E. FRANCO-LARA et al., Specification and Supervision of System Performance …, *Chem. Biochem. Eng. Q.* **21** (4) 335–344 (2007)

335

Specification and Supervision of System Performance in Yeast Bioprocesses

E. Franco-Lara[*], R. Krull, D.-C. Hempel, and A. Haarstrick
Institut für Bioverfahrenstechnik, Technische Universität Braunschweig,
Gaußstr. 17, 30106 Braunschweig, Germany.

Three different case studies addressing the specification of system performance and estimation of optimal cultivation policies are examined for different yeast cultivation systems. In the first case, the industrial production of baker's yeast in bubble columns serves as illustration of primary model conception and exemplifies task-oriented reduction of model complexity. Within this simplified model-supported framework, different optimization techniques, that maximize process productivity, are demonstrated and compared in a second case study, the invertase production with a recombinant *Saccharomyces cerevisiae* strain. Finally, in the third case, the integration of specific process constrains of physical nature proper of the bioprocess is demonstrated on the optimal production of the protein GAL80 with the yeast *Kluyveromyces lactis*.

Key words:
Yeast processes, system performance, model-based optimization, soft-sensor

Introduction

Every R&D agenda of biotechnology-oriented industry and university addresses to date the problem of estimating optimal policies for microbial processes. Certainly heuristic trial-and-error approaches of process development are still applied, e. g. strain optimization via selection of UV-mutated phenotypes. However, since its approval by the FDA and meanwhile widespread use of Process Analytical Technology (PAT-Guidelines),[1] an increased interest to apply model-supported methods, especially in primary pharmaceutical manufacturing, focus more and more on the mechanistic understanding of the biochemical systems.

Yeasts constitute one of the most important industrial sources of a great variety of products. These range from small molecules, like ethanol, to complex recombinant therapeutic proteins. Biomass itself can also be the primary product of such a bioprocess. For example, in case of beer or industrial baker's yeast production, yeast propagation takes up a central step and emprises an important economical and technological factor in practice. However, not only the final harvest of the bulk product is important but also the vitality and quality of the propagated yeast, which exerts a relevant influence on the subsequent fermentation run and the resulting product quality.

A total of three different yeast cultivation systems were reviewed in this paper. Exemplified on the industrial production of baker's yeast in bubble columns, the first principle model conception of a yeast bioprocess is detailed described. Subsequently and in order to make the model manageable for optimization a reduction of its complexity is carried out. The simplification is performed by assuming weak dispersion effects and by transforming from the Euler to the Lagrange coordinates, which changes the original partial differential equation into an ordinary differential equation. According to some metabolic specifications, baker's yeast represents an interesting system that is used in cultivation processes of industrial size. Besides other process parameters like substrate feeding, oxygen is observed to be the most demanding. Thus, operating at maximum oxygen transfer capacity in large production reactor units is a challenging procedure which requires a reliable management of resources and processes.

After establishing the appropriate modelling framework, the invertase production with recombinant strain of *Saccharomyces cerevisiae* serves as example of an off-line model-supported optimization. The cultivation takes place in a stirred tank bioreactor and the maximization of the total invertase activity at a fixed cultivation time is attained by manipulating the substrate feed rate. For this purpose, the feeding policy was estimated via an evolutionary strategy using an artificial neural network (ANN), whereby the parameters were calculated with a genetic algorithm (GA). The process efficiency reached with this technique is furthermore compared to other two optimization strategies, showing better performance.

Finally, the production of the recombinant protein GAL80 with a recombinant yeast strain of *Kluyveromyces lactis* rounds up the survey. This

*Corresponding author (E-Mail: ibvt@tu-bs.de)

online optimization is exemplified on fed-batch cultivations, which were run under specific growth rate-controlled conditions. Since the recombinant protein shows a linear dependency with the viable yeast concentration, the optimization goal was to maximize the biomass production per total fermentation time. While the model-supported optimization was achieved with the iterative online readjustment of the Monod-like kinetic parameters, an ANN-based soft-sensor was used to independently monitor the process performance. That is, even though the macro-model serves as basis for the optimization procedure, it does not guarantee that an optimal state may be reached just by adjusting the parameters using only the available biomass concentration measurements. The alternative ANN-based soft sensor integrate additional online variables, like the oxygen uptake rate (*OUR*) and the carbon dioxide evolution rate (*CER*), enabling the online analysis of the influence, that several process constraints may have on the final process performance.

Problem statement: Modelling the industrial production of baker's yeast in bubble column bioreactors

Especially due to high costs associated with many fermentation processes and the increasingly competitive industry, prompt application of any modern control technique to industrial bioprocesses is more than desirable. Unfortunately, those control techniques are very often hampered by the absence of continuous, on-line measurement of the most relevant process variables, as well as by the lack of adequate mathematical models. Usually, in production-scale bioreactors microorganisms are exposed to a dynamically changing and heterogeneous environment. This may cause a loss in viability, reduction of the yield of biomass or desired metabolites, and a probable increase of undesired by-products. In fed-batch cultivations of baker's yeast, gradients in the reactor may occur in substrate and oxygen concentrations and in pH value. Baker's yeast production for food industry targets high-cell-densities cultivations, i. e. the achievement of maximum biomass yields in space and time. This also implies that the carbohydrate feedstock does not play a marginal role in the case of the time variable. Thus an optimization is required because the ATP yield from possible alcoholic fermentation is much lower than that from respiratory sugar dissimilation and, consequently, the alcoholic fermentation negatively affects biomass yield and should be avoided during the yeast production. Alcoholic fermentation and respiration occur simultaneously whenever the specific growth rate and/or the sugar concentration exceed a critical value (Crabtree effect). For further

details the reader is referred to as most comprehensive discussions and derived kinetic model of this topic given by *Sonnleitner and Käppeli*.[2]

In order to optimize the yeast production, the Crabtree effect must be avoided. Fed-batch operation is for this case best suited controlling the critical specific growth rate via the supply of sugar concentration that has to be dosed. This is already important during the early stages, where the maximum feasible feed rate is dictated by the critical specific growth rate at which both, respirative and fermentative sugar metabolism, occur. In most cases, the specific growth rate has to be reduced when the process reaches high-cell-densities. This procedure is applied to avoid problems with limiting oxygen transfer and/or cooling capacity. When cultivations are carried out in industrial scale fermenters, gradients of substrates will always arise. The question is if, how, and to what extent these local gradient influence the microorganisms and the process as a whole.

