Advanced Monitoring & Control in Microbial Cultivation Processes for Recombinant Protein Production

Dissertation

zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)

genehmigt durch die

Mathematisch-Naturwissenschaftlich-Technische Fakultät
- Ingenieurwissenschaftlicher Bereich -
der Martin-Luther-Universität Halle-Wittenberg

von
Herrn Dipl.-Ing. Marco Jenzsch
geb. am 09. Dezember 1978 in Merseburg

Dekan der Fakultät: Prof. Dr. rer. nat. habil. H. Graener

Gutachter: Prof. Dr. rer. nat. habil. A. Lübbert
Prof. em. Dr. Dr. h.c. K. Schügerl

Halle (Saale), den 13. Juli 2006

urn:nbn:de:gbv:3-000010501
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000010501]

Danksagung

Diese Arbeit wurde im Zeitraum von November 2002 bis Mai 2006 am Zentrum für Bioverfahrenstechnik der Martin-Luther-Universität Halle-Wittenberg in der Arbeitsgruppe von Herrn Prof. Dr. Andreas Lübbert angefertigt.

Ich danke allen Mitarbeitern für das angenehme Arbeitsklima und die Unterstützung bei den verschiedensten Problemstellungen die sich während dieser Zeit ergaben.

Für eine hervorragende wissenschaftliche Betreuung danke ich Prof. Andreas Lübbert und Prof. Rimvydas Simutis. Die überaus fruchtbaren nächtlichen Brainstorming-Sitzungen bleiben unvergessen.

Ich bedanke mich insbesondere bei Herrn Prof. Lübbert für das entgegengebrachte Vertrauen bei der Übertragung wichtiger Industrieprojekte und bei der Betreuung von Diplomanden, sowie für die Freiheiten, die ich genoss und die unzählbaren Kongresse, die ich besuchen durfte.

Danken möchte ich Martina Anwand und Sylke Gutknecht, ohne deren Hilfe bei der Versuchsvorbereitung und Analytik, die große Anzahl an Experimenten alleine gar nicht zu bewältigen gewesen wären.

Ich danke Christian Sieblist, der immer ein verlässlicher Anlaufpunkt bei Software- und Computerfragen war. Obwohl er selbst meist knietief in eigener Arbeit steckte, hat er die Zeit für meine Probleme entbehren können.

Des Weiteren danke ich unserem Werkstattmeister Frank Ullmann, der all meine Ideen in praktischer Form professionell und mit einem Höchstmaß an Präzision und Funktionalität umsetzte.

Ich danke Frau Koch und Frau Homolya für die Motivationshilfen, die sie mir in Form von Schokolade und Gummitierchen immer wieder erfolgreich zuführten.

Bedanken möchte ich mich bei all den Projektstudenten und Diplomanden, die ich betreuen durfte und deren Projekte einen wesentlichen Teil zu dieser Arbeit beitrugen. Deshalb vielen Dank, Christan Berger, Silke Dietzmann, Stefan Gnoth, Markus Lange, Régis Lelou, Max Wellerdiek, Martin Werthschulte und allen anderen, die sich angesprochen fühlen.

Ich danke meinen Eltern für die finanzielle Unterstützung während des Studiums und für die nie versiegende Pipeline zum Kühlschrank.

Zuletzt möchte ich der Person danken, die es immer wieder geschafft hat, mit viel Liebe, leckerem Essen und einem offenen Ohr meine Gedanken zu ordnen. *Danke Kätchen!*

Contents

Chapter 1

Introduction

Abstract. Compared to the immense achievements in fundamental molecular biological sciences, the improvements in the fermentation and downstream processing technologies used in industry have been less spectacular over the last decade. Hence, there is a misbalance between new cellular systems and production technologies, resulting in a decreasing annual rate of approvals for protein manufacturing processes. In its process analytical technology (PAT) initiative, the U.S. Food and Drug Administration identified the issues that must be improved to compensate for this development and forces manufactures towards a more scientific approach of solving the problems. In this doctoral dissertation, methods of bioprocess engineering science have been used to meet the demands. Recombinant protein production processes, where *Escherichia coli* were used as host cells, are taken as a concrete example. Concretely, a design strategy for improved robust process operational procedures was developed that can be tightly supervised and automatically controlled.

Parts of this chapter have been published in *Engineering in Life Sciences*:
Jenzsch, M., Simutis, R., Lübbert, A. (2006), Optimization and Control of Industrial Microbial Cultivation Processes, *Eng. Life Sci.*, 6(2), 117-124

1 INTRODUCTION

In most industrial countries, biotechnology is a well recognized key technology for the years to come. Its products or services are applied in medical care, in food technology, in agriculture, and, increasingly, in fine chemistry. Currently most hopes are put into new pharmaceutical products, preferentially recombinant therapeutic proteins that allow curing diseases by directly combating the sources rather than only symptoms. Hence, much money has been investigated into the development of new biologics.

In contrast to the immense achievements in fundamental molecular biological sciences, the fermentation and downstream processing technologies used in industry have not been developed at the same pace. Since they did not receive as much public interest as the biological sciences, much less money and efforts flew into that domain. Hence, a misbalance between new cellular systems and production technologies appeared. This resulted in a decreasing annual rate of approved production processes over the last ten years (FDA 2005).

This development was well recognized and made public by the Food and Drug Administration (FDA) of the USA and essentially the same observation have been made by the European Medicines Agency (EMEA). As FDA is controlling most biologics production processes world wide, it has the best possibilities to judge about the state of the developments. The administration found that there are severe deficits in the production of pharmaceutical products. Consequently the agency is forcing manufacturers to pay more attention to the quality of their production processes. FDA does not only express its concern about stagnation in process development, it also made suggestions to the manufactures, e.g., to make use of recent developments in process supervision and control engineering sciences (FDA 2003). In its "process analytical technology (PAT)" initiative, FDA proposes concretely: (i) to make use of new measurement techniques for online supervision of the processes and (ii) to more exhaustively exploit the measurement data gathered from the processes to gain mechanistic understanding, predominantly knowledge about the interrelationships between the various process variables. According to the FDA-initiative, measurement data should (iii) be used online in order to recognize deviations from 'in control situations' before the processes run out of control. And (iv) in the case of significant deviations from the desired setpoint profiles, the processes must be drawn back to the predefined path by automatic feedback control in the engineering sense.

One of the most important new developments in the strategy of the FDA is that it now focuses attention on a consequent utilization of current mechanistic process know-how for continuous improvements of the process. This is an important paradigm change in so far as previously, approved processes were only changed if it becomes necessary by safety arguments. Otherwise the processes were operated according to fixed standard operational procedures (SOPs).

In future, the processes should be improved whenever a quality relevant improvement is possible. Particularly important is the development of robust process operational procedures, a requirement that was formulated by leading process quality experts in other fields since several decades (e.g., Taguchi 1981, 1987, and Shainin *et al.* 1988).

In this dissertation it is shown at several concrete examples, how many of the requirements of the FDA can be realized and installed at industrial production plants everywhere in biotechnology, in pharmaceutical as well as in fine chemicals industries, with reasonable expenditures. The examples mentioned are from the domain of recombinant protein production, but the results can be used in other domains, e.g. in the production of industrial enzymes or biotransformations with whole cells, as well. Concretely, the development of

process operational procedures will be discussed that allow a more tight supervision of the processes and an automatic control in cases where processes deviate from their setpoint profiles.

2 QUALITY BY DESIGN OF OPERATIONAL PROCEDURE

The quality of production processes and, thus, the quality of the products as well, is largely determined by the batch-to-batch reproducibility of the fermentations. In Figure 1, a couple of biomass concentration profiles arbitrarily chosen from a protein formation process are shown. Different repetitions of the process under the same operational conditions usually lead to significant deviations from the long-term mean. This can be characterized statistically by the standard deviation σ or the variance σ^2. It is obvious that a process depicts a higher batch-to-batch reproducibility or quality when this standard deviation from the mean is small. An essential goal in the realm of process quality assurance is reducing this standard deviation.

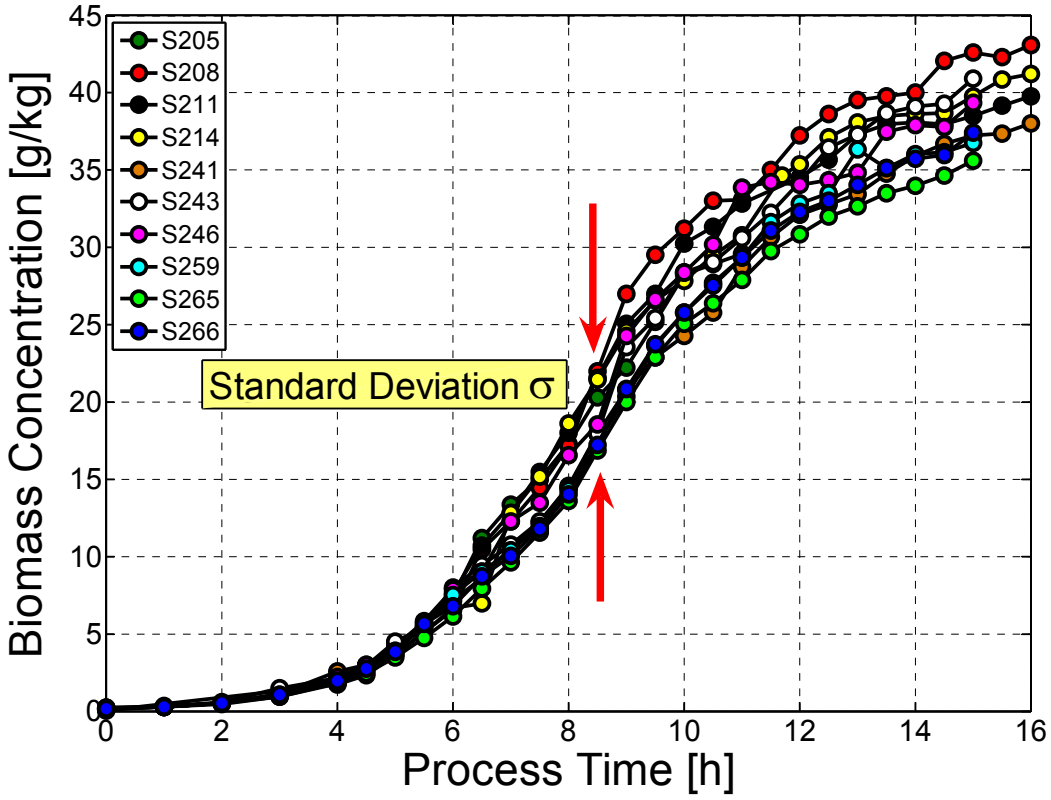

Figure 1. Profiles of biomass concentrations as a function of time for a set of *E.coli* fermentation runs performed under practically the same operational conditions.

This goal can be obtained using numerical process simulation, utilized in the sense of a sensitivity analysis. For this purpose, different possible alternatives for process operation can be investigated. In a first approach the influence of those process variables that are prone to be distorted in real fermenters on the standard deviation σ should be investigated.

In the example presented in Figure 1, it turned out that the standard deviation of the biomass concentration at induction time (t_{ind}=8 [h]) is primarily dependent of the amount of cells used for inoculation. Furthermore, it depends on the specific growth rate μ of these cells, i.e. the state at which they are harvested from the preculture.

In order to explore the influence of the biomass concentration at t=8 [h], the initial biomass concentration X_0 was varied and the growth conditions were searched for that lead to the smallest fluctuations at induction time. In such sensitivity analyses one can vary the initial biomass concentration, for instance in the interval X_0=0.2 … 0.5 [g/kg].

As fast growth is economically required in this initial biomass growth phase, it is straightforward to feed at every time instant the substrate replace that is currently consumed by the cells. This is realized by means of an exponential substrate feeding strategy. The simulations performed during the sensitivity analysis show, that - as opposed to current industrial practice - biomass should not be allowed to grow at maximal growth rate μ_{max}. The desired specific growth rate μ_{set} should be smaller that μ_{max}. This means an exponential feed rate $F(t) = F_0 \cdot \exp(\mu_{set} \cdot t)$ with $\mu_{set} < \mu_{max}$, where F_0 is known to be a function of X_0. Obviously, before the process is started, a decision about the feed rate profile must be made and installed in the programmable controller of the feed valve at the fermenter.

When then the initial biomass concentration X_0 is smaller than the one used to compute F_0, then each cell will recognize a higher substrate concentration than expected and it will thus increase its specific growth rate μ. In this way the process automatically compensates for the smaller initial biomass.

When the initial biomass concentration is too high, the individual cells will recognize a smaller substrate concentration and will reduce their growth speed. Again the process automatically removes the deviation. This manifests process robustness.

When, however, the fermentation would be operated with a feed profile corresponding to $\mu_{set}=\mu_{max}$, things run quite differently. When then a deviation in X_0 towards smaller values appears, the cells cannot increase their growth speed as they are already running at maximal specific growth rate. Hence, all these deviations lead to different growth profiles and finally to a significant standard deviation in the biomass concentrations at induction time t=8 [h] as can be seen in Figure 2.

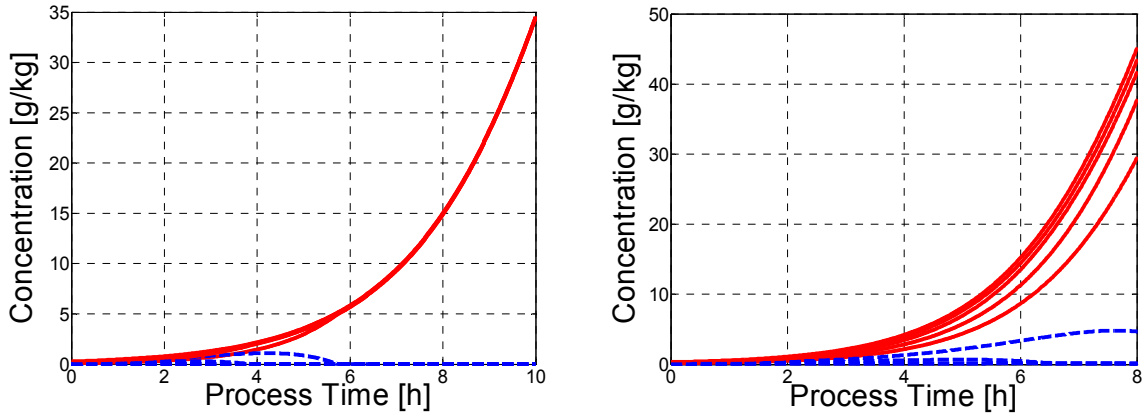

Figure 1. Influence of the initial biomass concentration on the variance of the biomass concentration at the end of the growth phase of a protein production process (biomass (full) and substrate concentrations (dashed)). Left: Fed batch cultivation with exponential feeding for a smaller specific growth rate. Right: Operation at maximal specific growth rate μ_{max}.

The goal is not only just to optimize the biomass at the end of the growth phase, but also to reduce the sensitivity of the operational procedure to uncontrollable factors or noise. Using numerical simulation experiments based on a reliable process model, such control actions can be searched for in a systematic way without too many expensive culture experiments in the laboratory. Needless to say that the computational results must finally be validated experimentally in order to make sure that they are correct (Jenzsch *et al.* 2006a).

This procedure of finding robust process variants by a combination of model-based simulation and experimental validation experiments is referred to as "quality-improvement-by-design". It is one of the main requirements of the FDA formulated within its PAT initiative (FDA 2003). The example presented here shows how this requirement can be met in fermentation technology practice.

3 PROCESS SUPERVISION

Currently, biotechnical production processes are generally operated with open loop control. The programmable controllers at the fermenter are programmed initially with the desired profiles of the manipulated variable. Most often this is the substrate feeding rate profile F(t).

In order to make sure that the process will follow the predetermined profile of the controlled variable, for instance the biomass concentration profile X(t), its current value must be measured online during the process. As there are no reliable online measuring devices available for most of the state variables and particularly not for the biomass concentration, indirect measurements are required.

In its PAT-Initiative, FDA suggests to use methods of statistical process control, particularly multivariate regression techniques. With such techniques, the information content in the signals from more than two variables is simultaneously exploited. In fermentation processes, it is straightforward to use the signals from the global measurement variables such as the oxygen and the carbon dioxide concentrations in the off gas as well as the total amount of base (NH_4^+) employed in pH control. From these signals, a quite accurate estimate of the actual biomass concentration value can be obtained.

A simple example for such a multivariate regression approach is:

$$X = a_0 + a_1 \cdot \sum CPR + a_2 \cdot \sum OUR + a_3 \cdot \sum Base \qquad (1)$$

Figure 3 shows the results of such a multivariate estimation of the biomass concentration based on 16 data records from protein production experiments with *E.coli* bacteria (Jenzsch *et al.* 2006b). It is obvious that even the simple linear regression with cumulative online measurement data lead to quite reliable online estimates of X.

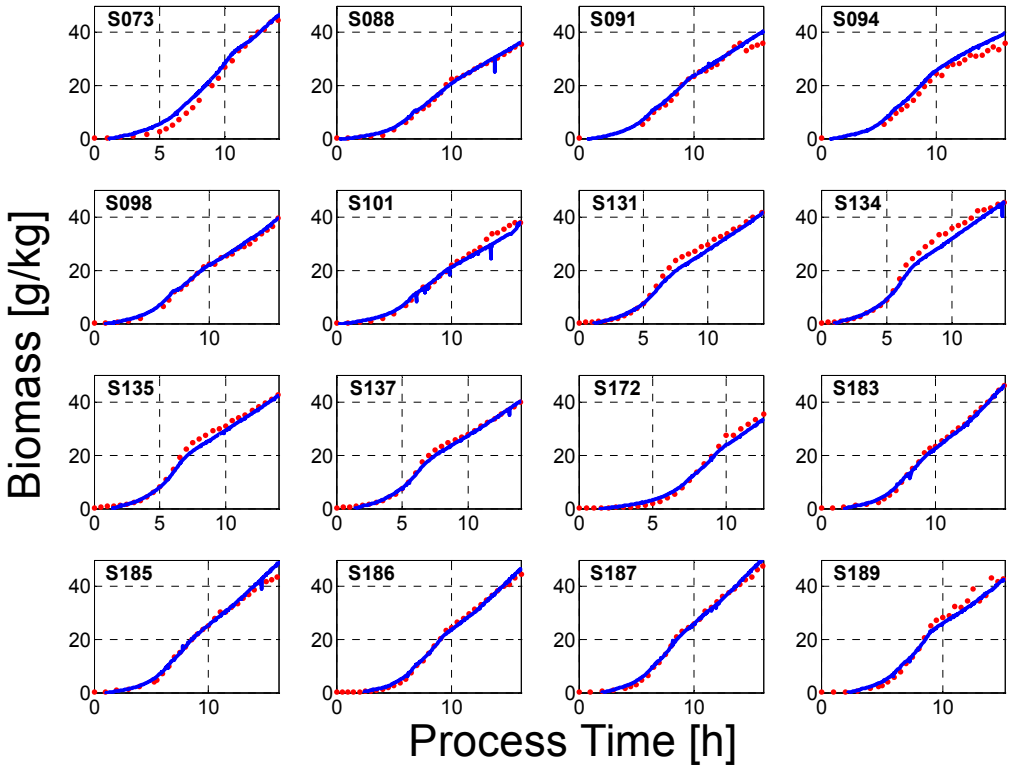

Figure 2. Linear cumulative regression used to estimate the biomass concentration during fermentation runs. 16 biomass concentration records from fermentations operated under similar conditions are depicted. The full lines are the estimates, the symbols the measurement values. The estimates are immediately available; the off-line measured values are available only after a longer time delay. The root-mean-square value of the deviations from the mean taken over all fermentations is 1.18 [g/kg].

If one takes into account that fermentation processes are known to be nonlinear, and so the relationships between the various influence variables, one must expect that nonlinear extensions of the simple correlations will lead to improved fits to the available data. As the numerical evaluation shows, the estimation error becomes significantly smaller even when only a simple quadratic regression model is used in the form

$$\begin{aligned} X = {} & a_0 + a_1 \cdot \sum CPR + a_2 \cdot \sum OUR + a_3 \cdot \sum Base + \\ & a_4 \cdot \left(\sum CPR\right)^2 + a_5 \cdot \left(\sum OUR\right)^2 + a_6 \cdot \left(\sum Base\right)^2 + \\ & a_7 \cdot \sum CPR \cdot \sum OUR + a_8 \cdot \sum CPR \cdot \sum Base + \\ & a_9 \cdot \sum OUR \cdot \sum Base + a_{10} \cdot \sum CPR \cdot \sum OUR \cdot \sum Base \end{aligned} \qquad (2)$$

The corresponding RMS error of the biomass estimation is 0.76 [g/kg] which is significantly lower than the one obtained with the linear approach. This result suggests using more flexible nonlinear approaches of multivariate data analyses for biomass estimation. A very universal nonlinear mapping technique is the method of artificial neural networks. When a larger set of data records is available, this technique leads to very accurate and reliable estimates.

The structure of the artificial neural networks that can be employed for this application can be kept quite simple so that the number of free parameters (network weights) can be kept rather low. In the example shown here, a network with three input nodes (OUR, CPR and cumulative base consumption) is taken. A simple feed forward architecture is sufficient. Here a network with a single hidden-layer containing 4 nodes and a single output node for the quantity to be estimated, the biomass concentration X was used. For network training, i.e. the search for optimal network weights, practically every computer program for nonlinear optimization can be used. In the examples shown, the Levenberg-Marquardt routine, available in many program libraries, e.g. Matlab, was employed. The fit was tested using cross validation procedures.

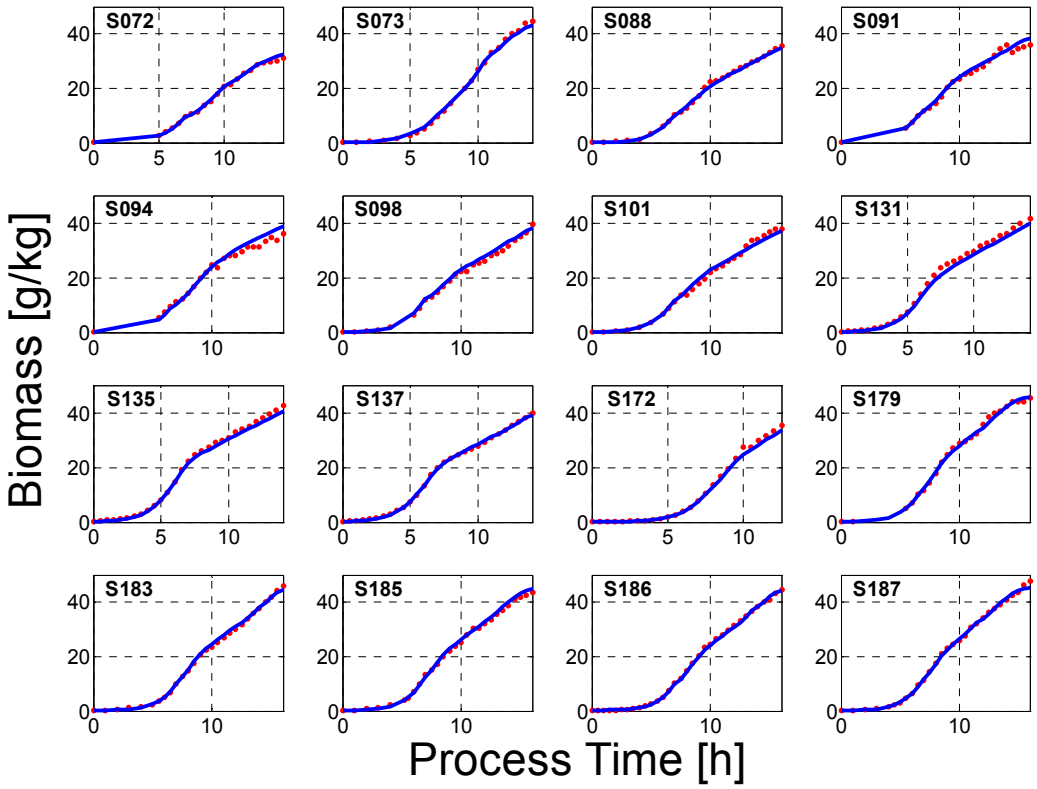

Figure 4. Estimation of the biomass concentration from the online measured signals of OUR, CPR und Base by means of a simple artificial neural network.

The result of the estimations with ANNs, applied to the same set of biomass concentration records as shown in the previous section is depicted in Figure 4. The application of artificial neural networks led to a significant improvement of the goodness of estimation. The RMS-value dropped to 0.46 [g/kg].

These examples show impressively that nonlinear multivariate regression techniques and cross validation lead to very reliable estimates of the biomass from the actual values of the variables measured online during the cultivation processes. This allows determining at every time instant online the current value of the key quantity biomass concentration X. This is a fundamental prerequisite of quality assurance in the sense of the PAT initiative of the FDA.

When it is possible to detect whether or not the process is in control, the consequent next question is what to do when an “out-of-control-situation” is detected. What are the possibilities to avoid significant deviations from the desired setpoint profile? For that purpose,

automatic feedback control is the matter of choice. In the PAT initiative of the FDA control in the engineering sense is demanded.

4 CONTROL OF THE SPECIFIC BIOMASS GROWTH RATE

When dealing with process control that process variable should be controlled that most sensitively influences the process performance. In the case of producing high value products, the amount of product that can be delivered to the downstream processing department after a given fermentation period [0 … t_e] is a good criterion.

The mass of product available at $t = t_e$ is dependent on the biomass x [kg] and the specific product formation rate π [kg product/kg biomass/h]:

$$m_P = \int_0^{t_e} \pi \cdot x \cdot dt \tag{3}$$

While biomass x is an extensive measure for the number of cells that can be employed for product creation, the specific product formation rate π is a measure of their mean performance. Needless to say that both variables must be kept at as high values as possible during the entire product formation phase in order to obtain a high product mass m_P.

In practice one often aims in running the process at predetermined profiles of the specific biomass growth rate μ as already discussed in the section about robust process operation, since both key variables, x and π, are primarily dependent on μ, as it is shown in Figure 5 for the concrete example of the expression of green fluorescent protein (GFP) in *E.coli*-host cells. Thus, it is straightforward to use μ as control variable, as then it is possible to keep the other state variables close to their setpoint profiles as well. The most convenient manipulated variable, by which the process state is influenced, is the substrate feed rate F(t) [kg/h].

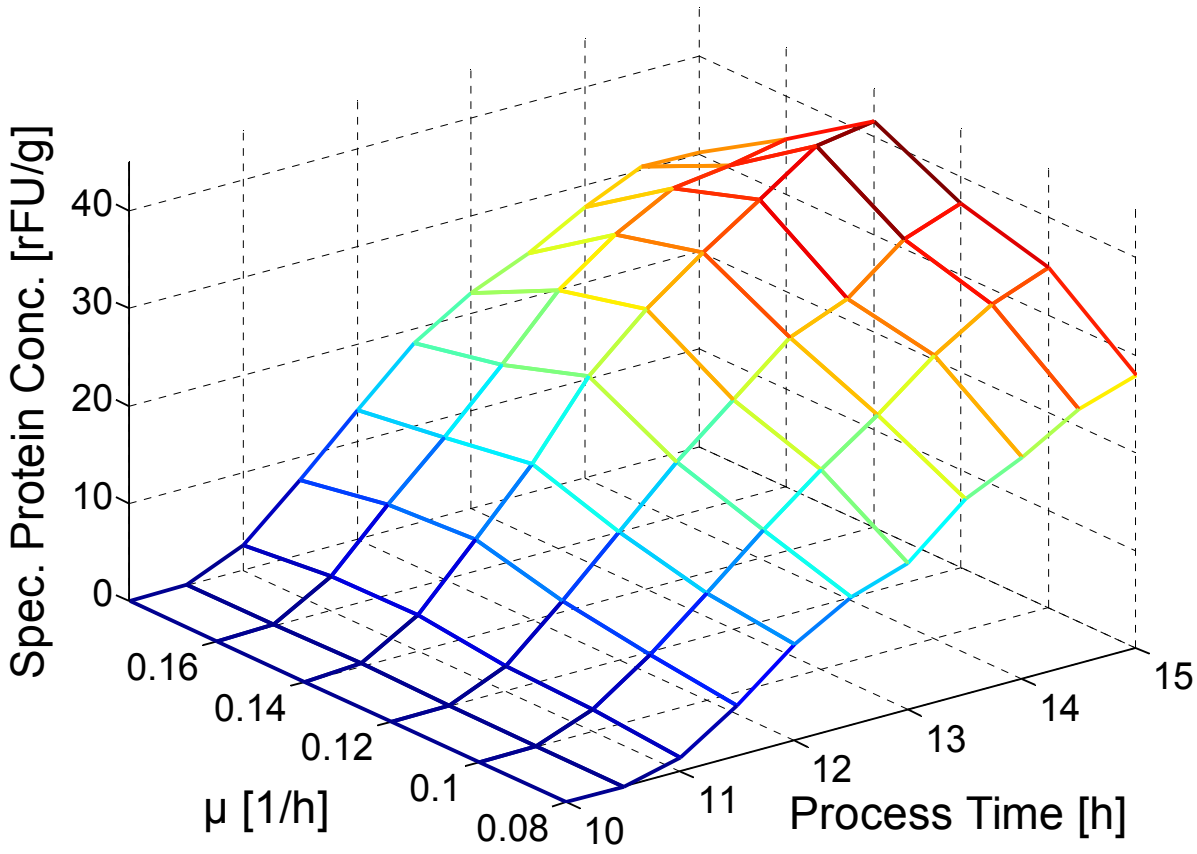

Figure 5. Specific target protein concentration profiles from 6 fed batch fermentation runs performed with different specific growth rates (0.08 … 0.18 [1/h]) after induction at 10 [h]. As can be seen, there is a real maximum in specific protein load of the cells at μ=0.14 [1/h].

Unfortunately, as control algorithms one cannot make use of the well-known simple PID algorithms. The reason is that the dynamics of bioprocesses is continuously changing and thus the optimal parameters of the PID controllers. This reduces the controller performance so much that they cannot be used in practice. This problem can be solved by adapting the controller parameters to the changes in the process' dynamics. However, for that purpose the changes must be estimated online during the cultivation. Thereto one needs a model describing the changes as a function of the current state of the process.

A rather simple control algorithm that is able to make use of such model information to adapt the controller parameters is the "Generic Model Control" (GMC). As could be shown experimentally (Jenzsch *et al.* 2006c) GMC controllers are able to keep the fermentation process on a profile $\mu_{set}(t)$ of the specific biomass growth rate. For that case the GMC algorithm provides a simple explicit expression for the manipulated variable F(t):

$$F = \left(\frac{\mu_{set} \cdot X \cdot W}{Y_{XS} \cdot S_F} + \frac{k_1 \cdot (\mu_{set} - \mu) + k_2 \cdot \int_0^t (\mu_{set} - \mu) dt}{Y_{XS} \cdot \sigma_{\max} \cdot \frac{(K_S - S^2/K_I)}{(K_S + S + S^2/K_I)^2}} \cdot \frac{W}{(S_F - S)} \right) \tag{4}$$

Where W is the reactor weight, S the substrate concentration, Y_{XS} the biomass per substrate yield, π_{max}, K_I and K_S are parameters of the general model for the substrate consumption kinetics. As becomes clear by inspection equation 4, F(t) is composed of two main terms: the first is a feed forward control expression leading to an exponential feeding profile corresponding to the setpoint value μset of the specific biomass growth rate. And the second term is a feedback controller which reacts on deviation of the current specific growth rate from its setpoint μ_{set}. As is easy to see from the numerator of this term, this controller is a PI controller with parameters k_1 and k_2. Both these parameters however are adapted to the current dynamic state of the process by their common denominator, which is essentially reflecting the substrate consumption kinetics in the culture.

The results of two protein production experiments controlled with such a generic model controller are depicted in Figures 6a and 6b. In the lower part of the figures, the μ-setpoint profiles are depicted together with the μ estimates made during the cultivation. In the upper part of the figures, the feeding rate profiles F(t) are depicted that were provided by the controller in order to keep the process on target.

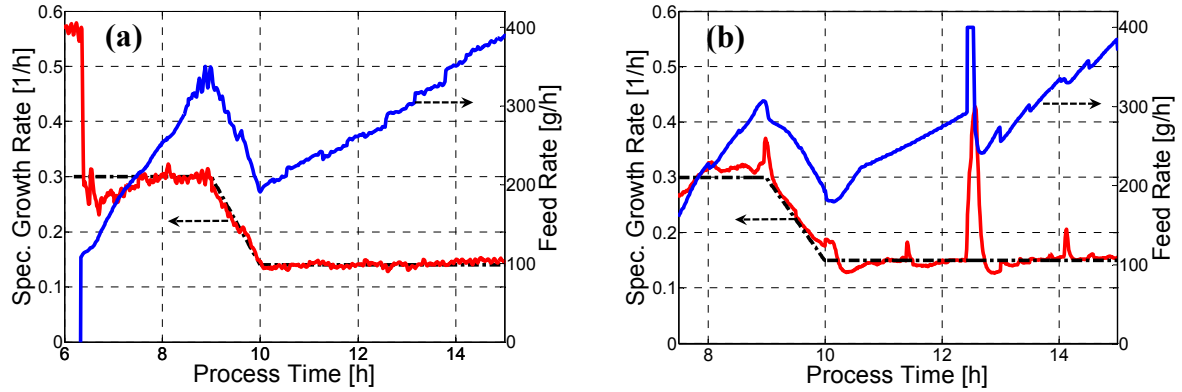

Figure 6. (a) Typical result of a μ-controlled culture producing a recombinant protein. In the upper part the manipulated variable, the feed rate F(t) is shown. In the lower part the controllers performance in keeping the process close to the predetermined μ(t) setpoint profile (dashed line) is depicted. The noisy line shows the online estimates of μ(t). (***b)*** The same results as shown in *(a)*, but for an experiment where defined process disturbances were arranged, e.g. at 12.5 [h], where the control pump for substrate supply was set on full power for 5 minutes. As can be seen the controller is able to bring the process back on the predefined optimal setpoint profile, even after such an intense disturbance.

These results clearly show that control of physiological key variables such as the specific biomass growth rate μ is possible with a quite high accuracy. In other projects this was also shown with different organisms and with different reactor sizes up to the large production level.

Other, even more sophisticated closed-loop controller, have been investigated, such as the model predictive control (MPC) algorithm. The increase in intricacy necessary to apply such sophisticated algorithms however did not yield in a comparable increase in controller performance (Jenzsch *et al.* 2005).

5 CONTROL OF THE TOTAL BIOMASS

Obviously, the final mass of target protein m_P should be kept as high as possible in an industrial manufacturing process; however in a production environment, robustness of the process behaviour with respect to occasionally appearing disturbances is highly important as well. Reproducibility is a very important constraint as it affects the downstream processing and thus quality of the final product. Direct control of μ was shown to be possible in recombinant production processes (Jenzsch *et al.* 2005, 2006c). This works perfectly as long as there are no severe disturbances in the process. When, however, some disturbances leads to a significant deviation of the biomass from its desired path, one must correct for it before one can proceed with the desired optimal or quasi–optimal μ-profile. Hence, one must look for more robust alternatives that are able to keep the process more tightly on the desired profiles of μ and other state variables that are important for process quality. From equation (3) it is straightforward to consider biomass x first.

Using start biomass and desired specific biomass growth rate profile it is easy to estimate total biomass profile during the cultivation

$$\frac{dx}{dt} = \mu_{set} \cdot x \tag{5}$$

A given biomass profile $x_{set}(t)$ is then in a close relationship with the specific biomass growth rate μ. Hence, controlling the process to an x profile should satisfy the corresponding specific growth rate profile $\mu_{set}(t)$ as well. In this case the cultivation process is more robust, because the deviations in biomass concentration can be eliminated by controlling the integral variable x.

μ as well as x cannot be measured directly with sensors that work reliably at a production fermenter. Both can be measured indirectly. As the estimation of biomass is much more reliable, x was chosen as the controlled variable. For indirect x-estimation, an artificial neural network was applied (Jenzsch *et al.* 2006b). The ANN was trained on 26 data sets measured during a process development period.

Preliminary simulations and experiments showed that the total biomass x(t) is better suited as the controlled variable than the biomass concentration X(t) itself. The control can then be performed with a simple adaptive control algorithm (Jenzsch *et al.* 2006d).

The controlled biomass profiles depicted in Figure 7 show that this simple control approach leads to a very good reproducibility of the total biomass profiles.

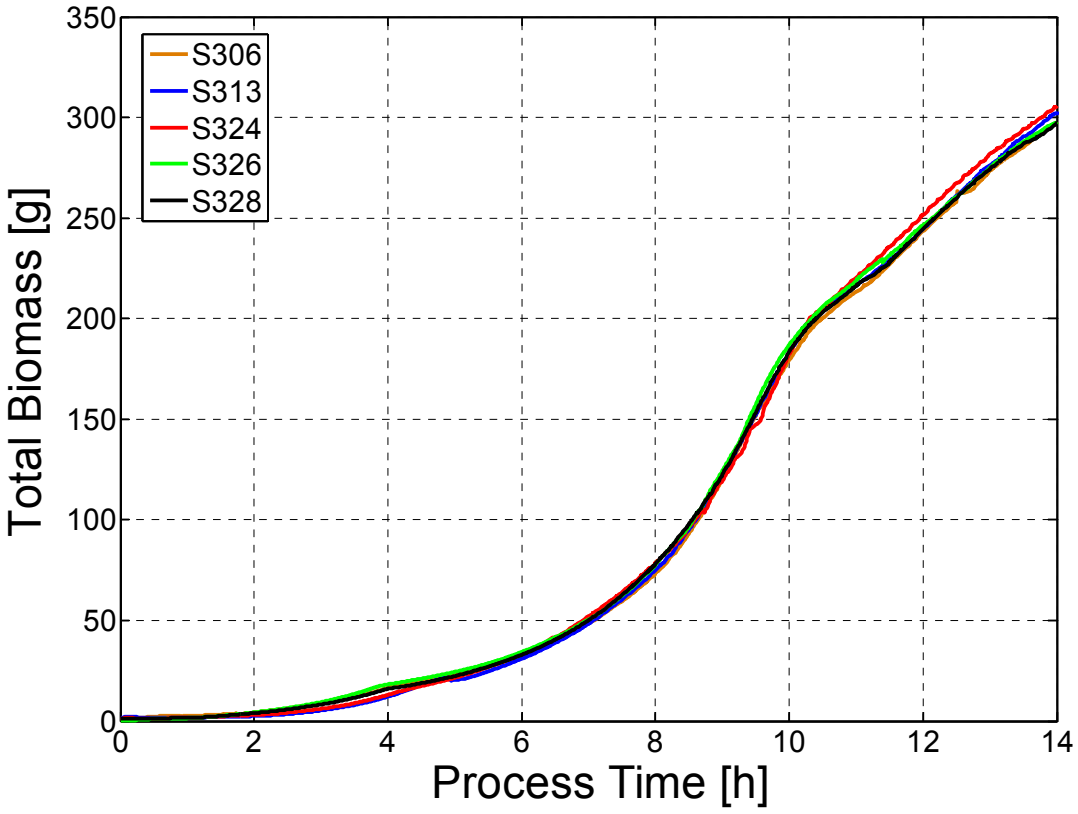

Figure 7. Total biomass signals from 5 fermentations performed sequentially using the same setpoint profile.

6 CONTROL OF THE TOTAL CARBON DIOXIDE PRODUCTION

The number of production reactors in biotechnology where process control by manipulating the feed rate profile is realized is negligible. It seems to be a rather big step to install process control in industry. Many of the control procedures that were proposed in literature are rather complex do not really downsize the activation barrier. Thus, simple, easy-to-use regulation procedures are required.

The main problem, however, remains that π, μ and even x-based control suffer from the fact that they cannot accurately enough be measured directly with sensors that can be installed at a production fermenter.

Apart from the difficulty of making available online values of μ, in a μ-controlled fermentation a further problem is the instability of that approach with respect to variations in the initial total biomass amount x_0. As depicted in Figure 8a, an initial deviation in x will be amplified when a fixed μ profile will be applied. Hence, merely keeping the process at a fixed $\mu(t)$ leads to an unfavourable batch-to-batch reproducibility.

This problem can be solved by relating the μ profile to a CPR(t)-profile and running CPR along the resulting trajectory by means of feedforward/feedback control. This variable that is tightly related to the specific growth rate and can easily be measured online. Instead of the carbon dioxide production rate CPR, one can also use the corresponding total CPR signal

tCPR=CPR·W, where W(t) is the corresponding culture mass signal, or even better, the total accumulative signal tcCPR of the total carbon dioxide production rate. The proposed method leads to a self-tuning, i.e. a robust process behavior. This is shown in Figure 8b. When the initial amount of total biomass is not on target, the specific growth rate will be adapted automatically. When the inoculated biomass is too high, the cells will produce more CO_2 than expected; hence the controller will reduce the feed rate and thus substrate concentration. Consequently, the biomass growth rate will become smaller. If on the other hand the initial amount of total biomass is smaller than expected, the cells will produce less CO_2 and the controller will increase the feeding rate. This will lead to an increase in the biomass growth rate. After a few hours, the specific growth rate is on target simply by controlling the total cumulative CPR to its corresponding profile.

