from a given fermentation is contained in the carbon balance. This is constructed from measurements of all material flows and analysis or estimation of broth constituents. From this it can be seen that in all the fermentations it is possible to account for about 90% of all carbon added to the fermentation. Also, it can be seen that the pattern of errors in the carbon balance is consistent. From this it can be concluded that there is a systematic error in the method used to determine carbon levels either in the inlet and outlet gas streams of the fermentation, or more likely in the fermentation broth.

There are some known reservoirs of carbon which are not being accounted for, including for example penicillin breakdown components. The most important such reservoir is the carbon locked up in insoluble components of corn steep liquor. Mou and Cooney (1983) present a detailed analysis of penicillin fermentations fed using glucose and corn steep liquor, they suggest that a proportion of these corn steep derived insolubles are fermentable. This is not of importance in fault detection as long as the rate of digestion and hence the mass balance errors are consistent. In effect the parameter measured is the completeness of the carbon balance. This can be compared with its completeness at a similar stage of other fermentations as an indication as to whether the batch is progressing satisfactorily.

9.2 The Use Of On-line Estimators To Enhance On-line Measurements.

A goal for this project was to establish a means of on-line supervision without requiring additional instrumentation or sampling of the broth. This implies that broth sample data will only be available at infrequent intervals (The project sponsors, SmithKline Beecham, typically sample their small scale fermenters at eight hour intervals). In order to enhance the information available to an on-line supervision system, it is possible to use models to estimate fermentation states at times when off-line sample data is not available.

In part the fermentation system used in this project was chosen because a model was already available. The model used is described in chapter 7. It can be seen that it predicts broth solids to ± 10% (Section 7.2). In effect this is utilising historic fermentation data as the model parameters are representations of observations from prior batches.

An effort was made to extend this using an on-line observer based on a Kalman filter. On average the filter performed marginally worse than the model. This was due to the fact that it was prone to destabilisation by unexpected fermentation events. Also the filter consistently underestimated broth solids during the closing stages of the fermentation. This is due to the problem that the model used was somewhat simple in that it did not account for non biologically active biomass, or other broth solids. (For a description of how this could be done, see Menezes et al (1994)). However this is not important for a fault detection system as long as the errors are consistent.

The advantage of using the filter is that it is responsive to unexpected process variations which result in a change in carbon dioxide evolution rate output. This is because the filter used in this work uses carbon dioxide evolution rate to generate feedback to correct estimates of broth solids and penicillin concentration. In this respect the filter captures the features of the fermentation better than the model.

9.3 The Impact of Fermentation Faults on Process Data.

Chapter 8 reports a number of case studies in which the impact of faults on the data output from fermentations are studied. On investigating the best means for identifying whether a fault has occurred, it becomes apparent that the most helpful indicators are generated by a comparison between data obtained from current fermentations with data obtained at equivalent times from past fermentations. The most effective approach is to use historic data in an on-line statistical process control system. This is the approach recommended here.

A simple and effective fault detection system can be established by selecting the appropriate fermentation parameters and comparing the value of each parameter at a given time with the distribution of parameters obtained from other broadly similar fermentations. The case studies presented in chapter 8 clearly show that the trajectory of parameters after the occurrence of a fault quickly exceed the expected range for that parameter, where that range is based on data from similar past fermentations. An excellent example is seen in figs 8.9, 8.17, 8.26 and 8.33, where the ratio of carbon dioxide evolution rate to glucose added is tracked. In each case a fault becomes apparent soon after it occurs. Of particular interest is the event shown in figure 8.33, which shows data from a fermentation in which the glucose feed rate was increased by 50% to simulate an overfeeding fault. This fault would not have been readily apparent from other fermentation data such as carbon dioxide evolution rate, broth solids concentration or penicillin concentration.

The parameter trajectory provides a simple means of fault detection in which the value of a fermentation parameter can be compared to interpolated distribution data from previous fermentations. If the value in question is outside the expected distribution, a fault is flagged.

Studies of real fault events, presented in chapter 8 show the relative merits of more highly processed data. In general, very simple data such as carbon dioxide output, or broth solids concentration are of little practical use because, as shown in chapter 6, the batch to batch variability of these data at specific times in a fermentation are very wide. In addition such data are subject to variability. This is particularly true of carbon dioxide data, as it is dependent on a number of systems functioning correctly to generate data. The use of this sort of combination means that there will always be a risk of false positive fault detection.

The simplest data transformation which can be made is to normalise data to permit better batch to batch comparison. The advantage of this can be seen in section 8.2.3 where two aeration faults are discussed. In one case the fault was readily apparent because the apparent carbon dioxide evolution rate observed was well above the expected level. In the other fault the carbon dioxide evolution rate was only marginally below the expected range. This fault was shown up much more clearly when the (apparent) carbon dioxide evolution rate / glucose addition rate was the parameter monitored.

The key conclusion from this research is that simple ratios, in particular carbon dioxide evolution / limiting nutrient feed give a sensitive indication of the occurrence of a fault. Tracking this parameter, and carrying out upper and lower limit tests will provide better fault indications than any other tests which can be applied. By definition a fault must impact on the course of a fermentation, whether by promoting or retarding metabolic activity. Carbon dioxide evolution is a direct measure of that metabolic activity, therefore any fault will be shown up by monitoring carbon dioxide evolution. The only exception to this is where biological contamination occurs, in which case carbon dioxide evolution, particularly in the steady state phase of a fermentation, is likely to be unaffected.

An alternative means of monitoring fermentations is to use a Kalman Filter to estimate broth solids. It can be seen in fig 8.11 that the aeration fault which ocurred resulted in Kalman Filter broth solids predictions outside the expected range. In addition the estimates of broth solids diverged sharply from the measured broth solids, both of these are indicative of a fault having ocurred. The filter responded more slowly to the fault event than did the carbon dioxide evolution rate data itself. This in some circumstances is an advantage, an examination of figs 5.11-5.17 shows that carbon

dioxide evolution rate data tends to be noisy and subject to occasional large fluctuations due to events such as automatic recalibration of equipment, the Kalman filter can in fact be recast into a data smoothing tool (Bozic 1979). The Kalman Filter is an excessively complex data smoothing algorithm, more direct signal processing techniques such as fairly simple non recursive digital filtering could be applied to raw carbon dioxide data.

