Modelling of a Fed-batch Culture Applying Simulated Annealing

Olympia Roeva* and Tanya Trenkova[†]
* Department of Bioinformatics and Mathematical Modelling
Institute of Biophysics and Biomedical Engineering, Sofia, Bulgaria
Email: olympia@biomed.bas.bg
[†] Department of Geography
National Institute of Geophysics, Geodesy and Geography, Sofia, Bulgaria
Email: ttrenkova@gmail.com

Abstract—In this paper the metaheuristic Simulated Annealing (SA) is applied for parameter identification of non-linear model of cultivation process. SA algorithm is a stochastic relaxation technique, using the Metropolis algorithm based on the Boltzmann distribution in statistical mechanics, for solving nonconvex optimization problems. A real *E. coli* MC4110 fed-batch cultivation process is considered. The mathematical model is presented by a system of five ordinary differential equations, describing the regarded cultivation process variables - biomass, substrate, acetate, dissolved oxygen and bioreactor volume increasing. The obtained criteria values show that the developed model is adequate and has a high degree of accuracy. The presented results are a confirmation of successful application of the SA algorithm and of the choice of SA algorithm parameters.

Keywords-metaheuristics; simulated annealing; optimization; E. coli; cultivation process

I. INTRODUCTION

Classical biotechnology is the science of production of human-useful products under controlled conditions, applying biological agents micro-organisms, plant or animal cells, their exo- and endo-products, e.g. enzymes, etc. [7]. Cultivation of recombinant micro-organisms e.g. *E. coli*, in many cases is the only economical way to produce pharmaceutical biochemicals such as interleukins, insulin, interferons, enzymes and growth factors. To maximize the volumetric productivity of bacterial cultures it is important to grow *E. coli* to high cell concentration.

In order to optimize a real biotechnological production process, the model must describe those aspects of the process that significantly affect the process performance. The cost of development of mathematical models for bioprocess improvements is often too high and the benefits are too low. The main reason for this is related to the intrinsic complexity and non-linearity of biological systems.

Mathematical forms and their parameters used to describe cell behavior constitute the key problem of bioprocess modelling. The model building leads to information deficiency and to non unique parameter identification. While searching for new, more adequate modeling metaphors and concepts, methods which draw their initial inspiration from nature have received the early attention.

In this paper a Simulated Annealing (SA) algorithm is proposed to identify the unknown parameters in a non-linear mathematical model of a fed-batch cultivation process. SA as an optimization technique first was introduced to solve problems in discrete optimization, mainly combinatorial optimization. Subsequently, this technique has been successfully applied to solve optimization problems over the space of continuous decision variables.

SA is a local search method where the search mechanism is modelled on the Metropolis et al. [4] algo-

---

rithm and the principles of thermodynamic annealing. Kirkpatrick et al. [3] and Cerny [2] were the first to follow such a technique to solve optimization problems. SA can deal with arbitrary systems and cost functions; statistically guarantees finding an optimal solution (SA has the ability to avoid getting stuck at local minima); guarantees a convergence upon running sufficiently large (infinite) number of iterations; is relatively easy to code, even for complex problems. This makes annealing an attractive option for optimization problems

This paper is organized as follows. Outline of the introduced SA algorithm is described in Section 2. In Section 3 the problem formulation is presented. In Section 4 a discussion of the obtained numerical results of *E. coli* cultivation process model parameter identification is presented. Concluding remarks are made in Section 5.

II. OUTLINE OF THE SIMULATED ANNEALING ALGORITHM

SA is a stochastic relaxation technique [3]. SA is named by analogy to the annealing of solids, in which a crystalline solid is heated to its melting point and then allowed to cool gradually until it is again in the solid phase at some nominal temperature. At the absolute zero final temperature, the resulting solid achieves its most regular crystal configuration - corresponding to a (global) minimal value of the system's energy.