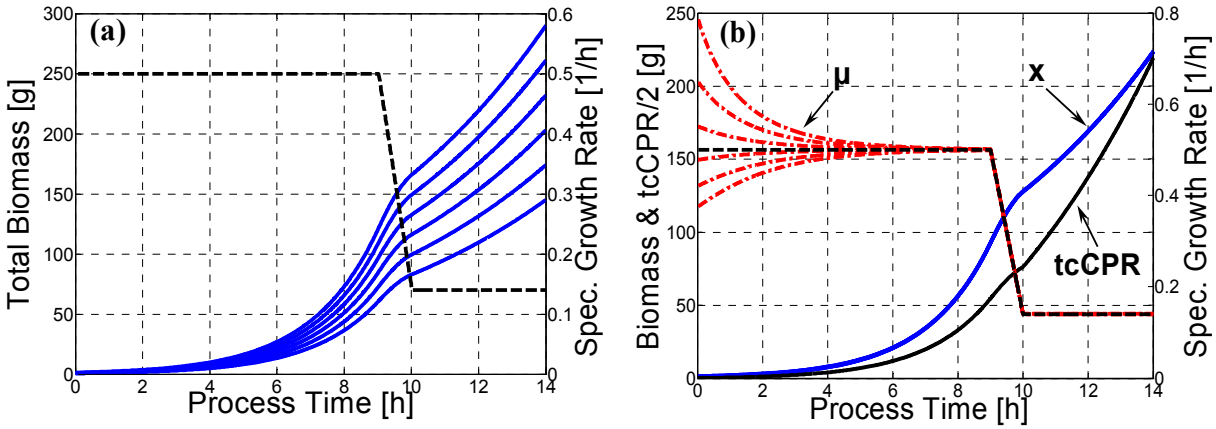

Figure 8. (a) Simulation of fed batch processes running with a fixed growth rate profile (dashed line) showing the effect of an initial variance in the total biomass x_0=0.65 … 1.35 [g] (full lines). The variance becomes amplified with time. (***b)*** This simulation shows the self-stabilizing effect of controlling total cumulative CPR in fed batch cultivations of a recombinant therapeutic protein production process. The same variations in x_0 were assumed.

As shown in the coarse simulation study depicted in Figure 8b, the adaptation appears in the first few hours. The essential effect of the adaptation the reduction in the variance of the biomass trajectories appears during the protein production phase, i.e. in the second half of the entire process time.

7 CONCLUSIONS

Fermentation processes in the manufacturing of biologics are known to fall short with respect to batch-to-batch reproducibility. In this regard there is a considerable lag as compared with manufacturing processes in other industries. Fermentations are running on practically the same level of control as 20 years ago. Since the product quality is dependent on the reproducibility the accrediting administrations demand improvements. For that purpose the FDA suggests to first analyze the processes more accurately in order to improve the mechanistic knowledge about the process dynamics and then to draw consequences in terms of robust design of process operational procedures, as well as process supervision and feedback control in the engineering sense.

This work focuses on the suggestions made by the FDA and shows how significant improvements of the batch-to-batch reproducibility can be obtained at recombinant protein production processes with *E.coli* cultures. Priority number one in the arsenal of methods is design of robust operational procedures. Here it is shown how the biomass concentration at induction time can be kept close to a given target value.

Obviously random distortions can nevertheless disturb the process. These can only be eliminated by automatically controlling the process with feedback controllers. These, however, require a well performing process state estimation technology.

For this purpose the multivariate data analysis techniques proposed by the FDA were shown to do a good job. However, more powerful than the established techniques of nonlinear multivariate data analysis are, as shown in this work, artificial neural networks. These proved to be well performing and stable, provided enough process data records were used for its training. The latter is no real problem in manufacturing plants where the processes running very often under roughly the same operational conditions.

The proposed specific growth rate control procedure is quite universal. It includes the state estimation algorithms for basic state variables and the generic model control algorithm providing the substrate feed rates used as action variable at the controller. The procedure allows realizing complicated specific growth rate profiles during the cultivation. This feature is very important for obtaining the information about the process' dynamics needed for bioprocess optimization, in particular the dependency of the specific product formation rates from the specific biomass growth rates.

With a generic model controller a good control performance was obtained. However, this requires a well performing process model. Whenever the process is changed, at least the model parameters must be adapted by identification procedures with a set of new data. This might be too a big expenditure for many manufacturers.

From the cell-physiological point of view the specific growth rate would be the most important variable to be controlled in fed-batch fermentation. Thus much work was put into control procedures to put μ-control into practice. However, this may lead to problems from the process reproducibility point of view. If batch-to-batch reproducibility is an issue, and this is what FDA is demanding from drug manufactures, control of biomass is clearly superior. In industrial practice this does not mean that the primary idea of control guiding the process along favourable profiles of the specific biomass growth rate must be abandoned. Instead, from the μ-profile a corresponding x-profile should be determined and this can then be controlled in a feedback fashion.

In a further approach, the total cumulative CPR was chosen as control variable, which has several important practical advantages. It is first of all robust as compared to the biomass concentration profiles. Following a profile of the total cumulative CPR (tcCPR) automatically corrects for deviations from the biomass concentration profile. The second important point is that tcCPR can quite accurately be measured online. CPR as measured via the usually applied offgas analysis is a global quantity that does not depict such heavily fluctuating signals as any variable locally measured with a probe within the multiphase flow of the culture. Finally, as a global measurement it is more representative than a locally measured variable, a property that is necessary in our models that all assume homogeneity or ideal stirred tank conditions in bioreactors. Hence, keeping the cultures operated during recombinant protein production on predefined tcCPR profiles, that are derived from the desired profiles of the specific growth rate $\mu(t)$ is a very good strategy for keeping fed batch processes on the optimal track.

Many control strategies have been discussed in literature but there is practically no industrial plant where control of variables other than pH, T, and may be pO_2 has been realized. Most of

the control approaches published so far are rather complex and there is some fear in industry to apply them. From that point of view, the control procedures developed during the work reported about here are extremely simple adaptive approaches. They have been tested in more than a hundred fermentation processes where recombinant proteins were generated with *Escherichia coli* and *Pichia pastoris* cells. It proved to be very robust and practicable even in large-scale bioreactors.

Upon the appearance of FDA's PAT-Initiative there is no longer any excusing for manufacturers the stay away from automatic control of the feeding to their fermenters. This is particularly true for production processes of biologics where the processes are approved together with the products themselves. It now becomes a must for manufacturers not only in pharmaceutical industries, to invest in knowledge-based process optimization and control.

REFERENCES

FDA (2003), *Guidance for Industry: PAT - A Framework for innovative pharmaceutical manufacturing and quality assurance*, August 2003, Food and Drug Administration, Rockville, MD

FDA (2005), FDA's Critical Path Initiative: History, Objectives, Approach, June 21, 2005, http://www.fda.gov/oc/initiatives/criticalpath/presentations/bio200501_files/slide4.html

Jenzsch, M., Simutis, R., Lübbert, A. (2005), Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9),* Pons MN, van Impe JFM, eds., IFAC/Elsevier, ISBN 0 08 044251 X

Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006a), Open loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J. Biotechnol.*, in press

Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A. (2006b), Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioproc. Biosyst. Eng.*, 29(1), 19-27

Jenzsch, M., Simutis, R., Lübbert, A. (2006c), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006d), Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile, *Bioproc. Biosyst. Eng.*, submitted

Shainin, D., Shainin, P. (1988), Better than Taguchi orthogonal tables, *Quality and Reliability Eng. Int.*, 4, 143-149

Taguchi, G. (1981), On-line quality control during production, *Japanese Standards Association*, Tokyo

Taguchi, G. (1987), System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs, *Vols. 1 & 2*, UNIPUB/Kraus International Publications, White Plains NY

Chapter 2

Open Loop Control of the Biomass Concentration within the Growth Phase of Recombinant Protein Production Processes

Abstract. Recombinant protein production processes are typically divided into two phases. In the first one, pure cell propagation takes place, while in the second one product formation is switched on within the cells by adding an inducer. In the initial biomass formation phase, the cell density is rather low and, hence, the measurement quantities that could be used to determine the process' state depict small values and are rather severely distorted by measurement noise. Because of these measurement problems, the fermentation cannot be reliably controlled by feedback control during this first production phase; instead, the process must be controlled in an open-loop fashion. The consequence, worked out in this paper, is to design substrate feed rate profiles for the growth phase in such a way that they are robust with respect to the main disturbances observed in practice. The robustness of the biomass formation is shown to be primarily dependent on the specific growth rate adjusted in the first hours. High batch-to-batch reproducibility can be obtained with exponential feeding profiles F(t) corresponding to specific growth rates μ_{set} well below the maximal specific growth rate μ_{max} of the organism. The reduction in the growth rate needed to obtain a robust process behavior depends on the inaccuracies in the initial biomass concentrations. Quantitative feed rate profiles were obtained by numerical simulation and these results were validated experimentally by means of a series of cultivation runs, where a recombinant pharmaceutical protein was produced. All experimental data confirmed the assumptions made in the robust process design study.

This chapter has been published in *Journal of Biotechnology*:
Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006), Open loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J. Biotechnol.*, *in press*

1 INTRODUCTION

Biologics are different from traditional drugs in several important respects. They are not only greater in size and more complex; also there are inextricable relationships between product properties and the biological processes that are used to make them. In other words, the therapeutic efficacy of a biologic drug may be dependent of variations in the operating procedure of the production process (e.g., PhRMA 2001). Particularly the fermentation part must be performed in a tightly reproducible way in order to guarantee product quality. This is the reason why the approval of these proteins as drugs is restricted to predefined production processes. Thus, in recombinant therapeutic protein production, measures that increase the batch-to-batch reproducibility have a direct impact on product quality.

There are two principal approaches to improve batch-to-batch reproducibility as underlined in the PAT initiative of the FDA (FDA 2003). The first is increasing the robustness of the operational procedure, i.e. to gain understanding of the process and use it for construction of process trajectories that are widely immune against common distortions. The other is feedback control in order to remove deviations of the actual trajectories from the desired one. Both approaches are complementary to each other.

Common industrial practice is to produce recombinant therapeutic proteins with organisms carrying inducible expression cassettes. These processes are divided into two phases, a biomass growth phase, which is usually operated by supplying optimal growth conditions to the cells, followed by a product formation phase, where the conditions are controlled in such a way that the cells can produce the desired product at a high formation rate over a sufficiently long time.

Here we focus our attention to the first phase where the objective is to produce sufficient biomass that can then be used for product formation. The usual advice is to increase productivity by running the process at maximal specific biomass growth rate μ_{max} (e.g., Shioya 1992). Some companies put this into practice by starting the process with a batch phase where the cells are growing at maximal speed almost all time. Others start the process in the fed-batch mode with a feed rate that keeps the specific growth rate practically equal to μ_{max}. In principle, during the fed-batch operation feedback control could be applied in the engineering sense in order to keep the process on the desired path and thus to increase the batch-to-batch reproducibility. However, in the biomass production phase of the processes, feedback control usually does not lead to satisfactory results. The reason is that the controlled variables cannot be determined accurately enough within the time intervals required for control. Hence, the process must be operated in an open loop way.

In open loop control of the biomass formation phase of protein production processes, a major problem is to find the optimal feeding strategy. In this paper we focus on the robustness of exponential feeding profiles in the biomass growth phase of fermentations. We will proceed in comparing the different approaches by means of numerical simulation and in presenting a quantitative design for a particular example of a recombinant therapeutic protein. The computational result is subsequently validated experimentally by a couple of fermentation runs performed under various initial conditions.

2 MATERIALS AND METHODS

Cultivation system

In the experimental part of this work, an *E.coli* bacterial culture was used. The host organism used in this report is an E.coli BL21 (DE3) strain carrying a pET28a, Kan+ (Kanamycin resistance) expression system (Novagen 2003). The T7 promoter on the plasmid controlling expression was induced with 1 [mM] IPTG. The plasmid carries a gene coding for a particular recombinant protein. The product appears in form of inclusion bodies (IBs) within the cytoplasm of *E.coli*.

Fermentations were carried out in a B. Braun Biotech International® fermenter (Biostat C) with 10 L working volume equipped with 3 standard Rushton turbine impellers and a balance to measure the culture weight continuously. pH was measured with a Mettler-Toledo® electrode, culture temperature with a Pt-100 platinum metal resistance thermometer. A Maihak® offgas analysis system, containing the two modules Oxor610 and Unor610, allowed measuring the gas volume fractions of O_2 and CO_2 in the vent line. Substrate feed rate as well as base consumption rate were measured by means of balances under the reservoir vessels. Biomass concentration was determined offline by means of an optical density measurement with a spectrometer. The values were adjusted with additional dry weight measurements. Such measurements were also made in order to check the long term stability of the optical density measurement. Product concentration was measured with SDS Page electrophoresis against a standard.

The fermentation processes were conducted in the following way. The fermenter is prepared with a substrate solution that contains all salts and trace elements required but does not contain glucose. Fermentation is started with inoculation. At the same time the feeding is started with a fixed feeding profile $F_f(t)$ which is computed by

$$F_f(t) = \frac{\mu_{set} \cdot X_0 \cdot W_0}{Y_{XS} \cdot (S_F - S_0)} \cdot \exp(\mu_{set} \cdot t) \tag{1}$$

from the desired initial biomass concentration $X_{0f} = 0.15$ [g/kg], and $\mu = \mu_{set}$ which was chosen to be either $\mu_{set} = 0.5$ [1/h] or $\mu_{set} = \mu_{max} = 0.67$ [1/h] respectively. $S_F = 300$ [g/kg] is the concentration of the substrate in the feed and the initial substrate concentration S_0 was chosen to be 0.0 [g/kg]. The initial culture mass was adjusted to $W_0 = 4.75$ [kg], and Y_{XS} was estimated from a couple of experimental data sets to be $Y_{XS} = 0.49$ [g/g].

Experiments with different initial biomass concentrations X_0 were performed in such a way that the preculture was prepared using a different number of vials taken from the working cell bank. Each vial is used to inoculate a prefermentation in a shake flask of 100 [mL]. These are cultured for 6 [h] at 37 [°C] and 170 [rpm]. Experiments with 12 vials were taken as the standard. Hence X_{0f} was taken from this choice. Deviations from the standard were prepared with 8, 16 or 20 vials. The preculture conditions were the same in all other respects.

Modeling the cultivation process

In order to design the process operational procedure and to investigate the robustness of the growth process one needs a dynamic process model for simulation. Parameters of the model must be identified on experimental data from previous runs of the process.

The dynamical model was based on the mass balances around the entire culture within the bioreactor. The state vector c in this model is composed of the biomass concentration X, and the substrate concentration S. Additionally, the entire mass of the culture W is considered in a separate equation.

$$\frac{d\vec{c}}{dt} = \vec{q} \cdot X + \frac{F_S}{W} \cdot \vec{c}_F - \frac{F}{W} \cdot \vec{c} \tag{2}$$

$$\frac{dW}{dt} = F - F_{sampl} \tag{3}$$

$$\vec{c} = \begin{bmatrix} X \\ S \end{bmatrix} \qquad \vec{c}_F = \begin{bmatrix} 0 \\ S_F \end{bmatrix} \qquad \vec{q} = \begin{bmatrix} \mu \\ -\sigma \end{bmatrix} \tag{4a, b, c}$$

$$\sigma = \sigma_{\max} \cdot \frac{S}{K_S + S} \cdot \left(1 - \exp\left(\frac{-t}{t_{lag}}\right)\right) \tag{5}$$

$$\mu = Y_{XS} \cdot \sigma \tag{6}$$

$$F = F_S - F_{evap} + F_{gas} + F_{base} \tag{7}$$

$$F_{gas} = \left(OUR - CPR\right) \cdot W \tag{8}$$

$S_F = 300$ [g/kg]

Figure 1 shows a typical set of trajectories of the process under consideration. Model solutions (full lines) are compared with the measurement values (symbols). The agreements between model and measurements are sufficient for validating the model assumptions.

Table 1. Values of the model parameters used for numerical simulations.

Parameter	*Value*	*Dimension*
σ_{max}	*1.370*	*g/g/h*
K_S	*0.008*	*g/kg*
Y_{XS}	*0.490*	*g/g*
t_{lag}	*0.370*	*h*
F_{sampl}	*0.025*	*kg/h*
F_{evap}	*0.013*	*kg/h*

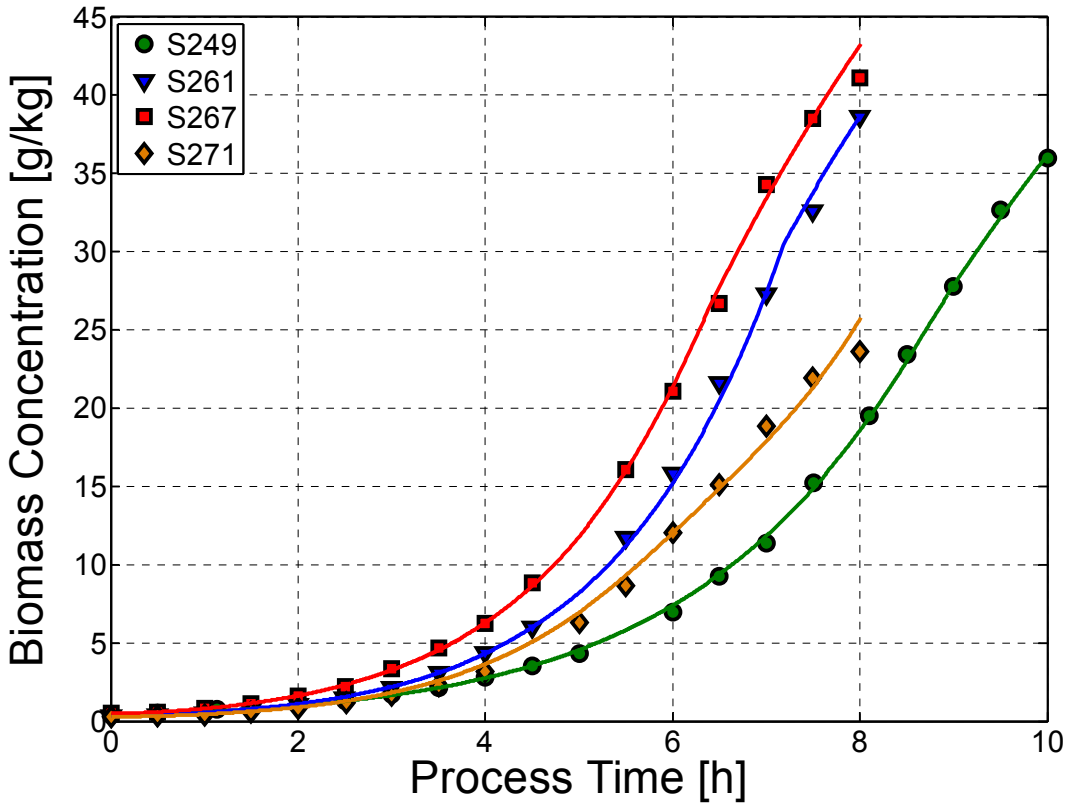

Figure 1. Typical examples of the model fit (full lines) to experimentally obtained biomass profiles (symbols) from 4 randomly chosen fermentation runs.

3 PROCESS ROBUSTNESS STUDY

One key sentence in FDA's PAT-related Guidance for Industry (FDA 2003) is: "quality cannot be tested into products; it should be built-in or should be by design". Several subsequently published comments of leading FDA officers specify that this is to be understood as robust design of the manufacturing process with respect to the process equipment design and its operational procedure. Designs have to be laid out in such a way that usual distortions do not affect the outcome of the process severely and, if possible, the process should be self adjusting upon distortions. Here we will show a simple example of how such a robust operational design can be made possible in practical fermentation technology.

We first identify which variables mainly influence the divergence of the biomass concentration trajectories X(t) in industrial protein production processes (and most other fermentation processes as well). Then we show by means of numerical experiments how to find sufficiently robust open-loop control profiles for X(t) and then we show experimentally that this designed operational procedure can be confirmed in real cultivations.

Variables mainly influencing the process quality

First we will show that the quantities mainly leading to deviations from the predetermined path of the process are the size of inoculum, i.e., initial biomass concentration X_0 and the initial specific growth rate μ of the cells used for inoculation.

In fermentation processes which are repeated many times, typically industrial production processes, the batch-to-batch reproducibility is rather low. Figure 2 shows a quite representative example for the biomass concentration profiles X(t) for fed-batch fermentations, where an E.coli strain was cultivated under similar conditions in order to produce a recombinant protein. In the initial biomass development phase, the cells were grown at maximal specific growth rate as usual in industrial practice.

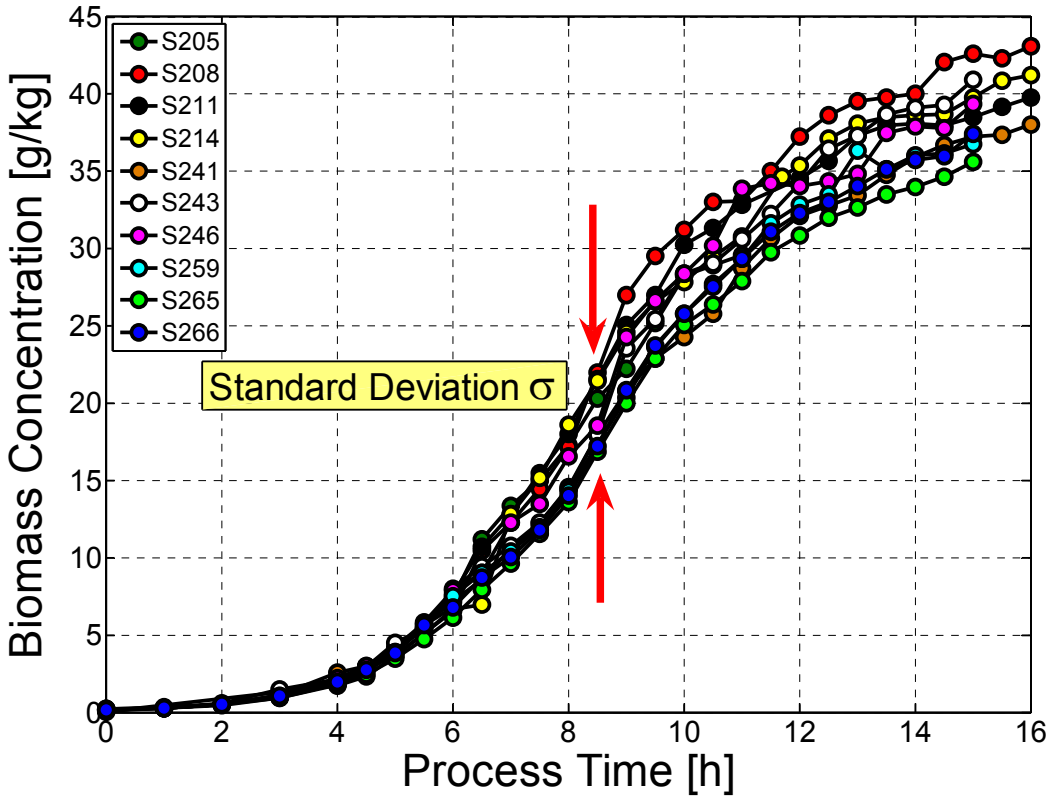

Figure 2. Typical set of biomass concentration profiles from 10 experiments with *E.coli* cells where recombinant proteins are produced.

Experience from many different cultivation systems shows, that the variances of the biomass concentrations $X(t_f)$, measured at given time instants t_f, are rather high, at least when compared with manufacturing plants in most other process industries. This fact essentially led to the PAT initiative of the FDA, which intends to improve the process quality in drug manufacturing processes.

A closer look at Figure 2 shows that most of the variability in the data is already present at induction time, which was $t_i = 8{,}5$ [h] in the examples shown in the Figure. Numerical least square fits of the model described above to the data in the time interval $[0 \ldots t_i]$, i.e. during the biomass formation phase, showed that the differences in the trajectories were primarily due to the initial biomass concentrations X_0. Second rank in the list of important influence variables is the specific growth rate μ at inoculation time, i.e. the growth phase at which they were taken from the prefermenter (in this case the shake flask).

Operation at reduced specific growth rates

In industrial practice, one usually tries to keep the biomass growth phase as short as possible in order to hold the total cultivation time low. Hence, the cells are grown at maximal growth rate. In order to avoid overfeeding and thus metabolic overflow, the substrate is added exactly at the amount the cells consume it. This means an exponential feeding profile adjusted to the maximal specific growth rate μ_{max}.

In order to search for the most robust operational procedure with respect to fluctuation in the size of the inoculum, numerical simulations were performed. For fixed specific growth rates μ, the influence of the initial biomass on the final biomass and specific growth rate was investigated. An important result is that the process appears to be very robust when the exponential feeding profile is taken for a specific growth rate μ that is clearly smaller than the maximal specific growth rate of the strain used.

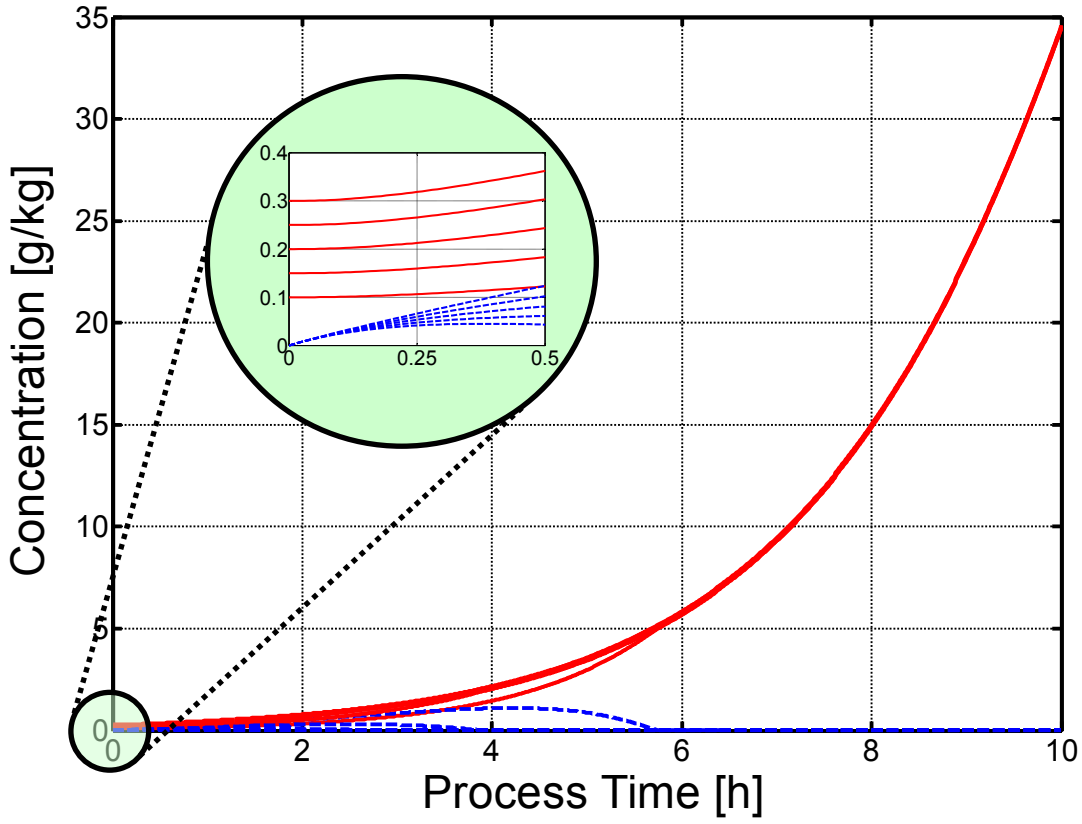

Figure 3. Simulations including biomass X (full lines) and substrate S (dashed lines) concentrations of the case that the feed rate in the experiments is computed for a specific growth rate set point of $\mu_{set} = 0.5$ [1/h], which is lower than $\mu_{max} = 0.67$ [1/h] of the cells. The initial biomass was varied in the interval 0.1 … 0.3 [g/kg].

The result of the simulations is shown in Figure 3. It is easy to explain: With a specific growth rate μ_{set} that is smaller than μ_{max} a self adjusting growth system appears. When the initial biomass concentration X_0 is smaller than the desired value X_{0f}, for which the feed rate profile F(t) was computed, then the cells recognize a higher substrate concentration than expected and thus they adapt their specific growth rate towards higher values. When, on the other hand the cell concentration X_0 starts with too a high value, the cells experience lower substrate concentrations than expected and reduce their growth rate. The integral effect is that all cells end up at the same concentration $X(t_f)$ at some time t_f.

How far the initial specific biomass growth rate must be reduced relative to the maximal specific growth rate of the cells depends on the variance of the biomass concentration at the inoculation time. The larger this variance, the lower the nominal specific growth rate set point for the feeding profile. When the variations in the initial biomass concentrations and the biomass at which induction will be performed are known, it is easy to estimate that specific growth rate to be adjusted by means of the exponential feeding profile $F_f(t)$ using the process model discussed in the last paragraph:

An appropriate algorithm is:

(i) Analyse the data on the initial biomass concentrations X_0 from previous experiments and determine the maximum X_{0max}

(ii) Estimate the specific biomass growth rate by $\mu_{set} = \frac{1}{t_f} \ln\left(\frac{X(t_f)}{X_{0\max}}\right)$

(iii) Estimate the substrate feed rate profile F(t) using equation (1)

Operation at high specific growth rates

The strategy that is predominant in industry is either to operate the process with an initial batch phase or just from the beginning in the fed-batch mode, where the feed rate is adjusted to μ_{max}. Hence, in both cases the process will initially run at maximal growth rate.

In order to shorten the biomass growth phase it was often proposed to run the process in this phase at the maximal specific growth rate μ_{max} of the cells (e.g., Shioya 1992). Other production fermenters are initially be run in the batch mode, which is equivalent with respect to running with the maximal growth rate. As will be shown in this paper, it is preferable to keep the specific growth rate lower. Then the process will be more robust, however, one must obviously pay for that advantage by a slightly longer biomass formation phase.

The reason is keeping the time needed to produce a sufficient amount of biomass as short as possible. However when the feed rate profile F(t) is computed with $\mu_{set} = \mu_{max}$, then deviations of the initial biomass concentration X_0 from the value X_{0f} assumed during the layout of the feeding profile F(t) may lead to significantly diverging biomass trajectories. In the case of deviations towards smaller values X_0, the cells cannot adjust their growth velocities towards higher μ as they are already growing at their maximal pace. Hence, the initial deviations in X are amplified during the growth phase so that we observe a considerable variance in X at t_f. This can be seen in Figure 4, which is computed for the same initial biomass concentrations than in the computations shown in Figure 3.

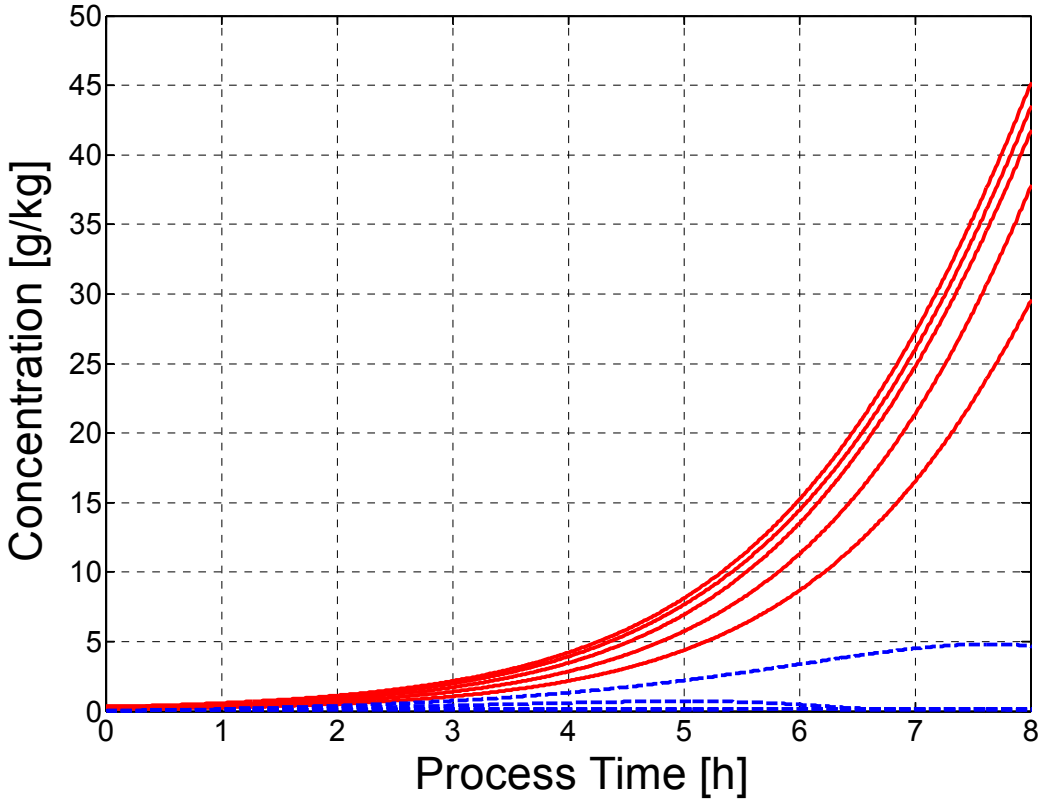

Figure 4. Result of a simulation based on the model equation using the same set of initial conditions as in the result depicted in Figure 3 with a feedrate profile F(t) computed for the same X_0 but $\mu_{set} = \mu_{max}$.

The result depicted in Figure 4 shows a considerable variance in the biomass concentrations at t = 8 [h]. Also, the specific growth rates are considerably different at that time. Consequently, the initial conditions of the protein production phase are much different and, thus, the process quality becomes rather low.

4 EXPERIMENTAL VALIDATIONS OF THE SIMULATED RESULTS

In order to confirm the computational results, a couple of experiments were performed with different initial biomass concentrations X_0. Two sets of experiments were conducted:

(i) a set in which the feed rate profile $F_f(t)$ was fixed by a constant $X_{0f} = 0.15$ [g/kg] and the specific growth rate to $\mu_{set} = 0.5$ [1/h] and

(ii) a set were it was fixed by $X_{0f} = 0.15$ [g/kg] and $\mu_{set} = \mu_{max} = 0.67$ [1/h]. The true initial biomass concentrations were varied in order to test the robustness of the open loop control.

For this purpose, inocula were generated from 8, 12, 16 or 20 vials that were previously stored in the working cell bank of the laboratory.

Biomass concentration profiles from the first set of experiments (for a nominal specific growth rate $\mu_{set} = 0.5$ [1/h]) are depicted in Figure 5. As can be seen, the curves converge to the trajectory $X_f(t)$ computed for the nominal feeding rate profile $F_f(t)$.

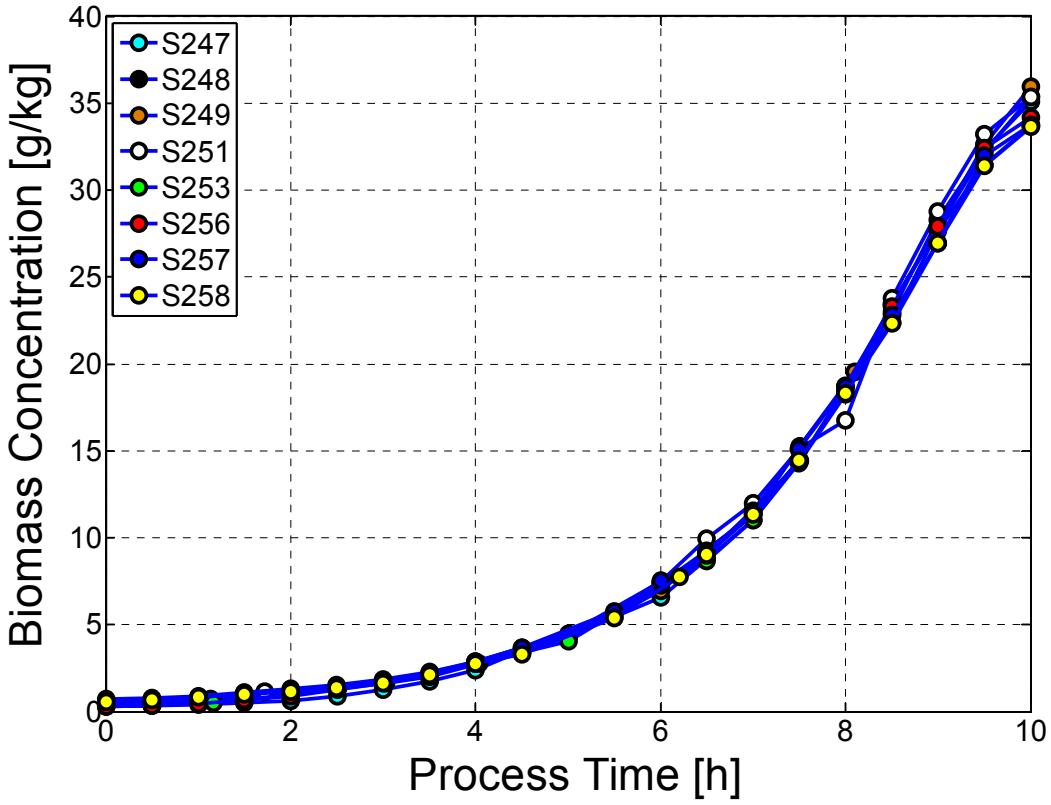

Figure 5. Biomass concentration profiles measured in a set of 8 experiments, where the feed rate was fixed according to $X_{0f} = 0.15$ [g/kg] and $\mu_{set} = 0.5$ [1/h]. The initial conditions in the experiments, however, varied as described in the text.

Since the effect of the initial biomass variations cannot be really seen in the original biomass data (Figure 5), the relative deviation of the X(t) profiles from the desired profiles $X_f(t)$ computed for the applied feeding profile $F_f(t)$ are shown in Figure 6. In this plot, the initial differences in X_0 become clearly apparent and it is seen that the trajectories converge to the nominal profile $X_f(t)$ within a few hours. Concretely, all signals remain within a ±5 [%] interval around $X_f(t)$ for all times longer than 8 [h].

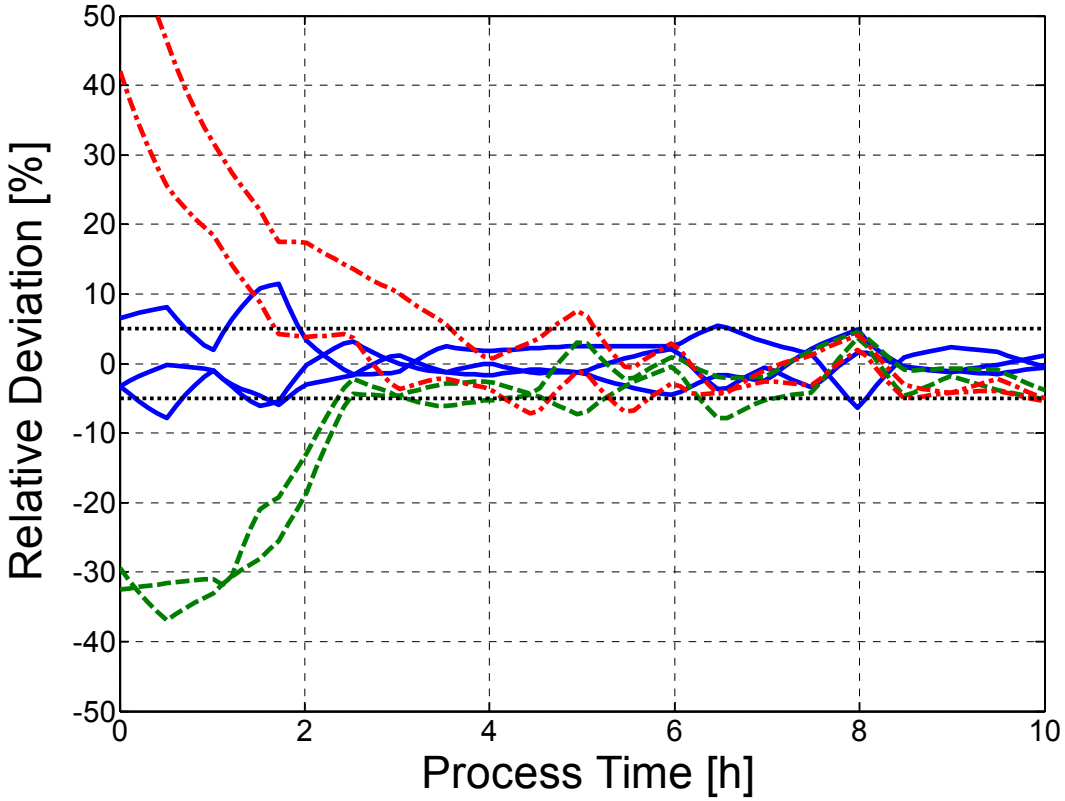

Figure 6. Relative deviation of the measured biomass concentration profiles depicted in Figure 5 from the nominal profile $X_f(t)$ computed with the feed rate $F_f(t)$.

