DESIGN OF DYNAMIC EXPERIMENTS FOR THE OPTIMIZATION OF BATCH FERMENTATION PROCESSES: THE CASE OF PENICILLIN

A thesis

submitted by

Fredy Afif

In partial fulfillment of the requirements

for the degree of

Master of Science

in

Chemical Engineering

TUFTS UNIVERSITY

Date: May 2011

ADVISER: Prof. Christos Georgakis

Abstract

This work aims to investigate the use of a systematic methodology to optimize the operating conditions of batch fermentation processes, presented by Georgakis (Georgakis, 2009). This methodology is a novel model-free technique, as opposed to model-based optimization techniques. The methodology consists of designing certain experiments, obtaining a response surface model, and optimizing the response surface model. This methodology has been called Design of Dynamic Experiments (Georgakis, 2009) and is an extension of the well-studied and widely used classical Design of Experiments technique (Montgomery, 2005)(Box & Draper, 2007). The main difference is that the DoDE methodology allows for the design of experiments in which at least one of the decision variables is a time-varying one. This allows us to explore several substrate feeding strategies, and to determine the optimal one. Two different designs of interest to fed-batch fermentations are studied. One in which the substrate is fed in a systematic fashion throughout the fermentation (centralized), and one in which the fermentation is split into two segments, corresponding to the growth phase and the production phase (decentralized). The results of the two designs are compared. The production of penicillin is used as a case study for this methodology, using a well-established and widely studied model by Bajpai and Reuss (Bajpai and Reuss, 1980). Centralized designs are found to be more efficient than decentralized designs. Using four dynamic subfactors gives the optimal penicillin production when using Centralized designs. Using three dynamic subfactors gives the optimal penicillin production when using Decentralized design. However, the number of experiments required for each optimal design is the same.

Centralized Design has the advantage of only needing to add one extra factor to test the significance of adding one more dynamic subfactor, whereas the Decentralized Design needs two extra factors to test the significance of adding one more dynamic subfactor to each phase.

Acknowledgements

First of all, I would like to thank my advisor, Professor Christos Georgakis, for his guidance, patience, and support throughout my time at Tufts. I would also like to thank my other thesis committee members, Professor Blaine Pfeifer, and Professor Christoph Borgers.

I would like to thank my research group members, Andrew Fiordalis, Fenia Makrydaki, Athina Pantelidou, and Sarah Lucier, for their support, help and feedback. I would also like to thank my colleague and friend, JiHae Sohn, for being there for me during the tough times.

I especially would like to thank my family, who have always been there for me, and supported me unconditionally. I would not have been able to get this far without them.

Tufts University

May 2011

Table of Contents

List of Tables

List of Figures

DESIGN OF DYNAMIC EXPERIMENTS FOR THE OPTIMIZATION OF BATCH FERMENTATION PROCESSES: THE CASE OF PENICILLIN

1 Introduction

Batch processes are usually optimized using traditional optimization techniques, which involve having a mathematical model representing the process. Experiments are performed in order to estimate parameters in the model. However, these methods can sometimes be quite complex and finding such a model can be very difficult and time-consuming. In many cases, a first-principles model is not available. One method around this is the use of the Design of Experiments technique (Montgomery, 2005)(Box & Draper, 2007). However, this methodology does not allow for the design of experiments in which at least one of the decision variables is a time-varying one. Therefore, this model-free optimization technique is not effective in processes that have an important time-varying factor. This is the case for crystallizations and fed-batch processes, including fermentations.

The Design of Dynamic Experiments methodology (Georgakis, 2009) employs many of the Design of Experiments ideas and methods, but has the added advantage of designing experiments in which at least one decision variable is a time-varying one.

In the case of fermentations, this methodology allows us to explore several substrate feeding strategies, and to determine the optimal one. In this work, this

methodology will be applied to the penicillin fermentation process, with the main goal of developing a framework for the application of the Design of Dynamic Experiments to a general batch fermentation process. Fermentation processes were considered as a suitable and attractive application for this methodology, since they usually operate in fed-batch mode, with the substrate being fed in a time-varying fashion. Also, fermentation processes have industrially significance, especially in the pharmaceutical industry, where they are used to produce antibiotics and other medications.

2 Background

2.1 Model-based Optimization

Traditional optimization techniques involve having a fundamental mechanistic model that describes the process. Some experiments are usually run to estimate the parameters in the model. Several different techniques are available for model-based optimizations. Fermentation processes have been optimized by several of these techniques. The penicillin fermentation optimization problem has been solved analytically(Lim, Tayeb, Modak, & Bonte, 1986), by successive quadratic programming and orthogonal collocation (Biegler, 1984), as well as by other methods (Cuthrell & Biegler, 1989) (Chiou & Wang, 1999). These methods all require a mechanistic model, and can be quite difficult to solve.

2.2 Design of Experiments

The Design of Experiments methodology is a well-established methodology (Box & Draper, 2007) (Montgomery, 2005). It is a systematic way of designing experiments in an efficient manner, such that these experiments provide the researcher with as much information as possible with a limited number of experiments. Design of Experiments enables the measurement of the interaction effect between different factors. In the one-factor-at-a-time approach, each factor

is varied individually, and thus any interaction between the factors cannot be estimated. The optimum operating conditions of the process may very well be different than the optimum obtained by this approach. For example, if temperature and pressure are the two factors being evaluated, and experiments are performed in which only one factor is varied at a time, the optimum temperature observed at a specific pressure may not be as optimal at a different pressure. This indicates an interaction effect between temperature and pressure, and the one-factor-at-a-time approach does not allow for such effects to be accounted for. However, Design of Experiments does allow for this by designing experiments in which multiple factors are varied together to evaluate such interaction effects. Design of Experiments then allows the development of a statistical model that relates the response variable to the factors being varied. This model can then be used for prediction of the response variable at different operating conditions within the design space tested. This response surface model (RSM) can be optimized by plotting the response surface over the range of the design space, and locating the global maximum or minimum, or by performing a nonlinear optimization numerically.

Design of Experiments allows for the optimization of a process without knowledge of the underlying process or the use of a mechanistic model describing the process. Also, optimization of such response surface models is much simpler and quicker than model-based optimization techniques.

There are several ways to design the experiments. Factorial designs are generally used when two or more factors are being evaluated. Such designs evaluate all possible combinations of the levels of the factors. Fractional factorial designs can also be used where only a fraction of the factorial design runs are used. Such designs do not usually allow for the inclusion of quadratic terms in the model. This is a potential problem in the case that the response exhibits curvature. There are designs available that overcome this obstacle. These are called response surface designs. Some examples of such designs are 'Central Composite Design', 'Box-Behnken Design', and computer-generated designs such as D-optimal designs. It is a good idea to start with a factorial design initially. By including center points in the design, i.e. all factor levels are zero, curvature can be detected, if it is present. If curvature is found to exist, then the factorial design can be augmented by some axial runs, which then represents a Central Composite Design (Montgomery, 2005).

3 Methodology

3.1 Design of Dynamic Experiments

The Classical Design of Experiments discussed above, only allows for factors to be set at constant values throughout the experiment. These will be referred to as static factors. However, many industrially significant processes involve factors that might need to be varied with time, such as cooling profiles in crystallizations, feeding profiles in fed-batch reactions, heating or cooling temperature curves etc. Applying the classical Design of Experiments technique to such cases where the optimal operating conditions are ones that change with time will not yield a very good optimum.

The Design of Dynamic Experiments technique (Georgakis, 2009) is a novel way of optimizing certain processes without the need for a first-principles model. It provides a systematic way of designing experiments that have one or more dynamic input variables. The classical Design of Experiments technique also provides a way to optimize processes without having a fundamental model. However, it is limited to designing and optimizing processes in which all the input functions are constant throughout the experiment. The Design of Dynamic Experiments technique utilizes many of the concepts of the classical Design of

Experiments. The experiments are designed in much the same manner, with the main difference being the ability to design experiments with time-varying factors.

The first step in performing a Design of Dynamic Experiments is to identify the dynamic decision variable. In the case of fermentations, the substrate-feeding rate is a key time-varying function. Next, a functional basis must be defined. An appropriate choice would be the shifted Legendre polynomials. These polynomials form a complete linearly independent set and can be used as a functional basis. This is an orthogonal basis. Time must be non-dimensionalized by dividing the time by the batch time.

The dynamic decision variable, u, can be expanded in terms of the functional basis (Georgakis, 2009):

$$u(\tau) = u_0 + \Delta U \sum_{i=1}^{N} a_i \phi_i(\tau)$$

$$u_0 = \left(u_{\max} + u_{\min}\right)/2$$

$$\Delta U = diag\left(\left(u_{i\max} - u_{i\min}\right)/2\right)$$

where a_i are the expansion coefficients. They shall be referred to as dynamic subfactors from hereon, as they are factors that will be used to design the dynamic variable, and sets of experiments can be designed with these coefficients as factors using the classical Design of Experiments technique. In this work, a design of experiments will be performed involving static as well as dynamic factors.

In this case, the basis-functions, ϕ_i, are the shifted Legendre polynomials, P_i. Here are the first five shifted Legendre polynomials (Horng & Chou, 1986):

$$P_0(\tau) = 1$$

$$P_1(\tau) = 2\tau - 1$$

$$P_2(\tau) = 6\tau^2 - 6\tau + 1$$

$$P_3(\tau) = 20\tau^3 - 30\tau^2 + 12\tau - 1$$

$$P_4(\tau) = 70\tau^4 - 140\tau^3 + 90\tau^2 - 20\tau + 1$$

The dynamic decision variable can then be calculated at each level of dynamic subfactors. By designing such experiments, the dynamic subfactors can be optimized using the same methods as the classical Design of Experiments. The optimization will yield optimal values for the dynamic subfactors, which can then be used to calculate the optimal dynamic decision variable.

It must be noted that some constraints must be applied on the dynamic subfactors. These constraints are imposed so that the coded dynamic factor stays within the range -1 to 1, for all time $0 \le \tau \le 1$, where τ is the non-dimensionalized time:

$$-1 \le \sum_{i=1}^{N} a_i \phi_i(\tau) \le 1$$

In the case where the basis-functions, ϕ_i, are chosen to be the shifted Legendre polynomials, P_i:

$$-1 \le \sum_{i=1}^{N} a_i P_i(\tau) \le 1$$

$$-1 \le a_0 P_0(\tau) + a_1 P_1(\tau) + a_2 P_2(\tau) + ... \le 1$$

For the inequality to be satisfied for all τ, it is sufficient to apply this constraint at τ=0, τ=1, and at any local minima and maxima located in between. In order to find such stationary points, the following equation must be solved:

$$\frac{d\left(a_0 P_0(\tau) + a_1 P_1(\tau) + a_2 P_2(\tau) + ...\right)}{dt} = 0$$

Then the constraints for the dynamic subfactors are:

$$-1 \le a_0 P_0(0) + a_1 P_1(0) + a_2 P_2(0) + ... \le 1$$

$$-1 \le a_0 P_0(1) + a_1 P_1(1) + a_2 P_2(1) + ... \le 1$$

$$-1 \le a_0 P_0(\tau^*) + a_1 P_1(\tau^*) + a_2 P_2(\tau^*) + ... \le 1$$

,where τ^* represents the non-dimensionalized time where the stationary point occurs.

When dealing with third order polynomials and higher order polynomials, solving the above equation for τ can become quite complex, and an explicit solution may not exist. Substituting the resulting expression for τ, if it exists, into the inequality constraint, could lead to highly nonlinear and complex constraint.

Many of the software used for performing Design of Experiments, such as Design Expert®, do not permit the inclusion of nonlinear constraints. One would then have to approximate the nonlinear constraint with a set of linear constraints, where more problems may arise, such as the violation of the constraint if the nonlinear constraint is convex.