The SA algorithm performs the following steps:

The algorithm generates a random trial point. Initial temperature is 100. The SA chooses the distance of the trial point from the current point by a probability distribution with a scale depending on the current temperature. The trial point distance distribution could be set as a function that generates a point based on the current point and the current temperature using different distributions. Here Boltzman distribution is used:

$$Pr\left\{E\right\} = \exp\left\{E\left(i\right)/kT\right\}/Z\left(T\right) \quad (1)$$

where E is the system energy; T – the system temperature; k – Boltzman's constant ($k = 1.380650 \times 10^{-23}m^2kgs^{-2}k^{-1}$); $Z(T)$ – a normalization function of the form:

$$Z\left(T\right) = \sum_{h} \exp\left\{E\left(h\right)/kT\right\} \quad (2)$$

The algorithm determines whether the new point is better or worse than the current point. If the new point is better than the current point, it becomes the next point. If the new point is worse than the current point, the algorithm can still make it the next point. The algorithm accepts a worse point based on an acceptance function. The probability of acceptance is:

$$\frac{1}{1 + \exp\left(\Delta/\max\left(T\right)\right)} \quad (3)$$

where Δ = new objective — old objective, T — current temperature. Since both Δ and T are positive, the probability of acceptance is between 0 and 1/2. Smaller temperature leads to smaller acceptance probability. Also, larger Δ leads to smaller acceptance probability.

The algorithm systematically lowers the temperature, storing the best point found so far. The function that the algorithm uses to update the temperature is:

$$T = T_0 0.95^r \quad (4)$$

where r denotes the annealing parameter.

Reannealing sets the annealing parameters to lower values than the iteration number, thus raising the temperature in each dimension. The annealing parameters depend on the values of estimated gradients of the objective function in each dimension:

$$r_i = \log\left(\frac{T_0}{T_i} \frac{\max_{j}\left(s_j\right)}{s_i}\right) \quad (5)$$

where r_i is the annealing parameter for component i, T_0 — the initial temperature of component i, T_i — current temperature of the component i, s_i — gradient of the objective in direction i times difference of bounds in direction i.

The algorithm stops when the average change in the objective function is sufficiently small with respect to the predefined tolerance.

The SA algorithm can be described by the scheme:

Find initial solution (by generating it randomly)
Set initial value for the parameter $T = T_0$
Set a value for r, the rate of cooling parameter
    $j = 0$
    Generate a new solution S'
    Calculate the difference in cost:
    $\Delta = cost(S') - cost(S)$
    Examine the new solution and decide: accept or reject
    If accepted, it becomes the current solution; otherwise, keep the old one;
    $j = j + 1$
    Reduce the T and generate a new solution
Until some stopping criterion is applied

III. PROBLEM FORMULATION

The important part of model building is the choice of a certain optimization procedure for parameter estimation, so that to calibrate the model with a given set of experimental data in order to reproduce the experimental results in the best possible way. The estimation of model parameters with high parameter accuracy is essential for successful model development.

A. E. coli MC4110 fed-batch cultivation model

The mathematical model of the considered process can be represented by [5]:

$$\frac{dX}{dt} = \mu_{max} \frac{S}{k_S + S} X - \frac{F_{in}}{V} X \qquad (6)$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}} \mu_{max} \frac{S}{k_S + S} X + \frac{F_{in}}{V}(S_{in} - S) \qquad (7)$$

$$\frac{dA}{dt} = \frac{1}{Y_{A/X}} \mu_{max} \frac{A}{k_A + A} X - \frac{F_{in}}{V} A \qquad (8)$$

$$\frac{dpO_2}{dt} = -\frac{1}{Y_{pO_2/X}} \mu_{max} \frac{pO_2}{k_{pO_2} + pO_2} X + \\ k_L a(pO_2^* - pO_2) - \frac{F_{in}}{V} pO_2 \qquad (9)$$

$$\frac{dV}{dt} = F_{in} \qquad (10)$$

where: X is biomass concentration, [g/l]; S - substrate concentration, [g/l]; A - acetate concentration, [g/l]; pO_2 - dissolved oxygen concentration, [%]; pO_2^* - saturation concentration of dissolved oxygen, [%]; F_{in} - feeding rate, [l/h]; V - bioreactor volume, [l]; S_{in} - substrate concentration in the feeding solution, [g/l]; μ_{max} - maximum value of the specific growth rate, $[h^{-1}]$; k_i - saturation constants; $k_L a$ - volumetric oxygen transfer coefficient, $[h^{-1}]$; $Y_{i/X}$ - yield coefficients, [-].