Gradients surge from combined effects of limited mixing and mass transfer, and yeast activities. They increase with reactor size and become more pronounced when the fed-batch technique is used, since the rate limiting substrate is fed as a concentrated solution, often in just one point of the reactor. Therefore, mixing of substrate feed and oxygen mass transfer are often problematic in industrial scale column reactors. At this scale, bubble columns and their periphery (sparger, pumps, etc.) form the most economical alternative bioreactor design for baker's yeast production. With regard to these reactors, in which mixing is accomplished by aeration alone, the flow behaviour is completely different from stirred tank reactors and in general less complex. The cultivation broth is transported in a loop, rising with the air in the central part of the column and flowing down in a loop along the column wall. One example of these phenomena can be seen in Fig. 1, which addresses the gas hold-up in a bubble column.[3] In this case, the increasing superficial gas velocity causes the apparition of heterogeneous flow patterns and, therefore, the presence of zones with different mass transfer characteristics, which again have an impact on cell growth.

These heterogeneities can be detailed described by complex computational fluid dynamic models; however, for optimization purposes a more convenient and simplified model acquainting for key variables, e. g. an average value for the dissolved oxygen concentration, must be made available. In a fed-batch mode operated airlift reactor, and according to the identified section of rising fluid, the balance equations for oxygen including uptake by microorganisms can be written as:
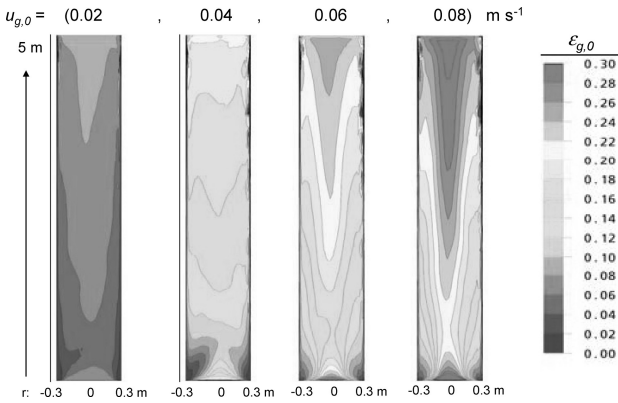
$u_{g,0} =$ (0.02   ,   0.04   ,   0.06   ,   0.08) m s^{-1}

F i g . 1 – *Local gas hold-up $\varepsilon_{g,0}$ in a bubble column (gassed liquid level: 5m, column diameter: 0.6 m) at different superficial gas velocities (0.02, 0.04, 0.06, and 0.08 m s^{-1}); the liquid used in this simulation was water with $\eta = 0.1$ mPas and $T = 20\ °C$. With increasing superficial gas velocity the simulated gas hold-up resulted in marked radial gradients. As most predominant characteristic of these heterogeneous flow, bubble agglomerations and big bubbles predominantly rise in the middle of the column and contribute to higher gas hold-ups.[3]*

$$\frac{\partial(C_{O2,L}V_R)}{\partial t} = \frac{V_R \cdot D_L}{\rho_L} \cdot \frac{\partial^2 C_{O2,L}}{\partial z^2} -$$
$$- \frac{U_L \cdot V_R}{1-\varepsilon} \cdot \frac{\partial C_{O2,L}}{\partial z} + \tag{1}$$
$$\frac{k_L a \cdot V_R}{1-\varepsilon} \cdot (C^*_{O2,L} - C_{O2,L}) - r_{O2} \cdot V_R,$$

with $C_{O2,L}$: oxygen concentration in the liquid, V_R: liquid volume, t: time, ρ_L: liquid density, D_L: dispersion coefficient, z: axial direction, U_L: superficial velocity of liquid, ε: gas hold-up, $k_L a$: mass transfer coefficient, $C^*_{O2,L}$: saturation concentration in the liquid, r_{O2}: oxygen uptake rate.

Zhang et al.[4] could show that the axial dispersion has no marked influence on the profile of oxygen concentration in the liquid phase and the assumption of oxygen gradients in gas phase is negligible. Considering a superficial gas velocity, U_G, the oxygen concentration in the gas phase at axial position, z, can finally be regarded as constant if the time interval $\Delta t > z/(U_G/(1-\varepsilon))$ and the dispersion coefficient can be neglected. That leads to the reduction of the partial differential equation (1):

$$\frac{\partial(C_{O2,L}V_R)}{\partial t} = \frac{k_L a \cdot V_R}{1-\varepsilon} \cdot (C^*_{O2,L} - C_{O2,L}) -$$
$$- \frac{U_L \cdot V_R}{1-\varepsilon} \cdot \frac{\partial C_{O2,L}}{\partial z} - r_{O2} \cdot V_R, \tag{2}$$

$$\frac{\partial(C_{O2,L})}{\partial t} = \frac{k_L a}{1-\varepsilon} \cdot (C^*_{O2,L} - C_{O2,L}) -$$
$$- \frac{U_L}{1-\varepsilon} \cdot \frac{\partial C_{O2,L}}{\partial z} - r_{O2} - D \cdot C_{O2,L}, \tag{3}$$

with: $\frac{\partial V_R}{\partial t} = F, F/V_R = D, r_{O2} = \frac{dC_{O2}}{dt} = -q_{O2} \cdot X$

and q_{O2}: specific oxygen uptake rate, X: biomass concentration, F: feed rate, D: dilution rate.

