Chapter 5

Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria

Abstract. Model predictive control is shown to be a reliable method to keep cultures of genetically modified bacteria very close to predetermined profiles of their key physiological variable, the specific biomass growth rate. This was shown experimentally by *E.coli* bacteria producing the recombinant model protein GFP, which can be monitored quickly and accurately by means of fluorescence spectrometry. In the experiments, the culture was shown to exactly follow a complicated path with considerable jumps in the specific growth rate. They were performed in a standard 10 L Biostat® C fermenter. Prior to the control experiments, a process model was developed and validated against data from the process under consideration. This model was then used to determine the corresponding optimal feeding profile required to keep the process at the desired profile of the specific growth rate by means of numerical optimization. The experiments showed that this model predictive control procedure can be routinely applied to protein production processes, when it is possible to provide sufficient online measurement information about the current state of the process. This was shown to be possible using Extended Kalman Filter algorithms running on a simple PC.

1 INTRODUCTION

Recombinant proteins are becoming economically important products in pharmaceutical industries (Walsh 2000). The demands with respect to product quality and reproducibility of their production processes are very high, particularly for therapeutic proteins. Indeed, it is not only the product itself but also the production process as such which must be validated in order to make sure that the final product meets the safety requirements. The most straightforward way to guarantee reproducibility of the production process is tight closed-loop, i.e. feedback control.

Biomass growth rate has ever been considered the central dynamic property of microbial physiology (Neidhard et al. 1990). As the specific biomass growth rate $\mu$ is also primary influencing the specific product development rate $\pi$ in most industrial protein production systems, it is straightforward to control this important variable to its optimal values as appropriate.

In fed-batch operation, the preferred mode in protein production, the culture is primarily controlled by the feed rate $F_S$ and the substrate concentration $S_F$ in the feed. Several control approaches have been proposed in literature (Lim and Lee 1991, Lee and Ramirez 1994, Schubert et al. 1994, Levisauskas et al. 1996, Sandoz et al. 1999, 2000, Akesson et al. 2001, DeLisa et al. 2001), however, only a few industrial cultivation bioprocesses have truly been controlled in a closed loop fashion. The most important reason why direct feedback control is seldom applied in practice is that it is difficult to obtain quickly and accurately enough sufficient information about the current state of the production process. In order to determine the state of the protein formation process, indirectly, i.e. model supported measurement procedures must be used. Hence, a discussion on control does not make sense without concrete concepts about getting the necessary information about the state of the process. Again, several approaches to state estimation have been discussed in literature. Extended Kalman Filter algorithms (e.g., Stephanopoulos et al. 1993) were shown to be well suited for this purpose.

Here, we report on an “Extended Kalman Filter” which estimates the biomass concentration $X$, the current culture mass $W$, and the value of the controlled variable, the specific biomass growth rate $\mu$. These estimates are then used in a model predictive controller which keeps the specific biomass growth rate $\mu$ close to its desired value. In order to make sure that the underlying process model describes the process sufficiently well, the model parameter that most significantly changes during a recombinant protein production process, the biomass yield on the substrate $Y_{XS}$, is adapted in a straightforward simple way.

We will demonstrate that it is possible with this adaptive model predictive control approach to force the process following very complicated profiles of the specific growth rate $\mu$. This is important for quickly and reliably establishing the relationship between the key quantities $\mu$ and the specific product formation rate $\pi$ and further for control of the protein production if the desired $\mu$-profile is already known.
2 EXPERIMENTAL

Strain and cultivation technique

*E. coli* BL21 (DE3) pET11a EGFP was used as the organism in the experiments. The organism is able to express the green fluorescent protein (GFP) under control of the T7 promoter (Sambrook *et al.* 2000). Product formation was induced with IPTG. A defined mineral salt medium was used with glucose as the energy and C-source.

Cultivations were carried out in a 10-L-fermenter (B. Braun’s Biostat C), operated in a fed-batch mode. The feeding solution consisted of glucose at a concentration of 200 [g/kg] and otherwise the same composition as the initial cultivation medium. After automatic inoculation over night, the process was first operated in a batch mode. After about 4 [h], substrate feeding was started. Induction with IPTG was performed after a predefined given time, concretely 10 [h] after inoculation.

B. Braun’s DCU was used as front-end controller. The entire process was monitored under control of MFCS-Win. The controller reported about in this article was operated on a separate PC connected to the MFCS-Win computer via a local network. Data exchange between MFCS and the controller was performed via a software interface allowing to read all relevant data from MFCS and to set values to all actors installed on the fermenter.