The corresponding experiments with a fixed feed rate profile computed for a nominal specific growth rate $\mu_{set} = \mu_{max} = 0.67$ [1/h] are depicted in Figure 7 (Δ).

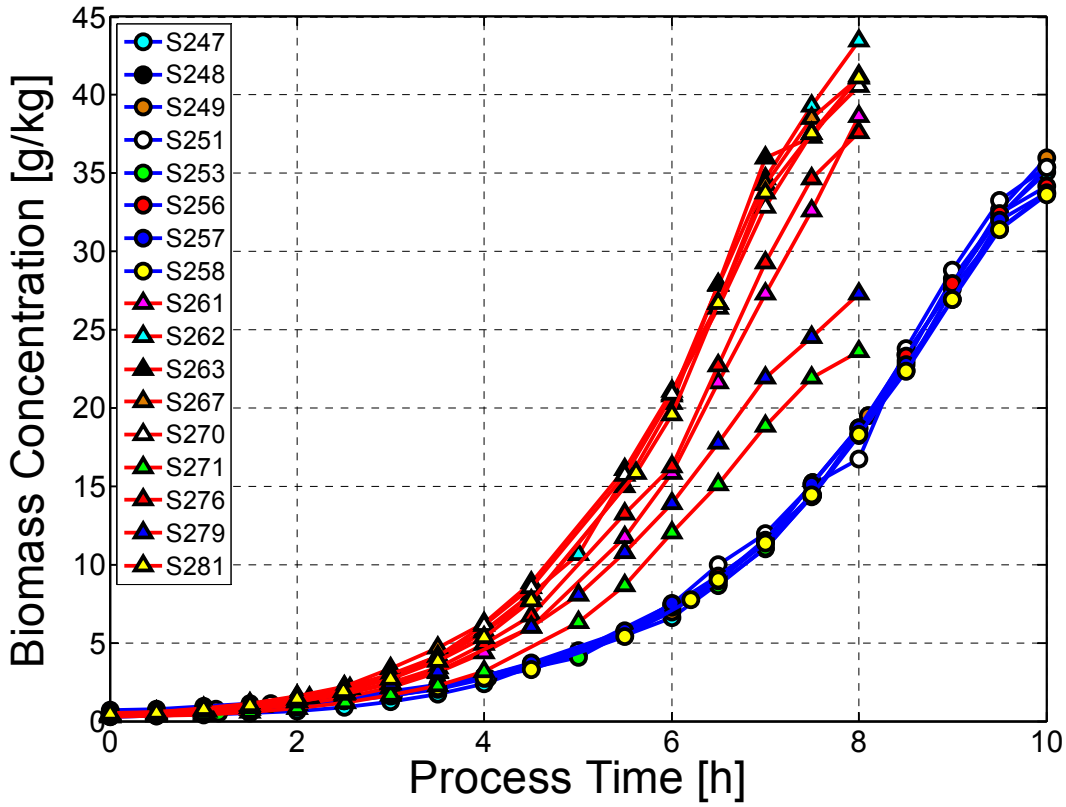

Figure 7. Biomass concentration profiles (Δ) measured in a set of 9 experiments, where the feed rate was computed with a fixed $X_0 = 0.15$ [g/kg] and a $\mu_{set} = 0.67$ [1/h]. The initial conditions for X_0 were the same as in the case $\mu_{set} = 0.5$ [1/h] (○).

Again, it is instructive to look for the corresponding relative deviation of X(t) from the reference profile $X_f(t)$. This is depicted in Figure 8.

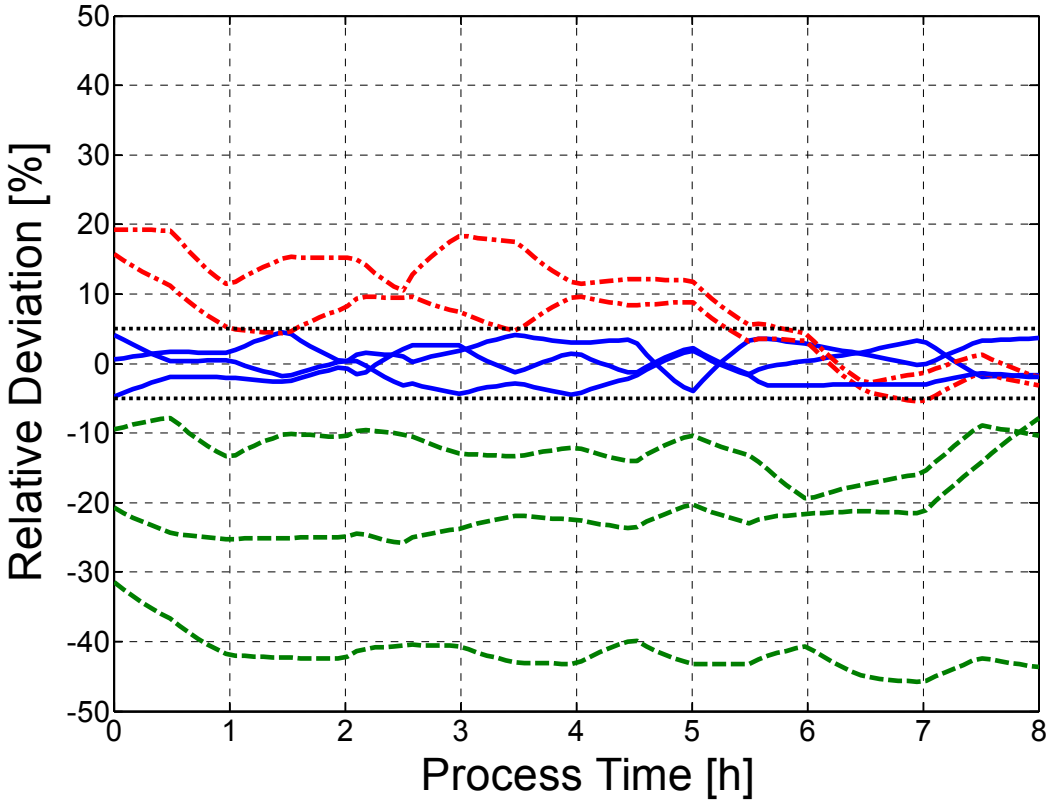

Figure 8. Relative deviations in offline measured biomass from fed batch experiments grown at maximum specific growth rate of $\mu_{max} = 0.67$ [1/h] with different initial biomass concentrations.[full line … 12 vials (standard procedure); dashed … 8 vials, dash-dot .. 16 vials].

In Figure 8 one can clearly see what happens with the different deviation. When there are no deviations (case of 12 vials), the profiles stay within the ±5 [%] interval around $X_f(t)$ for all times. When we start with two a high biomass concentration $X_0 > X_{0f}$ then the trajectory converge, as expected, towards the nominal X_{0f} profile. However, when the deviation is the other way around, meaning that too little initial biomass was taken as inoculum, the trajectories do not converge. This is exactly the effect expected from the simulations performed beforehand.

In other words, in experiments with a feed rate profile fixed beforehand, the trajectories starting with different initial biomass values do converge to the nominal trajectory when the cells are propagated at a specific growth rate smaller than the maximal one. When the process is run at maximal growth rate, significant variances in X are to be expected under otherwise the same experimental conditions.

Product formation

An important question now is whether or not the controlled biomass growth also has a positive effect on product formation. In the investigation reported here, the process was continued only about 2 hours after induction. And within this interval the product concentrations were also measured for some of the cultivation runs. The result is shown in Figure 9.

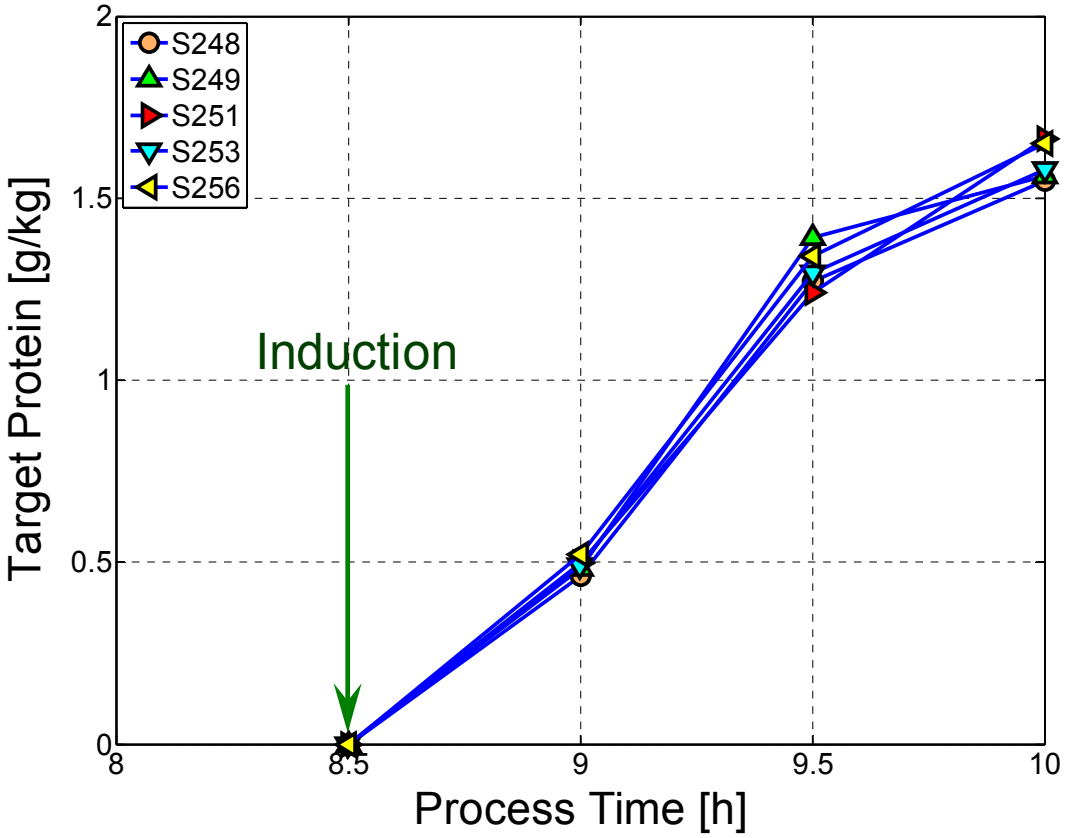

Figure 9. Target protein concentration profiles from 5 fed batch experiments grown at a specific growth rate of $\mu = 0.5$ [1/h] with different initial biomass concentrations. Protein formation was induced with IPTG at 8.5 hours.

As with the biomass profiles, the target protein trajectories are close together saying that the control has an immediate influence on the product formation.

Feasibility of closed loop control

One could pose the question whether or not the batch-to-batch reproducibility in the initial growth phase could have been easier obtained by means of a closed loop control of the cultivation process. The answer is no, the problem is that control requires reliable measurements of the state of the process in order to become able at all to decide whether or not there is, and, if, how large the deviation from the desired set point is. This answer can usually be supported by experimental results on the mean standard error of the measurement values available from the process.

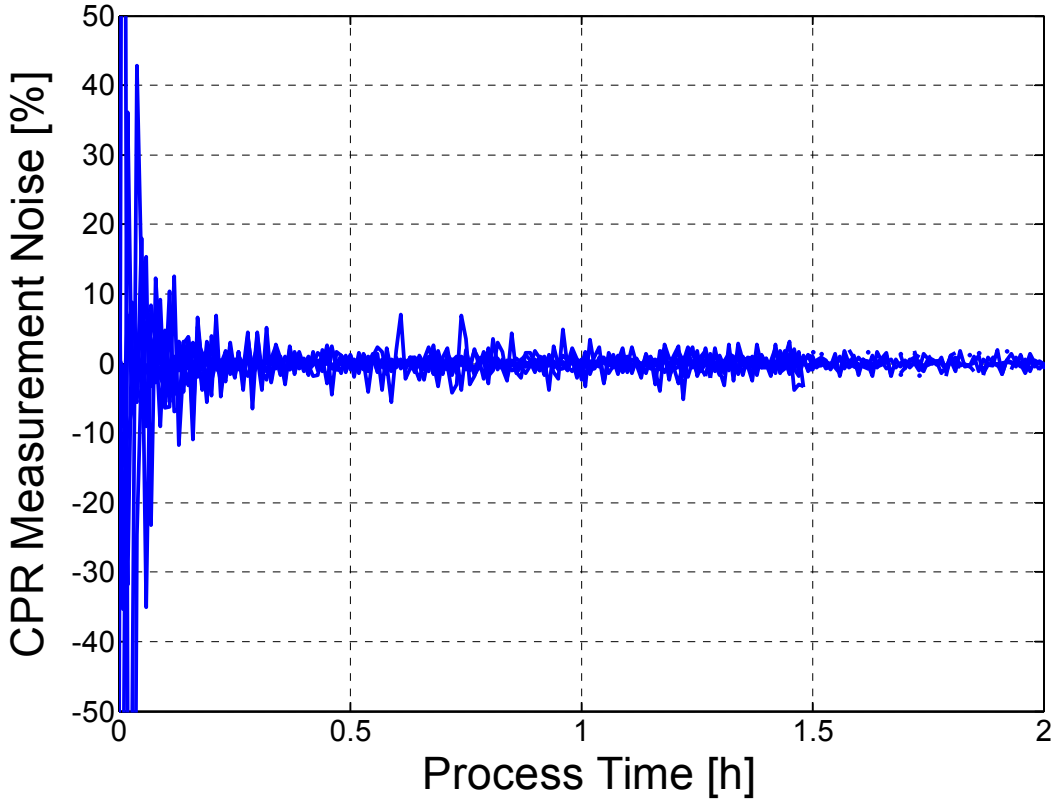

Figure 10. Relative noise in the CPR measurements in the beginning of the fermentation up to the 2nd fermentation process hour.

Initially the accuracy by which the key process variables such as the biomass concentration X can be determined is rather low but with time more biomass accumulates and the accuracy becomes higher. In order to access these accuracy problems, one can look for the accuracies of the measurement variables used for biomass estimation. As the latter is usually determined from the off-gas analysis, the relative error of the CPR signal plays an essential role. This is depicted in Figure 10. As can be seen from the figure, the error is rather high in the beginning of the fermentation but becomes significantly smaller after a few hours. The consequence from this measurement results is that the CPR can be used for control purposes only about 2 hours after inoculation. Consequently, the growth phase must be controlled in an open loop fashion, preferentially along a robust feed rate trajectory as discussed before.

Statistical process analysis

In its PAT-Initiative, FDA proposes to make use of multivariate statistical analysis of the data from manufacturing processes in order to assess the process quality. These techniques can obviously also be applied to the process data presented before and one expects that the analysis is able to discriminate between the various approaches tested during the validation of the numerical predictions.

Standard principal component analysis (PCA) is a well known multivariate statistical method (e.g., Jackson 1991) that was developed to study the variances and in particular the covariance structures in multivariate data records. One of the most important extensions of normal PCA is MPCA (e.g., Nomikos and MacGregor 1994, 1995), a multiple principal component

analysis based on a set of (fed-) batch experiments. The various approaches on monitoring batch processes with this technique as discussed in literature differ in the way they combine the three-dimensional data records [I batches; J variables; K times] into a two-dimensional data matrix that can be evaluated by standard PCA routines.

Here we propose to combine variables and batches to get a two dimensional matrix with [I rows; J*K columns] according to Nomikos and MacGregor (1994). Then we get a single point for each batch i in the space spanned by the principal components. This allows analyzing the variability among the various batches considered by summarizing the information in the data with respect to the variables and their time variation. By plotting the scores of the first two principal components versus each other once can easily see which of the process realizations have similarities in their measurements and thus form clusters (Nomikos and MacGregor 1995) and which are isolated form the others and therefore are outliers or unusual representations in other respects. The result for the 17 batches discussed before are depicted in Figure 11.

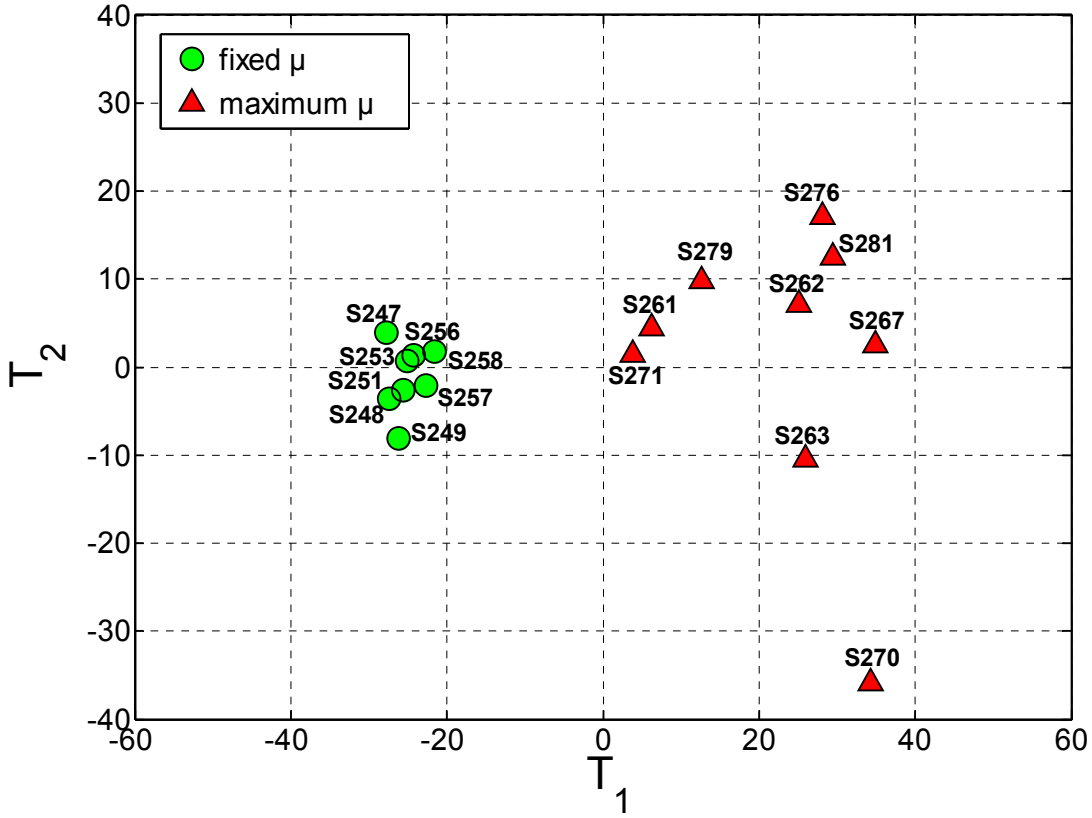

Figure 11. Multiple principal component analysis of the experiments discussed above. Only the scores of the first two principal components are shown in the figure.

It becomes clear that the different batches in the case of the well designed experiment are close together in this space and the batches performed with a feeding profile based on maximal specific growth rate are found in a different domain of the space and are much more different from each other. Hence, with such a technique one can see on the first glance whether or not the processes were under tight control.

One can now ask the question about the time development of the processes towards their ends. When we perform the same analysis for data records ending at different times K*Δt, we see the time developments of the clouds depicted in Figure 11. This is shown in Figure 12.

Additionally, in Figure 12 we also considered additional batches that were run under tightly controlled using feed back control.

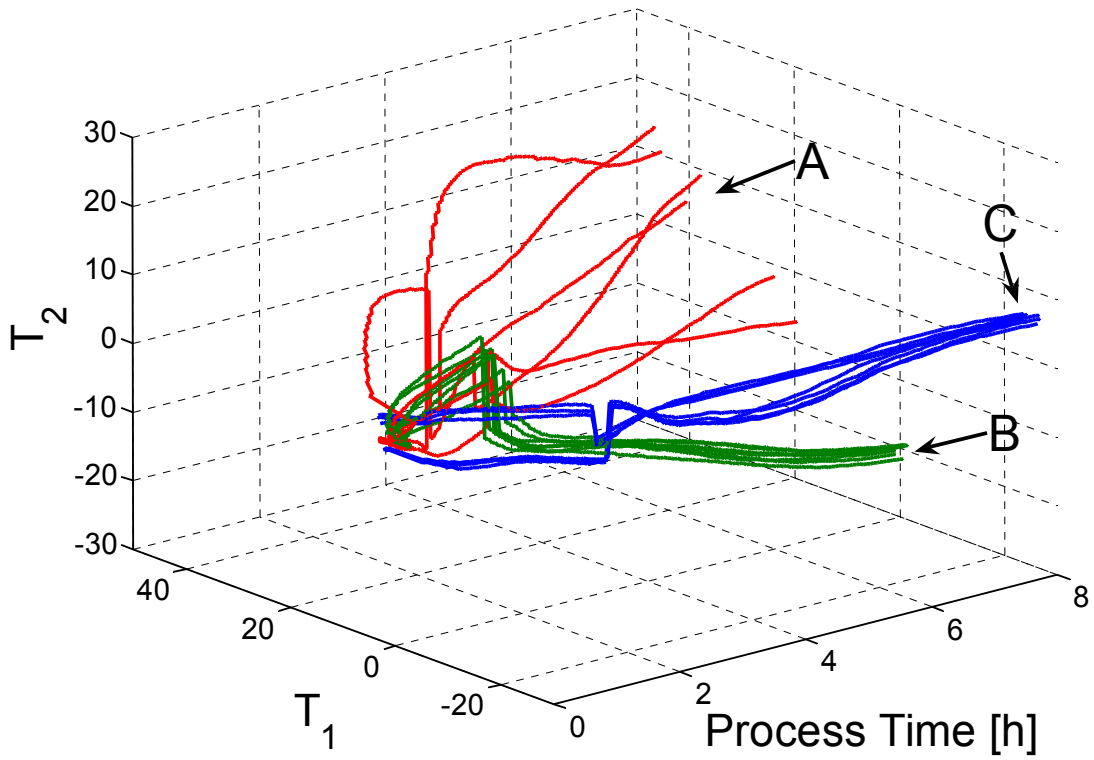

Figure 12. Time development of the clusters that appear in Figure 11 and an additional one from a set of experiments where the process was controlled during the product formation phase. **A** … fed-batch process where the cells are running at maximal growth rate, **B** … open loop control within the biomass growth phase, **C** … feedback-controlled cultivation.

It becomes clear that the controlled batches, either controlled in an open-loop fashion or in a feedback closed loop fashion are very close together in the three-dimensional principal component space and those realizations of the cultivation which are run at maximal growth rate in the beginning are diverging, thus depicting a low batch-to-batch reproducibility or process quality.

5 CONCLUSIONS

Recombinant protein production processes are being used to produce several high value proteins, so-called biologics. The manufacturing process for each biologic defines, to a significant extent, the product that is produced. Apparently small differences between manufacturing processes can cause significant differences in the clinical properties of the products of those processes (PhRMA 2001). The approval of these products by the authorities is thus bound to details in the operational procedure of the production process. Hence, from the product quality point of view, the batch-to-batch reproducibility must be number one priority in the manufacturing processes.

Batch to batch reproducibility can be improved by more robust process operating procedures and remaining distortions can be combated by means of feedback control. However, the latter can only be successful where sufficiently accurate online data is available to quickly enough determine the current value of the controlled variable. This is not possible in the first phase of a fermentation process as the biomass then is too low to produce accurately measurable rate effects. Hence, in the beginning one is restricted to open-loop control.

We showed at a concrete example of a recombinant protein production process that it is favourable to run such processes even within the biomass formation phase in the fed-batch operational mode. The feed rate F should follow an exponential profile corresponding to a specific growth rate which is significantly lower than the maximal specific growth rate of the cells. Then, a self-controlling behaviour of the cellular growth is to be expected resulting in a stable biomass concentration X and specific growth rate at induction time.

The cost of this procedure is a slightly longer cultivation time as the biomass growth phase will be about two hours longer in the particular case discussed in this paper.

It was shown, that this open loop control is able to cover the most important distortions appearing in industrial production systems, namely the variation in the initial biomass concentration upon inoculation. The procedure is rather robust with respect to these distortions as the plots show.

The developments reported in above are in the focus of FDA's initiative on Process Analytical Technology (PAT). This exactly aims to improve product quality and process performance in the manufacturing processes in pharmaceutical industry. With PAT, FDA particularly aims in improving manufacturing processes by means of a more knowledge-based analysis and control that are based on timely measurements during processes. In this paper we present methods by which these aims can be realized in industrial practice. The feasibility of the methods could be confirmed at the concrete example of a recombinant therapeutic protein production process.

Acknowledgements. This work was supported by "Land Sachsen-Anhalt" within the project "Wertschöpfung durch Proteine als Wirkstoffe und Werkzeuge". We gratefully thank for the financial support.

NOMENCLATURE

c	concentration vector
c_F	feed concentration vector
CPR	CO_2 production rate [g/kg/h]
F_{base}	mass flow rate in reactor due to base feed from pH control [kg/h]
F_{evap}	mass flow rate in reactor due to medium evaporation [kg/h]
F_f	fixed substrate feed rate profile [kg/h]
F_{gas}	mass flow rate in reactor due to carbon loss with CO_2 in off gas [kg/h]

F_S	substrate feed rate [kg/h]
F_{sampl}	sampling rate in reactor [kg/h]
I	number of batches
J	number of variables
K	number of time steps
K_S	Monod constant for substrate consumption [g/kg]
OUR	O_2 uptake rate [g/kg/h]
q	vector of specific reaction rates
S	substrate concentration [g/kg]
S_0	initial substrate concentration [g/kg]
S_F	substrate concentration in feed [g/kg]
t	current process time [h]
t_f	concrete time instant [h]
t_i	induction time [h]
t_{lag}	time lag constant for biomass growth [h]
W	culture mass [kg]
W_0	initial culture mass [kg]
X	biomass concentration [g/kg]
X_0	initial biomass concentration [g/kg]
X_{0f}	desired value of X_0 [g/kg]
X_{0max}	maximal value of X_0 [g/kg]
X_f	desired profile of X [g/kg]
Y_{XS}	biomass/substrate yield [g/g]

Greek letters

μ	*s*pecific biomass growth rate [1/h]
μ_{max}	maximal specific biomass growth rate [1/h]
μ_{set}	set point of specific biomass growth rate [1/h]
σ	specific substrate consumption rate [g/g/h]
σ_{max}	maximal specific substrate consumption rate [g/g/h]

REFERENCES

FDA (2003), *Guidance for Industry: PAT - A Framework for innovative pharmaceutical manufacturing and quality assurance*, August 2003, Food and Drug Administration, Rockville, MD

Jackson, J.E. (1991), A User's Guide to Principal Components, Wiley, New York

Nomikos, P., MacGregor, J.F. (1994), Monitoring Batch Processes Using Multiway Principal Component Analysis, *AIChE J.*, 40/8, 1361-1375

Nomikos, P., MacGregor, J.F. (1995), Multivariate SPC charts for Monitoring Batch Processes, *Technometrics*, 37, 41-59

Novagen (2003), pET System Manual, 10th ed., www.novagen.com

PhRMA (2001), Issues and Questions on Biologics, Presentation of the Organization of Pharmaceutical Research and Manufacturers of America, http://www.europabio.org/documents/QA-Pharma.doc

Shioya, S. (1992), Optimization and control in fed-batch bioreactors, *Adv. Biochem. Eng. Biotechnol.*, 46, 111-142

Chapter 3

Estimation of Biomass Concentrations in Fermentation Processes for Recombinant Protein Production

Abstract. Online biomass estimation for bioprocess supervision and control purposes is addressed. As the biomass concentration cannot be measured online during the production to sufficient accuracy, indirect measurement techniques are required. Here we compare several possibilities for the concrete case of recombinant protein production with genetically modified *E.coli* bacteria and perform a ranking. At normal process operation, the best estimates can be obtained with artificial neural networks (ANNs). When they cannot be employed, statistical correlation techniques can be used such as multivariate regression techniques. Simple model-based techniques, e.g., those based on the Luedeking/Piret-type are not as accurate as the ANN approach; however, they are very robust. Techniques based on principal component analysis (PCA) can be used to recognize abnormal cultivation behavior. For the cases investigated, a complete ranking list of the methods is given in terms of the root-mean-square error of the estimates. All techniques examined are in line with the recommendations expressed in the process analytical technology (PAT)-initiative of the FDA.

This chapter has been published in *Journal of Bioprocess & Biosystems Engineering*:
Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A. (2006), Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioproc. Biosyst. Eng.*, 29(1), 19-27

1 INTRODUCTION

Quality assurance is a key issue in pharmaceutical manufacturing, particularly in the production of high value recombinant therapeutic proteins. Quality of processes is best characterized by their batch-to-batch reproducibility. A primary goal in process quality assurance is a process design that leads to robust operation with respect to typically appearing distortions. In other words, among the possible operational variants, that one must be selected, that is least sensitive to distortions. Once such robust trajectories for the fermentation process are found, feed forward control is used to guide the process along these paths. This is the way in which most fermentation processes are currently operated in industrial practice. However, in order to make sure that the process will tightly follow the desired path, deviations resulting from random distortions must be compensated for by feedback control. This requires determining the current values of the controlled variables online. In a recently published recommendation (FDA 2003), manufacturers have been encouraged to use the latest scientific advances in process supervision and control to improve the quality in pharmaceutical manufacturing processes.

The response of several leading manufacturers to FDA's initiative was to start activities in process automation with feedback control. This primarily requires well performing online measurements of the key variables influencing the performance of the production processes. As these usually cannot be measured directly online, model-supported, indirect measurements are required. In recombinant protein production, one of the primary desired quantities is the mass of product m_P that can finally be placed onto the balance. m_P can explicitly be expressed as

$$m_P = \int_0^t \pi \cdot x \cdot dt \tag{1}$$

where x is the biomass and π the specific product formation rate. In order to get a high m_P, both, π and x, which are both time dependent, must be kept as high as possible over periods as long as possible. Hence, in order to monitor the protein production processes efficiently, powerful online methods must be available for both factors.

In this paper we examine indirect estimation techniques for biomass, i.e. measurements of the number of producers of the recombinant protein in production bioreactors where *E.coli* bacteria are cultivated. Several direct measurement methods have been proposed in literature such as dielectric spectroscopy, optical density, etc. These can do a good job in the laboratory environment but are seldom found as online sensors in large-scale production reactors. As they are local measurement devices, they usually provide less representative values of the biomass concentration at the large production scale than global data from integral measurement devices. We restrict the discussion to estimates based on those online variables that are commonly measured at production fermenters.

A large number of estimation techniques are mentioned in literature, however, most of these methods have not been used in real production environments and in particular not in fermentations for recombinant protein production.

In production fermenters, the problem is that only a small number of quantities are measured online during the cultivation. Under these special conditions it is not clear beforehand whether or not the established multivariate statistical estimation techniques can be applied successfully

and which of the techniques are to be preferred under the conditions prevailing in fermenters operated with genetically modified cells.

Indirect measurements require sufficiently accurate relationships between the variables to be monitored and those that are measured online. Here we will discuss models, based on simple mass balances, which allow estimating the biomass dry weight x or the corresponding concentration X from online measurement values such as oxygen uptake rate (OUR), carbon dioxide production rate (CPR) and the base consumption.

These direct model-based techniques are compared with data-driven techniques. The simplest variant considered is multivariate linear regression. Further, a special application of the principal component analysis (PCA) is tested. An additional approach uses the mapping properties of feed forward artificial neural networks. We will discuss simple feed forward networks with a single hidden layer, and we will also examine the more sophisticated auto associative neural networks. Generally all these methods require an adequate number of process data records to guarantee sufficiently accurate results.

The various biomass concentration estimators are on different levels of complexity. The question, in how far the more complex ones truly lead to performance improvements is addressed. We compare the techniques for the estimation of biomass concentrations at the concrete example of the production of the green fluorescence protein (GFP) with *Escherichia coli* bacteria. The data are taken from cultivation experiments on the laboratory (10-liter) scale.

In literature, further "advanced methods" are discussed such as Luenberger observers and Extended Kalman filters. These methods do a good job in scientific environments (Jenzsch *et al.* 2006). However, at industrial production environments the tuning load upon small changes in hard and software inhibit their use.

2 MATERIALS AND METHODS

E.coli BL21 (DE3) pET11a EGFP was used in all experiments considered. The strain is able to express the green fluorescent protein (GFP) under control of the T7 promoter (Sambrook and Russell 2001). The organisms were cultivated on a defined mineral salt medium with 20 g/kg glucose as the energy and carbon source. Product formation was induced with IPTG.

All cultivations were carried out in a B. Braun Biotech 10-L-bioreactor Biostat C, operated in the fed-batch mode. The feeding solution consisted of glucose at a concentration of 400 [g/kg] and otherwise the same composition as the initial mineral salt medium. The process was started by means of automatic inoculation over night, and, in the first phase, operated in the batch mode. Substrate feeding was started after the initially supplied glucose was exhausted (about 4 [h] after inoculation). Induction was performed at a predetermined time, between 7 – 10 [h] after inoculation.

B. Braun's DCU was used as front-end controller. The entire process was monitored under control of MFCS/win. All estimators reported on in this article were run on a separate PC connected to the MFCS/win computer via a local network. A commercially available server/client-software (IPCOS) was employed. Data exchange between MFCS and the estimators was carried out with a software interface (MATLAB®) that allows reading all relevant data from MFCS.

Off-gas analysis was performed with a paramagnetic sensor (Maihak Oxor 610) for O_2 and an infrared detector (Maihak Unor 610) for CO_2. pH values were measured with an Ingold pH probe. The dissolved oxygen concentration was monitored with a pO_2 electrode (Mettler-Toledo). Substrate and base addition rates were measured with balances (Sartorius). These online signals were recorded all 36 [s].

Off-line measurements were performed with a time increment of about half an hour. Optical density (OD_{600}) measurements were performed with a spectral photometer (Shimadzu UV-2102PC) at 600 [nm]. Biomass dry weight measurements were performed with standard procedures in order to establish and test the correlation between biomass dry weight and the OD_{600} values, which is needed to validate the estimation results. The biomass measurement data were obtained from offline OD_{600} measurements by multiplying the OD values by the calibration coefficient 0.33 [OD/(g DW/kg culture)]. An YSI 2700 analyzer was employed to measure the glucose concentration. The product concentration (GFP) was estimated from the fluorescence intensity measured from samples with fully intact cells in a spectro-fluorimeter (Hitachi F-2500).

Oxygen uptake rate OUR and carbon dioxide production rate CPR, are both derived from the offgas analysis. The total base consumption is computed from the mass signal measured with the balance below the base reservoir. Fermenter head pressure, pH and culture temperature T are generally measured online. However, these quantities are not considered in the biomass estimation as they were separately controlled to predefined values by controllers incorporated in the DCU. pO_2 and culture weight are further quantities for which sensors are available. More details about the experiments can be found in Jenzsch *et al.* 2005.

Biomass estimation

Online estimation of the current biomass concentration was made indirectly utilizing those quantities that can easily be measured during the running cultivation. Here we restrict the discussion to measurements that are commonly made at production bioreactors: Volume or mole fractions of oxygen and carbon dioxide in the bioreactors' vent line and the base addition rate that appears with the pH control in the bioreactor. From these quasi-continuously measured quantities the oxygen uptake rate, OUR, as well as the carbon dioxide production rate, CPR, are determined in a straightforward way. Hence, these three quantities (OUR, CPR, Base) can be considered online measured quantities. Below, we discuss various methods for estimating the biomass concentration online from these variables.

Estimations based on Luedeking/Piret-like relationships

In most aerobic cultivations, the relation between the biomass x and the oxygen uptake rate OUR and CPR in bioreactor can be modeled by means of a Luedeking/Piret-type relationships (Shuler and Kargi 2002, Luedking and Piret 1959 a, b)

$$OUR(t) = Y_{OX} \cdot R_X(t) + m_O \cdot \frac{x(t)}{W(t)}, \tag{2}$$

$$CPR(t) = Y_{CX} \cdot R_X(t) + m_C \cdot \frac{x(t)}{W(t)}, \tag{3}$$

where R_x [g*(X)*/kg(*Culture)*/h] is the biomass formation rate of the cellular system, x the total biomass [g], W the culture mass [kg], and Y_{OX} [g*(O₂)*/g*(X)*], Y_{CX} [g*(CO₂)*/g*(X)*] are yield expressions. m_O [g*(O₂)*/g*(X)*/h], and m_C [g*(CO₂)*/g*(X)*/h] are model parameters that are often termed maintenance contributions as m_o quantifies the growth-independent part of the specific oxygen consumption rate and m_c the growth-independent part of the specific carbon dioxide production rate.

As W(t) is measured online, the biomass balance around the reactor can be formulated by the ordinary differential equation.

$$\frac{dx}{dt} = R_X(t) \cdot W(t) \quad , \qquad [g(X)/h] \tag{4}$$

With equations (2) and (3) this equation can be rewritten to give

$$\frac{dx}{dt} = \frac{OUR(t) \cdot W(t) - m_O \cdot x(t)}{Y_{OX}}, \quad [g(X)/h] \tag{5}$$

$$\frac{dx}{dt} = \frac{CPR(t) \cdot W(t) - m_C \cdot x(t)}{Y_{CX}}, \quad [g(X)/h] \tag{6}$$

Since OUR, CPR are online measured as well, the simple ordinary differential equations (5) and (6) can be solved if the initial biomass x_0 as well as the coefficients Y_{OX}, m_O , Y_{CX} , and m_C are known. The 4 coefficients can be identified independently from data records W(t), OUR(t), CPR(t) and x(t) previously measured at the process under consideration using standard nonlinear parameter optimization techniques.

An analog equation can be formulated for the ammonium or base consumption rate during the cultivation

$$Base(t) = Y_{BX} \cdot R_X(t), \tag{7}$$

This equation is similar to equations (2) and (3), however, the maintenance term was ignored because the value of the maintenance term governed by m_B is negligibly small compared with $Y_{BX} R_X$.

Estimations based on multi-variable regression methods

A straightforward purely data-driven approach to biomass estimation is using correlations between the online measured signals and the biomass concentration. The obvious first choice is a simple linear multi-variable regression

$$X = a_0 + a_1 \cdot CPR + a_2 \cdot OUR + a_3 \cdot \sum Base \tag{8}$$

The a_i are the regression parameters, which can be estimated using classical least square methods. The sum in the Base-term results from the fact that the base consumption is usually available in terms of total mass of base pumped into the reactor up to the current time t; i.e. Σ_j Base, the cumulative addition of base up to time $t=t_j$, is directly available as (cumulative) measurement signal. The simple multiple linear regression approach can be improved using cumulative values of CPR and OUR as well. This can be considered a noise filter. In this case the equation for the biomass estimation reads

$$X = a_0 + a_1 \cdot \sum CPR + a_2 \cdot \sum OUR + a_3 \cdot \sum Base \tag{9}$$

The reason why the correlation between the cumulative variables and the biomass concentration performs better is thought to be a consequence of the fact that the biomass concentration is a cumulative variable by itself. Another reason for using cumulative variables is that the cumulation process eliminates a considerable part of the measurement noise. Hence, it is straightforward to use cumulative variables in equation (9). On the other hand, experience shows that OUR, CPR and base consumption rate BCR depict a better correlation with the biomass growth rate.

Although this simple model finally did not provide the best results, it is straightforward according to *Ockham's razor* to start a new estimation project with a very simple approach. And in many practical cases the accuracy might be sufficiently high. Anyway, it should be considered a first approach to biomass concentration which can be used to judge the advantage obtained with more sophisticated methods.

As the cultivation processes are nonlinear in nature, an improvement is to be expected by extending the linear cumulative variable model towards nonlinear one. A simple quadratic expression, formulated by

$$\begin{aligned} X = {} & a_0 + a_1 \cdot \sum CPR + a_2 \cdot \sum OUR + a_3 \cdot \sum Base + \\ & a_4 \cdot \left(\sum CPR\right)^2 + a_5 \cdot \left(\sum OUR\right)^2 + a_6 \cdot \left(\sum Base\right)^2 + \\ & a_7 \cdot \sum CPR \cdot \sum OUR + a_8 \cdot \sum CPR \cdot \sum Base + \\ & a_9 \cdot \sum OUR \cdot \sum Base + a_{10} \cdot \sum CPR \cdot \sum OUR \cdot \sum Base \end{aligned} \tag{10}$$

can be assumed as a first approximation to a nonlinear approach which is linear in its parameters. The 11 parameters can also be estimated using least square methods from the available process data.

Biomass estimation based on principal component analysis (PCA)

As the various variables that can be measured online during the fermentation, OUR, CPR, BASE, and pO_2, are rather strongly correlated with each other, the representation of the process' state by means of these variables is reducible. In such cases, principal component analysis (PCA) techniques (e.g.: Jollife 1986; Jackson 1991; Kourti and MacGregor 1995) can be used to simplify the representation. Its main idea is to represent the state of the process with a reduced number of independent variables. Principle component analysis orders these "latent variables" by their contribution in describing the variability inherent in the data. The component that represents most of the variance in the original data is named the first principal component. The variances with respect to all further components are smaller. Some only carry process or measurement noise and can thus be discarded. This is the idea behind the reduction of the dimensionality in the data representation.