For the parameter estimation problem real experimental data of the *E. coli* MC4110 fed-batch cultivation process are used. The cultivation experiments are performed in the Institute of Technical Chemistry, University of Hannover, Germany during the collaboration work with the Institute of Biophysics and Biomedical Engineering, BAS, Bulgaria, granted by DFG. The cultivation conditions and the experimental data are presented in details in [1], [6].

B. Optimization Criterion

The optimization criterion is a certain factor, which value defines the quality of an estimated set of parameters. The objective function is presented as a minimization of a distance measure J between experimental and model predicted values, represented by the vector **y**:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{m} \{[\mathbf{y}_{exp}(i) - \mathbf{y}_{mod}(i)]_j\}^2 \to min \qquad (11)$$

where n is the number of data for each state variable m; $\mathbf{y}_{exp}$ - the experimental data; $\mathbf{y}_{mod}$ - model predictions with a given set of the parameters.

IV. NUMERICAL RESULTS AND DISCUSSION

A series of parameter identification procedures for the considered model Eq. (6) - (10), using SA is performed. The computer specifications to run all optimization procedures are Intel Core i5-2320 CPU @ 3.00GHz, 8 GB Memory (RAM), Windows 7 (64bit) operating system, Matlab 7.5 environment.

A. Tuning of SA Algorithm Parameters

Each algorithm has its own influential parameters that affect its performance in terms of solution quality and computational time. In order to increase the performance of the SA it is necessary to provide the adjustments of the parameters depending on the problem domain. Parameters of the FA are tuned on the basis of a large number of pre-tests according to the parameter identification problem, considered here. The algorithm performance is tested varying the following function and parameters:

1) Annealing function AF
   a) generates a point based on the current point and the current temperature using Student's distribution (A_{fast});
   b) generates a point based on the current point and the current temperature using multivariate normal distribution (A_{boltz}).

2) Temperature function TF
   a) uses fast annealing by updating the current temperature based on the initial temperature and the current annealing parameter k (T_{fast});
   b) "*temperatureexp*" - uses exponential annealing schedule by updating the current temperature based on the initial temperature and the current annealing parameter k (T_{exp});

TABLE I
OBJECTIVE VALUES OBTAINED FOR DIFFERENT AF AND TF AT $IT = 100$ AND $RI = 100$

		AF	
		A_{fast}	A_{boltz}
TF	T_{fast}	6.6348	7.6207
	T_{exp}	6.3744	2.0137
	T_{boltz}	12.3012	9.3365

TABLE II
NUMERICAL RESULTS FROM PARAMETER IDENTIFICATION USING SIMULATED ANNEALING

Parameters	Average	Best	Worst
μ_{max}	0.47	0.46	0.45
k_S	0.01	0.012	0.02
k_A	0.56	0.59	0.51
k_{pO_2}	0.02	0.023	0.021
$Y_{S/X}$	0.5	0.497	0.52
$Y_{A/X}$	0.14	0.13	0.16
$Y_{pO_2/X}$	0.2	0.201	0.22
$k_L a$	57.71	55.87	50.23
J value	6.2661	6.1989	6.4022
execution time, s	88.723	80.954	91.298

c) "*temperatureboltz*" - implements Boltzman annealing by updating the current temperature based on the initial temperature and the current annealing parameter k (T_{boltz}).

3) Reannealing interval: $RI \in [1, 100]$.
4) Initial temperature: $IT \in [0.001, 100]$.

It is started with $IT = 100$ and $RI = 100$. The two AFs and the three TFs are tested for considered IT and RI. The obtained main objective functions J are presented in Table I.

On the basis of the obtained results functions A_{fast} for AF and T_{fast} for TF are chosen as more appropriate. Using these functions the IT and RI are varied in the considered ranges. First, the IT is decreased to 50. The resulting main J is 6.4575. Then the RI is decreased to 50. At $RI = 50$ J is 6.4262 and at both IT and RI set to 50 - $J = 6.3801$. Further decreasing of RI do not improve considerably the objective function - for $RI = 40$ mean J is 6.3798 and for $RI = 10$ mean J is 6.3857. After that further decreasing of IT is considered. For the values TI of 10, 1, 0.5, 0.1 and 0.001 the resulting main J are as follows: 7.4391, 6.1989, 6.2442, 6.4040 and 6.8768. The lowest mean value of J is obtained at $IT = 1$ and $RI = 50$.