Another available fermentation monitoring method shown in this case study is the use of a carbon balance. This is essentially an analytical redundancy, since a comparison is being made between the amount of carbon added to the fermentation, mainly due to glucose additions, and the sum of accumulation and carbon in the exit gas stream. The difference between these two quantities is presented in fig 8.10. A fault is indicated in both cases by the fact that the error observed is greater than the expected value. Clearly all of these methods indicate that a fault has occurred.

9.4 Summary of Fault Detection And Identification Methodologies.

At the start of this thesis a number of fault detection approaches were discussed. It is useful now to review their usefulness in the light of the cases reviewed in this chapter:

- Statistical process control: Any on line fault detection system is certain to utilise limit tests for example for temperature, pH, agitator RPM, airflow, dissolved oxygen and exit carbon dioxide. The upper and lower control limits for these parameters being set at appropriate fixed levels. In my view this will deliver a significant proportion of the benefit of an on line fault detection system for very little cost.

Statistical process control can be extended to utilise data from the trajectories from past fermentations for example of carbon dioxide evolution rate or broth solids. This data was shown to be an effective way of detecting contaminations, and hardware faults. However it has the problem that because the fermentations conducted differ significantly, the variance on the useful parameters is wide. A fault has to cause a very large deviation before it can be detected using these methods.

The solution to the problem of very wide control limits is to use simple knowledge about the fermentation process, for example that the carbon dioxide evolution rate is likely to be linked with the rate of glucose addition. Glucose levels are easily known outside the exponential part of the process as carbon is the growth limiting element. Using such simple ratios the usefulness of basic statistical process control techniques is much enhanced.

- Analytical Redundancy: Analytical redundancy can bring additional information to the statistical process control scheme. The example shown in this thesis is the carbon balance, which compares all carbon inputs with outputs and accumulation. This provides two ways of measuring the integrated carbon additions to the system. In practice the measured inputs are somewhat greater than the measured outputs, probably because some reservoirs of carbon, for example penicillin degradation products, are being overlooked.

The difference between measured inputs and measured outputs is not important for fault detection as long as it is consistent. This error term simply becomes another parameter on which statistical limit checks can be performed,

- Use of on line observers: The Kalman filter was shown to be able to provide effective estimates of broth solids (chapter 7) based on carbon dioxide evolution data and the parameters held in the Bajpai Reuss model. In fault detection terms it is questionable whether this estimate adds value versus using the carbon dioxide evolution data directly.

In the context of complex non-linear systems such as fed batch penicillin fermentations, it is unlikely that a sufficiently good process model can be devised such that the whiteness test or multiple hypothesis filter techniques described by Willsky (section 2.1) can be usefully deployed.

This thesis only looks at the Kalman filter in its most basic form. There is scope for further development of the filter, in particular in this research, there was a focus on developing a filter which tracked broth solids concentration. This is of interest as the level of broth solids is difficult to establish on line in real time. An alternative approach to the use of statistical observers would be to develop an observer based on something that could be monitored on line, for example dissolved oxygen concentration. Such an observer would therefore generate redundant data which could be used to detect faults.

The other area which is worth further investigation is the optimisation of the filter used in this work. This could take the form of finding a way of incorporating feedback based on low frequency time delayed data such as broth solids, or penicillin titre. Alternatively, the overall model could be extended to include a term for an additional parameter, such as dissolved oxygen which could be measured on line. Either of these modifications would lessen the dependence of

the filter on carbon dioxide evolution, and make it a more useful tool to provide comparisons with measured data.

- Neural Networks: The use of neural networks (described in chapter 3 section 4) was not attempted as part of this work as there was not time to develop sufficient data. Typical literature references (Montague 1994, Nan 1993, Srinivasin 1994) suggest that around 50 data sets are required to train an artificial neural network. It is likely that given sufficient data, a neural network based model could be used to estimate parameters in a given fermentation. This would have the benefit versus conventional modelling of being updatable with given some data from new production strains. Neural network modelling of new strains and processes will be simplified by the use of normalised data. This is because the degree of difference between existing models and the new systems will be minimised, reducing the size of the training task for a neural network.

  The disadvantage of using Neural networks is that there is no meaningful information to be obtained from the functions mapping inputs to outputs. Essentially neural network models treat the system they are observing purely as a black box.
  Neural networks would be used in a statistical process control scheme (Srinivasan 1994, Ignova 1994), parameters estimated by the network being expected to fall within a predetermined window if the process was functioning correctly.

- Multivariate Statistical Analysis: An alternative to the use of neural networks is the use of multivariate statistical analysis. Essentially, these are techniques for simplifying process data by reducing the dimensionality of a data set containing a large number of correlated data. This should be seen as a data pre-processing

step, the output from the analysis being used for example in statistical process control schemes which would be more effective than statistical process control based on single variables. This is because all of the data from a batch would be used in a single comprehensive comparison, thus relative changes within the data set are considered.

The use of Principal Component Analysis (PCA) is a promising technique which has been shown in the literature (Nomikos 1994, Kresta 1991, Skagerberg 1992)) to yield a simplified means of preparing a comprehensive statistical process control scheme. PCA can be significantly enhanced by the use of Partial Least Squares (PLS) regression (McGreggor 1991), which is essentially the PCA of both inputs and outputs, and multiple linear regression between the input and output principle components.

Using PLS techniques, it would be possible to combine the simple limit checks and the more complex time variant and ratio based SPC tests into a small set of parameters. This would avoid the need to set up specific software tests for each of potentially a very large number of inputs. These techniques are very data intensive, there was not sufficient data available to test them properly in the cases presented in this chapter. Initial results (graph 8.20) show that the technique is worth following up. Ultimately principal component analysis provides the benefit of a reduced set of parameters on which statistical process control can be applied.

- Artificial Intelligence. Both neural network and multivariate statistical analysis techniques have the weakness that they treat the fermentation process as a black box. They do not provide easily interpretable outputs, which would allow automatic fault identification.

This thesis includes both a priori and experimental studies of the chain of events following fermentation process faults. A priori information, supported by input from SKB and UCL based fermentation experts are summarised in section 1.4 and in table 1.2. Experimental data about the impact of faults is extensively analysed in chapter 8, this data is summarised in table 8.4. This is the key raw material for an expert system which could be used to identify faults.