While it is easier to deal with a smaller number of independent data, a disadvantage is that these new variables are generalized quantities that cannot be interpreted as easily as the well known physical input variables.

In PCA the original process data are mapped, with minimal loss of information, into a space of lower dimensions. Let **Y** represent a *(n,m)*- data matrix, where n is the number of observations and m the number of variables. PCA performs an optimal factorization of **Y** into two matrices, **S**, referred to as the scores matrix of size *(n,f)* and, **P,** called the loadings matrix of dimension *(m,f):*

$$\mathbf{Y} = \mathbf{S}\,\mathbf{P}^{T} + \mathbf{E} \tag{11}$$

where f is the number of factors ($f \leq m$) , and **E** the matrix of residuals, which is of size *(n,m)*. The condition of optimality of the factorization is that the Euclidian norm of the residual matrix is minimal for the given number f of factors. It is well known that for this case the columns of **P** must be the eigenvectors corresponding to the f largest eigenvalues of the covariance matrix of **Y**. In our application we interpret the PCA as a linear mapping of data from an m-dimensional into an f-dimensional space (Figure 1). For the case $f = m$ it is straightforward to estimate the square matrix **P** and inverse matrix $\mathbf{P}^{-1}$ for all f components. Then the scores matrix for all f-components can be estimated by

$$\mathbf{S} = \mathbf{Y}\,\mathbf{P}^{-1} \tag{12}$$

In real applications one uses only the first 2-3 components or columns of S for process monitoring. This is referred to as the restricted matrix $\mathbf{S}_R$. The trajectories of the first components in the feature space can provide essential information about the dynamics of the process. The information lost upon using only the first components of the scores matrix can be evaluated by reconstruction of the measurement vector using restricted matrixes $\mathbf{S}_R$ and $\mathbf{P}_R$

$$\mathbf{Y}_R = \mathbf{S}_R\,\mathbf{P}_R^{T} \tag{13}$$

where $\mathbf{Y}_R$ are the reconstructed measurement values. The smaller the dimension of the components of the feature space, the greater is the resulting error for the reconstructed

measurement data. This PCA analysis can easily be performed using general PC software. Subroutines for estimation of loading matrix P are included in all common statistical packages for PCs (e.g. MATLAB® subroutine *princomp*).

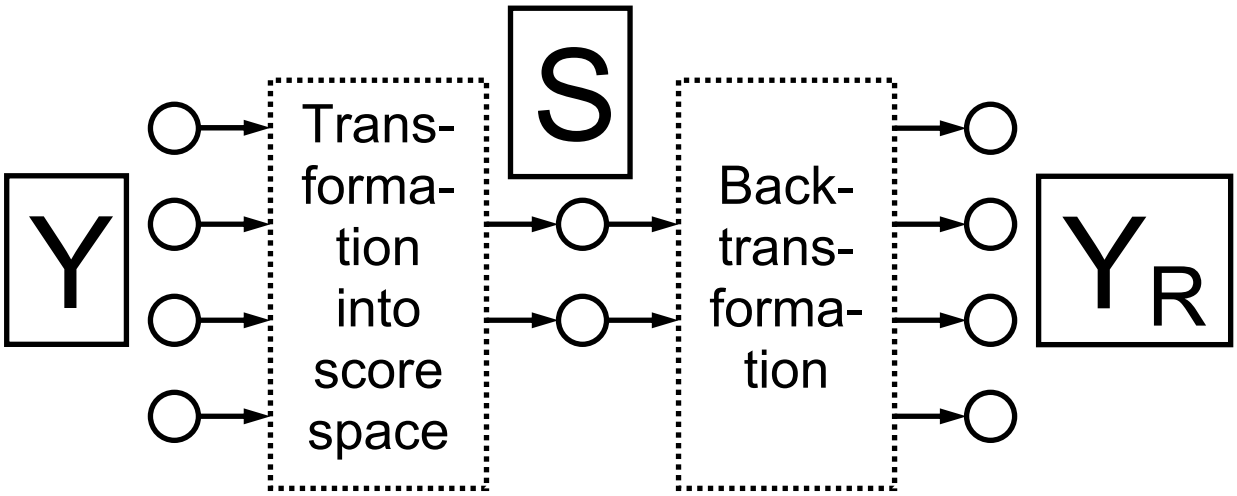

Figure 1. General concept of noise filtering by means of the PCA techniques described in the text. The input signals on the left are transformed into the scores space and then mapped back into the original data space.

Thus, the classical PCA technique as described so far only acts as filtering procedure and can not be used directly to estimate the biomass concentration. However, with PCA we determine the transformation matrix P. This can be used for biomass estimation in following way:

When we augment the data vectors [OUR, CPR, Base] used in the multivariate regression methods by the corresponding biomass concentration measurements, we can compute the PCA transformation matrix P and its inverse P^{-1}, as well as the corresponding restricted matrices S_R and P_R. According to Equation (13), this allows mapping the original data back again into the original data space spanned by [OUR, CPR, Base, X].

As all the 4 input variables are correlated with each other, an error in one input variable, e.g. the X-value, has the consequence that the corresponding score vector cannot be mapped back onto the input vector. However, one can easily recover the disturbed X-value. With a simple search one can find that X-value that fits best to the values of the other variables so that the entire input vector will again be mapped onto itself. This technique can be used to replace a faulty measurement value. This approach does not suffer from the disadvantage of dealing with "latent variables" that lack a physical meaning. Hence, the disadvantage of PCA of dealing with latent variables is removed.

In the context of biomass estimation, the faulty or even not measured variable is X. By means of a simple line search algorithm at every time instant, the X that fits best into the set of the actual online measured OUR, CPR, and Base values can be obtained, so that the reconstructed measurement values (Figure 1) depict a minimal deviation from the original ones:

$$|\mathbf{Y}-\mathbf{Y_R}| \rightarrow \min \qquad (14)$$

In this way the entire procedure can be used to estimate the biomass concentration X from the other online values.

Biomass estimation using artificial neural network

In order to express nonlinear relationships between process variables one often uses artificial neural networks (ANN) (Figure 2a). The number of applications of this technique is increasing and large number of software tools is available for establishing and training ANNs. For biomass estimation using ANNs we suggest two approaches: the direct mapping of the cumulative values of the online measurable quantities OUR, CPR, and Base onto biomass concentration values by means of a feed forward artificial neural network, and a technique based on the idea of auto associative artificial neural networks (AANN).

In the first approach, the online measured variables are directly mapped onto the biomass concentration X. This can be done with a simple feed forward ANN with 3 input nodes plus one bias node and some hidden layer nodes to estimate the single output value X. The number of nodes in hidden layer can be determined empirically using trial and error methods. In our investigations we found by cross-validation, that 4 neurons in hidden layer are sufficient to cover the nonlinearities between the observed input variables and biomass. The bias node is a constant input of the ANN, which becomes necessary to be able to model situations where the inputs of the ANN are zero. It then keeps the output values finite (e.g., Haykin 1999).

The second approach investigated is based on nonlinear principal component analysis (Hiden *et al*. 1999) using auto associative artificial neural networks (AANN). It has been shown that AANNs are useful in various recognition tasks including monitoring of chemical processes (Kramer 1991). AANNs are networks where the target output pattern is identical to the input pattern (Figure 2b). The network contains an internal "bottleneck" layer with fewer nodes than in the input layer, which forces the network to develop a compact representation of the input data. For the training of the AANN the same algorithms as for classical ANN can be used.

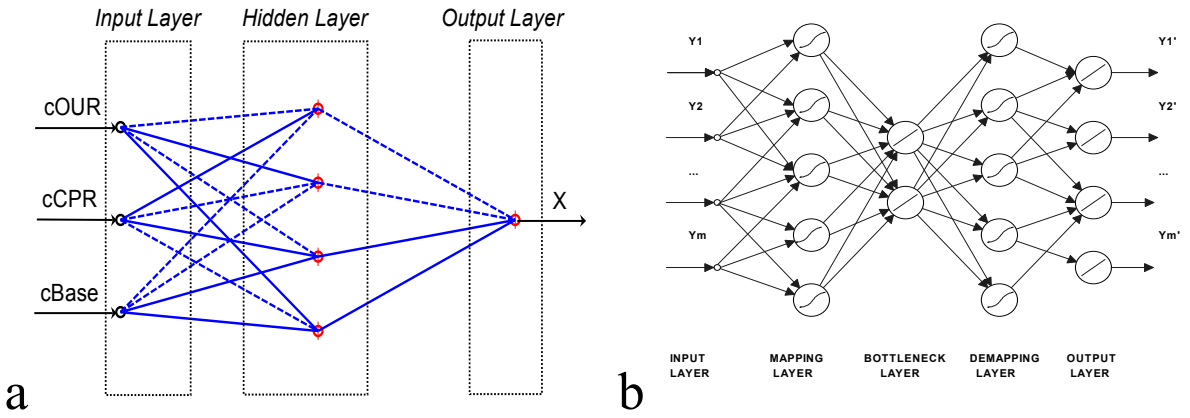

Figure 2. Structure of the ANN (a) and auto associative artificial neural network (b).

Auto associative network were used to estimate measurement values in cases where the physical sensors are distorted or even missing (Dong and McAvoy 1994). For biomass estimation one can use an AANN in the same way as discussed in the preceding section in connection with PCA transformations. The difference is that AANNs perform nonlinear mappings. In this application the design of the ANN underlying the AANN was based on a cross validation procedure. Its 4 inputs are the cumulative signals of OUR, CPR, and Base, as well as the X signal, with 4 hidden layer nodes and 2 nodes in the bottleneck layer. After training, the auto associative network was used to estimate biomass in exactly the same way as the P and P^{-1} matrices described for the PCA transformation.

3 RESULTS AND DISCUSSIONS

20 complete data records from *E.coli* cultures run to produce the green fluorescent protein were used to determine the parameters in the relationships mentioned in the previous section. The quality of the estimates obtained with these relationships was quantified by the root mean square error (RMSE) of the deviations between the original biomass data and the corresponding estimates. The RMSE values for the different methods obtained in the experiments described before are summarized in Table 1.

Table 1. RMSE values for the different biomass estimation methods as described in text.

Estimation procedure	*Biomass estimation error RMSE [g/kg]*
Biomass estimation by a feed forward ANN	*0.67*
Polynomial regression with cumulative input variables	*0.76*
AANN-based estimation	*0.78*
Luedeking/Piret based estimate on all online measurements (weighted)	*0.85*
PCA-based estimation	*0.86*
Luedeking/Piret-like relationship and CPR measurements	*0.89*
Multi-variable linear regression with cumulative input variables	*1.18*
Luedeking/Piret-like relationship and OUR measurements	*1.49*
Luedeking/Piret-like relationship and base consumption measurements	*1.83*
Multi-variable linear regression	*2.55*

The results presented in Table 1 are thought to be of general validity because most of the methods have been applied to many other bacterial and yeast strains modified with different expression systems. The results were in full agreement with those reported here. Furthermore, the methods have also been used in cultivations in several different reactors on very different scales even on the production scale with no noticeable difference.

Figure 3 depicts the biomass concentration profiles $X(t) = x(t)/W(t)$ as a function of time for 16 data records. It compares the computed profiles (lines) with the measured data. The model parameters used in the estimations were determined with MATLAB's *fminsearch* to be $[Y_{OX};\ m_O] = [0.572;\ 0.106]$. These parameters, and, also the parameters Y_{CX}, m_C, are not independent of each other as the corresponding estimation error surface, depicted in Figure 4, shows. For the purpose of biomass estimation, however, it is not very important to reach the exact minimum on this error surface. Thus one may use slightly different combinations of Y_{CX} and m_C to get practically the same accuracy of the biomass estimate.

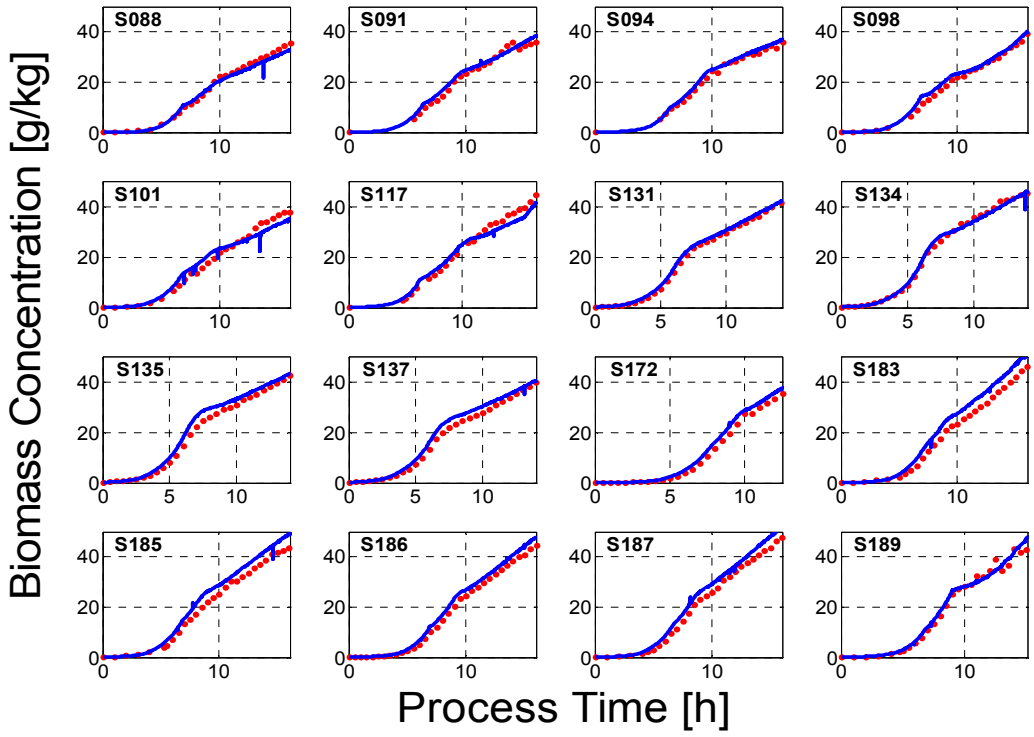

Figure 3. Utilizing oxygen uptake rate data for biomass estimation: Symbols represent the offline measured biomass concentrations X [g/kg], lines the model-based estimates.

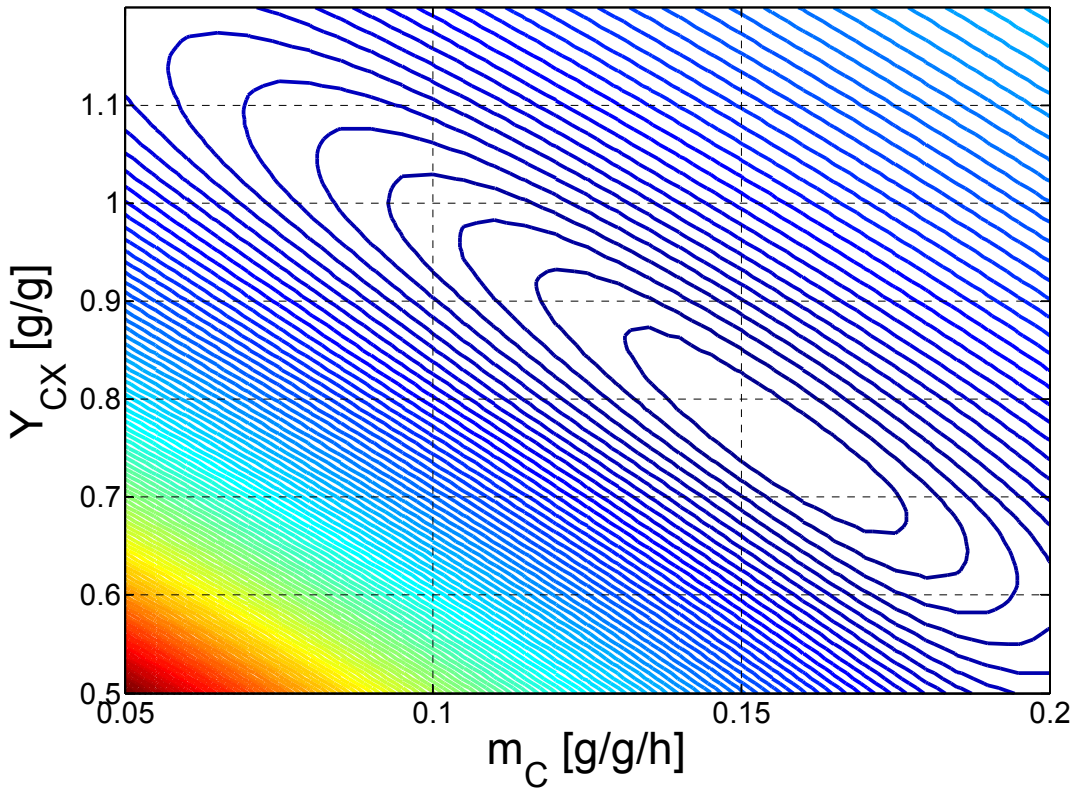

Figure 4. Error-surface for the identification of the parameters [Y_{CX} m_C] of the model relating the biomass to the CRP data. The absolute minimum is difficult to obtain as different parameter combinations lead to similar identification errors.

The biomass estimates resulting from the ANN-based approach are shown in Figure 5. It becomes clear that the online biomass values obtained from ANN-estimator closely predict the biomass concentrations which are measured offline much later.

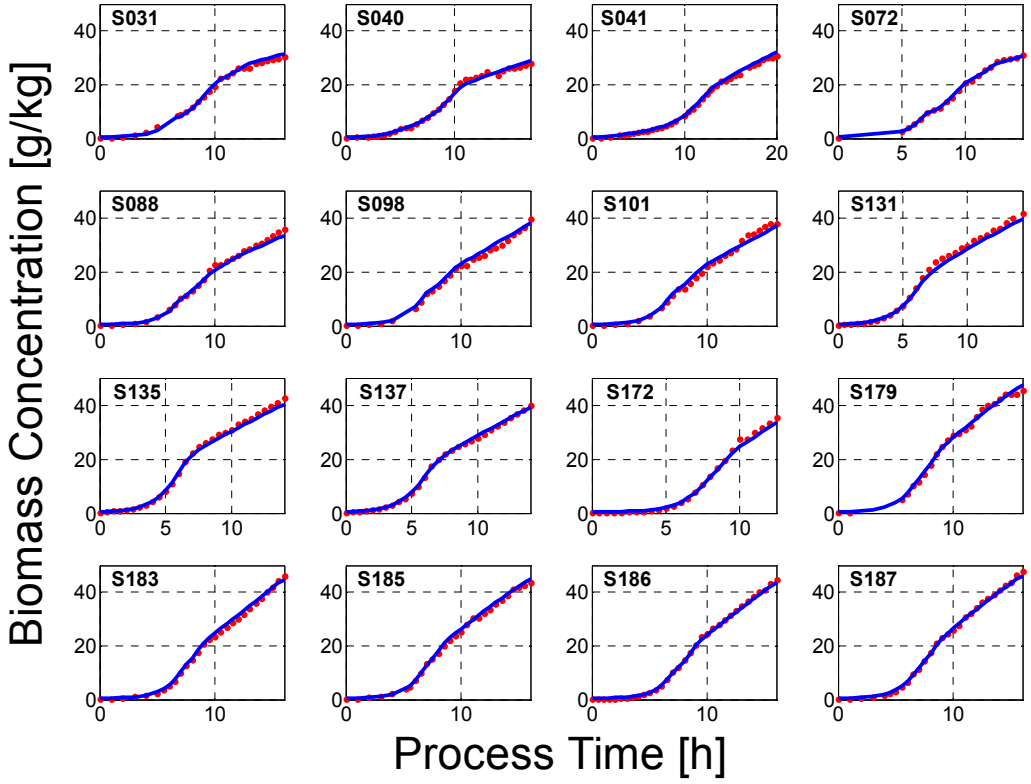

Figure 5. Biomass estimation by means of a simple feed forward artificial neural network with a single hidden layer (lines … estimates; symbols … offline measurements).

The main selection criteria for an application of one or the other of these analyzed techniques in real application are:

1. Biomass estimation performance (Accuracy, RMSE)
2. Simplicity of the computations

By simplicity of the computations we mean simplicity of an adaptation of the estimators upon small changes in the measurement hardware and the operational procedure. This is quite easy with the procedures discussed here which are basically static relationships, but would be very cumbersome with advanced model-based state estimators such as Extended Kalman Filters.

From that point of view, the simple feed forward ANN proved to be best in the experimental environment used for this investigation.

The final result of this investigation is: When the process is operated well insides the regime which was experienced before, i.e. from which fermentation data are available simple feed forward ANNs provide the best biomass estimates.

The decision whether or not an actual cultivation data set reflects normal behavior of the process can be made using PCA techniques. For that purpose one compares its trajectory in the score space with the mean trajectory computed from all available data records taken from experiments that did not depict significant faults.

In order to demonstrate such a case, an experiment was performed with the same biological system but with different initial conditions. Preculture conditions and the operating conditions in the first cultivation phase after inoculation were different as compared to the previously performed experiments: The fermentation was started without an initial batch phase and without any substrate in the initial medium. Substrate was only added with the feed.

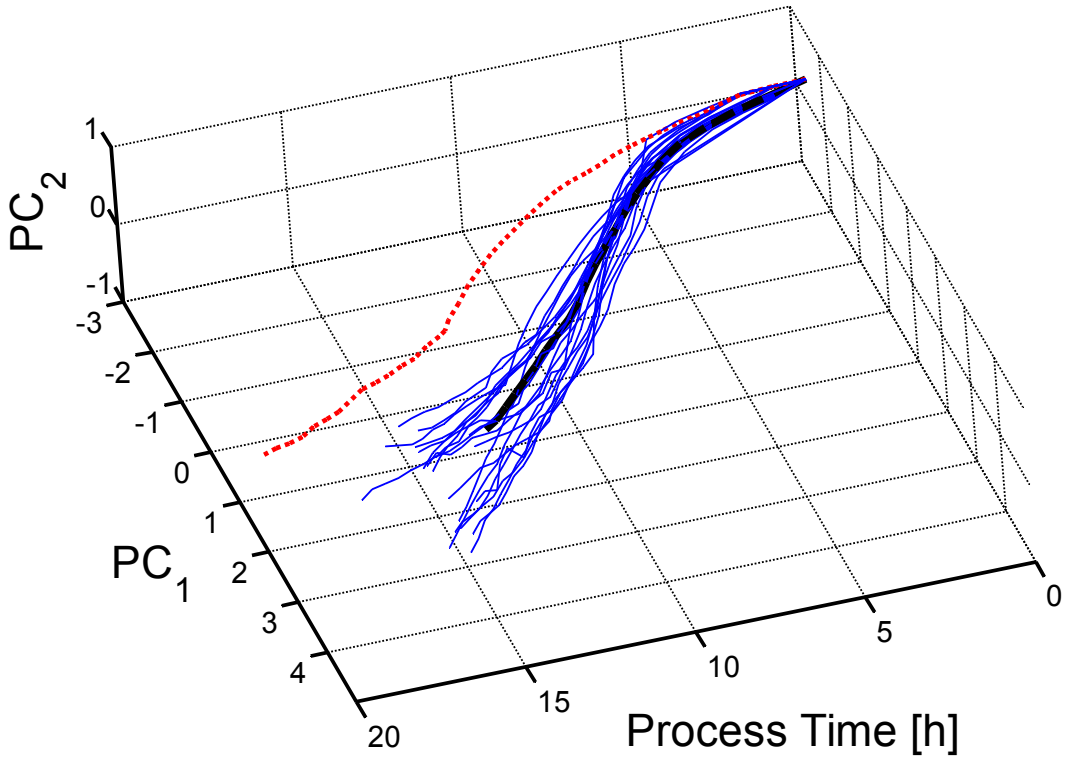

Figure 6. Trajectories of the 20 experiments in a three-dimensional space constructed from the two scores and the fermentation time (full lines) together with the mean of these experiments (bold line). Additionally an experiment made under completely different conditions (dashed line) is shown to deviate considerably from the group of normal experiments.

This different production run can immediately be recognized as such by means of a simple principal component analysis. This is depicted in Figure 6. The trajectory for this special experiment is seen to be significantly different to the others. By means of a least square deviation from the average trajectory, this particular trajectory can be identified to be strange.

An immediate consequence of the different fed-batch strategy is that the biomass estimation by means of the artificial neural network trained on data obtained from normal experiments is of significantly lower performance as the network was trained on a much different data set. The estimates for this case are depicted in Figure 7.

When, in a practical application, such a situation appears and a simple PCA analysis detects the fault, the straightforward reaction is to employ a more robust but less accurate biomass estimation technique. For this purpose, a model based estimation, for instance the estimator based on the Luedeking/Piret-type relationship can be used. This is depicted in Figure 7. While it is of lower accuracy in normal situations it performs better now in fault situations. The reason is that with such model-based approaches, missing information in the data is compensated for by the a priori knowledge supplied by the model.

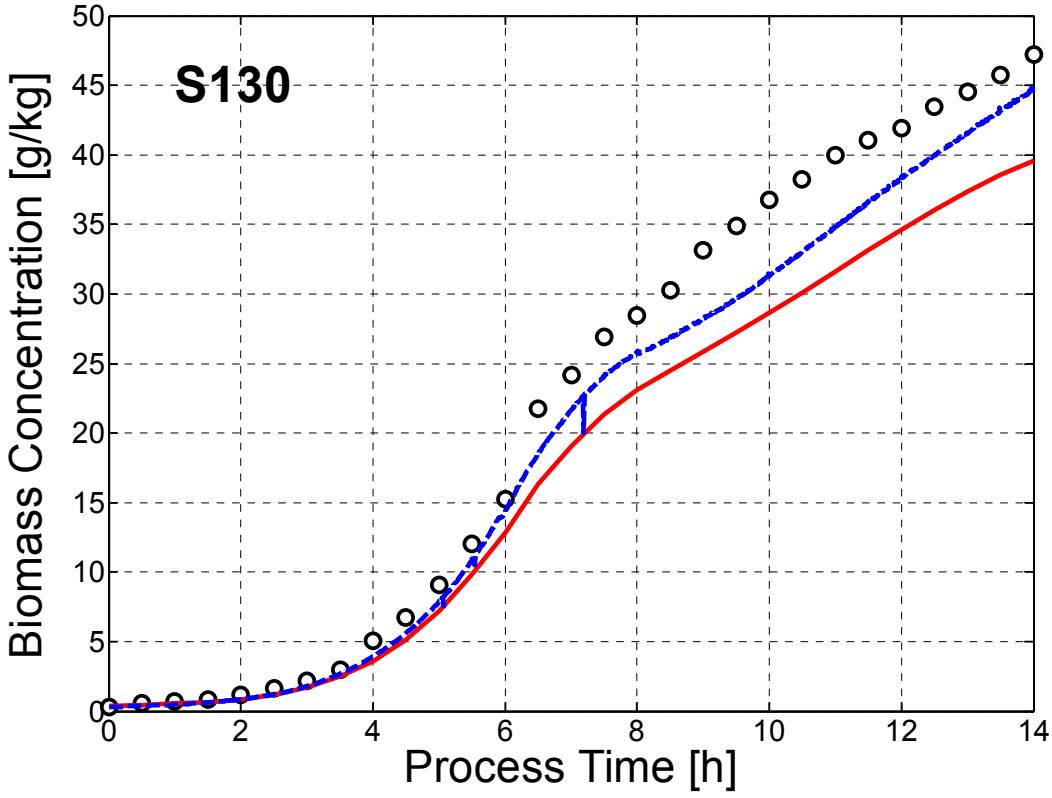

Figure 7. Estimate of the biomass by means of an ANN (full line) together with the estimation result from Luedeking-Piret relationship based on oxygen uptake rate data (dashed line) for the experiment that was performed under significantly different conditions. Offline measured biomass concentration data are given by the symbols.

In practice it is straightforward to combine the techniques discussed. A PCA should be used for fault analysis, or better for making sure that the current process is in control. When the current fermentation does not deviate from the set of "typical" fermentations, the artificial neural network is best used to estimate the biomass concentration. If not, the simple model-based approach should be used instead.

By obvious reasons, the estimation techniques presented in this paper were discussed at the example of laboratory-scale fermentation equipment. However, all the techniques have also been applied to large industrial production fermenters. The results were qualitatively of the same performance.

Here we discussed several estimation techniques that can immediately be used to estimate biomass concentrations in industrial production environments. There are other methods discussed in literature, for the most part more advanced methods such as nonlinear observers, Extended Kalman Filters, etc.. To our experience they are not used in biotechnological practice as they are too complicated. Hence, there is a fear that they cannot be maintained by the plant personnel. This reduces the acceptance of these methods considerably.

Acknowledgements. Part of the work presented has been financed by the "Kultusministerium des Landes Sachsen-Anhalt". We gratefully acknowledge this support.

REFERENCES

Dong, D., McAvoy, T.J. (1994), Non-linear principal component analysis – based on principal curves and neural networks, *Proc. Acc.*, 2, 1284-1288

FDA (1987), *Guideline on general principles of process validation*, Food and Drug Administration, Rockville, MD

FDA (2003), *Guidance for Industry: PAT - A Framework for innovative pharmaceutical manufacturing and quality assurance*, Food and Drug Administration, Rockville, MD

Hiden, H.G., Willis, M.J., Tham, M.T., Montague, G.A. (1999), Non-linear principal components analysis using genetic programming, *Comp. Chem. Eng.* 23, 413-425

Haykin, S. (1999), *Neural Networks, 2nd edn.* Prentice Hall, Englewood Cliffs

Jackson, J.E. (1991), *A User's Guide to Principal Components*, Wiley, New York

Jenzsch, M., Simutis, R., Lübbert, A. (2005), Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9),* Pons MN, van Impe JFM, eds., IFAC/Elsevier, ISBN 0 08 044251 X

Jenzsch, M., Simutis, R., Lübbert, A. (2006), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.* 122(4), 483-493

Jollife, I.T. (1986), Principal Component Analysis, Springer, New York

Kourti, T., MacGregor, J.F. (1995), Multivariate Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods – A Tutorial, *Chemometrics and Int. Lab. Sys.* 28, 3

Kramer, M.A. (1991), Nonlinear Principal Component Analysis Using Auto Associative Neural Networks, *AIChE J.* 37/2, 233-243

Luedeking, R., Piret, E.L. (1959a), A kinetic study of the lactic acid fermentation batch process at controlled pH, *J.Biochem. Microbiol. Technol. Eng.* 1, 393-412

Luedeking, R., Piret, E.L. (1959b), Transient and steady state continuous fermentation: Theory and experiment, *J. Biochem. Microbiol. Technol. Eng.* 1, 431-459

Nomikos, P., MacGregor, J.F. (1994), Monitoring Batch Processes Using Multiway Principal Component Analysis, *AIChE J.* 40/8, 1361

Sambrook, J.E., Russell, D.W. (2001), *Molecular Cloning*, 3rd edn., Cold Spring Habor Laboratory Press

Shuler, M.L., Kargi, F. (2002), *Bioprocess Engineering Basics Concepts*, 2nd edn., Upper Saddle River: Prentice Hall

Chapter 4

Generic Model Control of the Specific Growth Rate in Recombinant *Escherichia coli* Cultivations

Abstract. Generic model control is shown to be a powerful tool for keeping a microbial cultivation process close to its predetermined (optimized) control profile. This is demonstrated at the example of the green fluorescent protein (GFP) expressed in genetically modified *E.coli* host cells. It is shown that the process can be run very closely to a predefined complex profile of the specific cell growth rate μ(t). Controlling the experiments at many different growth conditions is a straightforward way of effectively collecting the data necessary for optimization of recombinant protein production systems. Although the process dynamics is rather complex, the model for the controller can be kept quite simple. The control technique, used here for specific growth rate control, is quite universal and can be applied for different biotechnological processes as well.

This chapter has been published in *Journal of Biotechnology*:
Jenzsch, M., Simutis, R., Lübbert, A. (2006), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

1 INTRODUCTION

Every-day task of bioprocess engineers in recombinant therapeutic protein production is improving performance and robustness of their processes. As the cells are not naturally optimized for producing recombinant proteins, protein production processes must be controlled by appropriately adjusting the cells' environment. In most production systems, the substrate feed rate is used to control growth and product formation rates of the organisms.

A general objective of process control in this field is to guarantee that the process will yield the desired amount of product at the end of the cultivation. It depends on the time integral of product of the biomass concentration X(t) and the specific product formation rate $\pi(t)$ along the production path. In most applications, the specific product formation rate $\pi(t)$ primarily depends on the specific growth rate μ. As the increase of the biomass concentration X also depends on μ, it is straightforward to control the value of μ during protein production. This can be done by appropriately changing the substrate feed rate F to the bioreactor.

Traditionally, protein production processes are controlled in an open-loop manner. This means that the control profile F(t) is determined beforehand and applied during the cultivation process without any feedback to the control device. Usually, however, cultivations are subject to distortions, may they result from deviations in the initial conditions, - e.g. the initial substrate or biomass concentrations, - from those assumed during the feed rate profile determination or from disturbances appearing during the process, - for instance feed rate variations or problems with the oxygen mass transfer.

Randomly appearing disturbances can only be coped with by means of feedback control. The main problems with feedback control in biotechnical production processes are:

- Lack of the information about the current value of the controlled variable or the state of the process. For instance, the current value of the crucial process variable - specific growth rate μ - cannot be measured with commercially available sensors. Instead, it must be determined indirectly from the online signals measured at the bioreactor. In order to determine its current value, a reliable quantitative relationship must be known that relates its current value to the available measurement data. This is one reason why we need process models in bioprocess engineering.
- Complicated and nonlinear dynamics of biotechnological processes. For instance, the simple control algorithms (P, PI, PID controllers) don't work properly in recombinant protein production systems because the dynamics of these processes are changing with time in a nonlinear way. This would require changing the control parameters automatically during the process.

To solve these problems and to control the biotechnological process reliably, special state estimation algorithms and adaptive or model based control algorithms are necessary. One of the most powerful tools for process state estimation in biotechnology remains the well know Extended Kalman Filter (EKF) algorithm (Stephanopoulos *et al.* 1993; Grewal and Andrews 2001; Simutis *et al.* 1993). In this paper we employ the EKF algorithm and show that it can be very useful in estimating the specific growth rate in recombinant protein production process.

To keep the process close to the predetermined trajectories, different schemes of model based control algorithms have been proposed (Istre 2004, Lee *et al.* 1999; Lee and Sullivan 1987). During the past decade there has been increasing awareness and use of model predictive approaches. Predictive control approaches are easy to understand and they provide a straightforward way to explicitly handle constraints. In model predictive control (e.g., Morari and Lee 1999) a dynamic process model is used to predict at the actual time t_a the effect of a

couple of future moves $F_i(t)$ of the manipulated variables on the process' behaviour within a time horizon $[t_a\ t_a+\Delta t]$. The result is a set of trajectories of the controlled variable, here $\mu_i(t)$ over this time horizon. For all of them the least square deviation from the desired trajectory $\mu_{set}(t)$ is computed. From them the j-th one, depicting the smallest deviation is chosen. Of the chosen control profile $F_j(t)$, only the first element $F_j(t_1)$ is taken as the value for the actuator of the controller at the next time step. This process is iteratively repeated after moving the time horizon one time step ahead, i.e. the controller works with a so-called receding time horizon. A major difficulty with this method is that it needs a reliable optimization procedure. Further it requires some computational time, which could be an issue when dealing with complex models and fast process dynamics. In these cases Generic Model Control (GMC) can serve as an alternative process control method (Lee and Sullivan 1987; Lee and Newell 1989).

As a model aided technique, generic model control also uses a process model for formulating the control law. Here, the control action is computed immediately from a one-step-ahead prediction/estimation. Generic model control algorithms are already successfully used in different applications (Lee 1993). There are also some trials to use generic model control algorithms in biotechnology: In their simulation study, DeLisa *et al.* (2001) discussed the efficiency of the GMC method for controlling induced protein expression in high cell density cultivations of *Escherichia coli*.

In this article we practically demonstrate for the first time the applicability of generic model control for guiding recombinant protein production processes along predefined profiles of the specific growth rate $\mu(t)$, provided a powerful estimation technique can be made available for estimating the key process characteristics online. We show that an extended Kalman filter (EKF) can be used for estimating the current biomass concentrations X and the specific biomass growth rates μ.

2 MATERIALS AND METHODS

Strain and Cultivation Technique

E.coli BL21 pET11a EGFP (DE3) 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.* 1989). 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 B. Braun 10-L-bioreactor 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. Then, after a small time delay, protein formation starts.

B. Braun's DCU was used as front-end controller. The entire process was monitored under control of MFCS/win2.0. The generic model 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 CO_2. pH was measured by means of an Ingold pH probe (Mettler Toledo). The dissolved oxygen concentration was monitored with an Ingold pO_2 probe (Mettler Toledo). pO_2 was controlled by the DCU-front-end-controller. A simple cascade control was employed where first the air flow rate and subsequently the stirrer speed was enhanced. Substrate addition was recorded by a balance (Sartorius). Base addition is also measured quasi online. These signals were filtered and provided to the controller with time increments of 36 [s].

Off-line measurements were performed with a time increment of about half an hour. Biomass concentration was estimated from an OD_{600} measurement performed with a spectral photometer (Shimadzu UV-2102PC). Additionally, some dry weight measurements were performed in order to test the correlation between biomass dry weight and the OD_{600} values. An YSI 2700 analyzer was employed to measure the glucose concentration enzymatically. The GFP concentration was estimated from the fluorescence intensity measured from samples with fully intact cell in a spectro-fluorimeter (Hitachi F-2500). The OD_{600} values as well as the glucose concentrations estimates are available about 5 minutes after drawing the culture sample. They were made available to the control program immediately after the values were known.

GMC Algorithm for Specific Growth Rate Control

In generic model control the nonlinear process model is directly incorporated into the control algorithm. The design framework is similar to other model based approaches such as internal model control (Ogunnaike and Ray 1994). However, rather than adopting a classical approach of comparing the trajectory of the process output against some desired trajectory, GMC defines the performance objective in terms of time derivatives of the process output, i.e., it minimizes the difference between the desired derivative of the process output and the actual derivative.

For the case of the specific growth rate control, we can choose the following approximation for the derivative of the desired (index d) specific growth rate trajectory for the time derivative of μ

$$\left(\frac{d\mu}{dt}\right)_d = k_1(\mu_S - \mu) + k_2\int_0^t(\mu_S - \mu)dt \qquad (1)$$

where k_1 and k_2 are design constants (to be specified) and μ_s is the setpoint profile for specific growth rate μ. This equation expresses the following requisitions to control algorithms: (i) The more the actual μ deviates from μ_s, the more quickly it should be drawn back to μ_s and (ii) the longer the deviation from the setpoint lasts, the more rapidly should it be moved back to μ_s. As shown by several authors, (for a recent compilation of the literature see Istre 2004), with this approach for the derivative, the value of μ will always converge to the setpoint value μ_s.

On the other hand we can obtain a corresponding equation for the derivative of the specific growth rate from process model. As we are interested in developing a universal algorithm for

specific growth rate control, the process model must be simple and universal. A wide class of fed-batch cultivation process can be described by the following simple set of mass balance equations:

$$\begin{aligned} \frac{dX}{dt} &= \mu X - \frac{F}{W} X \\ \frac{dS}{dt} &= -\sigma X + \frac{F}{W}(S_F - S) \\ \frac{dW}{dt} &= F \end{aligned} \tag{2}$$

where X is the concentration of biomass, S the concentration of the substrate, W the culture weight, F the manipulated variable, the substrate feed rate and S_F the substrate concentration in the feed. When the product concentration is much smaller than the biomass concentration, and this is the case in most recombinant protein formation experiments in practice, a separate equation for product formation is not necessary in such a general mass balance. For simplicity reasons such a balance is thus omitted in equation (2).

$\mu = Y_{XS}\,\sigma$ is the specific biomass growth rate, and σ the specific substrate consumption rate. Most often in practice, σ can be described by simple kinetic laws, e.g., by one considering substrate limitation and inhibition only:

$$\sigma = \frac{\sigma_{\max} S}{K_S + S + S^2 / K_I}, \tag{3}$$

Y_{XS} , σ_{max},K_S , K_I are model parameters.