One way of avoiding this is by using the following logic:

$$-1 \le a_0 P_0(\tau) + a_1 P_1(\tau) + a_2 P_2(\tau) + ... \le 1$$

$$\max\left(a_0 P_0(\tau) + a_1 P_1(\tau) + a_2 P_2(\tau) + ...\right) \le \max\left(a_0 P_0(\tau)\right) + \max\left(a_1 P_1(\tau)\right) + \max\left(a_2 P_2(\tau)\right)...$$

Therefore, the upper constraint can be satisfied by the following:

$$\max\left(a_0 P_0(\tau)\right) + \max\left(a_1 P_1(\tau)\right) + \max\left(a_2 P_2(\tau)\right)... \le 1$$

If $a_i>0$, for all i, then the above expression becomes:

$$a_0 + a_1 + a_2 + ... \le 1$$

If $a_i>0$, for all i, except $a_2<0$, then the expression becomes:

$$a_0 + a_1 - a_2 + ... \le 1$$

If all possible cases are considered, by the same logic, the upper constraint becomes:

$$\pm a_0 \pm a_1 \pm a_2 \pm ... \leq 1$$

Applying the same reasoning for the lower constraint yields:

$$1 \leq \pm a_0 \pm a_1 \pm a_2 \pm ...$$

Combining the above two inequalities results in the following constraint:

$$-1 \leq a_0 \pm a_1 \pm a_2 \pm ... \leq 1$$

This is a set of simple linear inequalities that can be easily input into any Design of Experiments software that handles constraints. This set of inequalities is sufficient, but not necessary, to ensure the constraint on the coded dynamic factor is met. The drawback to using this set of inequalities is that some portion of the original design space is not explored. However, we find that this is not a major drawback, and using this method is much simpler and more efficient than the previously mentioned method involving the location of the stationary points.

3.2 Optimal Designs

There are several different designs of experiments that can be used to fit models of a certain form. Different designs are better suited for different situations. Of particular interest for fermentations are computer-generated designs. Such designs include D-Optimal designs. These designs are generated by computer software

such as Design Expert®. Computer-generated designs are useful in cases where the design space is not a regular geometric shape. In the case of fermentations, there is usually a constraint on the volume at the end of the biosynthesis process. Applying this constraint will yield an irregular design region. Thus, standard designs such as factorial designs and central composite designs will not fit into such regions. Another advantage of computer-generated designs is that they often employ fewer experiments than factorial and central composite designs. This is particularly beneficial for processes that are costly and/or time-consuming. It is generally preferred to use the standard designs when the process is relatively inexpensive and not time-consuming, and when the design space is of regular geometric shape. This is because the standard designs are general and flexible (Montgomery, 2005).

Computer-generated designs are usually designed in a way to optimize certain criteria. One such design is the D-optimal design that minimizes the determinant $\left|(\mathbf{X}'\mathbf{X})^{-1}\right|$, where $\mathbf{X}$ is the design matrix. This could be interpreted to mean that the set of experiments will be designed such that the volume of the confidence ellipsoid related to the vector of regression coefficients is minimized. In other words, it improves the accuracy of the estimates of the regression coefficients. This is the most widely used computer-generated design. Other designs include A-optimal designs, which minimize the sum of the variances of the regression coefficients. G-optimal designs are generally used when prediction accuracy is

very important. Such designs minimize the maximum variance of the predictions (Montgomery, 2005).

Since fermentations frequently encounter reactor volume constraints, and since the application of the Design of Dynamic Experiments technique requires the application of constraints on the dynamic subfactors as previously explained, the design region is usually irregular in shape. Thus, optimal designs should be used for the application of DoDE to fermentation processes. Since optimization is the main goal of this study, and D-optimal designs are the most widely used computer-generated designs, D-optimal designs will be used in designing the experiments. Design-Expert® is a valuable tool in generating such designs, and allows the input of constraints on the design space.

3.3 Centralized and Decentralized Designs

Many factors affect fermentation processes. These include, but are not limited to, temperature, pH, substrate concentration being fed, initial volume in the reactor, initial cell concentration in the fermentor, batch time, initial substrate concentration, and the feed rate of the substrate. For this study, experiments will be designed with static and dynamic factors. The following factors will be focused on in this study: initial biomass concentration, initial substrate concentration, and substrate feeding profile.

Penicillin is an antibiotic that is usually a non-growth associated product. It is produced in the stationary phase as opposed to during the growth phase when the cells are growing. Glucose is the substrate fed for penicillin fermentations. Typical feeding strategies used for fermentations is feeding the substrate at a high rate initially for the cells to grow in the growth phase, then feed it at a lower rate that is enough for the cell maintenance, when the metabolite is produced. This is known as the production phase.

In this work, two main designs are studied: Decentralized and Centralized.

3.3.1 Decentralized Design

In the first design, which shall be called the "Decentralized Design", the general approach of feeding the substrate in two different phases is used. Two feeding profiles are designed: one for the growth phase, and one for the production phase.

For this design, two static factors are used: initial substrate concentration and initial biomass concentration. The feed switch time from growth to metabolite production will also be varied, but will not be an independent variable, as explained later on.

Since this design has two different substrate feeding profiles, two sets of dynamic subfactors are needed. We will use an incremental approach in testing how many dynamic subfactors, i.e. how many Legendre polynomials show an effect on the optimum production of the penicillin metabolite.

The next design will have the first two shifted Legendre polynomials, i.e two dynamic subfactors, which represent linear feeding profiles. As long as the dynamic subfactors corresponding to the higher order Legendre polynomials in the design are significant, then we will increase the number of dynamic factors until the added higher order polynomials has no improvement on the optimum metabolite production.

The dynamic feeding profiles must be designed about some base case. For the decentralized design, the base case used will be a step function. The base feeding profile during the growth phase is constant and relatively high, whereas that during the production phase is constants but relatively low. This is a customary base case to use, and is chosen because of the knowledge of secondary metabolite production.

3.3.2 Centralized Design

The second design used, is the so-called “Centralized Design”, in which the entire batch time is considered as one phase with only one continuous feeding profile.

For this design, only two static factors are used: initial substrate concentration and initial biomass concentration.

Since this design has only one substrate feeding profile, only one set of dynamic subfactors is needed. As in the case of the decentralized design, an incremental approach will be used in testing how many dynamic subfactors have an influence on the production of penicillin.

This design will follow the same procedure of designing experiments first with only one dynamic subfactor, then increasing until the addition of more dynamic subfactors shows no improvement on the optimum metabolite production.

Two different base cases are used for this design; a flat base case, and a linearly decreasing base case. The motivation behind using a decreasing linear base case comes from knowledge of fermentation processes. We know that secondary metabolites are usually produced after a substantial lag phase, so it is logical to feed more substrate initially for the cells to grow, then feed just enough substrate later so that the cells can maintain themselves, but at the same time, not inhibiting product formation. Thus a decreasing linear profile would be a smarter base case to design the profiles about. A flat base case may not yield the best optimal feeding profile, as it may not allow the exploration of as much of the optimal operating area as the linear one.

Therefore, we begin with a flat base case, with two dynamic subfactors, and obtain the optimum substrate feeding profile in that design space. This feeding profile is expected to be a linearly decreasing profile. This optimum substrate feeding profile then serves as the linearly decreasing base case for the next design. This is done so that we design our experiments closer to the expected optimum region.

3.4 Simulation and error

Since this is a preliminary study in testing the DoDE methodology for its effectiveness in the design of dynamic experiments for the optimization of fermentation processes, the process is simulated using the mechanistic model derived by Bajpai and Reuss (Bajpai & Reuss, 1980), rather than actually performing the experiments. The model is a set of ordinary differential equations that can be solved using the 'ODE45' function in MATLAB® when the system is non-stiff, and the 'ODE15s' function when the system is stiff. 'ODE45' uses a Runge-Kutta method to numerically solve the differential equations. 'ODE15s' is a multistep solver based on the numerical differentiation formulas. The system is solved for each designed 'experiment'.

The performance index, which is the objective function that we wish to maximize, will be the amount of secondary metabolite present at the end of the reaction, as this is the final product. This quantity is calculated by solving the system of ordinary differential equations for each experiment.

A random error term must be added to this simulated performance index to make it more realistic and to account for uncertainty in the parameters as well as measurement errors associated with measuring the amount of product formed at the end of the experiment. The error term added can be chosen as a random number from the normal distribution with mean of zero and standard deviation to be determine according to the amount of error desired. In this study, a 5% error on

the amount of product is used. Adding the error is critical for simulating reality accurately, also enabling the estimation of the lack of fit statistic for the response surface model. If the error added is very low, then the sum of squares due to pure error could be underestimated, which leads to the overestimation of the sum of squares associated with the lack of fit. An important rule-of-thumb is to make sure the residuals of the response surface model are normally distributed. This is an indication of a good fit. Several other statistics and plots should be looked at together to determine if the model is satisfactory.

3.5 Response Surface

After simulating the amount of product formed at the end of the reaction, i.e. the performance index, and adding the random error to them, the performance index is regressed with the values of the coded factors. When using an optimal design, or any other response surface design, quadratic terms can be estimated, in addition to main effects and interaction terms. Therefore, the data will be fit using a quadratic model. The terms in the model then need to be evaluated in order to determine which of them is significant, and which of them are not.

It is very cumbersome to evaluate all possible regressions, so a stepwise regression is performed. There are three types of stepwise regressions: forward selection, backward elimination, and mixed stepwise regression (Montgomery, Peck, & Vining, 2006).

Forward selection entails adding regressors to the model one at a time. The model initially has no regressors. The regressor with the largest F statistic, i.e. the most significant term, is added first, assuming the F statistic is larger than a preselected F statistic. The next regressor that will be added is the one that has the largest partial F statistic with the new model. This algorithm terminates when no more terms have partial F statistics larger than the preselected threshold.

Backward elimination works in quite the opposite way. As the name indicates, the model begins with the inclusion of all possible regressors, and works its way back. The term with the smallest partial F statistic in the model is eliminated as long as it is below a minimum F value. The partial F statistics of the remaining terms are recalculated after the first term is removed. The term with the smallest recalculated F value is removed also on condition that it is below the minimum F value chosen. This process is repeated until no more terms remaining in the model have partial F values below the minimum F value. This technique is favored over forward selection.

The stepwise regression is a combination of forward selection and backward elimination. This algorithm works by applying a forward selection, and after each step, a backward elimination step is performed to check if any of the terms that had been added during the forward selection step have become insignificant after the addition of a new regressor. This method requires two limiting F values: one for entering terms, and the other for eliminating them. Frequently, the two F

values are set to the same value. Some people also like to choose the F value for entering to be greater than the F value for exiting. This strategy makes it more difficult for a term to be added than to be removed (Montgomery, Peck, & Vining, 2006).

A mixed stepwise regression, also known simply as stepwise regression, will be used in developing the response surface models. A cutoff p-value of 0.05 for the addition and removal of terms will be used for all stepwise regressions in this case study.

3.6 Optimization

After obtaining a satisfactory response surface model, the next step is to optimize this RSM. One method of doing this would be to plot the response surface and observe where the maximum occurs. This is an effective method when there are only two factors. However, the designs used produce response surfaces that depend on more than two factors. This makes it difficult to observe the optimum graphically. Also, this is a constrained nonlinear optimization, since we have constraints on the dynamic factors as well as a constraint on the final volume. An easier and more accurate method is to use nonlinear programming. The function *fmincon* in MATLAB®, which employs a trust-region reflective algorithm, can be used. This function finds the minimum of constrained nonlinear multivariable

functions. Thus, the sign of the response surface model must be changed, since the objective is to maximize the performance index (amount of metabolite formed). By doing this, the maximum performance index will be the absolute value of the minimum of the negative function. This function will give the optimum operating conditions that will maximize the product formation. The amount of product can then be predicted using the response surface model at the optimum operating conditions. A good test of the model would be to simulate the optimum operating conditions calculated from the optimization, and compare it to that predicted by the model to assess the prediction power of the response surface model.

3.7 Volume Constraint

Fermentations will generally have physical constraints on the system. The volume at the end of the reaction cannot exceed the volume of the reactor; otherwise the fermentor will overflow. As a result, a constraint must be imposed on the final volume. The details of this volume constraint will be different for the Decentralized and Centralized designs.