An expert system approach to fault detection and identification would be complementary to analytical redundancy, neural network or multivariate statistical analysis methodologies. The expert system could be triggered by data falling outside one or other statistical process control limit. The expert system would provide the fault identification component of a fault detection and identification system.

9.5 Recommendations For An On-line Fault Detection System

As a result of my work I would recommend that a practical automatic fault detection system can be specified based on the measurement and analysis concepts presented in this thesis. Such a system could have the following structure.

1. Simple alarm limits. This level of the fault detection system will pick up sensor failures. A fault in the pH control hardware for example will be best detected by setting maximum and minimum divergence from set point criteria. The detection of static or very high noise levels can also be used to determine hardware faults such as the mass spectrometer fault clearly visible in fig 5.16. Tests at this level

should be high level, low level, rate of change, and also a test for electronic noise levels from sensors.

2. Statistical limit testing on key derived parameters: The control system will be informed of planned parameters including nutrient addition rates, aeration rates, and fermentation volumes. These will be used to derive scaled measurements such as carbon dioxide evolution rate / carbon addition, or specific rate parameters such as yield of biomass on glucose. These can be periodically compared to expected values for these parameters, data exceeding the expected parameter will be used to flag a fault. The case studies presented in chapter 8 show that this would pick up hardware faults such as aeration, agitation and feed system failures.

3. Use of Estimation and models: On-line observers such as a Kalman filter can be used to estimate current process parameters which are difficult to determine on-line, such as broth solids. These can be used in an overall mass balance, or compared with expected values for a given parameter. This approach may give some benefits over level 2, but is aimed at detecting essentially the same faults. These techniques would only be worth considering if the level of benefit derived from implementing level 2 was deemed to be insufficient. Given the difficulties associated with implementing this level, in most instances, hardware redundancy should be considered as an option instead.

4. Use of Off-line Measurements: Some fault types, particularly contaminations are most easily detected by measurement of broth components off line. I would recommend that the practice of sampling fermentations periodically should be maintained. Data from these off line analyses should be fed back into the fault detection system to allow it to determine whether they are outside expected

bounds given their time of sampling. The importance of off-line data for detecting contaminations is illustrated in section 8.2, where the first indicator of a contamination was a decline in penicillin titre.

5. I would recommend that an expert system based fault identification system should be developed, based on the data -presented in this thesis. This is important because such a system will provide a means of identifying the principal causes of fermentation failure. This data can be used to implement a continuous improvement program driving up the reliability of the fermentation suite.

The practical fault detection system, based on the case studies presented in chapter 8 will in practice require a smaller number of samples than would otherwise be the case to achieve fault detection in a given time. This is because faults can be detected when the data go outside predetermined bounds, but before the data become obviously wrong to process operators.

All the faults described in chapter 8 could have been detected by comparing the ratio of carbon dioxide evolution rate to glucose addition rate with its expected trajectory as derived from previous successful data, and by comparing penicillin titre with its expected trajectory. This latter test shows up contamination by β *Lactamase* producing organisms.

The logic for the effectiveness of carbon dioxide evolution rate as a fault marker is that carbon dioxide evolution rate is a direct indicator of the level of metabolic activity, that level is controlled in a fed batch fermentation by the rate of addition of the limiting nutrient, in this case carbon derived from glucose. All the fermentations presented in this thesis, despite being carried out on differing scales and with differing

nutrient feed rates, showed similar carbon dioxide evolution rate / glucose addition rate trajectories, this permits a simple detection of all hardware or nutrient addition faults.

The disadvantage of carbon dioxide evolution rate data is that it is subject to periodic fluctuations. Minimisation of the number of false alarms generated due to these phenomena is recommended by using moving average carbon dioxide evolution rate data rather than point by point data. A step in tuning a fault detection system would be to establish the size of the moving average sample which gave the optimal ratio of false alarms to missed fault events.

The project has shown that supporting evidence for the existence of faults can be obtained from using overall mass balancing. This level of fault detection is important in that it provides a consistency check for the fermentation. If the carbon balance error is within typical bounds it is possible to show that the planned nutrient addition and aeration rates were used. This will be helpful in validating data from a fermentation.

The data obtained in this work shows that the most effective fault detection procedures are simple comparisons between current and past batches. It is difficult to justify working on complex monitoring systems involving on-line parameter estimation in essence because this is not generating fresh data. For example, an on line Kalman filter could be used to estimate broth solids, essentially from carbon dioxide evolution rate data. This could be compared to an expected broth solids. However, this comparison could be made directly by comparing the ratio of carbon dioxide evolution rate / broth solids to past carbon dioxide evolution rate / broth solids ratios. The case studies on aeration faults (section 8.2.3) show that the broth

solids estimate produced by the filter in the steady state part of the fermentation contains little or no more information than the carbon dioxide evolution rate trajectory.

9.5.1 Summary

- There is sufficient information from a typical industrial fermentation to attempt to deploy automated fault detection systems.

- This data is available in a number of forms
  - On-line measurable parameters, pH, temperature, exit gas composition etc.
  - Off-line measured parameters, broth solids, penicillin titre etc.
  - Specified process conditions, nutrient feed rates, aeration rate etc.
  - Historical data from past fermentations.

- The most effective use of this data is in simple consistency and limit checks. The most effective of these is to determine whether the ratio of carbon dioxide produced to carbon added is consistent. This is because for this ratio to be consistent there must have been no hardware faults. If a fault occurs, it must interfere with metabolism, either generating more or less growth and metabolic activity than would otherwise be the case. This must be reflected in an aerobic fermentation in carbon dioxide output.

- The above conclusion can be generalised to fermentations in which another nutrient is limiting, by taking carbon dioxide output / limiting nutrient input as the parameter which is closely followed.

- Fermentation monitoring is amenable to enhancement by the use of mathematical models which combine an understanding of biological activities with data on past fermentations (encoded into model parameters).

- Non directly measurable fermentation parameters can be observed using a Kalman filter. This essentially translates information obtained from carbon dioxide evolution data into correction terms for a mathematical model of the system. Its usefulness in detecting faults in comparison with using techniques based on working directly with carbon dioxide evolution data is questionable.