With this assumption we can determine the time derivative of μ using chain rule:

$$\frac{d\mu}{dt} = \frac{d(Y_{xs}\sigma)}{dt} = Y_{xs}\sigma_{\max}\frac{\partial\sigma}{\partial S}\frac{dS}{dt}, \tag{4}$$

In our case

$$\frac{\partial\sigma}{\partial S} = \frac{(K_S - S^2 / K_I)}{(K_S + S + S^2 / K_I)^2}\ ,\ \frac{dS}{dt} = -\sigma X + \frac{F}{W}(S_F - S) \tag{5}$$

and

$$\frac{d\mu}{dt} = Y_{xs}\sigma_{\max}\frac{(K_S - S^2 / K_I)}{(K_S + S + S^2 / K_I)^2}\left(-\sigma X + \frac{F}{W}(S_F - S)\right) \tag{6}$$

A controller must force the system to follow the desired μ_s trajectory as closely as possible. Hence its task is to keep the difference between both approaches for $d\mu/dt$ as small as possible. Thus the performance index J to be minimized can be formulated as:

$$J = \int_0^t \left(\left(\frac{d\mu}{dt} \right)_d - \frac{d\mu}{dt} \right)^2 dt \tag{7}$$

The performance index (7) is minimal when

$$\left(\frac{d\mu}{dt} \right)_d = \frac{d\mu}{dt} \tag{8}$$

This equation is used to define the expression for the GMC control law:

$$\begin{aligned} &k_1(\mu_S - \mu) + k_2 \int (\mu_S - \mu) dt = \\ &Y_{XS} \sigma_{\max} \frac{\left(K_S - S^2 / K_I\right)}{\left(K_S + S + S^2 / K_I\right)^2} \left(-\sigma X + \frac{F}{W}(S_F - S) \right) \end{aligned} \tag{9}$$

and

$$F = \left(\sigma X + \frac{k_1(\mu_s - \mu) + k_2 \int_0^t (\mu_s - \mu) dt}{Y_{xs} \sigma_{\max} \frac{(K_S - S^2 / K_I)}{(K_S + S + S^2 / K_I)^2}} \right) \frac{W}{(S_F - S)} \tag{10}$$

Because $\mu = Y_{XS}\, \sigma$, the equation for feeding rate can be modified as follows:

$$F = \left(\frac{\mu}{Y_{XS}} X + \frac{k_1(\mu_s - \mu) + k_2 \int_0^t (\mu_s - \mu) dt}{Y_{xs} \sigma_{\max} \frac{(K_S - S^2 / K_I)}{(K_S + S + S^2 / K_I)^2}} \right) \frac{W}{(S_F - S)} \tag{11}$$

In experiments with stepwise changing setpoint profiles $\mu_s(t)$ it was found that the quality of control is better if one uses μ_s instead of μ in the first term of equation (11):

$$F = \left(\frac{\mu_s}{Y_{XS}} X + \frac{k_1(\mu_s - \mu) + k_2 \int_0^t (\mu_s - \mu) dt}{Y_{xs} \sigma_{\max} \frac{(K_S - S^2 / K_I)}{(K_S + S + S^2 / K_I)^2}} \right) \frac{W}{(S_F - S)} \tag{12}$$

Expression (12), which contains two terms, can easily be interpreted. The first term represents the feed rate derived from the model (2) under the assumption of an undisturbed process. This is feed forward control. The second term can be interpreted as a "PI-like" feedback controller, correcting for the deviations in μ due to process disturbances and slight model/process mismatches. In other words, this is a feedforward/feedback-controller. As can easily be seen, the controller parameters of the "PI-like" feedback controller part are dynamically changing: This is expressed by their common denominator, which is governed by the substrate consumption kinetics.

The proposed specific growth rate control algorithm is quite universal and can be applied for process control purposes in different biotechnological processes.

In order to apply this control algorithm in a real production process, it is necessary to estimate or to measure the process state variables X, S, W, and, of course, the current value of the controlled variable μ. The state estimation algorithm used in this paper is analyzed in experimental part.

3 SIMULATION STUDIES ON GMC ALGORITHM

Process Model

Prior to tests in real fermentations, the generic model control algorithm was studied by means of computer simulations. The GFP production process with genetically modified *E.coli* bacteria was analyzed and a suitable process model was constructed. For process simulation the following model was used:

$$\begin{aligned}
\frac{dX}{dt} &= \mu(S)X - \frac{(F + F_1)}{W} X \\
\frac{dS}{dt} &= -\sigma(S)X + \frac{F}{W} S_F - (F + F_1)\frac{S}{W} \\
\frac{dP_X}{dt} &= \pi(\mu, P_X) \\
\frac{dG}{dt} &= \gamma(\mu, G) \\
\frac{dW}{dt} &= F + F_1 + F_2
\end{aligned} \tag{13}$$

where X, S, P_X are biomass, substrate and specific product concentrations, G is the relative plasmid copy number and W is the reactor weight. S_F is the substrate concentration in the feed, F is the substrate feed rate, F_1 considers the changes in culture mass through base addition, CO_2 outflow through the vent line, and water evaporation. These components are computed using separate algebraic equation as described by Galvanauskas *et al.* (1998). F_2 considers the culture mass changes due to sampling for offline analysis.

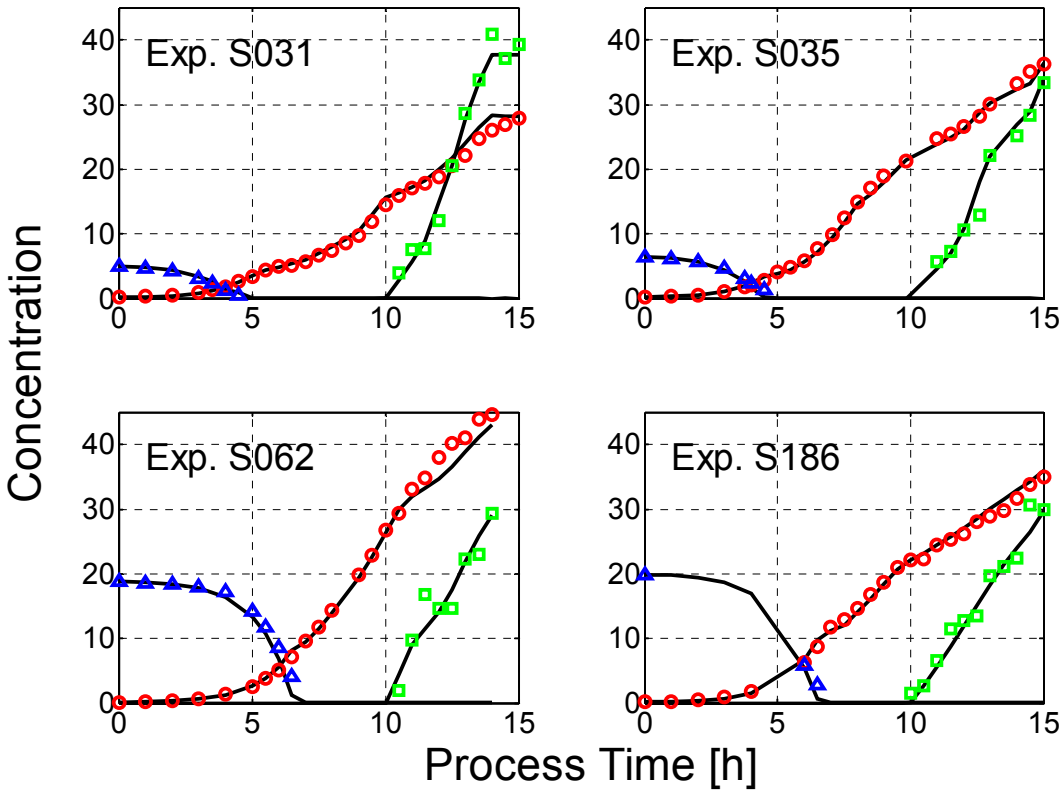

Figure 1. Comparison of experimental data (symbols) and simulation (lines) of the process model used for controller tests (Δ … substrate conc. [g/kg], ○ … biomass conc. [g/kg], and □ … specific product conc. [rFU/g biomass]).

The kinetic expressions were based on the specific substrate consumption rate

$$\sigma = \frac{\sigma_{\max} S}{K_S + S + S^2 / K_I}, \quad \mu = Y_{XS}\,\sigma \tag{14}$$

For the specific product development rate π we assume that the promoter is stringently controlled so that π is equal to zero before the culture was induced. After induction, π is assumed to be primarily dependent on the specific growth rate μ. Obviously the product formation rate depends on the plasmid copy number, here considered as a relative quantity. Additionally, the specific product development rate is reduced by dilution due to cell growth and by proteolytic degradation of the protein. So we describe product development by

$$\pi = \pi_{\max} \frac{G k_{pd}}{k_{pd} + \mu} - \mu P_x - k_{prot} P_x \tag{15}$$

As the specific plasmid concentration is dynamically changing, at least after induction, we need a corresponding expression for the specific relative plasmid copy number:

$$\gamma = k_3 \frac{\mu}{K_\mu + \mu} \frac{G}{K_{\mu i} + G} - \mu G \tag{16}$$

Again this is valid only for the protein formation phase. During the repression phase we assume constant plasmid concentrations, i.e., γ=0.

Table 1. Parameter values of the full process model.

Parameter	*Value*	*Dimension*
σ_{max}	*1.68*	*g/g/h*
K_S	*0.15*	*g/kg*
K_I	*16.50*	*g/kg*
Y_{XS}	*0.50*	*g/g*
π_{max}	*1.06*	*rFU/g/h*
k_{pd}	*2.19*	*1/h*
k_{prot}	*0.14*	*1/h*
k_3	*2.56*	*1/h*
Kμ	*0.34*	*1/h*
Kμi	*6.65*	*-*

The model parameters were estimated from a set of three experiments, its values are listed in Table 1. A solution of the equations (13) through (16) is depicted in Figure 1 together with the experimental result of an additional experiment performed to validate the model. As can be seen in Figure 1 the biomass X and substrate concentrations S, as well as the specific protein concentration Px are well described by this process model.

Simulation of Control Algorithm

The generic model control algorithm described in equation (12) was tested in simulation experiments where the process behavior was simulated with the model equations (13)-(16). The values of *X, S, W* and μ were used to determine the substrate feed rate necessary to keep the process at the desired specific growth rate profile. In the computations, the values of the "real" or "measured" μ were corrupted with random noise (mean M_N = 0.0 and standard deviation M_σ = *0.1* [1/h] and supplied to the controller with a time delay τ = *0.05* [h]. These disturbances are characteristic for indirect μ estimation techniques in real cultivation processes.

Figure 2 shows typical simulations results when using GMC controller (a) and standard PI-controller (b) to control the process to run along a predefined specific growth rate profile. In both cases the values of the controller parameters k_1 and k_2 were derived by random search optimization techniques. For the PI controller, the time interval from 7 to 10 [h] was chosen in this parameter optimization. Figure 3 shows the values of substrate feed rate when using GMC controller (a) and standard PI-controller (b). The simulations show, that the classical PI-controller works good in the dynamic range where it was optimized but is not able to achieve satisfactory control quality in other situations, because the process dynamics is changing severely. Generally, if the PI-controller is able to keep the process at one desired μ, it is usually not able to hold other μ values in other process phases without adapting the set of controller parameters. In contrast, the GMC controller provides a high control quality during the entire cultivation.

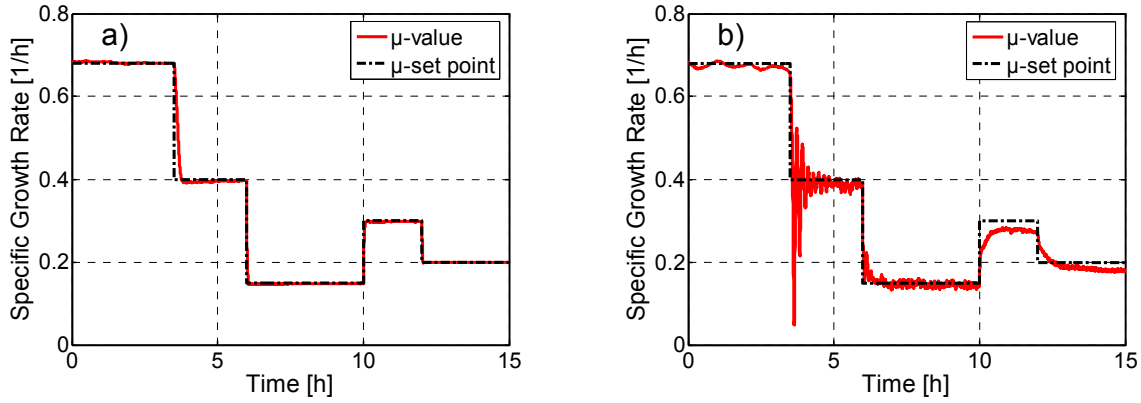

Figure 2a, b. Comparison the performance of GMC controller (k_1=*2.1,* k_2=*0.01*) and classical PI controller (k_1=*0.75,* k_2=*0.02*) when controlling the desired specific growth rate profile (μ_s ,- full line, μ - dashed line). (a) Generic model control, (b) PI-controller optimized to μ=0.13 [1/h] in the time interval 7…10 [h].

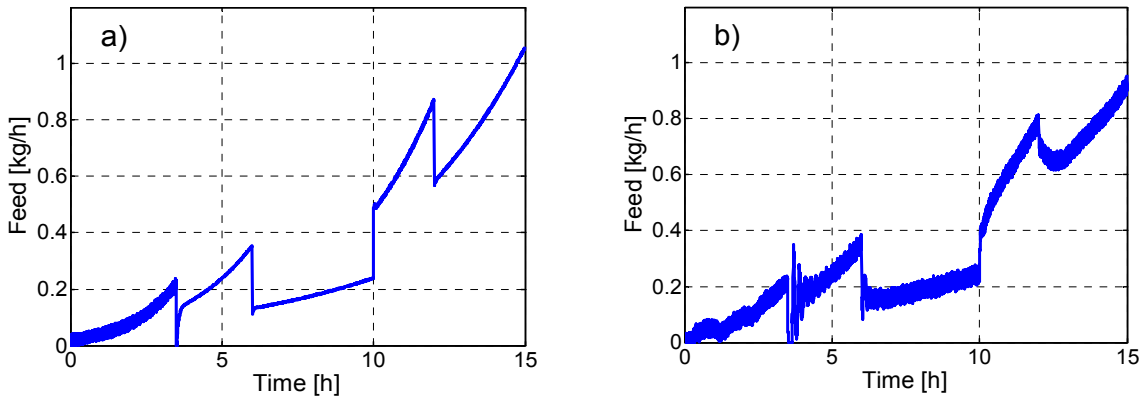

Figure 3a, b. Substrate feeding profiles corresponding to the results depicted in *Figure 2*: GMC controller (a) and standard PI controller (b).

4 RESULTS (Validation of GMC Algorithm in Real Fermentation)

State Estimation

As mentioned before, the real application of *GMC* algorithm is only possible when values of the state variables X, S, and W as well as of the specific growth rate values are available on-line during the cultivation.

Neither the state variables biomass concentration X and substrate concentration S nor the specific growth rate μ can be measured directly; hence they must be estimated from the available online signals. While W was measured online, an extended Kalman filter was applied that estimates X, S, and μ from the online measured off-gas volume ratios for O_2 and CO_2 as well as from the measured total amount of base consumed during the separately working pH control. Whenever offline measured values of the biomass concentration X become available during the cultivation, they are automatically incorporated into the online estimation procedure.

For the Kalman filter procedure we propose to use the same balance equations (2) already used by determination of GMC algorithm. In this case, however, we assume that the specific biomass growth rate μ is kept constant within the time intervals considered. Consequently the process model was augmented with a formal differential equation for specific growth rate μ:

$$\begin{aligned}
\frac{dX}{dt} &= \mu X - \frac{F}{W} X \\
\frac{dS}{dt} &= -\frac{\mu}{Y_{XS}} X + \frac{F}{W}(S_F - S) \\
\frac{dW}{dt} &= F \\
\frac{d\mu}{dt} &= 0
\end{aligned} \tag{17}$$

The measurement model was based on oxygen uptake rate (OUR), carbon dioxide production rate (CPR) and total base consumption (for pH control in reactor) measurements. Their relationships to the state variables are described in the measurement model:

$$\begin{aligned}
OUR &= \mu Y_{ox} X + m_{ox} X \\
CPR &= \mu Y_{cx} X + m_{cx} X \\
BASE &= Y_{bx}(XW - X_0 W_0)
\end{aligned} \tag{18}$$

Where Y_{ox}, m_{ox}, Y_{cx}, m_{cx}, Y_{bx} are model parameters. Its values are listed in Table 2.

Actually, the variables *OUR, CPR, BASE* serve as intermediate quantities and their values are computed via standard procedures from the original measurement signals of air flow, O_2, CO_2 and total base consumption measurements.

Table 2. Parameter values for the measurement model.

Parameter	*Value*	*Dimension*
Y_{OX}	1.100	g/g
Y_{CX}	1.275	g/g
m_{ox}	0.030	g/g/h
m_{cx}	0.065	g/g/h
Y_{BX}	1.100	g/g

Based on the process and the measurement model, Extended Kalman Filter algorithms determine estimates $\mathbf{C}_{est} = [X_{est}, S_{est}, W_{est}, \mu_{est}]^T$ of the state vector $\mathbf{C} = [X, S, W, \mu]^T$ as weighted averages between the corresponding values $\mathbf{C}_{pred}$, predicted by the model (17) and the corresponding information from measurements, represented by the difference between the actual estimates $\mathbf{H} = [OUR, CPR, BASE]^T$ and the corresponding model predictions $\mathbf{H}_{pred}$ (18) in the following way

$$\mathbf{C}_{est} = \mathbf{C}_{pred} + \mathbf{K}_g(\mathbf{H} - \mathbf{H}_{pred}) \tag{19}$$

The weighting factor K_g, referred to as the Kalman gain in literature, is estimated using standard Extended Kalman filter procedures. This includes the linearization of process and measurement model and makes use of the corresponding covariance matrices that represent the uncertainties in the model, the measurements and in the values of the state variables respectively. Extended Kalman filters are standard procedures, for details the reader is referred to basic literature (e.g. Grewal and Andrews 2001).

Application of the Controller in Real Fermentation

The validation experiments were automatically started with a batch phase overnight. In the morning, about 4 h after inoculation, the substrate feeding was started and at the same time the generic model controller was activated. Two experiments were performed to tune the controller parameters k_1 and k_2. Afterwards the processes were run with the fixed parameter set even when big changes in the μ-profile were applied. Figure 4 shows a typical result that allows to judge the actual controller performance.

It is interesting to note that the process very quickly follows the high step changes in the profile. Particularly down-up-changes led to notable overshoots in μ but these were damped down immediately so that the process proceeded with the new μ within half an hour.

Experiments, in which the μ-setpoints are changed forth and back, as depicted in Figure 4, are of importance in experiments in which the product formation kinetics is investigated in more detail (e.g. Levisauskas *et al.* 2003). Such a setpoint profile is chosen in investigations of the dependency of the specific product formation rate π as a function of the specific growth rate μ. In our specific case the step experiments were particularly used to tune the controller parameters k_1 and k_2.

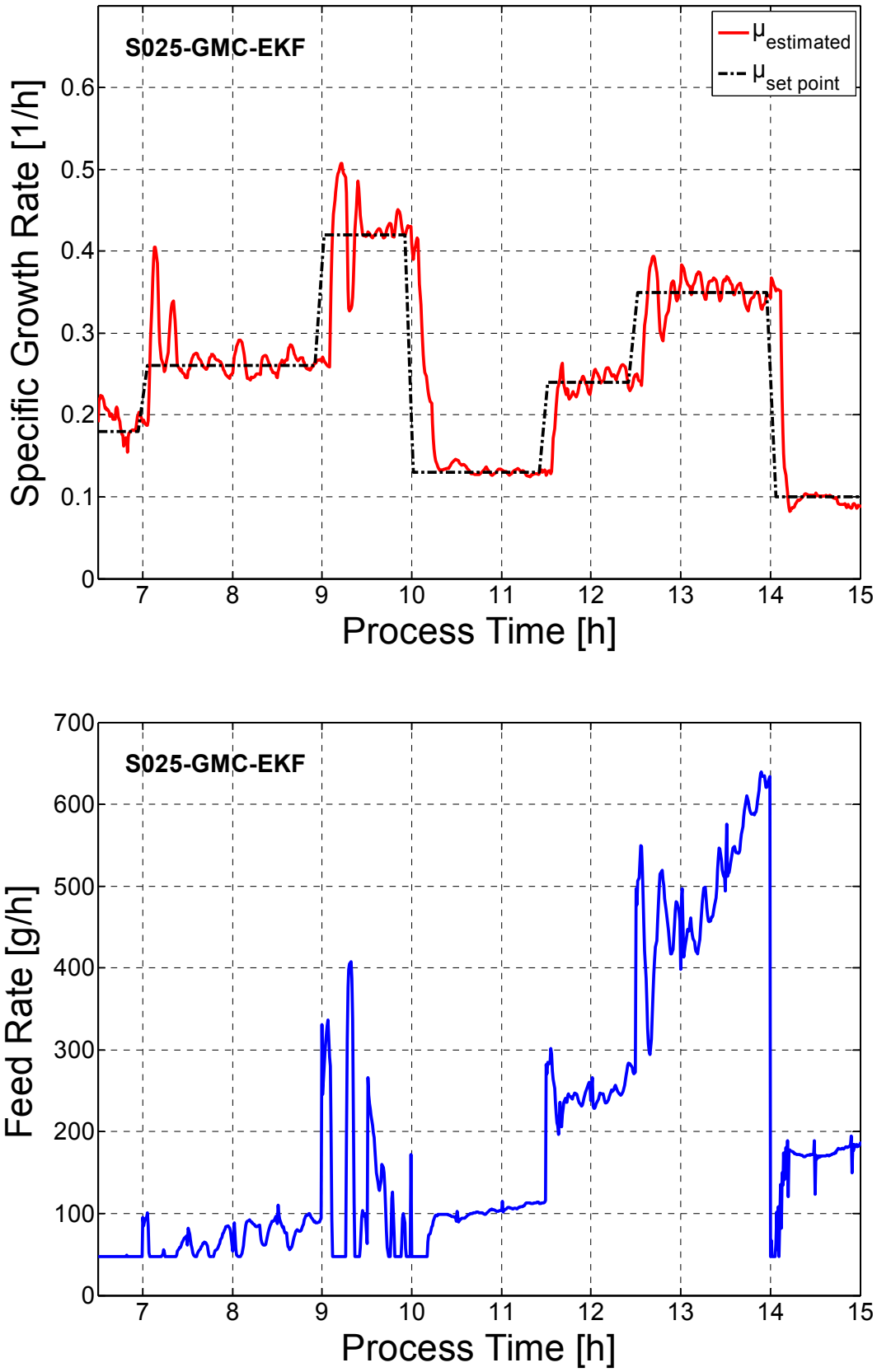

Figure 4. μ-Profile obtained in the controlled fermentation process (full line) together with the corresponding setpoint profile (dashed line). The setpoint profile is typical for investigations of the specific product formation rate π as a function of the specific growth rate μ. In the second part of the figure, the feed rate profiles F(t) are depicted set by the controller to achieve the μ-profile.

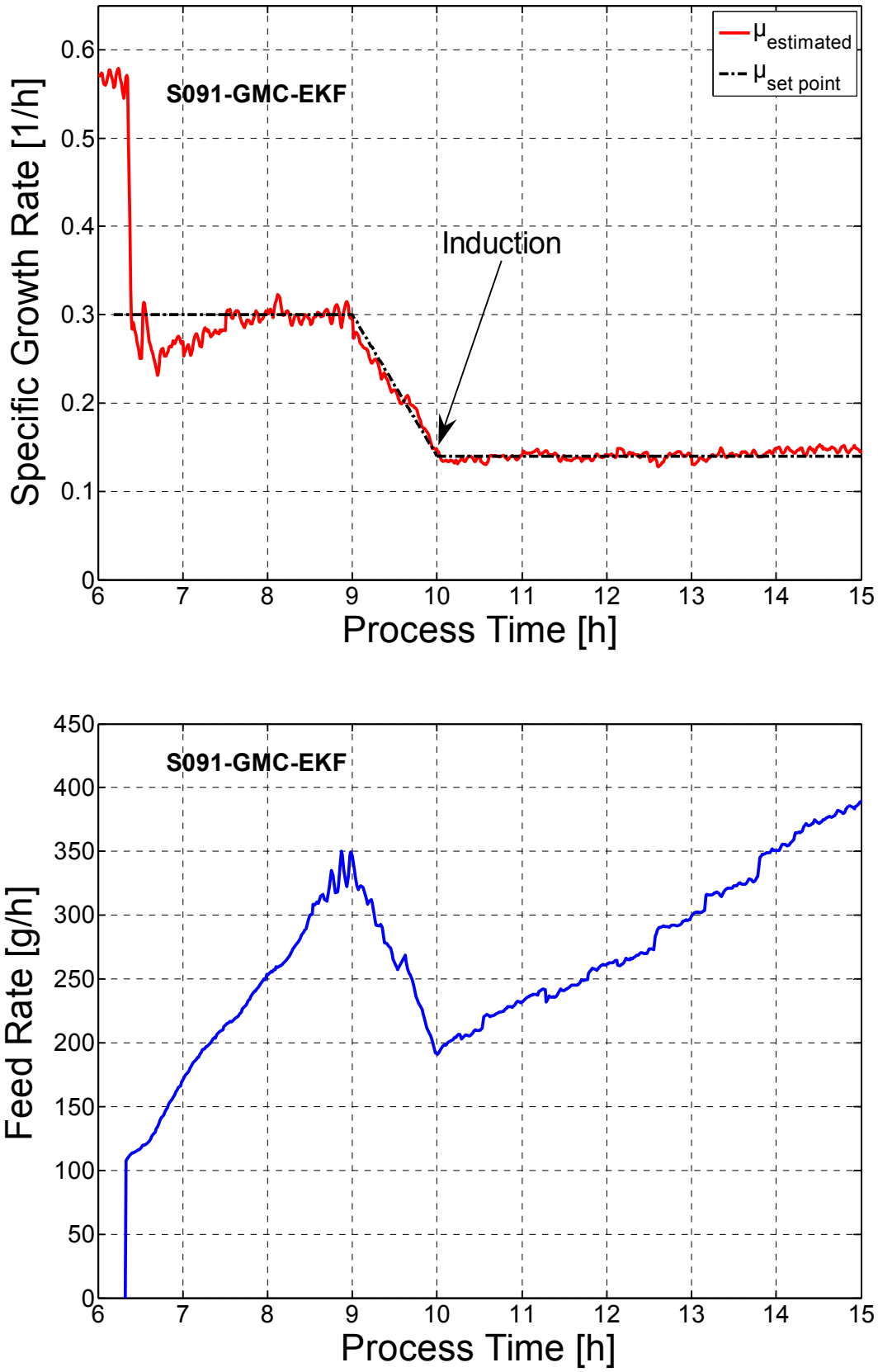

Figure 5. μ-Profile obtained in the controlled fermentation process (full line) together with the corresponding setpoint profile (dashed line). Again, in a further part of the figure the corresponding feed rate profile F(t) is depicted. Protein expression is induced in the 10th hour.

In recombinant protein production processes such step changes are performed at the induction time where the process is switched from a high growth rate during the biomass production phase to a low biomass growth rate and a corresponding high product formation rate during the production phase. In real production runs, the specific growth rate profile is chosen as shown in Figure 5. Initially it is not possible in real practice to estimate the state variables and the specific growth rate accurately enough to allow for feedback control, as the biomass is too small and the corresponding off gas values are thus unreliable. However, at the end of the biomass growth phase state estimation becomes reliable and control possible. At that time the process is usually run at higher specific growth rates. Then before induction, growth is gradually reduced to that specific biomass growth rate at which μ is intended to stay during protein production. To this value μ is controlled as long as this is possible with the culture. The full line in Figure 5 shows that the GMC controller can perfectly keep the specific growth rate μ on the desired value, in this particular case 0.13 [1/h].

The state variable biomass *X* and substrate concentration *S* corresponding to the data shown in Figure 4 and 5 are depicted in Figure 6.

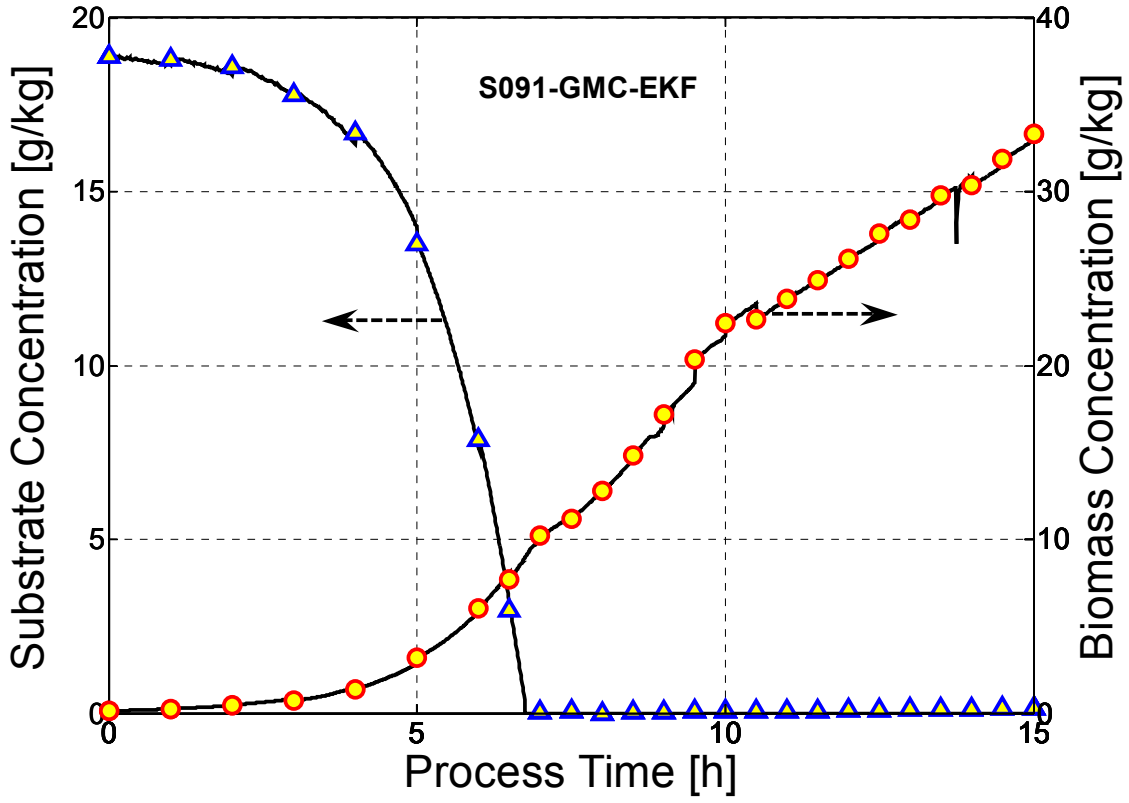

Figure 6. Biomass and substrate concentration profiles corresponding to the μ-profile depicted in Figure 5. The symbols are the estimations from the Kalman filter at times where new offline measurement values appear. The lines are model predictions in the meantime.

As can be seen from Figure 6, the relatively simple model used within the controller does all the way predict the process behavior fairly well. It is sufficiently accurate for its desired purpose: controlling μ to its setpoint profile.

5 DISCUSSION

Generic model control can be considered a simple adaptive feedback control algorithm which is combined with a feedforward controller. Its feedback part is similar to a PI controller in which the parameters are adjusted to the changing process dynamics by direct application of a simple dynamic process model.

Every controller needs reliable information on the current value of the controlled variable. The main reason for the fact that only a few cultivation processes are really closed-loop controlled is the difficulty in obtaining reliable online estimations about the process' state. Conceptually, Extended Kalman Filters are nearly ideal procedures for estimating the required values. They base their estimate on measurement information as well as on a priori model information and weight these data by the relative confidence one is attributing to measurements and model based predictions. The latter is expressed via covariance matrices.

As the results obtained in several *E.coli* cultivations showed, the GMC controller allows running the cultivation process quite closely along a predefined μ profile. The process quickly follows jumps in the μ-setpoints. As expected there are some overshoots in the real μ signals obtained indirectly from the simultaneously applied process state estimation procedure. However, the process is closely varying around the setpoint profile.

In the experiments performed, the process run automatically without operator interaction with respect to tuning control variables. The only manual activities were drawing culture samples of predefined volumes at the predefined time instances and performing the off-line measurements. The results shown in this article are typical for the application of the generic model controller at the laboratory scale. The controller was applied in several different fermentation systems where different proteins were produced with different bacterial strains and different μ-control profiles were followed. The controller performance was about the same in all experiments performed.

The proposed specific growth rate control procedure is quite universal. It includes the state estimation algorithms for basic state variables and the generic model control algorithm providing the substrate feed rates used as action variable at the controller. The procedure allows realizing complicated specific growth rate profiles during the cultivation. This feature is very important for obtaining the information about the process' dynamics needed for bioprocess optimization, in particular the dependency of the specific product formation rates from the specific biomass growth rates.

It becomes quite clear that the generic model controller leads to a good control performance. However, this is achieved at the expenditure of formulating a model and to determine the appropriate model parameters. During its application the process' state must be estimated.

Acknowledgements. The work was performed in the project "Technologies for the production of recombinant therapeutic proteins" performed 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).

NOMENCLATURE

$BASE$	Total Base Consumption [g]
CPR	CO_2 Production Rate [g/kg/h]
F	Substrate Feed Rate, Manipulated Variable [kg/h]
F_1	Mass Flow Rate in Fermenter due to Base Feed, Medium Evaporation, and Carbon Loss with CO_2 in Off-Gas [kg/h]
F_2	Sampling Rate in Fermenter [kg/h]
G	Relative Plasmid Copy Number [-]
J	Controller Performance Index, Relative Units
k_1	Controller Parameter, Proportional Part
k_2	Controller Parameter, Integral Part
k_3	*M*aximal Specific Plasmid Formation Rate [1/h]
K_I	Inhibition Constant for Substrate Consumption [g/kg]
k_{pd}	Inhibition Constant for Product Formation [1/h]
k_{prot}	Coefficient for Proteolytic Degradation of the Protein [1/h]
K_S	Monod Constant for Substrate Consumption [g/kg]
K_μ	Growth Limitation Constant for Plasmid Formation [1/h]
$K_{\mu i}$	Monod-Like Constant for Plasmid Formation [-]
m_{cx}	CO_2 Production Parameter Related to Biomass Maintenance [g/g/h]
m_{ox}	Oxygen Uptake Parameter Related to Biomass Maintenance [g/g/h]
OUR	O_2 Uptake Rate [g/kg/h]
P_X	Specific Protein Activity [rFU/g biomass]
rFU	Relative Fluorescence Units
S, S_{EST}	Real and Estimated Substrate Concentrations [g/kg]
S_F	Substrate Concentration in Feed [g/kg]
t	Current Process Time [h]
W, W_{EST}	Real and Estimated Culture Mass [kg]
X, X_{EST}	Real and Estimated Biomass Concentrations [g/kg]
Y_{BX}	Base Consumption per Biomass [g/g]
Y_{CX}	CO_2 Production Yield Related to Biomass [g/g]
Y_{OX}	Oxygen Uptake Yield Related to Biomass [g/g]
Y_{XS}	Biomass/Substrate Yield [g/g]
γ	Specific Plasmid Number Changing Rate [g/h]
μ, μ_{EST}	Real and Estimated Specific Biomass Growth Rate [1/h]
μ_S	Setpoint Value for Specific Biomass Growth Rate [1/h]
π	Specific Protein Accumulation Rate [rFU/g/h]
π_{max}	Maximal Specific Protein Accumulation Rate [rFU/g/h]
σ	Specific Substrate Consumption Rate [g/g/h]
σ_{max}	Maximal Specific Substrate Consumption Rate [g/g/h]

REFERENCES

DeLisa, M.P., Chae, H.J., Rao, G., Weigand, W.A., Valdes, J.J., Bentley, W.E. (2001), Generic Model Control of Induced Protein Expression in High Cell Density Cultivation of Escherichia coli Using On-line GFP-Fusion Monitoring, *Bioproc. Biosyst. Eng.*, 24, 83-91

Galvanauskas, V., Simutis, R., Volk, N., Lübbert, A. (1998), Model based design of a biochemical cultivation process, *Bioproc. Eng.*, 18, 227-234

Grewal, M.S., Andrews, A.P. (2001), Kalman filtering theory and practice using MATLAB. 2nd edition, NY: Wiley

Lee, J., Lee, S.Y., Park, S., Middelberg, A. (1999), Control of fed-batch fermentations, *Biotechnol. Adv.*, 17, 29-48

Lee, P.L., Sullivian, G.R. (1987), Generic Model Control (GMC), *Comput. Chem. Eng.*, 12, 573-580

Lee, P.L., Newell, R.B. (1989), Generic model control - a case study, *Can. J. Chem. Eng.*, 67, 478-483

Lee, P.L. (1993), Nonlinear Process Control: Applications of Generic Model Control, Advances in Industrial Control, Ed. Lee PL, London: Springer-Verlag

Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A. (2003), Model-Based Optimization of Viral Capsid Protein Production in Fed-Batch Culture of recombinant Escherichia coli, *Bioproc. Biosyst. Eng.*, 25, 255-262

Istre, J.M. (2004), Robust Generic Model Control for parameter interval systems, PhD-dissertation, University of Kentucky, Lexington

Morari, M., Lee, J.H. (1999), Model Predictive Control: Past, Present and Future, *Comput. Chem. Eng.*, 23, 667-682

Ogunnaike, B.A., Ray, W.H. (1994), Process dynamics, modeling, and control, NY: Oxford University Press

Sambrook, J.E., Fritsch, E.F., Maniatis, T. (1989), Molecular Cloning: A Laboratory Manual, 2nd edition. Plainview, NY: Cold Spring Harbor Laboratory Press

Simutis, R., Havlik, I., Lübbert, A. (1993), Fuzzy aided neural network for real time state estimation and process prediction in a production scale beer fermentation, *J. Biotechnol.*, 27, 203-215

Stephanopoulos, G.N., Konstantinov, K., Saner, U., Yoshida, T. (1993), Fermentation data analysis for diagnosis and control, pp. 355-400 in: *Biotechnology Vol. 3*: Bioprocessing, Stephanopoulos, G.N., ed., VCH, Weinheim

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.

This chapter has been published in *Computer Application in Biotechnology 2004 (CAB9)*:
Jenzsch, M., Simutis, R., Lübbert, A. (2005), Application of Model Predictive to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9), Pons, MN, van Impe, JFM, eds., IFAC/Elsevier*, ISBN 0 08 044251 X

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 μ is also primary influencing the specific product development rate π 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 μ. These estimates are then used in a model predictive controller which keeps the specific biomass growth rate μ 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 μ. This is important for quickly and reliably establishing the relationship between the key quantities μ and the specific product formation rate π and further for control of the protein production if the desired μ-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 CO_2. pH was measured by means of an Ingold pH probe. The dissolved oxygen concentration was monitored with an Ingold pO_2 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 OD_{600} 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 OD_{600} 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 OD_{600} 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:

i) Concentrations with respect to the culture mass $\mathbf{c} = [\mathrm{X\ S}]$

ii) cell-internal concentrations with respect to cell mass $\mathbf{c}_x = [\mathrm{G_x\ P_x}]$.

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

$$\begin{aligned} \frac{d\mathbf{c}}{dt} &= \mathbf{R} + \frac{F}{W}(\mathbf{c}_F - \mathbf{c}) \\ \frac{d\mathbf{c}_x}{dt} &= \mathbf{q} - \mu\,\mathbf{c}_x \end{aligned} \tag{1}$$

$\mathbf{c}_F$ 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:

$$\mathbf{R} = [\mu - \sigma]\, X \tag{2}$$

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

$$\mathbf{q} = [-\gamma\ \pi] \tag{3}$$

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

$$\sigma = \sigma_{max} \frac{S}{K_S + S}\ , \tag{4}$$

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

The corresponding specific growth rate is considered to be linearly dependent on σ:

$$\mu = Y_{XS}\,\sigma \tag{5}$$

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

$$\frac{dW}{dt} = F = F_S + F_1 + F_2 \tag{6}$$

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, CO_2 outflow through the vent line, and evaporation must also be considered (F_1). 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 π is the specific growth rate μ. π is described as

$$\pi = \pi_{max}\, G_x \frac{K_\mu}{K_\mu + \mu} - k_{prot}\, P_x \tag{7}$$

Where it is assumed that there is an inhibition effect with respect to μ 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

$$\gamma = \gamma_{\max} \cdot \frac{G_X}{K_{G_x} + G_X} \cdot \frac{\mu}{K_\mu + \mu} \tag{8}$$

which states that the rate plasmid concentration changes depend on the plasmid concentration G_x and the specific growth rate μ.

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).

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 μ 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 μ

$$\begin{aligned} \frac{dX}{dt} &= \mu \cdot X - \frac{F}{W} \cdot X + v_X \\ \frac{d\mu}{dt} &= 0 + v_\mu \end{aligned} \quad , \tag{9}$$

where v_X and v_μ 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 (o*xygen uptake rate), *CPR* (carbon dioxide production rate) and *BASE* (total base consumption during pH control), referred to as the measurement model, were

$$\begin{aligned} \text{OUR} &= Y_{ox}\, \mu\, X + m_o\, X \\ \text{CPR} &= Y_{cx}\, \mu\, X + m_c\, X \\ \text{BASE} &= Y_{bx}\, (X\, W - X_0\, W_0) \end{aligned} \tag{10}$$

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.