After tuning procedures the main SA parameters are set to the following optimal settings:
$AF - A_{fast}$; $TF - T_{fast}$; $IT = 1$ and $RI = 50$.

B. Results from Parameter Optimization Procedure

Because of the stochastic characteristics of the applied algorithm a series of 30 runs is performed. The best, the worst and the average results of the 30 runs, for the J value and execution time are observed. The obtained results from parameter identification procedures are presented in Table II.

The resulting model parameters values are in admissible range according to [8], [9], [10]. In Fig. 1 some graphical results about algorithms performance (parameters estimation through generations) are presented.

A quantitative measure of the differences between modelled and measured values is important criterion
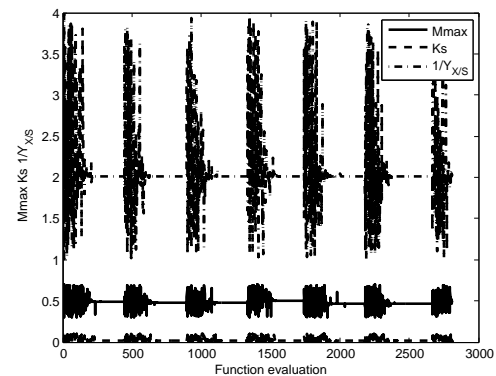

Fig. 1. Parameters estimation through generations

for the adequacy of a model. The model predictions of the state variables are compared to the experimental data points of the real *E. coli* MC4110 cultivation. The graphical results of the comparison between the model predictions of state variables, based on SA estimations, and the experimental data points of the real *E. coli* cultivation are presented in the Figs. 2–4. Model predicted data are presented with solid line.

The presented graphics show a very good correlation between the experimental and predicted data. The model predicts successfully the process variables dynamics during the fed-batch cultivation of *E. coli* MC4110. However, graphical comparisons can clearly show only the existence or absence of systematic deviations between model predictions and measurements. It is evident that a quantitative measure of the differences between calculated and measured values is an important criterion for the adequacy of a model. The most important criterion for the valuation of models is that the deviations between measurements and model calculations (J) should be as small as possible.

The obtained criteria values show that the developed model is adequate and has a high degree of accuracy. As

a result of the identification procedure accurate model parameters estimations are obtained. The presented results are a confirmation of successful application of the SA algorithm and of the appropriate choice of SA algorithm parameters.
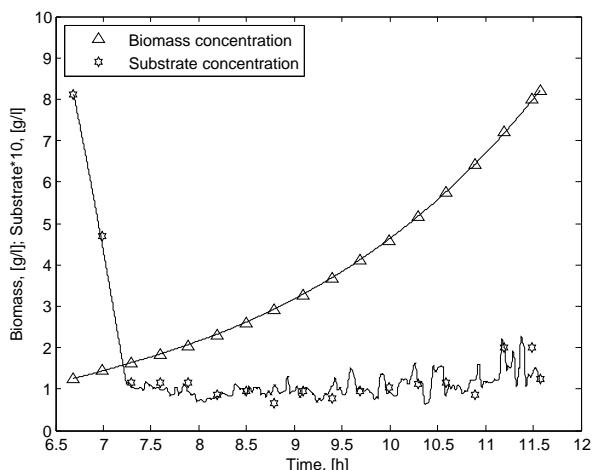

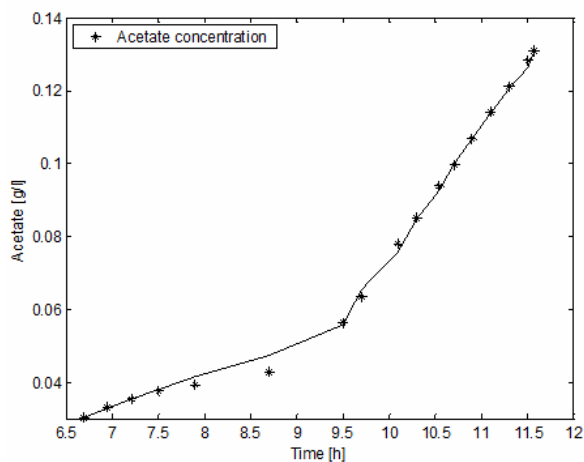Fig. 2. Biomass and substrate concentration - modelled and real experimental data