- It is likely that the approaches of using principal component analysis and scaled process variables would prove complementary, in that appropriate process variable scaling will reduce the variability of the output of the principal component analysis. Ignova (1994) makes an interesting case for the complementarity of the principal component analysis and neural network approach along the same lines.

- The data typically available around a fed batch fermentation would lend itself to utilisation in a practical expert system. Such a system could identify faults detected by statistical process control techniques.

9.6 Next Steps.

The key need to take this research forward is a large number of data sets from real fermentations. Given this data better statistical process control schemes could be established. Larger quantities of data would also permit more advanced processing techniques, such as artificial neural networks or multivariate statistical analysis to be deployed.

The second area worth investigating is to add a fault identification component to the fault detection studies presented in this work. The lead option for developing such a system would be through artificial intelligence techniques, although other approaches such as the establishment of banks of fault signatures have been suggested (Willsky 1976). This work has shown that the data required for an expert system is readily available for fed batch fermentation systems.

9.6.1 Utilisation of Additional Data For Statistical Process Control

The data presented here is from a small number of fed batch fermentations produced using similar organisms and media. The project has shown how a limited set of well defined fermentation data have value. This program would be enhanced by tracking the data output and faults from a large number of industrial fermentations. This would provide a large database of fermentation trajectories from which it would be possible to generate significantly improved fermentation trajectory data. Also with a sufficient number of fault events, it would be possible to establish an appropriate level of statistical significance after which divergence from the mean of a parameter population would generate a fault. The aim would be to establish an appropriate false alarm / missed fault ratio.

In addition, more data would permit a more sophisticated approach to batch / batch comparisons. For example, the timing for sampling data for making comparisons could be adjusted for each part of the fermentation (lag phase, exponential phase, steady state phase), depending on the starting time of that phase.

It would be useful to collect fault data from a large number of real industrial fermenters, and develop Pareto charts showing the relative importance of the faults which occur. This could be used to score the various available techniques on the basis of the proportion of faults they detect and fail to detect.

9.6.2 Neural Networks / Multivariate Statistical Analysis

Another means of collating data on previous fermentations would be through the use of this data to train a neural network model, or to identify the distribution of the parameters of a partial least squares model. These models would supply an alternative means of determining whether a parameter had diverged from an expected trajectory from the limit testing proposed in this work.

The most promising new technique for mathematically capturing the output from past fermentations is the use of Principal Component Analysis described by Geladi (1986), Kresta (1991, 1994), Nomikos (1994), Ignova (1994) and others. In this technique, the potentially large number of variables which can be subject to measurement around a fermentation are decomposed into a small number of underlying statistical quantities. This reflects the reality that most of the measurable variables will be highly correlated and will describe aspects of the same process changes rather than multiple separate events. Attempts to apply partial least squares in this project were frustrated by the relatively small amount of data available. Partial least squares applied to batch processes is a relatively data intensive method, as it is necessary to establish time varying confidence limits around the statistical parameters derived from the process data.

It is likely that the approach based on identifying key ratios presented in this thesis and the PLS methodology will prove complimentary. An issue with techniques such as partial least squares and neural networks, is that there is an assumption that the base process does not change. This is unlikely to be true in a development fermenter suite due to its intended workload. Combining parameters to minimise batch to batch data variability will permit narrower confidence limits using control charts based on partial least squares parameters, thus permitting earlier identification of outlier data.

An interesting study item is to determine whether partial least squares or neural networks represent the most efficient means of achieving feature extraction from multiple fermentations which can be used to set limits for use in statistical process control. Literature claims exist for the superiority of both these techniques, Ignova et al (1994) for example claim that neural networks are capable of differentiating between differing fermentation batches using less data and a lower level of complexity than is the case for principle component analysis, as neural networks are more adept at handling time varying non linear processes. Nomikos and MacGregor (1994) on the other hand point out the benefit that principle component analysis utilises all available information from a batch system. It can handle both high and low frequency data, and utilise both the trajectory of individual variables with time and also their interactions.

9.6.3 Fault Identification Via Expert Systems

In my view, the most practical way to extend the fault detection scheme I have presented in this work to a full fault detection and identification scheme would be via the use of an on line expert system. Neural network, multivariate statistical analysis and on line observers (usually in Kalman filter form) techniques have been described

in the literature (see Srinivasan (1994 and Nokimos (1992), Ignova (1994) and Fan (1993), Wilsky 1976, Chattaway 1989 respectively). The issue with all of these techniques is that considerable numbers of faults must occur or be simulated, to provide enough fault signatures to permit statistically meaningful identification criteria to become established.

In a typical fermentation development organisation, there exists significant amounts of non numerical data, and years of fermentation operating experience which combined with an understanding of the basic science behind penicillin fermentations can be included with available quantitative data to establish a fault identification system. Section 1.4 summarises the information, based on experience and knowledge of the biology of penicillin fermentations, available from staff at UCL and SKB in Sept - Dec. 1989.

This information provides the behavioural description and causality assumptions used to build rule based expert systems. These rules are arranged in a higherarchical structure, diagnostic rules for each node in the higherarchy are generated from these assumptions. Essentially each diagnostic rule is a heuristic model of the signature of a specific fault.

Data from specific faults can also be used to build an expert system. The case studies presented in chapter 8 show the type of information available. the advantage of building this data into expert systems is that small numbers of events can be used to determine the trends of process parameters after a fault has ocurred. These trends are what go up to make up the heuristic description of a given fault. This means that a comprehensive rule based expert system can be built up quickly relative to statistical fault identification methods.

With more data than was available to this work, it would be possible to investigate more sophisticated artificial intelligence techniques, such as the use of fuzzy logic or probability based representations of uncertainties associated with the fault models implicit in the expert system rule base (Bakshi 1993). However, formulation of probabilistic or fuzzy rules describing a fed batch fermentation may take as much effort as establishing the statistical identification methods described above.

The use of normalised data and analytical redundancy would complement expert systems in that the normalised data provides clearer fault detection by the ability to use narrower limits of detection. Analytical redundancy provides additional data for rule bases to utilise.

Most rule based systems used in the chemical industry have been developed for steady state systems. While fed batch fermentations are at least quasi steady state at the end of the exponential phase, there are key parts of the fermentation which are not. A number of approaches have been proposed in the literature which simplify and classify fermentation process trajectories (Cheung and Stephanopouos 1989, Bashkai and Stephanopoulos 1993), usually by looking for points of inflection in the fermentation parameter trajectories, and comparing the times and trends between these points.