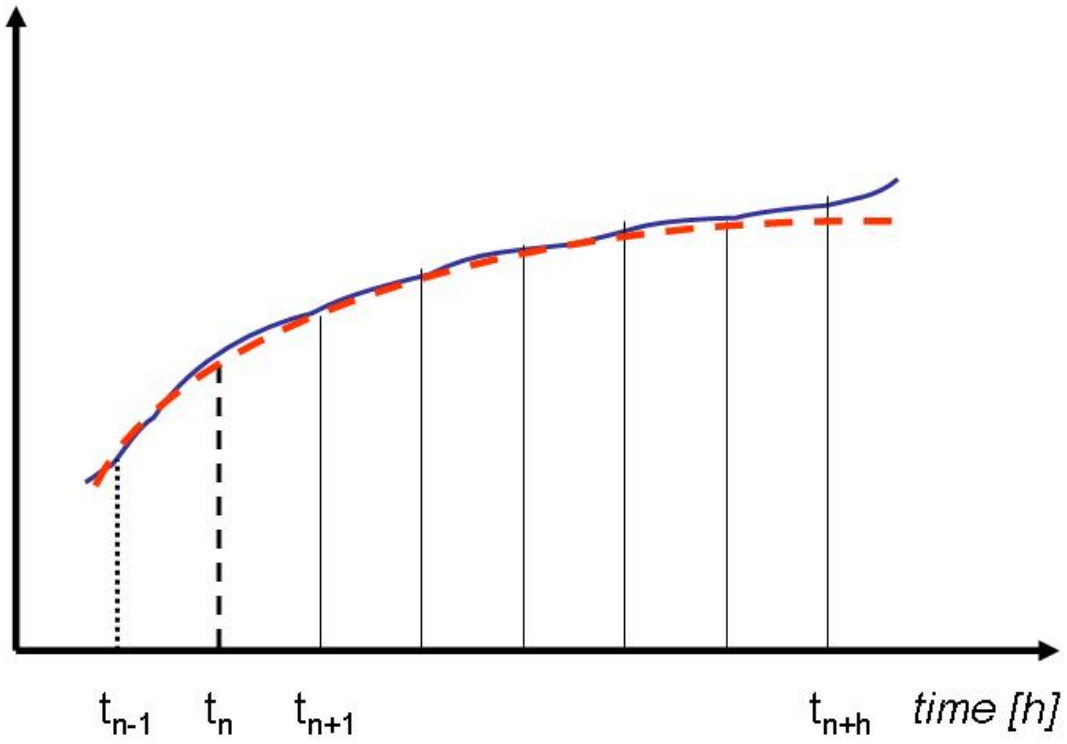

Figure 1. Schematic of the time steps discussed in the text.

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 μ 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.

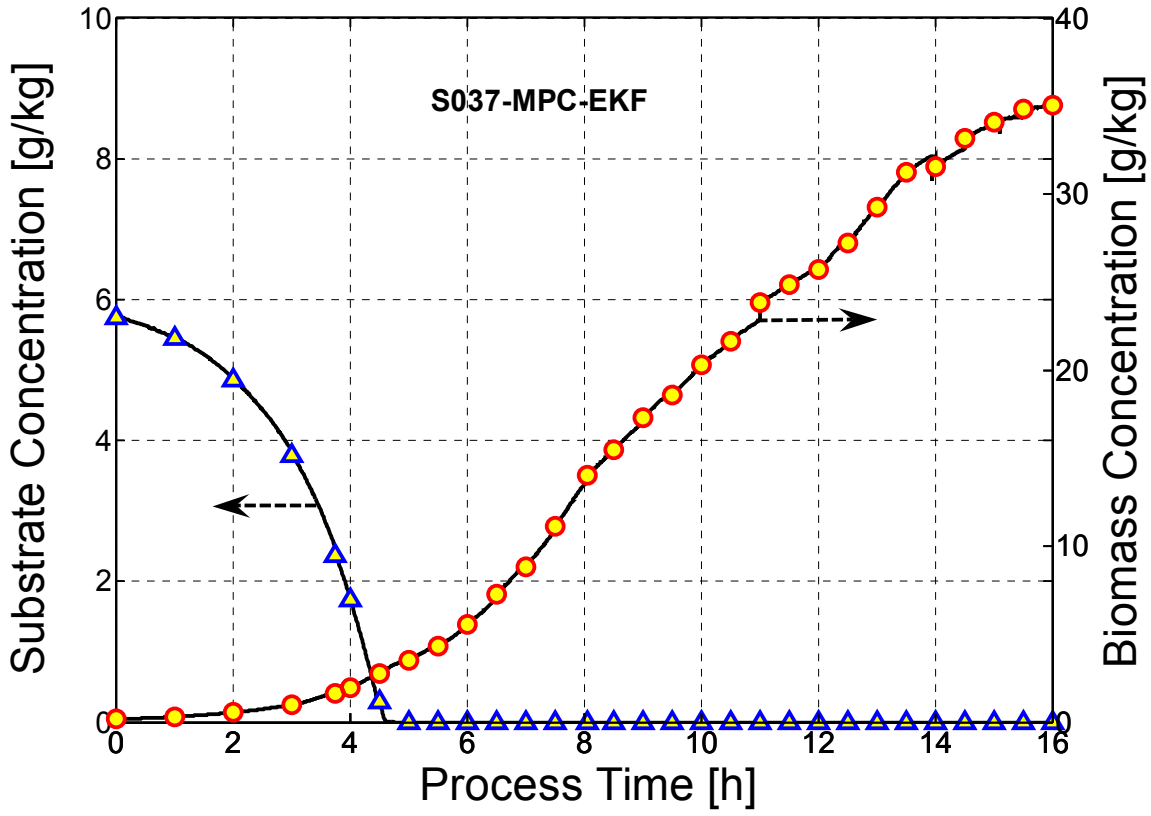

Figure 2. Evolution of the biomass and substrate concentration during controlled batch/fed-batch *E.coli* cultivation. Offline measured data (symbols) and model predictions (lines) are depicted.

When this model is used in a model predictive controller for the specific growth rate μ 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_{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.

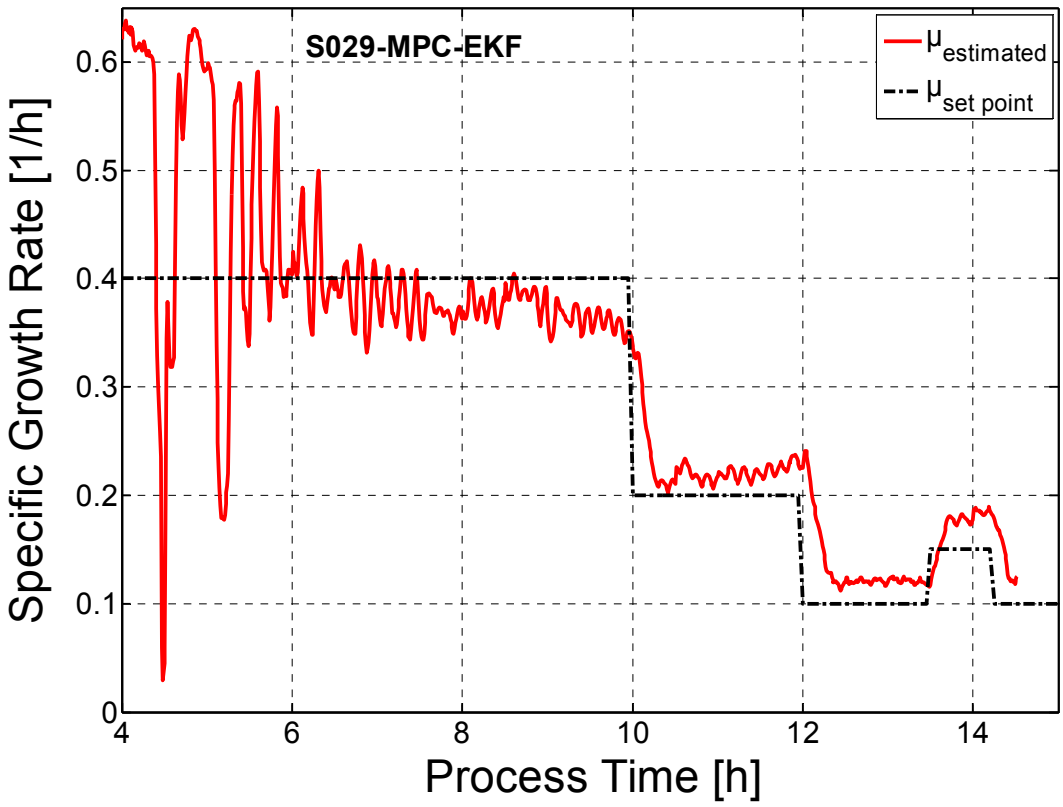

Figure 3. Specific biomass growth profile in the μ-controlled *E.coli* cultivation. In the second part of the cultivation systematic errors between setpoint and real values of specific growth rate can be observed.

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 μ, 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 μ 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_o 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.

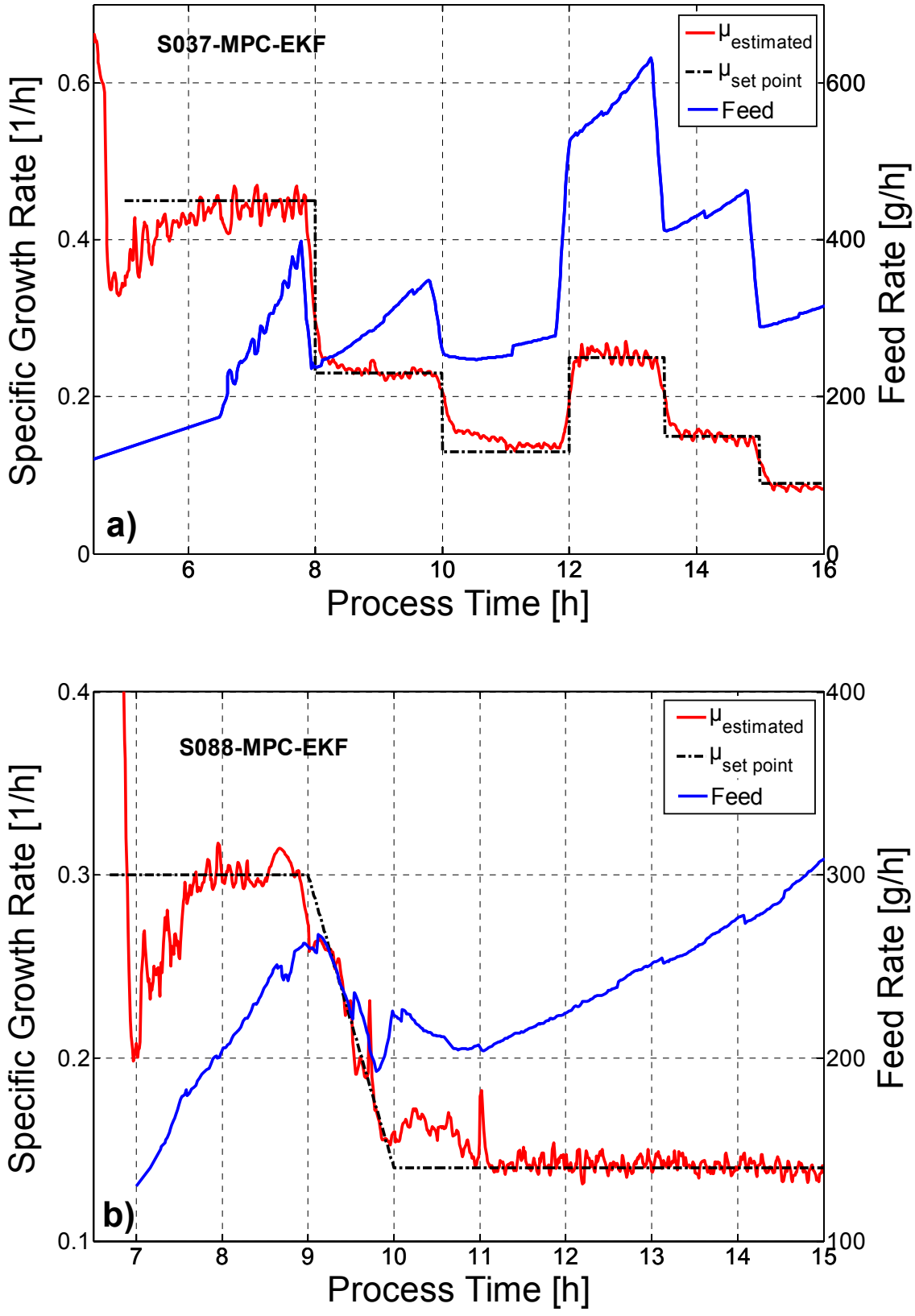

Figure 4. Specific biomass growth profiles in μ-controlled *E.coli* cultivations together with the feeding profiles set by the controller to achieve the μ-profiles [a) experiment with an "exotic" setpoint profile for controller tuning; b) reasonable setpoint profile for a protein production process]. The adaptation of the yield Y_{XS} 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 μ. 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

Akesson, M., Hagander, P., Axelsson, J.P. (2001), Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding, *Biotechnol. Bioeng.*, 73, 223-230

DeLisa, M.P., Chae, H.J., Rao, G., Weigand, W.A., Valdes, J.J., Bentley, W.E. (2001), Generic Model Control of Induced Protein Expression in High Cell Density Cultivation of *Escherichia coli* Using On-line GFP-Fusion Monitoring, *Bioproc. Biosyst. Eng.*, 24, 83-91

Jenzsch, M., Simutis, R., Lübbert, A. (2006), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

Kell, D.B., Sonnleitner, B. (1995), GMP-Good modelling practice: an essential component of good manufacturing practice, *Trends Biotechnol.*, 13, 481-492

Lee, J., Ramirez, W.F. (1994), Optimal fed-batch control of induced foreign protein production by recombinant bacteria, *AIChE J.*, 40(5), 899-907

Levisauskas, D., Simutis, R., Borvitz, D., Lübbert, A. (1996), Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis, *Bioproc. Eng.*, 15, 145 - 150

Lim, H.C., Lee, K.S. (1991), Control of Bioreactor Systems. pp. 510-560 in: *Biotechnology, Vol. 4*: Measuring, Modelling and Control, Schügerl, K., ed., VCH, Weinheim

Morari, M., Lee, J.H. (1999), Model Predictive Control: Past, Present and Future, *Comput. Chem. Eng.*, 23, 667-682

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

Sandoz, D.J., Desforges, M.J., Lennox, B., Goulding, P.R. (2000), Algorithms for industrial model predictive control, *Comput. Control Eng.*, 11(3), p125-134

Sandoz, D.J., Lennox, B., Goulding, P.R., Thorpe, P., Kurth, T., Desforges, M.J., Woolley, I.S. (1999), Innovation in industrial model predictive control, *Comput. Control Eng.*, 10(5), 189-197

Schubert, J., Simutis, R., Dors, M., Havlik, I., Lübbert, A. (1994), Bioprocess optimization and control: Application of hybrid modeling, *J. Biotechnol.*, 35, 51-68

Stephanopoulos, G.N., Konstantinov, K., Saner, U., Yoshida, T. (1993), Fermentation data analysis for diagnosis and control, pp. 355-400 in: *Biotechnology, Vol. 3*: Bioprocessing, Stephanopoulos, G.N., ed., VCH, Weinheim

Walsh, G. (2000), Biopharmaceutical benchmarks, *Nature Biotechnol.*, 18, 831-83

Chapter 6

Improving the Batch-to-Batch Reproducibility in Microbial Cultures During Recombinant Protein Production by Guiding the Process Along a Predefined Total Biomass Profile

Abstract. In industry *Escherichia coli* is the preferred host system for the heterologeous biosynthesis of therapeutic proteins that do not need posttranslational modifications. In this report, the development of a robust high-cell-density fed-batch procedure for the efficient production of a therapeutic hormone is described. The strategy is to guide the process along a predefined profile of the total biomass that was derived from a given specific growth rate profile. This profile might have been built upon experience or derived from numerical process optimization. A surprisingly simple adaptive procedure correcting for deviations from the desired path was developed. In this way the batch-to-batch reproducibility can be drastically improved as compared to the process control strategies typically applied in industry. This does not only apply to the biomass but, as the results clearly show, to the product titer as well.

This chapter has been submitted for publication in *Bioprocess & Biosystems Engineering*:
Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006), Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile, *Bioproc. Biosyst. Eng.*, submitted

1 INTRODUCTION

Biologics are known to be rather complex products. Apparently small changes in the manufacturing processes can cause significant differences in their clinical properties. Hence, production processes for biologics are approved by authorities only together with clearly defined constraints on their manufacturing procedures. Consequently, reproducibility is of utmost importance to these processes. From the engineering point of view there are two challenges in meeting this requirement: First of all, within the given constraints, the operational procedure, most robust with respect to typically appearing process fluctuations, must be found. And, secondly, while running the process along this robust path, the remaining randomly appearing disturbances must be eliminated by means of feedback control.

With respect to performance of process control, production processes for recombinant proteins are lagging far behind most other industrial processes. There are several reasons for this situation. From the technological point of view, the most pressing problem is the accuracy and reliability by which the state of the complex cultivation processes can be determined online. From the practical economical point of view, therapeutic production processes must be developed within a short period and after that the gross process layout is fixed. Since minimizing time-to-market is the overriding objective in this development period, most companies try to avoid any technical complications in the process. Their fear is a costly delay of the approval through discussions with authorities on new technology.

This situation considerably changed with the paradigm change in the FDA towards a new risk-based inspection procedure. An essential part of the new approach is FDA's PAT initiative (FDA 2004). With PAT, the agency concretely stimulates improving pharmaceutical and biologics production with respect to real time automated process monitoring and control. The way proposed is a more rigorous science-based approach to manufacturing since better understanding leads to more efficient process control, lower process variability, thus high product quality and finally patient safety. For biologics it is particularly important to keep the processes under control early in the product synthesis process (e.g., DePalma 2004). All major producers saw the strength of FDA's impact and nominated persons responsible for PAT affairs.

Typically, the profiles of the various state variables in recombinant therapeutic protein production reactors depict a batch-to-batch reproducibility comparable to the situation depicted in Figure 1 for the biomass and target protein concentrations. These data were measured during many fermentation runs under the same cultivation conditions with the biological system serving as the accompanying example in this paper. The situation depicted in Figure 1 is typical for many industrial production processes as well.

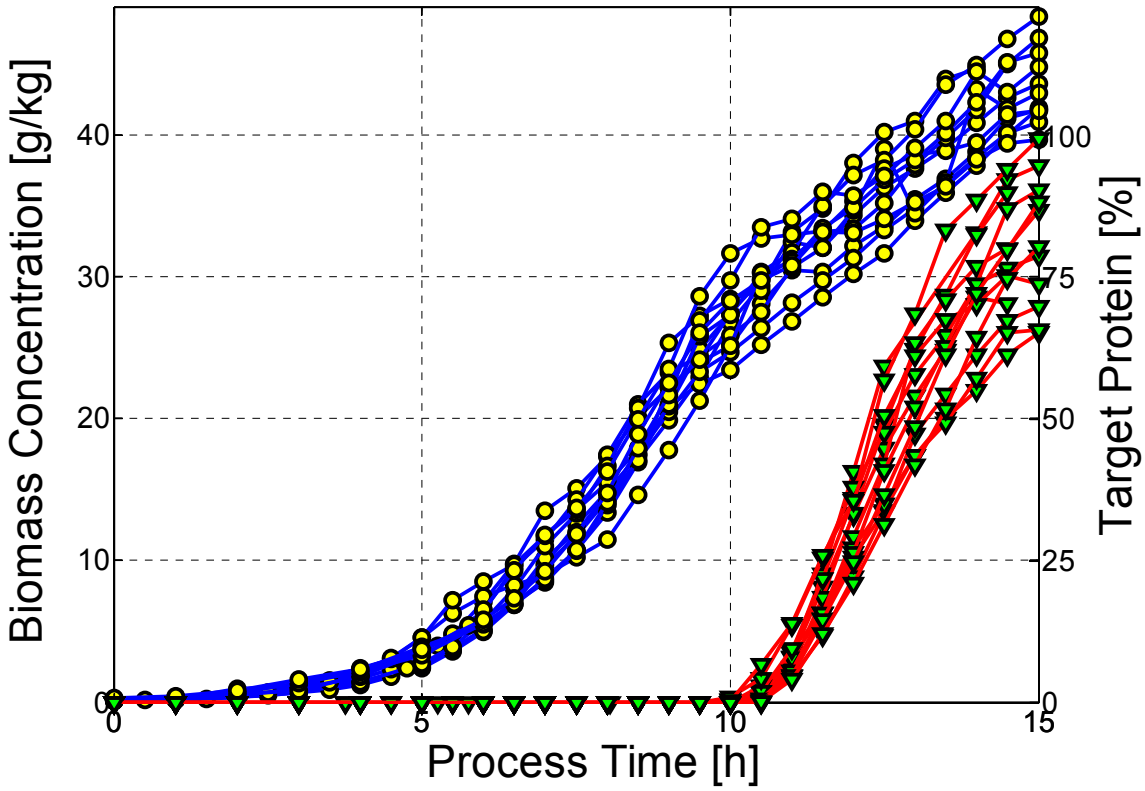

Figure 1. Biomass (○) and target protein (Δ) concentration profiles from 13 fed-batch fermentations for the production of a recombinant protein. Typically the batch-to-batch reproducibility of these production processes is rather low.

In manufacturing of recombinant therapeutic proteins, the primary objective is to produce a given mass m_P of protein within a more or less well defined production time t_P.

$$m_P = \int_0^{t_P} \pi \cdot x \cdot dt \tag{1}$$

where π is the specific product formation rate and x the total biomass. Both arguments of the integral are functions of time, but more importantly, both are primarily dependent on the specific biomass growth rate μ in most industrial production systems. The growth rate that a specific medium supports determines the physiological state of the cells and particularly the cell's protein-synthesizing machinery that is important to recombinant protein production is also under growth rate control (Neidhardt *et al.* 1990). Hence, the variable that rules the final outcome of the process is the specific biomass growth rate μ. Hence, design of effective cultivation processes should be based on an optimal or at least quasi-optimal profile of the specific growth rate. This can be obtained on by means of numerical optimization procedures (e.g. Levisauskas *et al.* 2003) or simply by deriving a profile from data records and experience with the production system under consideration. In order to make sure that the process follows this profile it is straightforward to control μ in the engineering sense (e.g. Pirt 1975, Shioya 1992, Levisauskas *et al.* 1996, Jenzsch *et al.* 2005, and 2006a).

Clearly, m_P should be kept as high as possible in an industrial manufacturing process; however in a production environment, robustness of the process behaviour with respect to occasionally appearing disturbances is highly important as well. Reproducibility is a very important constraint as it affects the downstream processing and thus quality of the final product. Direct control of μ was shown to be possible in recombinant production processes (Jenzsch *et al.* 2005, 2006a). This works perfectly as long as there are no severe disturbances in the process. When, however, some disturbances leads to a significant deviation of the biomass from its desired path, one must correct for it before one can proceed with the desired optimal or quasi–optimal μ-profile. Hence, one must look for more robust alternatives that are able to keep the process more tightly on the desired profiles of μ and other state variables that are important for process quality. From equation (1) it is straightforward to consider biomass x first.

Using start biomass and desired specific biomass growth rate profile it is easy to estimate total biomass profile during the cultivation

$$\frac{dx}{dt} = \mu_{set} \cdot x \tag{2}$$

A given biomass profile $x_{set}(t)$ is then in a close relationship with the specific biomass growth rate μ. Hence, controlling the process to an x profile should satisfy the corresponding specific growth rate profile $\mu_{set}(t)$ as well. In this case the cultivation process is more robust, because the deviations in biomass concentration can be eliminated by controlling the integral variable x.

μ as well as x cannot be measured directly with sensors that work reliably at a production fermenter. Both can be measured indirectly. As are able to estimate x much more reliably than the specific growth rate μ (Jenzsch *et al.* 2006b), x was chosen as the controlled variable.

For indirect x-estimation, there are several possibilities. The first we are considering here is using an extended Kalman filter. The alternative is estimating the biomass concentration by means of artificial neural networks (Jenzsch *et al.* 2006b). In production environments, where sufficiently many data records are available from the process under consideration, the latter method yields more accurate estimates. In the beginning of a new project, however, where the data is sparse, the Kalman filter is a better choice. Here we apply estimates by artificial neural networks (ANNs). The ANN was trained on 26 data sets measured during a process development project with the strain used in this work.

2 MATERIALS AND METHODS

Experiments were performed with *E.coli* BL21(DE3) as the host cell. The recombinant target protein was coded on the plasmid pET 28a and expressed under the control of the T7 promoter after induction with isopropyl-thiogalactopyranosid (IPTG). The strain was resistant against kanamycin. The product appears as inclusion body within the cytoplasm.

The main substrate was glucose. It was fed at a concentration of 600 [g/kg]. The other components are compiled in Table 1.

Table 1. Composition of the mineral medium

Mineral salt solution		***Trace element solution***	
Component	*Concentration [g/kg]*	*Component*	*Concentration [g/kg]*
K_2HPO_4	14.60	Na_2-EDTA	20.10
$NaH_2PO_4 \times H_2O$	3.60	$FeCl_3 \times 6\ H_2O$	16.70
$(NH_4)_2SO_4$	2.46	$CaCl_2 \times 2\ H_2O$	0.74
Na_2SO_4	2.00	$CoCl_2 \times 6\ H_2O$	0.21
$MgSO_4 \times 7\ H_2O$	1.20	$ZnSO_4 \times 7\ H_2O$	0.18
$(NH_4)_2$-H-Citrat	1.00	$CuSO_4 \times 5\ H_2O$	0.10
NH_4Cl	0.50	$MnSO_4 \times H_2O$	0.10
Kanamycin	0.10		
Thiamin	0.10		
Trace element solution	2 [mL/kg]		

All the experiments were performed within BBI Sartorius System's BIOSTAT® C 15-L-bioreactor with 8 [L] working volume. The fermenter was equipped with 3 standard 6-blade Rushton turbines that could be operated at up to 1400 [rpm]. The aeration rate could be increased up to 24 [sLpm]. Aeration rate and then stirrer speed were increased one after the other in order to keep the dissolved oxygen concentration at 25 [%] saturation. In order to suppress foam formation, increase oxygen solubility and reduce the risk of contamination the fermenter head pressure was kept at 0.5 [bar] above the ambient pressure in the laboratory.

The fermentations were operated in the fed-batch mode from the early beginning with an initial volume of 5 [L] at pH 7 and a temperature of 35 [°C]. All fermentations were started in the night by automatic transfer of the inoculation biomass from a refrigerator into the reactor. Substrate feeding starts immediately after inoculation with a fixed exponential profile. According to Jenzsch *et al.* (2006c), the specific growth rate setpoint was chosen to 0.5 [1/h]. After the biomass concentration reached values of about 35 [g/kg], the culture was induced with 1 [mM] IPTG. From the 7th fermentation hour on, the growth rate was kept under feedback control along an x-profile that was derived from a predefined profile of the specific growth rate. The corresponding profile of the substrate feed rate was determined from the profiles of μ and x. This was taken as reference feeding profile F_{ref} for all the fermentations described.

Temperature was measured with a Pt-100, pH with an Ingold-pH-probe, pO_2 with an Ingold pO_2-Clark-electrode, CO_2 in the vent line with MAIHAK®'s Unor 610, and O_2 there with MAIHAK®'s Oxor 610. Further, the total ammonia consumption during pH control was recorded by means of a balance beneath the base reservoir. All these quantities were measured online. Additionally, enhanced foam levels could be detected with a foam sensor and, if the critical level was reached, a silicone antifoaming emulsion (ROTH®) was added.

Biomass concentrations were measured offline (via the optical density at 600 [nm]) with a Shimadzu® photo-spectrometer (UV-2102PC). In preceding experiments a correlation was established between these values and the biomass dry weights which were determined with the standard drying and weighting technique. Glucose was determined enzymatically with a YSI 2700 Select Bioanalyzer. The product was measured with SDS PAGE after separation of the inclusion bodies and their solubilisation.

3 RESULTS

Preliminary simulations and experiments showed that the total biomass x(t) is better suited as the controlled variable than the biomass concentration X(t) itself. The control can then be performed with a simple adaptive control algorithm comprising the following steps:

i) Estimation of the total biomass x_{est} at the current time instant t using the artificial neural network trained before on the available data records (Jenzsch *et al.* 2006b). A simple feedforward artificial neural network was used which makes use of the online measured signals OUR, CPR, total base consumption resulting from pH control.

ii) Computing the deviation of estimated biomass x_{est} from its setpoint x_{set}

$$\Delta x = x_{set} - x_{est} \tag{3}$$

iii) Utilization of Δx to determine the controller variable α by which the controller is adapted to the current state of the process using the following proportional/integral approach:

$$\alpha = k_1 \cdot \Delta x + k_2 \cdot \int_{t_s}^{t} \Delta x \cdot dt \;; \qquad -0.15 \le \alpha \le 0.15 \tag{4}$$

The parameters were initially determined in simulation studies. Later, during preliminary control experiments, they were slightly adapted. There values are $k_1 = 0.1$ [1/kg(S)] and $k_2 = 0.02$ [1/kg(S)/h].

iv) Calculation of actual feed rate F using the corrected yield expression:

$$F = \frac{\mu_{set} \cdot x_{est}}{\left(Y_{XS} - \alpha\right) \cdot S_f} \;; \qquad 0.7 \cdot F_{ref} \le F \le 1.3 \cdot F_{ref} \tag{5}$$

The controlled biomass profiles depicted in Figure 2 show that this simple control approach leads to a very good reproducibility of the total biomass profiles.

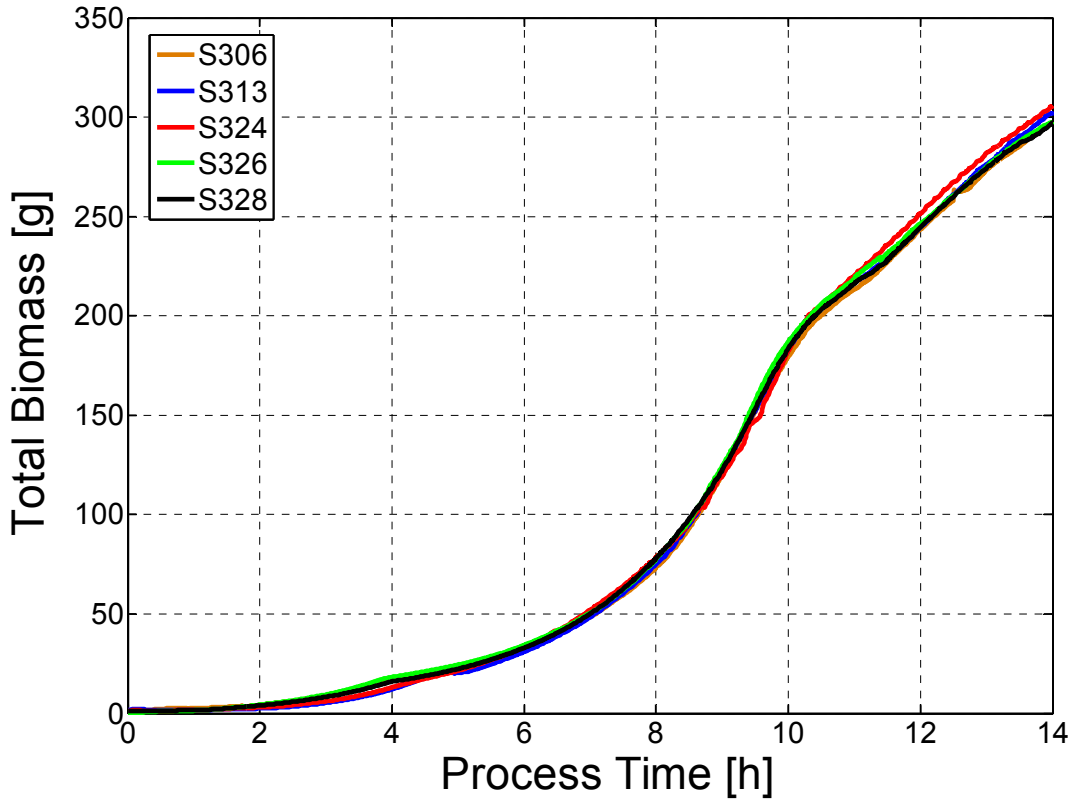

Figure 2. Total biomass signals from 5 fermentations performed sequentially using the same setpoint profile.

As one is usually not interested in the total biomass, the corresponding biomass concentrations are depicted in Figure 3. As can easily be seen from both plots, the batch-to-batch variability of the trajectories is rather small, i.e. the process total biomass control works well.

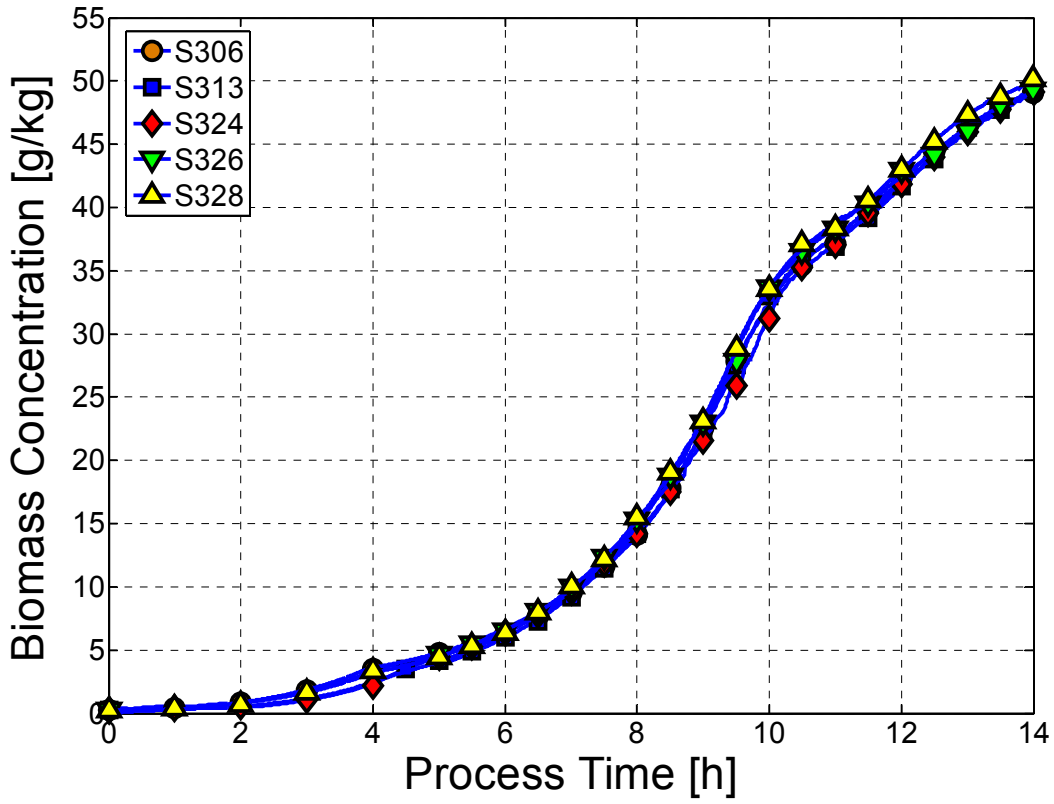

Figure 3. Measurements of the biomass concentration during five experiments in which the total biomass was controlled. The induction time was t_{ind}=11 [h] in all cases.

In order to get a better impression of the controller action, the deviations from the desired profile, i.e., the relative deviations between the total biomass and its setpoint is plotted in Figure 4. In all the experiments, the controller was switched on at t=7 [h]. In the first 4 hours thereafter, the relative deviations remain within a 5 [%] interval. The controller action then improves so that finally the relative deviation of total biomass remains within a 2 [%] interval around the mean.

There is one exception: In the last experiment (S330) feed pump was switched off from 3 to 5 hours in order to test controller performance under process conditions with an extremely hard disturbance. The controller appeared to be robust enough as to cope with this disturbance. Again in the end the relative deviation from the mean remained in the 2 [%] interval.

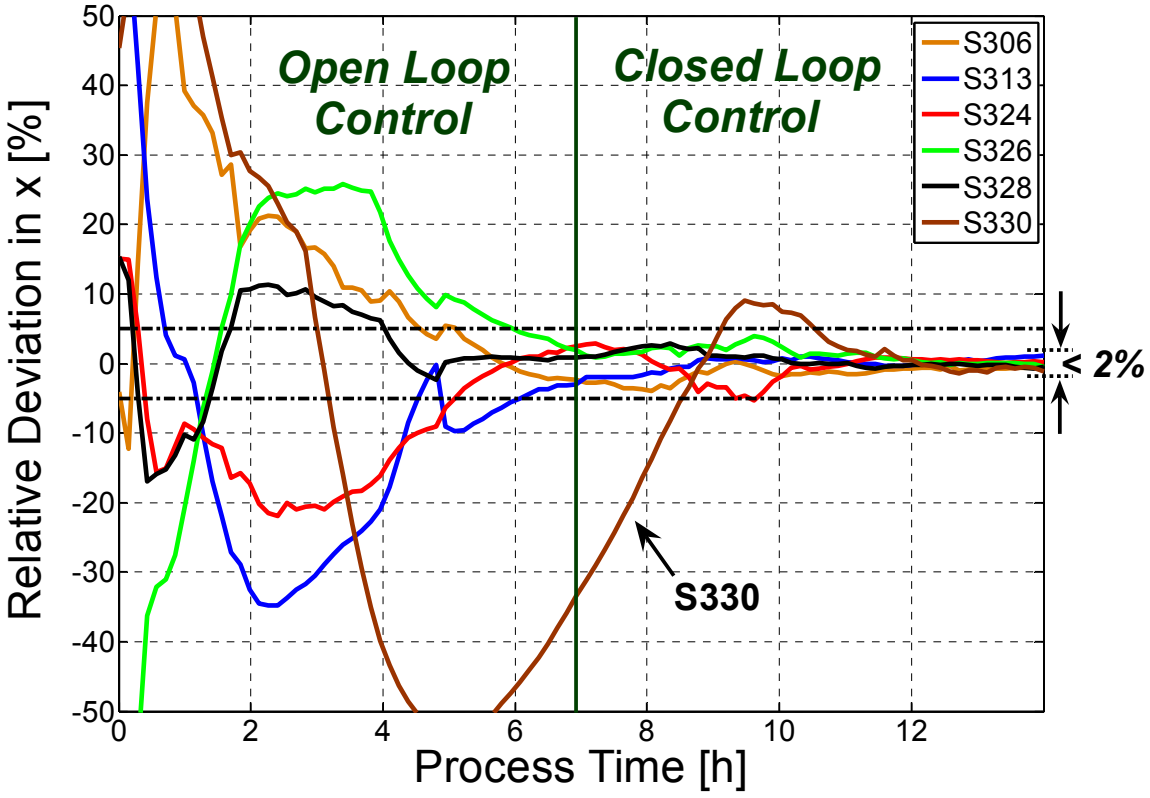

Figure 4. Relative deviations of the total biomass from the mean. The controller was switched on 7 [h] after the cultivation was inoculated.

The next question is what does this mean for the profiles of the specific biomass growth rate, which was initially chosen to determine the setpoint profiles for the total biomass. This comparison essentially should demonstrate the data consistency. The specific growth rate μ was determined from a Luedeking-Piret-type relationship between the biomass concentration, carbon dioxide production or oxygen uptake rate and the specific growth rate:

$$OUR = Y_{OX} \cdot \mu \cdot X + m_O \cdot X \tag{6}$$

$$CPR = Y_{CX} \cdot \mu \cdot X + m_C \cdot X \tag{7}$$

Both equations can be resolved for μ. The parameters were already known from a fit of the models to the fermentation data.

Hence, with the biomass concentration profiles estimated and the measured OUR(t) and CPR(t) signals, μ can easily be computed. The result is depicted in Figure 5. In the decisive phase of the process, the product formation phase, the resulting trajectories of the specific biomass formation rate μ(t) are quite well resembling the original setpoint profiles. It should be recalled that rather high noise is to be expected when computing μ(t). However, the results depicted in Figure 5 nevertheless clearly show that the data are consistent.

