Bioautomation, 2004, vol. 1, pp. 30-41

A Genetic Algorithms Based Approach for Identification of *Escherichia coli* Fed-batch Fermentation

Olympia Roeva, Tania Pencheva, Bernd Hitzmann*, Stoyan Tzonkov

Centre of Biomedical Engineering, Bulgarian Academy of Sciences
105, Acad. G. Bonchev Str., Sofia 1113, Bulgaria
E-mail:olympia@clbme.bas.bg,
tania.pencheva@clbme.bas.bg,
tzonkov@clbme.bas.bg

**Institut für Technische Chemie, Universität Hannover,*
Callinstr. 3, 30167 Hannover, Germany
E-mail: *hitzmann@iftc.uni-hannover.de*

Abstract: *This paper presents the use of genetic algorithms for identification of Escherichia coli fed-batch fermentation process. Genetic algorithms are a directed random search technique, based on the mechanics of natural selection and natural genetics, which can find the global optimal solution in complex multidimensional search space. The dynamic behavior of considered process has known nonlinear structure, described with a system of deterministic nonlinear differential equations according to the mass balance. The parameters of the model are estimated using genetic algorithms. Simulation examples for demonstration of the effectiveness and robustness of the proposed identification scheme are included. As a result, the model accurately predicts the process of cultivation of E. coli.*

Keywords: *Genetic algorithms, Modelling, Parameters estimation, Fed-batch fermentation process, Escherichia coli*

Introduction

Fermentation process are known to be very complex and their modelling may be a rather time consuming. However, it is neither necessary nor desirable to construct comprehensive mechanistic process models that can describe the system in all possible situations with a high accuracy. In order to optimize a real biotechnical production process, the model must be regarded as a step to reach more easily the final aim. The model must describe those aspects of the process that significantly affect the process performance.

Many mathematical models have been proposed for fermentations but just few have been used to optimize industrial fermentations [1, 3, 5, 8]. In this paper a mathematical model of a typical fed-batch fermentation of *Escherichia coli* is proposed. The model describes high yield of product (biomass), substrate consumption, acetate production, as well as the variations of dissolved oxygen and carbon dioxide concentrations. The parameters of the model are estimated using genetic algorithms, where the criterion is the error between the true model output and the identified model output. Recently, genetic algorithms (GAs) have been extensively used in solving many optimization-searching problems [6]. Compared with conventional optimization methods, GAs do not assume that the search space is differentiable or continuous. Also GAs do not require linearity in the parameters which is needed in iterative searching optimization techniques. These properties of GAs make them suitable for the parameter identification of fed-batch fermentation processes.

The appropriate implementation of GAs includes the following three aspects:

- definition of the objective function;
- definition and implementation of the genetic representation;
- definition and implementation of the genetic operators.

The simulation of the selected model has been accomplished with the help of *SIMULINK* under *MATLAB* environment as a modern and improved way for process simulation and possible control. The parametric identification of the process has been performed with the *SHEFFIELD MATLAB GENETIC ALGORITHM TOOLBOX* [4]. This is a novel instrument for implementing genetic algorithm methods as script files that could be changed according to the problem requirements.

Description of the *E. coli* fed-batch fermentation process

Cultivation of recombinant micro-organisms, e.g. *Escherichia coli*, in many cases is the only economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors. *Escherichia coli* is still the most important host organism for recombinant protein production. To maximize the volumetric productivities of bacterial cultures it is important to grow *E. coli* to high cell concentration. The use of fed-

batch cultivation in the fermentation industry takes advantage of the fact that residual substrate concentration may by maintained at a very low level in such a system. One of the most frequent used substrate for the cultivation of microorganisms is glucose.

The cultivation of *E. coli MC4110* was performed in the *Institut für Technische Chemie, Universität Hannover*. The cultivation was carried out in a 2l bioreactor (Bioengineering, Switzerland), using a mineral medium [2]. Before inoculation a glucose concentration of 2.5g/l was established in the medium. The concentration of glucose in feeding solution was 100 g/l. Initial liquid volume was 1350 ml, pH was controlled at 6.8 and the temperature was kept constant at 35^0C. The aeration rate was kept at 275 l/h air, stirrer speed at start was 900 rpm, after 11h the stirrer speed was increased in steps of 100 rpm and at end was 1500 rpm. Oxygen was controlled around 35%.