While the elimination of the dispersion term serves mainly to simplify the complexity of the model-based approach, the focus on further simplification of the advection term in equation 3 provide the juncture for a fed-batch optimization. For this purpose, the Eulerian framework has to be transformed to the Lagrange framework by coordinate transformation.[5] Based on this an infinitesimal fluid element is assumed that flows at a velocity of U_L. The axial position of the fluid element is given by ψ at time τ_0. Thus the partial differential equation (Equation 3) becomes an ordinary differential equation.

$$\frac{dC_{O2,L}(\psi,\tau)}{d\tau} = \frac{k_L a}{1-\varepsilon} \cdot (C^*_{O2,L} \cdot y - C_{O2,L}(\psi,\tau)) -$$
$$- r_{O2} - D \cdot C_{O2,L}(\psi,\tau) \tag{4}$$

$$z = \psi + u_L\tau, \ t = \tau + \tau_0, \ y = 1 - a \cdot \psi - a \cdot u_L \cdot \tau$$

With: ψ : Lagrange axial coordinate, τ : time in Lagrange coordinate, a: relative hydro-static pressure at given reactor height $[a = (1 - p(h)/p(0))/h]$.

In case of fed-batch operation conditions, the feeding and aeration process can be considered as key variables controlling and optimizing growth of yeast with the objective of obtaining maximum possible cell densities. This setting can be expressed according to *Na* et al.[6] by defining a performance function, P_F, to estimate the optimum feed rate, F_{opt}, at terminal time t_τ of the fed-batch cultivation:

$$P_F = \max(X(t_{END}) \cdot V_R(t_{END})) \tag{5}$$

For balancing and modelling fed-batch cultivation, additionally to the oxygen balance (Equation 4), three further equations can be applied for problem formulation:

$$\frac{d(X \cdot V_R)}{dt} = \mu \cdot X \cdot V_R, \tag{6}$$

$$\frac{d(S \cdot V_R)}{dt} = F \cdot S_F - (q_S + m) \cdot X \cdot V_R, \tag{7}$$

$$\frac{d(E \cdot V_R)}{dt} = (\alpha - q_E) \cdot X \cdot V_R, \tag{8}$$

With: S: glucose concentration, E: ethanol concentration, μ: specific growth rate, $q_S = \mu/Y_{X/S}$: specific glucose uptake rate, m: maintenance term, α: specific ethanol production rate, $q_E = \mu/Y_{X/E}$: specific ethanol uptake rate.

In addition to the process performance expressed by equation 5, maximum productivity should also be obtained when $0 = \alpha < \mu_{crit}$, whereby μ_{crit}: critical specific growth rate at which the yeast starts to produce ethanol, i. e. referred to feed rate, F, and reaction volume, V_R, the constraints are defined as,

$$V_{R,0} \leq V_R(t) \leq V_{R,max} = V_R(t_{END}) \quad \text{and}$$

$$0 = F_{min} \leq F_{opt}(t,\mu_{crit}) \leq F_{max}$$

Process performance specification and optimization policies

The specification of process performance is always coupled to the choice of the most convenient set of variables, such that a given productivity parameter can be maximized under certain specific process restrictions, but with the simultaneous minimization of the invested time and experimental expenditures. Since this is a problem of economic nature, it is necessary to formulate a performance function to estimate quantitatively the efficiency of any experiment with regard to its inherent costs. In this context, the use of model-supported designs to obtain suitable policies for efficient experiments can be regarded as a most effective tool to solve optimization problems.

Understood as a quantitative exploitable formulation of the current knowledge about a process under consideration, modelling is only justified when it helps to solve open issues of major relevance. Any approach fulfilling these characteristics can be used to study the influence of different control strategies by means of simulating numerically several promising scenarios of the real process. That is, once a model of a bio-system is considered to be satisfactory, it is straightforward to consider it for the optimization of the process. Such optimization policies can be applied either in open-loop (predictive) or in closed-loop (corrective) mode. In the first case, the main goal is to estimate off-line optimal trajectories of a key control variable as a function of time. In the second case, a process model is iteratively corrected, validated and finally used to keep an online set-point of the key control variable in order to optimize a predefined task. Still, the optimization of this performance function can only be accomplished within certain space of possible solutions, i. e., it is mostly subjected to a given set of constrains of physical, thermodynamic or chemical nature proper of the bioprocess. For example, in the concrete case of baker's yeast production, optimization has traditionally been done on the substrate feed rate by avoiding high concentrations of the sugar content and low oxygen concentrations to circumvent ethanol production. Such an optimization task may be formulated considering the total amount of biomass in the bioreactor as performance function, P_F, after a predefined cultivation time. Mathematically, the problem is then to maximize the product of the biomass concentration, X and the total reaction volume, V_R evaluated at certain economically convenient time, t_{END}: $P_F = \max(X(t_{END}) \cdot V_R(t_{END}))$, subjected to the constraints: $S \leq S_{max}$, $O_{2,L} \geq O_{2,L,min}$ and $V_R \leq V_{max}$, by manipulating the substrate feed, $F(t)$ to the culture media. S is the sugar content, S_{max} a value at which sugar fermentation metabolism occurs, $O_{2,L}$ is the dissolved oxygen concentration, $O_{2,L,min}$ a minimal value at which the dissolved oxygen becomes limiting and triggers the fermentative metabolism, V_R is the cultivation volume and V_{max} the maximal bioreactor volume.

The solution of such kind of optimization problems is not trivial, since most cultivation systems have a complex, time-variant, and highly nonlinear nature. In case of applying model-supported approaches, estimation of model errors due to complex nature of such bioprocesses frequently require that the kinetic model used in optimization must be re-parameterized or online updated for effective control of the fermentation process. Here, heuristic rules based on human experiences to simplify the complex modelling have been applied.[7,8] It is in the nature of these methods that the construction of a database has to precede the optimization. Fed-batch cultivation represents a highly dynamic process for which theoretical approaches to adaptive process control are usually adjusted to profile control rather than to set-point control. This requires dynamic model adaptation ability. As conditions in a fed-batch process change with every time step, online parameter estimation and optimization are essential to online adaptive optimization.