3.7.1 Centralized design

3.7.1.1 Flat Base Case

$$\frac{dV}{dt} = u(t)$$

Time is non-dimensionalized:

$$\frac{dV}{d\tau} = t_b \cdot u(\tau)$$

Separating and integrating yields:

$$V_f - V_0 = t_b \cdot \int_0^1 \left[u_m + \Delta u \left(a_0 P_0(\tau) + a_1 P_1(\tau) + ... \right) \right] d\tau$$

Integrating $P_1(\tau)$, $P_2(\tau)$, and higher order Legendre polynomials from 0 to 1 yields zero, since they are orthogonal with $P_0(\tau)=1$, assuming that Δu is not a function of time.

$$\int_0^1 P_0(\tau) P_i(\tau) d\tau = \int_0^1 P_i(\tau) d\tau = 0$$
$$i = 1,2,...$$

So we are left with:

$$V_f = V_0 + t_b \cdot \left(u_m + \Delta u \cdot a_0 \right)$$

Therefore, the inequality constraint on the final volume is:

$$V_0 + t_b \cdot \left(u_m + \Delta u \cdot a_0 \right) \le V_{reactor}$$

This constraint applies regardless of how many Legendre polynomials are used in designing the feeding profiles.

3.7.1.2 Linear Base Case

If we start with a linearly decreasing profile as the base case, and we choose Δu in such a way that it is larger at τ=0 than at τ=1, then we can represent u_m and Δu with the following expressions:

$$u_m(\tau) = A + B\tau$$

$$\Delta u(\tau) = C + D\tau$$

Δu is designed to be larger at τ=0 than at τ=1, since the base case is linearly decreasing, and thus u(τ) has a larger value at τ=0 than at τ=1.

Therefore,

$$V_f - V_0 = t_b \cdot \int_0^1 \left[u_m(\tau) + \Delta u(\tau)\left(a_0 P_0(\tau) + a_1 P_1(\tau) + ...\right)\right] d\tau$$

$$\frac{V_f - V_0}{t_b} = \int_0^1 \left[A + B\tau + (C + D\tau)\left(a_0 P_0(\tau) + a_1 P_1(\tau) + ...\right)\right] d\tau$$

$$= A + \frac{B}{2} + C \cdot a_0 + D \cdot \int_0^1 \left[\tau\left(a_0 P_0(\tau) + a_1 P_1(\tau) + ...\right)\right] d\tau$$

We can rewrite τ in terms of $P_0(\tau)$ and $P_1(\tau)$:

$$\tau = \frac{P_0(\tau) + P_1(\tau)}{2}$$

$$\Rightarrow \frac{\Delta V}{t_b} = A + \frac{B}{2} + C \cdot a_0 + \frac{D}{2} \cdot a_0 + \frac{D}{2} \cdot \int_0^1 \left[P_1(\tau) \cdot P_1(\tau)\right] d\tau$$

$$\Rightarrow \frac{\Delta V}{t_b} = A + \frac{B}{2} + C \cdot a_0 + \frac{D}{2} \cdot a_0 + \frac{D}{6} \cdot a_1$$

$$\Rightarrow V_f = V_0 + t_b \cdot \left(A + \frac{B}{2} + C \cdot a_0 + \frac{D}{2} \cdot a_0 + \frac{D}{6} \cdot a_1 \right)$$

Therefore, the inequality constraint on the final volume is:

$$\Rightarrow V_0 + t_b \cdot \left(A + \frac{B}{2} + \left(C + \frac{D}{2} \right) \cdot a_0 + \frac{D}{6} \cdot a_1 \right) \leq V_{reactor}$$

This constraint applies regardless of how many Legendre polynomials are used in designing the feeding profiles.

3.7.2 Decentralized Design

Assuming that Δu is not a function of time:

$$\Delta V = \Delta V_1 + \Delta V_2$$

$$\Delta V_1 = t_f \cdot \int_0^1 \left[u_{m1} + \Delta u_1 \left(a_{10} P_0(\tau) + a_{11} P_1(\tau) ... \right) \right] d\tau$$

$$\Delta V_1 = t_f \left(u_{m1} + \Delta u_1 a_{10} \right)$$

$$\Delta V_2 = \left(t_b - t_f \right) \cdot \left(u_{m2} + \Delta u_2 a_{20} \right)$$

$$V_f = V_0 + t_f \left(u_{m1} + \Delta u_1 a_{10} \right) + \left(t_b - t_f \right) \cdot \left(u_{m2} + \Delta u_2 a_{20} \right)$$

Therefore, the inequality constraint on the final volume is:

$$V_0 + t_f \left(u_{m1} + \Delta u_1 a_{10} \right) + \left(t_b - t_f \right) \cdot \left(u_{m2} + \Delta u_2 a_{20} \right) \leq V_{reactor}$$

This constraint also applies independent of the number Legendre polynomials are used in designing the feeding profiles.

Looking at the above inequality, we notice that it is a nonlinear inequality if the feed switch time is considered as a factor. However, most software used for designing experiments using the method of Design of Experiments, such as Design-Expert® and JMP®, do not allow the input of nonlinear constraints. Only linear constraints are permitted. Hence, a solution must be found in order to be able to do design the experiments in a way such that these nonlinear constraints are not violated. One way of doing this is by approximating the nonlinear constraints with a set of linearized constraints. Another solution is using MATLAB® to generate the design of experiments, since MATLAB® allows the input of nonlinear constraints when specifying the candidate set for the design.

4 Simulation Results for Penicillin

The above-proposed methodology is applied to a penicillin biosynthesis process as a case study. Due to time constraints and for the purposes of this study, it was decided to simulate the production of penicillin, rather than perform the experiments.

4.1 Model used for simulation

The model developed by Bajpai and Reuss (Bajpai & Reuss, 1980) was used as the mechanistic model for simulation purposes. This model was developed based on enzyme kinetics models. The model consists of a set of ordinary differential equations. The model is presented below:

$$\frac{dV}{dt} = \frac{U}{S_F}$$

$$\frac{dX}{dt} = \mu \cdot X - \frac{X}{V} \cdot \frac{dV}{dt}$$

$$\frac{dS}{dt} = -\mu \cdot \frac{X}{Y_{X/S}} - \rho \cdot \frac{X}{Y_{P/S}} - m_X \cdot X + \left(S_F - S\right) \cdot \frac{1}{V} \cdot \frac{dV}{dt}$$

$$\frac{dP}{dt} = \rho \cdot X - K \cdot P - \frac{P}{V} \cdot \frac{dV}{dt}$$

$$\frac{dC_L}{dt} = -\mu \cdot \frac{X}{Y_{X/O}} - \rho \cdot \frac{X}{Y_{P/O}} - m_O \cdot X + K_L a \cdot \left(C_L^{\,*} - C_L\right) + \frac{C_L}{V} \cdot \frac{dV}{dt}$$

where

$$\mu = \mu_x \cdot \left(\frac{S}{K_X X + S}\right) \cdot \left(\frac{C_L}{K_{OX} X + C_L}\right)$$

$$\rho = \mu_P \cdot \left(\frac{S}{K_P + S\left(1 + S/K_I\right)}\right) \cdot \left(\frac{C_L^{\,P}}{K_{OP} X + C_L^{\,P}}\right)$$

V represents volume, S is the substrate concentration, X is the biomass concentration, P is the product (penicillin) concentration, and C_L is the dissolved oxygen concentration.

The parameter values used for this simulation are the same ones used by Birol et al. (Birol, Undey, & Cinar, 2002), who created a modular simulation package for penicillin production.

Table 4.1.1: Penicillin Model Parameters

Parameter	Value	Unit
μ_X	0.092	h^{-1}
μ_P	0.005	h^{-1}
K_X	0.15	gS/gX
K_P	0.0002	g/l
K_I	0.1	g/l
K	0.04	h^{-1}
$Y_{X/S}$	0.45	gX/gS
$Y_{P/S}$	0.9	gP/gS
C_L^*	1.16	g/l
K_{OX}	0.02	gO/gX
K_{OP}	5.0×10^{-4}	gO^p/gX
K_La	125	h^{-1}
m_O	0.467	h^{-1}
$Y_{X/O}$	0.04	gX/gO
$Y_{P/O}$	0.2	gP/gO
p	2.74	N/A
m_X	0.014	h^{-1}.gS/gX

The fermentor was assumed to be a 10 liter reactor for the purposes of this study. This is a reasonable reactor size for a laboratory scale reactor.

The following initial conditions were used:

Table 4.1.2: Initial Conditions for Penicillin Model Simulation

Initial Condition	Value	Unit
S	1.5	gS/l
V	7	l
P	0	gP/l
C_L	1.16	gO/l

Some other variables need to be specified; the batch time and the substrate concentration of the feed:

Table 4.1.3: Variables Used for Penicillin Model Simulation

Variable	Value	Unit
t_b	150	h
S_f	600	gS/l

Error needs to be introduced into the simulated measurements. There is no exact amount of error that should be added. A reasonable error amount should be estimated to account for uncertainty in the parameters, as well as measurement errors. For the purpose of this study, a proportional error rate of 5% will be added. This is done by multiplying the simulated measurement by a random number from the normal distribution. In order to do this, the standard deviation required to sample random numbers that have a 95% chance of falling between -0.05 and +0.05 should be calculated.

$$P(-0.05 < X < 0.05) = 0.95$$

Normalizing:

$$P\left(\frac{-0.05}{\sigma} \le Z \le \frac{0.05}{\sigma}\right) = 0.95$$

$$2 * P\left(Z \le \frac{0.05}{\sigma}\right) - 1 = 0.95$$

$$P\left(Z \le \frac{0.05}{\sigma}\right) = 0.975$$

$$\sigma = \frac{0.05}{1.96} = 0.0255$$

Therefore, a normal distribution with a mean of zero and standard deviation of 0.0255 is used to generate the error fraction in the simulated measurements.

$$PI = PI * \left(1 + N\left(0, 0.0255\right)\right),$$

where PI is the performance index.

4.2 Centralized Design of Dynamic Experiments

First, we shall look at the centralized design where the substrate feeding profile is treated as one continuous function over the entire batch time.

4.2.1 Case 1

The first case we will look at is a design with a flat base case i.e. flat center point, with only two dynamic subfactors. Initially, we did not know where the optimum

region lies. So, we chose to start with the following region in order to locate the optimum operating region:

$$u_m = 0.008 l/h$$

$$\Delta u = 0.002 l/h$$

Since we are only working with two dynamic subfactors for this case, we can constrain the dynamic subfactors such that the flowrate is positive at τ=0 and τ=1 without having to deal with nonlinear inequalities. These inequalities are found to be:

$$a_0 + a_1 > -\frac{u_m}{\Delta u}$$

$$a_0 - a_1 > -\frac{u_m}{\Delta u}$$

The volume constraint derived earlier for a centralized design with a flat base case is:

$$V_0 + t_b \cdot \left(u_m + \Delta u \cdot a_0\right) \leq V_{reactor}$$

Substituting the values of V_0, $V_{reactor}$, t_b, u_m, and Δu into the above three inequalities yields:

$$a_0 + a_1 > -4$$

$$a_0 - a_1 > -4$$

$$a_0 \leq 6$$

Applying these three inequalities results in the following allowable design space (the triangle formed in the middle):

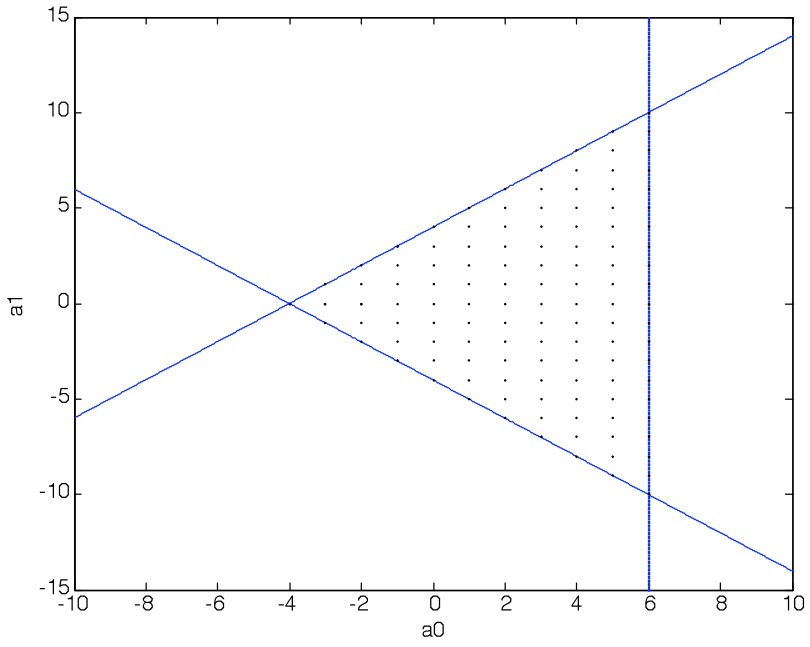

Figure 4.2.1: Design Space for Case 1

Even though we can operate in the entire space shown above, we cannot do one design that covers the entire region, as the model will not fit well at all. So we started off with a small design space at the left corner of the above design space (a_0=-4, a_1=0). The experimental results were simulated, a response surface model was developed, then optimized. The optimum was found to occur at the maximum value of a_0, and a certain negative value of a_1. This indicates that a new design

should be made, with a larger value of u_m, the average flow rate of the base case. The simulation and optimization procedure was repeated, and the optimum was again found to occur at the maximum value of a_0, and a certain negative value of a_1. For each design, the optimum always occurs at the maximum value of a_0, so it is clear that we should operate at the highest possible average flowrate value, which is constrained by the volume constraint. The largest corresponding average flowrate (u_m) is 0.020 l/h. This also indicates the optimum penicillin production will occur when the reactor is completely filled, i.e. the final volume should equal to 10 liters. Details of these calculations are provided in the Appendix.