In summary, the data exists around a fed batch penicillin fermentation to implement an expert system which could be used for fault identification. An identification system is an important component of a supervisory control scheme because it provides the data for a continuous improvement program aimed at driving up fermenter reliability. The best means of executing such an expert system throws up some interesting research questions:

- What is the best way to capture non steady state features of a fermentation.
- What is the cost / benefit ratio of using probabilistic rule bases vs. heuristic rule bases to describe fault signatures.
- What are the relative accuracy of heuristic, and probabilistic rule bases vs. alternative fault identification techniques.

9.6.4 Implementation Of A Fault Detection and Identification System

The key test for the ideas presented in this thesis is to implement them in a real fermentation suite, and to determine the productivity of that facility before and after the supervisory software is implemented.

9.6.5 Application To Production Scale Fermenters.

An area of study, outside the scope of this work, is how best to apply automated fault detection to production fermenters. There are likely to be significantly bigger financial benefits in detecting faults in full scale fermentations, which may contain 200 000 l of broth, vs. research fermenters containing 20 l. In that case, the interesting question becomes identification of steps to be taken to enable the fermentation to proceed with minimum loss of productivity once a fault has occurred. The other issue with implementation on a full scale is how to handle non homogeneity issues, due for example to differences in hydrostatic head, which are not present in smaller vessels.

On a production scale the issue of fault identification becomes much more significant, because it may well be possible to adjust the fermenter to save as much of the potential batch yield as possible.

Bibliography

Adlard, M.W., Gordon, B.M., Keshavarz, T., Bucke, C., Lilly, M.D., Bull A.T., and Holt, G., (1991)
An HPLC procedure for monitoring Penicillin G fermentations.
Biotechnology Techniques, 5, 2, 121-126

Alford, J.S., (1978)
Modelling of cell concentration in complex media.
Biotechnol. Bioeng., 20, 1873-1881

Aynsley, M., Ward, A.C., and Wright, A.R., (1990)
A mathematical model for the growth of mycelial fungi in submerged culture.
Biotechnol. Bioeng. 35, 820-830

Bastin, G., and Dochain, D., (1986)
On-line estimation of microbial specific growth rates.
Automatica, 22, (6), 705-709

Bailey, J.E. and Ollis D.F., (1986)
Biochemical Engineering Fundamentals 2nd ed.
McGraw Hill

Bajpai, R.K., and Reuss, M., (1980)
A mechanistic model for penicillin production.
J. Chem. Tech. Biotechnol. 30, 332-344.

Bajpai, R.K., and Reuss, M., (1981)
Evaluation of feeding strategies in carbon regulated secondary metabolite production through mathematical modelling.
Biotechnol. Bioeng., 23, 717-738

Basseville M., (1988)
Detecting Changes in Signals and Systems A Survey.
Automatica. 24 309-326

Bakshi B.R, and G.Stephanopoulos, (1994)
Representation of process trends - IV. induction of real-time patterns from operating data for diagnosis and supervisory control.
Computers Chem. Eng. 18, 4, 303-332

Benedict, R.G., Schmidt, W.H., Coghill, R.D., and Oleson, A.P., (1945)
Penicillin III. the stability of penicillin in aqueous solution.
J. Bact., 49, 85-95

Benedict, R.G., Schmidt, W.H., and Coghill, R.D., (1946)
The stability of penicillin in aqueous solution.
J. Bact., 52, 291-292

Bellgardt, K.-H., Kuhlmann, W., Meyer, H.D., Schügerl, K., and Thoma, M., (1986)
Application of an extended Kalman filter for state estimation of a yeast fermentation.
IEE Proceedings, 133, D, 5, 226-234.

Billings S.A., (1980)
Identification of Nonlinear Systems A Survey.
IEE Proc 127 136-145

Bozic, S.M., (1979)
Digital and Kalman Filtering.
1st Edition
Edward Arnold Ltd., London

Buckland, R.C., and Fastert, H., (1982)
Analysis of fermentation exhaust gas using a mass spectrometer.
ICCAFT-3, Applications in Fermentation Technology, London:
Society of Chemical Industry

Cagney, J.W., Chittur, V.K., and Lim, H.C., (1984)
Use of filtration measurements for estimation of cellular activity in penicillin production.
Biotechnol. Bioeng. 14, 419-434

Calam, C.T., and Ismail, B.A.K., (1980)
Investigation of factors in the optimization of penicillin production.
J. Chem. Technol. and Biotechnol., 30, 249-262

Calam, C.T., and Russell, D.W., (1973)
Microbial aspects of fermentation process development.
J. Appl. Chem. Biotechnol. 23, 225-237

Carleysmith, S.W., and Fox, R.I., (1984)
Fermenter instrumentation and control.
In: Advances in Biotechnological Processes 3, Pp. 1-51

Chattaway, T., and Stephanopoulos, G.N., (1987)
A new technique for estimation of bioreactors: adaptive state observers.
Proc. One day international Workshop, Control of Biotechnical Processes, University of Newcastle, England

Chow, E.Y. and Willsky A.S., (1984)
Analytical redundancy and the design of robust failure detection systems.
IEEE Trans. Aut. Control 29, 603-615

Clarke, D.J., Calder, M.R., Carr, R.J.G., Blake-Coleman, B.C., Moody, S.C., and Collinge, T.A., (1985)
The developement and application of biosensing devices for bioreactor monitoring and control.
Biosensors 1, 213-320

Constantinides, A., Spencer, J.L., and Gaden, E.L., (1970)
Optimization of batch fermentation processes. I. Development of mathematical models for batch penicillin fermentations.
Biotech. Bioeng., 12, 803-830

Contois, D.E., (1959)
Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures.
J. gen. Microbiol. 21, 40-50

Cooney, C.L., Wang, H.Y., and Wang, D.I.C., (1977)
Computer-aided material balancing for prediction of fermentation parameters.
Biotechnol. Bioeng., 19, 55-67

Cooney, C.L., and Acevedo, F., (1977)
Theoretical conversion yields for penicillin synthesis.
Biotecnol. Bioeng., 19, 1449-1462

Cooney, C.L., (1979)
Conversion yields in penicillin production: theory vs. practice.
Process Biochemistry, 5, 31-33.