Off-line analysis. For off-line glucose measurements as well as biomass and acetate concentration determination samples of about 10 ml were taken roughly every hour. Off-line measurements were performed by using the Yellow Springs Analyser (Yellow Springs Instruments, USA).

On-line analysis. For on-line glucose determination a flow injection analysis (FIA) system was employed using two pumps (ACCU FM40, SciLog, USA) for a continuous sample and carrier flow rate. To reduce the measurement noise the continuous-discrete extended Kalman filter were used [2].

Mathematical model of the process

The model of fed-batch fermentation of *E. coli* is based on the following assumptions:

- The bioreactor is completely mixed.
- The substrate (glucose) is consumed mainly oxidatively and its consumption can be described by Monod kinetics.
- The acetate production rate is assumed to be directly proportional to the biomass formation.

- Variation in the growth rate, acetate production and substrate consumption do not significantly change the elemental composition of biomass, thus balanced growth conditions are only assumed.

The rates of cell growth, glucose consumption, acetate formation, as well as the variations of dissolved oxygen and carbon dioxide concentrations in *Escherichia coli* fermentation are commonly described as follows according to the mass balance:

$$\frac{dX}{dt} = \mu X - \frac{F}{V} X \tag{1}$$

$$\frac{dS}{dt} = -q_S X + \frac{F}{V}(S_{in} - S) \tag{2}$$

$$\frac{dA}{dt} = q_A X - \frac{F}{V} A \tag{3}$$

$$\frac{dDO}{dt} = -q_{O_2} X + k_{la}^{O_2}(DO^* - DO) - \frac{F}{V} DO \tag{4}$$

$$\frac{dCO_2}{dt} = q_{CO_2} X + k_{la}^{CO_2}(CO_2^* - CO_2) - \frac{F}{V} CO_2 \tag{5}$$

$$\frac{dV}{dt} = F \quad , \tag{6}$$

where:

$$\mu = \mu_{\max} \frac{S}{k_s + S}, \; q_S = \frac{1}{Y_{S/X}} \mu, \; q_A = \frac{1}{Y_{A/X}} \mu, \; q_{O_2} = \frac{1}{Y_{O_2/X}} \mu, q_{CO_2} = \frac{1}{Y_{CO_2/X}} \mu . \tag{7}$$

where: X is the concentration of biomass, [g/l];

S - concentration of substrate (glucose), [g/l];

A - concentration of acetate, [g/l];

DO - concentration of dissolved oxygen, [%];

CO_2 - concentration of carbon dioxide, [%];

F - feeding rate, [l/h];

V - bioreactor volume, [l];

S_{in} - concentration of the feeding solution, [g/l];

μ - specific growth rate, [h^{-1}];

q_S - specific rate of substrate utilization, [h^{-1}];

q_A - specific rate of product formation, [h^{-1}];

q_{O_2} - specific rate of oxygen consumption, [h^{-1}];

q_{CO_2} - specific rate of carbon dioxide production, [h^{-1}];

μ_{max} - maximum growth rate, [h^{-1}];

k_S - the substrate concentration at which half a maximum specific growth rate is obtained (μ=0.5μ_{max}), [g/l];

$Y_{X/S}, Y_{A/S}, Y_{O_2/S}, Y_{CO_2/S}$ - yield coefficients, [gg^{-1}];

$k_{la}^{O_2}$ - volumetric oxygen transfer coefficient, [h^{-1}];

$k_{la}^{CO_2}$ - volumetric carbon dioxide transfer coefficient, [h^{-1}].