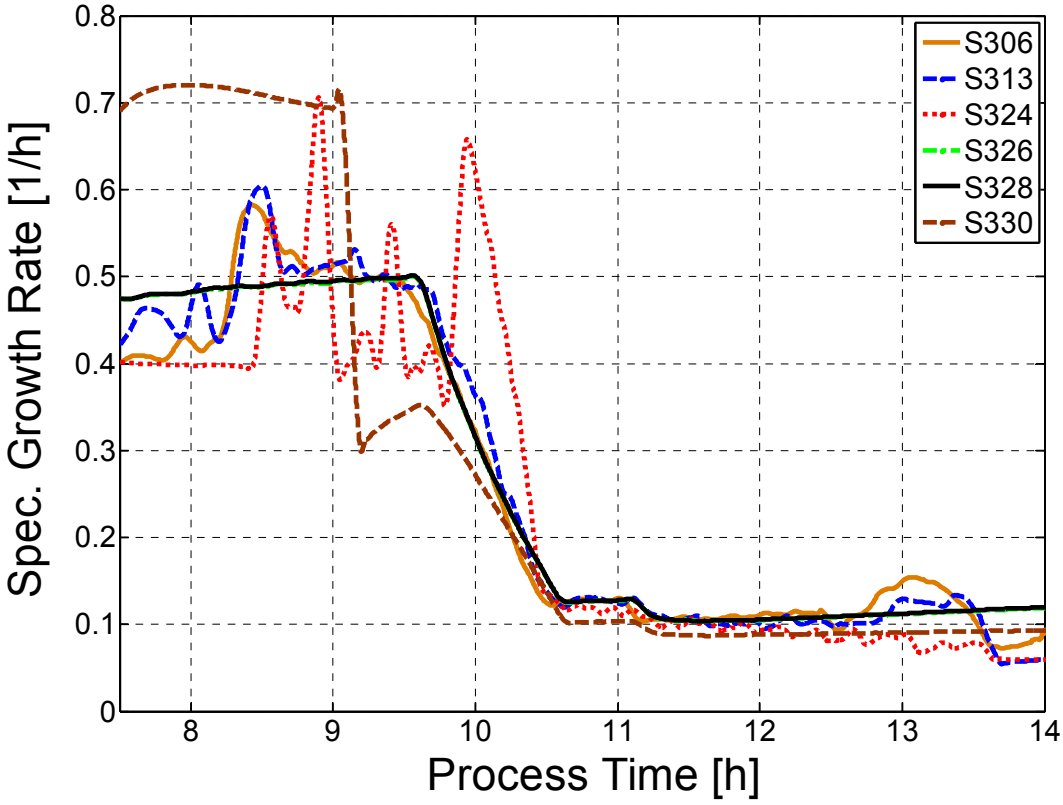

Figure 5. Specific biomass growth rate profiles determined from the measurement profiles of OUR, CPR as well as from the estimates for biomass concentration X(t). Note that study 330 was the test case where the substrate supply was interrupted for some period during the biomass growth phase.

Finally the question arises what does a high batch-to-batch reproducibility in the biomass profiles means for the variance in the product formation profiles. Results corresponding to the data shown before are depicted in Figure 6.

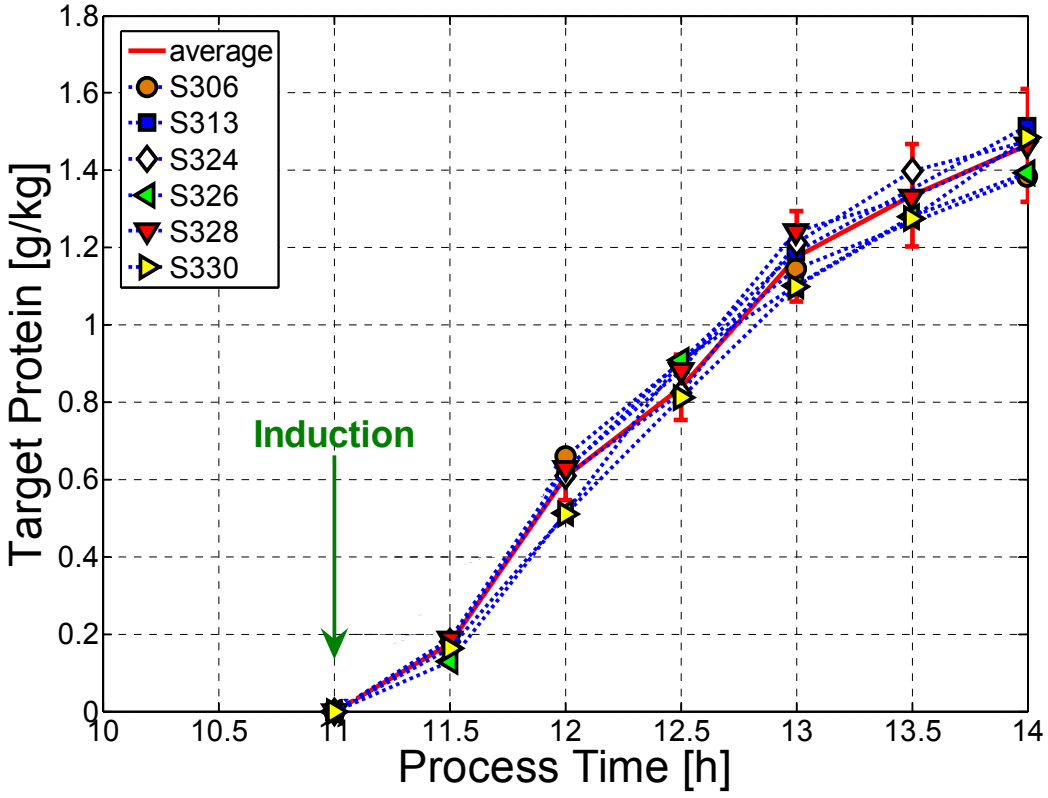

Figure 6. Product concentration profiles for the cultivation processes already mentioned. All protein data stay within the error bar ranges representing the confidential interval of the protein analysis method.

The corresponding protein concentration profiles are rather close together saying that the improved reproducibility in biomass profiles by means of x control also lead to an improved batch-to-batch reproducibility in the product concentration profiles. All profiles stay within the confidence interval of the analytical protein detection method. The relative error of ±15 [%] of target protein analysis by SDS-PAGE was estimated from a fivefold analysis using the same fermentation samples.

4 DISCUSSION

The development of industrial microbial cultivation processes for recombinant protein production is right now forced to proceed in the direction of achieving higher process and product quality levels. As recombinant therapeutic proteins are very complex molecules, actions performed to assure product quality are practically restricted to improvements of their manufacturing processes. Particular the process variability in the fermentation part is in the focus of these activities. Fermentation processes are thus approved with to tight constraints in their process operational procedures. Single violations may force the manufacturer to discard the entire culture. Hence, the companies are interested in a high reproducibility of their fermentation processes.

From the cell-physiological point of view the specific growth rate would be the most important variable to be controlled in fed-batch fermentation. Thus much work was put into control procedures to put μ-control into practice (Shioya 1992, Levisauskas *et al.* 1996, Jenzsch *et al.* 2005, 2006a). However, as shown in the beginning of this paper, this is not optimal from the process reproducibility point of view. If batch-to-batch reproducibility is an issue, and this is what FDA is demanding from drug manufactures, control of biomass is clearly superior. In industrial practice this does not mean that the primary idea of control guiding the process along favourable profiles of the specific biomass growth rate must be abandoned. Instead, from the μ-profile a corresponding x-profile should be determined and this can then be controlled in a feedback fashion.

Many control strategies have been discussed in literature but there is practically no industrial plant where control of variables other than pH, T, and may be pO_2 has been realized. Most of the control approaches published so far are rather complex and there is some fear in industry to apply them. From that point of view, the control procedure discussed here is an extremely simple adaptive approach. It was tested in more than a hundred fermentation processes where recombinant proteins were generated with E.coli and Pichia pastoris cell. It proved to be very robust and practicable even in large-scale bioreactors.

The developments discussed here are in the focus of FDA's "PAT Initiative" which aims in supporting innovative developments in pharmaceutical production, particularly with respect to knowledge-based real time process monitoring and control. With the process design and control techniques reported in this paper, we directly follow the simple philosophy behind the PAT initiative: Model-based design methods lead to more robust process control profiles; and together with the simple but efficient process control to a lower process variability, and, consequently, to a high product quality.

Finally it should be noted that batch-to-batch reproducibility of cultivation processes as achieved in this work is not only important to quality in the upstream process. It is important to downstream processes as well. The more reproducible product concentration at the end of the fermentation the better the reproducibility in protein separation and purification, and thus, the better is the quality of the desired product.

Acknowledgements. This work has been supported financially by the Ministry of Cultural Affairs of the state "Sachsen-Anhalt", Germany. We gratefully acknowledge this support.

REFERENCES

DePalma, A. (2004), PAT: Taking Process Monitoring to Next Level, *Gen. Eng. News*, 24(9), 46-47

FDA (2004), Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance, http://www.fda.gov/cvm/guidance/published.html

Jenzsch, M., Simutis, R., Lübbert, A. (2006a), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A. (2006b), Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioproc. Biosyst. Eng.*, 29(1), 19-27

Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006c), Open loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J. Biotechnol.*, in press

Jenzsch, M., Simutis, R., Lübbert, A. (2005), Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9),* Pons MN, van Impe JFM, eds., IFAC/Elsevier, ISBN 0 08 044251 X

Levisauskas, D., Galvanauskas, V., Henrich, S., Wilhelm, K., Volk, N., Lübbert, A. (2003), Model-Based Optimization of Viral Capsid Protein Production in Fed-Batch Culture of recombinant Escherichia coli, *Bioproc. Biosyst. Eng.*, 25, 255-262

Levisauskas, D., Simutis, R., Borvitz, D., Lübbert, A. (1996), Automatic control of the specific growth rate in fed-batch cultivations processes based on exhaust gas analysis, *Bioproc. Eng.*, 145-150

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

Pirt, S.J. (1975), *Principles of Microbe and cell cultivation*, Blackwell Scientific Publication, London, UK, 115-117

Shioya, S. (1992), Optimization and control in fed-batch bioreactors, *Adv. Biochem. Eng. Biotechnol.*, 46, 111-142

Chapter 7

Improving the Batch-to-Batch Reproducibility of Microbial Cultures During Recombinant Protein Production by Regulation of the Total Carbon Dioxide Production

Abstract. Batch-to-batch reproducibility of fermentation processes performed during the manufacturing processes of biologics can be increased by operating the cultures at feed rate profiles that are robust against typically arising disturbances. Remaining randomly appearing deviations from the desired path should be suppressed by automatically manipulating the feed rate. With respect to the cells' physiology it is best guiding the cultivations along an optimal profile of the specific biomass growth rate μ(t). However, there are two problems that speak for investigating alternatives: Upon severe disturbances that may happen during the fermentation, the biomass concentration X may significantly deviate from its desired value, then μ-control leads to a diminished batch-to-batch reproducibility. Secondly, the specific growth rate cannot easily be estimated online to a favourably high accuracy. The alternative discussed here solves both problems by keeping the process at the corresponding total cumulative carbon dioxide production-profile: It is robust against distortions in X and the controlled variable can accurately be measured directly and online during cultivations of all relevant sizes. As compared to the fermentation practice currently used in industry, the experimental results, presented at the example of a recombinant protein production with *E.coli* cells, show that CPR-based corrections lead to a considerably improved batch-to-batch reproducibility.

This chapter has been submitted for publication in *Journal of Biotechnology*:
Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006), Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of total carbon dioxide production, *J. Biotechnol.*, submitted

1 INTRODUCTION

With its PAT initiative (FDA 2004), FDA focused attention towards process supervision and regulation in manufacturing of biologics. Biologics are complex products where product quality is indispensably coupled with the quality of the manufacturing process. Apparently small changes in the operation of the processes can cause significant differences in the product's clinical efficacies. Hence, the production of a biologic is approved by authorities only together with clearly defined constraints on its manufacturing process. Consequently, the batch-to-batch reproducibility is of primary interest in manufacturing therapeutic proteins. From the engineering point of view there are two challenges in meeting this requirement: First of all, within the given constraints, the operational procedure most robust with respect to the possible process disturbances must be found. And, secondly, while running the process along this desired path, the remaining randomly appearing deviations must be eliminated by automatic corrections at the adjustable variables, for instance, the substrate feed rate.

With respect to batch-to-batch reproducibility, production processes for recombinant proteins are lagging behind most other industrial processes. From the technological point of view, the most important reason is the accuracy and reliably by which the actual states of the complex cultivation processes can be determined online.

From the economical point of view, therapeutic production processes must be developed within short periods and afterwards the gross process layouts are fixed. Since minimizing time-to-market is the overriding objective in this period, the companies avoid any technical complications in the process. Their main fear is a costly delay of the approval through discussions on new technology with the authorities. This situation considerably changed with the paradigm change in the FDA towards a new risk-based inspection procedure. An essential part of the new approach is FDA's PAT initiative (FDA 2004).

With PAT, the agency concretely stimulates improving pharmaceutical and biologics manufacturing processes with respect to real time automated process monitoring and control. The way proposed is a more rigorous science-based approach to manufacturing since it is generally understood that better understanding leads to more efficient process control, lower process variability, thus higher product quality and finally patient safety. For biologics it is particularly important to keep the processes under control early in the product synthesis process (e.g., DePalma 2004). This message was understood in industry and all major producers nominated persons responsible for PAT affairs.

The first question to be solved is about the strategy of manipulating the process in order to improving batch-to-batch reproducibility. In manufacturing recombinant therapeutic proteins, the primary objective is to produce a given mass m_P of protein within a more or less well defined production time t_P. From the product mass balance around the culture we obtain

$$m_P = \int_0^{t_P} \pi \cdot x \cdot dt \tag{1}$$

where π is the specific product formation rate and x the total biomass. Both arguments of the integral are functions of time, but more importantly, both are primarily dependent on the specific biomass growth rate μ in most industrial production systems. The growth rate that a

specific medium supports determines the physiological state of the cells, and particularly, the cell's protein-synthesizing machinery that is important to recombinant protein production is also under growth rate control (Neidhardt *et al.* 1990). Hence, the variable that rules the final outcome of the process is the specific biomass growth rate μ. Consequently, this variable should be kept under tight control (e.g. Pirt 1975, Shioya 1992, Levisauskas *et al.* 1996, Jenzsch *et al.* 2005, and 2006a).

The number of production reactors in biotechnology where process control by manipulating the feed rate profile is realized is negligible. It seems to be a relatively big step to install process control in this industry. Many of the control procedures that were proposed in literature are rather complex do not really downsize the activation barrier. Thus, simple, easy-to-use regulation procedures are required. The main problem, however, remains that π, μ and even x-based control suffer from the fact that they cannot accurately enough be measured directly with sensors that can be installed at a production fermenter.

In this paper we address the problem of designing a robust strategy of increasing the batch-to-batch reproducibility. From an underlying profile of the specific biomass growth rate we develop the profile of the total cumulative carbon dioxide production rate and show that guiding the process along this trajectory leads to a robust process behavior. We show experimentally at the example of a recombinant protein expressed by an *E.coli* strain that this procedure works perfectly. Further we show that the batch-to-batch reproducibility can be improved even more by automatically correcting random deviations of the process from this path.

2 EXPERIMENTAL

Strain and cultivation technique

All experiments were performed with *E.coli* BL21(DE3) as the host cell. The recombinant target protein was coded on the plasmid pET 28a and expressed under the control of the T7 promoter after induction with isopropyl-thiogalactopyranosid (IPTG). The strain was resistant against kanamycin. The product appears as inclusion body within the cytoplasm.

The cultivations were performed in mineral salt medium (Jenzsch *et al.* 2006d) with glucose as main substrate. It was fed at a concentration of 600 [g/kg].

All the experiments were performed within BBI Sartorius System's BIOSTAT® C 15-L-bioreactor with 8 [L] working volume. The fermenter was equipped with 3 standard 6-blade Rushton turbines that could be operated at up to 1400 [rpm]. The aeration rate could be increased up to 24 [SLpm]. Aeration rate and then stirrer speed were increased one after the other in order to keep the dissolved oxygen concentration at 25 [%] of its saturation value. In order to suppress foam formation, increase oxygen solubility and reduce the risk of contamination the fermenter head pressure was kept at 0.5 [bar] above the ambient pressure in the laboratory.

The fermentations were operated in the fed-batch mode from the early beginning with an initial volume of 5 [L] at pH 7 and a temperature of 35 [°C]. All fermentations were started in the night by automatic transfer of the inoculum from a refrigerator into the reactor. Substrate feeding starts immediately after inoculation with a fixed exponential profile. According to Jenzsch *et al.* (2006c), the specific growth rate setpoint was chosen to 0.5 [1/h]. After the biomass concentration reached values of about 35 [g/kg], the culture was induced with

1 [mM] IPTG. From the 7th fermentation hour on, the growth rate was kept under feedback control along a tcCPR-profile that was derived from a predefined profiles of the specific growth rate and the biomass x. The corresponding profile of the substrate feed rate was determined from the profiles of μ and x. This was taken as reference feeding profile F_{ref} for all the fermentations described.

Temperature was measured with a Pt-100, pH with an Ingold-pH-probe, pO_2 with an Ingold pO_2-Clark-electrode, CO_2 in the vent line with MAIHAK®'s Unor 610, and O_2 there with MAIHAK®'s Oxor 610. Further, the total ammonia consumption during pH control was recorded by means of a balance beneath the base reservoir. All these quantities were measured online. Additionally, enhanced foam levels could be detected with a foam sensor and, if the critical level was reached, a silicone antifoaming emulsion (ROTH®) was added.

Biomass concentrations were measured offline (via the optical density at 600 [nm]) with a Shimadzu® photo-spectrometer (UV-2102PC). In preceding experiments, a correlation was established between these values and the biomass dry weights which were determined with the standard drying and weighting technique. Glucose was determined enzymatically with a YSI 2700 Select Bioanalyzer. The product was measured with SDS PAGE after separation of the inclusion bodies and their solubilisation.

Following total CPR-trajectories

Apart from the difficulty of making available online values of μ, in a μ-controlled fermentation a further problem is the instability of that approach with respect to variations in the initial total biomass amount x_0. As depicted in Figure 1, an initial deviation in x will be amplified when a fixed μ profile will be applied. Hence, merely keeping the process at a fixed μ(t) leads to an unfavourable batch-to-batch reproducibility.

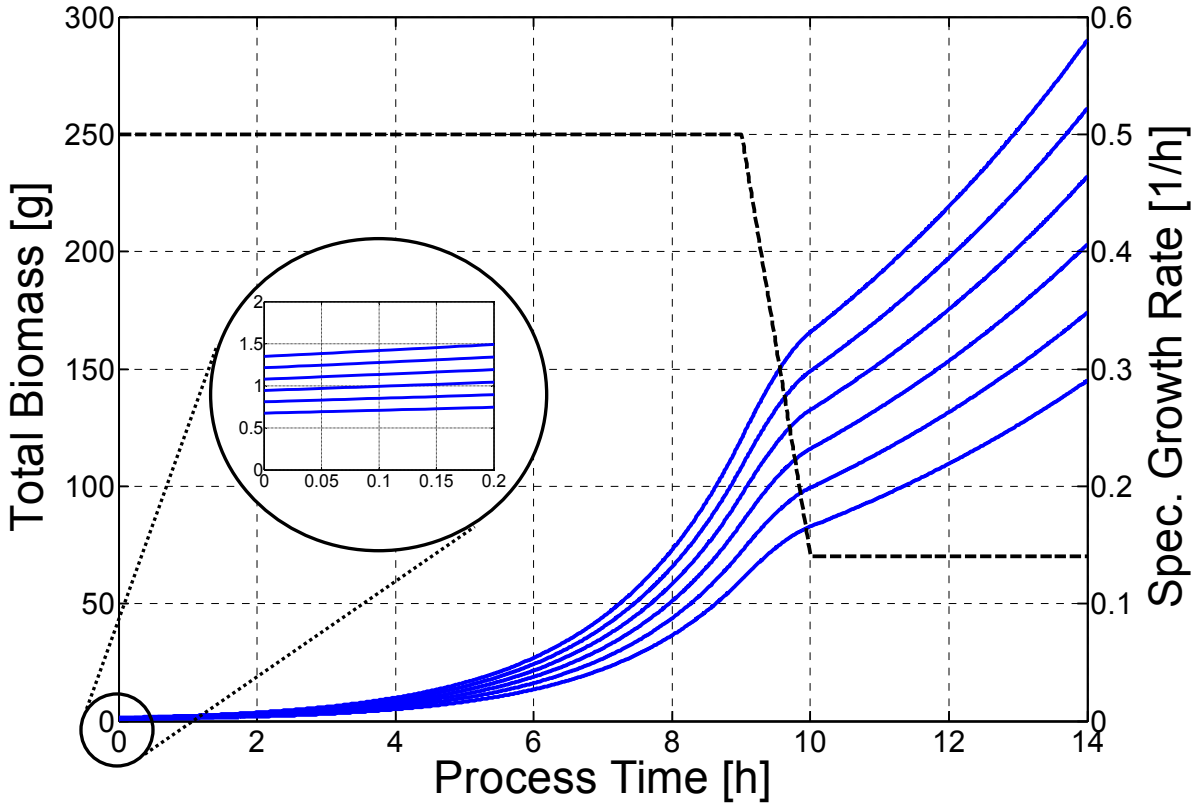

Figure 1. Simple simulation of fed batch processes running with a fixed growth rate profile (dashed line) showing the effect of an initial variance in the total biomass concentration x (full lines). The variance becomes amplified with time.

This problem can be solved by relating the μ profile to a CPR(t)-profile and running CPR along the resulting trajectory by means of feedforward/feedback control. This variable that is tightly related to the specific growth rate and can easily be measured online. Instead of the carbon dioxide production rate CPR, one can also use the corresponding total CPR signal tCPR=CPR·W, where W(t) is the corresponding culture mass signal, or even better, the accumulative signal tcCPR of the total carbon dioxide production rate.

Usually a simple relationship in terms of a Luedeking-Piret approach (Luedeking and Piret 1959a, b) is used to relate both quantities μ and CPR. Here we show that a slightly extended analogue of a Luedeking-Piret-type relationship can be used to relate total CPR to the specific biomass growth rate μ and total biomass x:

$$tCPR = \left(Y_{CX} \cdot \mu + m_C\right) \cdot x \cdot K_{ind} \qquad (2)$$

where the scalar factor K_{ind} is 1.0 during the biomass formation phase and K_{ind} >1 during the protein formation rate. This extension is reflecting the experimentally observed fact that more CO_2 is formed per volume, time and biomass during the product formation phase.

Equation (2) we can be rewritten in the following way

$$\mu = \frac{tCPR}{x \cdot Y_{CX} \cdot K_{ind}} - \frac{m_C}{Y_{CX}} \tag{3}$$

Using the balance equation for the total biomass

$$\frac{dx}{dt} = \mu \cdot x \tag{4}$$

the parameters of the equation (2) can simply be identified by means of a numerical fit using the experimental data for x and tCPR measured in a couple of experiments. The fit can be performed with any optimization routine. In Figure 2 an example is shown where the function x(t,tCPR) was fitted to a typical experimental data record using Matlab®'s routine *lsqcurvefit* (The MathWorks, Inc.).

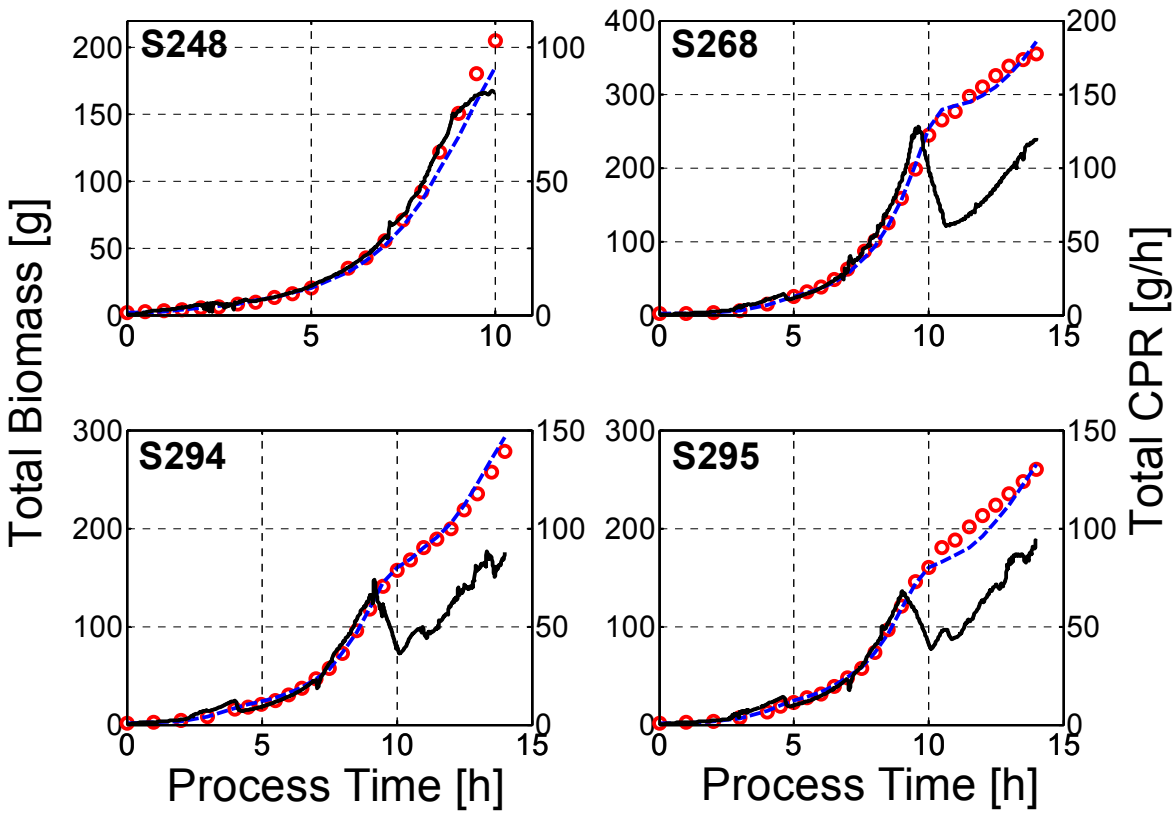

Figure 2. Fit of equation (2) to experimental data obtained in 4 experiments in which a recombinant protein was produced in *E.coli.* Full lines are the online measured total carbon dioxide production rates used to estimate (dashed lines) the total biomass (○).

This simple linear model results in a fairly good fit with a root mean square error in the total biomass of RMSE =7.5 [g] corresponding to about 3 [%] of the final value. The parameters of equation (3) obtained in the fit are:

Y_{CX} = 0.712 [g/g];
m_C = 0.136 [1/h];
K_{ind} = 1.384 [-] for $t>t_{ind}$.

An appropriate algorithm for guiding the fermentation along a total CPR profile is:

(i) Decide on an optimal μ profile for the cultivation (using experts' knowledge or a numerical optimization based on a mathematical model of the process)

(ii) Choose an appropriate inoculation size (initial amount of total biomass x)

(iii) Estimate the total CPR profile using equations (2) and (4)

(iv) Integrate the total CPR profile to get the cumulative total CPR setpoint trajectory

(v) Run the cultivation along the resulting tcCPR profile by means of feedforward/feedback control, manipulating the substrate feed.

The proposed method leads to a self-tuning, i.e. a robust process behavior. This is shown in Figure 3. When the initial amount of total biomass is not on target, the specific growth rate will be adapted automatically. When the inoculated biomass is too high, the cells will produce more CO_2 than expected; hence the controller will reduce the feed rate and thus substrate concentration. Consequently, the biomass growth rate will become smaller. If on the other hand the initial amount of total biomass is smaller than expected, the cells will produce less CO_2 and the controller will increase the feeding rate. This will lead to an increase in the biomass growth rate. After a few hours, the specific growth rate is on target simply by controlling the total cumulative CPR to its corresponding profile.

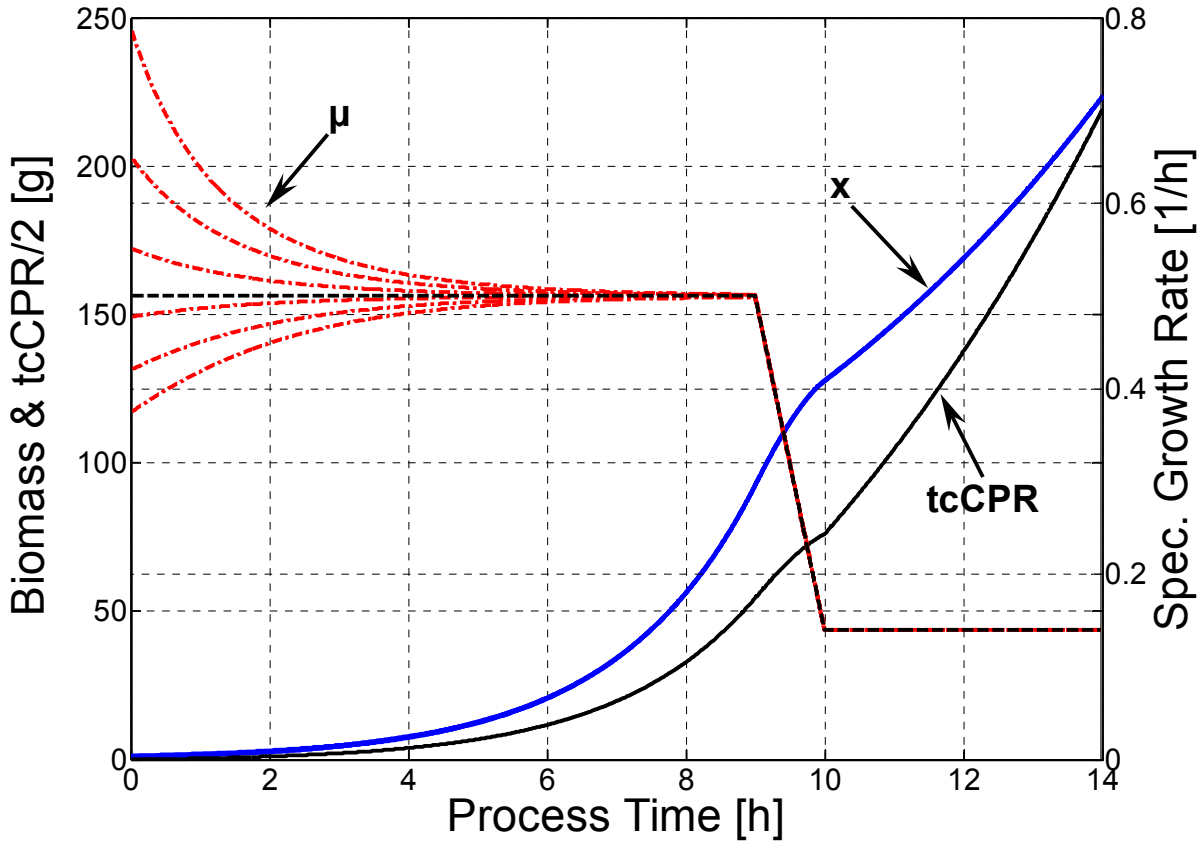

Figure 3. Simple simulation showing the self-stabilizing effect of controlling the total cumulative CPR in a fed-batch cultivation of a recombinant therapeutic protein production process. The simulated seed biomass varies between 0.65 and 1.35 [g].

As shown in the coarse simulation study depicted in Figure 3, the adaptation appears in the first few hours. The essential effect of the adaptation, the reduction in the variance of the biomass trajectories, appears during the protein production phase, i.e. in the second half of the entire process time. Now the essential problem is how to formulate the relationship between the specific growth rate μ, total biomass x and the total carbon dioxide production rate tcCPR.

Nonlinear approach based on artificial neural networks

As the metabolic details relating biomass growth to carbon dioxide production and vice versa are quite complicated and are changing upon induction of the product formation, a straightforward approach is to employ an artificial neural network to describe the relationship.

What we need is a network that connects the carbon dioxide production rate CPR with the specific biomass growth rate μ. In other words we need a data driven model that is able to determine μ from the measurement data.

CPR can be measured online. However, to substantiate the measurement information it is desirable to make use of further measurements that are related to the growth rate as input data for such a network. A straightforward choice is the biomass x from which it was already shown that it can be estimated online quite reliably (Jenzsch et al. 2006b). As a third input variable, t_{ind}, the induction time can be used. The reason for taking t_{ind} as a further input is that the carbon dioxide production is larger in the product formation phase as already mentioned in the discussion of the linear model. That's the inputs of the artificial neural network. For network training, however, one also needs the corresponding values of the specific growth rate μ at network output. As μ cannot be measured directly, it is necessary to develop methods for ANN training using available data, for instance the biomass measurement data.

The idea is to train the network indirectly using the residuals between the measured biomass values x_{data} and those computed with the μ-network via the following equation defining μ:

$$\mu = \frac{1}{x}\frac{dx}{dt} \tag{5}$$

which is used in the form of equation (4)

We choose a simple feedforward artificial neural network (ANN) with a single hidden layer that describes μ as a function of online input data and network weights $\boldsymbol{w}$.

$$\mu_{NET} = ANN(Data, w) \tag{6}$$

As mentioned above, *Data* is composed of the total CPR data, the total biomass x and the induction time t_{ind}.

$$Data = \left[tCPR,\ x_{on},\ t_{ind}\right]^{T} \tag{7}$$

tCPR can directly be derived from the online-measured CPR signal, and t_{ind} is an adjustable variable. x_{on}, could be taken from indirect measurements routines. Here, however, it is easier to estimate $x_{on}(t)$ online directly by integrating the differential equation (4) as will be shown later.

In order to train the network, we minimize the root-mean-square of the residuals

$$res = x_{data} - x_{NET} \tag{8}$$

i.e. the performance measure J

$$J = res^T \cdot res \tag{9}$$

by searching for the best weight matrices $\boldsymbol{w}$ for the artificial neural network (6). This search can be performed with any parameter optimization or nonlinear function fitting procedure. However, because of the large number of elements in $\boldsymbol{w}$, it is of advantage to make use of routines that can make use of the sensitivity of the μ with respect to the biomass x and the sensitivity of the biomass to the network weights $\boldsymbol{w}$. The routine used here was Matlab®'s *lsqcurvefit*.

In order to determine *res* we must formulate a function computing x_{NET} from the actual network output (using the weights $\boldsymbol{w}$) corresponding to a given measured biomass x_{data}.

$$x_{NET} = xfunction(w, x_{data}) \tag{10}$$

In simple feedforward artificial neural networks with a single hidden layer, $\boldsymbol{w}$ is composed of two partial matrices: $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$, connecting input with hidden and hidden with output layer.

With given weights, μ_{NET} can be determined for each online input data vector Data (7).

$$h_1 = \tanh(w_1 \cdot Data) \tag{11}$$

where h_1 is an auxiliary vector whose elements are the responses of the nodes of the hidden layer on the Data at the input-nodes. The network output μ_{NET} is then taken as

$$\mu_{NET} = w_2 \cdot h_1 \tag{12}$$

This μ_{NET} can be used recursively in an Euler integration loop to solve equation (4):

$$x_{NET}(i) = x_{NET}(i-1) \cdot \left(1 + dt \cdot \mu_{NET}(i-1)\right) \tag{13}$$

With the initial value $x_{NET}(1)=x_0$, the amount of biomass after inoculation, equation (4) can easily be solved numerically.

As the online data are available with rather small time increments dt, the missing biomass at $t=t_{i-1}$ can be approximated by the last network estimate:

$$Data(i-1) = \left[tCPR(i-1),\ x_{NET}(i-1),\ t_{ind}(i-1)\right]^{T} \tag{14}$$

In order to determine the Jacobian matrix, needed to support a gradient search for the best $\boldsymbol{w}$ in case of a given data record, we again make use of differential equation (4) and differentiate the equation partially with respect to $\boldsymbol{w}$

$$\frac{\partial}{\partial w}\frac{dx}{dt} = \frac{\partial}{\partial w} \cdot \mu(x, w) \cdot x \tag{15}$$

$$\frac{d}{dt}\frac{\partial x}{\partial w} = \left(x \cdot \frac{\partial \mu}{\partial x} + \mu\right)\frac{\partial x}{\partial w} + x \cdot \frac{\partial \mu}{\partial w} \tag{16}$$

This is an ordinary differential equation in $S = \frac{\partial x}{\partial w}$, the so-called sensitivities, that can be solved with the initial conditions

$$\frac{\partial x(t=0)}{\partial w} = S(t=0) = 0 \tag{17}$$

as there is no influence of $\boldsymbol{w}$ on x in the beginning. Hence this approach is termed the sensitivity equation approach.

However, in order to numerically compute the solution of equation (16) we need expressions for the other sensitivities $\frac{\partial \mu}{\partial x}$ and $\frac{\partial \mu}{\partial w}$.

The first expression is one component of the sensitivity of the network output with respect to changes in its input. This output/input-sensitivity S_{IO} of a simple feedforward network can be determined from equations (11) and (12). Because of the special choice of the activation function (11) of the network nodes, a rather simple expression appears:

$$S_{IO} = w_1^{\,T} \cdot \left(w_2^{\,T} \cdot \left(1 - h_1 \cdot h_1\right)\right) \tag{18}$$

We only need the sensitivity with respect to biomass concentration X

$$\frac{\partial \mu}{\partial x} = S_{IO}(2) \tag{19}$$

The other component $\frac{\partial \mu}{\partial w}$ = S_{OW} is the sensitivity of the network output with respect to the network parameters, the weights. In the networks with single hidden layer, this is according to equations (11) and (12)

$$\frac{\partial \mu_{net}}{\partial w_2} = h_1 \quad \text{and} \quad \frac{\partial \mu_{net}}{\partial w_1} = w_2^{\,T} \cdot \left(1 - h_1 \cdot h_1\right) \cdot Data = h_2 \tag{20}$$

Combining both parts of equation (20) yields

$$\frac{\partial \mu_{net}}{\partial w} = \begin{bmatrix} h_2 \\ h_1 \end{bmatrix} \tag{21}$$

The quantities $\frac{\partial \mu}{\partial x}$, $\frac{\partial \mu}{\partial w}$ and $\frac{\partial x}{\partial w}$ can directly be determined within the Euler loop (13) for integrating $\frac{dx_{NET}}{dt}$.

So far we assumed that the network weights ***w*** were already given. In practice, they must be determined by means of a curve fitting procedure. We used Matlab®'s *lwqcurvefit* for this purpose. This function was used in the mode where the Jacobian can be exploited.

Using the trained neural net relationship for μ in equation (4), the total biomass x(t) can be derived much more accurately from the measured total carbon dioxide production rate signal. This is shown in Figure 4 for the same set of experimental data that were used to generate Figure 2.

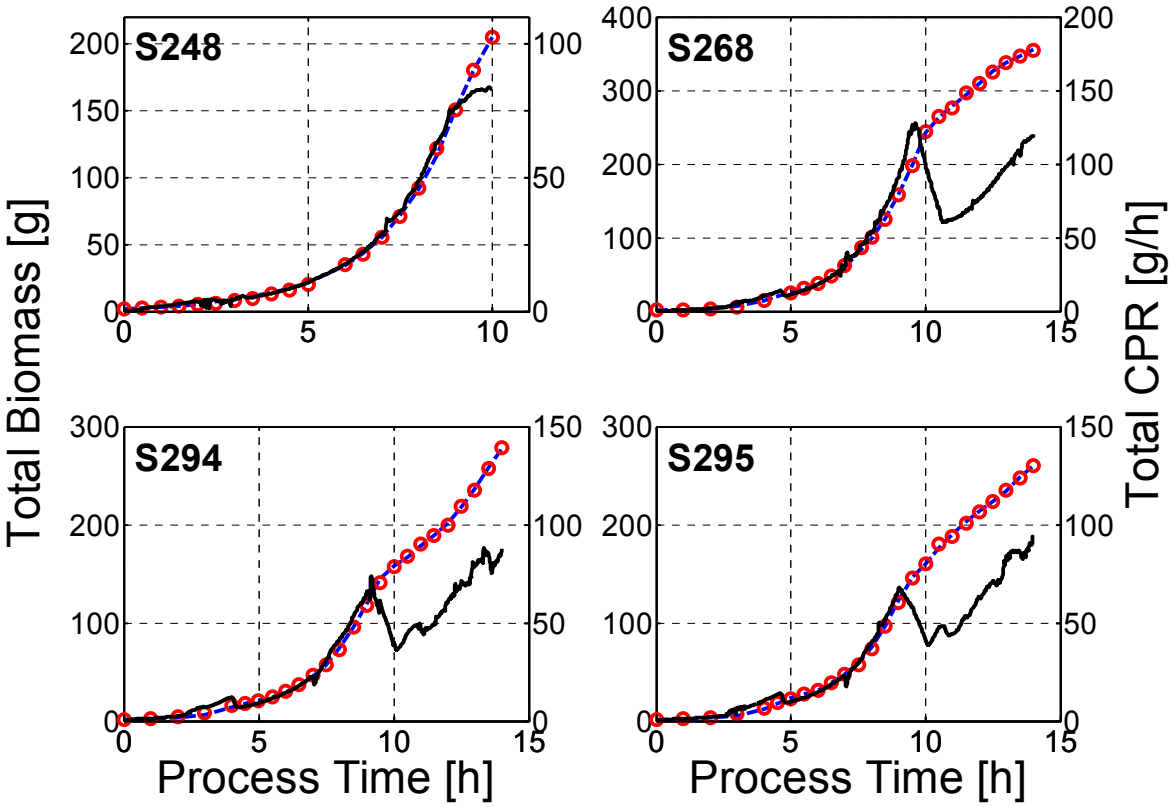

Figure 4. Total biomass trajectories (dashed lines) derived from the total carbon dioxide production rate signals (full lines) by means of the artificial neural network described in the text. The symbols (○) are the offline measured total biomass values.

The agreement between the computed x values and the biomass measurement data is clearly better than in the case depicted in Figure 2. The root mean square error with respect to the total biomass is now 1.4 [g] which is about 0.5 [%] of the final biomass value or a fifth of the RMSE resulting from the more simple linear model based on Luedeking and Piret's approach.