Dochain, D., and Bastin, G., (1985)
Stable adaptive algorithms for estimation and control of fermentation processes.
Proc. 1st IFAC Symp. Mod. Control Biotechnol. Process., Noordwijkerhout, The Netherlands, December 1985 Pp 37-42. A Johnson (Ed.)

Dohnal M., (1985)
Fuzzy Bioengineering Models.
Biotech Bioeng 27, .1146-1151

Dubach, A. Ch., and Markl, H., (1992)
Application of an extended Kalman filter method for monitoring high density cultivation of *Escherichia coli.*
J. Ferment. Bioeng., 73, 396-402

Dhurjati, P.S., and Leipold, R.J., (1990)
Biological modelling.
Chapter 8
Computer Control of Fermentation Processes pp 207-220
CRC Press, Florida
Ed. Daniel R Omstead

Esener, A.A., Roels, J.A., and Kossen, N.W.F., (1981)
Fed batch culture: modeling and applications in the study of microbial engineering.
Biotechnol. Bioeng., 23, 1851-1871

Esener, A.A., Roels, J.A., and Kossen, N.W.F., (1981)
On the statistical analysis of batch data.
Communications to the editor, Biotechnol. Bioeng., 23, 2391-2396.

Fan,J,Y, Nikolaou,M, White,R.E, (1993)
An Approach to Fault Diagnosis of Chemical Processes via Neural Networks.
AIChE Journal Vol 39, 82-88.

Fishman, V.M., and Biryukov, V.V., (1974)
Kinetic model of secondary metabolite production and its use in computation of optimal conditions.
Biotecnol. Bioeng. Symp., 4, 647-662

Flynn, D.S.,(1982)
Instrumentation in fermentation processes.
Proceedings of the 1st IFAC Conference on Modelling and Control of Biotechnical Processes, August, Helsinki,(A. Halme, Ed.) pp 5-12.
Pergamon Press, Oxford.

Gbewonyo, K., Jain, D., Hunt, G., Drew, S.W., and Buckland, B.C., (1989)
On-line analysis of avermectin fermentation cell growth kinetics in an industrial pilot plant.
Biotecnol. Bioeng. 34, 234-241

Hegewald, E., Wolleschensky, B., Guthke, R., and Knorre, W.A., (1981)
Instabilities of product formation in a fed-batch culture of *Penicillium chrysogenum.*
Biotechnol. Bioeng. 23, 1563-1572

Heijnen, J.J., Roels, J.A., and Stouthamer, A.H., (1979)
Application of balancing methods in modelling the penicillin fermentation.
Biotechnol. Bioeng. 26, 2175-2201.

Heijnen, J.J., and Roels, J.A., (1981)
A macroscopic model describing yield and maintenance relationships in aerobic fermentation processes.
Biotechnol. Bioeng., 23, 739-763

Heinzle E. and Dunn, I.J., (1991)
Methods and instrumentation in fermentation gas analysis.
Chapter 2 in Biotechnology 4: Measuring, Modelling, and Control, 2nd Edition, Volume editor Karl Schügerl. VCH Weinheim

Hudson S. (1988)
Qualitative Reasoning: A Demonstration of Qualitative Simulation with a Case Study in Fermentation Processes.
Thesis. Dept of Computer Science U.C.L.

Humphrey, A.E., (1987)
Algorithmic monitors for computer control of fermentations.
In Horizons of Biochemical Engineering, Ed. Aiba, S., 203-211
Oxford University Press, Oxford, U.K.

Ignovia,M., Glassey, Y., Montague, G.A., Morris, A.J. and Kiparissides, C. (1994)
Neural networks and non linear SPC
Proc. 3rd IEEE Conference in Control Applications, 2, 1271 - 1276

Isermann R. (1984)
Process Fault Detection Based On Modelling and Estimation Methods - A Survey.
Automatica, 20, 387-404

Jazwinski, A.H., (1969)
Adaptive filtering.
Automatica, 5, 475-485,

Johnson, A., (1987)
The control of fed-batch fermentation processes, a survey.
Automatica, 23, (6), 691-705

Kalman, R.E. (1960)
A new approach to linear filtering and prediction problems.
J.Basic Eng. 35-45

Kalman, R.E., and Bucy, R.S., (1961)
New results in linear filtering and prediction theory.
J. Basic Eng., 95-108

Kennedy, M.J., Thakur, M.S., Wang, D.I.C., and Stephanopoulos, G.N., (1992)
Estimating cell concentration in the presence of suspended solids: a light scatter technique.
Biotechnol. Bioeng., 40, 875-888

Kramer M.A. (1987)
Malfunction diagnosis using quantitative models with non-boolean reasoning in expert systems.
AIChE Journal 33,1,130-140

Kramer M.A. and Palowitch B.L. (1987(
A rule based approach to fault diagnosis using the signed directed graph.
AIChE Journal 33,7,1067-1078

Kresta,J.V.MacGregor, JF and ,Marlin, T.E. (1991)
Multivariate statistical monitoring of process operating performance.
C.Journ.Chem.Eng 69, 35 - 47

Kresta,J.V.,Marlin, T.E. and MacGregor, J.F. (1994)
Development of infrential process models using PLS.
Computers Chem. Eng., 7, 597 - 611.