The specific growth rate μ is generally found to be a function of three parameters: the concentration of limiting substrate S, the maximum growth rate μ_{max} and the substrate-specific constant k_S. The value for k_S is generally very low. Maximum specific growth rates are of considerable industrial importance. The specific growth rates vary between 0.09-0.61 h^{-1}. According to Monod kinetics, residual substrate should decrease as dilution rate decreases resulting in an increase in the cell concentration. Over most of the range of μ which will operate in fed-batch culture, S_{in} will be much higher than k_S, so that, for all practical purpose, the change in residual substrate concentration would be extremely small and may be considered as zero [9].

Parameter identification using genetic algorithms

Real parameter optimization of simulation models has especially become a research field of great interests in recent years. Such problems have widespread application. Nevertheless, this task still represents a very difficult problem. In this case only direct optimization strategies can be applied, because they exclusively use information about values of the goal function. Additional information about the goal function like gradients, etc., which may be used to accelerate the optimization process, is not available. Since an evolution of a goal for one string is provided by one simulation run, proceeding of an optimization algorithm may require a lot of computational time. Nowadays the most common direct methods used for global optimization are evolutionary algorithms such as genetic algorithms.

Genetic algorithms are a directed random search technique, which can find the global optimal solution in complex multidimensional search space. GAs are first proposed by Holland [7] and have been applied successfully to many engineering and optimization problems [1, 3, 5, 8, 10]. GAs employ different genetic operators [6] to manipulate individuals in a population of solutions over several generations to improve gradually their fitness.

Experimental data of cultivation of *Escherichia coli* obtained in *Institut für Technische Chemie, Universität Hannover* are used. On the basis on feeding rate data (Fig. 1), on-line data for dissolved oxygen and carbon dioxide, as well as off-line measurements of biomass, substrate (glucose) and acetate, a parameter identification of proposed mathematical model (1)-(7) is carried out.

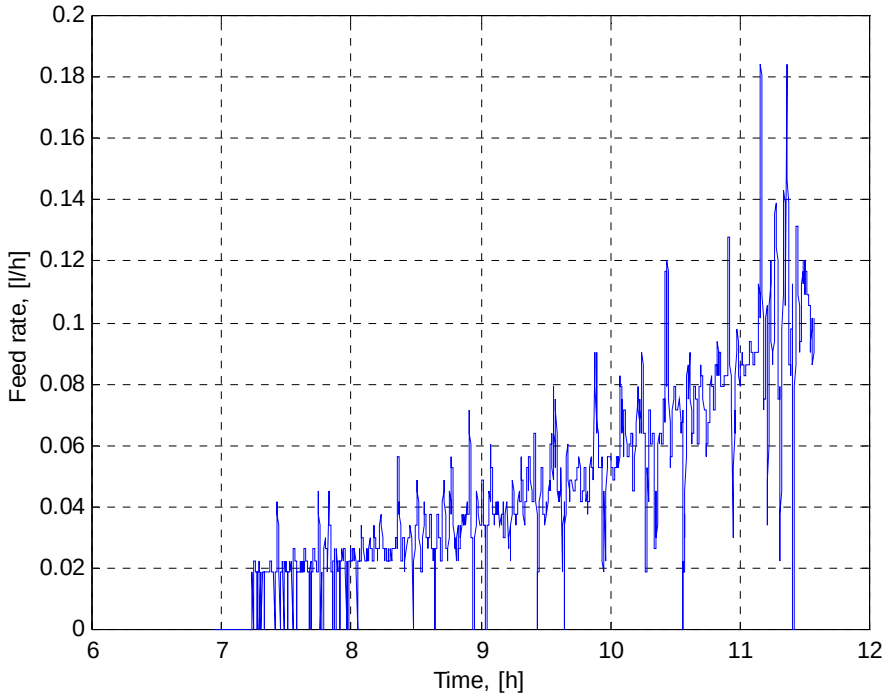

Fig. 1 Feed rate in a fed-batch culture of *Escherichia coli*

Under *Matlab* 5.3 environment a *Simulink* model of the *Escherichia coli* fed-batch fermentation has been developed. The model describes the differential equations (1)-(7), parameters and initial values in S-functions. In order to optimize the *Simulink* model, a script containing the necessary instructions for the *Genetic Algorithm Toolbox* [4] has been developed.