Process measurements

Vital to the success of process control are well performing measurement techniques allowing monitoring of process’ state continuously. Online, the off gas analysis was performed with a paramagnetic oxygen sensor (Maihak Oxor 610) and an infrared detector (Maihak Unor 610) for CO₂. pH was measured by means of an Ingold pH probe. The dissolved oxygen concentration was monitored with an Ingold pO₂ probe. Substrate addition was recorded by a Sartorius balance. Base addition is also measured quasi online. These signals were sampled all 36 [s].

Off-line measurements were performed with a time increment of about half an hour. Biomass concentration was estimated from an OD₆₀₀ measurement performed with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were performed in order to verify the correlation between biomass dry weight and the OD₆₀₀ values. A YSI 2700 analyzer was employed to measure the glucose concentration enzymatically. The GFP concentration was estimated from the fluorescence intensity measured from samples with intact cells in a spectro-fluorimeter (Hitachi F-2500). The OD₆₀₀ values as well as the glucose concentrations estimates are available about 5 minutes after sampling. The data were made available to the control program immediately after their values were known.

Basic Process Model

Obviously, in model predictive control algorithms, the process model is of prominent importance. The models suitable to process control must be as simple as possible in order to make sure that they can be validated with the restricted amount of data that can be made available in real industrial applications.
Generally these models are based on mass balances of the major players in the culture, the
values of which are significantly changing during the process. In protein production processes
these are predominantly biomass, product and culture weight. During the fed-batch process,
the substrate fed to the cells is most often consumed immediately. Therefore its concentration
is very small throughout most part of a fed-batch process. Instead of the substrate
concentration, the feed rate of the substrate is important variable for process monitoring.

In the process described here, the cultivation was started as a batch reaction and substrate
feeding was started only after its concentration in the medium fell below a predefined value.
There the main state variables are the concentrations $X$ and $S$ of biomass and substrate in the
cultivation medium, and the culture weight $W$. The culture weight is used instead of the
traditionally used culture volume as it is a variable that can be measured directly and online
by means of a balance. Correspondingly, as usual in industrial practice, all concentrations are
based on culture weight. After induction two additional variables get importance. Then we
assume the cells to be structured by taking the cell-internal product concentration $P_x$ as well
as the relative plasmid copy number $G_x$ as additional state variable. Hence the state vector
consists of two basic parts:

1. Concentrations with respect to the culture mass $c = [X \ S]$
2. Cell-internal concentrations with respect to cell mass $c_x = [G_x \ P_x]$.

Thus the model is the classical model for fed-batch processes that reads in vector notation

\[
\begin{align*}
\frac{dc}{dt} &= R + \frac{F}{W}(c_F - c) \\
\frac{dc_x}{dt} &= q - \mu c_x
\end{align*}
\]

(1)

c_F is the concentrations in the feed, $F$ is the feeding rate.

The complex metabolic network kinetics is lumped into simple models that describe the
specific growth or substrate consumption rate in terms of the macroscopic state variables:

\[
R = [\mu - \sigma] X
\]

(2)

$R$ is the net biochemical conversion rates of the components compiled in $c$. And

\[
q = [-\gamma \ \pi]
\]

(3)

$q$ is the corresponding cell internal, i.e. the biomass concentration related quantity describing
the dynamics of $c_x$.

$R$ is considered to be dominated by the specific substrate uptake rate

\[
\sigma = \sigma_{max} \frac{S}{K_S + S}
\]

(4)
which is specified by the classical Monod-type kinetic expression.

The corresponding specific growth rate is considered to be linearly dependent on \( \sigma \):

\[
\mu = Y_{XS} \sigma
\]  

(5)

As we are dealing with a fed-batch process, we must consider the weight changes:

\[
\frac{dW}{dt} = F = F_S + F_1 + F_2
\]  

(6)

\( W \) is the culture weight and \( F \) the total rate of change of the culture weight \( W \). Besides the substrate feeding rate \( F_S \), changes in culture mass through base addition, \( \text{CO}_2 \) outflow through the vent line, and evaporation must also be considered \( (F_j) \). The same applies for the amount of mass withdrawn for probing the state of the culture offline \( (F_2) \).