So by doing a design with u_m=0.016, and Δu=0.004, and allowing a_0 to range from -1 to 1, and a_1 from -1 to 1, we obtain an optimum at a_0=-0.35, a_1=-1. In this case, no inequality constraints on the dynamic subfactors are necessary. This optimum occurs at the edge of the design space with respect to a_1.

This indicates that there is a possibility that a lower value of a_1 could result in a higher optimum. So next, we repeat this design, except now let's allow a_1 to go from -2 to -1, in order to allow steeper decreasing profiles to be examined.

The following design of experiments for this case was generated using Design Expert®. Also included in this table is the simulated performance index (mass of penicillin produced) for each of the runs.

Table 4.2.1: D-Optimal Design of Experiments for Case 1 with Simulated Results

Run	a_0	a_1	PI (g)	PI with error
1	-1.00	-2.00	2.69	2.67
2	-1.00	-2.00	2.69	2.63
3	-0.33	-2.00	4.68	4.64
4	1.00	-2.00	11.78	11.64
5	1.00	-2.00	11.78	11.66
6	0.00	-1.75	9.08	8.96
7	-1.00	-1.50	5.97	6.16
8	-1.00	-1.50	5.97	6.06
9	1.00	-1.50	20.57	20.60
10	-0.50	-1.25	15.51	14.93
11	0.50	-1.25	20.15	19.32
12	-1.00	-1.00	15.96	15.16
13	-1.00	-1.00	15.96	17.02
14	0.00	-1.00	18.39	18.84
15	1.00	-1.00	12.61	12.69
16	1.00	-1.00	12.61	12.29

After performing a stepwise regression, the following acceptable model in terms of coded factors was obtained with an R^2_{adj} value of 0.9889 and a lack of fit statistic value of 0.3695:

$$PI = 14.29 + 7.19a_0 + 9.30a_1 + -3.12a_0a_1 - 1.17{a_0}^2 - 2.32{a_1}^2 - 2.95{a_0}^2a_1 - 5.86a_0{a_1}^2 - 2.79{a_1}^3$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.2.2: Parameter Estimates with Confidence Intervals for Case 1

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	14.29	13.38	15.20
a_0	7.19	6.34	8.04
a_1	9.30	6.98	11.62
a_0a_1	-3.12	-3.64	-2.61
a_0^2	-1.17	-2.09	-0.26
a_1^2	-2.32	-3.22	-1.42
$a_0^2 a_1$	-2.95	-4.18	-1.72
$a_0a_1^2$	-5.86	-6.86	-4.85
a_1^3	-2.79	-5.42	-0.16

The model was optimized using the fmincon function in MATLAB®. The optimum was found to occur at a_0=+1, and a_1=-1.41, which is at an intermediate value of a_1, therefore, this is the optimum linear profile that can be obtained for this case, without violating the volume constraint.

The optimum performance index obtained from the response surface is 20.61± 1.31 grams. The performance index obtained by simulating the optimum profile was found to be 20.80 grams, which is within the confidence interval of the predicted optimum.

4.2.2 Case 2

Now that we have a good idea of the optimum region, Case 2 will be designed using the optimum profile from Case 1 as the linear base case. For this case, u_m,

as well as Δu, will be functions of dimensionless time, $\tau=t/t_b$. This can be accomplished by choosing $u_m(\tau)$ to be:

$$u_m(\tau) = 0.0256 - 0.0112 \cdot \tau$$

, since at τ=0, u=0.0256, and at τ=1, u=0.0112, according to the optimum profile obtained from Case 1.

Δu was chosen in a way such that it is larger at τ=0 than at τ=1:

$$\Delta u(\tau) = 0.007 - 0.003 \cdot \tau$$

This results in design space that looks like this:

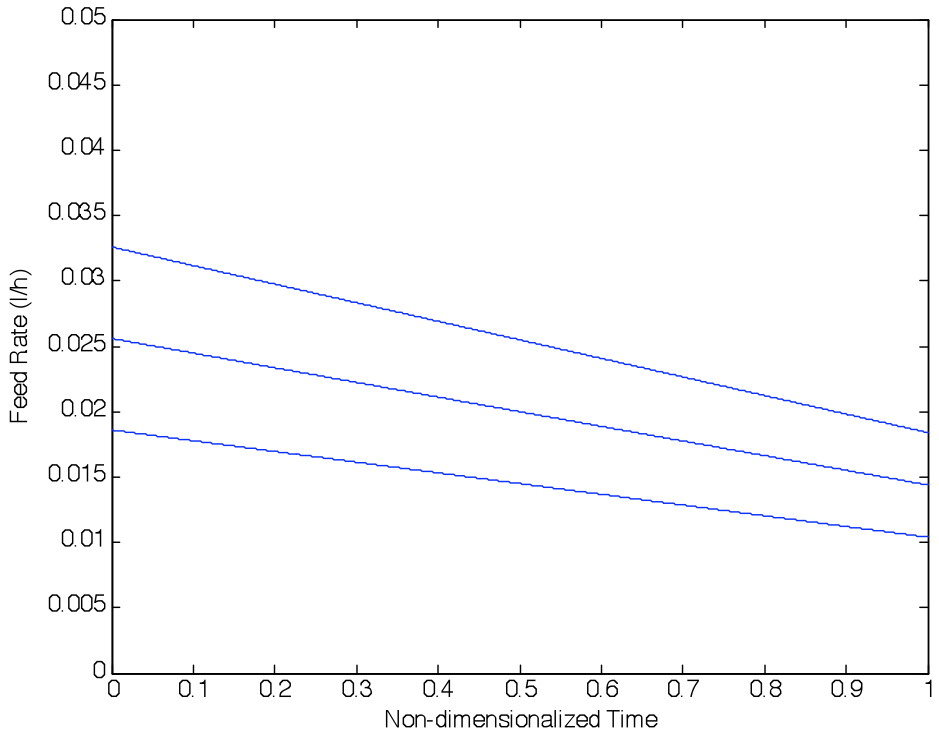

Figure 4.2.2: Feed Rate Range for Case 2

From the analysis of Case 1, it is clear that filling the reactor completely by the end of the batch time will give the maximum production. Therefore, instead of constraining the final volume to be less than the size of the reactor, we shall set the final volume equal to the reactor volume. Referring back to section 3.7.1.2, the constraint becomes:

$$V_0 + t_b \cdot \left(A + \frac{B}{2} + \left(C + \frac{D}{2} \right) \cdot a_0 + \frac{D}{6} \cdot a_1 \right) = V_{reactor}$$

In this case, A=0.0256, B=-0.0112, C=0.007, and D=-0.003.

Substituting these values as well as those of V_0 and $V_{reactor}$:

$$7 + 150 \cdot \left(0.0256 - \frac{0.0112}{2} + \left(0.007 - \frac{0.003}{2} \right) \cdot a_0 - \frac{0.003}{6} \cdot a_1 \right) = 10$$

$$\Rightarrow a_1 = 11 \cdot a_0$$

This reduces the number of independent factors by 1, and reduces the number of experiments needed.

For this case, we shall evaluate the first 3 shifted Legendre polynomials. Also, for the rest of the cases, we will also evaluate the initial biomass concentration and the initial substrate concentration. Examining the first 3 shifted Legendre polynomials requires including only 2 dynamic subfactors in the design of experiments. Therefore, the four independent factors that will be included in the design will be a_0, a_2, X_0, and S_0.

The constraints on the dynamic subfactors are:

$$-1 \le a_0 + a_1 \pm a_2 \le 1$$

$$-1 \le a_0 - a_1 \pm a_2 \le 1$$

Substituting the expression for a_1 in terms of a_0 simplifies the constraints to:

$$-1 \le 12 \cdot a_0 + a_2 \le 1$$

$$-1 \le 12 \cdot a_0 - a_2 \le 1$$

$$-1 \le -10 \cdot a_0 + a_2 \le 1$$

$$-1 \le -10 \cdot a_0 - a_2 \le 1$$

The range of values tested for X_0 and S_0 are:

$$X_0 = 0.10 \pm 0.05$$

$$S_0 = 0.75 \pm 0.75$$

A D-optimal design was generated using Design Expert®, with the above constraints on the dynamic subfactors. The designed dynamic experiments were subsequently simulated in MATLAB®. The designed experiments along with the simulated penicillin production are presented below:

Table 4.2.3: D-Optimal Design for Case 2 with Simulated Results

Run	a_0	a_2	S_0	X_0	PI	PI with error
1	-0.08	0.00	-1.00	-1.00	2.79	2.83
2	0.08	0.00	-1.00	-1.00	7.91	8.28
3	0.08	0.00	-1.00	-1.00	7.91	7.46
4	0.00	1.00	-1.00	-1.00	10.67	10.91
5	0.00	-1.00	0.00	-1.00	6.96	7.02
6	-0.08	0.00	1.00	-1.00	2.51	2.43
7	-0.08	0.00	1.00	-1.00	2.51	2.48
8	0.08	0.00	1.00	-1.00	8.06	8.13
9	0.08	0.00	1.00	-1.00	8.06	8.79
10	0.00	0.50	0.00	-0.05	14.11	15.11
11	0.00	0.00	-1.00	0.00	20.79	20.08
12	-0.04	-0.50	0.00	0.00	7.47	8.05
13	0.00	-1.00	1.00	0.00	7.93	8.08
14	0.00	1.00	1.00	0.00	10.17	10.15
15	0.00	-1.00	-1.00	1.00	8.49	8.65
16	-0.08	0.00	-1.00	1.00	6.69	6.66
17	0.00	1.00	-1.00	1.00	9.56	9.52
18	0.00	-1.00	0.00	1.00	8.39	8.70
19	-0.08	0.00	0.00	1.00	6.59	6.83
20	0.08	0.00	0.00	1.00	7.38	7.64
21	0.08	0.00	0.00	1.00	7.38	7.50
22	-0.04	-0.50	1.00	1.00	8.02	7.77
23	-0.04	-0.50	1.00	1.00	8.02	8.17
24	0.04	-0.50	1.00	1.00	16.77	17.47
25	0.00	1.00	1.00	1.00	9.75	9.87

After performing a stepwise regression, the following acceptable model was obtained with an R^2_{adj} value of 0.9594 and a lack of fit statistic value of 0.0237:

$$PI = 18.68 + 19.16a_0 + 0.97a_2 + 0.57X_0 - 213.65a_0a_2 - 13.68a_0X_0 - 1802.32a_0^2 - 9.77a_2^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.2.4: Parameter Estimates with Confidence Intervals for Case 2

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	18.68	17.53	19.82
a_0	19.16	12.73	25.59
a_2	0.97	0.39	1.54
X_0	0.57	0.17	0.97
a_0a_2	-213.65	-260.90	-166.40
a_0X_0	-13.68	-20.05	-7.31
a_0^2	-1802.32	-1989.46	-1615.17
a_2^2	-9.77	-11.13	-8.42

As can be noticed from the above model, S_0 has not been included, since it is statistically insignificant. The initial substrate concentration has no effect on the penicillin production over the range examined. This is expected since the substrate feed rate can be adjusted accordingly depending on how much the initial substrate concentration is. As a result, for the rest of the study, the initial substrate concentration shall be set at a fixed value of 1.5 g/l, and will not be varied as a factor in the experiments.