To implement the genetic algorithms, the model's parameters have to be presented in terms of chromosomes. Decimal numbers for the parameter values have been used to represent this principle instead of a binary profile. Each chromosome corresponds to one different objective function value. In the literature the most often used optimization criterion is defined as a modelling error, i.e. the mean square deviation between the model output and the corresponding data obtained during the fermentation. The optimization criteria are presented as follows:

$$J_Y = \sum (Y - Y^*)^2 \rightarrow min \tag{8}$$

where Y is experimental and Y^* is simulated values, respectively for X, S, A, DO and CO_2.

Initial tests with a large number of individuals and generations, 1500 individuals and 250 generations have been done. The initial values of other algorithm's parameters and functions are [3]:

- Generation gap how many new individuals are created in every generation) – 0.8;
- Precision of binary representation – 5;
- Selection function - Roulette Wheel Selection;
- Recombination function - Crossover Double Point;
- Crossover rate – 0.8;
- Mutation rate – 1/160.

Several optimizations have been made when the initial parameters related to the genetic algorithm have been changed. After different runs some of parameters have been changed as follows:

- Number of individuals – 100;
- Number of generations – 100;
- Generation gap – 0.97;
- Precision of binary representation – 20;
- Crossover rate – 0.7.

The estimated values of parameters are shown in Table 1.

Table1 Parameters values

Parameter	Value	Parameter	Value
μ_{max}	0.46 h^{-1}	$Y_{pO_2/X}$	1.25 gg^{-1}
k_S	0.012 g/l	$Y_{CO_2/X}$	0.0365 gg^{-1}
$Y_{S/X}$	0.4975 gg^{-1}	$k_{la}^{O_2}$	290.06 h^{-1}
$Y_{A/X}$	66.67 gg^{-1}	$k_{la}^{CO_2}$	68.57 h^{-1}
$Y_{O_2/X}$	23.31 gg^{-1}		

Values of criteria are:

J_X = 1.1532, J_S = 5.3860, J_A = 3.2911e-004, J_{DO} = 0.1493, J_{CO2} = 0.0778.

Both the experimental fermentation trajectories (*) and the simulated ones (—) for the fed-batch culture of *Escherichia coli* are presented in Fig. 2 - 6, respectively - Fig.2 - for biomass, Fig.3 - for substrate, Fig.4 – for acetate, Fig.5 – for dissolved oxygen and Fig.6 – for carbon dioxide.