Fig. 3. Acetate concentration - modelled and real experimental data

V. CONCLUSION

In this paper the metaheuristic SA is applied for parameter identification of non-lineal model of cultivation process. Real *E. coli* MC4110 fed-batch cultivation process is considered. The mathematical model is presented by a system of ordinary differential equations, describing the regarded cultivation process variables. Particular procedure for model parameter identification is performed using SA. Numerical and simulation results reveal that correct and consistent results are obtained.
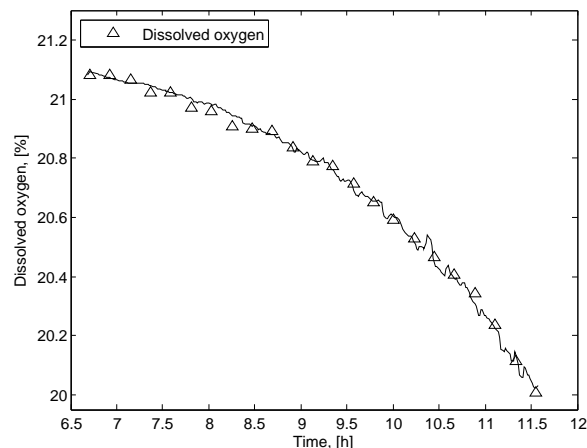

Fig. 4. Dissolved oxygen - modelled and real experimental data

Resulting non-linear model predicts adequately and to a high degree of approximation the variation of the considered state variables. Simulation results reveal that accurate and consistent estimates can be obtained using SA algorithm.

REFERENCES

[1] M. Arndt and B. Hitzmann, "Feed Forward/Feedback Control of Glucose Concentration during Cultivation of *Escherichia coli*", 8th IFAC Int. Conf. on Comp. Appl. in Biotechn., Canada, 2001, pp. 425–429.

[2] V. Cerny, "Thermodynamical Approach to the Travelling Salesman Problem: An Efficient Simulation Algorithm", Journal of Optimization Theory and Applications, vol. 45, 1985, pp. 41–51.
http://dx.doi.org/10.1007/BF00940812

[3] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Optimization by Simulated Annealing*, Science, New Series, vol. 220(4598), 1983, pp. 671–680.

[4] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and M. Teller, "Equation of State Calculations by Fast Computing Machines", Journal of Chemical Physics, vol. 21, 1953, pp. 1087–1092.
http://dx.doi.org/10.1063/1.1699114

[5] O. Roeva, "Parameter Estimation of a Monod-type Model based on Genetic Algorithms and Sensitivity Analysis", LNCS, Springer, vol. 4818, 2008, pp. 601–608.

[6] O. Roeva, T. Pencheva, B. Hitzmann and St. Tzonkov, "A Genetic Algorithms Based Approach for Identification of *Escherichia coli* Fed-batch Fermentation", Int J Bioautomation, vol. 1, 2004, pp. 30–41.

[7] U. Viesturs, D. Karklina , I. Ciprovica, "Bioprocess and Bio-engineering", Jeglava, 2004.

[8] O. Georgieva, M. Arndt and B. Hitzmann, "Modelling of *Escherichia coli* Fed-Batch Fermentation", In International Symposium "Bioprocess Systems'2001 — BioPS'2001", 1–3.X, Sofia, Bulgaria, 2001, pp. I.61–I.64.

[9] D. Levisauskas, V. Galvanauskas, S. Henrich, K. Wilhelm, N. Volk and A. Lubbert, "Model-based Optimization of Viral Capsid Protein Production in Fed-batch Culture of recombinant *Escherichia coli*", Bioprocess and Biosystems Engineering, vol. 25, 2003, pp. 255–262.

[10] B. Zelic, D. Vasic-Racki, C. Wandrey and R. Takors, "Modeling of the Pyruvate Production with *Escherichia coli* in a Fed-batch Bioreactor", Bioprocess and Biosystems Engineering, vol. 26, 2004, pp. 249–258.
http://dx.doi.org/10.1007/s00449-004-0358-0