Application of evolutionary algorithms and most concretely of GA's can be quite effective when searching for best possible solution on big parameter spaces. The terms evolutionary and genetic stand for a class of stochastic search methods that imitate the process of biological evolution.[9] Biological evolution itself can be regarded as an optimization process where simple reproductive elements are optimized by mutation, crossover, and selection. A GA operates on a population, P, which includes a parameter set of Q_i parameters. These could be, for instance, some sort reactions rates or feed flow rates on different time intervals. In a further step the parameter set is encoded using a binary code. The binary codings are called individuals or chromosomes. Each of this individuals consists of binary codes (or genes) of certain length and represents a particular solution to the problem. To produce better solutions to the problem the population, P, is evolved over generations. The input data for the GA

may contain, for instance, upper and lower values of the parameter set that has to be optimized. With these data, an initial population P_0 is created using a random number generator. The concrete example of parameter estimation in model fitting can also be regarded as an optimization problem. In this case, the objective is to find the optimal values of the model parameter set, which minimize a specific model/experiment error criterion. In a first step, the individuals of P_0 are assessed according to their fitness, which means that every generated individual is evaluated and the simulation results are compared to the experimental data, e. g. by calculating their squared sum of the residuals (*SSR*). The value of *SSR* then determines the rank of the fitness. The individual with the lowest *SSR* has the best fitness in the current population and, therefore, gets the highest rank. This initialization is followed by the generational loop. In this loop, operators like mutation of individuals, but also selection and crossover between individuals are applied. These operations lead to a temporary population, which individuals have to be assessed and ranked again. The reinsertion of the individuals from the temporary set with the highest fitness into the original population leads to a new population, P_1. The whole loop is performed according to the procedure described until the given number of generations is reached. In terms of model fitting, the solution of interest is the individual with the lowest squared sum of residuals ($I_{SSR}=\min(SSR)$).

Regarding the cultivation of baker's yeast, *Na* et al.[6] developed a genetic algorithm code (GA) to maximize cell production using the feed flow rate. The GA was designed to search for the optimal parameter set of discrete feed flow rates for different time intervals, $F_i(t)$. The strategy considered constant lengths for each time interval, i. e. $\Delta t = t_1 - t_0 = t_2 - t_1 = \ldots = t_n - t_{n-1}$. After a random initialization of individuals describing several feed flow rate sets, the GA compares the single criteria $X(t_{END}) \cdot V_R(t_{END})$ for every individual and, as described before, several generation loops are performed. Thus the GA can be used to estimate most convenient strategies for the state variables X, S, E, $C_{O2,L}$, V_R and their correlated model parameters μ, μ_{crit}, q_S, q_E and α. Based on this background, it is possible to determine an optimal feed rate, F_{opt}, which maximizes the performance function of Equation 5.

Results and discussions

Off-line optimization: Invertase production with *Saccharomyces cerevisiae*

Once the model of a bioprocess is validated and considered to be satisfactory, it is straightfor-ward to consider it for optimization. The present case was first examined by *Patkar* et al.[10] The aim of the work was the estimation of a feeding strategy that maximizes the invertase productivity of fed-batch cultivations. The micro-organism used was a recombinant *Saccharomyces cerevisiae* strain containing the plasmid pRB58, which codes for the enzyme invertase. The gene SUC2, responsible for the invertase expression, is repressed when the glucose concentration in the medium is high. This means, that the regulation and induction of invertase production is attained by regulation of the feed rate in order to maintain low glucose levels (< 1 g L^{-1}).

Glucose can be degraded in *S. cerevisiae* by two different routes from the bifurcation point at pyruvate. The proposed model considers essentially the existence of two main metabolic pathways from pyruvate, namely the respiratory and the fermentative fluxes. When the concentration of glucose in the media is low, the respiratory pathway predominates increasing the cell mass yield (high specific growth rate); in the fermentative pathway glucose is mainly converted to ethanol. The kinetics of these metabolic pathways were described mathematically by Monod-type formulation, including a logical rule for the course of glucose, that is, the amount of glucose entering the glycolytic chain of the cells. Here, a form of goal-oriented model approach was used i. e. the hypothesis and inferred kinetic model was explicitly conceived for latter process optimization. In the model the cells are considered as optimal strategists within the bioreactor environment, postulating the following:

1. The objective of the reactions within the cell is to maximize the cell growth rate.

2. The objective of the cell is to maximize the flux through the respiratory pathway.

3. The respiratory pathway gets saturated at glucose levels lower than those required for saturating glucose uptake rate, being also the saturated respiratory flux smaller than the saturated glucose uptake rate.

4. The rate of transcription of SUC2 gene is a function of the extra-cellular glucose concentration.

Based on these postulates, the expression used for the invertase production kinetics was assumed to follow a substrate limitation and inhibition of the Haldane-type. The state variables are described by a system of four differential equations (see Appendix for details). As performance function the maximization of the total invertase activity at the end of the fermentation time was defined:

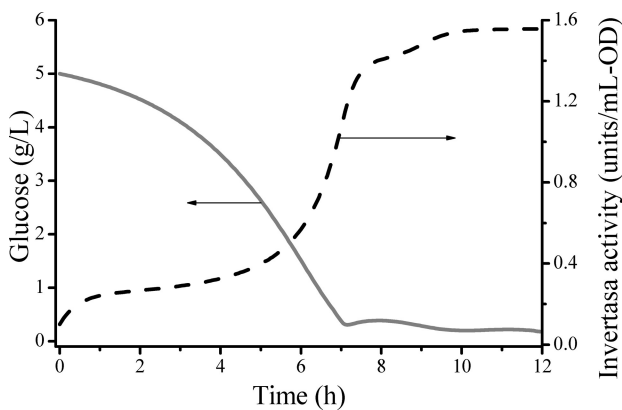$$P_F(t_{END}) = \max(P(t_{END}) \cdot X(t_{END}) \cdot V_R(t_{END})) \quad (9)$$