This model was then optimized by performing a nonlinear constrained optimization in MATLAB®. The optimum factor levels found were

$$a_0 = -0.004 \quad a_1 = -0.044 \quad a_2 = 0.0933 \quad X_0 = 1.00$$

The corresponding substrate feeding profile, along with the simulated substrate concentration, biomass concentration, and product concentration behaviors with time are shown below:

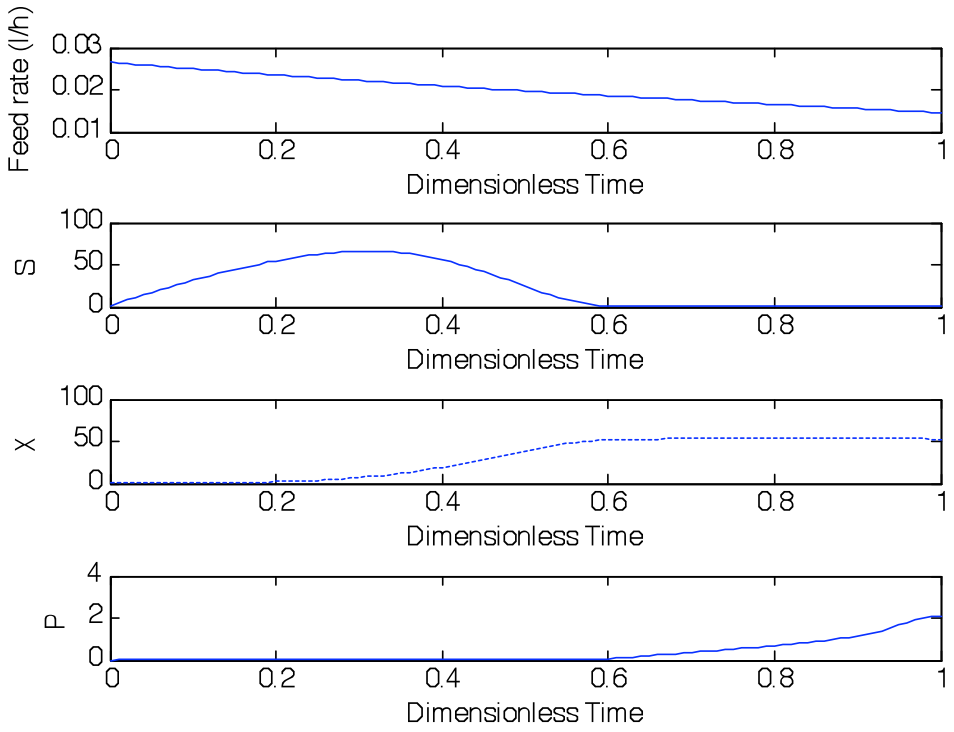

Figure 4.2.3: Optimum Profile for Case 2

The optimum performance index obtained from the response surface model was 19.28±1.17 grams. The performance index obtained by simulating the optimum profile was found to be 20.26 grams, which is within the confidence interval of the predicted optimum.

4.2.3 Case 3

The next step is to repeat Case 2, except now we will include a third dynamic subfactor in our design corresponding to the fourth shifted Legendre polynomial. Also, the initial substrate concentration will no longer be included as a factor, and will be set to a fixed value of 1.5 g/l.

The volume constraint remains the same as that of Case 2, since it applies regardless of how many polynomials are considered.

$$\Rightarrow a_1 = 11 \cdot a_0$$

For this case, we will have four independent factors that will be included in the design: a_0, a_2, a_3, and X_0.

We now have 8 constraints on the dynamic subfactors:

$$-1 \le 12 \cdot a_0 + a_2 \pm a_3 \le 1$$

$$-1 \le 12 \cdot a_0 - a_2 \pm a_3 \le 1$$

$$-1 \le -10 \cdot a_0 + a_2 \pm a_3 \le 1$$

$$-1 \le -10 \cdot a_0 - a_2 \pm a_3 \le 1$$

The range of values tested for X_0 remains:

$$X_0 = 0.10 \pm 0.05$$

A D-optimal design was generated using Design Expert®, with the above constraints on the dynamic subfactors. The designed dynamic experiments were subsequently simulated in MATLAB®. The designed experiments along with the simulated penicillin production are presented below:

Table 4.2.5: D-Optimal Design for Case 3 with Simulated Results

Run	a_0	a_2	a_3	X_0	PI	PI with error
1	0.00	0.00	-1.00	-1.00	8.47	8.54
2	0.00	-0.50	-0.50	-1.00	8.04	7.88
3	0.00	0.50	-0.50	-1.00	17.87	18.49
4	0.04	0.50	0.00	-1.00	9.17	8.77
5	0.04	0.50	0.00	-1.00	9.17	9.15
6	-0.08	0.00	0.00	-1.00	2.51	2.50
7	-0.08	0.00	0.00	-1.00	2.51	2.53
8	0.00	-1.00	0.00	-1.00	6.88	6.93
9	0.00	0.50	0.50	-1.00	14.20	13.88
10	0.00	0.00	0.50	-0.05	30.97	30.95
11	0.00	0.00	-1.00	0.00	9.01	8.97
12	-0.03	-0.33	-0.33	0.00	8.49	8.63
13	0.08	0.00	0.00	0.00	7.65	7.87
14	0.08	0.00	0.00	0.00	7.65	7.87
15	0.00	1.00	0.00	0.00	10.17	9.94

16	0.00	-1.00	0.00	0.00	7.93	7.95
17	0.00	0.00	1.00	0.00	27.96	27.10
18	0.00	0.00	-1.00	1.00	9.18	8.92
19	0.04	0.00	-0.50	1.00	14.21	14.21
20	0.04	0.00	-0.50	1.00	14.21	14.77
21	0.00	1.00	0.00	1.00	9.75	9.56
22	-0.08	0.00	0.00	1.00	6.48	6.54
23	-0.08	0.00	0.00	1.00	6.48	6.44
24	0.00	-1.00	0.00	1.00	8.33	8.57
25	0.04	0.00	0.50	1.00	9.55	9.29

A square root transformation was necessary to obtain a good model. After performing a stepwise regression on the square root of the response variable, the following acceptable model was obtained, with an R^2_{adj} value of 0.9422, and a lack of fit statistic value <0.0001:

$$\sqrt{PI} = 5.09 + 1.03a_3 - 42.50a_0a_2 - 41.07a_0a_3 - 6.48a_0X_0 - 3.10a_2a_3$$
$$-365.19{a_0}^2 - 1.93{a_2}^2 - 0.86{a_3}^2 - 0.39{X_0}^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.2.6: Parameter Estimates with Confidence Intervals for Case 3

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	5.09	4.69	5.48
a_3	1.03	0.80	1.25
a_0a_2	-42.50	-61.61	-23.39
a_0a_3	-41.07	-54.36	-27.78
a_0X_0	-6.48	-9.09	-3.87
a_2a_3	-3.10	-4.18	-2.03
a_0^2	-365.19	-424.17	-306.21
a_2^2	-1.93	-2.35	-1.50
a_3^2	-0.86	-1.36	-0.36
X_0^2	-0.39	-0.60	-0.17

This model was then optimized by using the fmincon function in MATLAB®. The optimum factor levels found were:

$$a_0 = -0.0181 \quad a_1 = -0.1991 \quad a_2 = -0.1039 \quad a_3 = 0.6783 \quad X_0 = 0.1508$$

The corresponding substrate feeding profile, along with the simulated substrate concentration, biomass concentration, and product concentration behaviors with time are shown below:

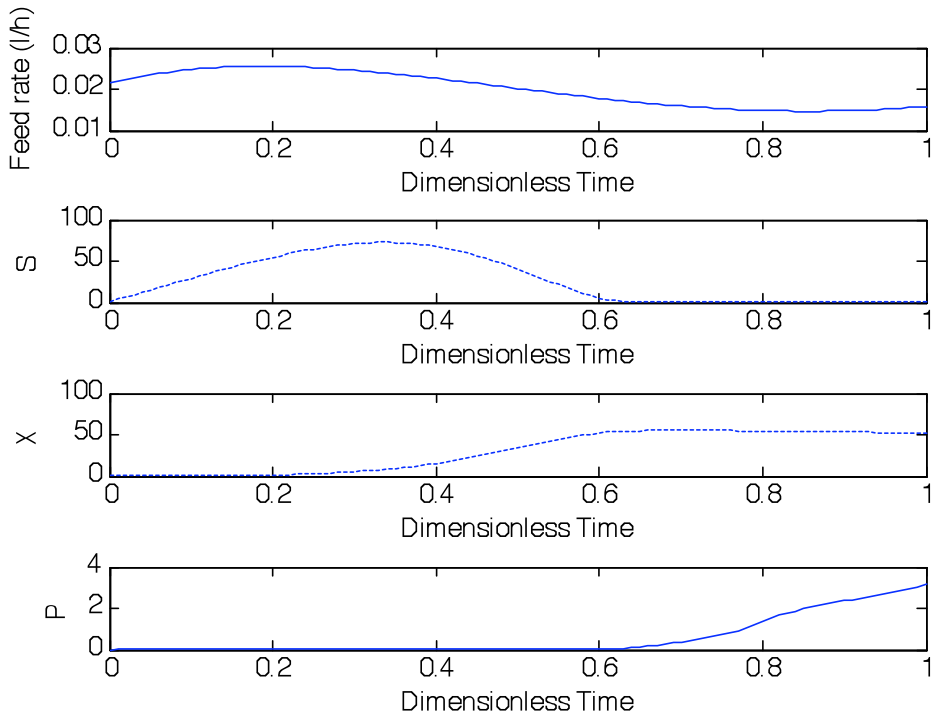

Figure 4.2.4: Optimum Profile for Case 3

The optimum performance index obtained from the response surface model was 34.86±4.14 grams. The performance index obtained by simulating the optimum profile was found to be 32.03 grams, which is within the confidence interval of the predicted optimum. This is a significant improvement from Case 2.

4.2.4 Case 4

The next step is to repeat Case 3, with the inclusion of a fourth dynamic subfactor in the design corresponding to the fifth shifted Legendre polynomial.

As before, the volume constraint remains the same as that of Case 3, since it applies regardless of how many polynomials are considered.

$$\Rightarrow a_1 = 11 \cdot a_0$$

For this case, we will have five independent factors that will be included in the design: a_0, a_2, a_3, a_4, and X_0.