Leigh, J.R., and Ng, M.H.,(1984)
Estimation of biomass and secondary product in batch fermentation.
Lecture notes in control and information science, 63, 456-467,

Lenz, R., Boelcke, C., Peckmann, U., and Reuss, M., (1985)
A new automatic sampling device for the determination of filtration characteristics
and the flow coupling of an HPLC to fermenters.
Proc. 1st IFAC symposium on modelling and control of fermentation processes
Noordwijkerhout, The Netherlands. Ed. A Johnson, Pergamon Oxford

Lichfield, R.J., Campbell, K.S., and Locke, A., (1979)
The application of several Kalman filters to the control of a real chemical reactor.
Trans. IChemE., 57, 113-120

Di Massimo, A.C.G., Saunders, A.J., and Montague, G.A., (1989)
Non-linear estimation and control of mycelial fermentations.
Proc. American Control Conference, Pittsburgh, June 1989, pp 1994-1999

Mehra, R.K., (1970)
On the identification of variances and adaptive Kalman filtering.
IEEE Trans. Auto. Control, AC-15, 2, 175-183

Mehra, R.K .and Peschon J. (1971)
An innovations approach to fault detection and diagnosis in dynamic systems.
Automatica 7, 637-640

Menezes, J.C. Sebastiao S.A., Lemos,J.M. and de Azevedo S.F., (1994)
Mathematical modelling of industrial pilot plant penicillin-G fed-batch fermentations.
J.Chem. Tech. Biotechnol> 61, 123-128

Molle D.T. and Himmelblau D.M., (1987)
Fault Detection in a Single Stage Evaporator via Parameter Estimation Using The Kalman Filter.
Ind.Eng.Chem.Res. 26 145-157

Montague, G.A., Morris, A.J., Wright, A.R., Aynsley, M., and Ward, A.C., (1985)
Parameter adaptive control of the fed-batch penicillin fermentation.
Proc. 1st IFAC Symposium Modelling and Control of Biotechnological Processes.
(A. Johnson ed.) Pp. 69-73

Montague, G.A., Morris, A.J., Wright, A.R., Aynsley, M., and Ward, A.C., (1986a)
Modelling and adaptive control of fed-batch penicillin fermentation.
Canadian J. Chem. Eng., 64, 567-580

Montague, G.A., Morris, A.J., Wright, A.R., Aynsley, M., and Ward, A.C., (1986b)
On-line estimation and adaptive control of penicillin fermentation.
IEE Proc.(133 Pt.D), 240-246

Montague, G.A., Morris, A.J., and Tham, M.T., (1989 a)
Adaptive inferential estimation and its application to biomass control.
Proc. 4th International Congress on Computer Applications in Fermentation technology, Modelling and Control of Bioprocesses
Cambridge, Ed. N.M.Fish, R.I.Fox, and Thornhill, N.F., 179-186

Montague, G.A., Morris, A.J., and Ward, A.C., (1989 b)
In Biotechnology and Genetic Engineering Reviews - Vol. 7
Chapter 4 Fermentation monitoring and control: a perspective.

Montague, G. and Morris J. (1994)
Neural network contributions in biotechnology.
Tibtech. 12, 312-323

Mou, D.G., and Cooney, C.L., (1983)
Growth monitoring and control through computer-aided on-line mass balancing in a fed-batch penicillin fermentation.
Biotechnol. Bioeng. 25, 225-255.

Mou, D.G., and Cooney, C.L., (1983)
Growth monitoring and control in complex medium: A case study employing fed-batch penicillin fermentation and computer-aided on-line mass balancing.
Biotechnol. Bioeng. 25, 257-269.

Nahlik, J., and Burianec, Z., (1988)
On-line parameter and state estimation of continuous cultivation by extended Kalman filter.
Appl. Microbiol. Biotechnol., 28, 128-134

Nelligan, I., and Calam, C.T., (1983)
Optimal control of penicillin production, using a mini-computer.
Biotechnol. Lett., 5, (8), 561-566

Nestaas, E., and Wang, D.I.C., (1981a)
A new sensor, the "filtration probe," for quantitative characterisation of the penicillin fermentation. I. mycelial morphology and culture activity.
Biotechnol. Bioeng., 23, Pp. 2803-2813.

Nestaas, E., Wang, D.I.C., Suzuki, H., and Evans, L.B., (1981b)
A new sensor, the "filtration probe," for quantitative characterisation of the penicillin fermentation. II. the monitor of mycelial growth.
Biotechnol. Bioeng., 23, Pp. 2815-2824.

Nestaas, E., and Wang, D.I.C., (1983 a)
Computer control of the penicillin fermentation using the filtration probe in conjunction with a structured process model.
Biotechnol. Bioeng., 25, 781-796

Nestaas, E., and Wang, D.I.C., (1983 b)
A new sensor - the "filtration probe" -for quantitative characterisation of the penicillin fermentation. III. an automatically operating probe.
Biotechnol. Bioeng., 25, Pp. 1981-1987

Nihtila, M., and Virkkunen, J., (1977)
Practical identifiability of growth and substrate consumption models.
Biotech. Bioeng., 19, 1831-1850

Nomikos P, and MacGreggor, J.F., (1994)
Monitoring Batch Processes Using Multiway Principal Component Analysis.
A.I.Ch.E.Journal, 40, 1361-1375.

Ono T. Kumamaru T. Maeda A. Sagara S. and Kumamaru K., (1987)
Influence Matrix Approach to Fault Diagnosis of Parameters in Dynamic Systems.
IEEE Trans. Ind. Elec. 34, 222-230.

Perlman, D., (1970)
The evolution of the penicillin manufacturing process.
In Chapter III, The history of penicillin production,
Chem. Eng. Symp. Series, 100, 66, 24-30

Pigott, A., (1989), Ph.D thesis, University College London

Pirt, S.J., (1965)
The maintenance energy of bacteria in growing cultures.
Proc. Roy. Soc., B163, 224-231

Roels, J.A. (1980)
Application of Macroscopic principles to microbial metabolism.
Biotech. Bioeng. 22 2457-2514

Sage, A.P., and Husa, G.W., (1969)
Algorithms for sequential adaptive estimation of prior statistics.
Proc. IEEE Symp. on Adaptive Processes, Decision and Control

San, K-Y., and Stephanopoulos, G., (1989)
Optimization of fed-batch penicillin fermentation: a case of singular optimal control with state constraints.
Biotechnol. Bioeng., 34, 72-79

Shioya, S., Shimizu, H., Ogata, M, and Takamatsu, T. (1986)
Simulation and experimental studies of the profile control of the specific growth rate in a fed-batch culture.
Proc. 1st IFAC Symp. Mod. Control Biotechnol. Process., Noordwijkerhout, The Netherlands, December 1985 Pp 79-84. A Johnson (Ed.)