In order to determine a the setpoint profile for a tcCPR-controller for a given μ(t)-profile, we must use a model that maps in the reverse direction. For that purpose we took a simple feedforward artificial neural network with a single hidden layer trained on the μ-data from the network described above together with the corresponding biomass x(t) and tcCPR(t) data.

Control algorithm

The control can then be performed with a very simple adaptive control algorithm comprising the following steps:

(i) Calculation of total cumulative carbon dioxide production rate tcCPR

$$tcCPR(t) = \int_0^t CPR(t) \cdot W(t) \cdot dt \qquad (22)$$

(ii) Computing the deviation of measured tcCPR from the set point $tcCPR_{set}$

$$\Delta tcCPR = tcCPR_{set} - tcCPR_{meas} \tag{23}$$

(iii) Determination of the correction γ for the manipulated variable F

$$\gamma = 1 + k_1 \cdot \Delta tcCPR + k_2 \cdot \int_{t_s}^{t} \Delta tcCPR \cdot dt \;; \quad 0.7 \le \gamma \le 1.3 \tag{24}$$

(iv) Calculation of current feed rate F

$$F = \gamma \cdot F_{ref} \tag{25}$$

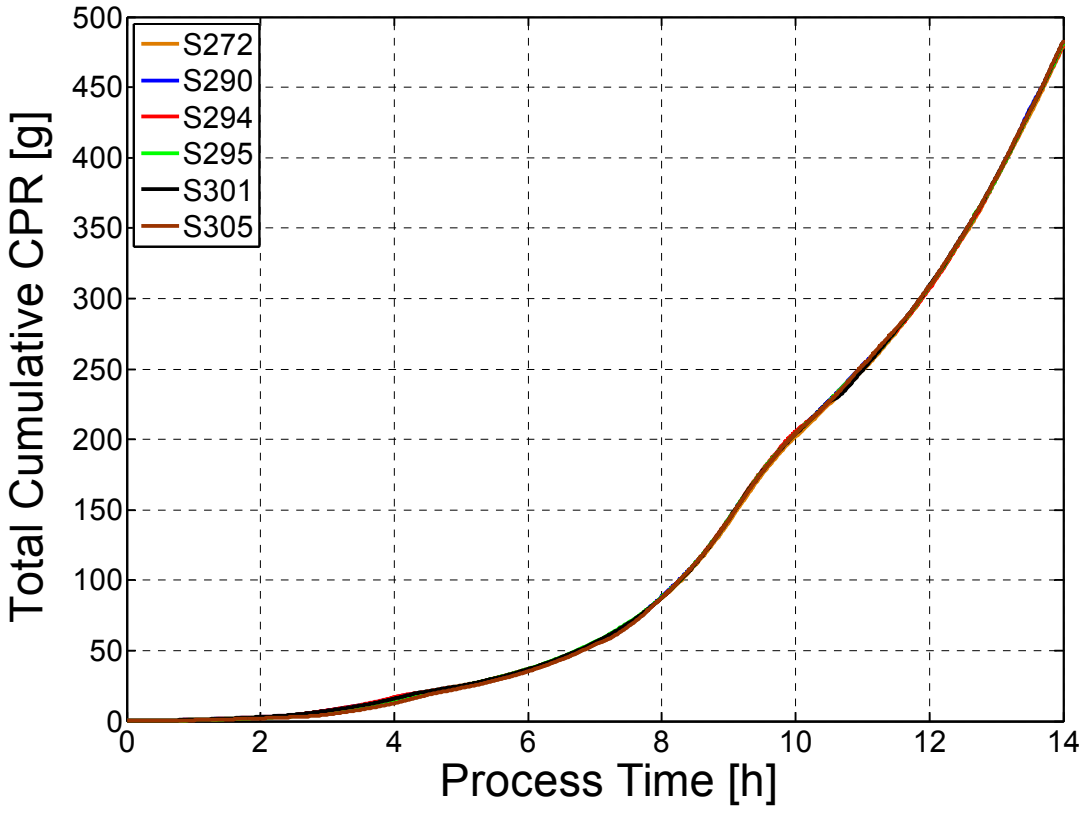

Figure 5. Total cumulative CPR profiles of six tcCPR-controlled fed batch fermentation runs.

As the CPR and thus the total cumulative signal as well are directly available without any noticeable time delay, the flow rates F can be adjusted immediately. As can be seen in Figure 5, where the results of 6 successively cultivations are shown, the tcCPR profiles can

nearly ideally be kept on track. Again, the relative deviations show, that the deviations from the mean tcCPR-profile are kept below 1 [%] within one hour after the controller was switched on (Figure 6).

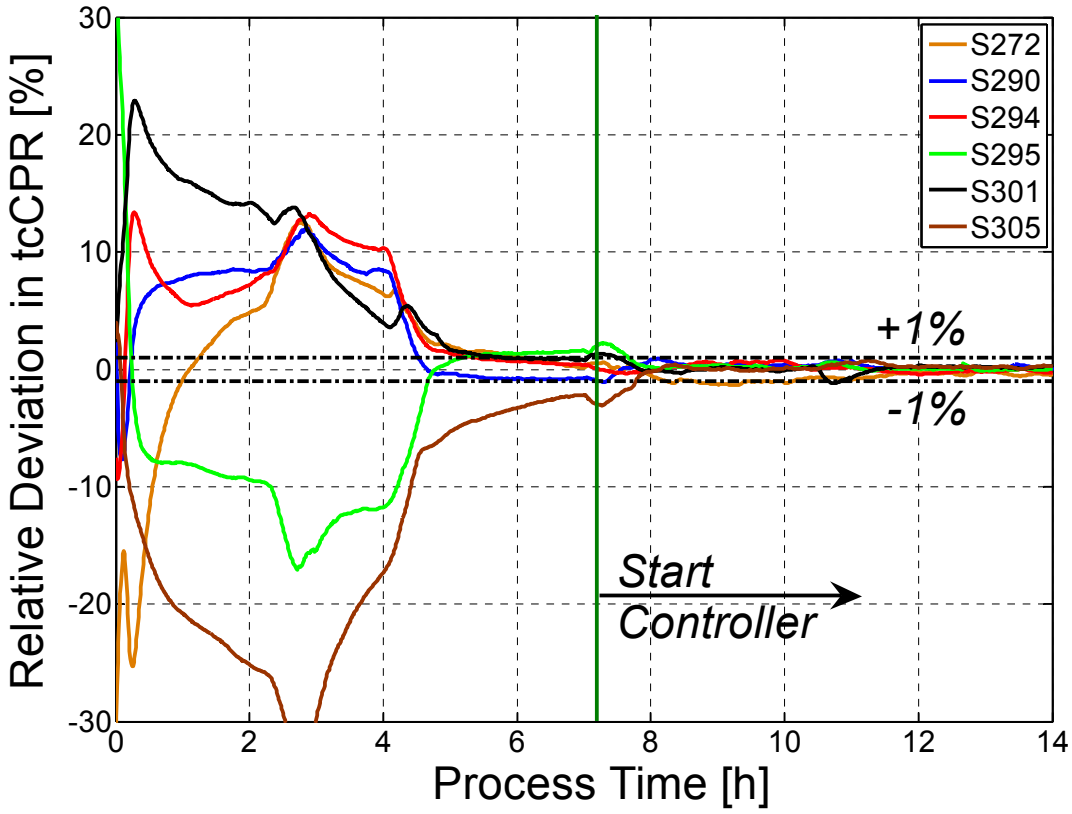

Figure 6. Relative deviations of the cumulative total carbon dioxide produced from their means.

Thus, the tcCPR-control is perfect. However, the questions are: what are the corresponding total biomass profiles, and how do they deviate from their mean. The answers are given in Figures 7 and 8. In the beginning the measurement errors are large with respect to the current biomass values; hence the fluctuations appear to be very high. Nevertheless the trajectories in Figure 8 finally remain within the ±5 [%] margins. This confirms that we do not only keep the controlled variable on target but more importantly one of the key state variable, the biomass x.

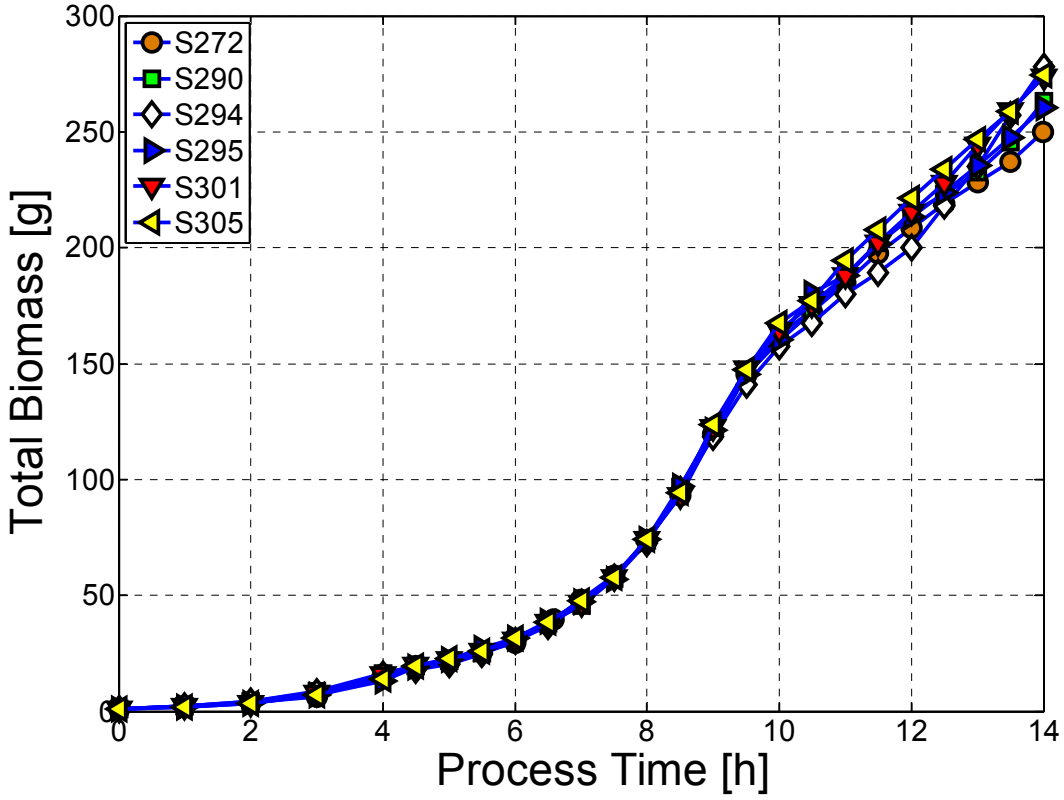

Figure 7. Total biomass profiles of six tcCPR-controlled fed batch fermentation runs.

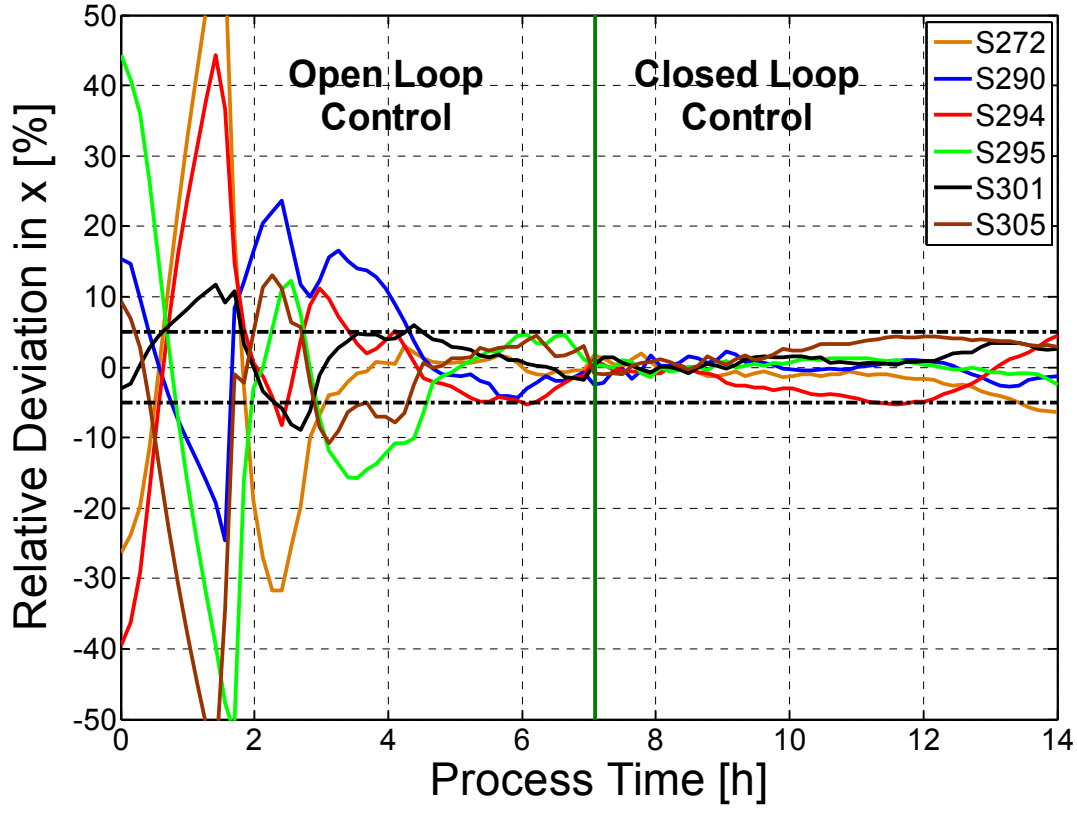

Figure 8. Relative deviations between the corresponding trajectories for the total biomass x from their mean.

The next question is whether or not the corresponding specific growth rate profile is really the one initially taken to determine the x and tcCPR profiles. Figure 9 depicts the μ estimates computed with the full model, described by Gnoth *et al.* (2006).

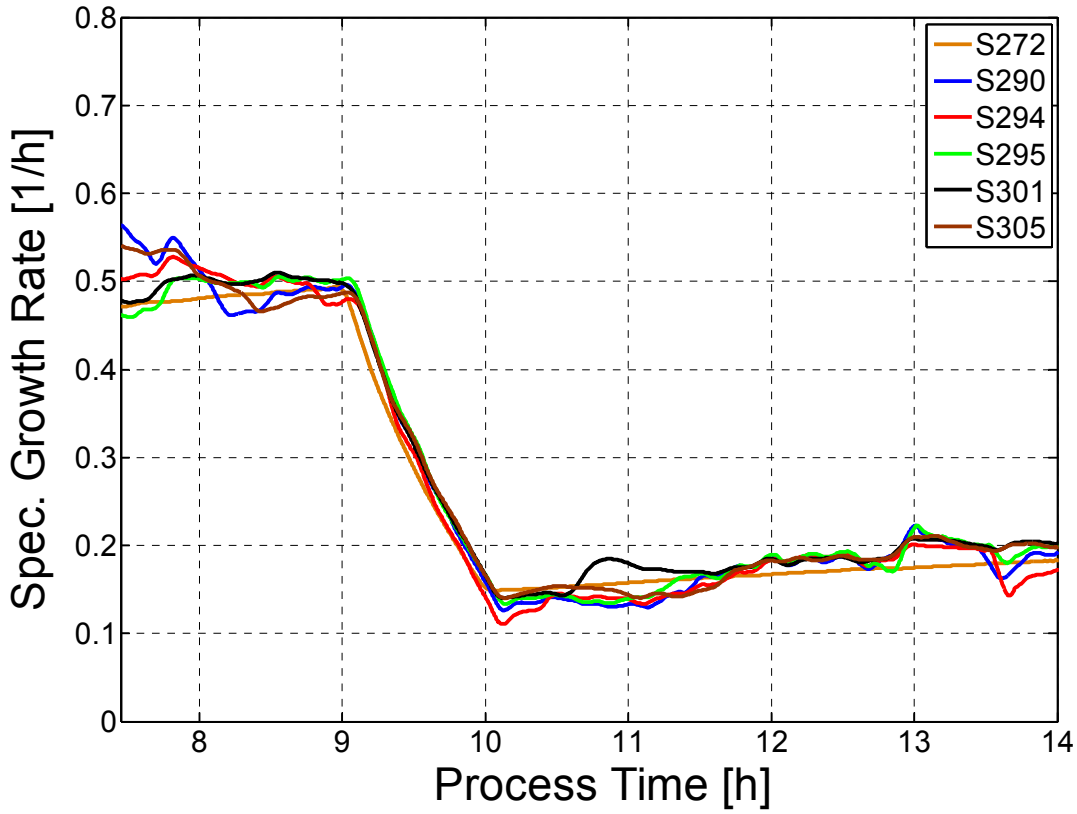

Figure 9. Specific growth rate profiles of six tcCPR-controlled fed batch fermentation runs.

Figure 10 finally depicts the measured product concentrations for the tcCPR-controlled experiments. The divergence in the protein data is slightly higher than those in the biomass profiles. However, one should keep in mind that protein measurement accuracies are drastically lower than the accuracies in the other measurement data. The estimated relative error of the measurements is ±15 [%]. All the data shown in Figure 10 stay well within these error bounds around the mean profile. One of the cultivations was performed without inducing the protein formation. This was done in order to see whether the promoter is really tight and in order to get an impression on the metabolic load of the cells by foreign protein formation.

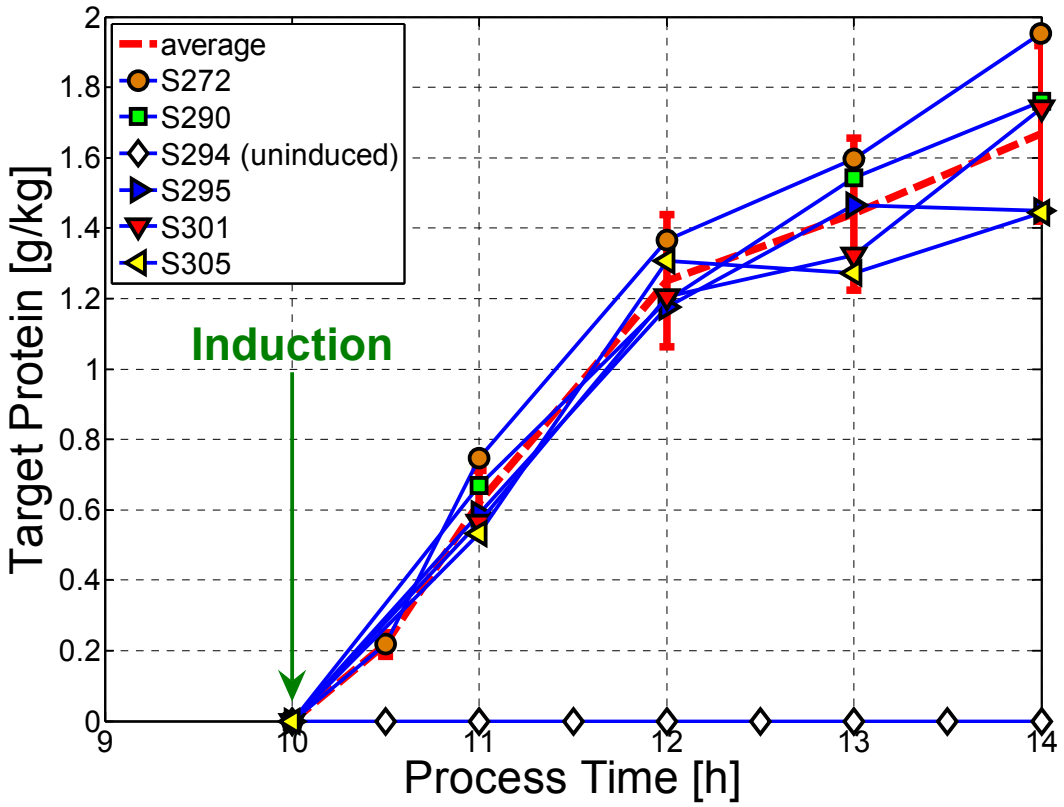

Figure 10. Concentration profiles of target protein in six tcCPR-controlled fed batch fermentation runs together with an average drawn only over the records from the induced fermentations. The error bars showing the confidential interval of ± 15 [%] resulting from the technique of protein analysis.

3 DISCUSSION

As in most cases the physiology of the cells and particularly the protein formation is mainly influenced by the specific growth rate of the cells (Neidhardt *et al.* 1990), it is straightforward to guide a protein production process along an optimal μ(t) profile. As was shown, such μ-controlled fermentation processes depict problems in the batch-to-batch reproducibility. Hence, alternative variables must be monitored and manipulated in order to make sure that the process is tightly reproduced in each batch.

Choosing the total cumulative CPR-profile derived from a given profile of the specific growth rate μ(t) has several important practical advantages. It is first of all robust as compared to the biomass concentration profiles. Following a tcCPR-profile automatically corrects for deviations from the biomass concentration profile. The second important point is that tcCPR can quite accurately be measured online. CPR as measured via the usually applied offgas analysis is a global quantity that does not depict such heavily fluctuating signals as any variable locally measured with a probe within the multiphase flow of the culture. Finally, a global measurement it is more representative than a locally measured variable. This property is favourable in our models that all assume homogeneity or ideal stirred tank conditions in bioreactors.

As the results depicted in Figure 5 show, the batch-to-batch reproducibility with respect to the tcCPR is very good. Additionally, the robustness of the cultivation strategy is also surprisingly good as shown in Figure 3. All deviations within the decisive product formation phase after induction could be kept below ±1 [%]. Even the biomass profiles were simultaneously found to follow closely the profiles observed in the previous runs. And, last but not least, the product titers were observed to be very reproducible.

Hence, keeping the cultures operated during recombinant protein production on predefined tcCPR profiles, that are derived from the desired profiles of the specific growth rate μ(t) is a very good strategy for keeping fed batch processes on the optimal track.

Acknowledgements. We have been supported financially by the ministry of cultural affairs of the German state "Sachsen-Anhalt". We gratefully acknowledge this support.

REFERENCES

DePalma, A. (2004), PAT: Taking Process Monitoring to Next Level, *Gen. Eng. News*, 24(9), 46-47

FDA (2004), Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Manufacturing and Quality Assurance, http://www.fda.gov/cvm/guidance/published.html

Gnoth, S., Jenzsch, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006), Robust process operation of microbial cultures for recombinant protein production improving batch-to-batch reproducibility, to be published

Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006c), Open loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J. Biotechnol.*, in press

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A. (2006d), Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile, *Bioproc. Biosyst. Eng.*, submitted

Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A. (2006b), Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioproc. Biosyst. Eng.*, 29(1), 19-27

Jenzsch, M., Simutis, R., Lübbert, A. (2005), Application of Model Predictive Control to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9),* Pons MN, van Impe JFM, eds., IFAC/Elsevier, ISBN 0 08 044251 X

Jenzsch, M., Simutis, R., Lübbert, A. (2006a), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

Levisauskas, D., Simutis, R., Borvitz, D., Lübbert, A. (1996), Automatic control of the specific growth rate in fed-batch cultivations processes based on exhaust gas analysis, *Bioproc. Eng.*, 15(3), 145-150

Luedeking, R., Piret, E.L. (1959a), A kinetic study of the lactic acid fermentation. Batch process at controlled pH, *J. Biochem. Microbiol. Technol. Eng.*, 1, 393-412

Luedeking, R., Piret, E.L. (1959b), Transient and steady state continuous fermentation. Theory and experiment, *J. Biochem. Microbiol. Technol. Eng.*, 1, 431-459

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

Pirt, S.J. (1975), *Principles of Microbe and cell cultivation*, Blackwell Scientific Publication, London, UK, 115-117

Shioya, S. (1992), Optimization and control in fed-batch bioreactors, *Adv. Biochem. Eng. Biotechnol.*, 46, 111-142

Summary

In contrast to the immense achievements in fundamental molecular biological sciences, the fermentation and downstream processing technologies used in industry have not been developed at the same pace. Since they did not receive as much public interest as the biological sciences, much less money and efforts flew into that domain.

With respect to process quality, there is a considerable lag as compared with manufacturing processes in other industries. Fermentations are running on practically the same level of control as 20 years ago. Since the product quality is dependent on the reproducibility, the accrediting administrations demand improvements. For that purpose the FDA suggests to first analyze the processes more accurately in order to improve the mechanistic knowledge about the process dynamics and then to draw consequences in terms of robust design of process operational procedures as well as process supervision and feedback control in the engineering interpretation.

This work has shown how significant improvements of the batch-to-batch reproducibility can be achieved at recombinant protein production processes with *E.coli* cultures. Priority number one in the arsenal of methods is the design of robust operational procedures. Here it was shown how the biomass concentration at induction time can be kept close to a given target value.

Obviously, random distortions can nevertheless disturb the process. These can only be eliminated by automatically controlling the process with feedback controllers. These, however, require a well performing process state estimation technology.

In this respect the multivariate data analysis techniques were shown to do a good job. However, more powerful than the established techniques of nonlinear multivariate data analysis are, as shown in this work, artificial neural networks. These proved to be well performing and stable, provided enough process data records were used for its training. The latter is no real problem in manufacturing plants where the processes running very often under nearly the same operational conditions.

The proposed specific growth rate control procedures are quite universal. They allow realizing complicated specific growth rate profiles during the cultivation. This feature was very important for obtaining the information about the process' dynamics needed for bioprocess optimization, in particular the dependency of the specific product formation rates from the specific biomass growth rates. With these controllers good control performance was obtained.

From the cell-physiological point of view the specific growth rate would be the most important variable to be controlled in fed-batch fermentation. Thus, much work was put into control procedures to put μ-control into practice. However, this may lead to problems from the process reproducibility point of view. If batch-to-batch reproducibility is an issue, control of biomass is clearly superior. Instead, from the μ-profile a corresponding x-profile was controlled in a feedback fashion.

In a further approach, the total cumulative CPR was chosen as control variable, which had several important practical advantages. It was first of all robust as compared to the biomass concentration profiles. Following a profile of the total cumulative CPR (tcCPR) automatically corrects for deviations from the biomass concentration profile. The second important point was that tcCPR can quite accurately be measured online. Finally, as a global measurement it is more representative than a locally measured variable, a property that is necessary in our

models that all assume homogeneity or ideal stirred tank conditions in bioreactors. Hence, keeping the cultures during recombinant protein production on predefined tcCPR profiles that are derived from the desired profiles of the specific growth rate μ(t) was a very good strategy for keeping fed batch processes on target.

Many control strategies have been discussed in literature but there is practically no industrial plant where control of variables other than pH, T, and may be pO_2 has been realized. Most of the control approaches published so far are rather complex and there is some fear in industry to apply them. From that point of view, the control procedures developed during the work reported about here are extremely simple adaptive approaches that can immediately implemented at any fermentation process. They have been tested in more than a hundred fermentations where recombinant proteins were generated with *Escherichia coli* and *Pichia pastoris* cells. It proved to be very robust and practicable even in large-scale bioreactors.

There is no longer any excusing for manufacturers the stay away from automatic control of the feeding to their fermenters. This is particularly true for production processes of biologics where the processes are approved together with the products themselves. It now becomes a must for manufacturers not only in pharmaceutical industries, to invest in knowledge-based process monitoring and control.

Zusammenfassung

Im Gegensatz zum rasanten Fortschreiten der Entwicklungen in den molekularbiologischen Wissenschaften der letzten Jahrzehnte, konnten die in der industriellen Praxis angewandten Fermentations- und Aufarbeitungstechniken bei weitem nicht mithalten. Zudem ist das öffentliche Interesse hier sehr viel geringer, als in den Biowissenschaften, was letztlich dazu führte, dass der Bereich der Bioprozesstechnik sehr viel weniger Aufmerksamkeit und finanzielle Unterstützung erhielt.

Nicht zuletzt das ist der Grund dafür, dass der Technisierungsgrad biotechnologischer Produktionsanlagen, verglichen mit dem in anderen industriellen Produktionsprozessen, erheblich zurückliegt. Fermentationsprozesse zur Herstellung pharmazeutischer Wirkstoffe laufen fast auf dem gleichen Entwicklungsstand, wie vor 20 Jahren.

Bei der Produktion therapeutischer Wirkstoffe mit Hilfe rekombinanter Organismen, wird die Produktqualität in erster Linie von der Prozessqualität beeinflusst. Da die Prozessqualität direkt von der Reproduzierbarkeit der Produktionsprozesse abhängt, fordern die für die Zulassung pharmazeutischer Produktionsanlagen zuständigen Behörden, wie FDA oder EMEA, die Betreiber dieser Anlagen jetzt auf, mehr Entwicklungsaufwand in die Verbesserung ihrer Produktionstechnologien zu investieren. Dabei werden nicht nur Forderungen gestellt, sondern auch konkrete Vorschläge gemacht, wie diese Verbesserungen aussehen können.

Diese Arbeit zeigt konkret, wie signifikante Verbesserungen bei der „batch-to-batch"-Reproduzierbarkeit von mikrobiellen Fermentationsprozessen zur Herstellung rekombinanter therapeutischer Proteine erzielt werden können. Einer der zentralen Punkte ist dabei die Entwicklung robuster Prozessführungsstrategien. Es wird gezeigt, wie die Fermentation in der ersten Prozessphase vor der Induktion, durch eine robuste Steuerung reproduzierbar gefahren werden kann. Diese Prozessführung erlaubt zudem die Kompensation kleinerer Schwankungen im Prozess.

Da zufällige Störungen nie ausgeschlossen werden können, müssen die Produktionsanlagen mit automatischen Reglern ausgestattet werden, welche die Abweichungen von den optimalen Prozesstrajektorien eliminieren können. Um diese Abweichung überhaupt detektieren zu können, sind geeignete Überwachungsmethoden, die den aktuellen Prozesszustand hinreichend genau abbilden, notwendig.

Für diesen Zweck, wurden verschiedene multivariate Analyseverfahren als Online-Prozessmonitore entwickelt und erfolgreich in bestehende Produktionsanlagen im Labor- und Produktionsmaßstab implementiert. Die besten Ergebnisse bei der indirekten Online-Messung kritischer Zustandsgrößen, wie z.B. der Biomasse, konnten mit Hilfe künstlicher neuronaler Netze erzielt werden.

Aus zellphysiologischer Sicht ist die spezifische Wachstumsrate μ die wichtigste zu regelnde Größe, da diese eine zentrale Schlüsselfunktion im Stoffwechsel der Produktionsorganismen einnimmt und damit hauptsächlich die Produktivität des Prozesses bestimmt. Deshalb wurde viel Arbeit in die Entwicklung verschiedener Regelungskonzepte für die spezifische Wachstumsrate investiert.

Mit Hilfe der in dieser Arbeit entwickelten Regelungsverfahren konnten auch sehr komplizierte μ-Profile umgesetzt werden, was das Verständnis über die Prozessdynamik und

vor allem über den Zusammenhang zwischen Wachstum und Produktbildung erheblich verbesserte.

Eine alleinige Regelung der Wachstumsrate kann jedoch zu Problemen bei der Reproduzierbarkeit führen. Deshalb wurden erweiterte Regelungskonzepte entwickelt, die z.B. ein Gesamtbiomasseprofil, welches auf der Grundlage des zuvor definierten μ-Profils berechnet wurde, regeln.

In einem weiteren Ansatz, wurde das integrale Signal der Kohlendioxidproduktionsrate als Regelgröße verwendet, was mehrere entscheidende Vorteile mit sich brachte. Zum einem ist diese Regelung vergleichsweise robust, wie der Algorithmus für die Gesamtbiomasse, und zum anderen wird hier eine Größe geregelt, die aus einem direkten Messsignal abgeleitet werden kann, d.h. ein zusätzliches indirektes Messverfahren wird nicht benötigt. Zudem ist die integrale CPR, welche sich aus der CO_2-Bilanz um den Reaktor ergibt, und eine globale Messgröße darstellt, viel besser geeignet den Prozess repräsentativ abzubilden, als irgendein anderes lokales Messsignal. Die erzielten Reglungsergebnisse zeigten die enorme Leistungsfähigkeit dieses Ansatzes im Bezug auf die Reproduzierbarkeit des Prozesses.

Viele Reglungskonzepte wurden bisher in der Literatur besprochen, allerdings gibt es praktisch keine einzige industrielle Produktionsanlage für „Biologics“ in der, abgesehen von Temperatur, pH und pO_2, eine geschlossene Regelung installiert wurde. Diese Zurückhaltung begründet sich u. a. mit der Komplexität der meisten bisher in der Literatur diskutierten Ansätze. Im Gegensatz dazu basieren die Verfahren, die in dieser Arbeit vorgestellt wurden, auf extrem einfachen und robusten Regelungskonzepten. Diese wurden in weit über hundert Fed-Batch-Fermentationen zur Produktion rekombinanter therapeutischer Proteine mit *Escherichia coli* und *Pichia pastoris*, sowohl im Pilot-, als auch Produktionsmaßstab von mehreren Kubikmetern erfolgreich getestet.

Es gibt daher für die Hersteller von Fermentationsprodukten keine Entschuldigung mehr dafür, auf eine automatische Regelung der Zufütterung ihrer Produktionsfermenter zu verzichten. Das trifft vor allem auf die Hersteller von „Biologics“ zu, bei denen die Produktionsverfahren zusammen mit den Produkten selbst zugelassen werden. Es ergibt sich damit die zwingende Notwendigkeit, dass diese Hersteller in erweiterte Prozessüberwachung und Regelung ihrer Anlagen investieren.

Publications List

The results of this work were published in 11 papers.

[1] Jenzsch, M., Lange, M., Bär, J., Rahfeld, J.U., Lübbert, A. (**2004**), Bioreactor retrofitting to avoid aeration with oxygen in Pichia pastoris cultivation processes for recombinant protein production, *Chemical Engineering Research and Design*, 82(A9), 1144-1152

[2] Jenzsch, M., Simutis, R., Lübbert, A. (**2004**), Knowledge or Model-Based Control of the Specific Growth Rate in Fed-Batch Cultivations for Recombinant Protein Production, 81-88 in: *BATCHPRO-Symposium Proceedings*, Poros, Greece

[3] Jenzsch, M., Simutis, R., Lübbert, A. (**2004**), Model-supported control of recombinant protein production fermenters: Pros and Cons of Different Approaches, 84 in: *Proceedings of the European Symposium on Biochemical Engineering Science*, ESBES-5, Stuttgart

[4] Jenzsch, M., Simutis, R., Lübbert, A. (**2005**), Application of Model Predictive to Cultivation Processes for Protein Production with Genetically Modified Bacteria, 511-516 in: *Computer Application in Biotechnology 2004 (CAB9)*, Pons M.N., van Impe J.F.M., eds., IFAC/Elsevier, ISBN 0 08 044251 X

[5] Jenzsch, M., Simutis, R., Lübbert, A. (**2005**), How to cope with FDA's PAT-initiative with respect to fermentation process monitoring and control, *J. Biotechnol.*, 118(1), Suppl.1, 49-50

[6] Jenzsch, M., Simutis, R., Lübbert, A. (**2005**), New developments in model-supported optimization and control of microbial cultivation processes, 97-111 in: *Advances in Science and Industrial Production of Baker's Yeast*, Proc. VH-Yeast Conference, Berlin

[7] Jenzsch, M., Simutis, R., Lübbert, A. (**2006**), Optimization and Control of Industrial Microbial Cultivation Processes, *Eng. Life Sci.*, 6(2), 117-124

[8] Jenzsch, M., Simutis, R., Lübbert, A. (**2006**), Generic model control of the specific growth rate in recombinant Escherichia coli cultivations, *J. Biotechnol.*, 122(4), 483-493

[9] Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I., Lübbert, A. (**2006**), Estimation of biomass concentrations in fermentation processes for recombinant protein production, *Bioproc. Biosyst. Eng.*, 29(1), 19-27

[10] Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., Lübbert, A. (**2006**), Open loop control of the biomass concentration within the growth phase of recombinant protein production processes, *J. Biotechnol.*, in press

[11] Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., Lübbert, A. (**2006**), Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile, *Bioproc. Biosyst. Eng.*, submitted

The results obtained over the last four years have been presented in lectures **(L)** and on posters **(P)** on the following national and international scientific conferences and meetings:

DECHEMA-Jahrestagung, Wiesbaden, Germany
June 11th-13th, 2002
„Modellgestützte Optimierung der periplasmatischen Produktion eines rekombinanten Proteins in *E.coli*“ **(P)**

1st Halle Conference on Recombinant Protein Production, Halle (Saale), Germany
February 27th- March 1st, 2003
“Model-Supported Optimization of *E.coli*-Cultivations for the Production of Therapeutic Proteins” **(L)**

11th European Congress on Biotechnology (ECB11), Basel, Switzerland
August 24th-29th, 2003
“Optimization and Control of Fermentations in New Recombinant Protein Production Processes” **(L)**

Bioprocess Engineering Course 2003, Supetar, Island Brač, Croatia
September 7th-13th, 2003

9th International Symposium on Computer Application in Biotechnology (CAB9), Nancy, France
March 28th-31st, 2004
“Application of Model Predictive to Cultivation Processes for Protein Production with Genetically Modified Bacteria” **(P)**

BIOPERSPECTIVES 2004, Wiesbaden, Germany
May 4th-6th, 2004
“Generic Model Control of the Specific Growth Rate in a Recombinant Protein Production Process with *Escherichia coli*“ **(P)**
“Avoiding Pure Oxygen Supply in a High Cell Density *Pichia pastoris* Cultivation Process by Bioreactor Retrofitting“ **(P)**

DECHEMA/ GVC Vortrags- und Diskussionstagung “Simultane und integrierte Bioprozess-entwicklung”, Eisenach, Germany
May 17th-19th, 2004
„Vergleich verschiedener Regelungsmöglichkeiten für die spezifische Wachstumsrate von *Escherichia coli* bei der Produktion rekombinanter Proteine“ **(L)**

BATCHPRO Symposium, Island Poros, Greece
June 6th-9th, 2004
“Knowledge or Model-based Control of the Specific Growth Rate in Fed-batch Cultivations for Recombinant Protein Production” **(L)**

ESBES 5, Stuttgart, Germany
September 8th-11th, 2004
“Model-supported Control of Recombinant Protein Production Fermenters: Pros and Cons of Different Approaches” **(L & P)**

12th International Biotechnology Symposium (IBS12), Santiago, Chile
October 17th-22nd, 2004
"Comparison of Different Model-Supported Control Procedures for Recombinant Protein Production Fermenters" **(L)**

2nd Halle Conference on Recombinant Protein Production, Halle (Saale), Germany
March 3rd-5th, 2005
"Responses to FDA's PAT-initiative with respect to fermentation process monitoring and control" **(L)**

BIOPERSPECTIVES 2005, Wiesbaden, Germany
May 10th-12th, 2005
"Experiences with process monitoring and control in recombinant protein production" **(P)**

ECI - BIOCHEMICAL ENGINEERING XIV, Harrison Hot Springs, B.C., Canada
July 10th-14th, 2005
"Mixing, process monitoring and control" **(L)**

12th European Congress on Biotechnology (ECB12), Copenhagen, Denmark
August 21st-24th, 2005
"How to cope with FDA's PAT-Initiative with Respect to Fermentation Process Monitoring and Control" **(P)**

ACHEMA 2006, Frankfurt am Main, Germany
May 15th-19th, 2006
"Quality Improvements at Recombinant Protein Production Processes Following the Proposals of FDA`s PAT-Initiative" **(L)**

DECHEMA/ GVC Vortrags- und Diskussionstagung "Industrielle Biotechnologie und Gewinnung von Produkten", Würzburg, Germany
May 22nd-24th, 2006
„Steigerung der Reproduzierbarkeit bei der Herstellung rekombinanter Proteine durch robuste Prozessführung" **(L)**

ECI - Natural Products Discovery and Production: New Challenges, New Opportunities, Santa Fe, New Mexico, USA
June 4th-8th, 2006
"Improvement of the reproducibility of fermentation processes by robust process operational design and control: an answer on FDA's PAT–Initiative" **(P)**

Curriculum Vitae

General	*Name:*	Marco Jenzsch
	Date of birth:	December 9th, 1978
	Place of birth:	Merseburg, Germany
	Citizenship:	German

School Education

09/1984 – 07/1990	Erweiterte Oberschule Spergau
08/1990 – 06/1997	Novalis-Gymnasium Bad Dürrenberg

Military Service

07/1997 – 04/1998	Basic military service in Kempten and Wildflecken

University Education

09/1998 – 11/2002	Graduation in Biochemical Engineering at Martin-Luther-Universität Halle-Wittenberg
11/2002	Degree "Diplom-Ingenieur" (final mark 96/100)
since 11/2002	Ph.D. student at Center of Bioprocess Engineering, Halle (Saale)

External Projects

07 – 08/2000	NASA Research Centre in Huntsville & University of Alabama in Huntsville, USA *"Production of PHA's with bacteria in simulated microgravity*"
02 – 03/2003	Institute of Applied Microbiology at University of Natural Resources and Applied Life Sciences, Vienna, Austria *"Analysis of plasmid stability in E.coli host cells using Agilent's Bioanalyzer®"*

Erklärung

Hiermit erkläre ich an Eides statt, dass ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe und dass diese Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht sind.

Halle (Saale), den 31. Mai 2006 Marco Jenzsch