F i g . 2 – *Glucose concentration and invertase activity optimal profiles. The invertase activity is repressed while glucose is at concentration levels above 2g/L; however, the activity reaches a maximum value when glucose is depleted and maintained almost constant at concentration levels below 1 g L⁻¹.*

F i g . 3 – *Optimized profiles for growth and invertase expression rates. The evolutionary optimization strategy consists mainly in two parts: the biomass maximization (high growth and low invertase expression rates) and the product maximization (low growth and high invertase expression rates).*

To solve the problem several optimization procedures were applied: first order gradient method,[10] an evolutionary strategy[11] and the generalized reduced gradient method.[12,13] Fig. 2 displays the profiles for glucose and invertase activity obtained after applying an evolutionary strategy.[11] The method is supported by a GA and used an artificial neural network (ANN) to estimate the feeding strategy. After a random initialization of the weights and biases (ANN parameters), the GA performs a systematic variation of these parameters through mutation, recombination and selection for several generations according to the schema described previously. In the feed policy obtained three phases are clearly identified. In the first phase, the concentration of glucose is at high level keeping the activity of invertase at a low and almost constant level. The second phase takes place when glucose is gradually depleted, while the specific activity of invertase increases steadily until both reach pseudo-steady-states, the third phase, where they remain around their optimal minimal and maximal values, respectively.

The evolutionary strategy leaded to higher values of invertase activity in a shorter period of time compared to the first order gradient optimization method of *Patkar et al.*[10] and the generalized reduced gradient method of *Chaudhuri and Modak*[12]. This situation is most advantageous, since it allows the simultaneous reduction of the total fermentation time.

All three phases can also be interpreted considering the postulates used to formulate the model. The maximization of cell growth rate (postulate one) can only be fulfilled with the existence of a fermentative flux (in its non-saturated form), even when the objective of the cell is to maximize the
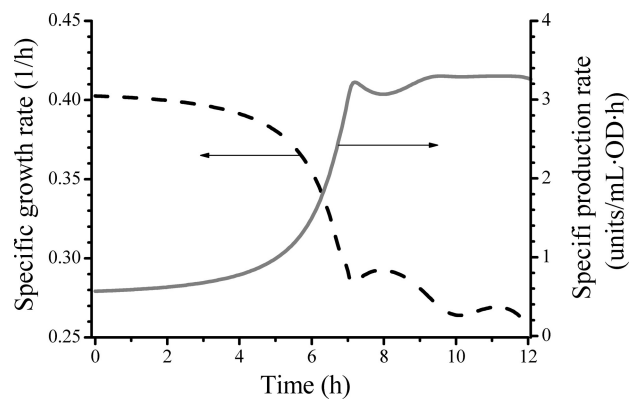
flux of glucose through the respiratory pathway (postulate two). *Patkar* et al.[10] explain the apparent dilemma arguing saturation in respiratory fluxes, as the reason why the cell does not exclusively use the respiratory pathway. Fig. 3 shows the specific growth and invertase expression rates, where is particularly clear that glucose is rapidly consumed in the first phase to maximize biomass as the main objective (postulate one). The second phase is that when glucose uptake rate is directed exclusively to the respiratory pathway (postulate three), situation that provokes a sharply increase in invertase activity (postulate four) and a decrease in the growth rate that, from hereon, is linear function only of the respiratory rate. The goal in the third phase is the maximization of invertase expression rate. This aim is attained at around the optimal glucose concentration when the respiratory pathway is not saturated and leads to the highest possible value of invertase expression rate, level that is reached and maintained in a pseudo-steady-state as depicted in Fig. 4.
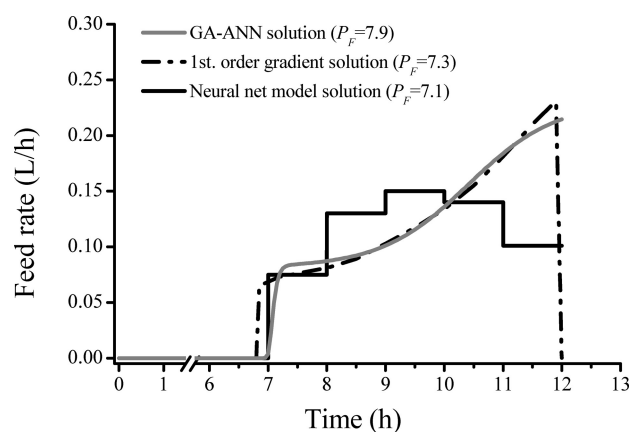

F i g . 4 – *Comparison of optimal feeding profiles for the invertase production. The performance function (P$_F$) is given in units of activity at time, t$_{END}$ =12 h*

The evolutionary feeding strategy yielded 7.9 units of activity for the invertase as optimal value. That is an improvement of about 8 % when compared with the result obtained by the gradient optimization solution[10] and an improvement of about 11 % compared to the generalized reduced gradient method.[12] The optimal feeding strategies are compared in Fig. 4. Two strategies, the evolutionary and the first order gradient method, display obvious similarities with a gradual increment of feed following a S-shaped profile. An important additional feature is also distinguishable: both strategies delivered different induction times for the expression of invertase, being the time for the gradient method shorter than for the evolutionary solution. This is the reason why the later reaches the maximal of invertase expression rate earlier.