We now have 16 constraints on the dynamic subfactors:

$$-1 \le 12 \cdot a_0 \pm a_2 \pm a_3 \pm a_4 \le 1$$

$$-1 \le -10 \cdot a_0 \pm a_2 \pm a_3 \pm a_4 \le 1$$

The range of values tested for X_0 remains:

$$X_0 = 0.10 \pm 0.05$$

A D-optimal design was generated using Design Expert®, with the above constraints on the dynamic subfactors. The designed dynamic experiments were subsequently simulated in MATLAB®. The designed experiments along with the simulated penicillin production are presented below:

Table 4.2.7: D-Optimal Design for Case 4 with Simulated Results

Run	a_0	a_2	a_3	a_4	X_0	PI	PI with error
1	0.00	0.00	0.00	-1.00	-1.00	9.59	9.72
2	0.04	0.00	0.00	-0.50	-1.00	13.50	14.13
3	0.04	0.00	0.00	-0.50	-1.00	13.50	12.72
4	0.00	-1.00	0.00	0.00	-1.00	6.88	7.03
5	-0.04	0.00	-0.50	0.00	-1.00	7.95	8.02
6	-0.04	0.00	0.50	0.00	-1.00	15.56	15.04
7	0.00	1.00	0.00	0.00	-1.00	10.94	10.82
8	0.00	0.00	1.00	0.00	-1.00	31.11	31.38
9	0.00	0.00	1.00	0.00	-1.00	31.11	33.95
10	0.04	0.00	0.00	0.50	-1.00	12.54	13.42
11	0.04	0.00	0.00	0.50	-1.00	12.54	12.10
12	0.00	0.00	0.00	1.00	-0.50	27.88	30.03
13	0.00	-0.25	-0.25	0.00	-0.25	13.13	13.37
14	0.00	0.00	0.50	-0.50	0.00	18.36	18.33
15	0.00	-1.00	0.00	0.00	0.00	7.93	8.08
16	0.00	0.50	0.00	0.50	0.00	14.34	14.27
17	0.00	0.00	0.50	0.50	0.00	29.44	29.34
18	0.00	0.00	-1.00	0.00	0.00	9.01	9.35
19	0.00	0.50	0.00	-0.50	0.50	14.65	15.18
20	0.00	0.00	-0.50	0.50	0.50	20.58	21.32
21	0.00	0.00	0.00	-1.00	1.00	9.89	10.06
22	-0.04	-0.50	0.00	0.00	1.00	8.02	7.77
23	-0.08	0.00	0.00	0.00	1.00	6.48	6.59
24	-0.08	0.00	0.00	0.00	1.00	6.48	6.74
25	0.00	0.50	-0.50	0.00	1.00	16.62	16.83
26	0.04	-0.50	0.00	0.00	1.00	16.77	17.21

27	0.08	0.00	0.00	0.00	1.00	7.43	7.57
28	0.08	0.00	0.00	0.00	1.00	7.43	7.38
29	0.00	0.50	0.50	0.00	1.00	11.59	11.68
30	0.00	0.00	1.00	0.00	1.00	15.92	15.60
31	0.00	0.00	0.00	1.00	1.00	27.60	28.22

After performing a stepwise regression, the following acceptable model was obtained with an R^2_{adj} value of 0.9487, and a lack of fit statistic value of 0.073:

$$PI = 16.18 + 7.67a_3 + 9.87a_4 - 226.48a_0a_2 + 238.37a_0a_3 - 252.84a_0a_4$$
$$-21.55a_2a_4 - 9.28a_3X_0 - 1395.02a_0^2 - 7.51a_2^2 + 3.33a_4^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.2.8: Parameter Estimates with Confidence Intervals for Case 4

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	16.18	14.92	17.44
a_3	7.67	5.96	9.37
a_4	9.87	8.14	11.60
a_0a_2	-226.48	-353.72	-99.24
a_0a_3	238.37	96.71	380.03
a_0a_4	-252.84	-351.90	-153.77
a_2a_4	-21.55	-32.71	-10.40
a_3X_0	-9.28	-11.31	-7.26
a_0^2	-1395.02	-1733.89	-1056.15
a_2^2	-7.51	-10.10	-4.93
a_4^2	3.33	0.96	5.69

This model was then optimized by using the fmincon function in MATLAB®. The optimum factor levels found were:

$$a_0 = 0.0041 \quad a_1 = 0.0451 \quad a_2 = 0 \quad a_3 = 0.9507 \quad a_4 = 0 \quad X_0 = -1$$

The corresponding substrate feeding profile, along with the simulated substrate concentration, biomass concentration, and product concentration behaviors with time are shown below:

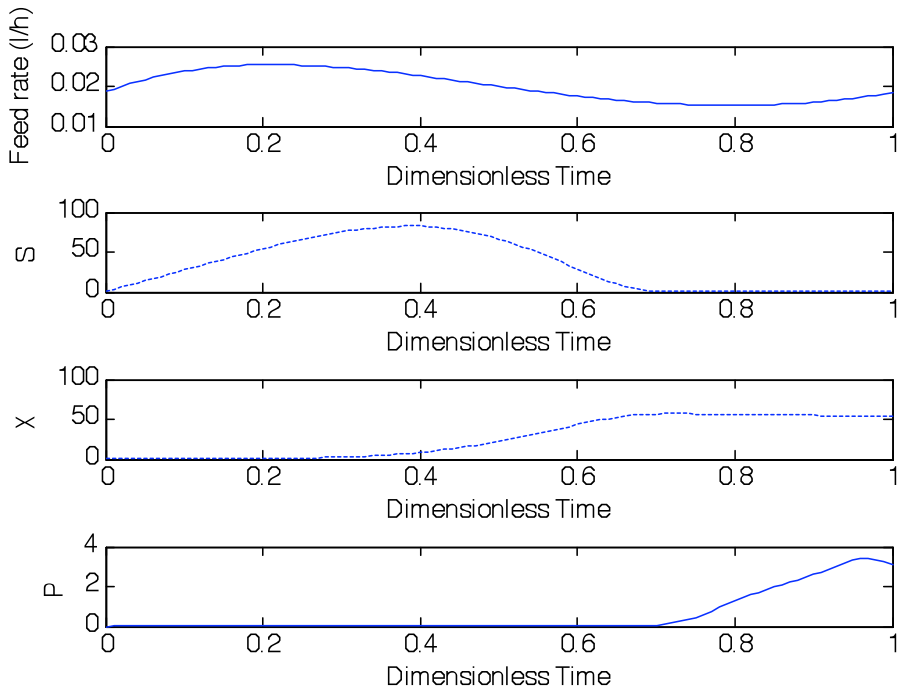

Figure 4.2.5: Optimum Profile for Case 4

The optimum performance index obtained from the response surface model was 33.20±2.50 grams. The performance index obtained by simulating the optimum

profile was found to be 31.76 grams, which is within the confidence interval of the predicted optimum. This is almost identical to what was obtained in Case 3. In addition, the optimum value for the dynamic subfactor corresponding to the fifth shifted Legendre polynomial is zero. Therefore, we can conclude that adding this term in our design has no improvement on the production of penicillin.

We notice that towards the end of the batch time, the concentration of penicillin begins to decrease. This indicates that the product is being metabolized. This may occur when the biomass concentration goes below a certain value, as can be seen in the fermentation model.

4.3 Decentralized Design of Dynamic Experiments

Next, we shall look at the decentralized design where the substrate feeding profile is split into two parts; a growth phase and a production phase.

From the optimum found in Case 3, we can approximate an appropriate base case to start with. For the decentralized designs, two separate feeding profiles must be designed. The values of the variables defining the range of the two profiles were chosen as follows:

Table 4.3.1: Values of variables defining the range of the substrate feeding profiles in both phases

Variable	Value	Unit
u_{m1}	0.025	l/h

Δu_1	0.005	l/h
u_{m2}	0.015	l/h
Δu_2	0.005	l/h

The volume constraint for the decentralized case is:

$$V_0 + t_f\left(u_{m1} + \Delta u_1 a_{10}\right) + \left(t_b - t_f\right)\cdot\left(u_{m2} + \Delta u_2 a_{20}\right) \le V_{reactor}$$

Since we are setting the final volume to be equal to the maximum working reactor volume, we can rearrange the constraint to obtain an expression for the feed switch time as a function of the dynamic subfactors:

$$t_f = \frac{V_{reactor} - V_0 - t_b\cdot\left(u_{m2} + \Delta u_2 a_{20}\right)}{u_{m1} - u_{m2} + \Delta u_1 a_{10} - \Delta u_2 a_{20}}$$

Substituting the values of the variables:

$$t_f = \frac{0.75 - 0.75\cdot a_{20}}{0.01 + 0.005\cdot\left(a_{10} - a_{20}\right)}$$

However, some other restrictions must be places on the feed switch time. This must be done in order to prevent the feed switch time from exceeding the batch time. Also, in order to make a smaller design space, we shall restrict the feed switch time to fall between $0.2*t_b$ and $0.4*t_b$. This is where the optimum switch time usually falls (Riascos & Pinto, 2004):

$$\Rightarrow 30 \le \frac{0.75 - 0.75 \cdot a_{20}}{0.01 + 0.005 \cdot (a_{10} - a_{20})} \le 60$$

This yields the following two inequality constraints:

$$a_{10} + 4 \cdot a_{20} \le 3$$

$$0.5 \le a_{10} + 1.5 \cdot a_{20}$$

These constraints apply disregarding how many Legendre polynomials are used in designing the feeding profiles.

4.3.1 Case 1

For this case, we shall evaluate the first 2 shifted Legendre polynomials, as well as the initial biomass concentration. Examining the first 2 shifted Legendre polynomials requires including 2 dynamic subfactors for each profile in the design of experiments. This results in five independent factors that will be included in the design: a_{10}, a_{11}, a_{20}, a_{21}, and X_0.

The constraints on the dynamic subfactors are:

$$-1 \le a_{10} + a_{11} \le 1$$

$$-1 \le a_{10} - a_{11} \le 1$$

$$-1 \le a_{20} + a_{21} \le 1$$

$$-1 \le a_{20} - a_{21} \le 1$$

A D-optimal design was generated using Design Expert®, with the above constraints on the dynamic subfactors, as well as the constraints on the feed switch time. The designed dynamic experiments were simulated in MATLAB®. The designed experiments along with the simulated penicillin production are presented below:

Table 4.3.2: D-Optimal Design for Case 1 with Simulated Results

Run	a_{10}	a_{11}	a_{20}	a_{21}	X_0	PI	PI with error
1	0.50	-0.50	0.00	-1.00	-1.00	1.40	1.37
2	0.50	-0.50	0.00	-1.00	-1.00	1.40	1.43
3	0.50	0.50	0.00	-1.00	-1.00	1.40	1.43
4	1.00	0.00	0.25	-0.75	-1.00	7.60	7.55
5	0.00	-1.00	0.75	-0.25	-1.00	10.19	10.24
6	0.00	1.00	0.75	-0.25	-1.00	10.19	9.89
7	1.00	0.00	-0.33	0.00	-1.00	3.99	3.87
8	1.00	0.00	-0.33	0.00	-1.00	3.99	4.00
9	1.00	0.00	0.50	0.50	-1.00	7.28	7.76
10	1.00	0.00	0.50	0.50	-1.00	7.28	7.16
11	0.00	-1.00	0.33	0.67	-1.00	8.14	8.18
12	0.50	0.50	0.00	1.00	-1.00	9.72	9.70
13	1.00	0.00	0.00	-1.00	0.00	2.72	2.58
14	0.00	-1.00	0.75	-0.25	0.00	9.44	9.34
15	0.00	1.00	0.75	-0.25	0.00	9.45	9.01
16	0.27	0.00	0.15	0.00	0.00	20.04	20.47
17	0.27	0.00	0.15	0.00	0.00	20.04	19.59
18	0.00	-1.00	0.33	0.67	0.00	7.89	7.78
19	0.00	1.00	0.33	0.67	0.00	7.89	7.95

20	0.00	1.00	0.33	0.67	0.00	7.89	7.77
21	1.00	0.00	0.00	1.00	0.00	8.88	8.99
22	1.00	0.00	-0.17	-0.83	1.00	2.88	2.94
23	0.00	-1.00	0.33	-0.67	1.00	20.14	21.02
24	0.00	1.00	0.33	-0.67	1.00	20.12	20.02
25	1.00	0.00	0.50	-0.50	1.00	21.46	20.29
26	0.50	-0.50	0.63	0.38	1.00	6.64	6.87
27	0.50	0.50	0.63	0.38	1.00	6.64	6.46
28	1.00	0.00	-0.33	0.67	1.00	31.32	32.09

Three of the runs were removed, as they caused the denominator term in the expression for the feed switch time to be zero. This occurs when a_{10}=-1 and a_{20}=+1. This means that in the range of substrate feeding profiles we are examining, when a_{10}=-1 and a_{20}=+1, it is impossible to fill the reactor to 10 liters. This is because:

$$V_0 + t_f\left(u_{m1} + \Delta u_1 a_{10}\right) + \left(t_b - t_f\right)\cdot\left(u_{m2} + \Delta u_2 a_{20}\right) = V_{final}$$

becomes,

$$8.5 = V_{final}$$

Design Expert® does not allow the input of an equality constraint, so we cannot specify the design space to include such points. One way around this, would be to redefine the range of the substrate feeding profiles in a way that excludes these runs.