Skagerberg, B, MacGregor, J.F., and Kiparissides, C. (1992)
Multivariate data analysis applied to low density polyethylene reactors
Chemometrics and Intelegent Laboratory Systems, 14, 341-356,

Srinivasan, A, and Batur, C, (1994)
Hopfield/ART-1 Neural Network Based Fault Detection and Isolation.
IEE Trans. on Neural Networks,Vol5, No6, 890-899

Staniskis, J., and Simutis, R., (1986)
A measuring system for biotechnical processes based on discrete methods of estimation.
Biotecnol. Bioeng., 28, 362-371

Stephanopoulos, G., and San, K-Y., (1981)
State estimation for computer control of biochemical reactors.
Adv. Biotechnol. 1, 399-406

Stephanopoulos, G., and San, K-Y., (1984)
Studies on on-line bioreactor identification. I. Theory.
Biotechnol. Bioeng. 26, 1176-1188.

Stephanopoulos, G., (1986)
Application of macroscopic balances and bioenergetics of growth to on-line identification and biological reactors.
Biochemical Eng. IV NY Acad. Sci. Vol.469, 332-349

Stephanopoulos, G. and Park, S., (1991)
Bioreactor state estimation.
Chapter 7 in Biotechnology 4: Measuring, Modelling, and Control,
2nd Edition, Volume editor Karl Schügerl
VCH Weinheim Pp 225-249

Stone, K.M, Roche, F.W., and Thornhill, N.F., (1992)
Dry weight measurement of microbial biomass and measurement variability analysis.
Biotechnology Techniques, 6, 207-212

Stone, K.M., Roche, F.W., Thornhill N.F., and Fish, N.M., (1992)
Using off-line measurements in an on-line estimator of biomass concentration for a penicillin fermentation, and its effects on the quality of the estimates.
Presented at the 5th International Conference on Computer applications in fermentation technology (ICCAFT5), Keystone, Colorado, U.S.A., March 1992.
Conference proceedings published by Pergamon Press, Oxford

Swartz, J.R. and Cooney, C.L., (1979)
Indirect fermentation measurement as a basis for control.
Biotechnol. Bioeng. Symp., 9, 95-101

Swinarski, R., Lesniewski, A., Dewshi, M.A.M., Ng, M.H., and Leigh, J.R., (1982)
Progress towards estimation of biomass in a batch fermentation process.
Proc. IFAC Workshop, Modelling and Control of Biotechnical Processes, Helsinki, Finland. Pp.231-242

Tarbuck, L.A., Ng, M.H., Leigh, J.R., and Tampion, J., (1985)
Estimation of the progress of Streptomyces clavuligerus fermentations for improved on-line control of antibiotic production.
Proc. 1st IFAC Symp. Modelling and Control Biotechnol. Processes, Holland (Pergammon Press)

Tarbuck, L.A., Ng, M.H., Tampion, J., and Leigh, J.R., (1986)
Development of strategies for on-line estimation of biomass and secondary product formation in growth-limited batch fermentations.
IEE Proceedings, Vol.133, Pt. D, No.5.

Tayeb, Y.J. and Lim, H.C., (1986)
Optimal glucose feed rates for fed-batch penicillin fermentation: an efficient algorithm and computational results.
Biochemical Eng. IV NY Acad. Sci. Vol.469, 382-403

Thomas CR. Paul G.C. and Kent C.A., (1993)
Structured model for the penicillin fermentation.
6th Eur. Cong. Biotechnol. abst TU003.

Tiller V., Meyerhoff J. Sziele D. Schugerl K. and Bellgardt K., (1994)
Segregated mathematical model for the fed batch cultivation of a high-producing strain of *Penicillium chrysogenum*.
Journ. Biotechnol. 34, 119-131

Thompson, M.L. and Kramer M.A., (1994)
Modelling chemical processes using prior knowledge and neural networks.
AIChE Journal 40, 1328-1340.

Thornhill, N.F., Manela, M., Campbell, J.A., and Stone, K.M., (1994)
Two methods of selecting smoothing splines applied to fermentation data.
AIChE Journal 40,5,716-725

Thornhill, N.F, and Royce, P.N.C. (1994)
Modelling Fermentaers for Control
Chapter in Measurement and Control in Bioprocessing, Elsevier Applied Science.

Tsuchiya, H.M., Fredrickson, A.G., and Aris, R., (1966)
Dynamics of microbial cell populations.
Adv. Chem. Eng., 6, 125. - requires the page numbers

Van der Heijden, R.T.J.M., Hellinga, C., Luyben K.Ch.A.M., and Honderd, G., (1989)
State estimators (observers) for on-line estimation of non-measurable process variables.
Trends Biotechnol. 7, 205-209

Wang, H.Y., Cooney, C.L., and Wang, D.C., (1979)
Computer control of bakers yeast production.
Biotecnol. Bioeng., 21, 975-995

Wang, N.S., and Stephanopoulos, G.N., (1983)
Application of macroscopic balances to the identification of gross measurement errors.
Biotecnol. Bioeng., 25, 2177-2208

Wang, N.S., and Stephanopoulos, G.N., (1984)
A new approach to bioprocess identification and modelling.
Biotechnol. Bioeng. Symp., 14, 635-656

Wang, N.S., and Stephanopoulos, G.N.,(1986)
Computer applications to fermentation processes.
CRC Critical Reviews in Biotechnology, vol.2, issue 1.

Willsky A.S., (1976)

A Survey of Design Methods for Failure Detection in Dynamic Systems.

Automatica. 12,601-607.

Wise, B.M., and Gallagher, N.B., (1996)

The process chemometrics approach to process monitoring and fault detection.

J.Proc.Cont, 6, 329-348.

Wishner, R.P., Tabaczynski, J.A., and Athans, M., (1969)

A comparison of three non-linear filters.

Automatica, 5, 487-496

Young, P.C., (1981)

Parameter estimation for continuous-time models-a survey.

Automatica 17, 23-39

Yu, J., Zhang, S., and Fang, B., (1987)

Retrospect on practical and engineering problems in fermentation industry.

In Horizons of Biochemical Engineering, Ed. Aiba, S., 303-316

Oxford University Press, Oxford, U.K.

Zabriskie, D.W., and Humphrey, A.E., (1978)

Real-time estimation of aerobic batch fermentation biomass by component balancing.

AIChE J., 24, (1), 138-146

Zabriskie, D.W., (1985)

Data analysis.

Comprehensive Biotechnology Vol.2

Moo-Young, M., Ed.,

Pergammon Press, New York, Pp.175-190