Online performance monitoring: Production of the recombinant protein GAL80 with *Kluyveromyces lactis*

Fed-batch cultivations of a recombinant strain of the *Kluyveromyces lactis* yeast were run under specific growth rate-controlled conditions. The optimization goal was to maximize the biomass production per total fermentation time. The strain *K. lactis* RUL 1888 D80ZR-pEAHG80 was aerobically grown on glucose as only carbon source. The product, the recombinant protein GAL80,[14] is merged with a selective histidine affinity tag (HIS-TAG) and was constitutively expressed by an ADG promoter (ADG: recombinant adenovirus expressing green fluorescence protein). Since the product is biomass associated, the cultivation was run under fed-batch modus, where a feeding profile was adjusted to control the specific growth rate calculated from:

$$F(t) = \frac{\mu_{SET}}{Y_{X/S}} \cdot X \cdot \frac{W}{(S_F - S)} \qquad (10)$$

The variables biomass concentration (X), glucose concentration (S), and culture broth weight (W) were estimated using the method proposed by *Claes* and *Van Impe*.[15] $Y_{X/S}$ and S_F represent the yield coefficient and the glucose concentration in the fresh medium stream, respectively. The mathematical model of the cultivation process[16] was cyclically adjusted once a biomass concentration measurement was available. Based on it, the feeding profile was adjusted accordingly. The best results were achieved by controlling the specific growth rate in order to decrease the intracellular metabolic overflow during the cultivation. In the case of the product concentration P, a linear correlation with the biomass concentration was observed: $P = \alpha X$. The optimization task considered was the maximi-

zation of the amount of biomass (X) per cultivation time (t_{END}).

However, this sequential model readjustment, prediction and feed set-point control strategy still does not guarantee that the optimum may be reached; i. e. the estimated feed rate is actually also an inherent function of the success of the iterative readjustment procedure, which does not unambiguously considers the possible process constraints. Especially in the case of correlated process variables, which are not explicitly took into account by the control algorithm and which have to be kept within certain ranges, independent observers must be considered to supervise the control and process performance. To overcome this difficulty, an ANN-based soft-sensor was proposed[17] to independently monitor the process performance. This alternative monitoring system is based exclusively on real-time available process measurements. The specific growth rate is formulated via a feed forward neural network with a single hidden layer and a single hidden node, which are complemented with a lag time term. The observer integrates online variables, like the oxygen uptake rate (OUR), enabling the instantaneous analysis of the influence that several process constraints may have on the final process performance.

The core of the soft-sensor is based on the online estimation of the specific growth rate, which is incorporated into the mass balances equations. For its estimation two modified online measurements from the exhaust gas analysis were used: the oxygen uptake rate and the carbon dioxide evolution rate (CER). When these variables are expressed as a function of the total biomass present in the bioreactor, two new useful correlations can be defined and used as inputs to the ANN: the biomass specific oxygen consumption rate, q_{XO2} and the biomass specific carbon dioxide evolution rate, q_{XCO2}. The biomass specific oxygen uptake rate, q_{XO2} is given by:

$$q_{XO2} = \frac{OUR}{X \cdot W} \qquad (11)$$

where X is the at-line estimated biomass concentration (g Biomass/kg culture broth) and W (kg) is the online measured broth weight.

For the case of the biomass specific carbon dioxide evolution rate, q_{XCO2} it is estimated from:

$$q_{XCO2} = \frac{CER}{X \cdot W} \qquad (12)$$

Additionally to the exhaust gas measurements, the online electrode signal for dissolved oxygen,

pO_2 and the at-line measurements of the culture broth optical density (*OD*) were taken as inputs to the ANN soft-sensor. Using one at-line estimated variable, the biomass concentration and three online measured variables, q_{XO2}, q_{XCO2} and pO_2 the tuning of the sensor (training of the ANN parameters) was also made online by means of a chemotaxis algorithm[18]. It is worth to note that in previous experiments it was found that the q_{XO2} was closely related with the accurate estimation of the biomass concentration. However, when used as single input for the soft sensor it provided a poor description of the substrate concentration. In contrast, the q_{XCO2} provided a high precision for the estimation of the substrate concentration. Nevertheless, it delivered a quite unsatisfactory estimation of the biomass concentration when it was used as single input for the soft sensor. For this reason, both variables must be used always in conjunction and not separately, since even when the removal of one of these from the soft sensor would improve the accuracy of a single variable, it would do it at expenses of worsening the estimation of the other.

To optimize the process, the performance function considered, P_F, was given by:

$$P_F(t_{END}) = \max\left(\frac{X(t_{END})}{t_{END}} - f_C(\mu, F, OUR)\right) \quad (13)$$

where f_C is a general penalization function containing the process constraints:

$$f_C(\mu, F, OUR) = f_{C1}(\mu) + f_{C2}(F) + f_{C3}(OUR) \quad (14)$$

$f_{C1}(\mu)$ is the constraint function that penalizes deviations from the specific growth rate set point, μ_{SET} at which the fermentation was controlled. $f_{C2}(F)$ and $f_{C3}(OUR)$ are functions that penalize, respectively, any situation in which the physical limits for the feeding function (pump) or for the oxygen uptake rate (*OUR*) may be exceeded. The single constraint functions f_{C1}, f_{C2}, f_{C3} are given by:

$$f_{C1}(\mu) = K_1 \cdot \sum (\mu(t) - 0.13)^2 \quad (15)$$
$$\text{if } \mu \neq 0.13 \text{ h}^{-1}$$

$$f_{C2}(F) = K_2 \cdot \sum (F(t) - 0.8)^2 \quad (16)$$
$$\text{if } F(t) > 0.8 \text{ kg h}^{-1}$$

$$f_{C3}(OUR) = K_3 \cdot \sum (OUR(t) - 12.0)^2 \quad (17)$$
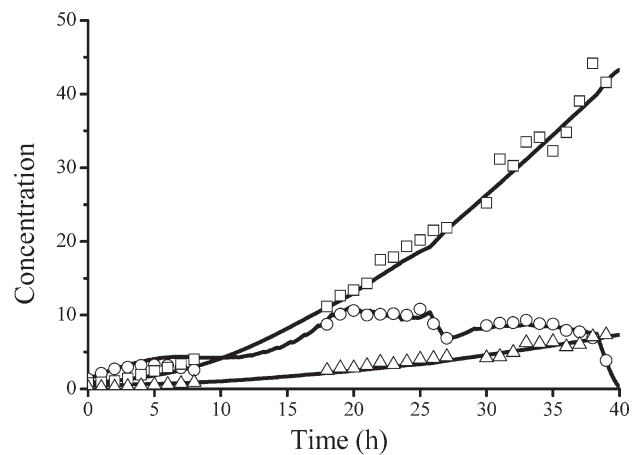$$\text{if } OUR(t) > 12.0 \text{ g kg}_X^{-1} \text{ h}^{-1}$$