A natural log transformation was necessary to obtain a good model. After performing a stepwise regression on the natural log of the response variable, the following acceptable model was obtained, with an R^2_{adj} value of 0.9838, and lack of fit statistic value of 0.0052:

$$\ln(PI) = 3.13 - 0.41a_{10} + 0.77a_{21} + 0.45X_0 + 1.29a_{10}a_{20} - 3.41a_{20}a_{21} - 0.67a_{20}X_0 - 0.14a_{21}X_0 - 2.17a_{20}^2 - 1.17a_{21}^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.3.3: Parameter Estimates with Confidence Intervals for Case 1

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	3.13	2.99	3.27
a_{10}	-0.41	-0.52	-0.30
a_{21}	0.77	0.69	0.86
X_0	0.45	0.39	0.52
$a_{10}a_{20}$	1.29	1.06	1.51
$a_{20}a_{21}$	-3.41	-3.72	-3.09
$a_{20}X_0$	-0.67	-0.82	-0.51
$a_{21}X_0$	-0.14	-0.22	-0.049
a_{20}^2	-2.17	-2.52	-1.82
a_{21}^2	-1.17	-1.33	-1.01

The second dynamic subfactor for the growth phase feeding profile was found to be insignificant. This is in agreement with the optimization results of Riascos and Pinto (Riascos and Pinto, 2003).

This model was then optimized by performing a nonlinear constrained optimization in MATLAB®. The optimum factor levels found were:

$$a_{10} = 0.62 \quad a_{20} = -0.08 \quad a_{21} = 0.30 \quad X_0 = 1.00$$

The corresponding substrate feeding profile, along with the simulated substrate concentration, biomass concentration, and product concentration behaviors with time are shown below:

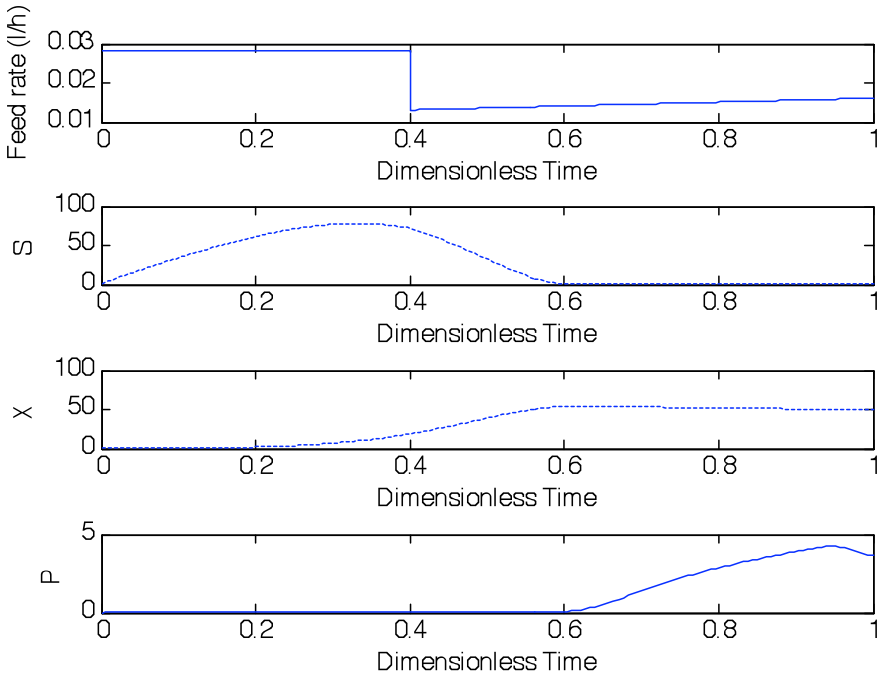

Figure 4.3.1: Optimum Profile for Case 1

The optimum performance index obtained from the response surface model was 32.18±4.63 grams. The performance index obtained by simulating the optimum profile was found to be 36.12 grams, which is within the confidence interval of the predicted optimum.

Again, we notice that towards the end of the batch time, the concentration of penicillin begins to decrease. This indicates that the product is being metabolized. This may occur when there is not enough biomass concentration in the system.

4.3.2 Case 2

The next step is to repeat Case 1, except now we will only include one factor for the growth phase, and increase the number of subfactors for the production phase by 1. This results in five independent factors that will be included in the design: a_{10}, a_{20}, a_{21}, a_{22}, and X_0.

The constraints on the dynamic subfactors are:

$$-1 \le a_{20} + a_{21} + a_{22} \le 1$$

$$-1 \le a_{20} + a_{21} - a_{22} \le 1$$

$$-1 \le a_{20} - a_{21} + a_{22} \le 1$$

$$-1 \le a_{20} - a_{21} - a_{22} \le 1$$

A D-optimal design was generated using Design Expert®, with the above constraints on the dynamic subfactors, as well as the constraints on the feed switch time. The designed dynamic experiments were simulated in MATLAB®. The designed experiments along with the simulated penicillin production are presented below:

Table 4.3.4: D-Optimal Design for Case 2 with Simulated Results

Run	a_{10}	a_{20}	a_{21}	a_{22}	X_0	PI	PI with error
1	1.00	0.00	0.00	-1.00	-1.00	8.46	8.24
2	-0.25	0.50	0.00	-0.50	-1.00	15.93	15.46
3	1.00	0.50	-0.50	0.00	-1.00	24.64	24.30
4	1.00	0.50	0.50	0.00	-1.00	7.28	7.46
5	1.00	0.50	0.50	0.00	-1.00	7.28	7.38
6	0.50	0.00	1.00	0.00	-1.00	9.72	9.72
7	1.00	0.00	-0.50	0.50	-1.00	12.37	12.36
8	1.00	0.00	-0.50	0.50	-1.00	12.37	12.12
9	-0.25	0.50	0.00	0.50	-1.00	9.83	10.08
10	1.00	0.00	0.50	0.50	-1.00	20.60	20.53
11	1.00	0.00	0.50	0.50	-1.00	20.60	20.22
12	0.50	0.00	-1.00	0.00	-0.75	2.02	2.09
13	1.00	-0.33	0.00	-0.67	0.00	9.61	9.56
14	1.00	-0.33	-0.67	0.00	0.00	0.76	0.75
15	0.60	0.60	0.00	0.00	0.00	9.04	8.97
16	0.50	0.00	1.00	0.00	0.00	9.58	9.31
17	1.00	0.00	0.00	1.00	0.00	20.65	21.98
18	0.50	0.00	0.00	-1.00	1.00	10.00	10.42
19	1.00	0.50	0.00	-0.50	1.00	15.26	15.38

20	1.00	0.50	0.00	-0.50	1.00	15.26	14.77
21	1.00	0.00	-1.00	0.00	1.00	4.21	4.12
22	1.00	0.00	-1.00	0.00	1.00	4.21	4.19
23	-0.25	0.50	-0.50	0.00	1.00	19.72	20.12
24	0.00	0.75	0.25	0.00	1.00	6.50	6.11
25	-0.25	0.50	0.50	0.00	1.00	7.21	6.94
26	1.00	0.00	1.00	0.00	1.00	8.71	8.78
27	1.00	0.50	0.00	0.50	1.00	8.28	8.36
28	1.00	-0.33	0.00	0.67	1.00	0.27	0.27

Three of the runs were removed, as they caused the denominator term in the expression for the feed switch time to be zero. This occurs for the same reasons explained in Case 1.

A square root transformation was necessary to obtain a good model. After performing a stepwise regression on the square root of the response variable, the following acceptable model was obtained, with an R^2_{adj} value of 0.8518, and a lack of fit statistic value <0.0001:

$$\sqrt{PI} = 2.09 - 2.46a_{10} + 2.22a_{20} + 0.71a_{21} - 0.48a_{10}X_0 - 5.56a_{20}a_{21} - 1.06a_{22}X_0 + 0.33{a_{10}}^2 + 0.84{a_{21}}^2 + 1.52{a_{22}}^2 - 0.73{X_0}^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 4.3.5: Parameter Estimates with Confidence Intervals for Case 2

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	2.09	1.38	2.80
a_{10}	-2.46	-3.81	-1.12
a_{20}	2.22	1.34	3.09
a_{21}	0.71	0.38	1.03
$a_{10}X_0$	-0.48	-0.70	-0.25
$a_{20}a_{21}$	-5.56	-7.09	-4.02
$a_{22}X_0$	-1.06	-1.49	-0.63
a_{10}^2	3.33	1.80	4.86
a_{21}^2	0.84	0.06	1.62
a_{22}^2	1.52	0.64	2.40
X_0^2	-0.73	-1.25	-0.20

This model was then optimized by performing a nonlinear constrained optimization. The optimum factor levels found were:

$$a_{10} = 1.00 \quad a_{20} = 0.50 \quad a_{21} = -0.50 \quad a_{22} = 0.00 \quad X_0 = -0.80$$

The corresponding substrate feeding profile, along with the simulated substrate concentration, biomass concentration, and product concentration behaviors with time are shown below:

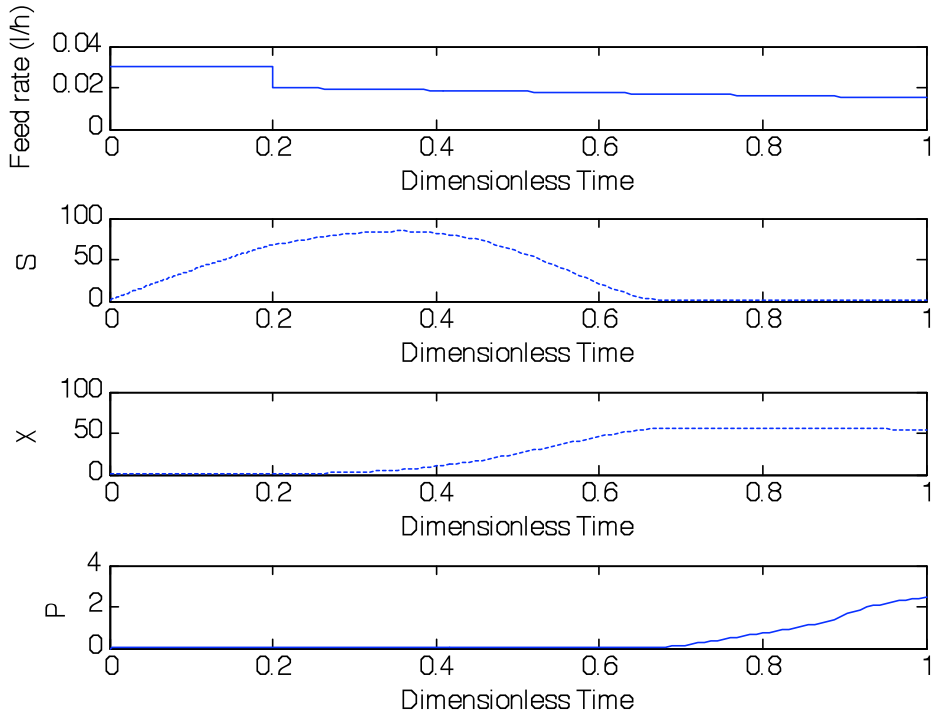

Figure 4.3.2: Optimum profile for Case 2

The optimum performance index obtained from the response surface model was 27.31±6.01 grams. The performance index obtained by simulating the optimum profile was found to be 25.48 grams, which is within the confidence interval of the predicted optimum.

The optimum value of the third dynamic subfactor for the production phase feeding profile is zero, which indicates a linear production phase feeding profile is preferable over a quadratic profile.

When comparing the optimum of Case 2 with the optimum of Case 1, it appears that the optimum of Case 2 is lower than that of Case 1. However, the RSM optimum prediction from Case 2 has a larger uncertainty than that of Case 1, and when comparing the two confidence intervals, they are very close. The simulated optimum from Case 2, however, is lower than that of Case 1. This can be attributed to the large prediction uncertainty of the RSM from Case 2. Another reason for the difference in the optimum predictions could be due to the different design spaces used for each case. This could result in slightly different response surfaces, with slightly different optimum values.