The variable essentially influencing the specific product development rate \( \pi \) is the specific growth rate \( \mu \). \( \pi \) is described as

\[
\pi = \pi_{\text{max}} G_x \frac{K_\mu}{K_\mu + \mu} - k_{\text{prot}} P_x
\]  

(7)

Where it is assumed that there is an inhibition effect with respect to \( \mu \) and a degradation effect which was described as a first order process with respect to the cell internal product concentration \( P_x \). \( G_x \) is the relative number of plasmids within the cells that carry the gene sequence for the desired protein. It became evident from the experiments that the number of active plasmids decreased during the process. This was described by the corresponding specific rate

\[
\gamma = \gamma_{\text{max}} G_y \frac{\mu}{K_\gamma + G_y} \frac{K_\mu + \mu}{K_\mu}
\]  

(8)

which states that the rate plasmid concentration changes depend on the plasmid concentration \( G_x \) and the specific growth rate \( \mu \).

The model parameters were identified by means of numerical optimization routines (Nelder Mead in combination with Random Search to avoid sticking in suboptimal regions) at several data sets obtained in the laboratory under similar experimental conditions as later during the controlled experiments. The model parameter values and other details about the process model are presented in (Jenzsch et al. 2006).
Chapter 5. Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria

State Estimator

Kalman filters are one-step-ahead predictors that make use of the last estimate, current measurement values and the prediction based on a model of the process. Weighting of the measured and modeled process values is based on covariance matrices for modeling and measurement noises. The model used here for estimation biomass concentration \( X \), culture weight \( W \) and the specific biomass growth rate \( \mu \) was simplified with respect to the model described just before, as not all of the dynamics considered are relevant for computing the estimates.

The simplified model connecting these quantities is the pure mass balance for the biomass, \( X \) and \( \mu \)

\[
\frac{dX}{dt} = \mu \cdot X - \frac{F}{W} \cdot X + v_X, \quad \frac{d\mu}{dt} = 0 + v_\mu
\]

(9)

where \( v_X \) and \( v_\mu \) are model inadequacies (model noises).

It is assumed that the specific growth rate is kept constant at its setpoint during the small time period, and the deviation from this assumption is considered as modelling noise. The required one-step-ahead prediction was performed using Euler’s simple integration rule. The actual values of \( F \) and \( W \) were measured during the cultivation and directly used during estimations.

The relationships between the state variables of the simplified model and the actually available measurement variables \( OUR \) (oxygen uptake rate), \( CPR \) (carbon dioxide production rate) and \( BASE \) (total base consumption during pH control), referred to as the measurement model, were

\[
\begin{align*}
OUR &= Y_{ox} \mu X + m_o X \\
CPR &= Y_{cx} \mu X + m_c X \\
BASE &= Y_{bx} (X W - X_0 W_0)
\end{align*}
\]

(10)

\( Y_{ox} \) is the oxygen consumption per unit biomass formed, \( m_o \) - the oxygen maintenance term, \( Y_{cx} \) - the carbon dioxide production per biomass formation and \( m_c \) - the corresponding maintenance term. \( Y_{bx} \) is the base per biomass yield, \( X_0 \) and \( W_0 \) are the initial biomass concentration and culture mass respectively. The parameter values of this measurement model are presented in (Jenzsch et al. 2006).

As Kalman filters also need estimates of the uncertainties of the model as well as on the measurement data, exemplified by covariance matrices, it is required to formulate and tune the matrices appropriately. This tuning process was done empirically, based on our experience with the application of extended Kalman filters to bioprocesses.
Model Predictive Controller

Once it is recognized that tight process monitoring can only be performed on a model aided basis it is straightforward to make use of this dynamic information for process control as well. From the many controllers proposed the moving horizon approach (e.g., Morari and Lee 1999) is particularly attractive as it comes much closer to the way in which living entities control their behavior, namely in a way considering the path of the system they expect for the near future.

![Figure 1](image-url)  
*Figure 1. Schematic of the time steps discussed in the text.*

At each time instant $t_n$, such a controller performs predictions for a time interval $[t_{n+1} \leq t \leq t_{n+h}]$ covering the coming $h$ time steps, with different assumptions about the action profile $F(t)$. From these predictions, the one that led to the smallest least square deviation from the desired control profile (dashed curve in Figure 1) is considered optimal. As the only relevant information needed for control is the action to be made at time $t_{n+1}$, the corresponding feed rate value $F(t_{n+1})$ is transferred to the valve in the feed line. Then the procedure is repeated for a time horizon moved one step ahead, i.e. for the interval from $t_{n+2}$ to $t_{n+h+1}$. In contrast to conventional controllers, the comparison between setpoint profile and process trajectory is performed for a finite interval instead for a single time instant.