F i g . 5 – *Time trajectories of the concentration of biomass (□), glucose (○) and recombinant protein (△) concentration. Lines account for simulated values while symbols represent the measurements*

The constants K_1, K_2, K_3 are constraint parameters that were determined empirically to manipulate the influence of the single penalty function on the optimization profit function. Fig. 5 depicts the modelled and measured state variables biomass, glucose and recombinant protein concentration. It can be stated, that the ANN-based soft sensor was able to properly describe the variables with high accuracy, even in the absence of measured data for a period of 9 h (from 8 h to 17 h).

The specific growth rate obtained is presented in Fig. 6. Even when the modelling can be considered satisfactory, the optimal set-point for the specific growth rate, $\mu = 0.13$ h^{-1} was only partially fulfilled. The reason for this mismatch can be only speculated, but since in the process an extremely sensitive recombinant strain was cultivated, a lost of the plasmid or even the emergence of
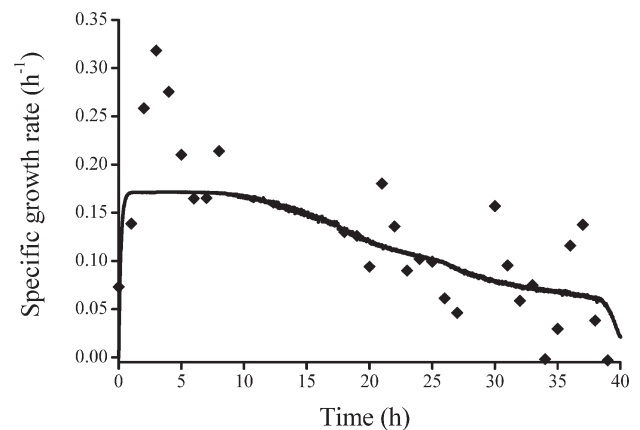

F i g . 6 – *Time trajectories of the specific growth rate (♦) calculated with the ANN-based soft-sensor. Lines account for simulated values while symbols represent the estimation derived from biomass measurements.*
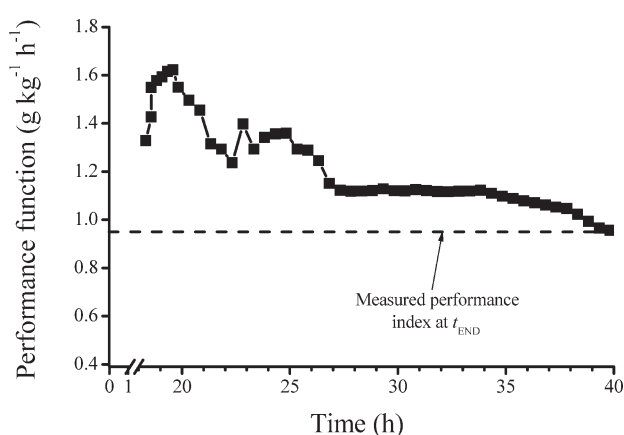
F i g . 7 – *Time trajectory of the expected performance func-*
tion. The line represents the measured value at the
end of the fermentation (t_{END}= 40 h).

heterogeneity in the yeast population, and there-
fore the apparition of a more stable mutant with
different kinetic characteristics, can not be dis-
carded.

Fig. 7 compares the expected values of the
profit function using the ANN-based soft-sensor
with the final measured total protein content per
fermentation time. After a period of adaptation of
about 20 h, the trajectory of the simulation for the
expected performance index showed an almost con-
stant trend in its prediction. The prediction also
showed good agreement with the final measurement
of the performance index obtained for the cultiva-
tion.

As discussed before, the optimization was con-
strained by controlling the specific growth rate at
the set point, $\mu = 0.13$ h^{-1} (see Equation 15). Fig. 8
presents the different time trajectories for the penal-

ties of the performance index, described by Equa-
tions 15 to 17. The penalties for maximal *OUR* and
feeding rate approached to zero during the course
of the cultivation. Contrastingly, the penalty func-
tion of the controlled specific growth rate did not
vanish and actually increased in the final part of the
fermentation. It is worth to remark, that an apparent
relationship between the estimated performance in-
dex and the vanishing of the feeding constraint was
observed.