5 Conclusions

Design of Dynamic Experiments is an effective technique for optimizing batch fermentation processes. This technique was applied to two main cases: centralized and decentralized designs. We observed that initial substrate concentration was not an important factor in the region we examined. However, initial biomass concentration had an impact on the final production of penicillin, though not a large one.

For the centralized design, we observed a big improvement in penicillin production when going from using 3 shifted Legendre polynomials to describe the substrate feed flow rate, to using 4 shifted Legendre polynomials. However, increasing the number of dynamic subfactors to include 5 shifted Legendre polynomials had no improvement on the amount of penicillin produced. In effect, the dynamic subfactor attributed to the 5^{th} shifted Legendre polynomial was found to be zero for optimal penicillin production.

For the decentralized design, we observed that only one dynamic subfactor was significant in characterizing the growth phase feeding profile. However, for the production phase, it was found that the second dynamic subfactor was significant, and that the optimal profile was a linearly decreasing one. The addition of a third dynamic subfactor for the production phase feeding profile was optimized to be

zero for optimal penicillin production. This indicates that there is no improvement in antibiotic production when using the 3rd shifted Legendre polynomial.

Comparing the best case from the centralized design, with the best case from the decentralized design, the optimum penicillin production is almost the same. Both designs also contain the same number of runs, so using either design will require the same amount of effort. However, centralized designs have the advantage of using more shifted Legendre polynomials with less experiments than the decentralized designs, since two feeding profiles need to be designed. Thus, to test the same number of polynomials, the decentralized design will require almost double the number of dynamic subfactors to be included in the design, which increases the number of experiments needed to be run. In reality, if these experiments were to be run, it would take a significant amount of time to run the decentralized designs with several dynamic subfactors for each feeding phase.

Having had a model that could be simulated to describe penicillin fermentation meant time was not a factor in doing these designs. However, in reality, if such a process is to be optimized, especially if no first-principles model is available to describe the system, the experiments will have to be run. Time will be of great importance in such a case. One will not have the liberty to run all the different designs that were done in this study. However, from the outcomes and conclusions of this study, we can recommend what design to run. I would recommend using the centralized design with dynamic subfactors to represent up

to the fourth Legendre polynomial. This will require 31 runs if using a D-optimal design.

6 Bibliography

Bajpai, R. K., & Reuss, M. (1980). A Mechanistic Model for Penicillin Production. *Chemical Technology and Biotechnology , 30* (6), 332-344.

Biegler, L. T. (1984). Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering , 8*, 243-247.

Birol, G., Undey, C., & Cinar, A. (2002). A modular simulation package for fed-batch fermentation: penicillin production. *Computers and Chemical Engineering , 26*, 1553-1565.

Box, G. E., & Draper, N. R. (2007). *Response Surfaces, Mixtures, and Ridge Analyses* (2nd Edition ed.). Wiley-Interscience.

Chiou, J.-P., & Wang, F.-S. (1999). Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process. *Computers and Chemical Engineering , 23* (9), 1277-1291.

Cuthrell, J. E., & Biegler, L. T. (1989). Simultaneous optimization and solution methods for batch reactor control profiles. *Computers and Chemical Engineering , 13*, 49-62.

Georgakis, C. (2009). A Model-Free Methodology for the Optimization of Batch Processes: Design of Dynamic Experiments. *Proceedings of the 8th International Symposium on Dynamics and Control of Process Systems.*

Horng, I. R., & Chou, J. H. (1986). The design of optimal observers with specified eigenvalues via shifted Legendre polynomials. *Optimization Theory and Application , 51*, 179-188.

Lim, H. C., Tayeb, Y. J., Modak, J. M., & Bonte, P. (1986). Computational algorithms for optimal feed rates for a class of fed-batch fermentation: Numerical results for penicillin and cell mass production. *Biotechnology and Bioengineering , 28*, 1408-1420.

Montgomery, D. C. (2005). *Design and Analysis of Experiments* (6th Edition ed.). John Wiley & Sons, Inc.

Montgomery, D. C., Peck, E. A., & Vining, G. (2006). *Introduction to Linear Regression Analysis.* Wiley-Interscience.

Riascos, C. A., & Pinto, J. M. (2004). Optimal control of bioreactors: a simultaneous approach for complex systems. *Chemical Engineering , 99*, 23-34.

7 Appendix

7.1 Case 1

The first case we will look at is at the left corner of the design space shown in Figure 1.

$$u_m = 0.008 l/h$$

$$\Delta u = 0.002 l/h$$

We shall choose the design space such that a_0 and a_1 range between:

$$-4 < a_0 < 0$$

$$-2 < a_1 < 2$$

We also need to impose the constraints derived in Section 4.2.1:

$$a_0 + a_1 > -4$$

$$a_0 - a_1 > -4$$

The following design of experiments for this case was generated using Design Expert®. Also included in this table is the simulated performance index (mass of penicillin produced) for each of the runs.

Table 7.1.1: D-Optimal Design of Experiments for Case 1 with Simulated Results

Run	a_0	a_1	PI (g)	PI with error
1	-2.00	-2.00	0.82	0.86
2	-2.00	-2.00	0.82	0.77
3	0.00	-2.00	5.40	5.52
4	0.00	-2.00	5.40	5.45
5	-3.00	-1.00	0.53	0.51
6	-1.30	-1.00	8.99	8.89
7	0.00	-0.67	9.77	9.85
8	-4.00	0.00	0.00	0.00
9	-4.00	0.00	0.00	0.00
10	-0.80	0.00	6.24	6.02
11	0.00	0.67	4.92	5.30
12	-1.30	1.00	3.73	3.80
13	-2.00	2.00	0.00	0.00
14	-2.00	2.00	0.00	0.00
15	0.00	2.00	3.27	3.26
16	0.00	2.00	3.27	3.26

After performing a stepwise regression, the following acceptable model in terms of coded factors was obtained with an R^2_{adj} value of 0.9082, and a Lack of Fit statistic value <0.0001:

$$PI = 1.62 + 1.54a_0 - 0.42a_1 - 1.10a_1^2$$

The parameter estimates along with the confidence intervals are presented below:

Table 7.1.2: Parameter Estimates with Confidence Intervals for Case 1

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	1.62	1.33	1.92
a_0	1.54	1.25	1.82
a_1	-0.42	-0.66	-0.17
a_1^2	-1.10	-1.54	-0.67

The model was optimized using a nonlinear constrained optimization in MATLAB®. The optimum was found to occur at a_0=0, and a_1=-0.38, which is at the maximum value of a_0, and at an intermediate value of a_1.

The optimum performance index obtained from the response surface is 10.22± 2.69 grams. The performance index obtained by simulating the optimum profile was found to be 8.22 grams, which is within the confidence interval of the predicted optimum.

7.2 Case 2

The next case we will look at is a repetition of Case 1, but with different ranges of the dynamic subfactors. We shall choose the design space such that a_0 and a_1 range between:

$$0 < a_0 < 4$$

$$-2 < a_1 < 2$$

The following design of experiments for this case was generated using Design Expert®. Also included in this table is the simulated performance index (mass of penicillin produced) for each of the runs.

Table 7.2.1: D-Optimal Design of Experiments for Case 2 with Simulated Results

Run	a_0	a_1	PI (g)	PI with error
1	0.00	-2.00	5.40	5.55
2	0.00	-2.00	5.40	5.56
3	4.00	-2.00	18.39	17.98
4	4.00	-2.00	18.39	18.42
5	2.00	-1.00	10.49	10.16
6	0.00	0.00	6.47	6.29
7	0.00	0.00	6.47	6.47
8	3.00	0.00	6.52	6.78
9	4.00	0.00	6.43	6.31
10	1.00	1.00	4.68	4.72
11	2.00	1.00	4.85	4.82
12	0.00	2.00	3.27	3.37
13	0.00	2.00	3.27	3.18
14	2.00	2.00	3.80	3.81
15	4.00	2.00	4.05	4.11
16	4.00	2.00	4.05	4.16

After performing a stepwise regression, the following acceptable model in terms of coded factors was obtained with an R^2_{adj} value of 0.9989, and a Lack of Fit statistic value of 0.5585:

$$PI = 7.04 - 5.76a_1 - 2.94a_0a_1 - 0.70a_0^2 + 1.45a_1^2 + 0.57a_0^2a_1 + 3.38a_0a_1^2 + 1.10a_1^3$$

The parameter estimates along with the confidence intervals are presented below:

Table 7.2.2: Parameter Estimates with Confidence Intervals for Case 2

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	7.04	6.83	7.24
a_1	-5.76	-6.33	-5.19
a_0a_1	-2.94	-3.07	-2.82
a_0^2	-0.70	-0.97	-0.42
a_1^2	1.45	1.22	1.69
$a_0^2a_1$	0.57	0.11	1.04
$a_0a_1^2$	3.38	3.25	3.51
a_1^3	1.10	0.34	1.85

The model was optimized using a nonlinear constrained optimization in MATLAB®. The optimum was found to occur at a_0=4.00, and a_1=-2.00, which is at the maximum value of a_0.

The optimum performance index obtained from the response surface is 18.20± 0.25 grams. The performance index obtained by simulating the optimum profile was found to be 18.16 grams, which is within the confidence interval of the predicted optimum.

7.3 Case 3

The next case we will look at is a repetition of Case 2, but with different ranges of the dynamic subfactors. We shall choose the design space such that a_0 and a_1 range between:

$$2 < a_0 < 6$$

$$-2 < a_1 < 2$$

This is also equivalent to choosing the following design, which will be used:

$$u_m = 0.016 l/h$$

$$\Delta u = 0.004 l/h$$

$$-1 < a_0 < 1$$

$$-1 < a_1 < 1$$

The following design of experiments for this case was generated using Design Expert®. Also included in this table is the simulated performance index (mass of penicillin produced) for each of the runs.

Table 7.3.1: D-Optimal Design of Experiments for Case 3 with Simulated Results

Run	a_0	a_1	PI (g)	PI with error
1	-1.00	-1.00	15.96	16.08
2	-1.00	-1.00	15.96	16.22
3	-0.33	-1.00	18.74	18.70
4	1.00	-1.00	12.61	12.78

5	1.00	-1.00	12.61	12.20
6	0.00	-0.50	9.06	9.31
7	-1.00	0.00	6.58	6.41
8	-1.00	0.00	6.58	6.27
9	1.00	0.00	6.21	6.43
10	-0.50	0.50	4.95	4.94
11	0.50	0.50	5.01	5.06
12	-1.00	1.00	3.80	3.77
13	-1.00	1.00	3.80	3.70
14	0.00	1.00	4.05	3.73
15	1.00	1.00	4.14	4.21
16	1.00	1.00	4.14	4.11

After performing a stepwise regression, the following acceptable model in terms of coded factors was obtained with an R^2_{adj} value of 0.9989, and a Lack of Fit statistic value of 0.6228:

$$PI = 1/(0.15 + 0.089a_1 - 0.011a_0a_1 - 0.0074a_1^2 - 0.015a_0^2a_1 - 0.0021a_0^3 + 0.018a_1^3)$$

The parameter estimates along with the confidence intervals are presented below:

Table 7.3.2: Parameter Estimates with Confidence Intervals for Case 3

Factor	Coefficient Estimate	95% CI Low	95% CI High
Intercept	0.15	0.15	0.16
a_1	0.089	0.079	0.098
a_0a_1	-0.011	-0.013	-0.0093

a_1^2	0.0074	0.0039	0.011
$a_0^2 a_1$	-0.015	-0.02	-0.01
a_0^3	-0.0021	-0.0039	-0.0003
a_1^3	0.018	0.0071	0.028

The model was optimized using a nonlinear constrained optimization in MATLAB®. The optimum was found to occur at a_0=-0.35, and a_1=-1.00, which is at the minimum value of a_1.

The optimum performance index obtained from the response surface is 18.60± 1.38 grams. The performance index obtained by simulating the optimum profile was found to be 18.98 grams, which is within the confidence interval of the predicted optimum.