In order to determine the optimal feeding profile $F(t)$ for the time interval $[t_n \leq t \leq t_{n+h}]$, the parameters of a polynomial approach to the rate function $F(t)$ of degree 2 were fitted to the desired control path within the interval. This was done by a simple direct search routine. Several tests showed that the Nelder/Mead algorithm (*Matlab’s fminsearch*), performed best in this particular application. As the parameters do not change dramatically from time step to
time step, such an optimization can easily be performed on a simple PC attached to the process control system within a second, even for the full model described above.

The information about the current value of the specific growth rate $\mu$ and the biomass $X$ is obtained by means of an extended Kalman filter algorithms using online measurements of OUR, CPR and the total base addition by the pH controller. The other state variables needed for model predictive control algorithms were estimated using the basic model equations (1-8). Figure 2 shows biomass and substrate estimations together with the offline measurement values that became available several minutes after sampling during one of the realized experiments. Obviously, the estimate closely matches the off-line data.

![Figure 2](image)

*Figure 2.* Evolution of the biomass and substrate concentration during controlled batch/fed-batch *E.coli* cultivation. Offline measured data (symbols) and model predictions (lines) are depicted.

When this model is used in a model predictive controller for the specific growth rate $\mu$ during an *E.coli* cultivation producing GFP the control does not work sufficiently well. A typical result is shown in Figure 3.

The reason for the deviations between the setpoint profile and the estimated specific growth rates was attributed to an additional problem appearing after the induction of the protein production at a fermentation time of $t_{\text{ind}}=10$ [h]. Due to the metabolic load change at that time, the biomass/substrate-yield changes. This change was not adequately considered in the model. Hence it is to be expected that the model supported controller will not work properly after induction. Figure 3 illustrates this situation.
Chapter 5. Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria

In principle, the offset in Figure 3 could be eliminated or at least reduced by the use of an integral action of the controller. However, the tuning of the gain for the integral part is quite complicated and can lead to controller’s instabilities. Therefore, this action was not considered further.

In order to avoid this problem, a simple adaptation procedure was incorporated into the controller. The actual biomass/substrate-yield was estimated directly from the substrate mass balance equation under steady state conditions in the fed-batch phase using the estimated values of $\mu$, $X$, $S$ and the measured values of $W$ and $F$.

$$Y_{XS} = \frac{\mu X W}{F_s \cdot (S_F - S)} \tag{11}$$

The value of the additional $S$ required to determine this yield was determined online using estimated value of $\mu$ and equations (4), (5). During the batch phase the $S$ concentration was computed using substrate mass balance equation and the estimated $X$ with a simple Euler integration approach, starting with the initial condition $S_0$ for the substrate concentration. As can be seen from Figure 2, this assumption led to good estimation results. The estimated biomass/substrate-yield was then filtered using a low pass filter and used in the model of the model predictive controller for estimating the feed rates. This simple procedure led to a significant improvement of the accuracy of the control as shown in Figure 4.
Chapter 5. Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria

Figure 4. Specific biomass growth profiles in μ-controlled E.coli cultivations together with the feeding profiles set by the controller to achieve the μ-profiles [a) experiment with an “exotic” setpoint profile for controller tuning; b) reasonable setpoint profile for a protein production process]. The adaptation of the yield $Y_{X}$ is shown to improve the performance with respect to the simple MPC algorithm that led to the result shown in Figure 3.
3 DISCUSSION

It was shown that recombinant protein production processes can be kept close to predefined setpoint profiles of the specific biomass growth rate $\mu(t)$. As the example displayed in Figure 4a shows, these profiles may be quite complex.

The model adaptation proposed in this paper is of interest in many production systems where genetically modified microorganisms are used to produce therapeutic proteins, as genetic stability may be a real problem in these systems. As $Y_{XS}$ often changes with time, its value cannot be easily changed by multiplying a constant. The adaptation mechanism described may help in many of these situations.

Complex step changes between piecewise constant specific biomass growth rate, as shown in Figure 4a, make sense in experiments designed to investigate the dependency of the specific product formation rate as a function of its principal influence parameter, the specific growth rate $\mu$. Reliable information about this $\pi(\mu)$-relationship is otherwise obtained from cultures operated in the continuous operation mode. Such experiments, however, would take much more time to obtain comparable results.

Acknowledgements. The work was performed in the project “Technologies for the production of recombinant therapeutic proteins” by Martin-Luther-University and local protein companies. We gratefully acknowledge the financial support of this project by the German Federal Ministry for Science and Education (BMBF).

REFERENCES


Chapter 5. Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria


Neidhardt, F.C., Ingraham, J.L., Schaechter, M. (1990), Physiology of the bacterial cell, Sinauer, Sunderland