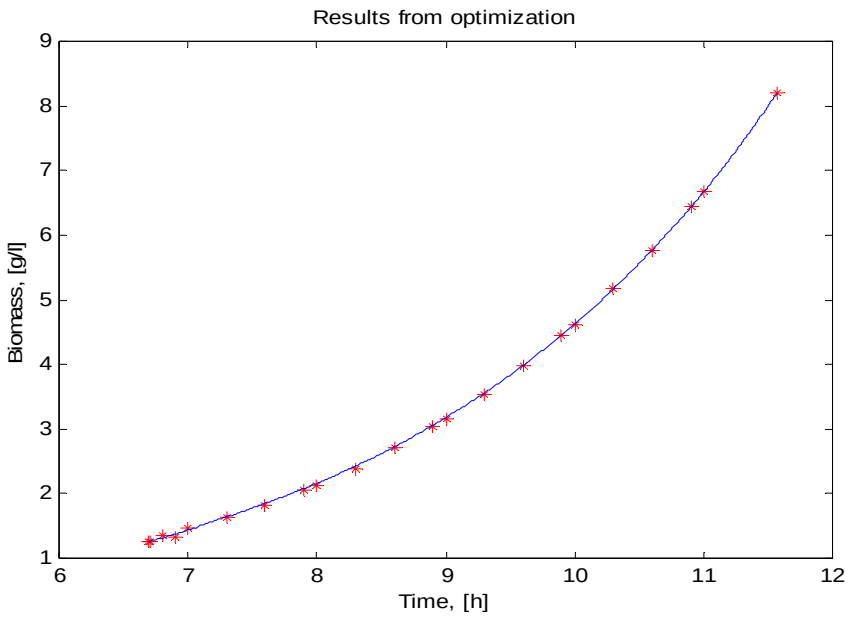

Fig. 2

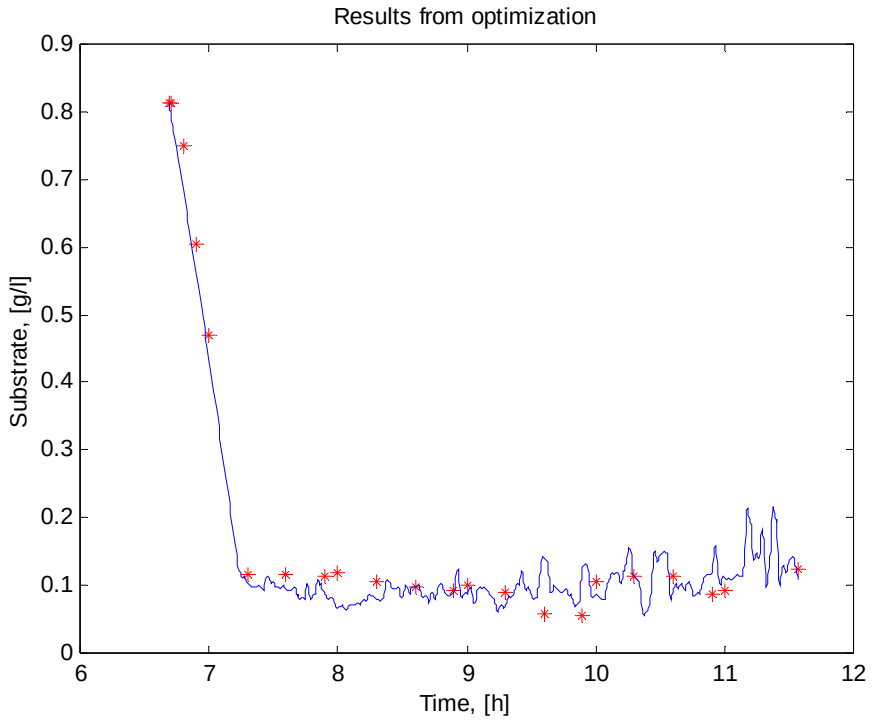
Results from optimization
Substrate, [g/l]
Time, [h]
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
6
7
8
9
10
11
12

Fig. 3

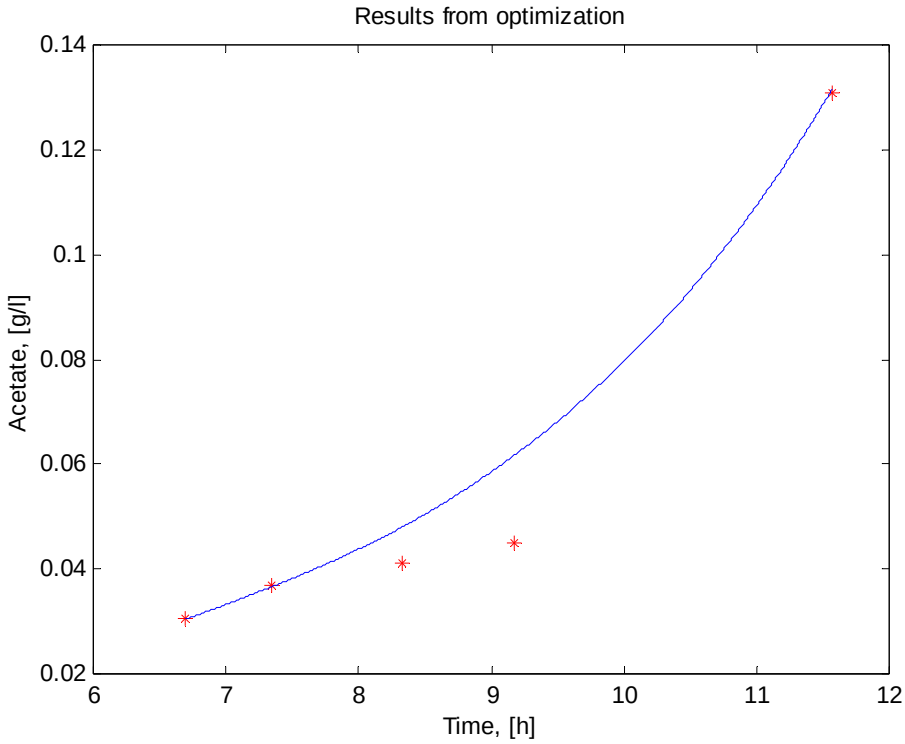
Results from optimization
Acetate, [g/l]
Time, [h]
0.02
0.04
0.06
0.08
0.1
0.12
0.14
6
7
8
9
10
11
12