Conclusions

Three different case studies addressing system
performance and estimation of optimal cultivation
policies of yeast cultivation systems were analyzed.
After the implementation of a detailed first prin-
ciple model approach to the cultivation of backer's
yeast in bubble columns, an optimization task-ori-
ented reduction of the complexity was carried out.
Using this simplified framework, the model-sup-
ported optimization of an invertase producing *S.
cerevisiae* cultivation was demonstrated. Several
techniques, including an evolutionary strategy, were
applied to maximize the invertase production and
it could be shown, that the evolutionary strategy
was superior even to well-established optimization
techniques, like the first order gradient method.
Finally, since the estimation of any optimal pro-
ductivity parameter is sometimes a conflictive prob-
lem of economic nature, the coupling of specific
process constrains to the choice of the most conve-
nient set of variables was exemplified on the opti-
mal production of the protein GAL80 with the yeast
K. lactis. The best results were achieved by keeping
the specific growth rate at a stable set point in order
to decrease the intracellular metabolic overflow
during the cultivation. However, the control of the
specific growth rate was based on an online esti-
mated feed rate, which is actually an intrinsic func-
tion of the iterative correction of the process model
parameters, and which did not unambiguously con-
sider the process constraints. Especially in the case
of correlated process variables that have to be kept
within certain ranges and which are not considered
explicitly in the control algorithm, independent
observers must be considered to supervise the con-
trol and process performance. For this purpose, an
independent online soft-sensor was implemented,
which allowed the monitoring of the performance
index and of the process constrains. The present
work illustrated on three different yeast systems the
use of model-supported designs to obtain suitable
policies for efficient experiments, which can be re-
garded as a most effective tool to solve optimiza-
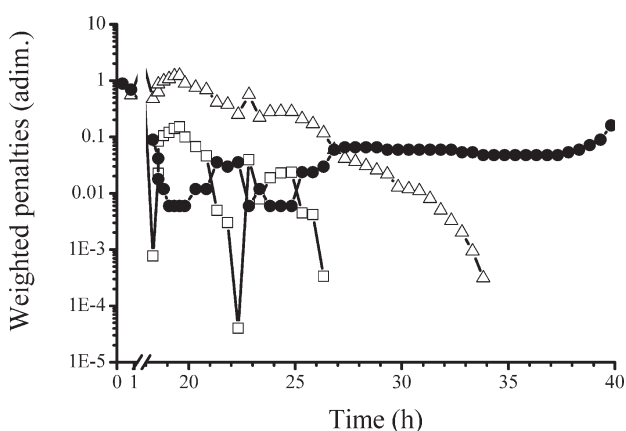tion problems.

F i g . 8 – *Time trajectories of the performance penalties. The*
penalty accounting for the maximal OUR (□) disappeared at
about 26 h of cultivation time, while that of maximal feeding
rate (△) at about 34 h. The penalty for deviation from the spe-
cific growth rate set point (●) does not vanish during the
whole process.

APPENDIX

Invertase production model[10].

$$\frac{d(XV)}{dt} = \mu XV,$$

$$\frac{d(SV)}{dt} = FS_F - R_t XV,$$

$$\frac{d(PXV)}{dt} = (\pi + k_d P)XV,$$

$$\frac{dV}{dt} = F,$$

where,

$$R_r = \frac{5.55S}{0.05 + S} \qquad R_t = \max\left(\frac{1.25S}{0.95 + S}, R_r\right),$$

$$\pi = \frac{6.25S}{0.1 + S + 2.0S^2} \qquad R_f = R_t - R_r \qquad \mu = (R_r Y_{sr} + R_f Y_{sf})$$

$$Y_{sr} = 0.6 \qquad Y_{xfn} = 0.15 \qquad k_d = 1.85$$

The performance function:

$$P_F = P(t_{END}) \cdot X(t_{END}) \cdot V(t_{END}),$$

subjected to the constrains,

$$V_{max} = 1.2L; \qquad 0 \le F \le 0.6 \ (\text{L/h})$$

References

1. *FDA*, PAT–A framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance (http://www.fda.gov/cder/guidance), 2004.
2. *Sonnleitner, B., Käppeli, O.,* Biotech. Bioeng. **28** (1986) 927-937.
3. *Schallenberg, J., Enss, J. H., Hempel, D. C.*, Chem. Eng. Sci. **60** (2005) 6027-6033.
4. *Zhang, T., Zhao, B., Wang, J.*, Chem. Eng. J. **119** (2006) 19-26.
5. *Warsi, Z. U. A.*, Fluid dynamics. Theoretical and computational approaches. CRC Press, Inc., Boca Raton, Florida, 1993.
6. *Simutis, R., Havlik, I., Lübbert, A.*, J. Biotechnol. **23** (1993) 203-215.
7. *Na, J.-G., Chang, Y. K., Chung, B. H., Lim, H. C.*, Bioproc. Biosys. Eng. **24** (2002) 299-308.
8. *Simutis, R., Havlik, I., Schneider, F., Dors, M., Lübbert, A.*, IFAC Computer Applications in Biotechnology (1995) 59-65.
9. *Goldberg, D. E.,* Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, Massachusetts, 1989.
10. *Patkar, A., Seo, J. H., Lim, H. C.*, Biotech. Bioeng. **41** (1993) 1066-1074.
11. *Franco-Lara, E., Weuster-Botz, D.*, Bioproc. Biosys. Eng. **27** (2005) 255-262.
12. *Chaudhuri, B., Modak, J. M.*, Bioproc. Eng. **19** (1998) 71-79.
13. *Reklatis, G. V., Ravindran, A., Ragsell, K.,* Engineering Optimization. Wiley, New York 1983.
14. *Zenke, F. T., Kapp, L., Breunig, K. D.*, Biol. Chem. **380** (1999) 419-430.
15. *Claes, J. E., Van Impe, J.F.*, Bioproc. Biosys. Eng. **21** (1999) 389-395.
16. *Volk, N., Franco-Lara, E., Galvanauskas, V., Lübbert, A.*, Model Supported Optimization of Fed-Batch Fermentations for Recombinant Protein Production, in *Merten, O.-W., Mattanovich, D., Lang, C., Larsson, G., Neubauer, P., Porro, D., Postma, P., Teixeira de Mattos, J., Cole J.* (eds.) Recombinant Protein Production with Prokaryotic and Eukaryotic Cells. A Comparative View on Host Physiology, pp 371-381, Kluwer Academic Publishers, 2001.
17. *Franco-Lara, E.*, PhD Dissertation, MLU Halle-Wittenberg, 2002. urn:nbn:de:gbv:3-000004115.
18. *Simutis, R. and Lübbert, A*, J. Biotechnol. **52** (1997) 245-256.