Fig. 4

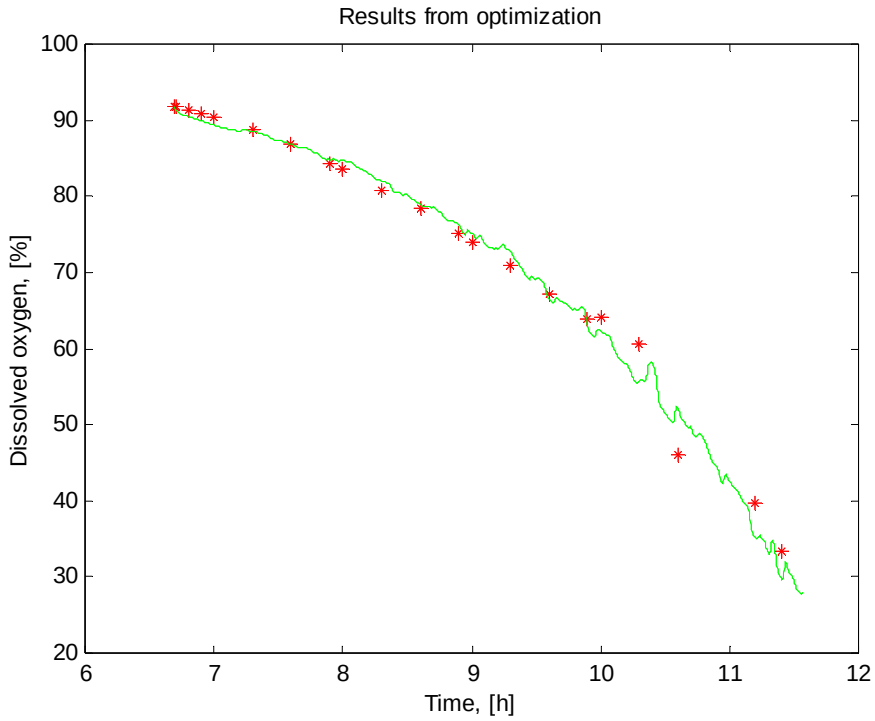
Results from optimization
Dissolved oxygen, [%]
Time, [h]
100
90
80
70
60
50
40
30
20
6
7
8
9
10
11
12

Fig. 5

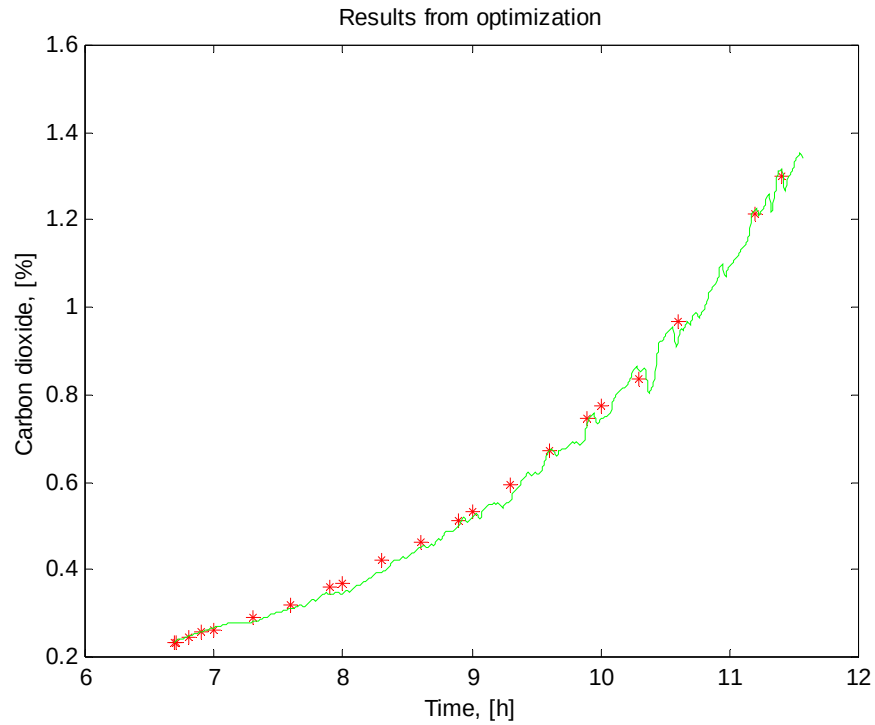
Results from optimization
Carbon dioxide, [%]
Time, [h]
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
6
7
8
9
10
11
12

Fig. 6

The figures show that the model predicts successfully the variation of glucose consumption, biomass concentration, acetate formation, oxygen consumption and carbon dioxide production during the fed-batch fermentation of *E. coli*.

Conclusions

In this paper a simple genetic algorithm-based scheme for the parameter identification of nonlinear system, namely *E. coli* fed-batch fermentation process, is proposed. It is accomplished through the formulation of the identification problem as an optimization problem and the application of GAs in order to estimate the unknown parameters from input-output data.

Simulation results reveal that accurate and consistent results can be obtained. It is evident that the model predicts successfully the variation of glucose consumption, biomass concentration, acetate formation, dissolved oxygen consumption and carbon dioxide production. The use of genetic algorithms for the identification of fed-batch fermentation process demonstrates the effectiveness and robustness of the proposed identification scheme.

Acknowledgements

This work is partially supported from DFG Project №113/135, NSF-TN-1314/2003 and Project №03-1046/2003 with Thracian University- Stara Zagora.

References

1. Al-Duwaish H. N. (2000). A Genetic Approach to the Identification of Linear Dynamical Systems with Static Nonlinearities. *International Journal of Systems Science*, 31, No. 3, pp. 307-313.
2. Arndt M., B. Hitzmann (2001). Feed Forward/feedback Control of Glucose Concentration during Cultivation of *Escherichia coli. Proceedings of the 8th IFAC Int. Conf. on Comp. Appl. in Biotechn., Quebec City, Canada,* pp. 425-429.
3. Carrillo-Ureta G. E., P. D. Roberts, V. M. Becerra (2001). Genetic Algorithms for Optimal Control of Beer Fermentation. *Proceedings of the 2001 IEEE International Symposium on Intelligent Control, Mexico City, Mexico*, pp. 391-396.

4. Chipperfield, A., P. J. Fleming, H. Pohlheim, C. M. Fonseca (1994). Genetic Algorithm Toolbox for Use with MATLAB. User's Guide. Version 1.2. Dept. of Automatic Control and System Engineering, University of Sheffield, U. K.
5. Cockshott A. R., I. D. L. Bogle (1999). Modelling the Effects of Glucose Feeding on a Recombinant *E. coli* Fermentation. *Bioprocess Engineering*, v. 20, pp. 83-90.
6. Goldberg D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Weslcy Publishing Company, Reading, Massachusetts.
7. Holland J. (1975). Adaptation in Natural and Artificial Systems. MIT Press.
8. Ranganath M., S. Renganathan, C. Gokulnath (1999). Identification of Bioprocesses Using Genetic algorithm. *Bioprocess Engineering*, v. 21, pp. 123-127.
9. Standbary P.F., A. Whitaker (1984). Principles of Fermentation Technology. Pergamon Press.
10. Vassileva, S., I. Sekaj, T. Švantnerová (1999). GA applications in parameter optimization of complex processes. *Proceeding of 12th International Conference "Process Control'99"*, May 31 – June 3, Slovak Republic.